A whole-exome case-control association study to characterize the contribution of rare coding variation to pancreatic cancer risk

Yao Yu,1 Kyle Chang,1 Jiun-Sheng Chen,1 Ryan J. Bohlender,1 Jerry Fowler,1 Di Zhang,1 Maosheng Huang,1 Ping Chang,2 Yanan Li,2 Justin Wong,1 Huamin Wang,1 Jian Gu,1 Xifeng Wu,4 Joellen Schildkraut,5 Lisa Cannon-Albright,6,7,8 Yuanqing Ye,4 Hua Zhao,9 Michelle A.T. Hildebrandt,10 Jennifer B. Permuth,11 Donghui Li,2 Paul Scheet,1,* and Chad D. Huff1,*

Summary

Pancreatic cancer is a deadly disease that accounts for approximately 5% of cancer deaths worldwide, with a dismal 5-year survival rate of 10%. Known genetic risk factors explain only a modest proportion of the heritable risk of pancreatic cancer. We conducted a whole-exome case-control sequencing study in 1,591 pancreatic cancer cases and 2,134 cancer-free controls of European ancestry. In our gene-based analysis, ATM ranked first, with a genome-wide significant p value of $1 \times 10^{-8}$. The odds ratio for protein-truncating variants in ATM was 24, which is substantially higher than prior estimates, although ours includes a broad 95% confidence interval (4.0–1000). SIK3 was the second highest ranking gene ($p = 3.84 \times 10^{-6}$, false discovery rate or FDR = 0.032). We observed nominally significant association signals in several genes of a priori interest, including BRCA2 ($p = 4.3 \times 10^{-4}$), STK11 ($p = 0.003$), PALB2 ($p = 0.019$), and TP53 ($p = 0.037$), and reported risk estimates for known pathogenic variants and variants of uncertain significance (VUS) in these genes. The rare variants in established susceptibility genes explain approximately 24% of log familial relative risk, which is comparable to the contribution from established common susceptibility variants (17%). In conclusion, this study provides new insights into the genetic susceptibility of pancreatic cancer, refining rare variant risk estimates in known pancreatic cancer susceptibility genes and identifying SIK3 as a novel candidate susceptibility gene. This study highlights the prominent importance of ATM truncating variants and the underappreciated role of VUS in pancreatic cancer etiology.

Pancreatic ductal adenocarcinoma (PDAC), which accounts for 95% of all diagnosed pancreatic cancers, is a leading cause of cancer-related deaths in the world, with a dismal 5-year survival rate of 10%. Genome-wide association studies (GWAS) of common genetic variation have identified 19 loci associated with pancreatic cancer, which together have been estimated to explain approximately 13% of heritable risk.1 Multiple pancreatic susceptibility genes harboring rare pathogenic variants have been identified from familial studies, highlighting the contribution of rare genetic variation to the genetic architecture of this disease.2 Recent studies have evaluated these genes using targeted gene panels in sporadic pancreatic cancer cases, but interpretation of these results has been complicated by control cohorts that lack cancer history data and are not matched on age or sequencing technology, as well as insufficient genetic information to control for subcontinental population stratification.3–7 To date, the majority of variation contributing to the heritable risk of pancreatic cancer remains unidentified.

Here, we describe a whole-exome case-control study to evaluate the genic-level contribution of rare protein-coding variation to pancreatic cancer risk. We conduct gene-based tests to identify genes with an excess of rare, potentially damaging coding variation among cases, weighting each variant by its estimated degree of dysfunction. We estimate the risks conferred by rare, protein-coding variants according to functional annotation as well as the proportion of pancreatic cancer familial relative risk (FRR) explained by these variants. We also evaluate the patterns of elevated familial risk of other cancers associated with the variants of interest in known and candidate susceptibility genes.

Our case-control study included 1,591 cases with pancreatic cancer ductal adenocarcinoma (PDAC) and 2,134 controls, including 1,591 controls age-matched to cases (+/− 3 years). Participants were recruited at The University of Texas MD Anderson Cancer Center (MDA), H. Lee Moffitt Cancer Center & Research Institute, The University of Utah School of Medicine, and Duke University. All cases

1Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 2Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 3Department of Anatomical Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 4Center for Clinical Big Data and Analytics, Bioinformatics and Big Data, The Second Affiliated Hospital and School of Public Health, Zhejiang University School of Medicine, PR China; 5Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA; 6Genetic Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA; 7George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT, USA; 8 Huntsman Cancer Institute, Salt Lake City, UT, USA; 9Department of Family Medicine and Population Health, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA; 10Department of Lymphoma and Myeloma, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 11Departments of Cancer Epidemiology and Gastrointestinal Oncology, Moffitt Cancer Center, Tampa, FL, USA

*Correspondence: pasheet@mdanderson.org (PS.), chuff1@mdanderson.org (C.D.H.)
https://doi.org/10.1016/j.xhgg.2021.100078.
© 2021 This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Cross-Platform Association Toolkit (XPAT) to perform detect the copy number variations (CNVs). We used variant quality score recalibration. We used XHMM to genome alignment, joint variant genotype calling, and CNVs. We incorporated the first 10 PCs from a principal component analysis and gender as covariates. We observed no inflation in type I error ($\lambda = 1.009$, Figure S5). One gene, ATM, was genome-wide significant after Bonferroni correction (unadjusted $p < 1 \times 10^{-8}$, $\alpha = 0.05/16,721 = 2.99 \times 10^{-6}$, Figure 1A). The association signal in ATM was driven primarily by truncating variants that introduce a stop codon or disrupt splicing structure, with 24 such variants observed in cases compared to one in controls. Additionally, two duplications, spanning multiple ATM exons and likely disrupting a copy of the gene, were present in cases compared to zero in controls (Figure 2B). The second highest ranking gene was SIK3, one of three members of the salt inducible kinase family (SIK1–3). Although not genome-wide significant ($p = 3.84 \times 10^{-6}$), the false discovery rate (FDR) for SIK3 was 0.032.

The top 15 genes in the gene-based analysis with $p \leq 0.001$ are shown in Table 1. This list includes four additional genes frequently amplified or otherwise substantially dysregulated in pancreatic cancer, FZD8 ($p = 1.03 \times 10^{-4}$), $^{13}$ ALPP ($p = 4.2 \times 10^{-4}$), $^{14}$ INPP4A ($p = 1.0 \times 10^{-3}$), $^{15}$ and BRCA2 ($p = 4.25 \times 10^{-4}$). BRCA2 is among the most well studied cancer predisposition genes; pathogenic variants in this gene confer an increased risk of numerous cancers, including pancreatic cancer. In addition to these findings, we observed nominally significant association signals in three additional established pancreatic cancer susceptibility genes, STK11 ($p = 0.003$), PALB2 ($p = 0.019$), and TP53 ($p = 0.037$) (Table S7).

Although the primary goal of this study was to evaluate the genic-level contribution of rare protein-coding variation to pancreatic cancer risk, the data collected also enabled a traditional single marker GWAS of variants across the exome (Figures 1B and Table S8) using logistic regression with the first 10 PCs and gender as covariates. Our results replicated a previously identified association in the TERT region (rs273698, $p = 5.7 \times 10^{-8}$, odds ratio [OR] = 0.69, 95% confidence interval [CI]: 0.60–0.79). We also observed an independent ($r^2 = 0.0023$), nominally significant association for the three base-pair TERT deletion $\Delta$E441 (rs377639087, $p = 5.9 \times 10^{-3}$, OR = 2.9). Excluding the TERT region, the most significant association was a missense variant in PRHOXB (rs9579139, $p = 7.4 \times 10^{-6}$, OR = 0.79). This variant is within 600,000 base pairs...
of a prior GWAS association identified near PDX1, although it is not in linkage disequilibrium ($r^2 = 0.029$) with the index SNP (rs9581943).

To better characterize the contribution of rare genetic variation to pancreatic cancer risk, we evaluated variant effect sizes according to a variety of functional classifications, considering their amino acid changes, affected protein domains, in silico functional predictions, and variant annotation status in ClinVar (Version 20,190,916). For ClinVar annotations, we included likely pathogenic variants in the known pathogenic category, likely benign in the benign category, and uncertain or unknown significance as VUS. We also classified variants absent from ClinVar as VUS. Because our goal was to estimate risk rather than identify associations, all risk estimates are derived from comparisons with our age-matched controls, except where specified (Table S9). We calculated ORs and CIs using logistic regression, using the first 10 PCs and gender as covariates. For categories with fewer than three allele copies in either cases or controls, we calculated ORs and CIs using a Fisher's Exact Test. Generally, across known pancreatic cancer susceptibility genes, we observed that VUS with higher in silico assessments of pathogenicity tended to have higher contributions to pancreatic cancer risk (Figures S6–S8).

The OR for all ATM coding variants was 1.8 (95% CI: 1.4–2.3) (Figure 3A and Table 2). We observed that truncating variants in ATM conferred a surprisingly high risk of pancreatic cancer, with an OR of 24 (95% CI: 4.0–1000). To confirm the carrier frequency estimate for truncating variants in ATM among cases, we validated all 24 SNVs and small INDELs using Sanger sequencing (Table S10). In contrast, the OR for known pathogenic missense variants in ATM was only 2.0 (95% CI: 0.29–22) while the OR for VUS was 1.4 (95% CI: 1.1–1.9). Because pathogenic missense variants are known to be highly enriched in ATM protein domains (TAN, FAT, PI3/4K, and FATC), we evaluated missense variants in these regions separately (Figure 2B). VUS in these regions exhibited reduced effect sizes relative to truncating and pathogenic variants, with an OR of 2.7 (95% CI: 1.2–5.9) in the four protein domains. Although pathogenic variants have not previously been reported to be enriched in ATM armadillo (ARM) repeat regions, we observed a significant excess of VUS in these regions ($p = 0.0042$, OR = 1.8 [95% CI: 1.2–2.8]). We observed no evidence of association for benign missense variants or VUS outside of domain or ARM repeat regions (Figure S9).

The OR for all rare missense variants in SIK3 was 2.9 (95% CI: 1.5–5.9) (Figure 3B). Variants in the two domains of SIK3, STKc and UBA, were enriched among cases, particularly in the UBA domain (Figure 2C). The effect size estimates for rare variants in BRCA2 were largely consistent with previous reports, with an OR of 4.0 (95% CI: 1.1–9.2) for truncating variants and no evidence of enrichment among cases for other variant categories (Figure 3C). VUS in the protein domain regions of STK11 and PALB2 were enriched among cases, with ORs of 5.0 (95% CI: 0.56–240) and 2.0 (95% CI: 0.76–5.0), respectively. VUS outside of the protein domains of STK11 and PALB2 exhibited no elevated risk (Figures 3D and 3E). Although the gene-based association test was not significant for CHEK2, the effect size estimates for VUS in CHEK2 were consistent with recent reports suggestive of increases in risk (OR = 1.7, 95% CI: 0.83–3.4). We identified no truncating variants in TPS3 but observed an

Figure 2. Lollipots of rare variants in ATM, SIK3, and TP53. (A) Rare truncating variants and CNVs in ATM, (B) rare VUS and missense variants in ATM, and rare coding variants in (C) SIK3 and (D) TP53. Each lollipop represents one variant in ATM. The upper (lower) area presents the variants identified in cases (controls). The number in each dot represents the number of carriers of each variant. The axis (height of dots) presents the conservation-controlled AAS matrix scores of variants calculated using VAAST 2, based on amino acid substitution severity and phylogenetic conservation. The plot was made by an R package named “trackViewer.” Domain structures were obtained from InterPro (https://www.ebi.ac.uk/interpro/) and NCBI (https://www.ncbi.nlm.nih.gov/).
enrichment for VUS among cases, with an OR of 3.1 (95% CI: 0.82–11) (Figures 2D and 3F). BRCA1 variants exhibited no evidence of increased risk, with an OR upper bound for truncating variants of 1.8 at the 95% confidence level (Figure 3G).

We also evaluated the patterns of familial cancer risk associated with the variants of interest identified in this study (Table S11). Overall, the relative risk (RR) of a positive family history of pancreatic cancer among cases was 3.3 relative to controls, with positive family history defined as a cancer diagnosis in a first- or second-degree relative. A positive family history of breast cancer was enriched among TP53 VUS carriers and SIK3 carriers as well as ATM and BRCA2 truncating and pathogenic missense variants carriers, although only ATM and BRCA2 were nominally significant (Figure 3H and Tables S12–S14). SIK3 carriers exhibited a modest familial enrichment for colon cancer, leukemia, and non-specific lymphoma, although the enrichment was not statistically significant. We also observed unexpected, nominally significant familial enrichment among BRCA2 carriers for kidney cancer and cervical cancer. All cases with a VUS in TP53 carriers had a positive family history in one or more cancers, and in general, the patterns of elevated familial risk were consistent with Li-Fraumeni syndrome (LFS) (Figure S10).

We estimated the proportion of pancreatic cancer FRR explained by known genetic risk factors with an approach used by previous studies, assuming an overall RR of PDAC to a first-degree relative of 1.7623 (for details, see supplemental methods). We combined estimates from the genes characterized in this study with common susceptibility alleles identified in published GWAS (Table 3). We estimated the proportion of log FRR explained by all

Table 1. Top 15 and known PDAC genes in case-control association tests using VAAST2

| Gene   | Rank | p Value    | FDR         |
|--------|------|------------|-------------|
| ATM    | 1    | 1.00 x 10^-8 | 0.0002      |
| SIK3   | 2    | 3.84 x 10^-6 | 0.0321      |
| FZD8   | 3    | 1.03 x 10^-4 | 0.5741      |
| PANX1  | 4    | 1.79 x 10^-4 | 0.7053      |
| PRPF3  | 5    | 3.42 x 10^-4 | 0.7053      |
| C10orf140 | 6  | 3.54 x 10^-4 | 0.7053      |
| FRMD4B | 7    | 4.08 x 10^-4 | 0.7053      |
| ALPP   | 8    | 4.20 x 10^-4 | 0.7053      |
| BRCA2  | 9    | 4.25 x 10^-4 | 0.7053      |
| GPC2   | 10   | 4.44 x 10^-4 | 0.7053      |
| AK3    | 11   | 4.64 x 10^-4 | 0.7053      |
| OSGIN2 | 12   | 5.72 x 10^-4 | 0.7357      |
| IGFLR1 | 13   | 5.72 x 10^-4 | 0.7357      |
| SMARCAL1 | 14 | 8.01 x 10^-4 | 0.8820      |
| INPP4A | 15   | 1.00 x 10^-3 | 0.8820      |
| STK11  | 45   | 0.0030      | 0.8820      |
| PMS1   | 144  | 0.0082      | 0.8820      |
| PALB2  | 361  | 0.0194      | 0.8900      |
| TPS3   | 666  | 0.0368      | 0.9121      |
| MSH2   | 1241 | 0.0709      | 0.9495      |
| MLH1   | 2525 | 0.1480      | 0.9639      |
| CHEK2  | 3881 | 0.2270      | 0.9659      |
| MSH6   | 3939 | 0.2300      | 0.9659      |
| SPINK1 | 4330 | 0.2520      | 0.9720      |
| CDKN2A | 8369 | 0.4980      | 0.9903      |
| PMS2   | 14,720 | 0.8900   | 1.0000      |
| BRCA1  | 14,921 | 0.9070   | 1.0000      |
| PRSS1  | 15,497 | 0.9470   | 1.0000      |
common variants that have been identified with genomewide statistical significance in one or more GWAS to be 16.9% (Table S15), which is consistent with a recent prior estimate (13% with 95% CI: 4–22%). Our point estimate for the proportion of log FRR explained by truncating variants in ATM is 16.6%. Although this estimate is imprecise given uncertainty in effect sizes, ATM truncating variants explain at least 0.4% of log FRR with 95% confidence. We estimate that known pathogenic missense variants in ATM account for 0.1% of log FRR, while VUS in protein domains and ARM regions account for 2.3%, strongly suggesting that most pathogenic missense variants in ATM are either currently classified as VUS or have yet to be identified. Truncating variants in BRCA2 and PALB2 explain an estimated 3.0% and 0.5% of log FRR, respectively, while VUS in TP53, PALB2, and STK11 explain 0.6%, 0.3%, and 0.2% of log FRR, respectively. Overall, we estimate that rare variants in ATM, BRCA2, TP53, PALB2, STK11, and CHEK2 explain 23.8% of log FRR.

Our OR estimate for the risk conferred by heterozygote ATM truncating variants (OR = 24.3) is consistent with a study of 593 familial PDAC families (OR = 31.9; Table S16), although the effect sizes from this study are expected to be higher due to familial enrichment. Other prior studies that did not utilize cancer-free controls matched on technology, population, or age have reported ORs between 5 and 9.3. Our proportion of cases carrying ATM truncating variants (1.5%) is comparable to those reported in previous studies of PDAC ranging from 0.9% to 3.2% (Table S17). Our proportion of controls carrying truncating variants (0.06%) is substantially lower than in prior studies (0.1%–0.3%) and public databases, although this is to some extent expected given that our controls were age-matched and cancer-free (Tables S16 and S18). In contrast to truncating variants, the ORs we observed for known pathogenic missense variants and VUS in domain regions were only 2.0 and 1.4, respectively. An attenuation in risk for pathogenic missense variants relative to pathogenic truncating
| Gene  | Variant type                                                      | Number of case carriers (N = 1,591) | Number of control carriers OR (95% CI) | Matched controls (N = 1,591) | OR (95% CI) | All controls (N = 2,134) | OR (95% CI) |
|-------|------------------------------------------------------------------|------------------------------------|---------------------------------------|-----------------------------|-------------|--------------------------|-------------|
| ATM   | All variants and CNVs                                           | 184                                | 1.77 (1.37–2.29)                      | 148                         | 1.71 (1.35–2.15) |                         |             |
|       | Truncating, CNV, and pathogenic                                  | 30                                 | 10.49 (3.17–34.77)                    | 3                           | 14.14 (4.28–46.66) |                         |             |
|       | Truncating and CNV                                              | 26                                 | 26.4 (4.33–1078.32)                   | 1                           | 26.3 (4.31–1073.81) |                         |             |
|       | CNV event                                                        | 2                                  | Inf (0.19–Inf)                        | 0                           | Inf (0.25–Inf)   |                         |             |
|       | Protein truncating                                              | 24                                 | 24.34 (3.96–997.52)                   | 1                           | 32.65 (5.31–1335.58) |                         |             |
|       | Pathogenic Missense                                             | 4                                  | 2.0 (0.29–22.16)                      | 2                           | 2.69 (0.38–29.73) |                         |             |
|       | Missense VUS                                                   | 134                                | 1.44 (1.08–1.92)                      | 124                         | 1.42 (1.1–1.85)  |                         |             |
|       | Inside domain                                                   | 27                                 | 2.66 (1.2–5.87)                       | 15                          | 2.2 (1.14–4.22)  |                         |             |
|       | In ARM repeat                                                   | 63                                 | 1.82 (1.19–2.78)                      | 49                          | 1.76 (1.2–2.59)  |                         |             |
|       | Outside domain/ARM repeat                                       | 55                                 | 0.91 (0.61–1.36)                      | 73                          | 0.93 (0.65–1.34)  |                         |             |
|       | Benign missense                                                 | 24                                 | 1.63 (0.84–3.19)                      | 24                          | 1.34 (0.75–2.39)  |                         |             |
| SIK3  | All variants and CNVs                                           | 36                                 | 3.03 (1.51–6.06)                      | 19                          | 2.49 (1.41–4.4)  |                         |             |
|       | Truncating                                                      | 1                                  | Inf (0.03–Inf)                        | 0                           | Inf (0.03–Inf)   |                         |             |
|       | Missense                                                        | 35                                 | 2.94 (1.47–5.91)                      | 19                          | 2.42 (1.37–4.3)  |                         |             |
|       | Missense inside domain                                          | 8                                  | Inf (1.71–Inf)                        | 0                           | Inf (2.3–Inf)    |                         |             |
|       | Missense outside domain                                         | 28                                 | 2.37 (1.15–4.88)                      | 19                          | 1.96 (1.08–3.56)  |                         |             |
| BRCA2 | All variants and CNVs                                           | 210                                | 1.04 (0.84–1.28)                      | 253                         | 1.11 (0.91–1.36)  |                         |             |
|       | Truncating and pathogenic                                        | 35                                 | 3.98 (1.73–9.15)                      | 9                           | 4.56 (2.16–9.6)  |                         |             |
|       | Truncating                                                      | 35                                 | 3.98 (1.73–9.15)                      | 9                           | 4.56 (2.16–9.6)  |                         |             |
|       | NM_000059.4: c.9976A>T (p.Lys3326Ter)                            | 34                                 | 1.24 (0.75–2.05)                      | 47                          | 1.05 (0.67–1.65)  |                         |             |
|       | Pathogenic                                                      | 0                                  | NA                                   | 0                           | NA             |                         |             |
|       | Missense VUS                                                   | 173                                | 0.94 (0.75–1.18)                      | 234                         | 0.99 (0.8–1.22)  |                         |             |
|       | Missense VUS inside domain                                      | 42                                 | 0.84 (0.55–1.28)                      | 69                          | 0.83 (0.56–1.23)  |                         |             |
|       | Missense VUS outside domain                                     | 133                                | 0.98 (0.76–1.28)                      | 167                         | 1.06 (0.83–1.35)  |                         |             |
|       | Benign missense                                                 | 9                                  | 0.68 (0.28–1.65)                      | 15                          | 0.77 (0.33–1.8)  |                         |             |
| STK11 | All variants and CNVs                                           | 25                                 | 1.42 (0.76–2.65)                      | 29                          | 1.19 (0.69–2.06)  |                         |             |
|       | Truncating, CNV, and pathogenic                                  | 0                                  | NA                                   | 0                           | NA             |                         |             |
|       | Missense VUS                                                   | 25                                 | 1.42 (0.76–2.65)                      | 29                          | 1.19 (0.69–2.06)  |                         |             |
|       | Missense VUS inside domain                                      | 5                                  | 5.01 (0.56–236.97)                    | 2                           | 3.36 (0.55–35.33) |                         |             |
|       | Missense VUS outside domain                                     | 20                                 | 1.21 (0.62–2.36)                      | 27                          | 1.03 (0.57–1.86)  |                         |             |
|       | Benign missense                                                 | 0                                  | NA                                   | 0                           | NA             |                         |             |
| PALB2 | All variants and CNVs                                           | 62                                 | 1.54 (1.01–2.34)                      | 57                          | 1.43 (0.98–2.07)  |                         |             |
|       | Truncating                                                      | 9                                  | 2.76 (0.73–10.48)                     | 4                           | 2.88 (0.88–9.5)  |                         |             |
|       | Missense VUS                                                   | 35                                 | 1.48 (0.86–2.57)                      | 33                          | 1.41 (0.87–2.3)  |                         |             |
|       | Missense VUS inside domain                                      | 13                                 | 1.96 (0.76–5.04)                      | 9                           | 2.08 (0.87–4.94)  |                         |             |
|       | Missense VUS outside domain                                     | 23                                 | 1.36 (0.7–2.64)                       | 24                          | 1.24 (0.69–2.22)  |                         |             |
|       | Benign missense                                                 | 18                                 | 1.18 (0.57–2.46)                      | 21                          | 1.06 (0.56–2.03)  |                         |             |

(Continued on next page)
variants has not been reported in other ATM-associated cancers and suggest that ATM truncating variants play a unique role in pancreatic cancer susceptibility.

In our gene-based analysis, SIK3 ranked second genome-wide with an FDR of 0.032. The effect size we observed for all missense variants with MAF < 0.5% in SIK3 (OR = 2.9) is comparable to an estimate of predicted rare damaging missense variants with MAF < 1% among cases from a prior case-control study of PDAC \(^4\) (OR = 1.8) (Table S19). SIK1–3 are negatively regulated by GNAS, a gene frequently mutated and amplified in pancreatic cancer.\(^27\) Murine models have shown that SIK1–3 genes are pancreatic cancer tumor suppressors, and that inhibition of SIK1–3 activity is an important mechanism through which mutant GNAS promotes pancreatic cancer tumorigenesis.\(^28\) SIK3 has also been implicated in obesity and diabetes, two well-established pancreatic cancer risk factors.\(^29,30\). Specifically, reduced SIK3 expression has been associated with obesity and insulin resistance,\(^31,32\) and an intronic SIK3 variant has been associated with obesity and dyslipidemia in a Mexican population.\(^33\) Given these observations, we also anticipated that tumor SIK3 expression would be associated with measures of pancreatic cancer tumorigenesis. We confirmed this expectation in an analysis of TCGA PDAC RNA-sequencing data clinical data of TCGA samples from the Human Protein Atlas\(^34\) (for details, see supplemental methods), identifying associations with increased SIK3 expression and lower tumor stage (p < 2.3 × 10\(^{-4}\)) as well as increased overall survival (p = 0.037) controlling for stage, gender, and age (Table S20).

We also observed suggestive association signals (p < 0.001) for three additional novel candidate genes dysregulated in pancreatic cancer, FZD8, ALPP, and INPP4A. FZD8 is downregulated in KRAS mutant pancreatic cancer; restoration of FZD8 expression has been shown to suppress malignancy in pancreatic cancer cell lines.\(^35\) Other studies have shown that ALPP is epigenetically silenced\(^14\) and

### Table 2. Continued

| Gene     | Variant type | Number of case carriers (N = 1,591) | Number of control carriers (N = 1,591) | Matched controls (N = 1,591) | All controls (N = 2,134) |
|----------|--------------|------------------------------------|----------------------------------------|----------------------------|--------------------------|
|          |              | Number of control carriers OR (95% CI) | Number of control carriers OR (95% CI) |                            |                          |
|          |              |                                     |                                        |                            |                          |
| TP53     | All variants and CNVs | 11 | 3 | 3.38 (0.92–12.38) | 4 | 3.59 (1.13–11.44) |
|          | Truncating, CNV, and pathogenic | 0 | 0 | NA | 0 | NA |
|          | Missense VUS inside domain | 6 | 3 | 1.97 (0.48–8.09) | 4 | 2.08 (0.58–7.47) |
|          | Missense VUS outside domain | 4 | 0 | Inf (0.66–Inf) | 0 | Inf (0.89–Inf) |
|          | Benign missense | 1 | 0 | Inf (0.03–Inf) | 0 | Inf (0.03–Inf) |
| CHEK2    | All variants and CNVs | 41 | 21 | 1.82 (1.05–3.15) | 38 | 1.26 (0.8–1.99) |
|          | Truncating, CNV, and pathogenic | 15 | 8 | 1.9 (0.78–4.61) | 16 | 1.21 (0.59–2.49) |
|          | CNV | 1 | 1 | 1.0 (0.01–78.5) | 3 | 0.45 (0.01–5.57) |
|          | Truncating | 12 | 7 | 1.7 (0.65–4.43) | 12 | 1.28 (0.57–2.91) |
|          | NM_007194.4: c.1100del (p.Thr367fs) | 11 | 5 | 2.31 (0.78–6.83) | 9 | 1.60 (0.65–3.94) |
|          | Pathogenic | 2 | 0 | Inf (0.19–Inf) | 1 | 2.68 (0.14–158.31) |
|          | Missense VUS inside domain | 19 | 11 | 1.42 (0.65–3.09) | 17 | 1.19 (0.6–2.34) |
|          | Missense VUS outside domain | 6 | 2 | 3.01 (0.54–30.51) | 2 | 4.03 (0.72–40.93) |
|          | Benign missense | 1 | 0 | Inf (0.03–Inf) | 3 | 0.45 (0.01–5.57) |
| BRCA1    | All variants and CNVs | 68 | 72 | 0.91 (0.64–1.29) | 104 | 0.85 (0.62–1.17) |
|          | Truncating, CNV, and pathogenic | 7 | 8 | 0.59 (0.19–1.84) | 13 | 0.51 (0.19–1.35) |
|          | CNV | 0 | 0 | NA | 1 | 0.0 (0.0–52.27) |
|          | Truncating | 7 | 8 | 0.59 (0.19–1.84) | 12 | 0.54 (0.2–1.46) |
|          | Pathogenic | 0 | 0 | NA | 0 | NA |
|          | Missense VUS inside domain | 18 | 24 | 0.78 (0.42–1.47) | 37 | 0.67 (0.38–1.19) |
|          | Missense VUS outside domain | 35 | 35 | 0.96 (0.59–1.57) | 47 | 0.99 (0.63–1.55) |
|          | Benign missense | 8 | 7 | 1.12 (0.39–3.2) | 9 | 1.2 (0.46–3.17) |
INPP4A is markedly downregulated\textsuperscript{15} in pancreatic cancer cell lines. The inactivation of INPP4A has been shown to promote cell migration and inhibit cell apoptosis.\textsuperscript{15} Additionally, our OR estimate for missense variants with MAF <0.5% in ALPP (OR = 1.9) is consistent with the estimate for predicted damaging missense variants with MAF <1% among cases from a prior study\textsuperscript{4} (OR = 2.2; Table S19). In addition to our gene-based results, we also observed a nominally significant association with the three-base-pair TERT deletion ΔE441 (p = 5.9 × 10^{-3}, OR = 2.9). Although we did not identify this variant as candidate \textit{a priori}, it is known to causally reduce telomerase activity and is associated with acute myeloid leukemia and liver cirrhosis.\textsuperscript{36,37}

Previous studies have observed an approximate 7-fold increased risk of developing pancreatic cancer in patients with LFS, which is caused by pathogenic variants in \textit{TP53}.\textsuperscript{38–41} The mutational pattern of pathogenic LFS variants is distinct, with approximately 27% predicted to truncate the protein and an additional 39% concentrated within seven codons (37, 125, 158, 175, 248, 273, and 282), four of which are in the DNA binding domain.\textsuperscript{42} In our study, we observed no variants within these seven codons, no truncating variants, and no enrichment for variants in the DNA binding domain. Nevertheless, overall, VUS were moderately enriched among cases (OR = 3.1). The observed dearth of truncating variants and lack of codon enrichment in our study suggests that the spectrum of pathogenic \textit{TP53} variants in pancreatic cancer may be distinct from LFS broadly, particularly for individuals without a prior cancer diagnosis.

We acknowledge that our OR and FRR estimates for \textit{ATM} truncating variants should be interpreted with cautious intrigue given the relatively low number of variant observations and the resulting wide CIs. The divergence in our estimates and prior studies may be due in part to the lack of phenotype information and the inability to control for population structure and technology differences in case-only candidate gene sequencing studies. GWAS have shown the importance of conducting matched case-control studies that control for technological and population stratification biases. Failure to properly match on these features could move estimates lower or higher, depending on the underlying confounding factors. Although these factors may have elucidated a heretofore underappreciated effect of truncation coding variation in \textit{ATM}, subsequent larger, well-matched studies will refine these estimates with greater precision that may edge again closer to prior estimates. Such large-scale studies are needed, in populations of European ancestry and particularly in underrepresented populations, to improve variant pathogenicity classification and more accurately estimate the risks conferred by rare variants in established susceptibility genes. Expansion of these studies from cancer gene panels to whole exomes or genomes will enable the identification of novel susceptibility genes and will provide the evidence needed to critically evaluate candidate susceptibility genes, including SIK3. Together, these efforts will lead to improvements in risk stratification for early detection and screening programs in pancreatic cancer.

In conclusion, our study identifies SIK3 as a novel candidate pancreatic cancer susceptibility gene and highlights the risk contributions from VUS as well as the prominent importance of \textit{ATM} truncating variants in pancreatic cancer etiology. Our results also demonstrate that rare protein-coding variants account for a substantial fraction of the familial risk of pancreatic cancer.

Data and code availability
The data from this study can be accessed through dbGaP (accession number pending).

Supplemental information
Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2021.100078.
Acknowledgments

This work was supported by US National Institutes of Health grants R01 CA181244, R01 CA195614, R01 CA188943, R01 CA177935, 1S00OD024977, and P30 CA016672 (ATGC) and in part by Tissue Core Shared Resource at the H. Lee Moffitt Cancer Center & Research Institute, an NCI designated Comprehensive Cancer Center (P30 CA076292). The author(s) acknowledge the support of the High-Performance Computing for research facility at the University of Texas MD Anderson Cancer Center for providing computational resources that have contributed to the research results reported in this paper.

Declaration of interests

The authors declare no competing interests.

Received: July 13, 2021
Accepted: December 8, 2021

References

1. Klein, A.P., Wolpin, B.M., Risch, H.A., Stolzenberg-Solomon, R.Z., Mocci, E., Zhang, M., Canzian, F., Childs, E.J., Hoskins, J.W., Jermusyk, A., et al. (2018). Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat. Commun. 9, 556.

2. Zhan, W., Shelton, C.A., Greer, P.J., Brand, R.E., and Whitcomb, D.C. (2018). Germline variants and risk for pancreatic cancer: a systematic review and emerging concepts. Pancreas 47, 924–936.

3. Hu, C., Hart, S.N., Polley, E.C., Gnaolivu, R., Shimelis, H., Lee, K.Y., Lilyquist, J., Na, J., Moore, R., Antwi, S.O., et al. (2018). Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. JAMA 319, 2401–2409.

4. Grant, R.C., Denroche, R.E., Borgida, A., Virtanen, C., Cook, N., Smith, A.L., Connor, A.A., Wilson, J.M., Petersen, G., Roberts, N.J., et al. (2018). Exome-wide association study of pancreatic cancer risk. Gastroenterology 154, 719–722.e3.

5. Hu, C., LaDuca, H., Shimelis, H., Polley, E.C., Lilyquist, J., Hart, S.N., et al. (2018). Multigene hereditary cancer panels reveal high-risk pancreatic cancer susceptibility genes. JCO Precis. Oncol. 2, 1–28.

6. Roberts, N.J., Norris, A.L., Petersen, G.M., Bondy, M.L., Brand, R., Gallinger, S., Kurtz, R.C., Olson, S.H., Rustgi, A.K., Schwartz, A.G., et al. (2016). Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. Cancer Discov. 6, 166–175.

7. Shindoh, K., Yu, J., Suenaga, M., Fesharikazadeh, S., Cho, C., Macgregor-Das, A., Siddiqui, A., Wittmer, P.D., Tamura, K., Song, T.J., et al. (2017). Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J. Clin. Oncol. 35, 3382–3388.

8. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.

9. Fromer, M., and Purcell, S.M. (2014). Using XHMM software to detect copy number variation in whole-exome sequencing data. Curr. Protoc. Hum. Genet. 81, 7.23.1–7.23.21.

10. Yu, Y., Hu, H., Bohler, R.J., Hu, F., Chen, J.S., Holt, C., Fowler, J., Guthery, S.L., Scheet, P., Hildebrandt, M.A.T., et al. (2018). XPAT: a toolkit to conduct cross-platform association studies with heterogeneous sequencing datasets. Nucleic Acids Res. 46, e32.

11. Hu, H., Huff, C.D., Moore, B., Flygare, S., Reese, M.G., and Yandell, M. (2013). Vaast 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. Genet. Epidemiol. 37, 622–634.

12. Yandell, M., Huff, C., Hu, H., Singleton, M., Moore, B., Xing, J., Jorde, L.B., and Reese, M.G. (2011). A probabilistic disease-gene finder for personal genomes. Genome Res. 21, 1529–1542.

13. Li, Q., Ye, L., Zhang, X., Wang, M., Lin, C., Huang, S., Guo, W., Lai, Y., Du, H., Li, J., et al. (2017). FZD8, a target of p53, promotes bone metastasis in prostate cancer by activating canonical Wnt/beta-catenin signaling. Cancer Lett. 402, 166–176.

14. Kumagai, T., Akagi, T., Desmond, J.C., Kawamura, N., Gery, S., Imai, Y., Song, J.H., Gui, D., Said, J., and Koefler, H.P. (2009). Epigenetic regulation and molecular characterization of C/EBPalpha in pancreatic cancer cells. Int. J. Cancer 124, 827–833.

15. Wang, C., Feng, Z., Jiang, K., and Zuo, X. (2017). Upregulation of MicroRNA-935 promotes the malignant behaviors of pancreatic carcinoma PANC-1 cells via targeting inositol polyphosphate 4-phosphatase Type I gene (INPP4A). Oncol. Res. 25, 559–569.

16. Iqbal, J., Ragone, A., Lubinski, J., Lynch, H.T., Molli, P., Ghadirian, P., Foulkes, W.D., Armel, S., Eisen, A., Neuhausen, S.L., et al. (2012). The incidence of pancreatic cancer in BRCA1 and BRCA2 mutation carriers. Br. J. Cancer 107, 2005–2009.

17. Breast Cancer Linkage Consortium (1999). Cancer risks in BRCA2 mutation carriers. J. Natl. Cancer Inst. 91, 1310–1316.

18. Mohelnikova-Duchonova, B., Havranek, O., Hlavata, I., Foretova, L., Klebl, Z., Pohlreich, P., and Soucek, P. (2010). CHEK2 gene alterations in the forkhead-associated domain, 1109delC and del5395 do not modify the risk of sporadic pancreatic cancer. Cancer Epidemiol. 34, 656–658.

19. Obazee, O., Archibugi, L., Andrulli, A., Soucek, P., Malecka-Panas, E., Ivanaukas, A., Johnson, T., Gazouli, M., Pausch, T., Lawlor, R.T., et al. (2019). Germline BRCA2 K3326X and CHEK2 I157T mutations increase risk for sporadic pancreatic ductal adenocarcinoma. Int. J. Cancer 145, 686–693.

20. Frank, C., Sundquist, J., Hemminki, A., and Hemminki, K. (2017). Risk of other cancers in families with melanoma: novel familial links. Sci. Rep. 7, 42601.

21. Jenkins, M.A., Makalic, E., Dowty, J.G., Schmidt, D.F., Dite, G.S., MacInnis, R.J., Ait Ouakrim, D., Clendenning, M., Flander, L.B., Stanesby, O.K., et al. (2016). Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening. Future Oncol. 12, 503–513.

22. Bahcall, O. (2013). Common variation and heritability estimates for breast, ovarian and prostate cancers. Nat. Genet. 35, 23–25.

23. Jacobs, E.J., Chanock, S.J., Fuchs, C.S., Lacroix, A., McWilliams, R.R., Steplewski, E., Stolzenberg-Solomon, R.Z., Arslan, A.A., Bueno-de-Mesquita, H.B., Gross, M., et al. (2010). Family history of cancer and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium (PanScan). Int. J. Cancer 127, 1421–1428.

24. Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-Freeman, L., Bracci, ...
P.M., Buring, J., et al. (2014). Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. 46, 994–1000.

25. Helgason, H., Rafnar, T., Olafsdottir, H.S., Jonasson, J.G., Sigurdsson, A., Stacey, S.N., Jonasdottir, A., Tryggyadottir, L., Alexiusdottir, K., Haraldsson, A., et al. (2015). Loss-of-function variants in ATM confer risk of gastric cancer. Nat. Genet. 47, 906–910.

26. Roberts, N.J., Jiao, Y., Yu, J., Kopelovich, L., Petersen, G.M., Bondy, M.L., Gallinger, S., Schwartz, A.G., Syngal, S., Cote, M.L., et al. (2012). ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2, 41–46.

27. Hollstein, P.E., and Shaw, R.J. (2018). GNAS shifts metabolism in pancreatic cancer. Nat. Cell Biol. 20, 740–741.

28. Patra, K.C., Kato, Y., Mizukami, Y., Widholz, S., Boukhali, M., Revenco, I., Grossman, E.A., Ji, F., Sadreyev, R.I., Liss, A.S., et al. (2018). Mutant GNAS drives pancreatic tumourigenesis by inducing PKA-mediated SIK suppression and reprogramming lipid metabolism. Nat. Cell Biol. 20, 811–822.

29. Berrington de Gonzalez, A., Sweetland, S., and Spencer, E. (2003). A meta-analysis of obesity and the risk of pancreatic cancer. Br. J. Cancer 89, 519–523.

30. Hassan, M.M., Bondy, M.L., Wolff, R.A., Abbruzzese, J.L., Vauthey, J.N., Pisters, P.W., Evans, D.B., Khan, R., Chou, T.H., Lenzi, R., et al. (2007). Risk factors for pancreatic cancer: case-control study. Am. J. Gastroenterol. 102, 2696–2707.

31. Horike, N., Takemori, H., Katoh, Y., Doi, J., Min, L., Asano, T., Sun, X.J., Yamamoto, H., Kasayama, S., Muraoka, M., et al. (2003). Adipose-specific expression, phosphorylation of Ser794 in insulin receptor substrate-1, and activation in diabetic animals of salt-inducible kinase-2. J. Biol. Chem. 278, 18440–18447.

32. Sall, J., Pettersson, A.M., Bjork, C., Henriksson, E., Wassersstrom, S., Linder, W., Zhou, Y., Hansson, O., Andersson, D.P., Eklund, M., et al. (2017). Salt-inducible kinase 2 and -3 are downregulated in adipose tissue from obese or insulin-resistant individuals: implications for insulin signalling and glucose uptake in human adipocytes. Diabetologia 60, 314–323.

33. Ko, A., Cantor, R.M., Weissglas-Volkov, D., Nikkola, E., Reddy, P.M., Sinzheimer, J.S., Pasaniuc, B., Brown, R., Alvarez, M., Rodrigoz, A., et al. (2014). Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat. Commun. 5, 3983.

34. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Stjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.

35. Wang, M.T., Holderfield, M., Galeas, J., Delrosario, R., To, M.D., Balmain, A., and McCormick, F. (2015). K-ras promotes tumorigenicity through suppression of non-canonical wnt signaling. Cell 163, 1237–1251.

36. Calado, R.T., Regal, J.A., Hills, M., Yewedell, W.T., Dalmazzo, L.F., Zago, M.A., Lansdorp, P.M., Hogge, D., Chanock, S.J., Estey, E.H., et al. (2009). Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. Proc. Natl. Acad. Sci. U S A 106, 1187–1192.

37. Calado, R.T., Brudno, J., Mehta, P., Kovacs, J.J., Wu, C., Zago, M.A., Chanock, S.J., Boyer, T.D., and Young, N.S. (2011). Constitutional telomerase mutations are genetic risk factors for cirrhosis. Hepatology 53, 1600–1607.

38. Varley, J.M. (2003). Germline TP53 mutations and Li-Fraumeni syndrome. Hum. Mutat. 21, 313–320.

39. Sivarastava, S., Zou, Z.Q., Pirollo, K., Blattner, W., and Chang, E.H. (1990). Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. Nature 348, 747–749.

40. Frebourg, T., Barbier, N., Yan, Y.X., Garber, J.E., Dreyfus, M., Fraumeni, J., Jr., Li, F.P., and Friend, S.H. (1995). Germ-line p53 mutations in 15 families with Li-Fraumeni syndrome. Am. J. Hum. Genet. 56, 608–615.

41. Ruijs, M.W., Verhoef, S., Rookus, M.A., Pruntu, R., van der Hout, A.H., Hogervorst, F.B., Kluijt, I., Sijmons, R.H., Aalfs, C.M., Wagner, A., et al. (2010). TP53 germ-line mutation testing in 180 families suspected of Li-Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes. J. Med. Genet. 47, 421–428.

42. Bougeard, G., Renaux-Petel, M., Flaman, J.M., Charbonnier, C., Ferney, P., Belotti, M., Gauthier-Villars, M., Stoppa-Lyonnet, D., Consolino, E., Brugieres, L., et al. (2015). Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. J. Clin. Oncol. 33, 2345–2352.