

ARTICLE

DOI: 10.1038/s41467-018-06136-x

OPEN

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

Jiang Chang  et al.<sup>#</sup>

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,  $P = 5.35 \times 10^{-10}$ ), rs2242241 in *DOK2* (OR = 1.85, 95% CI: 1.50–2.27,  $P = 4.34 \times 10^{-9}$ ), and rs183117027 in *APOB* (OR = 2.34, 95% CI: 1.72–3.16,  $P = 4.21 \times 10^{-8}$ ). 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.

Correspondence and requests for materials should be addressed to C.W. (email: [chenwu@cicams.ac.cn](mailto:chenwu@cicams.ac.cn)) or to D.L. (email: [lindx@cicams.ac.cn](mailto:lindx@cicams.ac.cn)) or to X.M. (email: [miaoxp@hust.edu.cn](mailto:miaoxp@hust.edu.cn)). <sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.

**P**ancreatic ductal adenocarcinoma (PDAC) is one of the most lethal human cancers, with a 5-year overall survival rate of only ~5%<sup>1,2</sup>. 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<sup>3,4</sup>. Cigarette smoking, type 2 diabetes, obesity and several hereditary cancer syndromes represent major risk factors for PDAC<sup>2,5–7</sup>. Based on accumulating evidence, germline variants also play an important role in the development of this disease<sup>8</sup>.

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<sup>9–15</sup>. 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<sup>16,17</sup>. 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<sup>16,18</sup>. 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<sup>19–21</sup>. 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 samples including 2142 cases and 4697 controls. We identify 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

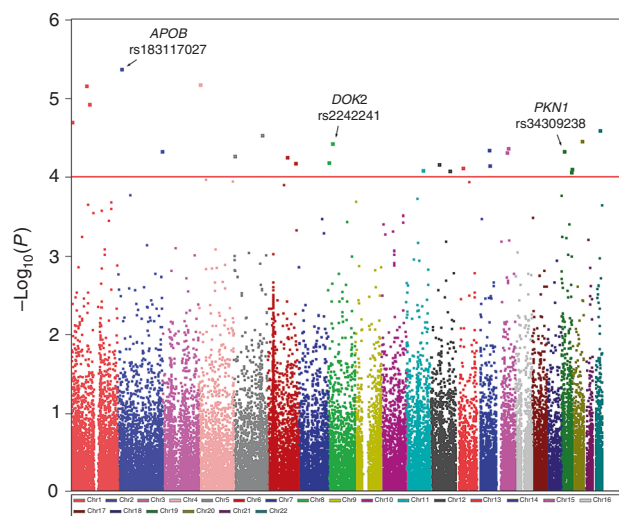
### Three low-frequency missense SNPs were identified for PDAC.

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,  $P = 4.34 \times 10^{-9}$ ) and 2.34 (95% CI 1.72–3.16,  $P = 4.21 \times 10^{-8}$ ) 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

**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
						Cases	Controls		
19p13.12	rs34309238	PKN1	C>A	p.Leu555Ile	Discovery	0.033	0.017	1.96 (1.44-2.67)	$1.79 \times 10^{-5}$
					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)	$5.35 \times 10^{-10}$
					Discovery	0.030	0.013	2.47 (1.75-3.49)	$2.94 \times 10^{-7}$
8p21.3	rs2242241	DOK2	T>G	p.Ser394Ala	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)	$4.34 \times 10^{-9}$
2p24.1	rs183117027	APOB	G>A	p.Val400Glle	Discovery	0.016	0.006	2.95 (1.83-4.74)	$8.05 \times 10^{-6}$
					Replication I	0.011	0.005	2.20 (1.21-3.98)	0.0093
					Replication II	0.011	0.006	2.01 (1.17-3.48)	0.0120
					Combined	0.012	0.006	2.34 (1.72-3.16)	$4.21 \times 10^{-8}$

P values are two sided and were calculated by an additive model in logistic regression analysis adjusted for sex and age  
Chr chromosomal region, MAF minor allele frequency, OR odds ratio, CI confidence interval, Allele Reference allele > Effect allele

test and a sequence kernel association test (SKAT). A total of 24,636 variants enriched in 9647 genes were analysed. Five genes (*DOK2*, *PKN1*, *TAS1R3*, *MLPH* and *CHPT1*) passed our significance threshold in the discovery stage with  $P < 1 \times 10^{-4}$  in either burden test or SKAT (Supplementary Table 6). However, all of the top variants of these genes were identified and replicated in the single variant analysis. We did not find novel significant susceptible variants enriched in genes.

#### 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 database<sup>22</sup> 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 OncoPrint 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>Ile) 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 for PANC-1 cells after transfection with a pcDNA3.1 plasmid containing *PKN1* rs34309238[C], rs34309238[A] 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<sup>Tyr397</sup>), PI3K regulatory subunit alpha (PI3KR1<sup>Ser83</sup>), 3-phosphoinositide-dependent protein kinase 1 (PDK1<sup>Ser241</sup>) and serine/threonine kinase 1 (AKT<sup>Ser473</sup>) upon transfection with *PKN1*[A] in the iTRAQ-based screen (Fig. 2b). The phosphorylation level of FAK<sup>Tyr397</sup> and AKT<sup>Ser473</sup> were further successfully validated using western blot (Fig. 2c). Meanwhile, knockdown of *PKN1* by

small interfering RNAs (siRNAs) reduced the levels of phosphorylated FAK<sup>Tyr397</sup> and AKT<sup>Ser473</sup> (Fig. 2c).

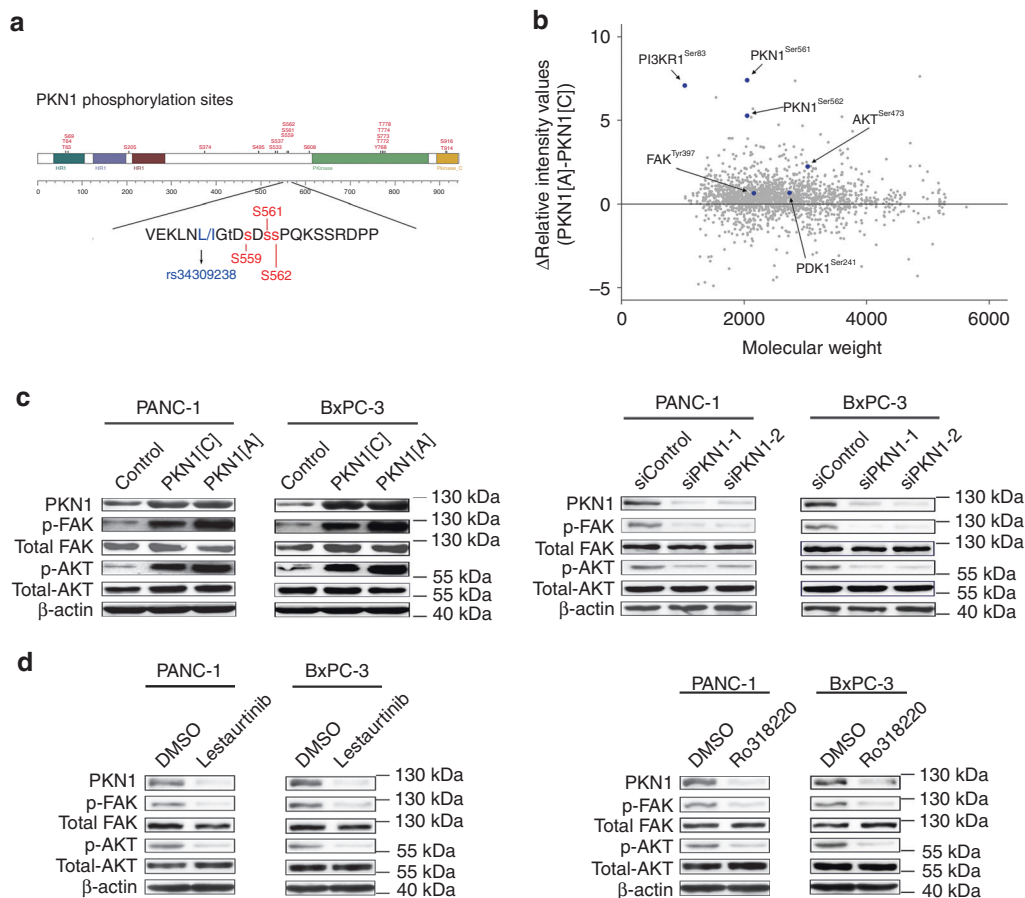
#### The rs34309238-A promotes PDAC cells' proliferation.

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)<sup>23-25</sup> 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 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<sup>26</sup>. 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)<sup>27</sup>, which is required for KRAS-induced pancreatic tumorigenesis in mice<sup>28</sup>. 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



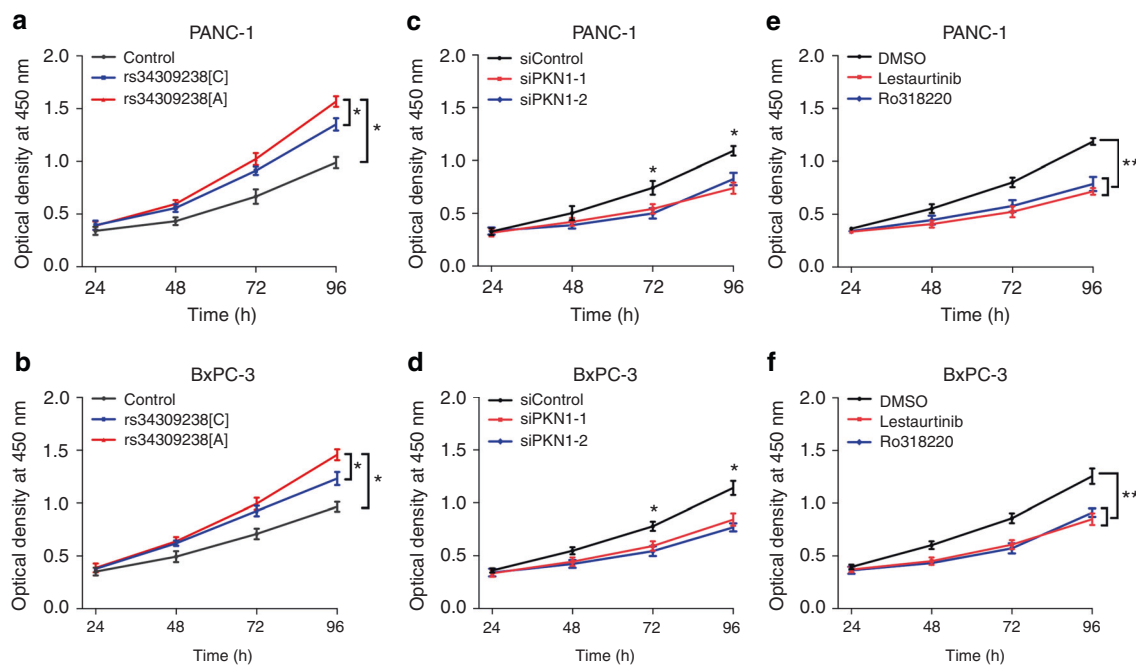
**Fig. 2** *PKN1* rs34309238 variant influences pancreatic cancer risk by altering the level of phosphorylated *PKN1* and thus affecting the FAK/PI3K/AKT signalling pathway. **a** Protein modification sites of *PKN1*. Annotations were obtained from the PhosphoSitePlus database. **b** Result of the iTRAQ-based comparative proteomics screen. PANC-1 cells were seeded in six-well plates after transfection with *PKN1*[A], *PKN1*[C] or control vector. The raw intensity values of cells transfected with *PKN1*[A] or *PKN1*[C] were divided by the intensity values of cells transfected with control vector to obtain the relative intensity values. The y axis shows the relative intensity values of cells' transfection with *PKN1*[A] minus the relative intensity values of cells' transfection with *PKN1*[C]. The x axis shows the molecular weight of detected peptides. The proteomics screen experiment was repeated independently for two times with similar results. **c** Levels of phosphorylated FAK and AKT were affected by the *PKN1* rs34309238 variant. Cells were seeded in six-well plates after transfection with *PKN1*[A], *PKN1*[C] or control vector (left) and *PKN1*-targeting siRNAs or control siRNA (right). **d** Levels of phosphorylated FAK and AKT were reduced by the *PKN1* inhibitors. Cells were seeded in six-well plates after transfection with *PKN1* inhibitors Lestaurtinib and Ro318220 or DMSO as control. For **c**, **d**, the western blot experiment was repeated independently for three times with similar results

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*<sup>29</sup>, and its phosphorylation enhances cancer development by activating the PI3K/AKT signalling pathway<sup>30–33</sup>. 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<sup>34–36</sup>. 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<sup>37,38</sup>. 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*<sup>39</sup>.

The *APOB* gene is located in the 2p24.1 region, which contains a Barrett's oesophagus susceptibility gene, the growth differentiation factor 7<sup>40</sup>. 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<sup>41,42</sup>. Germline mutations in *APOB* also potentially cause diseases associated with lipid metabolic disorders, such as hypobetalipoproteinaemia and hypercholesterolaemia<sup>43–45</sup>. Common polymorphisms in the *APOB* gene are associated with low-density lipoprotein cholesterol metabolism or the risk of coronary heart disease<sup>46,47</sup>. 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<sup>48</sup>.



**Fig. 3** *PKN1* rs34309238 variant influences pancreatic cancer cells' proliferation. **a, b** Overexpression of *PKN1*[A] substantially enhanced the rate of cell proliferation in PANC-1 (**a**) and BxPC-3 (**b**) cells. Cells were seeded in 96-well plates after transfection with *PKN1*[A], *PKN1*[C] or control vector. **c, d** Knockdown of *PKN1* significantly reduced the proliferation of PANC-1 (**c**) and BxPC-3 (**d**) cells. Cells were seeded in 96-well plates after transfection with *PKN1*-targeting siRNAs or control siRNA (siControl). **e, f** *PKN1* inhibitors significantly reduced the rate of cell proliferation in PANC-1 (**e**) and BxPC-3 (**f**) cells. Cells were seeded in 96-well plates after transfection with *PKN1* inhibitors Lestaurtinib and Ro318220 or DMSO as control. For **a–f**, cell numbers were determined every 24 h for 96 h using CCK-8 assays and the results present means  $\pm$  s.e.m. from three independent experiments and each had six replications. \* $P < 0.05$ , \*\* $P < 0.01$ , compared with the control by two-sided unpaired Student's *t* test

As abnormal lipid metabolism promotes pancreatic tumorigenesis<sup>49,50</sup> and obesity is an important risk factor for the PDAC<sup>7</sup>, rs183117027 variant may affect the development of PDAC by disturbing the lipid metabolic function of *APOB*.

Recently, an exome-wide association study based on exome and genome sequencing data identified that rare variants enriched in *BRCA2* was significantly associated with PDAC risk in the European population<sup>51</sup>. However, all of the identified *BRCA2* rare risk variants in this European study did not have frequency in the Chinese population. We also did not find significant *BRCA2*-susceptible variants using either single variant analysis or gene-based analysis (all  $P > 0.05$ ). These results suggested that the rare susceptibility variants were usually population-specific<sup>48</sup>.

In summary, our study has identified three low-frequency coding variants that are significantly associated with PDAC susceptibility. In further functional analyses, we found for the first time that the rs34309238 variant increases the risk of PDAC by enhancing the level of phosphorylated *PKN1* and thus activating the FAK/*PI3K*/*AKT* signalling pathway. These findings highlight the significance of rare coding variants in the development of PDAC and may be useful for the prevention and treatment of this disease in future.

## 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 recruited 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<sup>11,52–55</sup>. 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<sup>56</sup>.

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  $P$  value <0.0001 in a Hardy–Weinberg equilibrium test among the control subjects (320 variants). Five PDAC cases and three controls were excluded because they (1) had an overall genotyping rate of <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 of 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<sup>57</sup> based on 4431 autosomal informative ancestry markers included in the exome chip<sup>56</sup>. 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 ( $PI\_HET > 0.25$ ) 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 Haploview<sup>58</sup>, and Q–Q plot was generated with the R software (version 3.3.0).

**Genotype imputation.** We phased the haplotypes with SHAPEIT<sup>59</sup> and performed imputations with the IMPUTE2<sup>60,61</sup> 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 measure (Is)  $< 0.40$  (output from IMPUTE2 info file) were excluded from subsequent analyses. Regional plots were created using LocusZoom<sup>62</sup> 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 0.1 mg/ml streptomycin) at 37 °C in a humidified atmosphere of 5% CO<sub>2</sub>. Both cell lines used in this study were authenticated by short tandem repeat profiling and tested for the absence of mycoplasma contamination (MycAlert, Lonza Rockland, ME, USA).

**Construction of reporter plasmids and transfections.** The full-length *PKN1* cDNA containing the rs34309238[C] allele or rs34309238[A] allele was commercially synthesized (Genewiz, 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 Lestaurtinib (ab142107) and Ro318220 (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  $1 \times 10^6$  cells per well, and 3  $\mu$ g 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 7900 sequence detection system (Applied Biosystems, USA) with Power SYBR<sup>TM</sup> Green PCR Master Mix (Applied Biosystems, USA), 50 ng of cDNA templates and 0.5  $\mu$ M gene-specific primers in a 10- $\mu$ l reaction mixture. Reactions were performed with an initial 30-s denaturation step at 95 °C, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s. Independent experiments were performed in triplicate. GAPDH was employed as an internal control. The sequences of all specific primers used in qPCR are upon request.

**Antibodies and western blot.** The PKN1 antibody (ab195264) was purchased from Abcam (Cambridge, UK). FAK (#3285), phospho-FAK (Tyr397, #8556), Akt (#4685) and phospho-Akt (Ser473, #4060) antibodies were purchased from Cell Signaling Technology (Beverly, MA, USA). The beta-actin antibody (60008-1-Ig) was purchased from Proteintech (Wuhan, China). All antibodies were validated by the company using western blots of human cell lines. The working dilution for each antibody is listed in Supplementary Table 8. The western blot experiments were repeated independently for three times with similar results. The original full western blot plots are shown in Supplementary Fig. 9.

**Phosphopeptide enrichment and labelling.** PANC-1 cells were seeded in 15-cm diameter cell plates at a density of  $1 \times 10^8$  cells per well and transfected with

pcDNA3.1 empty vector (Control), PKN1[C] and PKN1[A], respectively. After 48 h, the cells were harvested and lysed using RIPA buffer containing protease (Complete Mini, Roche, IN, USA) and phosphatase inhibitors (PhosSTOP, Roche, IN, USA). After 10-min sonication in ice water bath, the lysed samples were centrifuged at 15,000  $\times g$  for 30 min at 4 °C to remove cell debris. The proteins were precipitated by adding four volumes of cold acetone. The resulting protein pellets were resuspended in 8 M urea/100 mM TEAB (pH 8.0). The proteins were then reduced with 10 mM dithiothreitol at 56 °C, alkylated with 50 mM iodoacetamide in the dark and diluted 4 times with 100 mM TEAB (pH 8.0) prior to digestion overnight with trypsin (1:50 w/w). The resulting peptides were desalted with Sep-Pak C18 column (Waters, MA, USA).

iTRAQ-based quantification was used for proteomics screen. The three peptide samples were first subjected to phosphopeptides enrichment using an Immobilized Metal Affinity Chromatography method<sup>63</sup>. Then these samples were labelled with an eight-plex iTRAQ Kit. The labelled samples were mixed, desalted, vacuum dried and fractionated into 12 fractions using a high-pH reversed-phase chromatography method<sup>64</sup> on the Ultimate 3000 HPLC system (Thermo Scientific, CA, USA). All the samples were vacuum dried and stored at  $-80$  °C until the liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis.

**The LC-MS/MS analysis.** LC-MS/MS analysis was performed on a TripleTOF 5600+ system (SCIEX, MA, USA). Each sample was first loaded onto a trap column and then chromatographed using a 90-min gradient on an analytical column (75  $\mu$ m  $\times$  15 cm, Magic C18 AQ 3- $\mu$ m 120- $\text{\AA}$ ). The eluted peptides were sprayed into mass spectrometer and scanned. MS data were acquired in DDA mode. MS1 spectra were collected in the range 350–1250 *m/z* for 250 ms. The 30 most intense precursors with charge state 2–5 were selected for fragmentation, and MS2 spectra were collected in the range 50–2000 *m/z* for 100 ms; precursor ions were excluded from reselection for 15 s.

Raw data were processed with ProteinPilot version 4.5 (SCIEX, MA, USA). MS and MS/MS spectra were searched against SwissProt human database using Paragon algorithm<sup>65</sup>. The database search was performed with the following parameters: iTRAQ 8plex (peptide labelled) for Sample type according to the experiments; Iodoacetamide for Alkylation; Trypsin for Digestion; Phosphorylation emphasis for Special factors; Thorough ID for Search effort. ProteinPilot calculates a percentage of confidence that reflects the probability that the hit is a false positive. Only high-quality peptide assignments with confidence  $>95\%$  were considered in this study.

**Analysis of cell proliferation.** Cells were seeded in 96-well flat-bottomed plates, and each well contained 2000 cells in 100  $\mu$ l of cell suspension. 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  $\pm$  s.e.m. Mean values from two groups were compared using Student's *t* test. The variance similar between the groups were tested using one-way analysis of variance. All statistical analyses of 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](https://ega-archive.org/studies/EGAS00001003040).

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

## References

- Wolfgang, C. L. et al. Recent progress in pancreatic cancer. *CA Cancer J. Clin.* **63**, 318–348 (2013).
- Li, D., Xie, K., Wolff, R. & Abbruzzese, J. L. Pancreatic cancer. *Lancet* **363**, 1049–1057 (2004).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA Cancer J. Clin.* **67**, 7–30 (2017).
- Chen, W. et al. Cancer statistics in China, 2015. *CA Cancer J. Clin.* **66**, 115–132 (2016).
- Hidalgo, M. Pancreatic cancer. *N. Engl. J. Med.* **362**, 1605–1617 (2010).
- 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).
- Eibl, G. et al. Diabetes mellitus and obesity as risk factors for pancreatic cancer. *J. Acad. Nutr. Diet* **118**, 555–567 (2017).
- Amundadottir, L. T. Pancreatic cancer genetics. *Int. J. Biol. Sci.* **12**, 314–325 (2016).

9. Childs, E. J. et al. Common variation at 2p13.3, 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**, 1090–1097 (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* **5**, e11824 (2010).
14. Diergaard, B. et al. Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. *Pancreatol.* **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* **140**, 329–336 (2017).
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. Kozlitina, 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**, 338–343 (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-dependent 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. A role for WDR5 in integrating threonine 11 phosphorylation to lysine 4 methylation on histone H3 during androgen signaling and in prostate cancer. *Mol. Cell* **54**, 613–625 (2014).
26. Safaee, M. et al. CD97 is a multifunctional leukocyte receptor with distinct roles in human cancers (Review). *J. Oncol.* **43**, 1343–1350 (2013).
27. Torbett, N. E., Casamassima, A. & Parker, P. J. Hyperosmotic-induced protein kinase N 1 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. PI3K regulation of RAC1 is required for KRAS-induced pancreatic tumorigenesis in mice. *Gastroenterology* **147**, 1405.e7–1416.e7 (2014).
29. Collazos, 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 findings 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., Gurlo, T., Costes, S., Huang, C. J. & Butler, P. C. Cyclin-dependent kinase 5 promotes pancreatic beta-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. Wistuba, I. I. et al. Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. *Cancer Res.* **59**, 1973–1979 (1999).
36. Kim, T. M. et al. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res.* **23**, 217–227 (2013).
37. Berger, A. H. et al. Identification of DOK genes as lung tumor suppressors. *Nat. Genet.* **42**, 216–223 (2010).
38. Coppin, E. et al. Mutational analysis of the DOK2 haploinsufficient tumor suppressor gene in chronic myelomonocytic leukemia (CMML). *Leukemia* **29**, 500–502 (2015).
39. Gu, J. et al. GFRa2 prompts cell growth and chemoresistance through down-regulating tumor suppressor gene PTEN via Mir-17-5p in pancreatic cancer. *Cancer Lett.* **380**, 434–441 (2016).
40. Palles, C. et al. Polymorphisms near TBX5 and GDF7 are associated with increased risk for Barrett's esophagus. *Gastroenterology* **148**, 367–378 (2015).
41. Cancer Genome Atlas Research Network et al. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341. e23 (2017).
42. Fernandez-Banet, J. et al. Decoding complex patterns of genomic rearrangement in hepatocellular carcinoma. *Genomics* **103**, 189–203 (2014).
43. Zhong, S. et al. Nonsynonymous mutations within APOB in human familial hypobetalipoproteinemia: evidence for feedback inhibition of lipogenesis and postendoplasmic reticulum degradation of apolipoprotein B. *J. Biol. Chem.* **285**, 6453–6464 (2010).
44. Welty, F. K. Hypobetalipoproteinemia and abetalipoproteinemia. *Curr. Opin. Lipidol.* **25**, 161–168 (2014).
45. Motazacker, M. M. et al. Advances in genetics show the need for extending screening strategies for autosomal dominant hypercholesterolaemia. *Eur. Heart J.* **33**, 1360–1366 (2012).
46. Shahid, S. U. et al. Effect of SORT1, APOB and APOE polymorphisms on LDL-C and coronary heart disease in Pakistani subjects and their comparison with Northwick Park Heart Study II. *Lipids Health Dis.* **15**, 83 (2016).
47. Waterworth, D. M. et al. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **30**, 2264–2276 (2010).
48. Lu, X. et al. Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat. Genet.* **49**, 1722–1730 (2017).
49. Guillaumond, F. et al. Cholesterol uptake disruption, in association with chemotherapy, is a promising combined metabolic therapy for pancreatic adenocarcinoma. *Proc. Natl. Acad. Sci. USA* **112**, 2473–2478 (2015).
50. Ying, H. et al. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev.* **30**, 355–385 (2016).
51. Grant, R. C. et al. Exome-wide association study of pancreatic cancer risk. *Gastroenterology* **154**, 719–722 (2018).
52. Zhu, B. et al. Genetic variants in the SWI/SNF complex and smoking collaborate to modify the risk of pancreatic cancer in a Chinese population. *Mol. Carcinog.* **54**, 761–768 (2014).
53. Zhu, B. et al. A single nucleotide polymorphism in the 3'-UTR of STAT3 regulates its expression and reduces risk of pancreatic cancer in a Chinese population. *Oncotarget* **7**, 62305–62311 (2016).
54. Gong, J. et al. A functional polymorphism in lnc-LAMC2-1:1 confers risk of colorectal cancer by affecting miRNA binding. *Carcinogenesis* **37**, 443–451 (2016).
55. Li, J. et al. A low-frequency variant in SMAD7 modulates TGF-beta signaling and confers risk for colorectal cancer in Chinese population. *Mol. Carcinog.* **56**, 1798–1807 (2017).
56. Guo, Y. et al. Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* **9**, 2643–2662 (2014).
57. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
58. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
59. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
60. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
61. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).
62. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
63. Albuquerque, C. P. et al. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol. Cell. Proteomics* **7**, 1389–1396 (2008).
64. Bath, T. S., Francavilla, C. & Olsen, J. V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **13**, 6176–6186 (2014).
65. Shilov, I. V. et al. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **6**, 1638–1655 (2007).

## Acknowledgements

This work was supported by grants from the National Key Research and Development Plan Program [2016YFC1302702 to X.M., 2016YFC1302701 to C.W. and 2016YFC1302703 to R.Z.]; The Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences [2016-I2M-3-001 to L.W.]; The National Natural Science Foundation of China [81673256 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

Jiang Chang<sup>1</sup>, Jianbo Tian<sup>1</sup>, Ying Zhu<sup>1</sup>, Rong Zhong<sup>1</sup>, Kan Zhai<sup>2,3</sup>, Jiaoyuan Li<sup>1</sup>, Juntao Ke<sup>1</sup>, QiangQiang Han<sup>4</sup>, Jiao Lou<sup>1</sup>, Wei Chen<sup>1</sup>, Beibei Zhu<sup>1</sup>, Na Shen<sup>1</sup>, Yi Zhang<sup>1</sup>, Yajie Gong<sup>1</sup>, Yang Yang<sup>1</sup>, Danyi Zou<sup>1</sup>, Xiating Peng<sup>1</sup>, Zhi Zhang<sup>5</sup>, Xuemei Zhang<sup>6</sup>, Kun Huang<sup>7</sup>, Ming Yang<sup>8</sup>, Li Wang<sup>9</sup>, Chen Wu<sup>2</sup>, Dongxin Lin<sup>2</sup> & Xiaoping Miao<sup>1</sup>

<sup>1</sup>Department 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. <sup>2</sup>Department of Etiology and Carcinogenesis, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 100021 Beijing, China. <sup>3</sup>Medical Research Center, Beijing Chao-Yang Hospital, Capital Medical University, 100020 Beijing, China. <sup>4</sup>Wuhan GeneCreate Biological Engineering Co., Ltd, 430075 Wuhan, China. <sup>5</sup>Department of Chemotherapy and Radiotherapy, Tangshan Gongren Hospital, 063210 Tangshan, China. <sup>6</sup>Department of Molecular Genetics, College of Life Science, North China University of Science and Technology, 063210 Tangshan, China. <sup>7</sup>Tongji School of Pharmacy, Huazhong University of Science and Technology, 430030 Wuhan, China. <sup>8</sup>Shandong Provincial Key Laboratory of Radiation Oncology, Cancer Research Center, Shandong Cancer Hospital affiliated to Shandong University, Shandong Academy of Medical Sciences, 250117 Jinan, China. <sup>9</sup>Department 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.