

Associations between Genetically Predicted Blood Protein Biomarkers and Pancreatic Cancer Risk



Jingjing Zhu¹, Xiang Shu², Xingyi Guo², Duo Liu^{1,3}, Jiandong Bao², Roger L. Milne^{4,5,6}, Graham G. Giles^{4,5,6}, Chong Wu⁷, Mengmeng Du⁸, Emily White^{9,10}, Harvey A. Risch¹¹, Nuria Malats¹², Eric J. Duell¹³, Phyllis J. Goodman¹⁴, Donghui Li¹⁵, Paige Bracci¹⁶, Verena Katzke¹⁷, Rachel E. Neale¹⁸, Steven Gallinger¹⁹, Stephen K. Van Den Eeden²⁰, Alan A. Arslan²¹, Federico Canzian²², Charles Kooperberg⁹, Laura E. Beane Freeman²³, Ghislaine Scelo²⁴, Kala Visvanathan²⁵, Christopher A. Haiman²⁶, Loïc Le Marchand¹, Herbert Yu¹, Gloria M. Petersen²⁷, Rachael Stolzenberg-Solomon²³, Alison P. Klein^{25,28}, Qiuyin Cai², Jirong Long², Xiao-Ou Shu², Wei Zheng², and Lang Wu¹

ABSTRACT

Background: Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal malignancies, with few known risk factors and biomarkers. Several blood protein biomarkers have been linked to PDAC in previous studies, but these studies have assessed only a limited number of biomarkers, usually in small samples. In this study, we evaluated associations of circulating protein levels and PDAC risk using genetic instruments.

Methods: To identify novel circulating protein biomarkers of PDAC, we studied 8,280 cases and 6,728 controls of European descent from the Pancreatic Cancer Cohort Consortium and the Pancreatic Cancer Case-Control Consortium, using genetic instruments of protein quantitative trait loci.

Results: We observed associations between predicted concentrations of 38 proteins and PDAC risk at an FDR of < 0.05, including 23

of those proteins that showed an association even after Bonferroni correction. These include the protein encoded by *ABO*, which has been implicated as a potential target gene of PDAC risk variant. Eight of the identified proteins (*LMA2L*, *TM11D*, *IP-10*, *ADH1B*, *STOM*, *TENCI*, *DOCK9*, and *CRBB2*) were associated with PDAC risk after adjusting for previously reported PDAC risk variants (OR ranged from 0.79 to 1.52). Pathway enrichment analysis showed that the encoding genes for implicated proteins were significantly enriched in cancer-related pathways, such as *STAT3* and *IL15* production.

Conclusions: We identified 38 candidates of protein biomarkers for PDAC risk.

Impact: This study identifies novel protein biomarker candidates for PDAC, which if validated by additional studies, may contribute to the etiologic understanding of PDAC development.

Introduction

Pancreatic cancer, 95% of which is pancreatic ductal adenocarcinoma (PDAC), is the second most commonly diagnosed gastrointestinal malignancy and the third leading cause of cancer-related death in

the United States (1). With a 5-year survival of 8%, the incidence of pancreatic cancer keeps increasing in the United States (2). Because pancreatic cancer is typically asymptomatic in early stages, most patients are diagnosed at an advanced stage, which precludes the

¹Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, Hawaii. ²Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, Tennessee. ³Department of Pharmacy, Harbin Medical University Cancer Hospital, Harbin, China. ⁴Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia. ⁵Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia. ⁶Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia. ⁷Department of Statistics, Florida State University, Tallahassee, Florida. ⁸Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. ⁹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington. ¹⁰Department of Epidemiology, University of Washington, Seattle, Washington. ¹¹Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut. ¹²Spanish National Cancer Research Centre (CNIO) and CIBERONC, Madrid, Spain. ¹³Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO-IDIBELL), L'Hospitalet de Llobregat, Spain. ¹⁴SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington. ¹⁵Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas. ¹⁶Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California. ¹⁷Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁸Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, Queensland,

Australia. ¹⁹Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. ²⁰Division of Research, Kaiser Permanente Northern California, Oakland, California. ²¹Department of Obstetrics and Gynecology, New York University School of Medicine, New York, New York. ²²Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²³Division of Cancer Epidemiology and Genetics, NCI, NIH, Rockville, Maryland. ²⁴Genetic Epidemiology Group, Section of Genetics, International Agency for Research on Cancer, World Health Organization, Lyon, France. ²⁵Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland. ²⁶Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California. ²⁷Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota. ²⁸Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, Maryland.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Lang Wu, University of Hawaii at Manoa, 701 Ilalo Street, Building B, Room 520, Honolulu, HI 96813. Phone: 808-564-5965; Fax: 808-586-2982; E-mail: lwu@cc.hawaii.edu

Cancer Epidemiol Biomarkers Prev 2020;29:1501-8

doi: 10.1158/1055-9965.EPI-20-0091

©2020 American Association for Cancer Research.

possible application of curative surgery. Therefore, identifying biomarkers that would contribute to screening or early diagnosis in high-risk populations may improve pancreatic cancer outcomes. Serum CA 19-9 is currently the only biomarker for pancreatic cancer used in clinical settings. However, it is mainly used for diagnosing symptomatic patients, and monitoring disease prognosis and response to treatment (3). Besides CA 19-9, several other blood circulating proteins have been reported to be potentially associated with pancreatic cancer risk, such as CA242, PIVKA-II, PAM4, S100A6, OPN, RBM6, EphA2, and OPG (4–7), but the results in those studies are inconsistent. For example, those studies often only involved a small sample size and evaluated a few candidate proteins, and were often limited by a lack of external validation. In addition, due to the observational study design, they were potentially subject to selection bias and residual and unmeasured confounding.

Mendelian randomization (MR) analysis is a widely applied design using genetic variants as instruments to evaluate the potential causal relationship between exposure and outcome (8–12). The nature of random assortment of alleles from parents to offspring during gamete formation makes such a design using genetic instruments to be less susceptible to biases encountered by conventional epidemiologic studies (8, 13).

In this study, we aimed to use genetic variants as an instrument to study blood concentrations of proteins and to assess their associations with PDAC risk. Genome-wide association studies (GWAS) have identified hundreds of protein quantitative trait loci (pQTL; refs. 14, 15), many of which can serve as strong instrumental variables. To our knowledge, this is the first large-scale study to comprehensively evaluate the associations between genetically predicted blood concentrations of a wide range of proteins and PDAC risk. We used data for 8,280 cases and 6,728 controls of European descent from the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case-Control Consortium (PanC4).

Materials and Methods

We conducted an extensive literature search to identify studies examining the associations between genetic variants at genome-wide scale and blood protein concentrations and our analysis based on a recently published comprehensive study (16). Focusing on a total of 3,301 healthy European descent individuals (2,481 and 820 in each of two subcohorts) in the INTERVAL study, Sun and colleagues identified 1,927 associations between 1,478 proteins and 764 genomic loci. In brief, 3,622 proteins in plasma were quantified by an aptamer-based multiplex protein assay (SOMAscan). Genotyping was performed using the Affymetrix Axiom UK Biobank genotyping array, with subsequent imputation based on a combined 1000 Genomes phase 3-UK10K reference panel. After quality control, pQTL analyses for 3,283 SOMAmers were conducted separately for each subcohort with adjustment for age, sex, duration between blood draw and processing, and the first three principal components. The results from these two subcohorts were combined by fixed-effects inverse-variance meta-analysis. The estimated associations between genetic variants and protein concentrations were considered significant only if they meet all three criteria: (i) $P < 1.5 \times 10^{-11}$ in the meta-analysis (5×10^{-8} /3,283 aptamers tested); (ii) $P < 0.05$ in both subcohorts; and (iii) consistent effect across subcohorts. The pQTLs identified in this study were used to generate the instrumental variables for evaluating the associations between genetically predicted proteins concentrations in blood and pancreatic cancer risk. When protein concentrations were associated with more than one pQTL variant located at the same

chromosome, the correlations between these single-nucleotide polymorphisms (SNP) were estimated using the pairwise linkage disequilibrium (LD) function of SNIIPA (http://snipa.helmholtz-muenchen.de/snipa/index.php?task=pairwise_ld). Only independent SNPs ($R^2 < 0.1$ based on 1000 Genomes Project phase 3 version 5 data for European descendants) were included to create a single instrument for each protein.

In this study, we used data from GWAS conducted in the PanScan and PanC4 consortia downloaded from the database of Genotypes and Phenotypes (dbGaP), including 8,280 PDAC cases and 6,728 controls of European ancestry. Detailed information on GWAS from PanScan and PanC4 can be found elsewhere (17–22). In brief, four GWAS studies including PanScan I, PanScan II, PanScan III, and PanC4 were genotyped using the Illumina HumanHap550, 610-Quad, OmniExpress, and OmniExpressExome arrays, respectively. Standard quality control was performed according to the guidelines of each consortium (20). We excluded study participants who were related to each other, had gender discordance, had genetic ancestry other than European, had a low call rate (less than 98% and 94% in PanC4 and PanScan, respectively), or had missing information on age or sex. We removed duplicated SNPs, and those with a high missing call rate (at least 2% and 6% in PanC4 and PanScan, respectively) or with violations of Hardy-Weinberg equilibrium ($P < 1 \times 10^{-4}$ and $P < 1 \times 10^{-7}$ in PanC4 and PanScan, respectively). For SNP data from PanC4, we additionally excluded those with minor allele frequency < 0.005 , with more than two discordant calls in duplicate samples, with more than one Mendelian error in HapMap control trios, and those with sex difference in allele frequency > 0.2 or in heterozygosity > 0.3 for autosomes/XY in European descendants. Genotype imputation was conducted using Minimac3 after prephasing with SHAPEIT from a reference panel of the Haplotype Reference Consortium (r1.1 2016; refs. 23–25). Imputed SNPs with an imputation quality of at least 0.3 were retained. We then assessed associations between individual variants and PDAC risk after adjustment of age, sex, and top 10 principal components (Supplementary Materials; Supplementary Table S1).

On the basis of the summary statistics from the above-mentioned pQTL study (16) and the analyses of PanScan/PanC4 GWAS datasets, we used the inverse variance weights (IVW) method to assess the association between genetically predicted blood protein concentrations and PDAC risk (26, 27). The beta coefficient of the association between each protein and PDAC risk was estimated using the formula of $\sum_i \beta_{i,GX} * \beta_{i,GY} * \sigma_{i,GY}^{-2} / (\sum_i \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})$, and its corresponding SE was calculated by $1 / (\sum_i \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})^{0.5}$. Here, $\beta_{i,GX}$ represents the beta coefficient adopted from the pQTL study for the association between the i th SNP and concentration of the protein of interest; $\beta_{i,GY}$ and $\sigma_{i,GY}$ represent the estimated beta coefficient and SE of the association between the i th SNP and PDAC risk in PanScan/PanC4 GWAS. We further computed ORs and confidence intervals (CI) by exponentiation of the beta coefficients. A Benjamini-Hochberg FDR of < 0.05 was used to define statistical significance. We also performed the analyses using individual level data. For this analysis, first we generated the predicted protein concentration for each subject in PanScan/PanC4 GWAS based on the individual-level genetic data and the beta coefficient from the pQTL study for the association between pQTL SNP and protein of interest. We then assessed the associations between predicted protein concentrations and PDAC risk. We further conducted conditional analysis with adjustments for previously identified risk variants to assess whether the observed

associations between genetically predicted protein concentrations and PDAC risk in our main analyses were independent of the risk variants identified in GWAS studies. Previously reported PDAC risk SNPs that are available in the current dataset (rs2816938, rs3790844, rs1486134, rs2736098, rs35226131, rs401681, rs17688601, rs78417682, rs6971499, rs2941471, rs10094872, rs1561927, rs505922, rs9581943, rs9543325, rs4795218, rs11655237, rs1517037) were adjusted for in the conditional analysis. In addition, we performed sensitivity analyses using data from different subgroups by consortium to assess the robustness of the significant associations.

For the proteins that were associated with PDAC risk, we performed an enrichment analysis of the genes encoding these proteins to examine whether they are enriched in specific pathways, functions or networks, by using Ingenuity Pathway Analysis (IPA) software. Detailed information of the methods has been described by the tool developer (28). In brief, the level of enrichment was estimated by assessing the overlap of the observed tested gene sets and the predicted regulated gene sets using Fisher exact test.

Results

We were able to assess associations between genetically predicted protein levels and PDAC risk for 1,226 proteins using pQTLs as instruments. Using the IVW method, we identified 38 proteins for which the genetically predicted concentrations showed associations with PDAC risk at a FDR of < 0.05 (23 proteins after Bonferroni correction; **Tables 1** and **2**); eight that remained significant after adjusting for known PDAC risk variants identified in previous GWAS (**Table 1**). Positive associations were observed for seven of these proteins, including Beta-crystallin B2 (CRBB2), Dedicator of cytokinesis protein 9 (DOCK9), VIP36-like protein (LMAN2L), Erythrocyte band 7 integral membrane protein (STOM), Tensin-2 (TENC1), Transmembrane protease serine 11D (TM11D), and Alcohol dehydrogenase 1B (ADH1B; ORs ranging from 1.17 to 1.52; **Table 1**). We observed a negative association between predicted protein concentration of C-X-C motif chemokine 10 (IP-10) and PDAC risk (OR per one SD increase in genetically predicted protein = 0.79; 95% CI, 0.69–0.91; $P = 1.19 \times 10^{-3}$; **Table 1**).

The associations for the other 30 proteins were substantially attenuated after adjusting for previously identified PDAC risk variants, potentially due to (i) the previously identified associations of risk SNPs with PDAC at these loci may be mediated through these proteins identified in this study, or (ii) confounding effects. Of these 30 proteins, 14 were positively associated with PDAC risk, including Histo-blood group ABO system transferase (BGAT), C1GALT1-specific chaperone 1 (C1GLC), Cadherin-5, Platelet glycoprotein 4 (CD36-ANTIGEN), Desmoglein-2, Protein FAM3B, CD209 Antigen (DC-SIGN), GDNF family receptor alpha-like (GFRAL), D-glucuronyl C5-epimerase (GLCE), Neurogenic locus notch homolog protein 1 (Notch1), Tollid-like protein 1 (TLL1), N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2 (B3GN2), Carbohydrate sulfotransferase 11 (CHSTB), and Angiotensin-1 receptor, soluble (sTie-2; ORs ranging from 1.12 to 3.62; **Table 2**). Conversely, an inverse association between predicted protein concentrations and PDAC risk was identified for P-Selectin, Intestinal-type alkaline phosphatase, Endoglin, Insulin-like growth factor 1 receptor (IGF-IR), IL3 receptor subunit alpha (IL3Ra), Insulin receptor (IR), Protein jagged-1 (JAG1), Leukemia inhibitory factor receptor (LIF-sR), Hepatocyte growth factor receptor (Met), E-selectin (sE-Selectin), Carbohydrate sulfotransferase 15 (CHST15), Thrombospondin type-1 domain-containing protein 1 (THSD1),

Table 1. Genetically predicted protein concentrations that are independently associated with pancreatic cancer risk after adjustment for previously identified risk SNPs.

Protein	Protein full name	Protein-encoding gene	Region for protein encoding gene	Instrument variants	Type of pQTL	OR (95% CI) ^a	P	FDR P^b	P value after adjusting for risk SNPs ^c
LMA2L	VIP36-like protein	LMAN2L	2q11.2	rs2271893	cis	1.39 (1.15–1.68)	6.47×10^{-4}	3.17×10^{-2}	7.72×10^{-4}
TM11D	Transmembrane protease serine 11D	TMPRSS11D	4q13.2	rs3197999	trans	1.17 (1.06–1.29)	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}
IP-10	C-X-C motif chemokine 10	CXCL10	4q21.1	rs11548618	cis	0.79 (0.69–0.91)	1.19×10^{-3}	3.93×10^{-2}	9.71×10^{-4}
ADH1B	Alcohol dehydrogenase 1B	ADH1B	4q23	rs13085791	trans	1.22 (1.08–1.37)	1.28×10^{-3}	4.14×10^{-2}	2.81×10^{-3}
STOM	Erythrocyte band 7 integral membrane protein	STOM	9q33.2	rs6770670	trans	1.19 (1.07–1.33)	1.05×10^{-3}	3.78×10^{-2}	2.27×10^{-3}
TENC1	Tensin-2	TNS2	12q13.13	rs3197999	trans	1.25 (1.09–1.42)	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}
DOCK9	Dedicator of cytokinesis protein 9	DOCK9	13q32.3	rs3197999	trans	1.32 (1.12–1.56)	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}
CRBB2	Beta-crystallin B2	CRYBB2	22q11.23	rs3197999	trans	1.52 (1.18–1.95)	1.11×10^{-3}	3.78×10^{-2}	2.44×10^{-3}

^aOR and CI (confidence interval) per one SD increase in genetically predicted protein after adjustment for age, sex, and top 10 principle components.

^bFDR P : false discovery rate (FDR) adjusted P value; associations with an FDR $P \leq 0.05$ considered statistically significant.

^cAssociations were adjusted for risk SNPs include: rs2816938, rs3790844, rs1486134, rs2736098, rs35226131, rs401681, rs17688601, rs78417682, rs6971499, rs2941471, rs10094872, rs1561927, rs505922, rs9581943, rs9543325, rs4795218, rs11655237, and rs1517037.

Table 2. Genetically predicted protein concentrations in association with pancreatic cancer risk that are potentially influenced by previously identified risk SNPs.

Protein	Protein full name	Protein-encoding gene	Region for protein-encoding gene	Instrument variants	Type of pQTL	OR (95% CI) ^a	P	FDR P ^b	P value after adjusting for risk SNPs ^c
P-Selectin	P-Selectin	SELP	1q24.2	rs74227709 rs6136	trans cis	0.86 (0.80-0.92)	2.67 × 10 ⁻⁵	1.49 × 10 ⁻³	0.32
sE-Selectin	E-selectin	SELE	1q24.2	rs2519093	trans	0.84 (0.80-0.88)	4.80 × 10 ⁻¹³	3.68 × 10 ⁻¹¹	0.13
B3GN2	N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2	B3GN2	2p15	rs2519093	trans	1.97 (1.64-2.37)	4.80 × 10 ⁻¹³	3.68 × 10 ⁻¹¹	0.13
Alkaline phosphatase, intestine	Intestinal-type alkaline phosphatase	ALPI	2q37.1	rs550057	trans	0.43 (0.35-0.53)	1.91 × 10 ⁻¹⁵	4.68 × 10 ⁻¹³	0.56
VEGF sR2	Vascular endothelial growth factor receptor 2	KDR	4q12	rs34231037	cis	0.80 (0.74-0.87)	4.86 × 10 ⁻⁸	3.31 × 10 ⁻⁶	0.55
TLL1	Toll-like protein 1	TLL1	4q32.3	rs635634	trans	1.30 (1.11-1.52)	1.10 × 10 ⁻³	3.78 × 10 ⁻²	0.60
LIF-sR	Leukemia inhibitory factor receptor	LIFR	5p13.1	rs8176747	trans	0.49 (0.41-0.59)	2.10 × 10 ⁻¹³	2.57 × 10 ⁻¹¹	0.10
gp130, soluble	IL6 receptor subunit beta	IL6ST	5q11.2	rs635634	trans	0.73 (0.63-0.84)	2.94 × 10 ⁻⁵	1.57 × 10 ⁻³	0.70
GFRAL	GDNF family receptor alpha-like	GFRAL	6p12.1	rs11574765	cis	1.33 (1.13-1.57)	6.15 × 10 ⁻⁴	3.14 × 10 ⁻²	0.47
GP116	Adhesion G protein-coupled receptor F5	ADGRF5	6p12.3	rs72975088	trans	0.76 (0.71-0.82)	4.80 × 10 ⁻¹³	3.68 × 10 ⁻¹¹	0.13
CD36-ANTIGEN	Platelet glycoprotein 4	CD36	7q21.11	rs8176672	cis	1.27 (1.10-1.46)	1.01 × 10 ⁻³	3.78 × 10 ⁻²	0.61
Met	Hepatocyte growth factor receptor	MET	7q31	rs635634	trans	0.57 (0.49-0.66)	2.10 × 10 ⁻¹³	2.57 × 10 ⁻¹¹	0.10
sTie-2	Angiopoietin-1 receptor, soluble	TEK	9p21.2	rs8176693	trans	1.28 (1.10-1.48)	1.01 × 10 ⁻³	3.78 × 10 ⁻²	0.61
Endoglin	Endoglin	ENG	9q34.11	rs635634	trans	0.41 (0.32-0.52)	2.10 × 10 ⁻¹³	2.57 × 10 ⁻¹¹	0.10
BGAT	Histo-blood group ABO system transferase	ABO	9q34.2	rs505922	cis	1.20 (1.15-1.24)	5.74 × 10 ⁻²¹	2.35 × 10 ⁻¹⁸	NA ^d
Notch1	Neurogenic locus notch homolog protein 1	NOTCH1	9q34.3	rs8176743	trans	1.46 (1.16-1.83)	1.10 × 10 ⁻³	3.78 × 10 ⁻²	0.60
CHST15	Carbohydrate sulfotransferase 15	CHST15	10q26.13	rs550057	trans	0.52 (0.44-0.61)	1.91 × 10 ⁻¹⁵	4.68 × 10 ⁻¹³	0.56
CHSTB	Carbohydrate sulfotransferase II	CHSTII	12q23.3	rs687621	trans	3.62 (2.77-4.74)	5.57 × 10 ⁻²¹	2.35 × 10 ⁻¹⁸	0.46
THSD1	Thrombospondin type-1 domain-containing protein 1	THSD1	13q14.3	rs41292808	trans	0.74 (0.65-0.83)	5.85 × 10 ⁻⁷	3.77 × 10 ⁻⁵	0.13
F177A	Protein FAM177A1	FAM177A1	14q13.2	rs2519093	cis	0.63 (0.54-0.73)	1.54 × 10 ⁻⁹	1.11 × 10 ⁻⁷	0.45
GLCE	D-glucuronyl C5-epimerase	GLCE	15q23	rs550057 rs679574	Trans trans	1.13 (1.07-1.19)	1.85 × 10 ⁻⁵	1.08 × 10 ⁻³	0.44
IGF-IR	Insulin-like growth factor 1 receptor	IGFIR	15q26.3	rs2519093	cis	0.38 (0.29-0.49)	2.10 × 10 ⁻¹³	3.22 × 10 ⁻¹¹	0.10
Cadherin-5	Cadherin-5	CDH5	16q21	rs635634	trans	1.12 (1.05-1.20)	1.10 × 10 ⁻³	3.78 × 10 ⁻²	0.60
Desmoglein-2	Desmoglein-2	DSG2	18q12.1	rs8176746 rs2704050	trans trans	1.88 (1.59-2.23)	3.31 × 10 ⁻¹³	3.67 × 10 ⁻¹¹	0.99
IR	Insulin receptor	INSR	19p13.2	rs687621	cis	0.69 (0.63-0.77)	3.59 × 10 ⁻¹³	4.40 × 10 ⁻¹¹	0.12
DC-SIGN	CD209 antigen	CD209	19p13.2	rs507666	trans	1.32 (1.25-1.40)	5.74 × 10 ⁻²¹	2.35 × 10 ⁻¹⁸	NA ^d
JAG1	Protein jagged-1	JAG1	20p12.2	rs7041 rs550057	trans trans	0.70 (0.59-0.82)	9.08 × 10 ⁻⁶	7.95 × 10 ⁻⁴	0.41

(Continued on the following page)

Table 2. Genetically predicted protein concentrations in association with pancreatic cancer risk that are potentially influenced by previously identified risk SNPs. (Cont'd)

Protein	Protein full name	Protein-encoding gene	Region for protein-encoding gene	Instrument variants	Type of pQTL	OR (95% CI) ^a	P	FDR P ^b	P value after adjusting for risk SNPs ^c
FAM3B	Protein FAM3B	FAM3B	21q22.3	rs2608894 rs73226194	<i>cis</i> <i>trans</i>	1.15 (1.06–1.25)	8.01×10^{-4}	3.78×10^{-2}	0.64
IL-3Ra	IL3 receptor subunit alpha	IL3RA	Xp22.3	rs2519093	<i>trans</i>	0.78 (0.73–0.84)	4.80×10^{-13}	4.90×10^{-11}	0.13
C1GLC	C1GALT1-specific chaperone 1	C1GALT1C1	Xq24	rs7787942 rs2519093	<i>trans</i> <i>trans</i>	1.39 (1.28–1.50)	5.34×10^{-15}	1.09×10^{-12}	3.02×10^{-3}

^aOR and CI (confidence interval) per one SD increase in genetically predicted protein after adjustment for age, sex, and top 10 principle components.

^bFDR P: false discovery rate (FDR) adjusted P value; associations with an FDR $P \leq 0.05$ considered statistically significant.

^cAssociations were adjusted for risk SNPs include: rs2816938, rs3790844, rs1486134, rs2736098, rs35226131, rs401681, rs17688601, rs78417682, rs6971499, rs2941471, rs10094872, rs1561927, rs505922, rs9581943, rs9543325, rs4795218, rs11655237, and rs1517037.

^dInstrument SNP itself is a known PC risk SNP.

Adhesion G protein-coupled receptor F5 (GP116), IL6 receptor subunit beta (gp130, soluble), VEGF receptor 2 (VEGF sR2), and Protein FAM177A1 (F177A; ORs ranging from 0.38 to 0.86; **Table 2**).

On the basis of subgroup analyses, the associations of the identified 38 proteins, in general, were robust across the GWAS subsets (PanScan I, II, and III; PanScan I and II; PanC4 and PanScan I and II; and PanC4; Supplementary Table S2).

The IPA analysis showed enrichment in several cancer-related function pathways for the genes encoding the proteins identified by our study. The top canonical pathways identified included IL15 production ($P = 2.71 \times 10^{-6}$) and STAT3 ($P = 5.25 \times 10^{-6}$; **Table 3**).

Discussion

This is the first study with a large sample size to systematically evaluate the associations between genetically predicted circulating protein concentrations and PDAC risk using pQTLs as study instruments. Overall, we identified 38 proteins that were significantly associated with PDAC risk after FDR correction, including eight that showed an association with PDAC risk independently from the previously identified PDAC risk variants. If confirmed, our data suggest new knowledge on the etiology of PDAC, and provide a list of proteins as candidate blood biomarkers for assessing risk of PDAC, a malignancy with universally high case fatality.

Previous studies have suggested blood concentrations of CA242, PIVKA-II, PAM4, S100A6, OPN, RBM6, EphA2, and OPG to be associated with pancreatic cancer risk (4–7). However, with the exception of S100A6 and OPG, a pQTL was not identified for these proteins (16). Using the corresponding pQTL rs62143206 of S100A6 as an instrumental variable, we did not observe evidence of association for S100A6 (OR = 1.01; 95% CI, 0.91–1.13; $P = 0.86$) with PDAC. For OPG, by using the corresponding pQTL rs570618 as an instrumental variable, we observed an association (OR = 1.35; 95% CI, 1.04–1.76, $P = 0.03$), although this was not significant after correcting for multiple comparisons. Nevertheless, the direction of the association is consistent with that identified in previous work. Our inconsistent finding with previous studies for S100A6 might be explained by either the weak instrument used in our study or potential biases in previous studies that used a conventional observational design.

In this large study, we identified eight PDAC-associated proteins that are independent of PDAC risk variants previously identified in GWAS. Compared with GWAS, which aim to identify novel susceptibility variants by assessing the association between each genetic variant and disease risk across the genome, this study has improved statistical power by aggregating the effects of several SNPs into one continuous testing unit, the genetically predicted blood concentration of protein, when applicable. In this study, we used both *cis* and *trans* pQTL as genetic instruments whenever possible (**Tables 1** and **2**). Previous research has supported a potential role for some of the novel proteins identified in this study in pancreatic tumorigenesis. On the basis of an IHC analysis, significantly higher expression of tensin-2 was observed in pancreatic tumor tissues than in adjacent normal tissues (29). In the same study, there were also positive associations of tensin-2 with glucose metabolism-related insulin receptor substrate 1 and glucose transporter type 4, the proliferation marker ki-67, the angiogenesis marker CD31, and the mesenchymal markers N-cadherin and fibronectin, suggesting a potential role of tensin-2 in pancreatic cancer metabolism, proliferation, angiogenesis, and epithelial–mesenchymal transition process (29). Protein TM11D, encoded by gene *TMPRSS11D*, serves as an efficient activator of

Table 3. Canonical pathways, diseases, biofunctions, and networks associated with the genes encoding identified pancreatic cancer risk-associated proteins.

Top canonical pathways	Top diseases and disorders	Molecular and cellular functions	Top networks
IL15 production	Cancer	Cell-to-cell signaling and interaction	Cardiovascular system development and function, organismal development, and cellular movement
STAT3 pathway	Organismal injury and abnormalities	Carbohydrate metabolism	Cell signaling, cell-to-cell signaling and interaction, and cancer
Sperm motility	Dermatologic diseases and conditions	Cellular development	
Heparan sulfate biosynthesis (late stages)	Tumor morphology	Cellular function and maintenance	
Granulocyte adhesion and diapedesis	Inflammatory response	Cellular growth and proliferation	

macrophage stimulating protein (MSP). MSP can further stimulate the activation of its receptor, RON, which has been suggested to be overexpressed early in the progression of pancreatic malignancy (30, 31).

For the other 30 proteins identified in this study, for which associations with PDAC risk were mainly explained by previously reported PDAC risk variants, some were also suggested to play a role in pancreatic cancer development based on *in vitro/in vivo* human studies. For example, GWAS has identified the *ABO* gene as a susceptibility locus for PDAC risk (20). The protective T allele of rs505922, the instrument SNP for the protein encoded by *ABO*, is in LD with a single base-pair deletion that encodes the O antigen. Genotype-inferred O blood type was shown to be associated with a reduced risk of PDAC compared with other blood types, which was suggested to be possibly attributed to altered inflammation state, glycosyltransferase activity, or differentiated expression of blood group antigens (32, 33). On the basis of *in vitro* experiments, knockdown of *C1GALT1C1*, the encoding gene for protein C1GLC, promoted migration and survival but inhibited proliferation of pancreatic cancer cells (34). In contrast, for some of the proteins identified, it is worth noting that the directions of the observed associations are not consistent with those suggested in the literature. For example, *CHST15* is an enzyme that biosynthesizes chondroitin sulfate, which is known to be able to promote tumor invasion and metastasis. *CHST15* mRNA was found to be highly expressed in pancreatic cancer cell lines (35). Pancreatic tumor growth was inhibited after *CHST15* protein blood concentrations were reduced in both mice and humans (36). In this study, however, we found that a low level of genetically determined *CHST15* concentration was associated with an increased risk of pancreatic cancer. Possible explanations for this inconsistency may include that the focus of this study is the genetically regulated circulating protein concentrations, whereas the measured protein concentrations in previous studies may be influenced by both inherent and extrinsic factors. Additional well-designed studies with directly measured protein concentrations are warranted to better understand the relationship between the identified proteins and pancreatic cancer risk.

The strengths of our study include its large sample size for the main association analyses, providing high statistical power to detect proteins associated with PDAC risk. The use of genetic instruments potentially minimized several biases that are commonly encountered in conventional observational studies. However, several limitations of the current work need to be recognized. First, our results may be susceptible to potential pleiotropic effects. For example, rs3197999, the instrument for proteins CRBB2, DOCK9, TENC1, and TM11D, has also been associated with several other traits, including primary sclerosing

cholangitis, Crohn disease, and ulcerative colitis (37–39). Similarly, rs2519093, which was the instrument for proteins IL3Ra and sE-Selectin, as well as one of the variants constituting the instrument for P-Selectin, C1GLC, FAM3B, GLCE, and THSD1, was shown to be associated with coronary artery disease, allergy, and venous thromboembolism (40–42). Although most of these traits do not appear to be strongly related to pancreatic carcinogenesis, allergy is known to be potentially associated with pancreatic cancer risk (43, 44), and previous studies have linked Crohn disease and ulcerative colitis with pancreatic cancer risk (45, 46). Results of our MR-Egger regression analyses for protein FAM3B ($P = 0.55$) and P-Selectin ($P = 0.73$), which involved three variants as instrument, suggested that their associations were less likely to be influenced by potential directional pleiotropic effects (47). Second, in this study, we were only able to capture the genetically regulated components of circulating protein concentrations, so that their utility of as a biomarker is unclear due to the impact of environmental factors. Further prospective studies with measured circulating protein concentrations in predisease blood samples are warranted to validate the potential predicting role of our identified proteins in pancreatic cancer. Third, our analysis largely relies on the pQTLs identified by previous GWAS of circulating protein concentrations; thus our ability to evaluate candidate protein biomarkers for pancreatic cancer was limited by whether a pQTL had been identified for some of these proteins. We expect that additional protein biomarkers can be identified when new knowledge is generated regarding the pQTL for additional proteins. Fourth, research has suggested that specific variables, such as smoking and body weight, are related to protein levels in blood (48, 49). Ideally for our study the instrument pQTL SNPs would be identified in analyses with adjustment of relevant variables; however, this is not the case for the INTERVAL study. Further research is needed to validate our findings.

In summary, in this large study, we identified multiple novel protein biomarkers, for which the genetically predicted circulating concentrations were associated with PDAC risk. Our study may serve as a basis for future investigation of these proteins to better understand the underlying mechanisms of PDAC and to advance the development of effective biomarker panels for risk assessment of PDAC.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Authors' Contributions

Conception and design: R.L. Milne, G.G. Giles, G. Scelo, L. Wu

Development of methodology: X. Shu, X. Guo, L. Wu

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Bao, R.L. Milne, G.G. Giles, M. Du, E. White, H.A. Risch, N. Malats, D. Li, P. Bracci, R.E. Neale, S. Gallinger, S.K. Van Den Eeden, A.A. Arslan, F. Canzian, C. Kooperberg, L.E. Beane Freeman, G. Scelo, C.A. Haiman, L. Le Marchand, H. Yu, G.M. Petersen, R. Stolzenberg-Solomon, A.P. Klein, Q. Cai, X.-O. Shu, L. Wu

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Zhu, X. Guo, C. Wu, E. White, H.A. Risch, L. Wu

Writing, review, and/or revision of the manuscript: J. Zhu, X. Shu, D. Liu, R.L. Milne, G.G. Giles, C. Wu, E. White, H.A. Risch, N. Malats, E.J. Duell, P.J. Goodman, D. Li, P. Bracci, V. Katzke, R.E. Neale, S.K. Van Den Eeden, A.A. Arslan, C. Kooperberg, L.E. Beane Freeman, G. Scelo, K. Visvanathan, L. Le Marchand, H. Yu, G.M. Petersen, R. Stolzenberg-Solomon, A.P. Klein, Q. Cai, J. Long, X.-O. Shu, W. Zheng, L. Wu

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): E.J. Duell, A.A. Arslan, F. Canzian, L.E. Beane Freeman, G. Scelo, L. Le Marchand, R. Stolzenberg-Solomon, A.P. Klein, L. Wu

Study supervision: L. Wu

Other (provision of data from one of the studies included in the larger cohort): P.J. Goodman

Acknowledgments

The datasets used for the analyses described in this manuscript were obtained from dbGaP at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap> through dbGaP accession phs000206.v5.p3 and phs000648.v1.p1. The authors thank Laufey Amundadottir, Eric Jacobs, and Idan Ben-Barak for their help for this manuscript. The authors also thank all of the individuals for their participation in the parent studies and all the researchers, clinicians, technicians, and administrative staff for their contribution to the studies. L. Wu is supported by NCI R00CA218892. D. Liu is supported by the Harbin Medical University Cancer Hospital. The PanScan study was funded in whole or in part with federal funds from the NCI, US NIH

under contract number HHSN261200800001E. Additional support was received from NIH/NCI K07 CA140790, the American Society of Clinical Oncology Conquer Cancer Foundation, the Howard Hughes Medical Institute, the Lustgarten Foundation, the Robert T. and Judith B. Hale Fund for Pancreatic Cancer Research, and Promises for Purple. A full list of acknowledgments for each participating study is provided in the Supplementary Note of the manuscript with PubMed ID: 25086665. For the PanC4 GWAS study, the patients and controls were derived from the following PANC4 studies: Johns Hopkins National Familial Pancreas Tumor Registry, Mayo Clinic Biospecimen Resource for Pancreas Research, Ontario Pancreas Cancer Study (OPCS), Yale University, MD Anderson Case Control Study, Queensland Pancreatic Cancer Study, University of California San Francisco Molecular Epidemiology of Pancreatic Cancer Study, International Agency of Cancer Research, and Memorial Sloan Kettering Cancer Center. This work is supported by NCI R01CA154823. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN2682011000111. The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN26820160001C, HHSN26820160002C, HHSN268201600003C, and HHSN268201600004C. This manuscript was prepared in collaboration with investigators of the WHI, and has been reviewed and/or approved by the Women's Health Initiative (WHI). WHI investigators are listed at <https://www.whi.org/researchers/SitePages/Home.aspx> a Paper/Write a Paper-Resources/Acknowledgement Lists:Short Lists. SELECT study is supported by NIH grant award number U10 CA37429 (to C.D. Blanke), and UMI CA182883 (to C.M. Tangen/I.M. Thompson).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 17, 2020; revised March 15, 2020; accepted April 27, 2020; published first May 21, 2020.

References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018;68:7–30.
- Raimondi S, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: an overview. *Nat Rev Gastroenterol Hepatol* 2009;6:699–708.
- Ballehaninna UK, Chamberlain RS. Serum CA 19-9 as a biomarker for pancreatic cancer—a comprehensive review. *Indian J Surg Oncol* 2011;2:88–100.
- Koshikawa N, Minegishi T, Kiyokawa H, Seiki M. Specific detection of soluble EphA2 fragments in blood as a new biomarker for pancreatic cancer. *Cell Death Dis* 2017;8:e3134.
- Loosen SH, Neumann UP, Trautwein C, Roderburg C, Luedde T. Current and future biomarkers for pancreatic adenocarcinoma. *Tumour Biol* 2017;39:1010428317692231.
- Duan B, Hu X, Fan M, Xiong X, Han L, Wang Z, et al. RNA-binding motif protein 6 is a candidate serum biomarker for pancreatic cancer. *Proteomics Clin Appl* 2019;13:e1900048.
- Tartaglione S, Pecorella I, Zarrillo SR, Granato T, Viggiani V, Manganaro L, et al. Protein induced by vitamin K absence II (PIVKA-II) as a potential serological biomarker in pancreatic cancer: a pilot study. *Biochem Med* 2019;29:020707.
- Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133–63.
- Jia J, Dou P, Gao M, Kong X, Li C, Liu Z, et al. Assessment of causal direction between gut microbiota-dependent metabolites and cardiometabolic health: a bidirectional mendelian randomization analysis. *Diabetes* 2019;68:1747–55.
- Shu X, Wu L, Khankari NK, Shu X-O, Wang TJ, Michailidou K, et al. Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis. *Int J Epidemiol* 2019;48:795–806.
- 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;109:djx012.
- Langdon RJ, Richmond RC, Hemani G, Zheng J, Wade KH, Carreras-Torres R, et al. A phenome-wide mendelian randomization study of pancreatic cancer using summary genetic data. *Cancer Epidemiol Biomark Prev* 2019;28:2070–8.
- Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Stat Methods Med Res* 2007;16:309–30.
- Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017;8:14357.
- Enroth S, Johansson A, Enroth SB, Gyllenstein U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat Commun* 2014;5:4684.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2012;9:179–81.

24. 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.
25. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48:1279–83.
26. Shu X, Bao J, Wu L, Long J, Shu X-O, Guo X, et al. Evaluation of associations between genetically predicted circulating protein biomarkers and breast cancer risk. *Int J Cancer* 2020;146:2130–8.
27. Wu L, Shu X, Bao J, Guo X, Kote-Jarai Z, Haiman CA, et al. Analysis of over 140,000 European descendants identifies genetically predicted blood protein biomarkers associated with prostate cancer risk. *Cancer Res* 2019;79:4592–8.
28. Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2014;30:523–30.
29. Cheng L-C, Chen Y-L, Cheng A-N, Lee AY-L, Cho C-Y, Huang J-S, et al. AXL phosphorylates and up-regulates TNS2 and its implications in IRS-1-associated metabolism in cancer cells. *J Biomed Sci* 2018;25:80.
30. Kawaguchi M, Kataoka H. MST1 (macrophage stimulating 1 (hepatocyte growth factor-like)). *Atlas Genet Cytogenet Oncol Haematol* 2013 [cited 2019 Jul 31]. Available from: <http://hdl.handle.net/2042/51872>.
31. Camp ER, Yang A, Gray MJ, Fan F, Hamilton SR, Evans DB, et al. Tyrosine kinase receptor RON in human pancreatic cancer: expression, function, and validation as a target. *Cancer* 2007;109:1030–9.
32. Wolpin BM, Kraft P, Xu M, Steplowski E, Olsson ML, Arslan AA, et al. Variant ABO blood group alleles, secretor status, and risk of pancreatic cancer: results from the pancreatic cancer cohort consortium. *Cancer Epidemiol Biomark Prev* 2010;19:3140–9.
33. Wolpin BM, Chan AT, Hartge P, Chanock SJ, Kraft P, Hunter DJ, et al. ABO blood group and the risk of pancreatic cancer. *J Natl Cancer Inst* 2009;101:424–31.
34. Hofmann BT, Schlüter L, Lange P, Mercanoglu B, Ewald F, Fölster A, et al. COSMC knockdown mediated aberrant O-glycosylation promotes oncogenic properties in pancreatic cancer. *Mol Cancer* 2015;14:109.
35. Takakura K, Shibasaki Y, Yoneyama H, Fujii M, Hashiguchi T, Ito Z, et al. Inhibition of cell proliferation and growth of pancreatic cancer by silencing of carbohydrate sulfotransferase 15 in vitro and in a xenograft model. *PLoS One* 2015;10:e0142981.
36. Nishimura M, Yahagi N, Itoi T, Ochiai Y, Matsuda Y. A translational study to investigate the role of carbohydrate sulfotransferase 15 for pancreatic cancer biology from in vitro to first-in-human clinical research. *J Clin Oncol* 33, 2015 (suppl; abstr e22201).
37. Melum E, Franke A, Schramm C, Weismüller TJ, Gotthardt DN, Offner FA, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. *Nat Genet* 2011;43:17–9.
38. Julià A, Domènech E, Chaparro M, García-Sánchez V, Gomollón F, Panés J, et al. A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis. *Hum Mol Genet* 2014;23:6927–34.
39. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118–25.
40. Klarin D, Emdin CA, Natarajan P, Conrad MF; INVENT Consortium, Kathiresan S. Genetic analysis of venous thromboembolism in UK Biobank identifies the ZFPM2 locus and implicates obesity as a causal risk factor. *Circ Cardiovasc Genet* 2017;10. pii: e001643.
41. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016;48:709–17.
42. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res* 2018;122:433–43.
43. Gandini S, Lowenfels AB, Jaffee EM, Armstrong TD, Maisonneuve P. Allergies and the risk of pancreatic cancer: a meta-analysis with review of epidemiology and biological mechanisms. *Cancer Epidemiol Biomark Prev* 2005;14:1908–16.
44. Olson SH, Hsu M, Satagopan JM, Maisonneuve P, Silverman DT, Lucente-forde E, et al. Allergies and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Case-Control Consortium. *Am J Epidemiol* 2013;178:691–700.
45. Hemminki K, Li X, Sundquist J, Sundquist K. Cancer risks in ulcerative colitis patients. *Int J Cancer* 2008;123:1417–21.
46. Hemminki K, Li X, Sundquist J, Sundquist K. Cancer risks in Crohn disease patients. *Ann Oncol* 2009;20:574–80.
47. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
48. Wingerd J, Sponzilli EE. Concentrations of serum protein fractions in white women: effects of age, weight, smoking, tonsillectomy, and other factors. *Clin Chem* 1977;23:1310–7.
49. Yang M, Kohler M, Heyder T, Forsslund H, Garberg HK, Karimi R, et al. Long-term smoking alters abundance of over half of the proteome in bronchoalveolar lavage cell in smokers with normal spirometry, with effects on molecular pathways associated with COPD. *Respir Res* 2018;19:40.

BLOOD CANCER DISCOVERY

Associations between Genetically Predicted Blood Protein Biomarkers and Pancreatic Cancer Risk

Jingjing Zhu, Xiang Shu, Xingyi Guo, et al.

Cancer Epidemiol Biomarkers Prev 2020;29:1501-1508. Published OnlineFirst May 21, 2020.

Updated version Access the most recent version of this article at:
doi: [10.1158/1055-9965.EPI-20-0091](https://doi.org/10.1158/1055-9965.EPI-20-0091)

Supplementary Material Access the most recent supplemental material at:
<http://cebp.aacrjournals.org/content/suppl/2020/05/21/1055-9965.EPI-20-0091.DC1>

Cited articles This article cites 46 articles, 7 of which you can access for free at:
<http://cebp.aacrjournals.org/content/29/7/1501.full#ref-list-1>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/29/7/1501>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.