Exome-wide analysis identifies three low-frequency missense variants associated with pancreatic cancer risk in Chinese populations

Jiang Chang et al.

Germline coding variants have not been systematically investigated for pancreatic ductal adenocarcinoma (PDAC). Here we report an exome-wide investigation using the Illumina Human Exome Beadchip with 943 PDAC cases and 3908 controls in the Chinese population, followed by two independent replicate samples including 2142 cases and 4697 controls. We identify three low-frequency missense variants associated with the PDAC risk: rs34309238 in *PKN1* (OR = 1.77, 95% CI: 1.48–2.12, P = 5.35 × 10^{-10}), rs2242241 in *DOK2* (OR = 1.85, 95% CI: 1.50–2.27, P = 4.34 × 10^{-9}), and rs183117027 in *APOB* (OR = 2.34, 95% CI: 1.72–3.16, P = 4.21 × 10^{-8}). Functional analyses show that the *PKN1* rs34309238 variant significantly increases the level of phosphorylated PKN1 and thus enhances PDAC cells’ proliferation by phosphorylating and activating the FAK/PI3K/AKT pathway. These findings highlight the significance of coding variants in the development of PDAC and provide more insights into the prevention of this disease.
Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal human cancers, with a 5-year overall survival rate of only ~5%\textsuperscript{1,2}. The incidence of PDAC is rapidly increasing worldwide, and prevention or early diagnosis at a curable stage remains exceedingly difficult for this disease. Therefore, PDAC has become the fourth leading cause of cancer-associated death in both men and women\textsuperscript{3,4}. Cigarette smoking, type 2 diabetes, obesity and several hereditary cancer syndromes represent major risk factors for PDAC\textsuperscript{5–7}. Based on accumulating evidence, germline variants also play an important role in the development of this disease\textsuperscript{8}.

In previous genome-wide association studies (GWAS) from our group and other researchers, several susceptibility loci associated with PDAC risk were identified in populations of Asian and European ancestry populations\textsuperscript{9–15}. However, GWAS exclusively focused on common single-nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 5%, and the identified variants explained only a small fraction of the heritability for PDAC\textsuperscript{16,17}. Low-frequency variants (defined here as an MAF of 0.1%–5%) or rare variants (defined here as a MAF < 0.1%) have essential effect size and may substantially contribute to the “missing” heritability\textsuperscript{16,18}. Therefore, identifying additional low-frequency or rare variants that increase the susceptibility to PDAC will deepen our understanding of the aetiology of this disease.

The Illumina HumanExome Beadchip (referred to as “exome chip” hereafter) platform is one approach that primarily focuses on low-frequency or rare variants in the exon regions of genes, which has been successfully used in numerous studies to identify a series of functional coding variants\textsuperscript{19–21}. In this study, we performed an exome-wide association analyses using this chip with 943 individuals with PDAC and 3908 healthy controls to identify protein-coding susceptibility loci in the Chinese population, followed by two independent replicate sample including 2142 cases and 4697 controls. We identified three low-frequency missense variants in the protein kinase N1 (PKN1), the docking protein 2 (DOK2) and the apolipoprotein B (APOB) associated with the PDAC risk with genome-wide significance and relatively high effect sizes (odds ratio (OR) > 1.5) by an additive model in logistic regression analysis. Further functional analyses show that the PKN1 rs34309238 variant increases the level of phosphorylated PKN1 and thus enhances cells’ proliferation by phosphorylating and activating the focal adhesion kinase (FAK)/phosphatidylinositol-3 kinase (PI3K)/AKT signalling pathway. These findings highlight the significance of low-frequency missense variants in the development of PDAC and provide more insights into the prevention of this disease.

Results

Three low-frequency missense SNPs were identified for PDAC.

In the discovery stage of this study, we performed an exome-wide association analyses in 943 individuals with PDAC and 3908 healthy controls (Supplementary Fig. 1 and Supplementary Table 1), and the cases and controls of Han Chinese ancestry were well matched (Supplementary Figs. 2, 3). The overall association \( P \) values are presented in Fig. 1, and 25 variants exhibited a promising association, with \( P < 1 \times 10^{-4} \) by an additive model in logistic regression analysis (Supplementary Table 2). We therefore chose these 25 variants for further replication in 1048 cases and 2094 controls from Wuhan, and significant associations with four coding variants were verified (Supplementary Table 3). These four variants were then replicated in stage II with 1094 cases and 2603 controls from Shandong and Hebei provinces (Supplementary Table 3). These four variants were all associated with PDAC risk with same direction for both the two replication stages, and three of them with \( P < 0.05 \) in both the two replication stages by an additive model in logistic regression analysis. When combining the results from the discovery and replication stages, we identified three low-frequency coding variants that were significantly associated with the risk of PDAC and displayed \( P \) values reaching genome-wide significance by an additive model in logistic regression analysis (Table 1 and Supplementary Table 4). The most significant association was noted for rs34309238, which is located in the 11th exon of PKN1 in chromosome 19p13.12 (OR = 1.77, 95% confidence interval (CI) 1.48–2.12, \( P = 5.35 \times 10^{-10} \)) by an additive model in logistic regression analysis. The rs2242241 variant in the fourth exon of DOK2 and rs183117027 variant in the 28th exon of APOB were also associated with an increased risk of PDAC, with ORs being 1.85 (95% CI 1.50–2.27, \( P = 4.34 \times 10^{-8} \)) and 2.34 (95% CI 1.72–3.16, \( P = 4.21 \times 10^{-9} \)) by an additive model in logistic regression analysis, respectively.

No other independent signals in the significant regions. We performed an imputation analysis for the identified three regions to investigate whether the association of each of the three susceptibility regions with PDAC risk was completely explained by the index SNP. After imputation, we tested 6675 SNPs (108 directly genotyped and 6567 well-imputed SNPs) for the association with these three regions. Only two imputed variants passed our significance threshold in the discovery stage (\( P < 1 \times 10^{-4} \) by an additive model in logistic regression analysis), and they were all in high linkage disequilibrium (LD) with the identified SNP (Supplementary Figs. 4–6 and Table 5). After conditioning with each of the three SNPs, the \( P \) values for the association of those SNPs in LD with the identified SNP were not <0.05, suggesting that the association signals in these regions probably point towards these three SNPs identified by genotyping (Supplementary Table 5).

No other signals were identified by gene-based analysis. We performed a gene-based analysis to identify significant susceptible variants enriched in genes using two methods: a simple burden

![Fig. 1 Manhattan plot for associations between genetic variants and pancreatic cancer risk. A total of 43,045 variants that passed the quality control and with MAF > 0.1% in controls were analysed and plotted. The associations (\(-\log_{10}(P)\) values, y axis) are plotted against genomic position (x axis by chromosome and the chromosomal position of NCBI build 37). The red horizontal line corresponds to a \( P \) value threshold of 1.00 \( \times \) 10\(^{-4}\). Variants that passed the threshold and were successfully verified in this study were annotated](https://www.nature.com/naturecommunications/)

\[\text{ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06136-x} | \text{www.nature.com/naturecommunications}\]
The rs34309238 variant may affect PKN1 phosphorylation. Among the three identified variants, the rs34309238 variant (Leu555Ile change) in PKN1 exhibited the most significant signal by an additive model in logistic regression analysis in this study. This variant was predicted to be probably damaging (score = 0.997 calculated by PolyPhen2). Three previously reported phosphorylation sites (S559, S561 and S562, annotated by the PhosphoSitePlus database22 with >5 references) are located near this variant, and thus their phosphorylation levels may be affected (Fig. 2a). Furthermore, by using data obtained from the Oncomine database and the Human Protein Atlas database, we found that PKN1 expression was upregulated in numerous cancers, such as glioma, kidney cancer, ovarian cancer, prostate cancer, and PDAC (Supplementary Fig. 7). Therefore, we conjectured that the rs34309238 C>A (Leu>lile) change might affect PDAC risk by affecting the level of phosphorylated PKN1 and the phosphorylation of PKN1 activates its downstream signalling pathway.

The rs34309238-A activates the PKN1/FAK/PI3K/AKT pathway. To elucidate the function of rs34309238 in the development of PDAC, we performed an isobaric Tag for Relative and Absolute Quantitation (iTRAQ)-based comparative proteomics screen of PDAC, we performed an isobaric Tag for Relative and Absolute Quantitation (iTRAQ)-based comparative proteomics screen for PANC-1 cells after transfection with a pcDNA3.1 plasmid containing PKN1 or the control vector. The levels of phosphorylated PKN1 at Ser561 and Ser562 (near the rs34309238 Leu555Ile change) were increased upon transfection with PKN1[A], compared with cells transfected with PKN1[C] or the control vector (Fig. 2b). We also observed increased levels of phosphorylated FAK (FAK Tyr397), PI3K regulatory subunit alpha (PI3K1β), 3-phosphoinositide-dependent protein kinase 1 (PDK1) and serine/threonine kinase 1 (AKT) upon transfection with PKN1[A] in the iTRAQ-based screen (Fig. 2b). The phosphorylation level of FAK Tyr397 and AKT Ser473 were further successfully validated using western blot (Fig. 2c). Meanwhile, knockdown of PKN1 by small interfering RNAs (siRNAs) reduced the levels of phosphorylated FAK Tyr397 and AKT Ser473 (Fig. 2c).

The rs34309238-A promotes PDAC cells’ proliferation. To further characterize the function of PKN1 variant in PDAC cells, we overexpressed different PKN1 variants in PANC-1 and BxPC-3 cells and tested the rate of cell proliferation. The PKN1[A] overexpression significantly enhanced PANC-1 and BxPC-3 cells’ proliferation compared with overexpression of PKN1[C] or the control vector by two-sided unpaired Student’s t test (Fig. 3a, b). In contrast, knockdown of PKN1 by two PKN1 siRNAs reduced PANC-1 and BxPC-3 cells’ proliferation (Fig. 3c, d). We also selected two previously reported PKN1 inhibitors (Lestaurtinib and Ro318220)23–25 and tested whether they can reduce PDAC cells’ proliferation. The results exhibited that both Lestaurtinib and Ro318220 inhibited PANC-1 and BxPC-3 cells’ proliferation by reducing the levels of phosphorylated FAK/PI3K/AKT signalling pathway (Figs. 2d and 3e, f).

Discussion

In this study, we used the illumina HumanExome Beadchip to perform an exome-wide interrogation of coding susceptibility loci for PDAC. This chip was designed based on exome sequencing data of ~12,000 individuals from the European, African, Chinese and Hispanic population. The chip consists of >240,000 markers that is estimated to include 97–98% of the nonsynonymous variants detected in average genome through exome sequencing. By using this approach, we identified three low-frequency missense variants associated with PDAC risk in a total of 3085 cases and 8605 controls. No common variants were successfully verified under the significant threshold (P < 0.0001 by an additive model in logistic regression analysis) in the discovery stage of this study. Potential susceptible variants with P values between 0.05 and 0.0001 still need to be investigated in future studies.

Among these three identified variants, the rs34309238 variant (Leu555Ile change) in PKN1 exhibited the most significant signal. The PKN1 located in the 19p13.12 region, which contains an important G protein-coupled receptor and cancer metastasis-related gene, CD97. The PKN1 belongs to the protein kinase C superfamily and is activated upon binding a member of the Rho family of small G proteins, such as Ras-related C3 botulinum toxin substrate 1 (RAC1), which is required for KRAS-induced pancreatic tumorigenesis in mice26. Through a series of functional analyses, we found that the rs34309238 variant influences the risk of PDAC by altering the level of phosphorylated PKN1 at small interfering RNAs (siRNAs) reduced the levels of phosphorylated FAK Tyr397 and AKT Ser473 (Fig. 2c).

Table 1 The identified variants associated with pancreatic cancer risk in the discovery, replication and combined samples

| Chr | SNP    | Gene | Allele | Variation | Stage           | MAF  | OR (95% CI)  | P      |
|-----|--------|------|--------|-----------|----------------|------|--------------|--------|
|     |        |      |        |           | Discovery       |      |              |        |
| 19p13.12 | rs34309238 | PKN1 | C>A    | p.Leu555Ile | Discovery       | 0.033| 0.017        | 1.96 (1.44–2.67) | 1.79 × 10⁻⁵ |
|       |        |      |        |           | Replication I   | 0.031| 0.019        | 1.62 (1.16–2.26) | 0.0043  |
|       |        |      |        |           | Replication II  | 0.032| 0.018        | 1.75 (1.28–2.40) | 0.0005  |
|       |        |      |        |           | Combined        | 0.032| 0.018        | 1.77 (1.48–2.12) | 5.35 × 10⁻⁵ |
| 8p21.3 | rs2242241 | DOK2 | T>G    | p.Ser394Ala | Discovery       | 0.030| 0.013        | 2.47 (1.75–3.49) | 2.94 × 10⁻⁷ |
|       |        |      |        |           | Replication I   | 0.023| 0.014        | 1.64 (1.13–2.40) | 0.0101  |
|       |        |      |        |           | Replication II  | 0.022| 0.014        | 1.51 (1.06–2.15) | 0.0210  |
|       |        |      |        |           | Combined        | 0.025| 0.014        | 1.85 (1.50–2.27) | 4.34 × 10⁻⁹ |
|       |        |      |        |           | Discovery       | 0.016| 0.006        | 2.95 (1.83–4.74) | 8.05 × 10⁻⁶ |
|       |        |      |        |           | Replication I   | 0.011| 0.005        | 2.20 (1.21–3.98) | 0.0093  |
|       |        |      |        |           | Replication II  | 0.011| 0.005        | 2.01 (1.17–3.48) | 0.0120  |
|       |        |      |        |           | Combined        | 0.012| 0.006        | 2.34 (1.72–3.16) | 4.21 × 10⁻⁵ |

Legend: MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.
Ser561 and Ser562. The enhanced phosphorylation of PKN1 activates the FAK/PI3K/AKT signalling pathway and enhances PDAC cells’ proliferation. These results are consistent with previous findings that FAK contains a phosphorylation consensus motif for PKN1, and its phosphorylation enhances cancer development by activating the PI3K/AKT signalling pathway. Our results also suggested that PKN1 inhibitors Lestaurtinib and Ro318220 reduces PDAC cells’ proliferation by inhibition of PKN1-associated FAK/PI3K/AKT signalling pathway. Therefore, these two PKN1 inhibitors may serve as potential drugs for the treatment of PDAC.

In addition to the functional variant at PKN1, we also identified two low-frequency coding variants that were significantly associated with the risk of PDAC: rs2242241 in DOK2 and rs183117027 in APOB. The DOK2 gene is located on chromosome 8p21.3, which is frequently lost in multiple human cancers. The DOK family members DOK1, DOK2 and DOK3 are substrates of dozens of crucial protein tyrosine kinases and function in negative-feedback signalling loops that tightly modulate the duration and intensity of growth factor signalling. DOK2 is a tumour-suppressor gene in lung adenocarcinoma and myelomonocytic leukaemia. Except for DOK2, this region contains the glial cell line-derived neurotrophic factor family receptor alpha 2, which could prompt pancreatic cancer cell growth and chemoresistance through downregulating tumour-suppressor gene PTEN via Mir-17-5p.

The APOB gene is located in the 2p24.1 region, which contains a Barrett’s oesophagus susceptibility gene, the growth differentiation factor 7. The APOB encodes the main apolipoprotein of chylomicrons and low-density lipoproteins. APOB exists in plasma as two main isoforms: APOB-48 and APOB-100. The former is synthesized exclusively in the gut, and the latter is synthesized exclusively in the liver. Significant APOB alterations induced by somatic mutations (~10%) or downregulation by hypermethylation likely result in hepatocellular carcinoma by diverting energy into cancer-relevant metabolic pathways. Germline mutations in APOB also potentially cause diseases associated with lipid metabolic disorders, such as hypobetalipoproteinaemia and hypercholesterolaemia. Common polymorphisms in the APOB gene are associated with low-density lipoprotein cholesterol metabolism or the risk of coronary heart disease. Recently, an exome-wide associated study also identified that low-frequency coding variants in APOB is associated with plasma lipid level in 47,532 East Asian individuals.
Methods

Study subjects. We conducted a three-stage case–control study in the present work, and the study subjects and work flow are summarized in Supplementary Fig. 1 and Table 1. In the discovery stage, 943 patients with PDAC were recruited from Cancer Hospital, Chinese Academy of Medical Sciences in Beijing. The sample size has 52–98% power to detect variants with MAF ranging from 0.01 to 0.05 (OR = 1.8). In the replication stage I, 1048 patients with PDAC were recruited from multiple hospitals in Wuhan. The replication stage II contains 1094 patients from multiple hospitals in Shandong and Hebei province. The PDAC diagnosis was confirmed histopathologically or cytologically by at least two local pathologists, according to the World Health Organization classification. A subset of individuals was included in our previous studies. All the controls were cancer-free individuals selected from a community nutritional survey in the same region during the same period the patients were recruited. Demographic data were obtained from the medical records and interviews. Informed consent was obtained from all participants, and this study was approved by the institutional review board of each participating institution.

Genotyping and quality control. In the discovery stage, samples were genotyped using the Illumina HumanExome Beadchip system to identify potential susceptibility variants. The case and control samples were mixed and randomly allocated in the plates. All initial genotyping reactions of cases and controls were performed simultaneously on the same platform, and researchers performing the assays were blinded to the case/control status. Genotype calling and quality control procedures were performed according to a standard protocol.

In summary, a total of 174,391 variants were excluded from subsequent analyses because they (1) had duplicate variants on the chip (831 variants), (2) were mitochondrial variants or were located on the X or Y chromosome (1338 variants), (3) were monomorphic in our study subjects (171,141 variants), (4) had a call rate of <95% (761 variants), or (5) presented a minor allele frequency (MAF) of <0.1%, and <95% (Supplementary Fig. 1). We further excluded 30,434 variants with extremely rare MAF (<0.1%), and finally, 43,045 variants were analysed and plotted in the Manhattan plot (Supplementary Fig. 1).

Genotyping consistency in the discovery stage was assessed based on 300 replicate samples genotyped using both exome chip and Sequenom MassArray platform (San Diego, CA, USA) for the 25 promising variants, and the concordance rate of each variant was between 99.7% and 100%. A principal component analysis was performed using EIGENSOFT to determine ancestry and population stratiﬁcation based on 4431 autosomal informative ancestry markers included in the exome chip. We determined identity-by-state similarity to estimate the cryptic relatedness or duplication for each pair of samples using the PLINK software and no duplicated individuals (PL_HET > 0.25) were identiﬁed in this study.

In the first replication stage, 25 promising SNPs were genotyped in 1048 PDAC cases and 2094 controls using the Sequenom genotyping platform. In the second replication stage, 4 promising SNPs were genotyped in 1094 PDAC cases and 2603 controls using TaqMan assays platform (ABI 7900HT system, Applied Biosystems). Several genotyping quality controls were implemented in the replication stage, including (i) case and control samples were mixed in the plates, and persons who performed the genotyping assay were unaware of the case or control status; (ii)
positive and negative (no DNA) samples were included on every 384-well assay plate, and (iii) direct sequencing of PCR products was employed to replicate sets of 50 randomly selected, Sequenom-genotyped and TaqMan-genotyped samples, and the concordance rate of Sequenom and TaqMan platforms for each variant was between 98.0% and 100%.

Association analysis. We then performed an association analysis using an additive model in a logistic regression analysis with adjustments for age and sex as well as the first three principle components. A quantile–quantile (Q–Q) plot exhibited a good match between the distributions of observed P values and those expected by chance (inflation factor \( \lambda = 1.036 \); Supplementary Fig. 3). Variants with \( P < 0.0001 \) by an additive model in logistic regression analysis were considered significant and selected for further replication (Supplementary Table 2). In the replication stage, association analyses were performed using a logistic regression model adjusted for age and sex. Four variants with \( P < 0.05 \) in the first replication stage were further genotyped in the second replication stage (Supplementary Table 3). The association analysis was performed with PLINK (version 1.90) and R software (version 3.3.0). The Manhattan plot was generated using Haploview58, and Q–Q plot was generated with the R software (version 3.3.0).

Genotype imputation. We phased the haplotypes with SHAPEIT59 and performed imputations with the IMPUTE260,61 software to impute ungenotyped SNPs in a 1-Mb region centred on the three identified SNPs (Supplementary Figs. 4-6 and Table 4). This analysis was based on the LD and haplotypes information from the 1000 Genomes Project Phase 3 ASN samples as references. Poorly imputed variants with an information score (\( i(\theta)=0.40 \) output from IMPUTE2 info file) were excluded from subsequent analyses. Regional plots were created using LocusZoom62 with hg19/1000Genomes Nov 2014 ASN for the LD analysis.

Cell lines. PANC-1 and BxPC-3 cells were obtained from the China Center for Type Culture Collection (Shanghai, China) and were cultured in Dulbecco’s Modified Eagle’s Medium ( Gibco, Grand Island, NY, USA) supplemented with 10% foetal bovine serum (Gibco, Grand Island, NY, USA) and 1% antibiotics (100 U/ml penicillin and 1 mg/ml streptomycin) at 37 °C in a humidified atmosphere of 5% CO2. Both cell lines used in this study were authenticated by short tandem repeat profiling and tested for the absence of mycoplasma contamination (MycoAlert, Lonza Rockland, ME, USA).

Construction of reporter plasmids and transfections. The full-length PKN1 cDNA containing the rs34390238[C] allele or rs34390238[A] allele was commercially synthesized (Geneviz, Suzhou, China) and subcloned into the BamHI and XhoI sites of the pcDNA3.1(−) vector (Invitrogen, USA) to construct a vector expression of the human PKN1 gene (Gene ID: 5585). The resulting vectors were named PKN1[C] and PKN1[A]. The siRNA oligonucleotides targeting PKN1 and non-targeting siRNA (Supplementary Table 7) were purchased from Ribobio (Guangzhou, China). The PKN1 inhibitors Lenti-puro (sh42707) and sh42708 (HY-13866A) were purchased from Abcam (Cambridge, UK) and MedChem Express (NJ, USA), respectively.

PANC-1 and BxPC-3 cells were seeded in 6-well plates at a density of 1 × 10^6 cells per well, and 3 μg of plasmids were cotransfected into cells using Lipofectamine 3000 (Invitrogen, USA). Total protein was extracted with radioimmunoprecipitation assay (RIPA) buffer (moderate strength) and supersonic decomposition. Transfection efficiency was detected by quantitative reverse transcriptase–PCR (qRT-PCR; Supplementary Fig. 8). Total RNA was extracted from cells with TRIzol (Invitrogen, USA), according to the manufacturer’s instructions. First-strand cDNAs were synthesized using the PrimeScript 1st Strand cDNA Synthesis Kit (Takara, Japan). Relative RNA expression levels were determined by qRT-PCR using the SYBR Green method on an ABI Prism cDNA Synthesis Kit (Takara, Japan). Relative RNA expression levels were calculated using the 2^-ΔΔCT method. The data that support the findings of this study have been deposited in the European Genome-phenome Archive (EGA) with accession number EGAS00001003040.

Analysis of cell proliferation. Cells were seeded in 96-well flat-bottomed plates, and each well contained 2000 cells in 10 μl of medium. After a certain time in culture, cell viability was measured using CCK-8 assays (Dojindo Laboratories, Tokyo, Japan). Each experiment included six replicates and was repeated three times.

Statistical analysis. For functional analyses, the results are presented as means ± s.e.m. Mean values from two groups were compared using Student’s t test. The values of 

\[
\text{A}_{1} - \text{A}_{0} = \text{A}_{1} - \text{A}_{0} = \frac{\text{A}_{1} - \text{A}_{0}}{\text{A}_{0}} \times 100\%
\]

Statistical analysis for the functional data were performed using the R software (3.3.0).

Data availability. The data that support the findings of this study have been deposited in the European Genome-phenome Archive (EGA) with accession number EGAS00001003040.

Received: 16 January 2018 Accepted: 30 July 2018
Published online: 11 September 2018

References
1. Wolfgang, C. L. et al. Recent progress in pancreatic cancer. CA Cancer J. Clin. 63, 318–348 (2013).
2. Li, D., Xie, K., Wolff, R. & Abbazzezze, J. L. Pancreatic cancer. Lancet 363, 1049–1057 (2004).
3. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. CA Cancer J. Clin. 67, 7–30 (2017).
4. Chen, W. et al. Cancer statistics in China, 2015. CA Cancer J. Clin. 65, 113–132 (2016).
5. Hidalgo, M. Pancreatic cancer. N. Engl. J. Med. 362, 1605–1617 (2010).
6. Zou, L. et al. Non-linear dose-response relationship between cigarette smoking and pancreatic cancer risk: evidence from a meta-analysis of 42 observational studies. Eur. J. Cancer 50, 193–203 (2014).
7. Elhai, G. et al. Diabetes mellitus and obesity as risk factors for pancreatic cancer. J. Acad. Nutr. Diet 118, 555–567 (2017).
8. Amundadottir, L. T. Pancreatic cancer genetics. Int. J. Biol. Sci. 12, 314–325 (2016).
9. Childs, E. J. et al. Common variation at 2p11.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat. Genet. 47, 911–916 (2015).
10. Wolpin, B. M. et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. 46, 994–1000 (2014).
11. Wu, C. et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene–environment interactions. Nat. Genet. 44, 1097–1102 (2012).
12. Petersen, G. M. et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat. Genet. 42, 224–228 (2010).
13. Low, S. K. et al. Genome-wide association study of pancreatic cancer in Japanese population. PLoS ONE 9, e11824 (2010).
14. De Streeger, B. et al. Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. Pancreatology 10, 194–200 (2010).
15. Amundadottir, L. et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat. Genet. 41, 986–990 (2009).
16. Park, J. H. et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat. Genet. 42, 570–575 (2010).
17. Dai, J. et al. Estimation of heritability for nine common cancers using data from genome-wide association studies in Chinese population. Int. J. Cancer 129, 329–336 (2011).
18. Manolio, T. A. et al. Finding the missing heritability of complex diseases. Nature 461, 747–753 (2009).
19. Huyghe, J. R. et al. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat. Genet. 45, 197–201 (2013).
20. Kodzinska, J. et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat. Genet. 46, 352–356 (2014).
21. Chang, J. et al. Exome-wide analyses identify low-frequency variant in CYP26B1 and additional coding variants associated with esophageal squamous cell carcinoma. Nat. Genet. 50, 384–393 (2018).
22. Hornbeck, P. V. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 40, D261–D270 (2012).
23. Kohler, J. et al. Lestaurtinib inhibits histone phosphorylation and androgen-dependence gene expression in prostate cancer cells. PLoS ONE 7, e34973 (2012).
24. Singh, N. K. et al. Protein kinase N1 is a novel substrate of NFATc1-mediated cyclin D1-CDK6 activity and modulates vascular smooth muscle cell division and migration leading to inward blood vessel wall remodeling. J. Biol. Chem. 287, 36291–36304 (2012).
25. Kim, J. Y. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 40, D261–D270 (2012).
26. Safaei, M. et al. CD97 is a multifunctional leukocyte receptor with distinct roles in human cancers (Review). J. Oncol. 43, 1343–1350 (2015).
27. Torbett, N. E., Casassasamina, A. & Parker, P. J. Hyperosmotic-induced protein kinase kinase N1 activation in a vesicular compartment is dependent upon Rac1 and 3-phosphoinositide-dependent kinase 1. J. Biol. Chem. 278, 32344–32351 (2003).
28. Wu, C. Y. et al. P38K regulation of Rac1 is required for KRAS-induced pancreatic tumorigenesis in mice. Gastroenterology 147, 1405.e7–1416.e7 (2014).
29. Collazo, A. et al. Site recognition and substrate screens for PKN family proteins. Biochem. J. 438, 535–543 (2011).
30. Zhao, J. & Guan, J. L. Signal transduction by focal adhesion kinase in cancer. Cancer Metastasis Rev. 28, 35–49 (2009).
31. Sulzmaier, F. J., Jean, C. & Schlaepfer, D. D. FAK in cancer: mechanistic insights and clinical applications. Nat. Rev. Cancer 14, 598–610 (2014).
32. Lane, D., Goncharenko-Khaider, N., Rancourt, C. & Piche, A. Ovarian cancer ascites protects from TRAIL-induced cell death through alphavbeta5 integrin-mediated focal adhesion kinase and Akt activation. Oncogene 29, 3519–3531 (2010).
33. Daval, M., Gurio, T., Costos, S., Huang, C. J. & Butler, P. C. Cyclin-dependent kinase 5 promotes pancreatic β-cell survival via Fak-Akt signaling pathways. Diabetes 60, 1186–1197 (2011).
34. Emi, M. et al. Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. Cancer Res. 52, 5368–5372 (1992).
35. Weidner, I. et al. Allieic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. Cancer Res. 59, 1973–1979 (1999).
Acknowledgements
This work was supported by grants from the National Key Research and Development Plan Program [2016YFC1302702 to X.M., 2016YFC1302703 to C.W. and 2016YFC1302703 to R.Z.]; The Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences [2016-IDM-3-001 to L.W.]; The National Natural Science Foundation of China [81672526 to X.M.]; National High-Tech Research and Development Program of China [2014AA020609 to X.M.]; Specialized Research Fund for the Doctoral Program of Higher Education [20130142110017 to X.M.] and National Program for Support of Top-notch Young Professionals for Xiaoping Miao.

Author contributions
X.M. and C.W. were the overall principal investigators in this study who conceived the study, obtained financial support, were responsible for the study design and supervised the entire study. J.C. performed statistical analyses, interpreted the results and drafted the initial manuscript. J.T., Y. Zhu, R.Z., K.Z., J. Li, J.K., J. Lou, W.C., B.Z., N.S., Y. Zhang, Y.G., Y.Y., D.Z. and X.P. performed laboratory experiments. Q.H. performed the iTRAQ-based comparative proteomics screen. M.Y. was responsible for patient recruitment from Shandong province and sample preparation. Z.Z. and X.Z. were responsible for patient recruitment from Hebei province and sample preparation. K.H., L.W. and D.L. reviewed the manuscript. All authors approved the final report for publication.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06136-x.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06136-x

ARTICLE

Jiang Chang1, Jianbo Tian1, Ying Zhu1, Rong Zhong1, Kan Zhai2,3, Jiaoyuan Li1, Juntao Ke1, QiangQiang Han4, Jiao Lou1, Wei Chen1, Beibei Zhu1, Na Shen1, Yi Zhang1, Yajie Gong1, Yang Yang1, Danyi Zou1, Xiating Peng1, Zhi Zhang5, Xuemei Zhang6, Kun Huang7, Ming Yang8, Li Wang9, Chen Wu2, Dongxin Lin2 & Xiaoping Miao1

1Department of Epidemiology and Biostatistics, Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Sciences and Technology, 430030 Wuhan, China. 2Department of Etiology and Carcinogenesis, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 100021 Beijing, China. 3Medical Research Center, Beijing Chao-Yang Hospital, Capital Medical University, 100020 Beijing, China. 4Wuhan GeneCreate Biological Engineering Co., Ltd, 430075 Wuhan, China. 5Department of Chemotherapy and Radiotherapy, Tangshan Gongren Hospital, 063210 Tangshan, China. 6Department of Molecular Genetics, College of Life Science, North China University of Science and Technology, 063210 Tangshan, China. 7Tongji School of Pharmacy, Huazhong University of Science and Technology, 430030 Wuhan, China. 8Shandong Provincial Key Laboratory of Radiation Oncology, Cancer Research Center, Shandong Cancer Hospital affiliated to Shandong University, Shandong Academy of Medical Sciences, 250117 Jinan, China. 9Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and School of Basic Medicine, Peking Union Medical College, 100730 Beijing, China. These authors contributed equally: Jiang Chang, Jianbo Tian, Ying Zhu, Rong Zhong.