OXFORD

JNCI J Natl Cancer Inst (2020) 112(10): djz246

doi: 10.1093/jnci/djz246
First published online April 5, 2020
Article

A Transcriptome-Wide Association Study Identifies Novel Candidate Susceptibility Genes for Pancreatic Cancer

Jun Zhong, PhD,[1] Ashley Jermusyk, PhD,[1] Lang Wu, PhD,[2] Jason W. Hoskins, PhD,[1] Irene Collins, PhD,[1] Evelina Mocci, PhD,[3] Mingfeng Zhang, MD, PhD,[1,4] Lei Song, MS,[5] Charles C. Chung, PhD,[5] Tongwu Zhang, PhD,[5] Wenming Xiao, PhD,[6,7] Demetrius Albanes, MD,[5] Gabriella Andreotti, PhD, MPH,[5] Alan A. Arslan, MD,[8,9,10] Ana Babic, PhD,[11] William R. Bamlet, MS,[12] Laura Beane-Freeman, PhD,[5] Sonja Berndt, PharmD, PhD,[5] Ayelet Borgida, MS,[13] Paige M. Bracci, PhD, MPH,[14] Lauren Brais, MPH,[11] Paul Brennan, PhD,[15] Bas Bueno-de-Mesquita, MD, MPH, PhD,[16,17,18,19] Julie Buring, PhD,[20,21] Federico Canzian, PhD,[22] Erica J. Childs, PhD,[3] Michelle Cotterchio, PhD, MPH, MS,[23,24] Mengmeng Du, PhD,[25] Eric J. Duell, PhD,[26] Charles Fuchs, MD, PhD,[27] Steven Gallinger, PhD,[13] J. Michael Gaziano, MD, PhD,[20,28,29] Graham G. Giles, PhD,[30,31,32] Edward Giovannucci, MD, PhD,[11] Michael Goggins, MD,[33] Gary E. Goodman, MD,[34] Phyllis J. Goodman, MS,[35] Christopher Haiman, PhD,[36] Patricia Hartge, PhD,[5] Manal Hasan, MD, MPH, PhD,[37] Kathy J. Helzlsouer, MD, MHS,[38] Elizabeth A. Holly, PhD, MPH,[39] Eric A. Klein, MD,[40] Manolis Kogevinas, PhD,[41,42,43,44] Robert J. Kurtz, MD,[45] Loic LeMarchand, MD, PhD,[46] Núria Malats, MD, PhD,[47] Satu Männistö, PhD,[48] Roger Milne, PhD,[30,31,49] Rachel E. Neale, PhD,[50] Kimmie Ng, MD, MPH,[11] Ofure Obazee, PhD,[22] Ann L. Oberg, PhD,[12] Irene Orlow, PhD, MS,[25] Alpa V. Patel, PhD,[51] Ulrike Peters, PhD, MPH,[34] Miquel Porta, MD, MPH, PhD,[42,43] Nathaniel Rothman, MD, MPH, MHS,[5] Ghislaine Scelo, PhD,[15,30,31] Howard D. Sesso, PhD, MPH,[20,21] Gianluca Severi, PhD,[52] Sabina Sieri, PhD,[53] Debra Silverman, PhD,[5] Malin Sund, MD, PhD,[54] Anne Tjønneland, MD, PhD, DMSc,[55,56,57] Mark D. Thornquist, PhD,[34] Geoffrey S. Tobias, BS,[5] Antonia Trichopoulou, MD, PhD, Stephen K. Van Den Eeden, PhD,[58] Kala Visvanathan, MD, MHS,[59] Jean Wactawski-Wende, PhD,[60] Nicolas Wentzensen, MD, PhD,[5] Emily White, PhD, MS,[34,61] Herbert Yu, MD, PhD,[46] Chen Yuan, PhD,[11] Anne Zeleniuch-Jacquotte, MD,[9,62] Robert Hoover, MD, PhD,[5] Kevin Brown, PhD,[1] Charles Kooperberg, PhD,[34] Harvey A. Risch, MD, PhD,[63] Eric J. Jacobs, PhD,[64] Donghui Li, PhD,[65] Kai Yu, PhD,[5] Xiao-Ou Shu, MD, PhD, MPH,[2] Stephen J. Chanock, MD,[5] Brian M. Wolpin, MD, MPH,[11] Rachael Z. Stolzenberg-Solomon, PhD, MPH, RD,[5] Nilanjan Chatterjee, PhD,[5,66] Alison P. Klein, PhD, MHS,[3,33] Jill P. Smith, MD,[67] Peter Kraft, PhD,[21,68] Jianxin Shi, PhD,[5] Gloria M. Petersen, PhD,[12] Wei Zheng, MD, PhD, MPH,[2] Laufey T. Amundadottir, PhD[1,*]

[1]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [2]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA; [3]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, USA; [4]US Food and Drug Administration, Silver Spring, MD, USA; [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [6]National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA; [7]Division of Molecular Genetics and Pathology, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA; [8]Department of Obstetrics and Gynecology, New York University School of Medicine, New York, NY, USA; [9]Department of Population Health, New York University School of Medicine, New York, NY, USA; [10]Department of Environmental Medicine, New York University School of Medicine, New York, NY, USA; [11]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; [12]Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, USA; [13]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada; [14]Department of Epidemiology and Biostatistics, University of California, CA, USA; [15]International Agency for Research on Cancer, Lyon, France; [16]Department for Determinants of Chronic Diseases, National Institute for Public Health and the Environment, BA, Bilthoven, The Netherlands; [17]Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands; [18]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; [19]Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; [20]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA; [21]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA;

Received: March 20, 2019; **Revised:** September 12, 2019; **Accepted:** December 30, 2019

[22]Genomic Epidemiology Group, German Cancer Research Center, Heidelberg, Germany; [23]Cancer Care Ontario, University of Toronto, Toronto, Ontario, Canada; [24]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [25]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA; [26]Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Bellvitge Biomedical Research Institute, Catalan Institute of Oncology, Barcelona, Spain; [27]Yale Cancer Center, New Haven, CT, USA; [28]Division of Aging, Brigham and Women's Hospital, Boston, MA, USA; [29]Boston VA Healthcare System, Boston, MA, USA; [30]Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, VIC, Australia; [31]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia; [32]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia; [33]Department of Pathology, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins School of Medicine, Baltimore, MD, USA; [34]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [35]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [36]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [37]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [38]Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [39]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA; [40]Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA; [41]ISGlobal, Centre for Research in Environmental Epidemiology, Barcelona, Spain; [42]CIBER Epidemiología y Salud Pública, Barcelona, Spain; [43]Hospital del Mar Institute of Medical Research, Universitat Autònoma de Barcelona, Barcelona, Spain; [44]Universitat Pompeu Fabra, Barcelona, Spain; [45]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA; [46]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA; [47]Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Center, Madrid, Spain; [48]Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland; [49]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, VIC, Australia; [50]Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, Australia; [51]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA; [52]Centre de Recherche en Épidémiologie et Santé des Populations (CESP, Inserm U1018), Facultés de Medicine, Université Paris-Saclay, UPS, UVSQ, Gustave Roussy, Villejuif, France; [53]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; [54]Department of Surgical and Perioperative Sciences, Umeå University, Umeå, Sweden; [55]Danish Cancer Society Research Center, Copenhagen, Denmark; [56]Department of Public Health, University of Copenhagen, Copenhagen, Denmark; [57]Hellenic Health Foundation, Athens, Greece; [58]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA; [59]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [60]Department of Epidemiology and Environmental Health, University at Buffalo, Buffalo, NY, USA; [61]Department of Epidemiology, University of Washington, Seattle, WA, USA; [62]Perlmutter Cancer Center, New York University School of Medicine, New York, NY, USA; [63]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA; [64]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA; [65]Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; [66]Department of Biostatistics, Bloomberg School of Public Health, Baltimore, MD, USA; [67]Department of Medicine, Georgetown University, Washington, DC, USA; and [68]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

*Correspondence to: Laufey T. Amundadottir, PhD, Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Gaithersburg, MD 20877, USA (e-mail: amundadottirl@mail.nih.gov).

Abstract

Background: Although 20 pancreatic cancer susceptibility loci have been identified through genome-wide association studies in individuals of European ancestry, much of its heritability remains unexplained and the genes responsible largely unknown. **Methods:** To discover novel pancreatic cancer risk loci and possible causal genes, we performed a pancreatic cancer transcriptome-wide association study in Europeans using three approaches: FUSION, MetaXcan, and Summary-MulTiXcan. We integrated genome-wide association studies summary statistics from 9040 pancreatic cancer cases and 12 496 controls, with gene expression prediction models built using transcriptome data from histologically normal pancreatic tissue samples (NCI Laboratory of Translational Genomics [n = 95] and Genotype-Tissue Expression v7 [n = 174] datasets) and data from 48 different tissues (Genotype-Tissue Expression v7, n = 74–421 samples). **Results:** We identified 25 genes whose genetically predicted expression was statistically significantly associated with pancreatic cancer risk (false discovery rate < .05), including 14 candidate genes at 11 novel loci (1p36.12: *CELA3B*; 9q31.1: *SMC2*, *SMC2-AS1*; 10q23.31: *RP11-80H5.9*; 12q13.13: *SMUG1*; 14q32.33: *BTBD6*; 15q23: *HEXA*; 15q26.1: *RCCD1*; 17q12: *PNMT*, *CDK12*, *PGAP3*; 17q22: *SUPT4H1*; 18q11.22: *RP11-888D10.3*; and 19p13.11: *PGPEP1*) and 11 at six known risk loci (5p15.33: *TERT*, *CLPTM1L*, *ZDHHC11B*; 7p14.1: *INHBA*; 9q34.2: *ABO*; 13q12.2: *PDX1*; 13q22.1: KLF5; and 16q23.1: *WDR59*, *CFDP1*, *BCAR1*, *TMEM170A*). The association for 12 of these genes (*CELA3B*, *SMC2*, and *PNMT* at novel risk loci and *TERT*, *CLPTM1L*, *INHBA*, *ABO*, *PDX1*, *KLF5*, *WDR59*, *CFDP1*, and *BCAR1* at known loci) remained statistically significant after Bonferroni correction. **Conclusions:** By integrating gene expression and genotype data, we identified novel pancreatic cancer risk loci and candidate functional genes that warrant further investigation.

ARTICLE

Pancreatic cancer is the third leading cause of cancer deaths in the United States (1) and seventh worldwide (2). Established risk factors include tobacco smoking, long-standing diabetes, obesity, heavy alcohol consumption, chronic pancreatitis, and family history of pancreatic cancer (3). Inherited rare mutations in hereditary cancer and pancreatitis genes, identified in families with a high incidence of disease, account for a small percentage of cases (4). At the other end of the spectrum, common risk variants with low penetrance have been discovered through genome-wide association studies (GWAS) (5–11). However, these loci explain a small fraction of genetic heritability for pancreatic cancer, and the genes underlying the associations at most of these are unknown (11–14).

Most susceptibility alleles discovered through GWAS reside in noncoding regions of the genome and likely function through allele-specific regulation of gene expression (15). A transcriptome-wide association study (TWAS) builds on this premise by imputing genetically predicted gene expression levels into GWAS datasets to discover genes whose *cis*-regulated expression is associated with complex traits (16–18). This approach has been applied to several common diseases, including melanoma, breast, prostate, and ovarian cancers (19–24). In this comprehensive TWAS for pancreatic cancer, we leveraged two expression quantitative trait loci (eQTL) datasets generated from histologically normal pancreatic tissue samples from individuals of European ancestry (25,26), with GWAS summary statistics [Pancreatic Cancer Cohort Consortium (PanScan) I–III and Pancreatic Cancer Case-Control Consortium (PanC4) (6–11)] to identify genes associated with risk of pancreatic cancer.

Methods

Transcriptome and Genotype Datasets

Two histologically normal-derived pancreatic expression datasets, the National Cancer Institute's Laboratory of Translational Genomics (LTG) (25) and the Genotype-Tissue Expression (GTEx, v7) (27), were used. Only samples with more than 80% European ancestry were included (LTG, n = 95; GTEx, n = 174). Alignment of RNA-seq reads from the LTG data (Illumina HiSeq 2000, phs001776.v1.p1) was performed using STAR v2.4.2a (28) based on GENCODE v19 gene annotations (GRCh37/hg19). For GTEx, gene expression read counts (Illumina HiSeq 2000/2500) for pancreatic tissue samples were obtained through controlled access (phs000424.v7.p2). The LTG and GTEx pancreatic transcriptome datasets were also combined for genes expressed in both datasets (see Supplementary Methods, available online).

Blood- or normal tissue–derived DNA samples (LTG dataset) were genotyped on Illumina OmniExpress or Omni1M arrays (25). After quality control, genotypes were imputed using the 1000 Genomes imputation reference dataset via the Michigan Imputation Server (29). Genotypes for GTEx samples were obtained via dbGaP (phs000424.v7.p2). Principal components were calculated for genotype data using SNPRelate (30). Gene expression values were adjusted for five principal components, probabilistic estimation of expression residuals factors (31), and gender.

Building Pancreatic Tissue Gene Expression Prediction Models

Expression prediction models were computed in FUSION (17) using variants ±500 kb of each gene. Genes with nominally significant *cis*–single-nucleotide polymorphism (SNP)-heritability (likelihood ratio test $P < .05$) and cross-validation ($R^2 > 0.01$) were used to train TWAS prediction models with a fivefold cross-validation. Prediction models were also computed using MetaXcan (16) for variants ±1 Mb of each gene. The model for each gene was implemented in the glmnet R package, with a ridge-lasso mixing parameter ($\alpha = 0.5$) and a penalty parameter lambda chosen through 10-fold cross-validation (32). Model performance was compared across the three expression datasets (LTG, GTEx, and LTG + GTEx) and two TWAS methods (FUSION and MetaXcan) showing good correlation (Supplementary Figure 1, available online). For cross-tissue TWAS, we used gene expression prediction models for 48 different human tissues from PredictDB (http://predictdb.org/) (16). These models were trained using GTEx (v7) data for European participants only, using PrediXcan (see Supplementary Methods, available online).

TWAS Association Analysis

The pancreatic cancer GWAS summary statistics included 9040 pancreatic ductal adenocarcinoma (PDAC) cases and 12 496 controls of European ancestry from PanScan I–III and PanC4 (11). Using FUSION and MetaXcan, associations between predicted expression and pancreatic cancer risk were estimated based on gene prediction model weights, GWAS summary statistics, and a SNP-correlation (linkage disequilibrium) matrix (17,22). A false discovery rate (FDR) corrected *P* value threshold of less than .05 was used for each analysis. Bonferroni correction for multiple testing was also used based on the number of tests in each analysis (Supplementary Figure 2, available online). Finally, we used Summary-MulTiXcan (SMulTiXcan) (33) to test associations between predicted gene expression levels and pancreatic cancer risk with cross-tissue models. Quantile-quantile plots are shown in Supplementary Figure 3 (available online).

We assessed statistical power by simulating gene expression and GWAS summary statistics using data from PanScan I–III and PanC4 (11). Parameters included the number of causal SNPs for gene expression in the *cis* region, the fraction of gene expression variance explained by causal SNPs, and the fraction of phenotypic variance explained by gene expression. We varied H^2 (h^2_g, 0.1, 0.3, and 0.5), causal SNPs (1, 1%, and 10%), and R^2 (h^2_{ge}, 0 to 0.001) and recomputed each configuration 100 times to assess how often the TWAS and GWAS tests were statistically significant (TWAS $P < 2.27 \times 10^{-6}$ [.05/22k]; GWAS $P < 5 \times 10^{-8}$) (Supplementary Figure 4 and Methods, available online).

Transcriptome Differences and Pathway Analyses for TWAS-Identified Genes

Transcriptome changes associated with high and low expression of genes identified by TWAS in the LTG and GTEx pancreatic datasets were assessed by comparing gene expression for samples in the bottom quartile with those in the top quartile of expression for each gene using EdgeR (34,35). Genes differentially expressed at FDR less than .05 and fold-change greater than twofold ($|\log FC| > 1$) were included in pathway analyses using DAVID (36,37) to identify enrichment in Gene Ontology (GO) biological processes, GO molecular functions, and Kyoto Encyclopedia of Genes and Genomes pathways (Supplementary Methods, available online).

Statistical Analyses

Expression prediction models (LASSO, Elastic Net, BLUP, BSLMM) were selected for genes with nominally significant SNP-heritability (*cis*-h^2_g LRT $P < .05$) and cross-validation ($R^2 > 0.01$). Logistic and linear (LASSO) models were used in GWAS and TWAS simulations for power estimates. TWAS *P* values were determined from calculated TWAS z scores and adjusted at FDR less than .05. For increased stringency, nominal TWAS *P* values were also compared with a Bonferroni corrected α threshold. Independence of SNPs and predicted expression effects on pancreatic cancer risk were tested by conditional (joint) tests using GWAS and TWAS summary statistics. Differential expression analyses used an empirical Bayes method (EdgeR) to estimate gene-level biological variation; exact test *P* values were corrected for multiple testing by FDR less than .05. All statistical tests were two-sided.

Results

Gene Expression Prediction Model Building

We performed TWAS by integrating pancreatic-specific and cross-tissue gene expression prediction models with results from a recent meta-analysis of pancreatic cancer GWAS data, performed within the PanScan and the PanC4, including 9040 PDAC cases and 12 496 controls (8,11). Two pancreatic transcriptome datasets from histologically normal pancreatic tissue samples were used: LTG (n = 95) (10) and GTEx v7 (n = 174) (27). Both the GWAS and gene expression datasets included only individuals of European ancestry. Two complementary TWAS approaches, FUSION (17) and MetaXcan (16,32), were used to

build robust gene expression prediction models (see Supplementary Methods and Table 1, available online). We first assessed the power of the TWAS as compared with GWAS by simulating causal SNP-expression-trait models to identify genome-wide significant signals. We found that TWAS substantially increased statistical power as compared with GWAS, particularly when multiple causal SNPs underlie signals (Supplementary Figure 4, available online).

After comparing gene prediction models (prediction performance $R^2 \geq 0.01$) in the three datasets (LTG, GTEx, and combined LTG + GTEx), we found that each had distinct and valuable characteristics for TWAS analysis. First, whereas some gene expression prediction models were common to all three datasets (FUSION: n = 1687; MetaXcan: n = 1408), a larger number was unique to one of these (FUSION: n = 658 885, and 1421 models for LTG, GTEx, and LTG + GTEx, respectively; MetaXcan: n = 648 975, and 1705 models for LTG, GTEx, and LTG + GTEx, respectively) (Supplementary Figure 2, available online). Second, a greater number of gene prediction models were observed for the combined LTG + GTEx dataset (FUSION: n = 5902; MetaXcan: n = 5775) as compared with the individual datasets (2440–4992 models for LTG and GTEx using FUSION and MetaXcan) (Supplementary Figure 2, available online). Third, among gene prediction models common to the three datasets, the number of models with improved performance in the combined LTG + GTEx dataset (n = 826–1283) was greater than those with poorer performance (n = 342–738) as compared with the individual LTG or GTEx datasets (Supplementary Figure 5, available online). Fourth, although gene prediction model performance was highly correlated between pancreatic tissue datasets (Pearson $r = 0.60–0.93$) and TWAS approaches (Pearson $r = 0.87–0.98$) (Supplementary Figure 1, available online), a substantial number of gene prediction models had improved performance in FUSION (n = 5730) or MetaXcan (n = 4267) (Supplementary Figure 6, available online). Finally, although both the LTG and GTEx datasets were derived from histologically normal pancreatic tissue samples, the former was generated mostly from samples adjacent to tumors, whereas the latter was generated using nondiseased tissues from rapid autopsy programs. Based on these factors, we performed the analysis using each of the three transcriptome datasets and the two TWAS methods.

Because a large proportion of *cis*-regulated gene expression is shared across multiple tissues (38,39), we also took advantage of publicly available gene expression models generated from 48 different tissues (n = 2043–21 422 models per tissue, n = 74–421 samples per tissue type; http://predictdb.org/) from 608 individuals of European ancestry (GTEx v7) (27) to discover additional pancreatic cancer susceptibility genes using PrediXcan (16). The quantile-quantile plots showed little evidence for inflation of the test statistics as compared with the expected distribution ($\lambda_{1000} = 1.004–1.025$) (Supplementary Figure 3, available online).

Association Analyses Between Genetically Predicted Gene Expression and PDAC Risk

We evaluated associations between genetically predicted gene expression and pancreatic cancer risk by an integrated analysis using FUSION (17), MetaXcan (16,32), and SMulTiXcan (33) (Figure 1; Tables 1 and 2). First, using FUSION and the LTG pancreas gene expression models (n = 2827), we found that genetically predicted expression of *ABO*, *CFDP1*, *PNMT*, *RCCD1*, and *PGAP3* was associated with PDAC risk (TWAS: $P < 9.11 \times 10^{-5}$, FDR < .05); in the GTEx (v7) pancreas gene expression models (n = 4992), we identified six additional PDAC risk-associated genes: *KLF5*, *SUPT4H1*, *BTBD6*, *CDK12*, *SMUG1*, and *CELA3B* (TWAS: $P < 6.96 \times 10^{-5}$, FDR < .05); in the combined LTG + GTEx pancreas dataset (n = 5902), we identified three additional genes: *SMC2*, *WDR59*, and *HEXA* (TWAS: $P < 6.52 \times 10^{-5}$, FDR < .05). Of these genes, *CELA3B*, *SMC2*, *ABO*, *KLF5*, *WDR59*, *CFDP1*, and *PNMT* were associated after Bonferroni correction (TWAS: $P < 1.77 \times 10^{-5}$ for LTG, $P < 1.00 \times 10^{-5}$ for GTEx, and $P < 8.47 \times 10^{-6}$ for LTG + GTEx).

Second, using MetaXcan and the LTG pancreas gene expression models (n = 2440), we observed predicted *ABO* expression to be associated with PDAC risk (TWAS: $P < 8.07 \times 10^{-27}$, FDR < .05); in the GTEx (v7) pancreas gene expression models (n = 4763), we identified six additional PDAC risk-associated genes: *PDX1*, *INHBA*, *CELA3B*, *PGAP3*, *SUPT4H1*, and *RP11-888D10.3* (TWAS: $P < 6.96 \times 10^{-5}$, FDR < .05); in the gene expression models (n = 5775) trained by the combined pancreas dataset, we identified two additional genes: *SMC2* and *PGPEP1* (TWAS: $P < 3.67 \times 10^{-5}$, FDR < .05). Of these genes, *CELA3B*, *INHBA*, *ABO*, and *PDX1* were associated after Bonferroni correction (TWAS: $P < 2.05 \times 10^{-5}$ for LTG, $P < 1.05 \times 10^{-5}$ for GTEx, and $P < 8.65 \times 10^{-6}$ for LTG + GTEx).

Finally, using SMulTiXcan (33) and gene expression models (n = 2043–21 422) trained in 48 different tissues (GTEx v7), we observed associations for predicted *SMC2*, *INHBA*, *PDX1*, and *CFDP1* expression and identified seven additional genes associated with PDAC risk: *TERT*, *CLPTM1L*, *BCAR1*, *ZDHHC11B*, *RP11-80H5.9*, *TMEM170A*, and *SMC2-AS1* (TWAS: $P < 1.39 \times 10^{-5}$, FDR < .05). Of these, *TERT*, *CLPTM1L*, *PDX1*, *CFDP1*, and *BCAR1* were statistically significant after Bonferroni correction (TWAS: $P < 2.33 \times 10^{-6}$).

Overall, we discovered 25 genes (Figure 1) whose genetically predicted gene expression was associated with PDAC risk (FDR < .05), including 14 genes at 11 novel loci—1p36.12 (*CELA3B*), 9q31.1 (*SMC2*, *SMC2-AS1*), 10q23.31 (*RP11-80H5.9*), 12q13.13 (*SMUG1*), 14q32.33 (*BTBD6*), 15q23 (*HEXA*), 15q26.1 (*RCCD1*), 17q12 (*PNMT*, *CDK12*, *PGAP3*), 17q22 (*SUPT4H1*), 18.q11.22 (*RP11-888D10.3*), and 19p13.11 (*PGPEP1*) (Table 1)—and 11 genes at six known risk loci (8,11)—5p15.33 (*TERT*, *CLPTM1L*, *ZDHHC11B*), 7p14.1 (*INHBA*), 9q34.2 (*ABO*), 13q12.2 (*PDX1*), 13q22.1 (*KLF5*), and 16q23.1 (*WDR59*, *CFDP1*, *BCAR1*, *TMEM170A*) (Table 2). Three TWAS genes identified at novel loci (*CELA3B*, *SMC2*, *PNMT*) and nine at previously reported GWAS loci (*TERT*, *CLPTM1L*, *INHBA*, *ABO*, *PDX1*, *KLF5*, *WDR59*, *CFDP1*, *BCAR1*) were statistically significant after Bonferroni correction (Tables 1 and 2). Genes showing positive and negative effects in different tissues are listed in Table 1 and Supplemental Figure 7 (available online).

We performed conditional tests at two loci containing more than one TWAS gene using pancreatic tissue models to determine if they represented conditionally independent signals. At chr17q12, three adjacent genes (Table 1; Figure 2A) were identified by TWAS: *PNMT*, *CDK12*, and *PGAP3*. The TWAS signal for *PNMT* and *PGAP3* dropped substantially after conditioning the analysis on predicted *CDK12* expression in the GTEx pancreas dataset (*PNMT* TWAS *P* value changed from 5.10×10^{-4} to .53 and *PGAP3* TWAS *P* value from 6.96×10^{-5} to .09). The GWAS signal at this locus also dropped markedly after conditioning on predicted expression of *CDK12* (Figure 2A) indicating that *CDK12* may explain a large part of the signal at this locus. The gene expression correlation for the three genes was low (*CDK12* and *PNMT*, Pearson $r = 0.09$ in both LTG and GTEx) to moderate (*PNMT* and *PGAP3*, Pearson $r = 0.33$ and $r = 0.27$; *CDK12* and *PGAP3*, Pearson $r = 0.66$ and $r = 0.29$ in the LTG and GTEx

ARTICLE

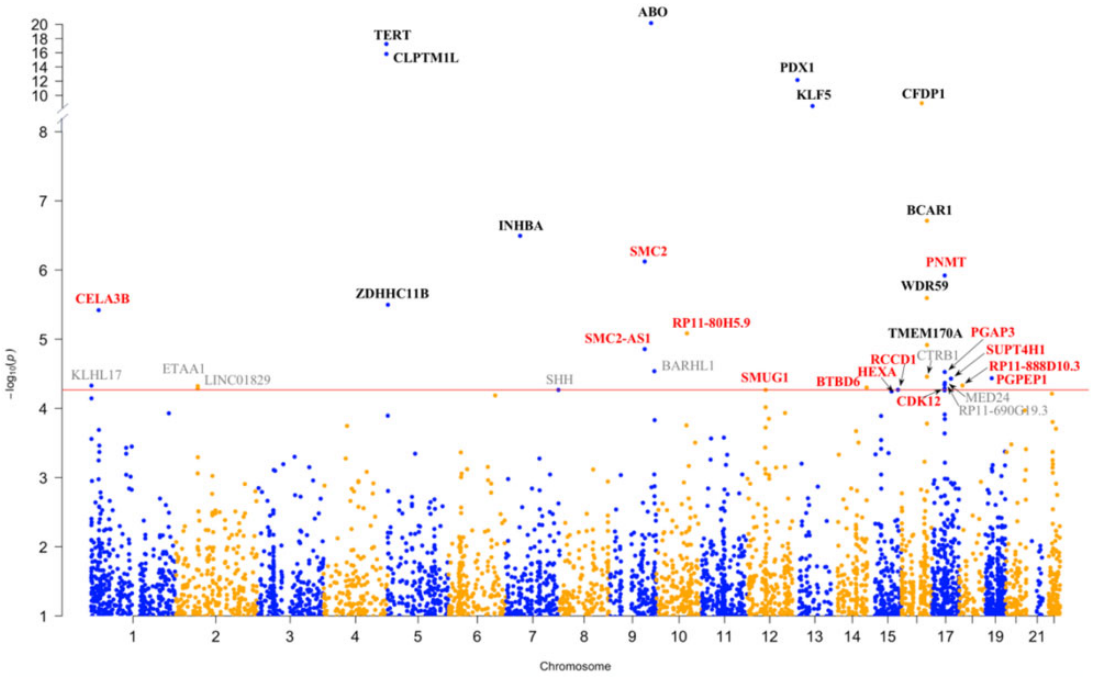

Figure 1. Manhattan plot of the results from the pancreatic cancer transcriptome-wide association study (TWAS). Each point corresponds to an association test between genetically predicted gene expression for a specific gene and pancreatic ductal adenocarcinoma risk. Genes listed in **red** are located at novel genomic loci, and those in **black** are known pancreatic cancer risk loci. Genes listed in **gray** did not pass the threshold for multiple testing (false discovery rate < 0.05) in the independent TWAS analyses. Genes with TWAS $P \leq 9.11 \times 10^{-5}$ in at least one analysis are annotated.

pancreas datasets, respectively) (Supplementary Table 2, available online). In contrast, the association with PDAC risk for two adjacent genes at chr16q23.1 (Table 2; Figure 2B), *WDR59* and *CFDP1*, appeared largely independent (TWAS *P* value changed from 2.54×10^{-6} to 5.60×10^{-4} for *WDR59* and from 8.47×10^{-9} to 1.70×10^{-6} for *CFDP1* in a joint analysis in the combined LTG + GTEx pancreatic dataset). The GWAS signal at this locus dropped dramatically after conditioning on predicted expression of *WDR59* and *CFDP1*, indicating that genetically predicted expression of *WDR59* and *CFDP1* together explain most of the signal at this locus (Figure 2B). The expression of these two genes was moderately to strongly correlated in the two datasets (Pearson $r = 0.52$–0.80) (Supplementary Table 2, available online).

To determine whether the associations between predicted gene expression and PDAC risk were independent of the lead GWAS-identified variants at each locus, we performed conditional analyses adjusting for the most statistically significant risk variants within ±1 Mb of TWAS-identified genes in the PanScan and PanC4 GWAS datasets. Among the 25 TWAS-identified genes, the association for three genes at novel loci (*PNMT*, *CDK12*, and *PGAP3*) and four genes at known loci (*TERT*, *CLPTM1L*, *ZDHHC11B*, and *KLF5*) remained statistically significant at the Bonferroni corrected *P* value threshold ($P < .05/25$ genes, ie, $P < .002$, Tables 1 and 2), indicating that these genes may be associated with PDAC risk independently of the GWAS-identified lead risk variants. Interestingly, at chr5p15.33, substantial loss of the TWAS signals for both *TERT* and *CLPTM1L* was seen after conditioning on three of the four GWAS-identified variants that mark independent pancreatic cancer risk signals at this locus (Table 3; Supplementary Table 3, available online) indicating that the underlying biology at this locus may involve both genes.

Transcriptome Changes Associated With High and Low Expression of Genes Identified by TWAS

To begin unravelling the potential consequences associated with different expression levels of TWAS-identified genes, we assessed transcriptome differences in samples in the top vs bottom quartiles of expression for each gene in the GTEx and LTG pancreatic datasets (see Supplementary Methods and Tables 4 and 5, available online) as previously described (40). As this analysis may be most relevant for well-expressed genes that are highly differentially expressed between samples in the top vs bottom quartile of expression, we focused on *CELA3B*, which was highly expressed and with a large difference in median expression (GTEx = 76%; LTG = 91%) in samples in the top and bottom quartiles (Supplementary Table 6 and Figure 8, available online). Pathway enrichment analyses for genes differentially expressed in samples at the top vs bottom quartile of *CELA3B* gene expression showed a negative correlation between expression of *CELA3B* and inflammatory and immune response genes (Table 4) indicating that low *CELA3B* expression may be associated with an inflammatory state in the pancreas.

Discussion

To identify novel susceptibility loci and putative causally relevant genes for pancreatic cancer development, we integrated eQTL datasets derived from pancreatic, as well as other tissues, with the largest currently available pancreatic cancer GWAS dataset (11) and identified 25 genes whose genetically predicted expression associated with pancreatic cancer risk. These genes localize to 17 genomic regions, of which 11 do not overlap with known PDAC risk loci.

ARTICLE

Table 1. Statistically significant expression–trait associations for genes at loci not previously identified by pancreatic cancer GWAS

Region	Gene name	Lead GWAS variant (±1 Mb)[†]	GWAS P[†]	Approach	Training tissue	$R^{2,‡}$	TWAS Z[§]	TWAS P[¶]	TWAS P after conditioning on lead GWAS variant[#]
1p36.12	*CELA3B**	rs61132601	2.27×10^{-7}	FUSION	GTEx pancreas	0.06	−3.98	6.89×10^{-5}	.08
				FUSION	Combined pancreas	0.04	−4.62	3.80×10^{-6}*	.03
				MetaXcan	GTEx pancreas	0.05	−4.43	9.38×10^{-6}*	.21
				MetaXcan	Combined pancreas	0.05	−4.29	1.83×10^{-5}	.03
9q31.1	*SMC2**	rs147699343	8.77×10^{-8}	FUSION	Combined pancreas	0.04	4.95	7.52×10^{-7}*	.08
				MetaXcan	Combined pancreas	0.02	4.93	8.19×10^{-7}*	.06
				SMulTiXcan	Cross-tissue	0.02–0.61	−3.34 to 5.35	8.50×10^{-6}	.66
9q31.1	*SMC2-AS1*	rs147699343	8.70×10^{-8}	SMulTiXcan	Cross-tissue	0.04-0.18	−4.9 to 4.8	1.39×10^{-5}	.61
10q23.31	*RP11-80H5.9*	rs7083351	5.22×10^{-5}	SMulTiXcan	Cross-tissue	0.02–0.20	−2.21 to 4.4	8.23×10^{-6}	.04
12q13.13	*SMUG1*	rs4759336	1.39×10^{-4}	FUSION	GTEx pancreas	0.28	−4.04	5.40×10^{-5}	.06
14q32.33	*BTBD6*	rs10638535	2.73×10^{-5}	FUSION	GTEx pancreas,	0.07	4.06	4.98×10^{-5}	.94
				FUSION	Combined pancreas	0.05	4.00	6.30×10^{-5}	.73
15q23	*HEXA*	rs11636684	2.35×10^{-5}	FUSION	Combined pancreas	0.02	−4.02	5.68×10^{-5}	5.31×10^{-3}
15q26.1	*RCCD1*	rs8028409	3.77×10^{-5}	FUSION	LTG pancreas	0.44	−3.98	6.94×10^{-5}	.87
				FUSION	GTEx pancreas	0.28	−4.04	5.38×10^{-5}	.86
				FUSION	Combined pancreas	0.37	−3.99	6.52×10^{-5}	.95
17q12	*PNMT**	rs12951693	6.17×10^{-7}	FUSION	LTG pancreas	0.02	4.86	1.20×10^{-6}*	4.01×10^{-5}
17q12	*CDK12*	rs12951693	6.17×10^{-7}	FUSION	GTEx pancreas	0.02	−4.05	5.15×10^{-5}	1.37×10^{-3}
17q12	*PGAP3*	rs12951693	6.17×10^{-7}	FUSION	LTG pancreas	0.10	3.91	9.11×10^{-5}	1.44×10^{-3}
				FUSION	GTEx pancreas	0.25	3.98	6.96×10^{-5}	2.16×10^{-4}
				MetaXcan	GTEx pancreas	0.24	4.11	3.03×10^{-5}	1.03×10^{-4}
				MetaXcan	Combined pancreas	0.18	4.17	2.98×10^{-5}	1.04×10^{-4}
17q22	*SUPT4H1*	rs6503868	2.15×10^{-5}	FUSION	GTEx pancreas	0.08	4.12	3.72×10^{-5}	3.32×10^{-3}
				MetaXcan	GTEx pancreas	0.07	4.11	3.90×10^{-5}	6.50×10^{-3}
18.q11.22	*RP11-888D10.3*	rs28637808	1.30×10^{-5}	MetaXcan	GTEx pancreas	0.09	−4.07	4.67×10^{-5}	.07
19p13.11	*PGPEP1*	rs12985909	3.48×10^{-5}	MetaXcan	Combined pancreas	0.06	−4.13	3.67×10^{-5}	.85

*Genes and corresponding TWAS P that are statistically significant after Bonferroni correction for multiple testing in each of the analyses. GTEx = Genotype-Tissue Expression; GWAS = genome-wide association studies; LTG = Laboratory of Translational Genomics; SMulTiXcan = Summary-MulTiXcan; TWAS = transcriptome-wide association study.

[†]The lead GWAS variant and GWAS P value indicates the most statistically significant GWAS variant within ± 1 Mb for each gene listed.

[‡]R^2: model prediction performance.

[§]TWAS Z: effect size and direction. Effect sizes for SMulTiXcan results in individual tissues are shown in Supplementary Figure 7 (available online).

[¶]TWAS P: P value from the TWAS for genes that passed the false discovery rate corrected P value ≤ 0.05 in each of the analyses.

[#]TWAS P values after conditioning on the lead GWAS variant within ± 1 Mb for each gene is shown in the last column.

ARTICLE

Several TWAS genes identified at novel loci function in DNA repair, chromosome organization, and cell division. *SMC2* (9q31.1) encodes structural maintenance of chromosomes protein 2, a core component of the condensin complex, which regulates chromosome organization during mitosis and meiosis and plays a critical role in single-strand break DNA repair (41–43). *SMUG1* (12q13.13) encodes a base excision repair enzyme (single-strand-selective monofunctional uracil-DNA glycosylase 1) that repairs several DNA-pyrimidine oxidation products, some of which are mutagenic (44). *RCCD1* (15q26.1) encodes RCC1 domain-containing protein 1, a partner of histone H3K36 demethylase KDM8; this complex is important for spindle organization, chromosome segregation, and accurate mitotic division (45). *CDK12* (cyclin-dependent kinase 12, 17q12) belongs to the cyclin-dependent kinase (CDK) family of serine and threonine protein kinases that regulate transcriptional and posttranscriptional processes, including DNA damage response, splicing, pre-mRNA processing, development, and differentiation (46,47). *CDK12* is mutated in some tumors and overexpressed in others, indicating that it may have context-dependent oncogenic and tumor suppressor functions (46). Decreased genetically predicted expression of *SMUG1*, *RCCD1*, and *CDK12* was associated with increased risk of pancreatic cancer, in agreement with their roles in maintaining genome stability. Conversely, increased *SMC2* expression was associated with pancreatic cancer risk, which is less consistent with its role in cell division and DNA repair but aligns with reports showing that its expression is regulated by the transcription factors β-catenin-TCF4 and that it is important for WNT-mediated cell proliferation in intestinal cells (48).

At chr1p36, another locus not previously reported in GWAS, genetically predicted *CELA3B* expression was inversely associated with risk of pancreatic cancer. This gene encodes chymotrypsin-like elastase family member 3B and, along with other pancreatic serine proteases, has a digestive function (49). Pathway enrichment analysis indicated that low expression of *CELA3B* may be associated with an inflammatory state in the pancreas, which is interesting in the light that inflammatory conditions, including pancreatitis, increase risk of pancreatic cancer (3).

Chr5p15.33 is a well-known multicancer risk locus with multiple independent signals reported in the *TERT-CLPTM1L* gene region for more than 10 different cancers, including pancreatic cancer (5,10,12,50–53). *TERT* encodes the catalytic subunit of the telomerase reverse transcriptase complex, whose major function is to maintain the ends of our chromosomes and overall chromosomal integrity (54–58). The *CLPTM1L* gene, relatively unknown until a few years ago, is now known to encode a multipass transmembrane protein that promotes growth and is frequently overexpressed in pancreatic and lung cancers (59–61). It

Table 2. Statistically significant expression–trait associations for genes at known pancreatic cancer risk loci previously identified by GWAS

Region	Gene name	Lead GWAS Variant (± 1 Mb)†	GWAS P†	Approach	Training tissue	R^{2},‡	TWAS Z§	TWAS P¶	TWAS P after conditioning on lead GWAS variant#
5p15.33	TERT*	rs31490	1.28×10^{-17}	SMulTiXcan	Cross-tissue	0.05–0.11	−8.24 to 4.20	5.80×10^{-18}*	3.37×10^{-4}
5p15.33	CLPTM1L*	rs31490	1.28×10^{-17}	SMulTiXcan	Cross-tissue	0.02–0.06	−8.33 to 0.91	1.48×10^{-16}*	6.10×10^{-4}
5p15.33	ZDHHC11B	rs31490	1.28×10^{-17}	SMulTiXcan	Cross-tissue	0.02–0.19	−1.16 to 3.13	3.18×10^{-6}	1.56×10^{-5}
7p14.1	INHBA*	rs12701838	3.59×10^{-09}	MetaXcan	GTEx pancreas	0.04	−5.11	3.20×10^{-7}*	.81
				SMulTiXcan	Cross-tissue	0.04–0.34	−5.11 to −0.72	4.10×10^{-6}	.02
9q34.2	ABO*	rs687621	2.37×10^{-27}	FUSION	LTG pancreas	0.37	9.38	6.71×10^{-21}*	.49
				FUSION	GTEx pancreas	0.56	6.96	3.44×10^{-12}*	.21
				FUSION	Combined pancreas	0.50	7.55	4.34×10^{-14}*	.23
				MetaXcan	LTG pancreas	0.30	10.72	8.07×10^{-27}*	.98
				MetaXcan	GTEx pancreas	0.55	7.08	1.41×10^{-12}*	.08
				MetaXcan	Combined pancreas	0.49	7.65	2.05×10^{-14}*	.07
13q12.2	PDX1*	rs2297316	4.43×10^{-13}	MetaXcan	GTEx pancreas	0.05	−7.18	6.85×10^{-13}*	.64
				SMulTiXcan	Cross-tissues	0.03–0.05	−7.18 to −6.59	4.87×10^{-12}*	.45
13q22.1	KLF5*	rs9573166	1.51×10^{-25}	FUSION	GTEx pancreas	0.05	4.91	9.17×10^{-7}*	4.91×10^{-4}
				FUSION	Combined pancreas	0.03	5.92	3.15×10^{-9}*	.02
16q23.1	WDR59*	rs72802357	1.32×10^{-16}	FUSION	Combined pancreas	0.01	−4.70	2.54×10^{-6}*	3.42×10^{-3}
16q23.1	CFDP1*	rs72802357	1.32×10^{-16}	FUSION	LTG pancreas	0.03	6.07	1.26×10^{-9}*	.06
				FUSION	Combined pancreas	0.12	5.76	8.47×10^{-9}*	.03
				MetaXcan	Combined pancreas	0.17	5.58	2.40×10^{-8}*	.05
				SMulTiXcan	Cross-tissue	0.03–0.20	2.50 to 6.89	2.02×10^{-8}*	.12
16q23.1	BCAR1*	rs72802357	1.32×10^{-16}	SMulTiXcan	Cross-tissue	0.02–0.21	−5.60 to 6.49	1.94×10^{-7}*	.22
16q23.1	TMEM170A	rs72802357	1.32×10^{-16}	SMulTiXcan	Cross-tissue	0.02–0.22	−3.69 to 2.86	1.21×10^{-5}	.41

*Genes and corresponding TWAS *P* values that are statistically significant after Bonferroni correction for multiple testing in each of the analyses. GTEx = Genotype-Tissue Expression; GWAS = genome-wide association studies; LTG = Laboratory of Translational Genomics; SMulTiXcan = Summary-MulTiXcan; TWAS = transcriptome-wide association study.
†The lead GWAS variant and GWAS *P* value indicates the most statistically significant GWAS variant within ±1 Mb for each gene listed.
‡R^{2}: model prediction performance.
§TWAS Z: effect size and direction. Effect sizes for SMulTiXcan results in individual tissues are shown in Supplementary Figure 7 (available online).
¶TWAS *P*: *P* value from the TWAS for genes that passed the false discovery rate corrected *P* value ≤ .05 in each of the analyses.
#TWAS *P* values after conditioning on the lead GWAS variant within ± 1 Mb for each gene is shown in the last column.

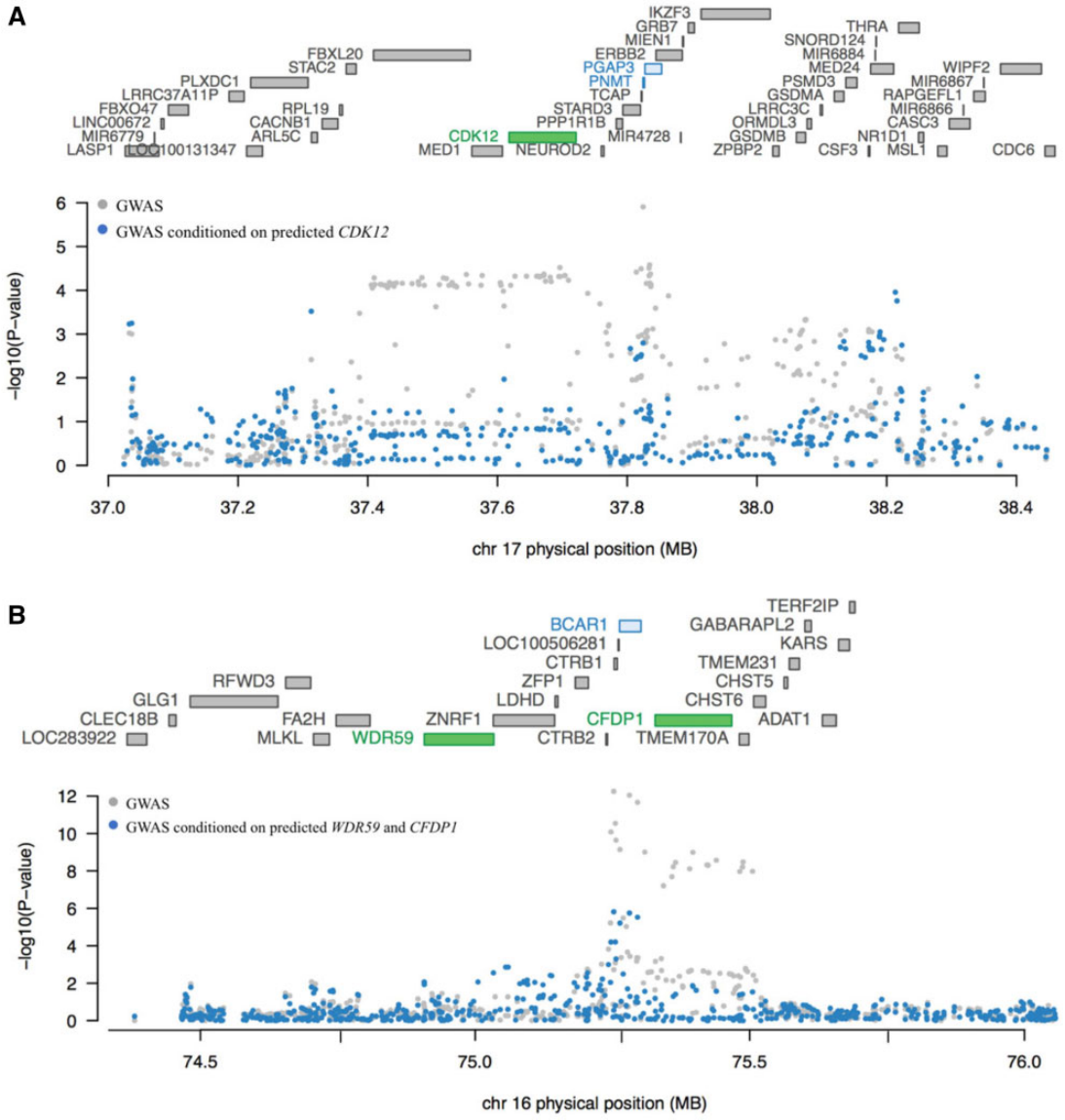

Figure 2. Conditional transcriptome-wide association study (TWAS) analysis for loci showing marginally statistically significant associations with multiple genes. **A)** Genes at chr17q12 are listed according to chromosomal location in the upper part of the plot with TWAS statistically significant genes shown in **green** (*CDK12*). The Manhattan plot for the genome-wide association studies (GWAS) results is shown in the lower part of the figure with pancreatic cancer GWAS *P* values before (**gray**) and after (**blue**) conditioning the analysis on imputed gene expression levels for *CDK12*. Genetically predicted expression of *CDK12* appears to explain a large part of the GWAS-identified signal at chr17q12 in the Genotype-Tissue Expression (GTEx) dataset. **B)** Similarly, genes at chr16q23 are listed according to chromosomal location in the upper part of the plot with TWAS statistically significant genes shown in **green** (*WDR59* and *CFDP1*). The Manhattan plot for GWAS results is shown in the lower part of the figure with pancreatic cancer GWAS *P* values before (**gray**) and after (**blue**) conditioning the analysis on imputed gene expression levels for *WDR59* and *CFDP1*. Genetically predicted expression of *WDR59* and *CFDP1* appears to explain the majority of the GWAS-identified signal at chr16q23.1. Marginal and joint analyses were performed in the GTEx pancreas dataset for chr17q12 and in the combined Laboratory of Translational Genomics + GTEx pancreas dataset for chr16q23.1.

Table 3. TWAS results on chr5p15.33 before and after conditional analyses for SNPs that mark independent GWAS risk signals on chr5p15.33

	TWAS P*				
		GWAS conditioned on:			
Gene name		rs31490	rs2736098	rs36115365	rs35226131
TERT	5.80×10^{-18}	3.37×10^{-4}	7.48×10^{-9}	9.86×10^{-6}	3.88×10^{-16}
CLPTM1L	1.48×10^{-16}	6.10×10^{-4}	4.83×10^{-8}	4.85×10^{-5}	4.04×10^{-14}
ZDHHC11B	3.18×10^{-6}	1.56×10^{-5}	8.41×10^{-3}	3.00×10^{-2}	6.50×10^{-3}

*Transcriptome-wide association study (TWAS) *P* values are shown prior to and after conditioning the Pancreatic Cancer Cohort Consortium–Pancreatic Cancer Case-Control Consortium genome-wide association studies analysis on four independent pancreatic cancer risk signals at chr5p15.33 per Wang et al. (50), Petersen et al. (7), Wolpin et al. (8), Zhang et al. (10), and Fang et al. (12).

ARTICLE

Table 4. Pathway enrichment analysis for genes expressed at higher levels in samples with low vs high *CELA3B* expression in the GTEx and LTG transcriptome datasets

Pathways enrichment for genes inversely associated with *CELA3B* expression						
				GTEx	LTG	
Category	Term	DE genes, No.*	Fold enrichment†	P‡	Fold enrichment†	P‡
GO biological process	Inflammatory response	72	6.3	3.8×10^{-33}	3.2	6.0×10^{-55}
GO biological process	Immune response	70	5.6	1.2×10^{-28}	3	2.7×10^{-50}
KEGG	Staphylococcus aureus infection	24	11.5	7.9×10^{-17}	4.5	3.5×10^{-21}
GO biological process	Cell adhesion	57	4.1	1.4×10^{-16}	2.7	1.2×10^{-40}
GO biological process	Innate immune response	52	4	1.6×10^{-14}	2.4	8.6×10^{-25}

*Genes expressed at twofold or higher levels in samples in the bottom vs top quartile of *CELA3B* (chymotrypsin-like elastase 3B) mRNA expression in the GTEx pancreas and LTG histologically normal pancreatic transcriptome datasets were included in a pathway enrichment analysis using DAVID. GTEx = Genotype-Tissue Expression; LTG = Laboratory of Translational Genomics; DE = Differentially expressed genes; GO = Gene Ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes.
†Fold enrichment for these genes in the pathways listed are shown as well as ‡Benjamini-Hochberg false discovery rate corrected *P* values.

is important for endoplasmic reticulum stress, apoptosis and cytokinesis, and *KRAS*-driven lung cancer (61). Using cross-tissue prediction models, we identified both *TERT* and *CLPTM1L* as pancreatic cancer TWAS genes with positive and negative effects, depending on the tissues. This type of pleiotropy for chr5p15.33 has been previously described by us and others (5,10,12,50–53).

Some of the genes identified in our study have been reported in TWAS for breast (*RCCD1*, *KLF5*), ovarian cancer (*RCCD1*), and type 2 diabetes (*RCCD1*) (20,22,62,63). *KLF5* is located at chr13q22.1, a pancreatic cancer risk locus in a large, nongenic region flanked by *KLF5* and *KLF12* (13). It encodes Kruppel Like Factor 5, a zinc finger transcription factor frequently overexpressed in pancreatic cancer, and is important for *Kras* mediated pancreatic tumorigenesis in mouse models (64). Because we have previously shown that *DIS3*, a gene that encodes a catalytic subunit of the nuclear RNA exosome complex that mediates RNA processing and decay, represents a functional gene at chr13q22.1 (13), our current findings indicate that *KLF5* may also play a role at this risk locus. None of the suggestive genes (unadjusted $P < .05$) reported in a recent but much smaller TWAS for pancreatic cancer (65) overlap with the genes reported in our study. Three loci overlap with our recent pathway-based analysis of pancreatic cancer GWAS data (chr9q31.1: *SMC2*; chr15q23: *HEXA*; and chr17q12: *PNMT*, *CDK12*, and *PGAP3*) and are suggestive in the GWAS analysis (66).

Although TWAS represents an attractive method to map risk loci that influence gene expression, this approach has advantages and disadvantages. Benefits include the reduced multiple testing burden and nomination of plausible candidate risk genes. However, identification of trait-associated gene expression differences by TWAS does not imply causality, and functional studies are needed to comprehensively determine underlying mechanisms of risk. Furthermore, coregulated genes can present as multiple associated genes at the same locus, even though only one gene underlies the signal. Finally, only *cis*-eQTLs are assessed, and genes whose genetically regulated gene expression cannot be predicted using SNPs are not evaluated. In the future, larger transcriptome and GWAS datasets for pancreatic cancer are likely to further improve statistical power for gene identification using this approach. Likewise, transcriptome datasets from specific cellular subtypes within the pancreas, such as acinar and ductal cells, could also improve future pancreatic cancer TWAS approaches.

In summary, we report 25 genes whose genetically predicted expression was associated with pancreatic cancer risk (FDR < .05), including 14 genes at 11 novel genomic loci. Twelve of these genes remained statistically significant after Bonferroni correction. Our findings provide new insights into the genetic basis of pancreatic cancer risk and identify target genes for future functional studies to thoroughly explore the mechanistic underpinnings of risk at each locus.

Funding

This work was supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health (NIH).

Notes

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication. The authors have no conflicts of interest to disclose. Acknowledgements, data access, and additional funding information are listed in the Supplementary Material (available online).

The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
3. Stolzenberg-Solomon RZ, Amundadottir LT. Epidemiology and inherited predisposition for sporadic pancreatic adenocarcinoma. *Hematol Oncol Clin North Am*. 2015;29(4):619–640.
4. Petersen GM. Familial pancreatic cancer. *Semin Oncol*. 2016;43(5):548–553.
5. Amundadottir LT. Pancreatic cancer genetics. *Int J Biol Sci*. 2016;12(3):314–325.
6. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet*. 2009;41(9):986–990.
7. Petersen GM, Amundadottir L, Fuchs CS, 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(3):224–228.
8. Wolpin BM, Rizzato C, Kraft P, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet*. 2014;46(9):994–1000.

ARTICLE

9. Childs EJ, Mocci E, Campa D, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet*. 2015;47(8):911–916.
10. Zhang M, Wang Z, Obazee O, et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget*. 2016;7(41):66328–66343.
11. Klein AP, Wolpin BM, Risch HA, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun*. 2018;9(1):556.
12. Fang J, PanScan Consortium, Jia J, Makowski M, et al. Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. *Nat Commun*. 2017;8(1):15034.
13. Hoskins JW, Ibrahim A, Emmanuel MA, et al. Functional characterization of a chr13q22.1 pancreatic cancer risk locus reveals long-range interaction and allele-specific effects on DIS3 expression. *Hum Mol Genet*. 2016;25(21): 4726–4738.
14. Zheng J, Huang X, Tan W, et al. Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat Genet*. 2016;48(7):747–757.
15. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 2018;19(8):491–504.
16. Gamazon ER, GTEx Consortium, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091–1098.
17. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48(3):245–252.
18. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet*. 2017;18(2):117–127.
19. Zhang T, Choi J, Kovacs MA, et al. Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Res*. 2018;28(11):1621–1635.
20. Wu L, Shi W, Long J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. *Nat Genet*. 2018;50(7):968–978.
21. Mancuso N, Gayther S, Gusev A, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun*. 2018; 9(1):4079.
22. Lu Y, Beeghly-Fadiel A, Wu L, et al. A transcriptome-wide association study among 97,898 women to identify candidate susceptibility genes for epithelial ovarian cancer risk. *Cancer Res*. 2018;78(18):5419–5430.
23. Gusev A, Mancuso N, Won H, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet*. 2018;50(4):538–548.
24. Theriault S, Gaudreault N, Lamontagne M, et al. A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis. *Nat Commun*. 2018;9(1):988.
25. Zhang M, Lykke-Andersen S, Zhu B, et al. Characterising cis-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. *Gut*. 2018;67(3):521–533.
26. Li X, Kim Y, Tsang EK, et al. The impact of rare variation on gene expression across tissues. *Nature*. 2017;550(7675):239–243.
27. Consortium G. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–213.
28. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
29. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284–1287.
30. Zheng X, Levine D, Shen J, et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012; 28(24):3326–3328.
31. Stegle O, Parts L, Durbin R, et al. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol*. 2010;6(5):e1000770.
32. Barbeira AN, GTEx Consortium, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun*. 2018;9(1):1825.
33. Barbeira AN, Pividori MD, Zheng J, et al. Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet. 2019;15(1):e1007889.
34. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.
35. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40(10):4288–4297.
36. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1): 44–57.
37. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1–13.
38. The GTEx Consortium. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648–660.
39. Wheeler HE, Shah KP, Brenner J, et al. Survey of the heritability and sparse architecture of gene expression traits across human tissues. *PLoS Genet*. 2016; 12(11):e1006423.
40. Cobo I, Martinelli P, Flandez M, et al. Transcriptional regulation by NR5A2 links differentiation and inflammation in the pancreas. *Nature*. 2018; 554(7693):533–537.
41. Takemoto A, Kimura K, Yokoyama S, et al. Cell cycle-dependent phosphorylation, nuclear localization, and activation of human condensin. *J Biol Chem*. 2004;279(6):4551–4559.
42. Schmiesing JA, Gregson HC, Zhou S, et al. A human condensin complex containing hCAP-C-hCAP-E and CNAP1, a homolog of Xenopus XCAP-D2, colocalizes with phosphorylated histone H3 during the early stage of mitotic chromosome condensation. *Mol Cell Biol*. 2000;20(18):6996–7006.
43. Kong X, Stephens J, Ball AR Jr, et al. Condensin I recruitment to base damage-enriched DNA lesions is modulated by PARP1. *PLoS One*. 2011;6(8):e23548.
44. Wood RD, Mitchell M, Sgouros J, et al. Human DNA repair genes. *Science*. 2001; 291(5507):1284–1289.
45. Marcon E, Ni Z, Pu S, et al. Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation. *Cell Rep*. 2014;8(1):297–310.
46. Paculova H, Kohoutek J. The emerging roles of CDK12 in tumorigenesis. *Cell Div*. 2017;12:7. doi: 10.1186/s13008-017-0033-x.
47. Dubbury SJ, Boutz PL, Sharp PA. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature*. 2018;564(7734):141–145.
48. Davalos V, Suarez-Lopez L, Castano J, et al. Human SMC2 protein, a core subunit of human condensin complex, is a novel transcriptional target of the WNT signaling pathway and a new therapeutic target. *J Biol Chem*. 2012;287(52): 43472–43481.
49. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733–D745.
50. Wang Z, Zhu B, Zhang M, et al. Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet*. 2014; 23(24):6616–6633.
51. Mocellin S, Verdi D, Pooley KA, et al. Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. *J Natl Cancer Inst*. 2012;104(11):840–854.
52. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*. 2013;45(4):371–384, 384e1–384e2.
53. Kote-Jarai Z, Saunders EJ, Leongamornlert DA, et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. *Hum Mol Genet*. 2013;22(12):2520–2528.
54. Armanios M, Blackburn EH. The telomere syndromes. *Nat Rev Genet*. 2012; 13(10):693–704.
55. Janknecht R. On the road to immortality: HTERT upregulation in cancer cells. *FEBS Lett*. 2004;564(1-2):9–13.
56. Cheung AL, Deng W. Telomere dysfunction, genome instability and cancer. *Front Biosci*. 2008;13(13):2075–2090.
57. Kim NW, Piatyszek MA, Prowse KR, et al. Specific association of human telomerase activity with immortal cells and cancer. *Science*. 1994;266(5193): 2011–2015.
58. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. *Eur J Cancer*. 1997;33(5):787–791.
59. Jia J, Bosley AD, Thompson A, et al. CLPTM1L promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res*. 2014;74(10):2785–2795.
60. Clarke WR, Amundadottir L, James MA. CLPTM1L/CRR9 ectodomain interaction with GRP78 at the cell surface signals for survival and chemoresistance upon ER stress in pancreatic adenocarcinoma cells. *Int J Cancer*. 2019;144(6): 1367–1378.
61. James MA, Vikis HG, Tate E, et al. CRR9/CLPTM1L regulates cell survival signaling and is required for RAS transformation and lung tumorigenesis. *Cancer Res*. 2014;74(4):1116–1127.
62. Hoffman JD, Graff RE, Emami NC, et al. Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. *PLoS Genet*. 2017;13(3):e1006690.
63. Torres JM, Barbeira AN, Bonazzola R, et al. Integrative cross tissue analysis of gene expression identifies 2 novel type 2 diabetes genes. BioRxiv 2017. doi: 10.1101/108134.
64. He P, Yang JW, Yang VW, et al. Kruppel-like factor 5, increased in pancreatic ductal adenocarcinoma, promotes proliferation, acinar-to-ductal metaplasia, pancreatic intraepithelial neoplasia, and tumor growth in mice. *Gastroenterology*. 2018;154(5):1494–1508 e13.
65. Gong L, Zhang D, Lei Y, et al. Transcriptome-wide association study identifies multiple genes and pathways associated with pancreatic cancer. *Cancer Med*. 2018;7(11):5727–5732.
66. Walsh N, Zhang H, Hyland PL, et al. Agnostic pathway/gene set analysis of genome-wide association data identifies associations for pancreatic cancer. *J Natl Cancer Inst*. 2019;111(6):557–567.