Associations between Genetically Predicted Blood Protein Biomarkers and Pancreatic Cancer Risk


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, TMI1D, IP-10, ADH1B, STOM, TENC1, 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 publication of early detection and effective treatment. Therefore, it is crucial to identify novel biomarkers to enable early diagnosis and effective patient management.

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, TMI1D, IP-10, ADH1B, STOM, TENC1, 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.
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 (SOMAmer). 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 SOMAmeres 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 (\( 5 \times 10^{-8} \) for 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 SNPiPA (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 data 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 \( P < 1 \times 10^{-3} \) 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 SHAPIT 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

\[
\hat{\beta}_{i,GX} = \frac{\sum_i \hat{\beta}_{i,GX} \pm \hat{\beta}_{i,GX} \times \hat{\sigma}_{i,GX} \times \hat{\sigma}_{i,GX}}{1/\{\hat{\beta}_{i,GX} \times \hat{\sigma}_{i,GX}\}^{0.5}}
\]

Here, \( \hat{\beta}_{i,GX} \) represents the beta coefficient adopted from the pQTL study for the association between the \( i \)-th SNP and concentration of the protein of interest; \( \hat{\beta}_{i,GX} \) and \( \hat{\sigma}_{i,GX} \) 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, rs39700844, rs1486134, rs2736098, rs35226131, rs401681, rs78417682, rs6971499, rs2941471, rs10094872, rs505922, rs9581943, rs9543325, rs7495218, rs11565237, 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 examining whether they are enriched in specific proteins, by using Ingenuity Pathway Analysis (IPA) software. 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 examining whether they are enriched in specific proteins, by using Ingenuity Pathway Analysis (IPA) software.

Table 1

<table>
<thead>
<tr>
<th>Protein Protein full name</th>
<th>Type of protein</th>
<th>Region of protein encoding gene</th>
<th>Protein-encoding gene</th>
<th>pQTL</th>
<th>p</th>
<th>P value after adjusting for risk SNPs</th>
<th>FDR α</th>
<th>OR ± SE/CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH1B Alcohol dehydrogenase 1B</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs2736098</td>
<td>4.14</td>
<td>1.19 × 10⁻⁴</td>
<td>1.39 (1.15—1.68)</td>
<td>1.11</td>
</tr>
<tr>
<td>STOM Erythrocyte band 7 integral membrane protein</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs39700844</td>
<td>4.27</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>TENC1 Tensin-2</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs1486134</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>DOCK9 Dedicator of cytokinesis protein 9</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs2736098</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>CRBB2 Beta-crystallin B2</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs35226131</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>LMAN2L VIP36-like protein</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs401681</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>CXCL10 C-X-C motif chemokine 10</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs17688601</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>IR Insulin receptor</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs78417682</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>JAG1 Protein jagged-1</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs2941471</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>STOM Erythrocyte band 7 integral membrane protein</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs10094872</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>FGRAL D-glucuronyl C5-epimerase</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs1561927</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>GDNF family receptor alpha-like</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs505922</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>DC-SIGN CD209 Antigen</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs9581943</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>TM11D Transmembrane protease serine 11D</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs2816938</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>TM11D Transmembrane protease serine 11D</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs3790844</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
<tr>
<td>TM11D Transmembrane protease serine 11D</td>
<td>Protein-encoding gene</td>
<td>TM11D Transmembrane protease serine 11D</td>
<td>TM11D</td>
<td>rs1486134</td>
<td>4.32</td>
<td>1.11 × 10⁻⁴</td>
<td>1.32 (1.12—1.56)</td>
<td>1.11</td>
</tr>
</tbody>
</table>

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 IV 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 R2 (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 × 10⁻⁴; 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 Hist-blood group ABO system transferase (BGAT), CIGALTI1-specific chaperone 1 (C1GLC), Cadherin-5, Platelet glycoprotein 4 (CD36-ANTITGEN), 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), Tolloid-like protein 1 (TLL1), N-acetyllactosaminide beta-1,3-N-acylglycerolcosaminyltransferase 2 (B3GNA2), Carbohydrate sulfotransferase 11 (CHST11), and Angiopoietin-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 2. Genetically predicted protein concentrations in association with pancreatic cancer risk that are potentially influenced by previously identified risk SNPs.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Protein full name</th>
<th>Protein-encoding gene</th>
<th>Region for protein-encoding gene</th>
<th>Instrument variants</th>
<th>Type of pGTL</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>FDR P</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-Selectin</td>
<td>P-Selectin</td>
<td>SELP</td>
<td>1q24.2</td>
<td>rs74227709</td>
<td>trans</td>
<td>0.86 (0.80-0.92)</td>
<td>1.49 x 10^-3</td>
<td>0.32</td>
</tr>
<tr>
<td>sE-Selectin</td>
<td>E-selectin</td>
<td>SELE</td>
<td>1q24.2</td>
<td>rs2519093</td>
<td>trans</td>
<td>0.84 (0.80-0.88)</td>
<td>4.80 x 10^-13</td>
<td>3.68 x 10^-11</td>
</tr>
<tr>
<td>B3GN2</td>
<td>N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2</td>
<td>B3GN2</td>
<td>2p15</td>
<td>rs2519093</td>
<td>trans</td>
<td>1.97 (1.64-2.37)</td>
<td>4.80 x 10^-13</td>
<td>3.68 x 10^-11</td>
</tr>
<tr>
<td>Alkaline phosphatase, intestine</td>
<td>Intestinal-type alkaline phosphatase</td>
<td>ALPI</td>
<td>2q37.1</td>
<td>rs550057</td>
<td>trans</td>
<td>0.43 (0.35-0.53)</td>
<td>1.91 x 10^-15</td>
<td>4.68 x 10^-13</td>
</tr>
<tr>
<td>VEGF sR2</td>
<td>Vascular endothelial growth factor receptor 2</td>
<td>KDR</td>
<td>4q12</td>
<td>rs3423037</td>
<td>cis</td>
<td>0.80 (0.74-0.87)</td>
<td>4.86 x 10^-8</td>
<td>3.31 x 10^-6</td>
</tr>
<tr>
<td>TLL1</td>
<td>Tolloid-like protein 1</td>
<td>TLL1</td>
<td>4q32.3</td>
<td>rs8176747</td>
<td>trans</td>
<td>1.30 (1.11-1.52)</td>
<td>3.78 x 10^-2</td>
<td>0.60</td>
</tr>
<tr>
<td>LIF-sR</td>
<td>Leukemia inhibitory factor receptor</td>
<td>LIFR</td>
<td>5p13.1</td>
<td>rs653634</td>
<td>trans</td>
<td>0.49 (0.41-0.59)</td>
<td>2.10 x 10^-10</td>
<td>2.57 x 10^-11</td>
</tr>
<tr>
<td>gp130, soluble</td>
<td>IL6 receptor subunit beta</td>
<td>IL6ST</td>
<td>5q12.2</td>
<td>rs653634</td>
<td>trans</td>
<td>0.73 (0.63-0.84)</td>
<td>2.94 x 10^-5</td>
<td>1.57 x 10^-5</td>
</tr>
<tr>
<td>GFRAL</td>
<td>GDNF family receptor alpha-like</td>
<td>GFRAL</td>
<td>1p21</td>
<td>rs8176672</td>
<td>cis</td>
<td>1.33 (1.13-1.57)</td>
<td>6.15 x 10^-4</td>
<td>3.14 x 10^-2</td>
</tr>
<tr>
<td>GPI16</td>
<td>Adhesion G protein-coupled receptor F5</td>
<td>ADGFRS</td>
<td>6p12.3</td>
<td>rs2519093</td>
<td>trans</td>
<td>0.76 (0.71-0.82)</td>
<td>4.80 x 10^-13</td>
<td>3.66 x 10^-11</td>
</tr>
<tr>
<td>CD36-ACTGEN</td>
<td>Platelet glycoprotein 4</td>
<td>CD36</td>
<td>7q21.1</td>
<td>rs8176693</td>
<td>trans</td>
<td>1.27 (1.10-1.46)</td>
<td>3.78 x 10^-2</td>
<td>0.61</td>
</tr>
<tr>
<td>Met</td>
<td>Hepatocyte growth factor receptor</td>
<td>MET</td>
<td>7q31</td>
<td>rs653634</td>
<td>trans</td>
<td>0.57 (0.49-0.66)</td>
<td>2.10 x 10^-10</td>
<td>2.57 x 10^-11</td>
</tr>
<tr>
<td>st1e-2</td>
<td>Angiopoietin-1 receptor, soluble</td>
<td>TIEK</td>
<td>9p21.2</td>
<td>rs8176693</td>
<td>trans</td>
<td>1.28 (1.10-1.48)</td>
<td>3.78 x 10^-2</td>
<td>0.61</td>
</tr>
<tr>
<td>Endoglin</td>
<td>Endoglin</td>
<td>ENG</td>
<td>9q34.11</td>
<td>rs653634</td>
<td>trans</td>
<td>0.41 (0.32-0.52)</td>
<td>2.10 x 10^-10</td>
<td>2.57 x 10^-11</td>
</tr>
<tr>
<td>BGAT</td>
<td>Histo-blood group ABO system transferase</td>
<td>ABO</td>
<td>9q34.2</td>
<td>rs505922</td>
<td>cis</td>
<td>1.20 (1.15-1.24)</td>
<td>5.74 x 10^-7</td>
<td>2.35 x 10^-6</td>
</tr>
<tr>
<td>Notch1</td>
<td>Neurogenic locus notch homolog protein 1</td>
<td>NOTCH1</td>
<td>2q34.3</td>
<td>rs8176743</td>
<td>trans</td>
<td>1.46 (1.16-1.83)</td>
<td>3.78 x 10^-2</td>
<td>0.60</td>
</tr>
<tr>
<td>CHST15</td>
<td>Carbohydrate sulfotransferase 15</td>
<td>CHST15</td>
<td>10q26.13</td>
<td>rs505057</td>
<td>trans</td>
<td>0.52 (0.44-0.61)</td>
<td>1.91 x 10^-10</td>
<td>4.68 x 10^-13</td>
</tr>
<tr>
<td>CHST1B</td>
<td>Carbohydrate sulfotransferase 11</td>
<td>CHST11</td>
<td>12q23.3</td>
<td>rs817621</td>
<td>trans</td>
<td>3.62 (2.77-4.74)</td>
<td>5.57 x 10^-21</td>
<td>2.35 x 10^-10</td>
</tr>
<tr>
<td>THSD1</td>
<td>Thrombospondin type-I domain-containing protein 1</td>
<td>THSD1</td>
<td>15q34.3</td>
<td>rs4129808</td>
<td>trans</td>
<td>0.74 (0.65-0.83)</td>
<td>5.85 x 10^-7</td>
<td>3.77 x 10^-5</td>
</tr>
<tr>
<td>FI77A</td>
<td>Protein FAM177AI</td>
<td>FAM177AI</td>
<td>14q13.2</td>
<td>rs2519093</td>
<td>cis</td>
<td>0.63 (0.54-0.73)</td>
<td>1.54 x 10^-9</td>
<td>1.11 x 10^-7</td>
</tr>
<tr>
<td>GLCE</td>
<td>D-glucuronol C5-epimerase</td>
<td>GLCE</td>
<td>15q23</td>
<td>rs679574</td>
<td>trans</td>
<td>1.13 (1.07-1.19)</td>
<td>185 x 10^-5</td>
<td>108 x 10^-3</td>
</tr>
<tr>
<td>IGF-IR</td>
<td>Insulin-like growth factor 1 receptor</td>
<td>IGF1R</td>
<td>15q26.3</td>
<td>rs653634</td>
<td>trans</td>
<td>0.38 (0.29-0.49)</td>
<td>2.10 x 10^-13</td>
<td>3.22 x 10^-11</td>
</tr>
<tr>
<td>Cadherin-5</td>
<td>Cadherin-5</td>
<td>CDH5</td>
<td>16q21</td>
<td>rs8176746</td>
<td>trans</td>
<td>1.12 (1.05-1.20)</td>
<td>3.78 x 10^-2</td>
<td>0.60</td>
</tr>
<tr>
<td>Desmoglein-2</td>
<td>Desmoglein-2</td>
<td>DS7G2</td>
<td>18q21.2</td>
<td>rs2704050</td>
<td>trans</td>
<td>1.88 (1.59-2.23)</td>
<td>3.31 x 10^-13</td>
<td>3.67 x 10^-11</td>
</tr>
<tr>
<td>IR</td>
<td>Insulin receptor</td>
<td>INSR</td>
<td>19p13.2</td>
<td>rs507666</td>
<td>trans</td>
<td>0.69 (0.63-0.77)</td>
<td>3.59 x 10^-10</td>
<td>4.40 x 10^-11</td>
</tr>
<tr>
<td>DC-SIGN</td>
<td>CD209 antigen</td>
<td>CD209</td>
<td>19p13.2</td>
<td>rs505922</td>
<td>trans</td>
<td>1.32 (1.25-1.40)</td>
<td>5.74 x 10^-21</td>
<td>2.35 x 10^-10</td>
</tr>
<tr>
<td>JAG1</td>
<td>Protein jagged-1</td>
<td>JAG1</td>
<td>20p12.2</td>
<td>rs7041</td>
<td>trans</td>
<td>0.70 (0.65-0.82)</td>
<td>9.08 x 10^-6</td>
<td>7.95 x 10^-4</td>
</tr>
</tbody>
</table>

(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)

<table>
<thead>
<tr>
<th>Protein encoding gene</th>
<th>Protein full name</th>
<th>Type of encoding</th>
<th>Instrument SNP</th>
<th>cis rsID</th>
<th>trans rsID</th>
<th>cis OR (95% CI)</th>
<th>trans OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAM177A1</td>
<td>Protein FAM177A1</td>
<td>cis</td>
<td>rs2608894</td>
<td>1.25 (0.96-1.62)</td>
<td>1.24 (0.95-1.61)</td>
<td>3.78 (2.98-4.76)</td>
<td>3.75 (2.94-4.73)</td>
</tr>
<tr>
<td>C1GALT1</td>
<td>C1GLC</td>
<td>trans</td>
<td>rs2519093</td>
<td>1.06 (0.92-1.23)</td>
<td>1.05 (0.92-1.20)</td>
<td>4.90 (2.93-7.42)</td>
<td>4.89 (2.92-7.38)</td>
</tr>
<tr>
<td>C1GALT1</td>
<td>C1GLC</td>
<td>trans</td>
<td>rs2519093</td>
<td>1.06 (0.92-1.23)</td>
<td>1.05 (0.92-1.20)</td>
<td>4.90 (2.93-7.42)</td>
<td>4.89 (2.92-7.38)</td>
</tr>
</tbody>
</table>

*P value after adjusting for risk SNPs.

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, RMR6, 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.

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 Adhesion G protein–coupled receptor F5 (GPR16), IL6 receptor subunit alpha (gp130, soluble), VEGF receptor 2 (VEGFR2), 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 × 10−7) and STAT3 (P = 5.25 × 10−6; Table 3).
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 *CIGALT1C1*, 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 aE-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 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
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Zhi, X. Guo, C. Wu, E. White, H.A. Risch, 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.p1 and phs006681.v1.p. 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 RO0CA218892. 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 HHSN26120080001E. 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 PAN4 studies: Johns Hopkins National Familial Pancreas Tumor Registry, Mayo Clinic Bioregion 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 HHSN268201100011. 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, HHSN268201600001C, HHSN268201600000C2, 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 CA137429 (to C.D. Blanke), and U54 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


Associations between Genetically Predicted Blood Protein Biomarkers and Pancreatic Cancer Risk

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


Updated version
Access the most recent version of this article at:
doi: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

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/29/7/1501.full#related-urls

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.