Am J Cancer Res 2020;10(7):2128-2144
www.ajcr.us /ISSN:2156-6976/ajcr0116403

Original Article

Associations of novel variants in *PIK3C3*, *INSR* and *MAP3K4* of the ATM pathway genes with pancreatic cancer risk

Ling-Ling Zhao[1,2,3], Hong-Liang Liu[2,3], Sheng Luo[4], Kyle M Walsh[2,5], Wei Li[1], Qingyi Wei[2,3,6]

[1]Cancer Center, The First Hospital of Jilin University, Changchun, China; [2]Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA; Departments of [3]Medicine, [4]Biostatistics and Bioinformatics, [5]Neurosurgery, [6]Population Health Sciences, Duke University School of Medicine, Durham, NC, USA

Received June 16, 2020; Accepted June 21, 2020; Epub July 1, 2020; Published July 15, 2020

Abstract: The ATM serine/threonine kinase (ATM) pathway plays important roles in pancreatic cancer (PanC) development and progression, but the roles of genetic variants of the genes in this pathway in the etiology of PanC are unknown. In the present study, we assessed associations between 31,499 single nucleotide polymorphisms (SNPs) in 198 ATM pathway-related genes and PanC risk using genotyping data from two previously published PanC genome-wide association studies (GWASs) of 15,423 subjects of European ancestry. In multivariable logistic regression analysis, we identified three novel independent SNPs to be significantly associated with PanC risk [*PIK3C3* rs76692125 G>A: odds ratio (OR)=1.26, 95% confidence interval (CI)=1.12-1.43 and $P=2.07\times10^{-4}$, *INSR* rs11668724 G>A: OR=0.89, 95% CI=0.84-0.94 and $P=4.21\times10^{-5}$ and *MAP3K4* rs13207108 C>T: OR=0.83, 95% CI=0.75-0.92, $P=2.26\times10^{-4}$]. The combined analysis of these three SNPs exhibited an increased PanC risk in a dose-response manner as the number of unfavorable genotypes increased ($P_{trend}<0.0001$). The risk-associated rs76692125 A allele was correlated with decreased *PIK3C3* mRNA expression levels, while the protective-associated rs11668724 A allele was correlated with increased *INSR* mRNA expression levels, but additional mechanistic studies of these SNPs are warranted. Once validated, these SNPs may serve as biomarkers for PanC risk in populations of European ancestry.

Keywords: Pancreatic cancer, single nucleotide polymorphism, risk analysis, ATM pathway

Introduction

Pancreatic cancer (PanC) is one of the deadliest cancers because of its rapid metastasis, accounting for 3.2% of all new cancer diagnoses [1]. An estimated 53% of PanCs are diagnosed at an advanced stage as a result of the lack of early symptoms and tumor-specific diagnostic tests, and the 5-year survival for metastatic PanC is only 2.9% [1]. It is estimated that PanC will be the second leading cause of cancer-related deaths in the United States by 2030 [2]. Hence, early prevention and detection may be the key to reduce PanC mortality.

Currently, several risk factors for PanC have been identified, including age, sex, smoking status, alcohol consumption, obesity, dietary factors, ethnicity, diabetes mellitus, family history of PanC and genetic susceptibility [3]. While the high-penetrant germline mutations are associated with an increased PanC risk, common genetic variants may also be the risk factors [4-9]. Over the past decade, although more than twenty single nucleotide polymorphisms (SNPs) have been identified as PanC-associated risk variants through genome-wide association studies (GWASs) that require a stringent *p* value of $<5.0\times10^{-8}$ due to the multiple-test inherent in this study method. Thus, some important functional susceptibility genes with a week effect may still remain unidentified. Recently, the pathway-based analyses of existing GWAS datasets have emerged as a hypothesis-driven approach of identifying novel functional SNPs within a biological context [10].

The ATM serine/threonine kinase (ATM) pathway is best known for its role in DNA damage response and plays important roles in mainte-

nance of genomic stability and suppression of tumorigenesis [11]. The *ATM* gene, also known as the ataxia telangiectasia mutated gene, is located on chromosome 11q22-23 and encodes a serine/threonine kinase member of the phosphatidylinositol 3 kinase family that is activated by DNA double-strand breaks; once activated, ATM phosphorylates multiple downstream effectors, including those involved in DNA damage repair, cell-cycle checkpoint arrest and apoptosis [12].

Recent studies have identified the link between the ATM pathway and PanC. For example, the activation of ATM and CHK1 has been reported to be associated with cell cycle arrest and apoptosis in human pancreatic cancer cells [13]. It has been reported that reduced levels of ATM coupled with oncogenic KRAS activation result in a higher number of dysplastic pancreatic lesions [14]. Accumulated evidence suggests that ATM can be activated independently from DNA damage through oxidative stress involved in autophagy induction [15]. The phosphoinositide-3-kinase class III (PIK3C3), a critical membrane marker for the autophagosome, has been reported to be associated with mediation of the autophagy in pancreatic cancer cells [16]. In addition, the ATM pathway genes can also affect insulin signaling function and glucose metabolism by regulating intracellular levels of reactive oxygen species, a known contributor to the onset of diabetes that is also a well-known risk factor of PanC [17].

In accordance with these findings, we hypothesize that genetic variants in the ATM pathway-related genes are associated with PanC risk. To test our hypothesis, we performed a comprehensive pathway gene-set-based analysis to identify potential functional SNPs that were associated with PanC risk by using the case and control subjects and genotyping data from two available GWAS datasets of pancreatic ductal adenocarcinoma in populations of European ancestry.

Material and methods

Study participants

We used the study participants with genotyping data from the two available PanC GWAS datasets, i.e., the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case Control Consortium (PanC4) studies that included 15,423 individuals of European ancestry. The PanScan GWAS dataset included 17 cohort studies and 11 case-control studies and had three phases: PanScan I (1,760 cases and 1,780 controls), PanScan II (1,457 cases and 1,666 controls) and PanScan III (1,538 cases and 0 controls) [4-6]. Because PanScan III lacks study-specific controls, the data from PanScan II and PanScan III were analyzed jointly, and the joint dataset was denoted as PanScan II/III (2,995 cases and 1,666 controls). The PanC4 GWAS dataset included nine studies from North America, Central Europe and Australia (3,722 cases and 3,500 controls) [7, 18, 19] (Table S1 and Figure S1).

All the studies obtained a written informed consent from study participants. The present study protocol was also approved by Duke University Health System Institutional Review Board (Pro00054575) and by the administration of the database of Genotypes and Phenotypes (dbGaP). The PanScan and PanC4 GWAS datasets are available from the dbGaP (accession #: phs000206.v5.p3 and phs000648.v1.p1, respectively).

Gene and SNP selection

We used the keyword "ATM" for the search in Molecular Signatures Database (MSigDB, v7.0) [20] and PathCards: multi-source consolidation of human biological pathways [21]. As a result, we obtained a total of 198 ATM pathway-related genes located in the autosomes from the online databases of BIOCARTA, PID and PathCards (Table S2), which were used for SNP extraction from the available GWAS datasets. The Illumina HumanHap550v3.0, the Human610_Quadv1_B, the HumanOmniExpress-12v 1.0 and the HumanOmniExpressExome-8v1 arrays [4-6] were used for genotyping in the GWAS datasets made available to the present study. We performed imputation by using IMPUTE2 with a buffer region of 500-kb up- and down-stream of these pathway genes and a reference panel from the 1000 Genomes Project (phase 3 release v5) [22]. We extracted the SNPs within the three GWAS datasets by using a boundary of 2-kb up- and down-stream of selected genes and performed quality assurance by using the following criteria: a SNP call rate of ≥ 95%, a minor allele frequency (MAF) of ≥ 0.01, a Hardy-Weinberg Equilibrium (HWE)

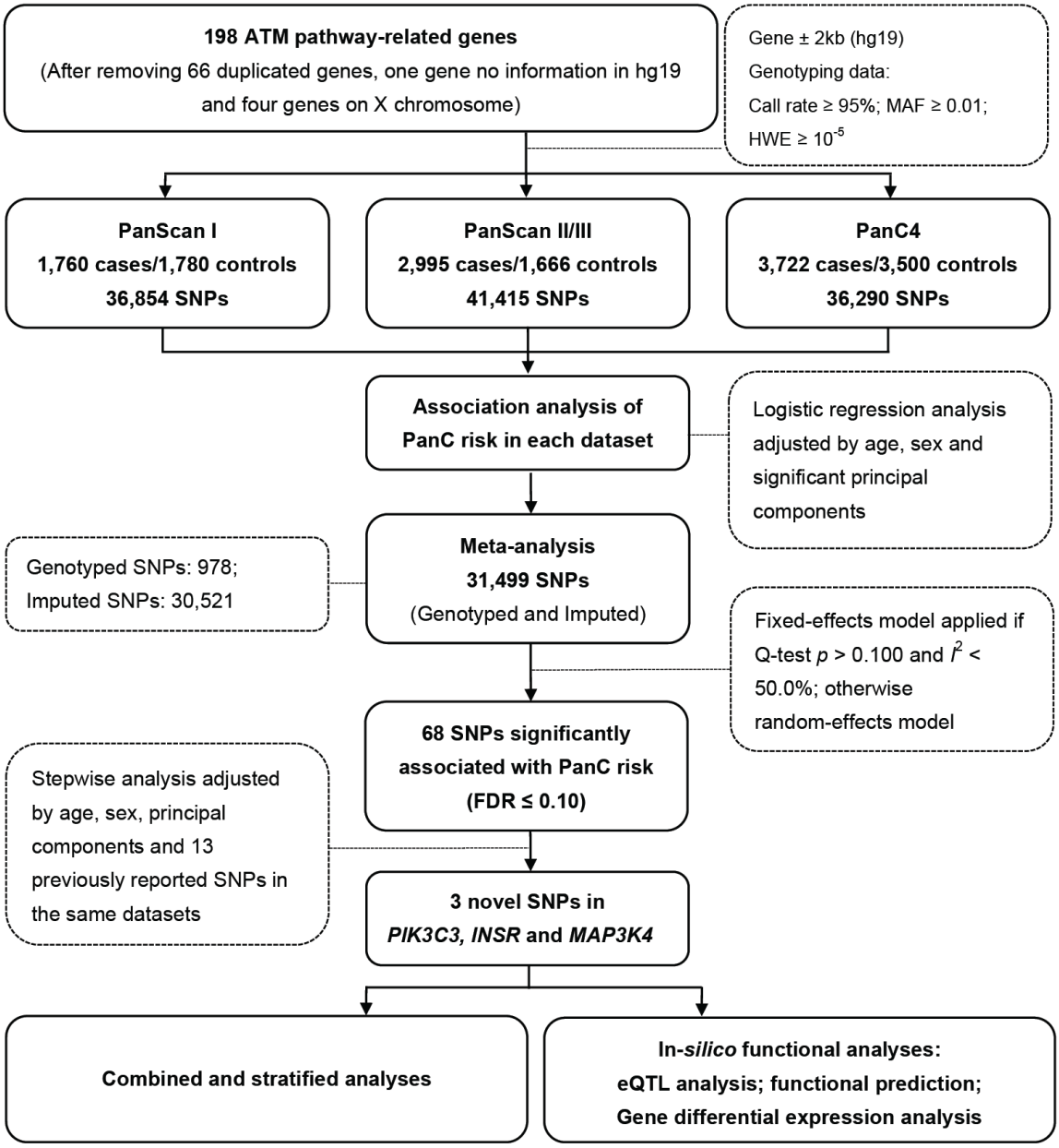


Figure 1. Flowchart of the present study. Abbreviations: kb, kilobase; SNP, single nucleotide polymorphism; MAF, minor allele frequency; HWE, Hardy-Weinberg equilibrium; PanC, pancreatic cancer; FDR, false discovery rate; *PIK3C3*, phosphatidylinositol 3-kinase catalytic subunit type 3; *INSR*, insulin receptor; *MAP3K4*, mitogen-activated protein kinase kinase kinase 4; eQTL, expression quantitative trait loci.

test of *P* value ≥ 1×10^{-5} among controls, and an imputation quality (INFO) score of ≥ 0.5 (**Figures 1** and S2).

Statistical analysis

We performed single-locus analysis between SNPs of the ATM pathway-related genes and PanC risk by using a multivariable logistic regression model with PLINK 1.9 [23] for each of the GWAS datasets separately. We constructed the models with adjustment for age, sex, and the top ancestry-informative principal components (PCs) from the genotyping data in PanScan I/II/III and PanC4. We estimated the effect sizes of SNPs by calculating odds ratio (OR) and 95% confidence interval (CI) accordingly. Then, we conducted a meta-analysis of the results of these three datasets by using the generic inverse variance method [24] with

PLINK 1.9. We used Cochran's Q test and I^2 to assess the heterogeneity between the datasets [25]. If the Cochran's Q-test *P*-value >0.100 and the I^2<50.0% were observed, a fixed-effects model was employed; otherwise, a random-effects model was adopted.

The false discovery rate (FDR) cutoff value was set at ≤ 0.10 to correct for multiple comparisons, which is sufficient for the tested SNPs largely in linkage disequilibrium (LD) [26]. We assessed the association between each SNP and PanC risk by using an additive genetic model. To identify independent-effect SNPs for PanC risk, we performed a stepwise logistic regression with adjustment for age, sex and the top five significant PCs as well as the 13 previously reported SNPs from the same datasets [27-30]. We constructed Manhattan plots by using Haploview v4.2 and the regional association plots to show all SNPs in LD by using LocusZoom [31]. To assess the joint effect of the SNPs, we also used the number of unfavorable genotypes (NUG) to calculate the cumulative effects of the independent SNPs and divided all the individuals into groups by the number of NUG for further analyses. Then, we used a general linear regression model in the exprression quantitative trait loci (eQTL) analysis to estimate the correlations between the independent SNPs and corresponding mRNA expression levels of the genes with the R software (v3.6.2). We performed all other statistical analyses with SAS software (v9.4; SAS Institute, Cary, NC, USA), if not specified otherwise.

Functional prediction and validation

We used three online bioinformatics tools, SNPinfo, RegulomeDB and HaploReg, to predict potential functions of the significant SNPs. SNPinfo integrates GWAS and candidate gene information into functional SNP selection for genetic association studies [32]; RegulomeDB helps identify DNA features and regulatory elements in non-coding regions of the human genome [33] and HaploReg is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks, such as candidate regulatory SNPs at disease-associated loci [34]. We also performed the eQTL analysis to estimate the associations between the genotypes of SNPs and the mRNA expression levels of the corresponding gene by using the mRNA expression data in the Genotype-Tissue Expression (GTEx) Project [35], the lymphoblastoid cells of 373 Europeans available from the 1,000 Genomes Project [36] and The Cancer Genome Atlas (TCGA) Project [37]. We used the PERFECTOS-APE [38, 39] and the PROMO online tools [40, 41] to predict the transcription factors in the promoter regions. We also obtained differences in the mRNA expression between tumor and adjacent normal tissues for the identified genes in TCGA dataset from the UALCAN web site [42, 43]. Finally, we conducted the overall survival analysis by using the TCGA dataset with the GEPIA web-portal tool [44, 45] and perform the mutations analysis by using publicly available data in the database of the cBioportal for Cancer Genomics [46, 47].

Results

Association analysis

The distributions of demographic characteristics between the three GWAS datasets are summarized in Table S1, including 8,477 cases and 6,946 controls of European ancestry [4-7, 18, 19]. As portrayed in **Figure 1**, after imputation and quality control, there were 36,854, 41,415 and 36,290 SNPs for PanScan I, PanScan II/III and PanC4, respectively. The single-locus analysis for the association between each of SNPs in the 198 ATM pathway-related genes (Table S2) and PanC risk was employed with the adjustment for age, sex and the top significant principle components (PCs) for each of the three datasets. A total of 2,177, 2,608 and 2,285 SNPs had a nominal *P*<0.050 in each dataset, respectively (Figure S3). In the meta-analysis of 31,499 SNPs (978 genotyped and 30,521 imputed) of the three datasets, we observed 1,901 SNPs to be associated with PanC risk with a nominal *P*<0.050, 68 of which were in eight genes (*WNT2B*, *MAP3K4*, *SMC2*, *ERBB2*, *TP53, PIK3C3*, *INSR* and *CHEK2*) with FDR ≤ 0.10 (**Figure 2A**). Then, we assessed these 68 SNPs in a multivariable stepwise logistic regression model to identify novel independent functional SNPs. As a result, we found that seven SNPs (i.e., rs13207108 in *MAP3K4*, rs76692125 in *PIK3C3*, rs11668724 in *INSR*, rs3838412 in *WNT2B*, rs2417487 in *SMC2*, rs2517955 in *ERBB2* and rs2236141 in *CHEK2*,) to be significantly associated with PanC risk (Table S3). Because some susceptibility

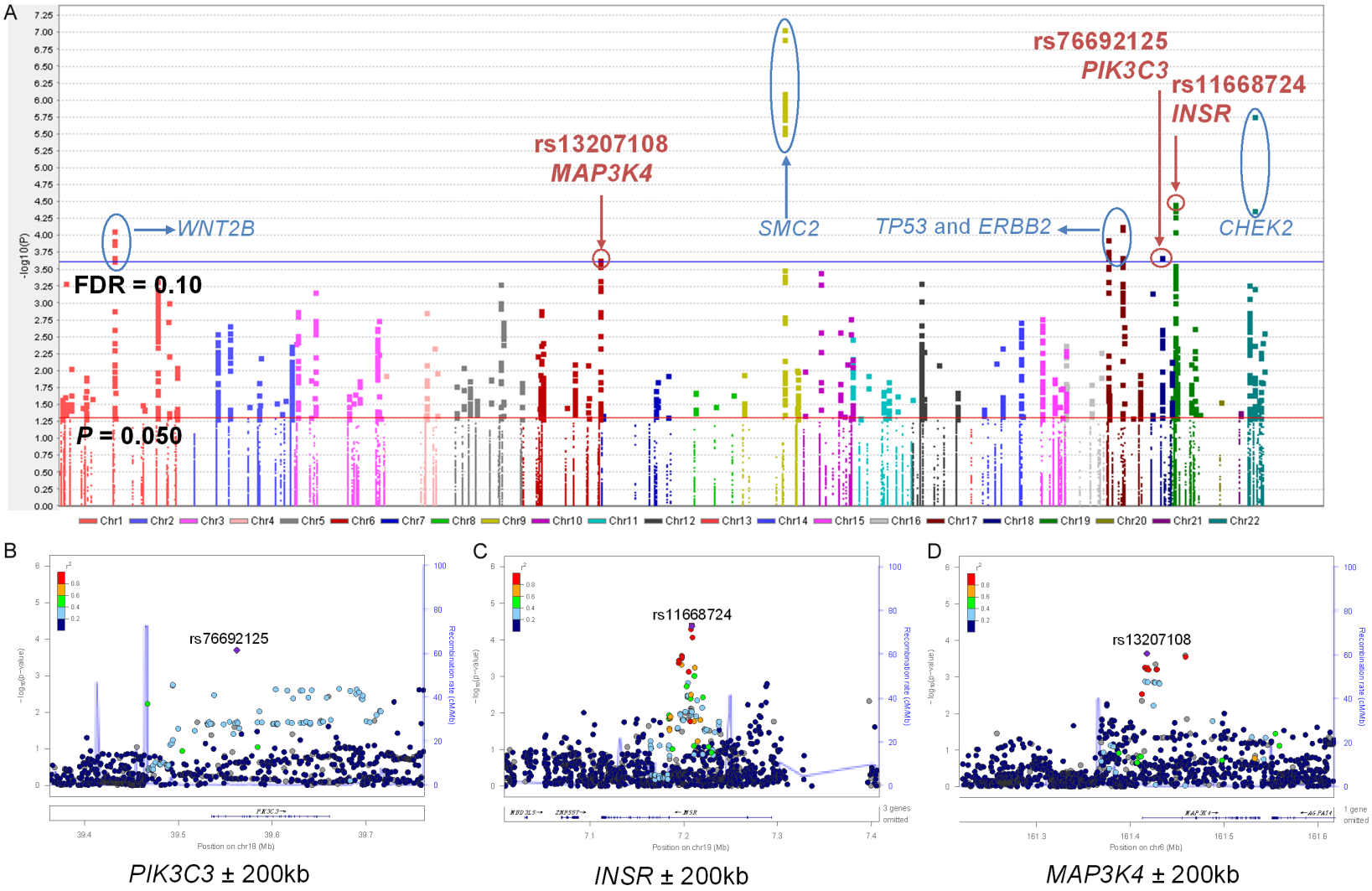


Figure 2. Association results of SNPs in the ATM-related pathway genes. (A) Manhattan plot of the association results. The statistical values across the autosomes of associations between 31,499 SNPs and PanC risk are plotted as -log10 *p* values in the meta-analysis of the three datasets. The red horizontal line indicates *P*=0.050. The blue horizontal line indicates FDR=0.10. There are 1,901 SNPs with *P*<0.050 and 68 SNPs with FDR ≤ 0.10. The three SNPs (i.e., rs76692125, rs11668724 and rs13207108) shown in the red bold are the novel findings in the present study. The five genes (i.e., *WNT2B*, *SMC2*, *ERBB2*, *TP53* and *CHEK2*) shown in blue are previously reported genes associated with PanC risk. Regional association plots for the three newly identified SNPs in the ATM pathway genes. SNPs in the regions of 200 kb up- and down-stream of (B) *PIK3C3*, (C) *INSR* and (D) *MAP3K4*.

Table 1. The three independent SNPs identified from the stepwise logistic regression analysis in the pooling PanC data

SNP	Allele[1]	Position	Gene	Region	OR (95% CI)[2]	*P*[2]
rs76692125	A/G	39562230	*PIK3C3*	18q12.3	1.31 (1.16-1.49)	<.0001
rs11668724	A/G	7208526	*INSR*	19p13.2	0.89 (0.84-0.94)	<.0001
rs13207108	T/C	161417626	*MAP3K4*	6q26	0.82 (0.74-0.91)	0.0002

Abbreviations: SNP, single nucleotide polymorphism; PanC, pancreatic cancer; OR, odds ratio; CI, confidence interval. [1]Effect (minor) allele/reference allele. [2]Stepwise analysis adjusted for age, sex, the top five principal components and 13 SNPs (rs35075084, rs2727572, rs34852782, rs62068300, rs3751936, rs3124761, rs17458086, rs1630747, rs5757573, rs6001516, rs79447092, rs9895829 and rs3818626) previously reported in the same dataset (PMID: 29168174, 30794721, 30972876 and 30997723, respectively).

loci located in *WNT2B*, *SMC2*, *ERBB2* and *CHEK2* had been reported to be associated with PanC risk through GWAS datasets analyses [6, 7, 9, 48], we used the three PanC risk loci [i.e., 18q12.3 (*PIK3C3* rs76692125 G>A), 19p13.2 (*INSR* rs11668724 G>A) and 6q26 (*MAP3K4* rs13207108 C>T)] newly identified in the present study for further analyses (**Table 1** and **Figure 2A-D**). As shown in **Table 2**, the *PIK3C3* rs76692125 A allele was associated with an increased risk of PanC (OR=1.26, 95% CI=1.12-1.43, $P=2.07\times10^{-4}$), while the *INSR* rs11668724 A allele and the *MAP3K4* rs13207108 T allele were associated with a reduced risk of PanC (OR=0.89, 95% CI=0.84-0.94, $P=4.21\times10^{-5}$ and OR=0.83, 95% CI=0.75-0.92, $P=2.26\times10^{-4}$, respectively) in the meta-analysis of the three GWAS datasets. Heterogeneity in these associations was not observed among the three datasets.

Combined and stratified analyses

As shown in **Table 3**, *PIK3C3* rs76692125 GA+AA genotypes were associated with an increased PanC risk (OR=1.31, 95% CI=1.16-1.48, *P*<0.0001), while *INSR* rs11668724 GA+AA genotypes and *MAP3K4* rs13207108 CT+TT genotypes were associated with a reduced PanC risk (OR=0.88, 95% CI=0.82-0.94, *P*<0.0001 and OR=0.84, 95% CI=0.76-0.93, *P*=0.0007, respectively). Therefore, we combined the unfavorable genotypes of rs76692125 GA+AA, rs11668724 GG and rs13207108 CC to assess the cumulative effect of these three SNPs. The joint analysis suggested that the NUG was significantly associated with risk of PanC in a dose-response manner (P_{trend}< 0.0001, **Table 3**). The multivariable logistic regression model incorporating the NUG exhibited that individuals with two-three NUGs had a higher risk of PanC, compared with those with zero-one NUGs (OR=1.22, 95% CI=1.15-1.30, *P*<0.0001, **Table 3**).

Furthermore, we employed stratified analysis for risk modification by age and sex in the present study, and no interaction was observed between the subgroups within the strata (P_{inter}>0.050, Table S4).

Function analysis

We performed the eQTL analysis for the three identified independent SNPs on mRNA expression by using the expression and genotyping data of 305 normal pancreatic tissues and 670 whole blood samples from the GTEx Project [35], the RNA-Seq and genotyping data of lymphoblastoid cell lines from 373 European descendants from the 1,000 Genomes Project [36], and the expression and genotyping data of 180 pancreatic tumor samples from the TCGA Project [37]. We found that the *PIK3C3* rs76692125 A allele was correlated with a decreased mRNA expression level in the whole blood cells, but not in the normal pancreatic tissues (**Figure 3A**), form the GTEx Project (n=670, $P=1.13\times10^{-5}$, Figure S4A), and in the lymphoblastoid cells from the 1,000 Genomes Project (n=373, $P=5.26\times10^{-3}$, **Figure 3B**). As shown in **Figure 3C**, the rs11668724 A allele was associated with an increased *INSR* mRNA expression level in normal pancreas tissues from the GTEx Project (n=305, $P=5.97\times10^{-6}$). This finding was not observed either in the whole blood cells from the GTEx Project (Figure S4B) or in the lymphoblastoid cells from the 1,000 Genomes Project (**Figure 3D**). No significant eQTL results for *MAP3K4* rs13207108 was observed either from the GTEx or from the 1,000 Genomes Projects (**Figures 3E**, **3F** and S4C). However,

Table 2. Associations between the three independent SNPs and PanC risk in the three PanC datasets

SNP	Allele[1]	Position	Gene	PanScan I 1,760/1,780[2]			PanScan II/III 2,995/1,666[2]			PanC4 3,722/3,500[2]			Meta-analysis 8,477/6,946[2]		Heterogeneity[5]		FDR
				MAF	OR (95% CI)	*P*[3]	MAF	OR (95% CI)	*P*[3]	MAF	OR (95% CI)	*P*[3]	OR (95% CI)	*P*[4]	Q	I^2	
rs76692125	A/G	39562230	*PIK3C3*	0.04	1.09 (0.86-1.39)	0.4808	0.03	1.22 (0.96-1.54)	0.0989	0.03	1.40 (1.17-1.67)	0.0003	1.26 (1.12-1.43)	2.07×10^{-4}	0.263	25.2	0.10
rs11668724	A/G	7208526	*INSR*	0.23	0.94 (0.84-1.05)	0.2812	0.24	0.83 (0.75-0.92)	0.0003	0.22	0.91 (0.84-0.98)	0.0185	0.89 (0.84-0.94)	4.21×10^{-5}	0.211	35.8	0.03
rs13207108	T/C	161417626	*MAP3K4*	0.05	0.93 (0.75-1.15)	0.5113	0.07	0.84 (0.70-1.00)	0.0597	0.06	0.78 (0.67-0.90)	0.0007	0.83 (0.75-0.92)	2.26×10^{-4}	0.395	0.0	0.10

Abbreviations: SNP, single nucleotide polymorphism; PanC, pancreatic cancer; GWAS, genome-wide association study; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; FDR, false discovery rate. [1]Effect (minor) allele/reference allele. [2]Number of case/number of control. [3]Adjusted for age, sex and significant principal components in each dataset. [4]Meta-analysis in the three datasets. [5]Heterogeneity assessed by Q-test or I^2: fixed-effects model if Q test $P>0.100$ and $I^2<50.0\%$; otherwise random-effects model.

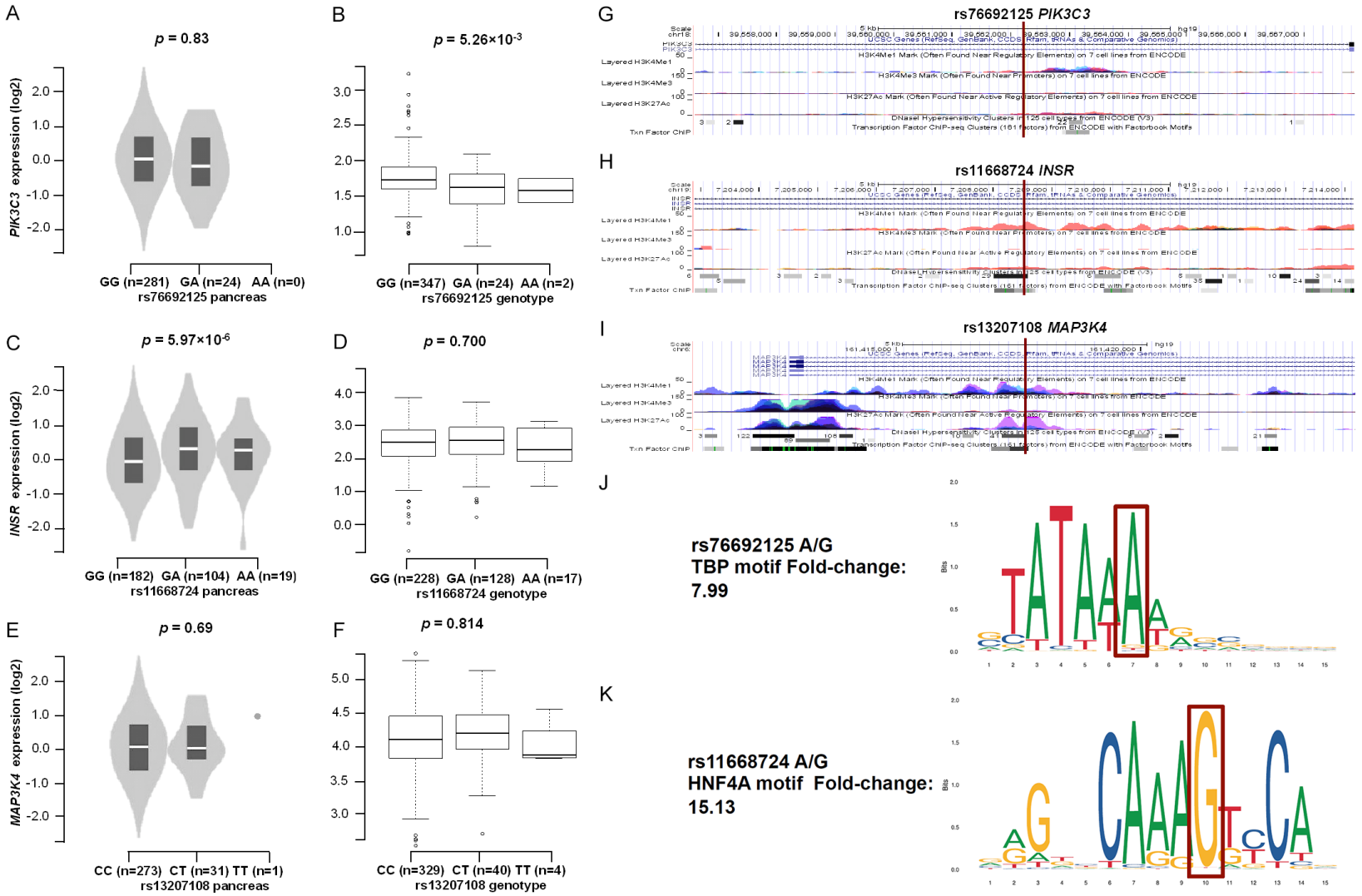
Table 3. Combined analysis of the three independent SNPs and PanC risk in the three PanC datasets

Genotype	Case (%)	Control (%)	OR (95% CI)[1]	*P*[1]
PIK3C3 rs76692125				
GG	7778 (91.8)	6490 (93.5)	1.0	
GA	682 (8.0)	444 (6.4)	1.30 (1.14-1.47)	<0.0001
AA	15 (0.2)	5 (0.1)	2.62 (0.95-7.25)	0.0627
Trend test				<0.0001
Dominant model				
GG	7778 (91.8)	6940 (93.5)	1.0	
GA+AA	697 (8.2)	449 (6.5)	1.31 (1.16-1.48)	<0.0001
INSR rs11668724				
GG	5322 (62.8)	4142 (59.6)	1.0	
GA	2774 (32.7)	2440 (35.1)	0.89 (0.83-0.95)	0.0005
AA	379 (4.5)	364 (5.2)	0.81 (0.69-0.94)	0.0050
Trend test				<0.0001
Dominant model				
GG	5322 (62.8)	4142 (59.6)	1.0	
GA+AA	3153 (37.2)	2804 (40.4)	0.88 (0.82-0.94)	<0.0001
Reversed model				
GA+AA	3153 (37.2)	2804 (40.4)	1.0	
GG	5322 (62.8)	4142 (59.6)	1.14 (1.06-1.22)	<0.0001
MAP3K4 rs13207108				
CC	7637 (90.1)	6130 (88.3)	1.0	
CT	818 (9.6)	793 (11.4)	0.84 (0.76-0.93)	0.0009
TT	22 (0.3)	23 (0.3)	0.79 (0.44-1.41)	0.4209
Trend test				0.0007
Dominant model				
CC	7637 (90.1)	6130 (88.3)	1.0	
CT+TT	840 (9.9)	816 (11.7)	0.84 (0.76-0.93)	0.0007
Reversed model				
CT+TT	840 (9.9)	816 (11.7)	1.0	
CC	7637 (90.1)	6130 (88.3)	1.19 (1.08-1.32)	0.0007
NUG[2]				
0	279 (3.3)	295 (4.3)	1.0	
1	3099 (36.6)	2813 (40.5)	1.14 (0.96-1.35)	0.1393
2	4734 (55.9)	3593 (51.8)	1.36 (1.15-1.61)	0.0004
3	361 (4.3)	238 (3.4)	1.59 (1.26-2.01)	<0.0001
Trend test				<0.0001
0-1	3378 (39.9)	3108 (44.8)	1.0	
2-3	5095 (60.1)	3831 (55.2)	1.22 (1.15-1.30)	<0.0001

Abbreviations: SNP, single nucleotide polymorphism; PanC, pancreatic cancer; OR, odds ratio; CI, confidence interval; NUG, number of unfavorable genotypes. [1]Adjusted by age, sex and the top five significant principal components. [2]Unfavorable genotypes were rs76692125 GA+AA, rs11668724 GG and rs13207108 CC.

the TCGA eQTL data of these three SNPs were not available for further analyses.

We assessed potential functions of the three independent PanC risk-associated SNPs by using *in silico* bioinformatics tools (Table S5). These three SNPs are located in intron regions, where active histone marks are enriched, including histone H3 mono methyl K4 (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac) (**Figure 3G-I**). We then evaluated the effects of these SNPs on the predicted tran-


A
p = 0.83
PIK3C3 expression (log2)
GG (n=281) GA (n=24) AA (n=0)
rs76692125 pancreas
B
p = 5.26×10-3
GG (n=347) GA (n=24) AA (n=2)
rs76692125 genotype
C
p = 5.97×10-6
INSR expression (log2)
GG (n=182) GA (n=104) AA (n=19)
rs11668724 pancreas
D
p = 0.700
GG (n=228) GA (n=128) AA (n=17)
rs11668724 genotype
E
p = 0.69
MAP3K4 expression (log2)
CC (n=273) CT (n=31) TT (n=1)
rs13207108 pancreas
F
p = 0.814
CC (n=329) CT (n=40) TT (n=4)
rs13207108 genotype
G
rs76692125 PIK3C3
H
rs11668724 INSR
I
rs13207108 MAP3K4
J
rs76692125 A/G
TBP motif Fold-change:
7.99
K
rs11668724 A/G
HNF4A motif Fold-change:
15.13

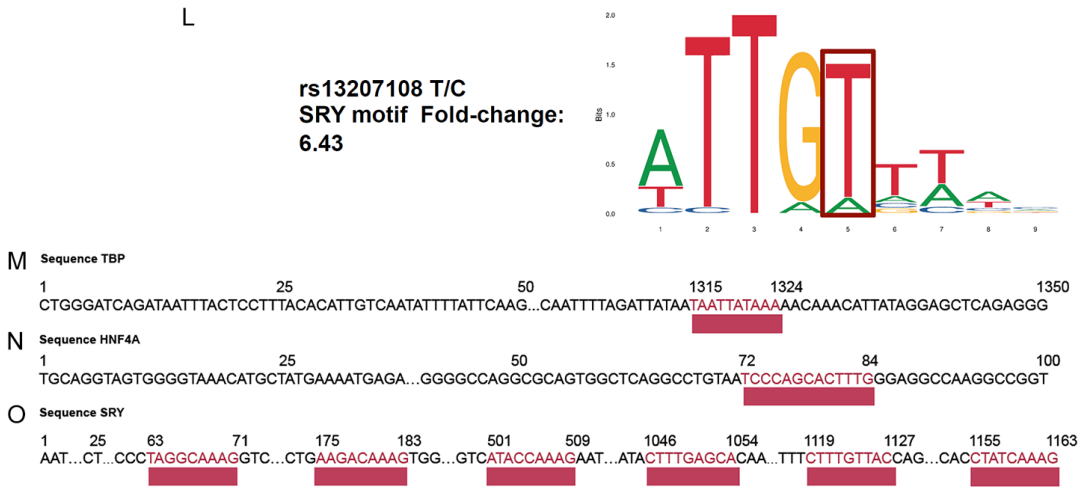


Figure 3. Functional analyses of the three identified SNPs. Expression quantitative trait loci (eQTL) analysis of SNP rs76692125 and *PIK3C3* mRNA expression levels (A) in the normal pancreatic tissues (n=305, P=0.83) from the GTEx Project and (B) in the transformed lymphoblastoid cells (n=373, P=5.26×10^{-3}) from the 1000 Genomes Project; the rs11668724 genotypes and *INSR* mRNA expression levels (C) in the normal pancreatic tissues (n=305, P=5.97×10^{-6}) from the GTEx Project and (D) in the transformed lymphoblastoid cells (n=373, P=0.700) from the 1000 Genomes Project; the rs13207108 genotypes and *MAP3K4* mRNA expression levels (E) the normal pancreatic tissues (n=305, P=0.69) from the GTEx Project and (F) in the transformed lymphoblastoid cells (n=373, P=0.814) from the 1000 Genomes Project. (G-I) Location and functional prediction of the three novel SNPs in the ENCODE Project. The H3K4Me1, H3K4Me3 and H3K27Ac tracks are associated with the enhancer and promoter regions. The Txn Factor ChIP tracks show regions of transcription factor binding of DNA; Effect analyses of the three SNPs on TF motifs predicated by using HOCOMOCO-11 collection from the PERFECTOS-APE online tools: (J) the rs76692125 A allele may alter the predicted TF-binding motifs for TBP; (K) the rs11668724 A allele may alter the predicted TF-binding motifs for HNF4A and (L) the rs13207108 T allele may alter the predicted TF-binding motif for SRY. The putative binding sites in promoter predicted by PROMO online tools: (M) the TBP binding site in *PIK3C3* promoter; (N) the HNF4A binding site in *INSR* promoter and (O) the SRY binding sites in *MAP3K4* promoter. Abbreviations: SNP, single nucleotide polymorphism; GTEx, Genotype-Tissue Expression; ChIP, Chromatin Immunoprecipitation; Txn Factor ChIP, Transcription Factor ChIP-seq from ENCODE; TF, transcriptional factor.

scription factor binding sites using the PERFECTOS-APE [38, 39] and the PROMO online tools [40, 41]. Several transcription factors, including TBP, HNF4A and SRY, are predicted to bind to DNA fragments containing these SNPs with different affinities (**Figure 3J-O**).

Furthermore, we performed differential mRNA expression analysis of *PIK3C3*, *INSR* and *MAP3K4* between pancreatic tumors and adjacent normal tissues and overall survival analysis from the TCGA database by using the UALCAN [42, 43] and GEPIA [44, 45] web-portal tools. The mRNA expression levels of *MAP3K4,* but not *PIK3C3* and *INSR*, were significantly decreased in pancreatic tumors (P= 0.047) (Figure S4D-F), compared with that of normal pancreatic tissues. High mRNA expression levels of *MAP3K4*, but not *PIK3C3* and *INSR*, were associated with a better survival of PanC patients, compared with low mRNA expression levels (P=0.045) (Figure S4G-I).

Mutation analysis

Finally, we assessed the mutation status of *PIK3C3*, *INSR* and *MAP3K4* in PanC tissues using the cBioPortal database for Cancer Genomics [46, 47]. The three genes had low somatic mutation rates in PanC: *PIK3C3* [0.56% (1/179), 0.52% (2/383), 0% (0/99) and 0% (0/109)]; *INSR* [1.12% (2/179), 0.26% (1/383), 0% (0/99) and 0% (0/109)]; and *MAP3K4* [0.56% (1/179), 0% (0/383), 0% (0/99), and 0% (0/109)] from TCGA PanCan 2018, Queensland Centre of Medical Genomics 2016, the International Cancer Genome Consortium and The University of Texas Southwest studies, respectively (Figure S5A-C).

Discussion

In the present study, we identified three independent and potentially functional SNPs (i.e., *PIK3C3* rs76692125 G>A, *INSR* rs11668724 G>A and *MAP3K4* rs13207108 C>T) that were associated with PanC risk in populations of European ancestry. The risk-associated rs76692125 A allele was correlated with decreased *PIK3C3* mRNA expression levels, while the protective-associated rs11668724 A allele was correlated with increased *INSR* mRNA expression levels. Because the ATM pathway-related genes play important roles in diverse biological processes and pathological disorders, such as DNA damage repair, autophagy regulation, metabolic disorders and cancers, including PanC [11-17], these genotype-mRNA expression correlations provided biological plausibility for the observed genotype-associated risk of PanC.

PIK3C3 is located at chromosome region 18q12.3 and encodes a phosphatidylinositol 3-kinase catalytic subunit type 3 (PIK3C3) protein, the member of the class III PI3K family that plays a positive role in autophagy through complex formation by generating phosphatidylinositol 3P. We found that the *PIK3C3* rs76692125A risk allele was significantly associated with an increased risk of PanC, possibly by a mechanism of decreasing *PIK3C3* mRNA expression levels. Studies showed that activation of ATM contributed to phosphorylation of Beclin1, followed by the complex formation with PIK3C3 [15, 16]. At the early stage of tumorigenesis, an inhibition of autophagy might contribute to continuous growth of precancerous cells, and the autophagy might act as a suppressor [49]. Although no reported study has demonstrated the role of *PIK3C3* in malignant transformation, the autophagy-associated *beclin1* gene as part of the PIK3C3 complex inhibit tumorigenesis in mice and expresses at decreased levels in human tumors [50, 51]. Another study indicated that the combined loss of autophagy and p53 dramatically promoted progression from early pancreatic Intraepithelial neoplastic lesions towards adenocarcinoma [52]. According to the functional analyses, the SNP rs76692115 may participate in modulating the interaction with DNA binding motifs and thus altering the *PIK3C3* expression, which may further affect the autophagy activity to increase the risk of PanC. However, this speculation deserves further investigation and additional functional studies.

INSR is located at chromosome region 19p13.2 and encodes an insulin receptor (INSR) that plays an essential role in the insulin signaling. In the present study, we found that the *INSR* rs11668724 A protective allele was associated with a significantly decreased risk for PanC and also was correlated with an increased mRNA expression level of *INSR* in normal pancreatic tissues. Because the mutation rate of *INSR* in PanC was as low as 1.12%, the expression levels of *INSR* in pancreatic tissues were likely to be effected by SNPs in the gene. For example, the *INSR* rs11668724 was predicted to change

the DNA binding capacities with the HNF4A motif that may be associated with transcriptional regulation of *INSR* expression and beta cell function in pancreatic tissues [53, 54]. One study identified that the pancreatic beta cell INSR knockout mice or cell lines lost insulin secretion, suggesting that INSRs in beta cells may play a critical role in beta cell dysfunction that is associated with insulin resistance and type 2 diabetes [55-57]. Epidemiological evidence also points to a link between some syndromes (e.g., insulin resistance and type 2 diabetes) and PanC risk [3]. Therefore, these results suggest that the *INSR* SNP rs11668724 may be involved in dysfunction of beta cells by regulating the *INSR* mRNA expression levels in the pancreas, leading to PanC risk. In addition, several previous GWASs identified that the *INSR* SNP rs7248104, which was in moderate linkage disequilibrium (LD) with rs11668724 (r^2=0.22), was associated with elevated blood triglycerides and fasting glucose levels [58] and an increased risk of thyroid cancer [59]. However, all these were of an indirect evidence, and more specific investigations are still required to investigate the mechanisms underlying the observed associations.

MAP3K4 is located at chromosome region 6q26 and encodes a mitogen-activated protein kinase kinase kinase 4 (MAP3K4) that is a member of the mitogen-activated protein kinase superfamily. In the present study, we identified that the *MAP3K4* rs13207108 T protective allele was associated with a significantly decreased risk of PanC, while a suggestive risk locus (*PARK2* rs3016539) was also identified at 6q26 in a PanC GWAS performed in a Japanese population [60]. Although there is no LD between rs3016539 and rs13207108 identified in the present study, it is likely indicative that the 6q26 region may be a potential PanC susceptibility locus in both populations of European ancestry and Japanese populations.

MAP3K4 acts as a mediator in the environmental stress-induced activation of the p38 and JNK signaling to regulate cell proliferation, apoptosis, and migration [61]. The *MAP3K4* genes have been reported to be associated with a variety of cancers, including PanC [62-64]. Studies have identified that *MAP3K4* may act as a suppressor in some cancers [63, 64]. One study reported that continuous activation of p38 through the Smad/GADD45β/MAP3K4 cascade might contribute to the tumor-suppressive effect in pancreatic cells [63]. This suppressor role has been supported by human intrahepatic cholangiocarcinoma [64]. Although no significant eQTL result was observed for this genetic variant, the data from TCGA study indicated that the PanC patients had lower *MAP3K4* expression levels that were associated with a worse survival. Taken together, multiple levels of evidence suggest that *MAP3K4* may be a candidate tumor suppressor for PanC.

Although the present study identified three potential susceptibility loci for PanC risk, it has some limitations. First, all the genotyping data used in the present study were exclusively from populations of European ancestry. Therefore, the findings may not be generalized to other ethnic populations. Secondly, except for age and sex, other clinical variables (e.g. treatment history, family history, smoking status, alcohol intake) were not available for further adjustment and stratified analyses. Finally, additional functional studies are required to verify the hypothesized biological mechanisms underlying our observed associations.

In summary, the present ATM pathway-based study analyzed genetic variants to be associated with PanC risk in populations of European ancestry. We identified three novel independent and potentially functional SNPs (i.e., *PIK3C3* rs76692125 G>A, *INSR* rs11668724 G>A and *MAP3K4* rs13207108 C>T) that were significantly associated with PanC risk. Additional larger population-based and functional studies are needed to validate our findings.

PanScan

The PanScan study was funded in whole or in part with federal funds from the National Cancer Institute (NCI) and National Institutes of Health (NIH), contract number HHSN2612-00800001E. Additional support was received from NIH/NCI K07CA140790, 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. The

dbGaP accession number for this study is phs000206.v5.p3.

PanC4

The cases and controls for the PanC4 study were drawn from the following 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. The PanC4 study was supported by NCI R01CA154823. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR was funded by a federal contract from the NIH to The Johns Hopkins University, contract number HHSN26820110-00111. The dbGaP accession number for this study is phs000648.v1.p1.

TCGA

The results published here are in whole or part based on data generated by the TCGA Project established by the NCI and the National Human Genome Research Institute (NHGRI). Information about TCGA and the investigators and institutions that constitute TCGA Research Network can be found at "http://cancergenome.nih.gov". The TCGA SNP data analyzed here are requested through dbGaP, accession number phs000178.v1.p1.

Acknowledgements

The authors acknowledge all the participants of the PanScan Study, PanC4 Study and the dbGaP repository for providing the cancer genotyping datasets that are described as follows. Qing-yi Wei was supported by the P30 Cancer Center Support Grant from the Duke Cancer Institute (Grant ID: NIH CA014236). Ling-Ling Zhao was supported by funds from The First Hospital of Jilin University in China.

Disclosure of conflict of interest

None.

Address correspondence to: Wei Li, Cancer Center, The First Hospital of Jilin University, 71 Xinmin Street, Changchun 130021, Jilin, China. Tel: +86-431-88782974; E-mail: jdyylw@163.com; Qingyi Wei, Duke Cancer Institute, Duke University Medical Center and Department of Population Health Sciences, School of Medicine, Duke University, 905 S LaSalle Street, Durham, NC 27710, USA. Tel: 919-660-0562; E-mail: qingyi.wei@duke.edu

References

[1] Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2016, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2016/, based on November 2018 SEER data submission, posted to the SEER web site, April 2019.

[2] Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM and Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014; 74: 2913-2921.

[3] Rawla P, Sunkara T and Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol 2019; 10: 10-27.

[4] Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, Bueno-de-Mesquita HB, Gross M, Helzlsouer K, Jacobs EJ, LaCroix A, Zheng W, Albanes D, Bamlet W, Berg CD, Berrino F, Bingham S, Buring JE, Bracci PM, Canzian F, Clavel-Chapelon F, Clipp S, Cotterchio M, de Andrade M, Duell EJ, Fox JW Jr, Gallinger S, Gaziano JM, Giovannucci EL, Goggins M, González CA, Hallmans G, Hankinson SE, Hassan M, Holly EA, Hunter DJ, Hutchinson A, Jackson R, Jacobs KB, Jenab M, Kaaks R, Klein AP, Kooperberg C, Kurtz RC, Li D, Lynch SM, Mandelson M, McWilliams RR, Mendelsohn JB, Michaud DS, Olson SH, Overvad K, Patel AV, Peeters PHM, Rajkovic A, Riboli E, Risch HA, Shu XO, Thomas G, Tobias GS, Trichopoulos D, Van Den Eeden SK, Virtamo J, Wactawski-Wende J, Wolpin BM, Yu H, Yu K, Zeleniuch-Jacquotte A, Chanock SJ, Hartge P and Hoover RN. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet 2009; 41: 986-990.

[5] Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, Arslan AA, Bueno-de-Mesquita HB, Gallinger S, Gross M, Helzlsouer K, Holly EA, Jacobs EJ, Klein AP, LaCroix A, Li D, Mandelson MT, Olson SH, Risch HA, Zheng W, Albanes D, Bamlet WR, Berg CD, Boutron-Ruault MC, Buring JE, Bracci PM, Canzian F, Clipp S, Cotterchio M, de Andrade M, Duell EJ, Gaziano JM, Giovannucci EL, Goggins

M, Hallmans G, Hankinson SE, Hassan M, Howard B, Hunter DJ, Hutchinson A, Jenab M, Kaaks R, Kooperberg C, Krogh V, Kurtz RC, Lynch SM, McWilliams RR, Mendelsohn JB, Michaud DS, Parikh H, Patel AV, Peeters PH, Rajkovic A, Riboli E, Rodriguez L, Seminara D, Shu XO, Thomas G, Tjønneland A, Tobias GS, Trichopoulos D, Van Den Eeden SK, Virtamo J, Wactawski-Wende J, Wang Z, Wolpin BM, Yu H, Yu K, Zeleniuch-Jacquotte A, Fraumeni JF Jr, Hoover RN, Hartge P and Chanock SJ. 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.

[6] Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, Arslan AA, Beane-Freeman L, Bracci PM, Buring J, Canzian F, Duell EJ, Gallinger S, Giles GG, Goodman GE, Goodman PJ, Jacobs EJ, Kamineni A, Klein AP, Kolonel LN, Kulke MH, Li D, Malats N, Olson SH, Risch HA, Sesso HD, Visvanathan K, White E, Zheng W, Abnet CC, Albanes D, Andreotti G, Austin MA, Barfield R, Basso D, Berndt SI, Boutron-Ruault MC, Brotzman M, Büchler MW, Bueno-de-Mesquita HB, Bugert P, Burdette L, Campa D, Caporaso NE, Capurso G, Chung C, Cotterchio M, Costello E, Elena J, Funel N, Gaziano JM, Giese NA, Giovannucci EL, Goggins M, Gorman MJ, Gross M, Haiman CA, Hassan M, Helzlsouer KJ, Henderson BE, Holly EA, Hu N, Hunter DJ, Innocenti F, Jenab M, Kaaks R, Key TJ, Khaw KT, Klein EA, Kogevinas M, Krogh V, Kupcinskas J, Kurtz RC, LaCroix A, Landi MT, Landi S, Le Marchand L, Mambrini A, Mannisto S, Milne RL, Nakamura Y, Oberg AL, Owzar K, Patel AV, Peeters PHM, Peters U, Pezzilli R, Piepoli A, Porta M, Real FX, Riboli E, Rothman N, Scarpa A, Shu XO, Silverman DT, Soucek P, Sund M, Talar-Wojnarowska R, Taylor PR, Theodoropoulos GE, Thornquist M, Tjønneland A, Tobias GS, Trichopoulos D, Vodicka P, Wactawski-Wende J, Wentzensen N, Wu C, Yu H, Yu K, Zeleniuch-Jacquotte A, Hoover R, Hartge P, Fuchs C, Chanock SJ, Stolzenberg-Solomon RS and Amundadottir LT. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat Genet 2014; 46: 994-1000.

[7] Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, Li D, Neale RE, Olson SH, Scelo G, Amundadottir LT, Bamlet WR, Bijlsma MF, Blackford A, Borges M, Brennan P, Brenner H, Bueno-de-Mesquita HB, Canzian F, Capurso G, Cavestro GM, Chaffee KG, Chanock SJ, Cleary SP, Cotterchio M, Foretova L, Fuchs C, Funel N, Gazouli M, Hassan M, Herman JM, Holcatova I, Holly EA, Hoover RN, Hung RJ, Janout V, Key TJ, Kupcinskas J, Kurtz RC, Landi S, Lu L, Malecka-Panas E, Mambrini A, Mohelnikova-Duchonova B, Neoptolemos JP, Oberg AL, Orlow I, Pasquali C, Pezzilli R, Rizzato C, Saldia A, Scarpa A, Stolzenberg-Solomon RZ, Strobel O, Tavano F, Vashist YK, Vodicka P, Wolpin BM, Yu H, Petersen GM, Risch HA and Klein AP. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet 2015; 47: 911.

[8] Amundadottir LT. Pancreatic cancer genetics. Int J Biol Sci 2016; 12: 314-325.

[9] Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, Canzian F, Childs EJ, Hoskins JW and Jermusyk A. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun2018; 9: 1-11.

[10] Cantor RM, Lange K and Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. Am J Hum Genet 2010; 86: 6-22.

[11] Bartkova J, Hořejší Z, Koed K, Krämer A, Tort F, Zieger K, Guldberg P, Sehested M, Nesland JM, Lukas C, Ørntoft T, Lukas J and Bartek J. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 2005; 434: 864-870.

[12] Armstrong SA, Schultz CW, Azimi-Sadjadi A, Brody JR and Pishvaian MJ. ATM dysfunction in pancreatic adenocarcinoma and associated therapeutic implications. molecular cancer Therapeutics 2019; 18: 1899.

[13] Sahu RP, Batra S and Srivastava SK. Activation of ATM/Chk1 by curcumin causes cell cycle arrest and apoptosis in human pancreatic cancer cells. Br J Cancer 2009; 100: 1425-1433.

[14] Russell R, Perkhofer L, Liebau S, Lin Q, Lechel A, Feld FM, Hessmann E, Gaedcke J, Güthle M, Zenke M, Hartmann D, von Figura G, Weissinger SE, Rudolph KL, Möller P, Lennerz JK, Seufferlein T, Wagner M and Kleger A. Loss of ATM accelerates pancreatic cancer formation and epithelial-mesenchymal transition. Nat Commun 2015; 6: 7677-7677.

[15] Stagni V, Cirotti C and Barilà D. Ataxia-telangiectasia mutated kinase in the control of oxidative stress, mitochondria, and autophagy in cancer: a maestro with a large orchestra. Front Oncol 2018; 8: 73-73.

[16] Maertin S, Elperin JM, Lotshaw E, Sendler M, Speakman SD, Takakura K, Reicher BM, Mareninova OA, Grippo PJ, Mayerle J, Lerch MM and Gukovskaya AS. Roles of autophagy and metabolism in pancreatic cancer cell adaptation to environmental challenges. Am J Physiol Gastrointest Liver Physiol 2017; 313: G524-G536.

[17] Ditch S and Paull TT. The ATM protein kinase and cellular redox signaling: beyond the DNA damage response. Trends Biochem Sci 2012; 37: 15-22.
[18] Borgida AE, Ashamalla S, Al-Sukhni W, Rothenmund H, Urbach D, Moore M, Cotterchio M and Gallinger S. Management of pancreatic adenocarcinoma in Ontario, Canada: a population-based study using novel case ascertainment. Can J Surg 2011; 54: 54.
[19] McWilliams RR, Bamlet WR, De Andrade M, Rider DN, Cunningham JM and Petersen GM. Nucleotide excision repair pathway polymorphisms and pancreatic cancer risk: evidence for role of MMS19L. Cancer Epidemiol Biomarkers Prev 2009; 18: 1295-1302.
[20] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP and Tamayo P. The molecular signatures database hallmark gene set collection. Cell Syst 2015; 1: 417-425.
[21] Belinky F, Nativ N, Stelzer G, Zimmerman S, Iny Stein T, Safran M and Lancet D. PathCards: multi-source consolidation of human biological pathways. Database 2015; 2015.
[22] Howie BN, Donnelly P and Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009; 5: e1000529.
[23] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW and Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-575.
[24] Begum F, Ghosh D, Tseng GC and Feingold E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Res 2012; 40: 3777-3784.
[25] Higgins JPT, Thompson SG, Deeks JJ and Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557-560.
[26] Benjamini Y, Drai D, Elmer G, Kafkafi N and Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 2001; 125: 279-284.
[27] Duan B, Hu J, Liu H, Wang Y, Li H, Liu S, Xie J, Owzar K, Abbruzzese J and Hurwitz H. Genetic variants in the platelet-derived growth factor subunit B gene associated with pancreatic cancer risk. Int J Cancer 2018; 142: 1322-1331.
[28] Feng Y, Liu H, Duan B, Liu Z, Abbruzzese J, Walsh KM, Zhang X and Wei Q. Potential functional variants in SMC2 and TP53 in the AURORA pathway genes and risk of pancreatic cancer. Carcinogenesis 2019; 40: 521-528.
[29] Yang W, Liu H, Duan B, Xu X, Carmody D, Luo S, Walsh KM, Abbruzzese JL, Zhang X and Chen X. Three novel genetic variants in NRF2 signaling pathway genes are associated with pancreatic cancer risk. Cancer Sci 2019; 110: 2022.
[30] Xu X, Qian D, Liu H, Cruz D, Luo S, Walsh KM, Abbruzzese JL, Zhang X and Wei Q. Genetic variants in the liver kinase B1-AMP-activated protein kinase pathway genes and pancreatic cancer risk. Mol Carcinog 2019; 58: 1338-1348.
[31] Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR and Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26: 2336-2337.
[32] Xu Z and Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 2009; 37: W600-W605.
[33] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC and Weng S. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012; 22: 1790-1797.
[34] Ward LD and Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2015; 44: D877-D881.
[35] Consortium GT. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348: 648-660.
[36] Lappalainen T, Sammeth M, Friedländer MR, Ac't Hoen P, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T and Ferreira PG. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013; 501: 506.
[37] Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014; 511: 543-550.
[38] Vorontsov IE, Kulakovskiy IV, Khimulya G, Nikolaeva DD and Makeev VJ. "PERFECTOS-APE-predicting regulatory functional effect of SNPs by approximate *P*-value estimation": SCITEPRESS), 102-108.
[39] PERFECTOS-APE. http://opera.autosome.ru/perfectosape/.
[40] Messeguer X, Escudero R, Farré D, Nuñez O, Martínez J and Albà MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics 2002; 18: 333-334.
[41] PROMO. http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3.
[42] Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVSK and Varambally S. UALCAN: a portal for facilitating tumor subgroup gene

expression and survival analyses. Neoplasia 2017; 19: 649-658.
[43] UALCAN. http://ualcan.path.uab.edu/index.html.
[44] Tang Z, Li C, Kang B, Gao G, Li C and Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017; 45: W98-W102.
[45] GEPIA. http://gepia.cancer-pku.cn/detail.php?gene.
[46] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-404.
[47] cBioPortal for Cancer Genomics (2015). https://www.cbioportal.org/.
[48] Walsh N, Zhang H, Hyland PL, Yang Q, Mocci E, Zhang M, Childs EJ, Collins I, Wang Z, Arslan AA, Beane-Freeman L, Bracci PM, Brennan P, Canzian F, Duell EJ, Gallinger S, Giles GG, Goggins M, Goodman GE, Goodman PJ, Hung RJ, Kooperberg C, Kurtz RC, Malats N, LeMarchand L, Neale RE, Olson SH, Scelo G, Shu XO, Van Den Eeden SK, Visvanathan K, White E, Zheng W; PanScan and PanC4 consortia, Albanes D, Andreotti G, Babic A, Bamlet WR, Berndt SI, Borgida A, Boutron-Ruault MC, Brais L, Brennan P, Bueno-de-Mesquita B, Buring J, Chaffee KG, Chanock S, Cleary S, Cotterchio M, Foretova L, Fuchs C, M Gaziano JM, Giovannucci E, Goggins M, Hackert T, Haiman C, Hartge P, Hasan M, Helzlsouer KJ, Herman J, Holcatova I, Holly EA, Hoover R, Hung RJ, Janout V, Klein EA, Kurtz RC, Laheru D, Lee IM, Lu L, Malats N, Mannisto S, Milne RL, Oberg AL, Orlow I, Patel AV, Peters U, Porta M, Real FX, Rothman N, Sesso HD, Severi G, Silverman D, Strobel O, Sund M, Thornquist MD, Tobias GS, Wactawski-Wende J, Wareham N, Weiderpass E, Wentzensen N, Wheeler W, Yu H, Zeleniuch-Jacquotte A, Kraft P, Li D, Jacobs EJ, Petersen GM, Wolpin BM, Risch HA, Amundadottir LT, Yu K, Klein AP and Stolzenberg-Solomon RZ. Agnostic pathway/gene set analysis of genome-wide association data identifies associations for pancreatic cancer. J Natl Cancer Inst 2019; 111: 557-567.
[49] Shintani T and Klionsky DJ. Autophagy in health and disease: a double-edged sword. Science 2004; 306: 990-995.
[50] Liang XH, Jackson S, Seaman M, Brown K, Kempkes B, Hibshoosh H and Levine B. Induction of autophagy and inhibition of tumorigenesis by beclin 1. Nature 1999; 402: 672-676.
[51] Galluzzi L, Pietrocola F, Bravo-San Pedro JM, Amaravadi RK, Baehrecke EH, Cecconi F, Codogno P, Debnath J, Gewirtz DA and Karantza V. Autophagy in malignant transformation and cancer progression. EMBO J 2015; 34: 856-880.
[52] Rosenfeldt MT, O'Prey J, Morton JP, Nixon C, MacKay G, Mrowinska A, Au A, Rai TS, Zheng L and Ridgway R. p53 status determines the role of autophagy in pancreatic tumour development. Nature 2013; 504: 296-300.
[53] Payankaulam S, Raicu AM and Arnosti DN. Transcriptional regulation of INSR, the insulin receptor gene. Genes 2019; 10: 984.
[54] Hansen SK, Párrizas M, Jensen ML, Pruhova S, Ek J, Boj SF, Johansen A, Maestro MA, Rivera F, Eiberg H, Andel M, Lebl J, Pedersen O, Ferrer J and Hansen T. Genetic evidence that HNF-1alpha-dependent transcriptional control of HNF-4alpha is essential for human pancreatic beta cell function. J Clin Invest 2002; 110: 827-833.
[55] Okada T, Liew CW, Hu J-J, Hinault C, Michael M, Krützfeldt J, Yin C, Holzenberger M, Stoffel MH and Kulkarni R. Insulin receptors in β-cells are critical for islet compensatory growth response to insulin resistance. Proc Natl Acad Sci U S A 2007; 104: 8977-8982.
[56] Kulkarni RN, Brüning JC, Winnay JN, Postic C, Magnuson MA and Kahn CR. Tissue-specific knockout of the insulin receptor in pancreatic β cells creates an insulin secretory defect similar to that in type 2 diabetes. Cell 1999; 96: 329-339.
[57] Wang J, Gu W and Chen C. Knocking down insulin receptor in pancreatic beta cell lines with lentiviral-small hairpin RNA reduces glucose-stimulated insulin secretion via decreasing the gene expression of insulin, GLUT2 and Pdx1. Int J Mol Sci 2018; 19: 985.
[58] Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML and Mora S. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013; 45: 1274.
[59] Son HY, Hwangbo Y, Yoo SK, Im SW, Kwak SJ, Park MS, Kwak SH, Cho SW, Ryu JS and Kim J. Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thyroid cancer. Nat Commun 2017; 8: 15966.
[60] Low SK, Kuchiba A, Zembutsu H, Saito A, Takahashi A, Kubo M, Daigo Y, Kamatani N, Chiku S and Totsuka H. Genome-wide association study of pancreatic cancer in Japanese population. PLoS One 2010; 5: e11824.
[61] Takekawa M, Posas F and Saito H. A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. EMBO J 1997; 16: 4973-4982.

[62] Liang Q, Yao X, Tang S, Zhang J, Yau TO, Li X, Tang CM, Kang W, Lung RWM and Li JW. Integrative identification of Epstein-Barr virus-associated mutations and epigenetic alterations in gastric cancer. Gastroenterology 2014; 147: 1350-1362.

[63] Takekawa M, Tatebayashi K, Itoh F, Adachi M, Imai K and Saito H. Smad-dependent GADD45β expression mediates delayed activation of p38 MAP kinase by TGF-β. EMBO J 2002; 21: 6473-6482.

[64] Yang LX, Gao Q, Shi JY, Wang ZC, Zhang Y, Gao PT, Wang XY, Shi YH, Ke AW and Shi GM. Mitogen-activated protein kinase kinase kinase 4 deficiency in intrahepatic cholangiocarcinoma leads to invasive growth and epithelial-mesenchymal transition. Hepatology 2015; 62: 1804-1816.

Table S1. Distributions of demographic characteristics among the three PanC datasets

Characteristic	PanScan I			PanScan II/III			PanC4[1]		
	Case (%)	Control (%)	p^2	Case (%)	Control (%)	p^2	Case (%)	Control (%)	p^2
All subjects	1760	1780		2995	1666		3722	3500	
Age			0.009			<.0001			<.0001
<60	330 (18.8)	265 (14.9)		820 (27.3)	564 (33.9)		1134 (30.5)	1260 (35.9)	
60-70	678 (38.5)	715 (40.2)		973 (32.5)	548 (32.9)		1344 (36.1)	1170 (33.4)	
>70	752 (42.7)	800 (44.9)		1202 (40.1)	554 (33.2)		1241 (33.4)	1068 (30.6)	
Sex			0.719			0.229			0.110
Male	899 (51.1)	921 (51.7)		1549 (51.7)	893 (53.6)		2144 (57.6)	1953 (55.7)	
Female	861 (48.9)	859 (48.3)		1446 (48.3)	773 (46.4)		1578 (42.4)	1547 (44.3)	

[1]Five missing date (three in cases and two in controls) of age in the PanC4 dataset were excluded. [2]Chi-square test.

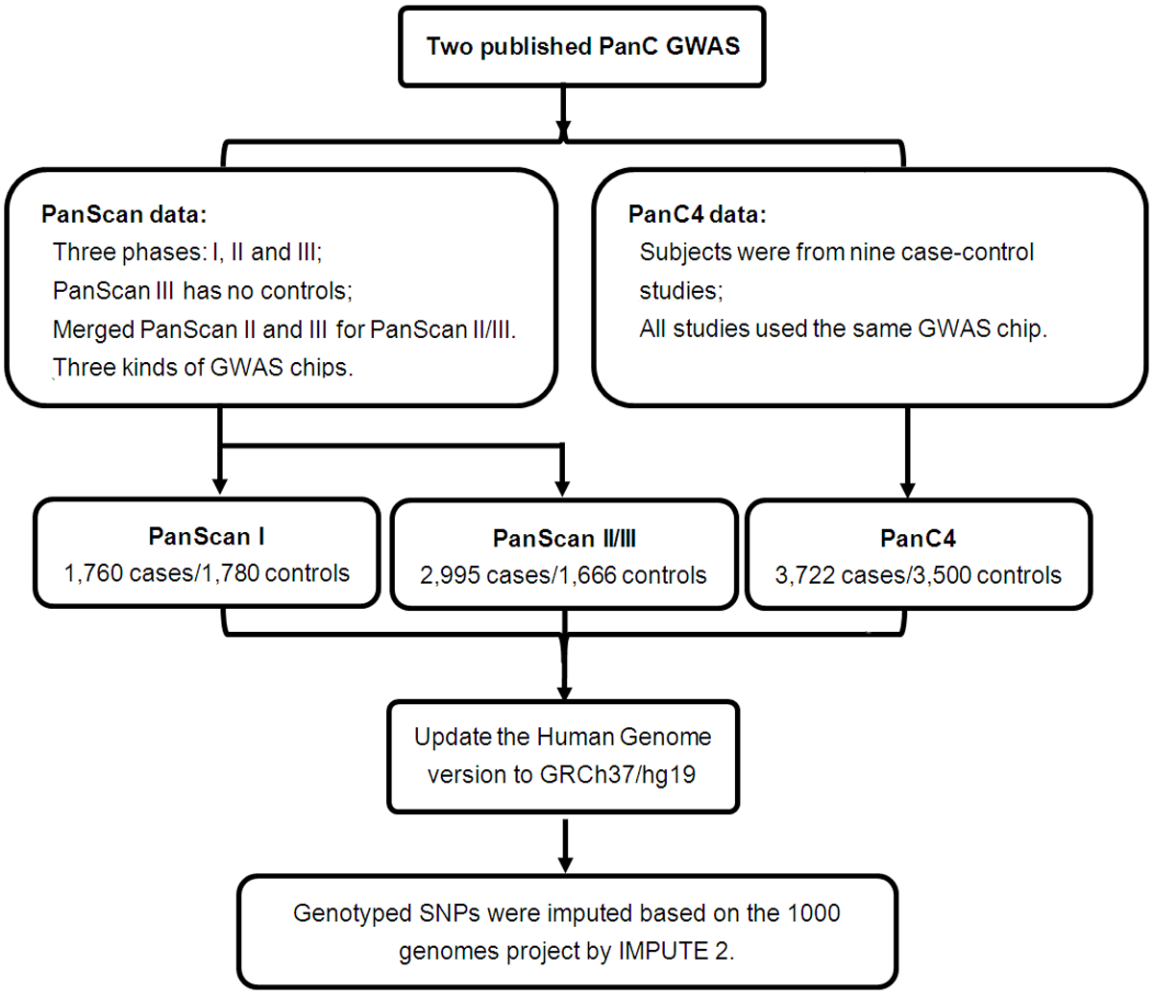


Figure S1. Workflow of the data process.

Table S2. List of 198 selected genes in the ATM pathway

Dataset	Name of the pathway	Gene Number	Selected genes[1]
MSigDB[2]			*ABL1, AKT1, AKT2, AKT3, AP3B2, APC, ATF2, ATF4, ATM, ATMIN, ATR, AURKB, AXIN1, BAD, BAK1, BAX, BBC3, BCL2, BCL2L11, BCL6, BID, BIK, BLM, BRCA1, BUB1, CASP2, CAT, CCNB1, CCNB2, CCND1, CCND2, CCND3, CCNE1, CCNG2, CDC25A, CDC25C, CDC42, CDK1, CDK2, CDK5, CDK5R1, CDKN1A, CDKN1B, CDKN2A, CEP63, CHEK1, CHEK2, CRADD, CREB1, CREB3, CREB3L4, CTBP1, CTNNB1, DCLRE1C, DVL1, DVL2, DVL3, ERBB2, FAM175A[4], FANCD2, FASLG, FOSL1, FOXO3, FRAT1, G6PC, GADD45A, GADD45B, GADD45G, GRB2, GSK3B, H2AFX[5], HDAC4, HMGB1, HMGN1, HRAS, HSPB1, INSR, IRS1, JUN, KAT5, KRAS, LBR, LDLR, LEF1, MAP3K1, MAP3K4, MAP3K5, MAP3K7, MAPK1, MAPK10, MAPK11, MAPK12, MAPK13, MAPK14, MAPK8, MAPK9, MDC1, MDM2, MDM4, MIR3191, MIR4741, MIR6808, MIR6886, MLKL, MRE11A[6], MTOR, MYC, NBN, NFKB1, NFKB2, NFKBIA, NRAS, PCK2, PDK1, PIDD1, PIK3C2A, PIK3C2B, PIK3C2G, PIK3C3, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R4, PIK3R5, PLAU, PMAIP1, PPM1D, PPP2R4[7], PPP2R5C, PPP2R5E, PPP5C, PRKAA1, PRKDC, PTEN, RAC1, RAC2, RAC3, RAD17, RAD50, RAD51, RAD9A, RASGRF1, RBBP8, RBL2, RELA, RFWD2[8], RHOA, RIF1, RIPK1, RNF168, RNF20, RNF40, RNF8, SCP2, SHC1, SMAD3, SMAD4, SMC2, SMC3, SOD2[9], SOS1, SOS2, STK11, TCF7, TCF7L1, TCF7L2, TERF2, TGFB1, TLK1, TOP3A, TP53, TP53BP1, TP73, TRIM28, TSC2, UBE2N, UIMC1, WNT1, WNT10A, WNT10B, WNT11, WNT16, WNT2, WNT2B, WNT3, WNT3A, WNT4, WNT5A, WNT5B, WNT6, WNT7A, WNT7B, XRCC4, YWHAB*
PID	ATM_PATHWAY	34	
BIOCARTA	ATM_PATHWAY	20	
GO	NO DATA	0	
REACTOME	NO DATA	0	
KEGG	NO DATA	0	
GeneCards[3]			
PathCards	ATM Pathway	54	
PathCards	DNA Damage Response (only ATM dependent)	115	
PathCards	ATM Signaling Network in Development and Disease	46	
Total		198	

Abbreviations: MSigDB, The Molecular Signatures Database. [1]After removing 66 replicated genes, four genes in X chromosome (*CCNB3, G6PD, IKBKG* and *SMC1A*) and one gene no information in hg19 (*LOC101929777*). [2]http://software.broadinstitute.org/gsea/msigdb/search.jsp (v7.0). [3]https://www.genecards.org/. [4]*FAM175A* also known as *ABRAXAS1*. [5]*H2AFX* also known as *H2AX*. [6]*MRE11A* also known as *MRE11*. [7]*PPP2R4* also known as *PTPA*. [8]*RFWD2* also known as *COP1*. [9]*SOD2* also known as *SOD2-OT1*. Keyword: ATM. Organism: Homo sapiens.

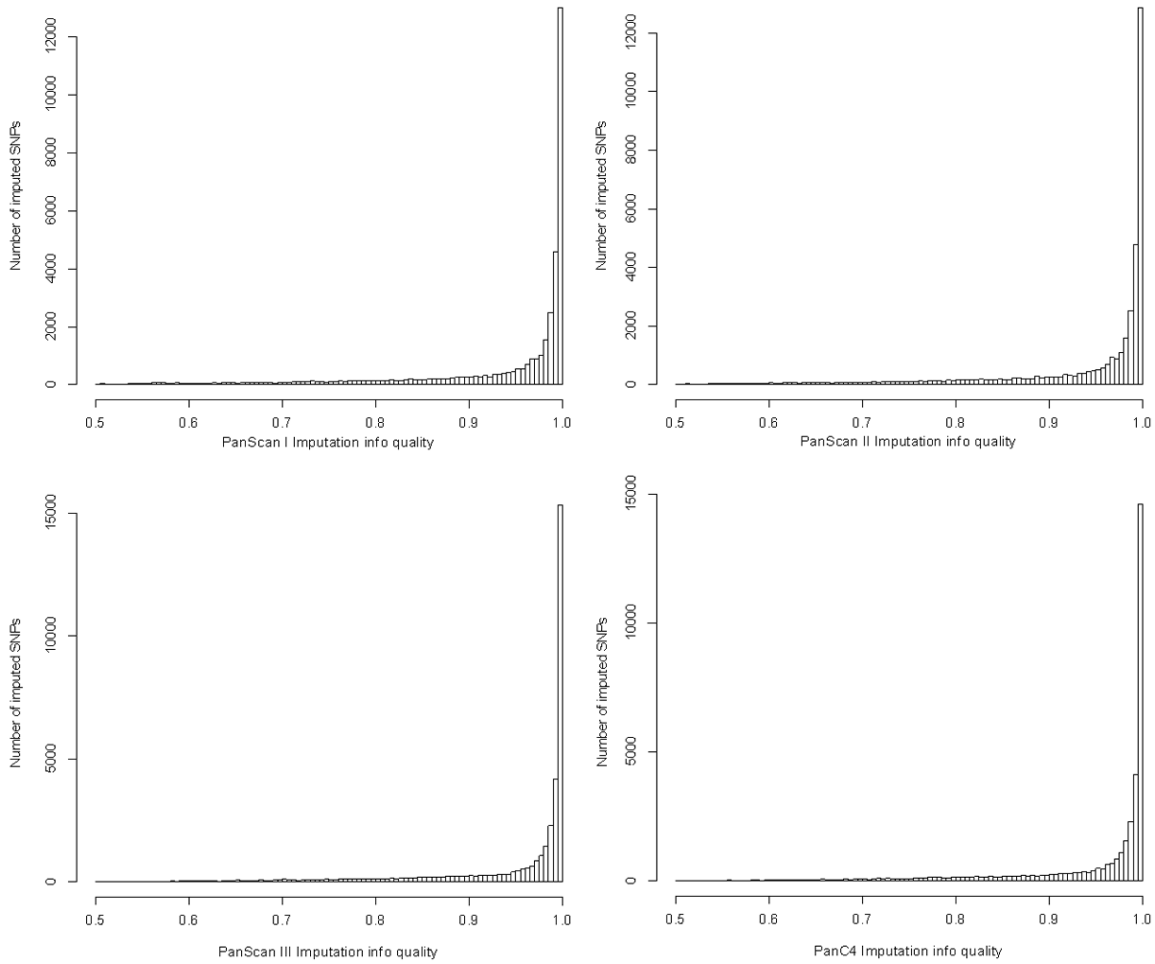
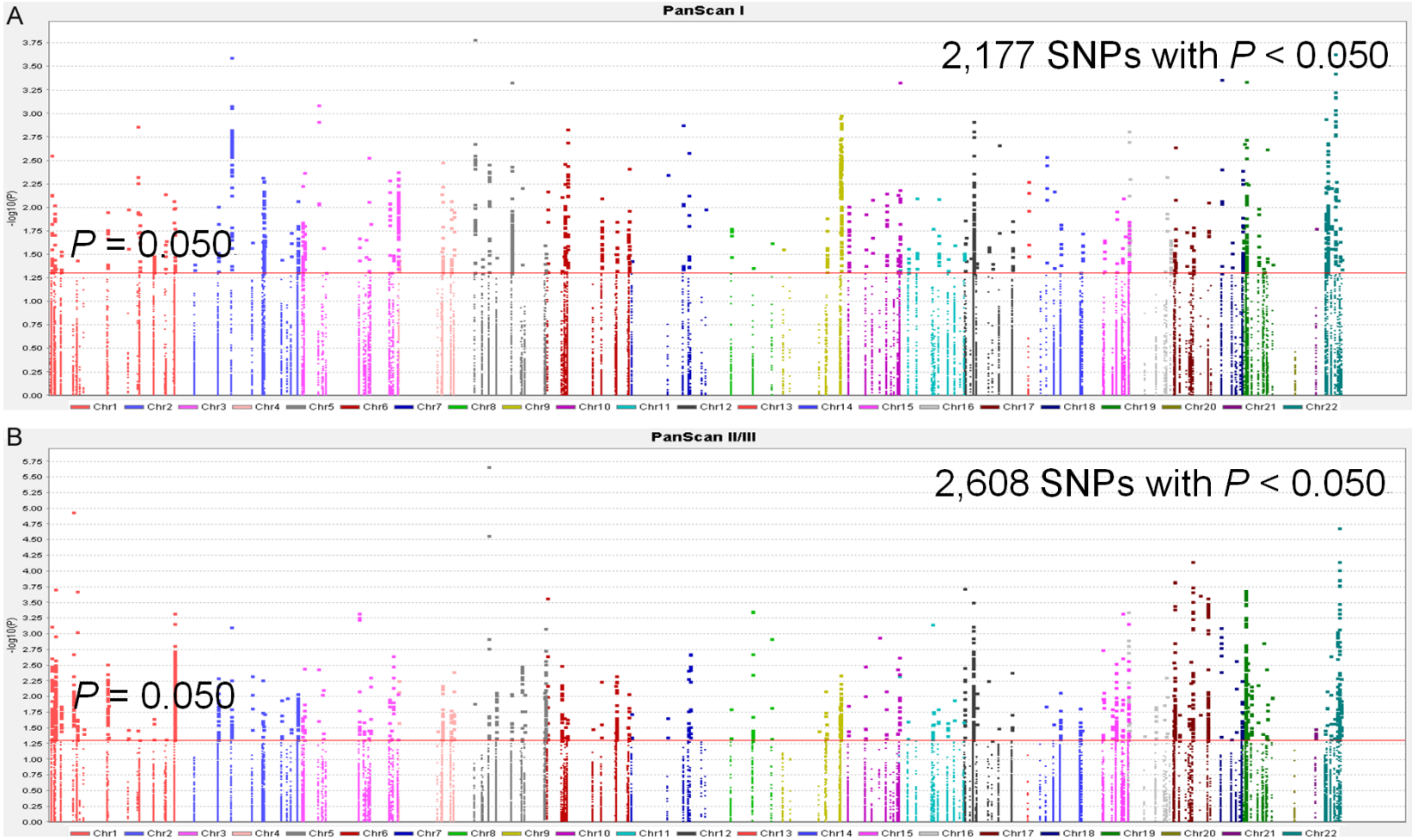


Figure S2. Distribution plot for imputation Info Quality in the present study.


A
PanScan I
2,177 SNPs with P < 0.050
P = 0.050
-log10(P)
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22
B
PanScan II/III
2,608 SNPs with P < 0.050
P = 0.050
-log10(P)
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22

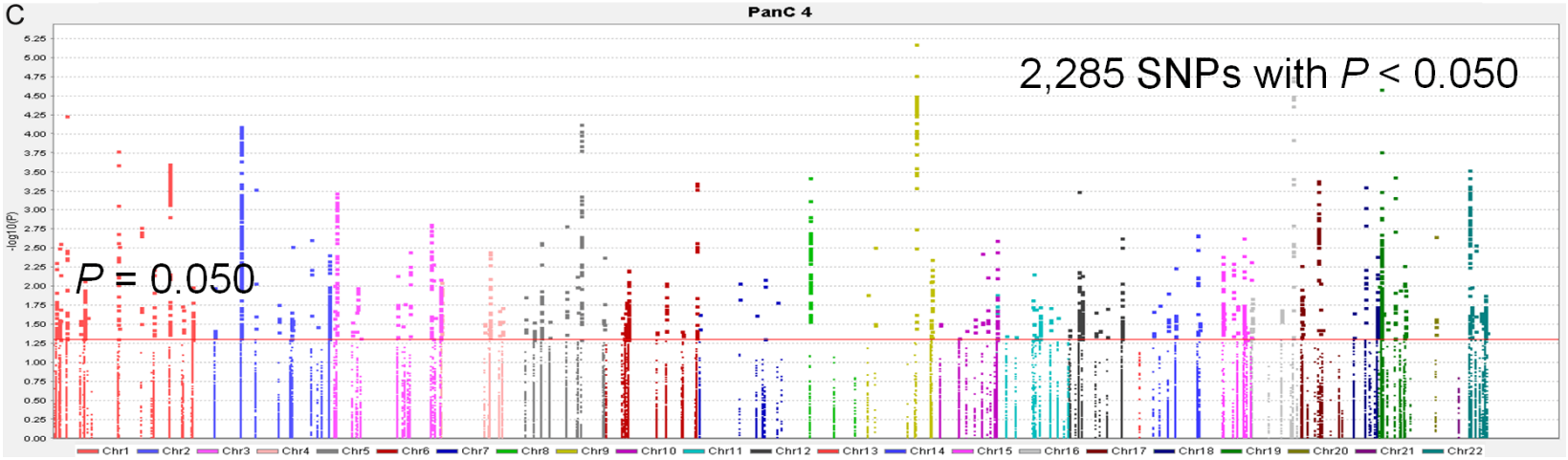


Figure S3. Manhattan plot of the association results in the three datasets. The statistical values across the autosomes of associations between SNPs and PanC risk are plotted as -log10 *p* values. The red horizontal line indicates *P*=0.050. There are (A) 2,177 SNPs with *P*<0.050 in PanScan I, (B) 2,608 SNPs with *P*<0.050 in PanScan II/III and (C) 2,285 SNPs with *P*<0.050 in PanC4. Abbreviations: SNP, single nucleotide polymorphism; PanC, pancreatic cancer.

Table S3. Independent PanC risk loci identified in both the present study and the previously reported studies[1]

Region[1]	Gene	SNP	Allele[2]	Position	*P*[3]
1p13.2	*WNT2B*	rs3838412	-/T	113046730	0.0003
9q31.1	*SMC2*	rs2417487	A/G	106887581	0.0040
17q12	*ERBB2*	rs2517955	C/T	37843681	0.0009
22q12.1	*CHEK2*	rs2236141	T/C	29137870	<.0001

Abbreviations: PanC, pancreatic cancer; SNP, single nucleotide polymorphism. [1]PMID: 26098869, 29422604, 30541042 and 25086665. [2]Effect (minor) allele/reference allele. [3]Adjusted by age, sex, the top five significant principal components and 13 reported SNPs. (rs35075084, rs2727572, rs34852782, rs62068300, rs3751936, rs3124761, rs17458086, rs1630747, rs5757573, rs6001516, rs79447092, rs9895829 and rs3818626) in the same model (PMID: 29168174, 30794721, 30972876 and 30997723, respectively).

Table S4. Stratified analysis for associations between NUG and PanC risk by age and sex

Characteristic	NUG 0-1		NUG 2-3		Univariate analysis		Multivariate analysis[1]		p_{inter}[2]
	Case (%)	Control (%)	Case (%)	Control (%)	OR (95% CI)	*p*	OR (95% CI)	*p*	
Age (years)									0.8218
<60	914 (27.1)	940 (30.3)	1367 (26.8)	1146 (29.9)	1.23 (1.09-1.38)	0.0009	1.23 (1.09-1.39)	0.0007	
60-70	1203 (35.6)	1082 (34.8)	1792 (35.2)	1350 (35.2)	1.19 (1.07-1.33)	0.0013	1.19 (1.07-1.33)	0.0016	
>70	1260 (37.3)	1084 (34.9)	1934 (38.0)	1335 (34.8)	1.25 (1.20-1.39)	<.0001	1.25 (1.12-1.39)	<.0001	
Sex									0.3064
Male	1825 (54.0)	1654 (53.2)	2764 (54.2)	2110 (55.1)	1.19 (1.09-1.30)	0.0001	1.19 (1.09-1.29)	0.0001	
Female	1553 (46.0)	1454 (46.8)	2331 (45.8)	1721 (44.9)	1.27 (1.15-1.39)	<.0001	1.27 (1.15-1.40)	<.0001	

Abbreviations: NUG: number of unfavorable genotypes; PanC, pancreatic cancer; OR, odds ratio; CI, confidence interval. [1]Adjusted for age, sex and the top five principal components. [2]P_{inter}: *P* value for interaction analysis between characteristics and NUG.

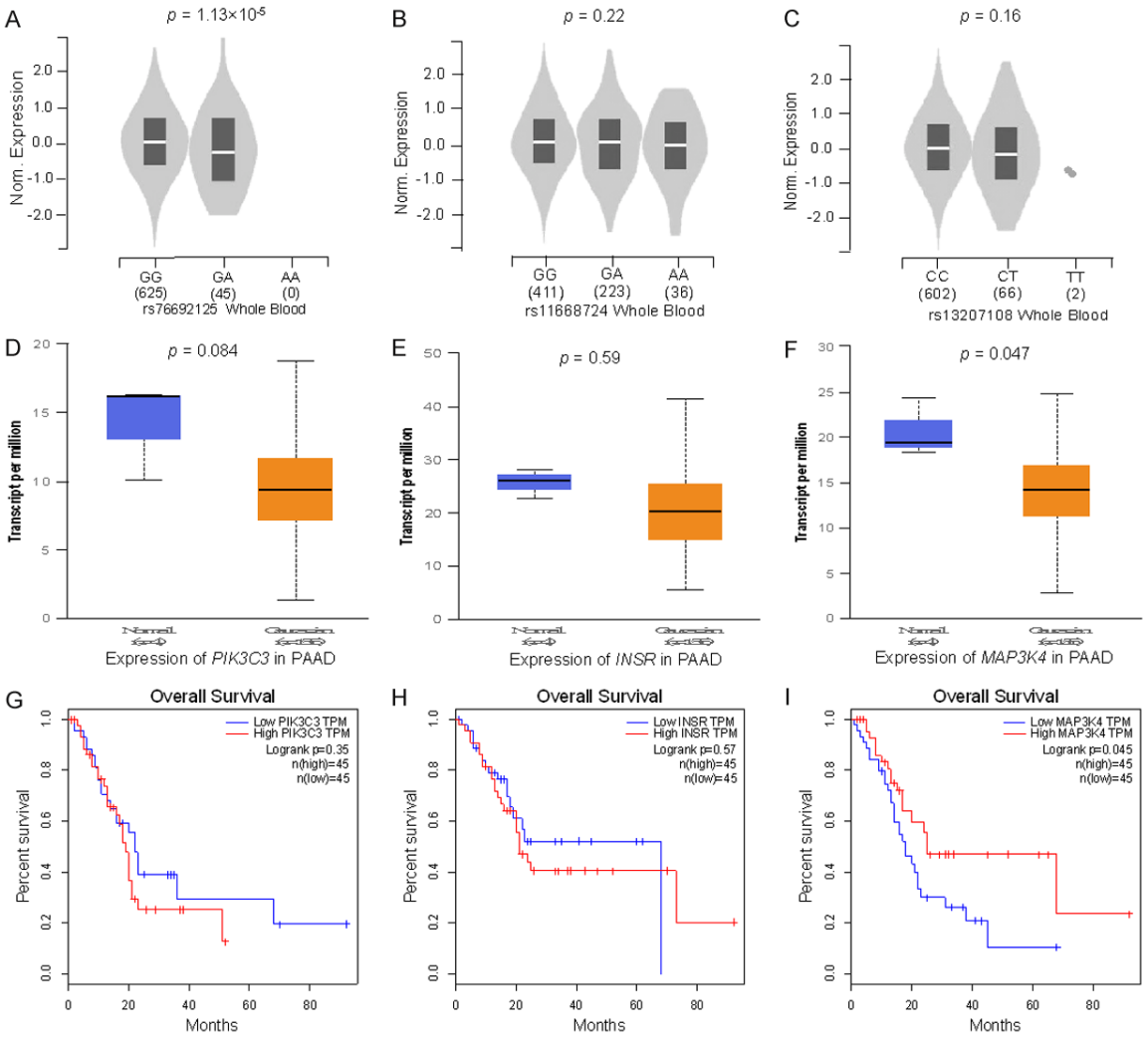


Figure S4. Functional analyses. Genotypes and mRNA expression levels in the whole blood cells from the GTEx Project: (A) *PIK3C3* rs76692125 (n=670, *P*=1.13×10^{-5}); (B) *INSR* rs11668724 (n=670, *P*=0.22) and (C) *MAP3K4* rs13207108 (n=670, *P*=0.16). The mRNA expression levels of (D) *PIK3C3* and (E) *INSR* and (F) *MAP3K4* between normal and pancre; tic adenocarcinoma (PAAD) tissues from the TCGA database (n=160, *P*=0.084, 0.59 and 0.047, respectively). Overall Survival analysis for PAAD patients in *PIK3C3*, *INSR* and *MAP3K4* from TCGA PAAD samples by using the GEPIA online survival analysis tool: (G) *PIK3C3* and (H) *INSR* expression were not significantly associated with survival of PAAD patients; (I) high *MAP3K4* expression was associated with a better survival of PAAD patients compared with low expression.

Table S5. Functional prediction of the three novel independent SNPs in the present study

SNP	Chr	Gene	SNPinfo[1]	RegulomeDB score[2]	HaploReg v4.1[3]						
					Promoter histone marks	Enhancer histone marks	DNase	Proteins bound	Motifs changed	Selected eQTL hits	dbSNP func annot
rs76692125	18q12.3	*PIK3C3*	--	6	--	BLD	--	--	FXR, Hltf, PLZF, STAT	--	intronic
rs11668724	19p13.2	*INSR*	--	4	--	10 tissues	8 tissues	TCF12	E2A	2 hits	intronic
rs13207108	6q26	*MAP3K4*	--	4	--	SKIN, BLD	15 tissues	8 tissues	CFOS	NF-Y	intronic

Abbreviations: SNP, single nucleotide polymorphism; chr, chromosome; eQTL, expression quantitative trait loci; dbSNP func annot, dbSNP function annotation. [1]https://snpinfo.niehs.nih.gov/snpinfo/snpfunc.html. [2]http://regulomedb.org/. [3]http://archive.broadinstitute.org/mammals/haploreg/haploreg.php.

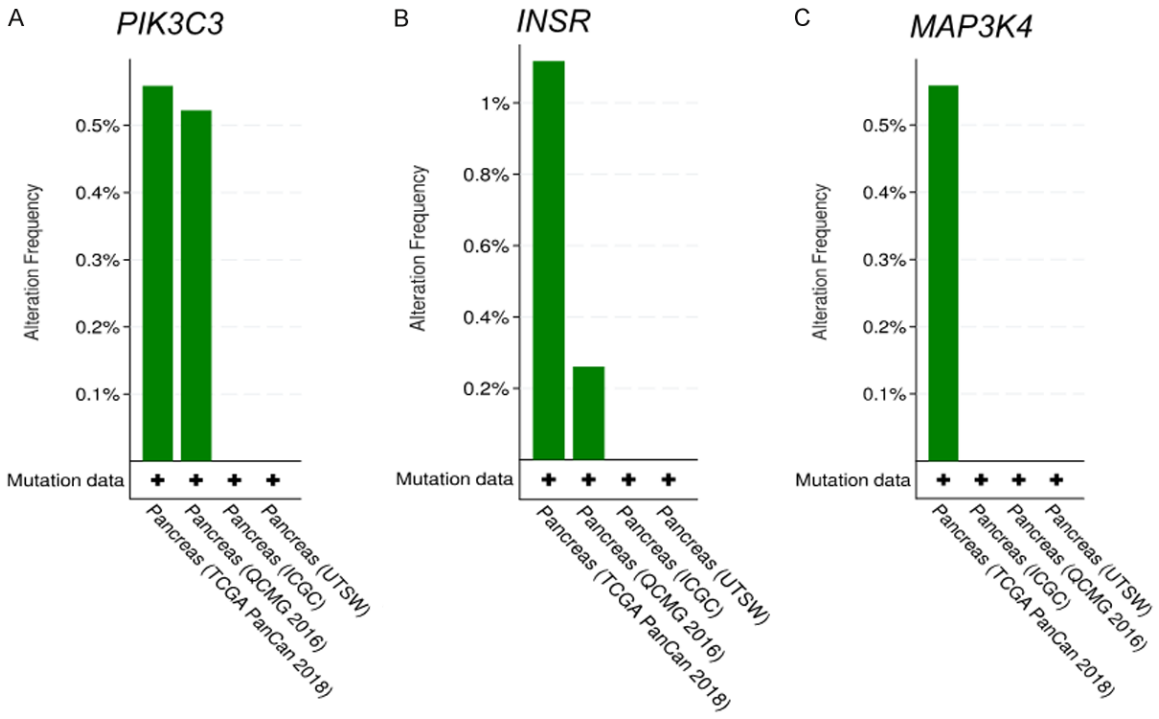


Figure S5. Mutation analyses of *PIK3C3*, *INSR* and *MAP3K4* genes in pancreatic adenocarcinoma (PAAD) tissues by using available data from the cBioportal for Cancer Genomics database. Mutation rates of (A) *PIK3C3* [0.56% (1/179), 0.52% (2/383), 0% (0/99) and 0% (0/109)], (B) *INSR* [1.12% (2/179), 0.26% (1/383), 0% (0/99) and 0% (0/109)] and (C) *MAP3K4* [0.56% (1/179), 0% (0/383), 0% (0/99) and 0% (0/109)] in PAAD from TCGA PanCan 2018, QCMG 2016, ICGC and UTSW studies, respectively.