

Systematic analysis on expression quantitative trait loci identifies a novel regulatory variant in ring finger and WD repeat domain 3 associated with prognosis of pancreatic cancer

Ying Zhu¹, Xiating Peng², Xiaoyang Wang², Pingting Ying², Haoxue Wang², Bin Li², Yue Li², Ming Zhang², Yimin Cai², Zequn Lu², Siyuan Niu², Nan Yang², Rong Zhong², Jianbo Tian¹, Jiang Chang², Xiaoping Miao¹

¹Department of Epidemiology and Biostatistics, School of Public Health, Wuhan University, Wuhan, Hubei 430072, China;

²Department of Epidemiology and Biostatistics, Key Laboratory for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Sciences and Technology, Wuhan, Hubei 430030, China.

Abstract

Background: Pancreatic adenocarcinoma (PAAD) is an extremely lethal malignancy. Identification of the functional genes and genetic variants related to PAAD prognosis is important and challenging. Previously identified prognostic genes from several expression profile analyses were inconsistent. The regulatory genetic variants that affect PAAD prognosis were largely unknown.

Methods: Firstly, a meta-analysis was performed with seven published datasets to systematically explore the candidate prognostic genes for PAAD. Next, to identify the regulatory variants for those candidate genes, expression quantitative trait loci analysis was implemented with PAAD data resources from The Cancer Genome Atlas. Then, a two-stage association study in a total of 893 PAAD patients was conducted to interrogate the regulatory variants and find the prognostic locus. Finally, a series of biochemical experiments and phenotype assays were carried out to demonstrate the biological function of variation and genes in PAAD progression process.

Results: A total of 128 genes were identified associated with the PAAD prognosis in the meta-analysis. Fourteen regulatory loci in 12 of the 128 genes were discovered, among which, only rs4887783, the functional variant in the promoter of Ring Finger and WD Repeat Domain 3 (*RFWD3*), presented significant association with PAAD prognosis in both stages of the population study. Dual-luciferase reporter and electrophoretic mobility shift assays demonstrated that rs4887783-G allele, which predicts the worse prognosis, enhanced the binding of transcript factor REST, thus elevating *RFWD3* expression. Further phenotypic assays revealed that excess expression of *RFWD3* promoted tumor cell migration without affecting their proliferation rate. *RFWD3* was highly expressed in PAAD and might orchestrate the genes in the DNA repair process.

Conclusions: *RFWD3* and its regulatory variant are novel genetic factors for PAAD prognosis.

Keywords: Pancreatic cancer; Survival; *RFWD3*; Genetic variation; Quantitative trait loci

Introduction

Pancreatic adenocarcinoma (PAAD) is a digestive malignancy with extremely poor prognosis and high fatality rate. Despite advances in medical technologies over the past decades, early detection and effective treatment are still lacking, resulting in a constantly rising of deaths attributed to PAAD worldwide.^[1] Besides clinical indicators such as the disease stage, genetic factors also have an important influence on the overall survival of PAAD patients.^[2,3]

For the purpose of discovering genetic variation associated with the survival of sporadic PAAD patients, a series of

genome-wide association studies (GWAS) have been performed.^[4-7] Notably, nearly nine tenths of the identified prognostic single nucleotide polymorphisms (SNPs) resides in the non-coding regions, which highlights the important role of regulatory genetic variations in PAAD progression.

Expression quantitative trait loci (eQTL) analysis is a method contributing to identifying regulatory SNPs as it links variations of gene expression to genotypes.^[8] By applying this method, many diseases associated SNPs were discovered.^[9,10] In general, the eQTLs are more likely prognostic when the corresponding genes are associated

Access this article online	
Quick Response Code: 	Website: www.cmj.org
	DOI: 10.1097/CM9.0000000000002180

Ying Zhu and Xiating Peng contributed equally to this work.

Correspondence to: Prof. Xiaoping Miao, School of Public Health, Wuhan University, Wuhan, Hubei 430072, China
E-Mail: xpmiao@whu.edu.cn

Copyright © 2022 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2022;135(11)

Received: 23-10-2021; Online: 13-07-2022 Edited by: Jing Ni

with PAAD survival. Therefore, it is probably a beneficial way to explore the prognostic genes before identifying the regulatory SNPs. There have been several studies measuring the gene expression profile that aims to comprehensively discover genes associated with PAAD prognosis or overall survival.^[11,12] However, the discoveries were not entirely consistent due to the limited sample sizes and the heterogeneity in studies design. To address these issues, meta-analysis of the existing results from the public databases is a welcome method.

In the current study, assuming that genetic variations can affect the PAAD survival by regulating the expression of prognostic genes, we firstly performed a meta-analysis to systematically discover the candidate genes of which the expression is associated with survival. Next, we targeted the SNPs regulating those genes through a tissue-specific eQTL analysis. Finally, we carried out a two-stage population study followed by a series of biological experiments to find and validate the PAAD survival associated functional SNP. What we found might contribute, to some extent, not only to identification of a prognostic biomarker but also to uncovering the mechanism of PAAD progression and finding the effective therapeutic targets.

Methods

Ethical approval

Informed consent was obtained from all participants at recruitment, and this study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology, No. [2016] IEC (S085).

Meta-analysis to identify survival associated genes

First of all, we systematically collected the gene expression data and survival information from three large-scale public databases: The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC) and Gene Expression Omnibus (GEO). Specifically, we downloaded the gene expression data and basic clinical information of all PAAD patients in TCGA and ICGC. As for GEO, we searched for the datasets containing the above information with the key words of “(survival OR prognosis OR progression) AND gene expression AND pancreatic cancer.” All the downloaded data were pre-processed as follows: (1) remove the duplicate samples, (2) discard gene expression data from normal tissue and cancer cell lines (consider only tumor tissues), (3) remove samples with incomplete survival information, (4) exclude the genes with average expression level of less than 1, (5) exclude the genes unless included in all datasets, (6) choose the transcript with the highest expression level to represent the expression level of the gene.

Next, the expression levels of each gene were divided into high and low group according to the median. The univariate Cox regression was implemented to determine the association of each gene expression with overall survival in each cohort. The hazard ratio (HR) and standard error (SE) for each gene in each cohort were acquired for further meta-analysis.

Finally, we applied a meta-analysis to integrate the gene survival analysis results of the seven datasets with the Metafor package in R (3.3.1) software. The random effect model was used to calculate the combined *P* value, HR and 95% confidence interval (CI).

Selection of SNPs

In order to systematically discover the regulatory SNPs, our team had utilized the genotypes and expression data from TCGA to investigate the eQTLs in 33 cancer types including PAAD.^[13] The Matrix eQTL methods in linear regression model were utilized and the SNPs with false discovery rates (FDR) <0.05 were regarded as eQTLs. We retrieved the PAAD-specific eQTLs that correlated to the prognostic genes suggested by our meta-analysis, and selected the tag SNPs from them with the Haploview software.

Study subjects of the two-stage survival analysis

We conducted a two-stage survival analysis on PAAD patients. In the discovery stage, 341 patients were recruited from Cancer Hospital, Chinese Academy of Medical Sciences in Beijing. As for the replication stage, 552 patients from multiple hospitals in Wuhan were included. The diagnosis of PAAD were pathologically confirmed. Demographic and clinical data including sex, age, and tumor stage were obtained from medical records. Survival time was measured from the date of surgery to last follow-up or death, and the follow-up information was obtained through telephone calls.

Genotyping and survival analysis

Genomic DNA was extracted from the peripheral blood samples of each subject. In the discovery stage, genotyping was performed with the GeneChip Human Mapping 6.0 (Affymetrix, Inc, Santa Clara, CA, USA) set as described previously.^[14] In the replication stage, the TaqMan-based assay was applied to genotype the candidate SNPs on the ABI PRISM 7900 HT platform (Applied Biosystems, Inc, Foster City, CA, USA). Then Cox proportional hazards regression adjusted for sex, age, and stage of disease was used to measure the effects of candidate SNPs on survival. Kaplan–Meier method was used to estimate and plot the survival time of different groups. The log-rank test was used to examine whether the distributions of lifetimes are distinguishable. Survival analyses were performed with R (3.3.1) using “coxph” function in “Survival package.”

Cell lines and culture

The human pancreatic cancer cell lines, CFPAC-1 (RRID: CVCL_1119) and BxPC-3 (RRID: CVCL_0186), were obtained from Procell Life Science & Technology, Wuhan, China. CFPAC-1 and BxPC-3 were maintained in RPMI-1640 and Dulbecco's Modified Eagle Medium (DMEM) cell culture media (Gibco, Waltham, MA, USA), respectively. Both culture media were supplemented with 10% (v/v) fetal bovine serum (PAN, Adenbach, Bavaria, Germany), 50 units/ml Penicillin-Streptomycin (Gibco,

USA). Cells were grown at 37°C in a humidified air and 5% CO₂ atmosphere. Both cell lines were characterized by Genetic Testing Biotechnology Corporation (Suzhou, China) using short tandem repeat markers within the last 3 years. All experiments were performed with mycoplasma-free cells.

Dual-luciferase reporter gene assays

The promoter region (1110 bp upstream of the transcription start site) of Ring Finger and WD Repeat Domain 3 (*RFWD3*) gene containing different alleles of SNP was cloned into the pGL3-Basic Luciferase Reporter Vectors (Promega, Madison, WI, USA). The Lipofectamine 3000 Transfection Reagent (Invitrogen, Waltham, MA, USA) was used to transfect constructed plasmids to cells according to the manufacturer's instruction. The transfected cells were assayed for relative luciferase activity, after 24-h incubation, with the Dual-Luciferase Reporter Assay System (Promega, USA). For each plasmid, three independent transfection experiments were performed with each in triplicate.

Electrophoretic mobility shift assays (EMSAs)

The probes containing different alleles were synthesized with/without biotin labeling at the 3' end (Takara, Kyoto, Japan). The nucleoproteins were extracted with the Nuclear and Cytoplasmic Protein Extraction Kit (Beyotime, Shanghai, China). The EMSAs were performed with the EMSA/Gel-Shift Kit (Beyotime, China) according to the instructions. For super-shift reactions, 1 µg of anti-REST antibody (Abcam, ab70300, US) was incubated with reaction mixtures for 20 min at room temperature before the addition of labeled probes.

Transient knockdown and overexpression of genes

We utilized the small interfering RNA (siRNA) and pcDNA3.1(+) vector (Invitrogen, USA) to transiently knock down and overexpress the genes, respectively. The siRNAs for *RFWD3* were purchased from the Guangzhou RiboBio Co. Ltd. (Guangzhou, China). The pcDNA3.1(+) vector containing full-length complementary DNA (cDNA) of *RFWD3* was commercially synthesized by Genewiz Biological Technology (Nanjing, China). CFPAC-1 and BxPC-3 were seeded in six-well plates the day before transfection. The siRNAs were transfected (50 pmol/well) to the cells using the Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, USA), while the pcDNA3.1(+) plasmids were transfected with the lipofectamine 3000 transfection reagent. After 48-h incubation, cells were harvested for following assay.

Cell viability assay

Cells were seeded in 96-well plates, with each well containing 3000 cells in 100 µL of cell suspension. After different periods of time in culture, cell viabilities were measured using Cell Counting Kit-8 (Dojindo, Japan) following the manufacturer's instructions. Each experiment with six replicates was repeated three times.

Transwell migration assay

Cells (2.5×10^4) in 100 µL serum-free medium were added to the transwell upper chamber (8 µm-pore polycarbonate membrane) of a 24-well plate (Corning, NY, USA). 650 µL of cell media containing 20% FBS was added to the lower compartment. After 20-h incubation at 37°C in a 5% CO₂, cells that migrate through the membrane were fixed with methanol, stained with 0.5% crystal violet, and photographed. Cell number on four random fields was counted. Each experiment was repeated three times.

Quantitative real-time polymerase chain reaction (PCR) analysis

Total RNA was extracted from cells with the Trizol reagent (Invitrogen, USA), and reversely transcribed to cDNA using the PrimeScript RT Master Mix (Takara, Japan). Real-time PCR amplification was carried out using the SYBR Green-based PCR Master Mix (Takara, Japan). The ViiA™ 7 Real-Time PCR System (Applied Biosystems, Inc, USA) was used for all reactions in a total volume of 10 µL. GAPDH was employed as an internal control for quantification of mRNA levels of target genes. The relative expression of the target gene was calculated by $2^{-\Delta\Delta C_t}$ method. The primers were provided as follows: *RFWD3* Forward Primer: 5'-CCTGACAGAAGTGGAGGT-3'; *RFWD3* Reverse Primer: 5'-CATTGAATGCAACGAAGA-3'; *GAPDH* Forward Primer: 5'-GGAGTCCACTGGCGTCTTCA-3'; *GAPDH* Reverse Primer: 5'-GTCATGAGTCCCTCCACGATACC-3'.

Bioinformatics analysis

Haploreg and RegulomeDB tools were used to perform the functional annotation for SNPs in non-coding regions.^[15,16] The ChIP-seq data of transcription factor (TF) and histones were downloaded and visualized with CistromeDB.^[17] Metascape and KOBAS were used to perform genes function annotation and pathway analysis.^[18,19]

Statistical analysis

Survival analyses were performed with the "survival" package in R software (3.3.1). Details of the statistical analysis for each experiment can be found in the relevant table notes or figure legends. Each experiment was conducted independently at least three times, and the values were presented as the means ± SE of the mean unless otherwise noted. For all biological experimental analyses, $P < 0.05$ was regarded as significantly different, and all tests were two sided.

Results

A total of 128 genes were identified to be associated with PAAD prognosis by a meta-analysis

In order to systematically identify the genes associated with the prognosis of PAAD patients, we performed a meta-analysis on seven independent public datasets, which collectively contained the expression data of 11,666

shared genes and the clinical information of 812 PAAD patients [Supplementary Table 1, <http://links.lww.com/CM9/B81>]. To facilitate integrating analyses, HR and SE acquired by univariate Cox regression for individual datasets were combined to yield meta HR, *P* value and FDR for the prognostic significance of each gene. As a result, a total of 128 genes were considered significantly correlated with the PAAD survival as $FDR < 1 \times 10^{-3}$ [Figure 1A and 1B, Supplementary Table 2, <http://links.lww.com/CM9/B81>]. Taking *RFWD3* gene for example, the high expression level was significantly associated with adverse prognosis with the combined HR of meta-analysis being 1.50 (95% CI = 1.25–1.79, $FDR = 6.9 \times 10^{-4}$) [Figure 1C]. The heterogeneity of included datasets was modest as $I^2 < 50\%$. No significant publication bias was found ($P = 0.29$ for Egger test). Sensitivity analysis demonstrated that the results are stable and reliable. Further Gene Ontology analysis on the 128 genes suggested that these PAAD prognostic genes were mainly enriched in cell cycle regulation and DNA damage response related function and process [Figure 1D].

Fourteen SNP-prognostic gene pairs were identified by eQTL analysis in PAAD

To further identify the regulatory SNPs for the prognostic genes, we performed an eQTL analysis with the gene expression profile and SNP genotype data of PAAD individuals in TCGA. Consequently, we found the tissue-specific eQTL relationships between 104,019 SNPs and 2468 genes, making up 113,810 SNP-gene pairs, at the

threshold of $FDR < 0.05$. Integrating with the results of above meta-analysis, we found a total of 727 SNPs associated with the expression of an identified prognostic genes. We then applied the linkage disequilibrium analysis to the 727 SNPs with Haploview to eliminate the redundancy, and obtained 14 probably functional eQTLs representing 14 blocks to which the 727 SNPs converged to [Figure 1A, Supplementary Table 3, <http://links.lww.com/CM9/B81>]. Accordingly, we presented 14 tag SNPs that potentially regulated the expression of the genes with prognostic significance in PAAD.

eQTLs of *RFWD3* tagged by rs4887783 were correlated with PAAD prognosis

To evaluate the prognostic significance of those eQTLs, we conducted a two-stage survival analysis in a total of 893 PAAD patients. The characteristics of the patients were delineated in Supplementary Table 4, <http://links.lww.com/CM9/B81>. In the discovery stage, we assessed the 14 tag SNPs using our GWAS data^[14] and identified that only rs4887783, the eQTL of *RFWD3*, exhibited association with PAAD survival ($FDR < 0.05$) [Table 1, Figure 2A]. We then tested this SNP within an independent cohort of 552 patients in the replication stage. As a result, it was successfully replicated with $P < 0.05$ [Figure 2B]. The per-allele HR adjusted for age, gender and stage of disease in the Cox regression of this variant was 1.34 (95% CI = 1.10–1.63; $P = 0.004$) and 1.20 (95% CI = 1.00–1.43; $P = 0.046$) in the discovery stage and replication stage, respectively. When combining the two

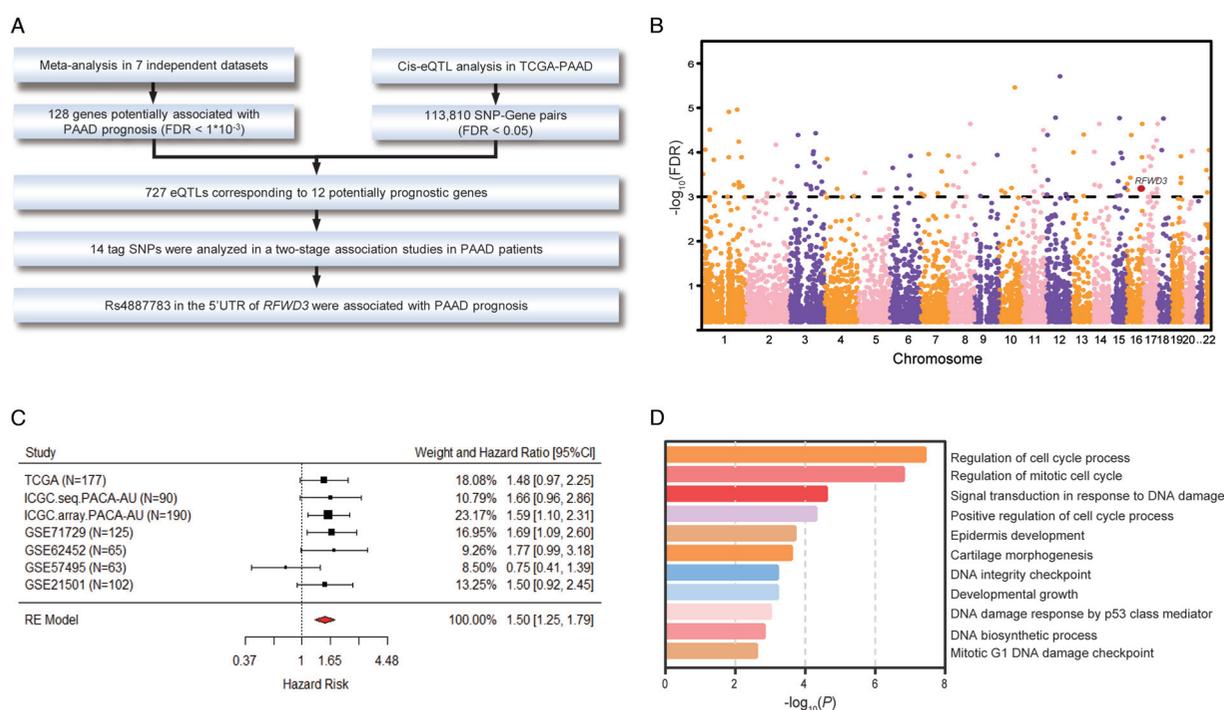


Figure 1: One hundred twenty-eight genes were identified to be associated with PAAD prognosis through a meta-analysis. (A) Flow chart of exploring the regulatory SNP associated with PAAD prognosis. (B) The Manhattan plot visualizes the association of gene expression with PAAD survival in the meta-analysis. The dots above the horizontal line indicates the genes with false discovery rate $< 1 \times 10^{-3}$. (C) The forest plot summarizes the association of *RFWD3* with PAAD prognosis in each dataset and the integrated meta-analysis. (D) The Gene Ontology annotation of the 128 prognostic genes. eQTL: Expression quantitative trait loci; FDR: False discovery rate; PAAD: Pancreatic adenocarcinoma; *RFWD3*: Ring finger and WD repeat domain 3; SNP: Single nucleotide polymorphism; TCGA: The Cancer Genome Atlas.

Table 1: The 14 tag single nucleotide polymorphisms (SNPs) in the discovery stage.

SNP	Alleles	MAF	Location	Type	Gene	FDR for eQTL analysis	HR (95% CI)	P values	FDR
rs12743200	G/A	0.29	Chr1: 224456358	Intronic	CNIH4	1.61×10^{-9}	0.94 (0.74–1.19)	0.592	0.6191
rs6426123	C/T	0.27	Chr1: 224436689	Intronic	CNIH4	8.52×10^{-10}	0.87 (0.54–1.40)	0.562	0.6119
rs10935606	C/T	0.20	Chr3: 146150361	Intronic	PLOD2	4.43×10^{-5}	0.88 (0.73–1.06)	0.184	0.4647
rs8180086	A/G	0.49	Chr3: 151320092	Intronic	GPR87	3.24×10^{-5}	1.02 (0.86–1.20)	0.778	0.6786
rs12663159	C/A	0.47	Chr6: 39314501	2 kb upstream	KCNK17	2.92×10^{-5}	0.92 (0.79–1.08)	0.299	0.5098
rs892845	G/A	0.22	Chr8: 142782660	–	LY6D	5.55×10^{-9}	1.06 (0.84–1.34)	0.608	0.6227
rs55977291	A/C	0.18	Chr8: 57539630	Intronic	FAM110B	2.46×10^{-5}	0.83 (0.59–1.16)	0.277	0.5036
rs2279863	G/T	0.45	Chr11: 66480225	5'-UTR	CTSF	4.86×10^{-5}	0.94 (0.81–1.09)	0.386	0.5544
rs7108388	G/C	0.23	Chr11: 66932151	Intronic	CTSF	7.81×10^{-6}	1.02 (0.81–1.28)	0.880	0.7049
rs4887783	C/G	0.17	Chr16: 74666845	5'-UTR	RFWD3	4.83×10^{-6}	1.34 (1.10–1.63)	0.004	0.0409
rs2661685	G/A	0.25	Chr17: 78221313	Intronic	BIRC5	2.60×10^{-5}	1.27 (0.86–1.88)	0.233	0.4883
rs62076114	A/G	0.22	Chr17: 82940215	Intronic	SLC16A3	3.95×10^{-5}	1.03 (0.86–1.22)	0.775	0.6778
rs8085523	C/T	0.14	Chr18: 31447794	5'-UTR	DSG3	2.47×10^{-7}	1.04 (0.83–1.32)	0.718	0.6609
rs4820590	T/C	0.47	Chr22: 24478729	Intronic	MIF	9.49×10^{-5}	0.87 (0.75–1.02)	0.079	0.3559

HR and P value were calculated using Cox proportional hazards regression with adjustment for age, gender, and stage of disease. Chr: Chromosome; CI: Confidence interval; eQTL: Expression quantitative trait loci; FDR: False discovery rate; HR: Hazard ratio; MAF: Minor allele frequency; SNP: Single nucleotide polymorphism; UTR: Untranslated region.

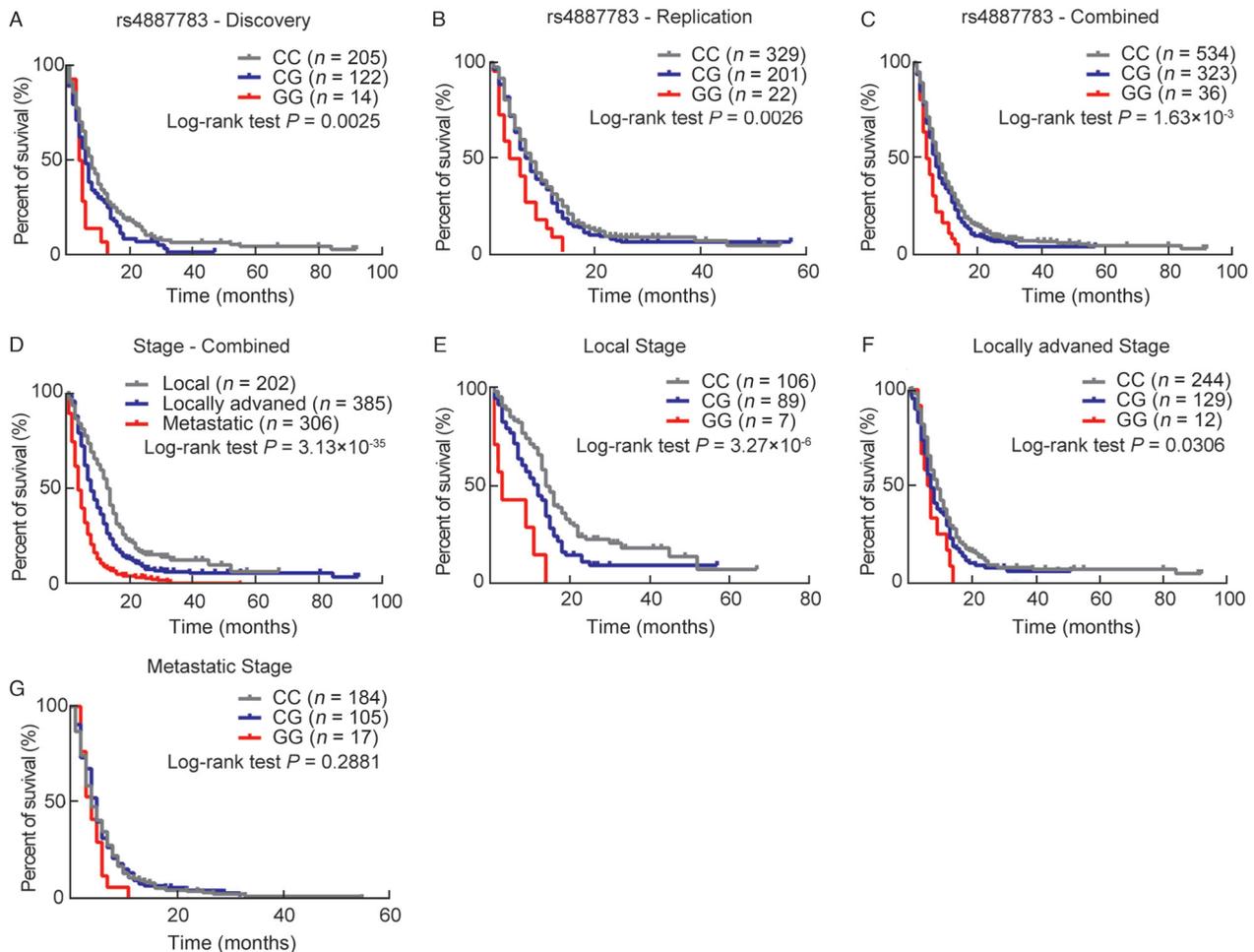


Figure 2: Rs4887783 was correlated with the overall survival of PAAD patients. (A–C) Kaplan–Meier survival estimates by different genotypes of rs4887783 in discovery, replication and combined stages, respectively. (D) Kaplan–Meier survival estimates by different stages of disease in combined stage. (E–G) Stratification analysis of different genotypes in local, locally advanced and metastatic stages. PAAD: Pancreatic adenocarcinoma.

Table 2: Association between rs4887783 and overall survival in pancreatic adenocarcinoma patients.

rs4887783	Discovery stage			Replication stage			Combined stage					
	N (%)	MST (months)	HR (95% CI)	P value	N (%)	MST (months)	HR (95% CI)	P value	N (%)	MST (months)	HR (95% CI)	P value
CC	205 (60.12)	8.4	1.30 (1.02-1.65)	0.0330	329 (59.60)	8.6	1.14 (0.95-1.38)	0.1530	534 (59.80)	8.5	1.20 (1.03-1.38)	0.0160
CG	122 (35.78)	6.8	1.86 (1.06-3.25)	0.0310	201 (36.41)	8.0	2.02 (1.30-3.14)	0.0020	323 (36.17)	7.4	1.96 (1.39-2.77)	0.0001
GG	14 (4.11)	5.0	1.35 (1.07-1.70)	0.0120	22 (4.00)	6.0	1.20 (1.00-1.43)	0.0460	36 (4.03)	5.0	1.25 (1.08-1.44)	0.0020
Additive model			2.02 (0.96-4.25)	0.0370			1.91 (1.24-2.93)	0.0030			1.84 (1.31-2.58)	0.0004
Recessive model			1.34 (1.10-1.63)	0.0040			1.24 (1.06-1.45)	0.0080			1.27 (1.12-1.43)	0.0001
Dominant model												

HR and P value were calculated using Cox proportional hazards regression with adjustment for age, gender and stage of disease. CI: Confidence interval; HR: Hazard ratio; MST: Median survival time.

stages, the rs4887783 was significantly associated with the overall survival of PAAD patients with HR being 1.25 (95% CI = 1.08-1.44; P = 0.002) [Table 2, Figure 2C]. Since the tumor stage was a prognostic factor of PAAD [Figure 2D], we further performed a stratification analysis based on this character. The result indicated that rs4887783 significantly correlated with survival especially in PAAD patients with local and locally advanced disease, with the HRs being 1.74 (95% CI = 1.31-2.30; P = 0.0002) and 1.26 (95% CI = 1.04-1.53; P = 0.0180), respectively [Figure 2E-G].

Rs4887783 variation increased RFWD3 expression through enhancing the binding of REST

For the purpose of discovering the functional variant, we performed fine-mapping for the prognostic locus tagged by rs4887783. Taking advantage of the phase three resource of 1000 genome project, we obtained 23 SNPs in linkage disequilibrium ($r^2 > 0.8$) with rs4887783, and subjected these variants to functional annotation and prediction. Consequently, rs4887783 was considered most promisingly functional as it is located in the 5' untranslated region of *RFWD3* and in the active region of transcriptional regulation marked by the H3K27ac and H3K4me3 [Figure 3A], thus indicating the possibly regulatory function of rs4887783 to the *RFWD3* promoter activity. In accordance with the functional annotation, the eQTL analysis in TCGA-PAAD samples presented that rs4887783 genotype was associated with the expression level of *RFWD3* [Figure 3B].

To investigate the regulatory effect of genetic variation rs4887783 on *RFWD3* promoter activity, we conducted a dual-luciferase reporter assay in CFPAC-1 and BxPC-3 human pancreatic cancer cell lines. We observed that the construct containing rs4887783-G exhibits significantly higher promoter activity than that containing the C allele [Figure 3C]. Since the variant in promoter region usually exerts its regulatory function through modulating the TF binding, we utilized the Cistrome database browser to predict the potential binding TF and found that rs4887783 maps within the binding motif of REST [Figure 3A]. The strong positive correlation of genes expression between *RFWD3* and REST (Pearson $r = 0.76$, $P < 0.0001$) also indicated a possibly underlying regulatory relationship [Figure 3D]. We then validated the allele-specific binding of REST with EMSAs. The results showed that the nuclear extracts bound selectively to the G allele ("G probe") rather than the C allele ("C probe") of an oligonucleotide probe representing the promoter segments of the rs4887783. Allele specificity of the binding was confirmed by cold competition experiments, since unlabeled G probe, but not unlabeled C probe, could gradually attenuate the binding signal of nuclear extract to the biotin-labeled G probe in a dose-dependent manner [Figure 3E]. Moreover, super-shift EMSA was carried out to further demonstrate the REST as the allele-specific binding TF, as the antibody for REST super-shifted the complex formation with the G probe [Figure 3F]. According to the above, we proved that rs4887783-G could increase *RFWD3* expression through enhancing the TF binding of REST.

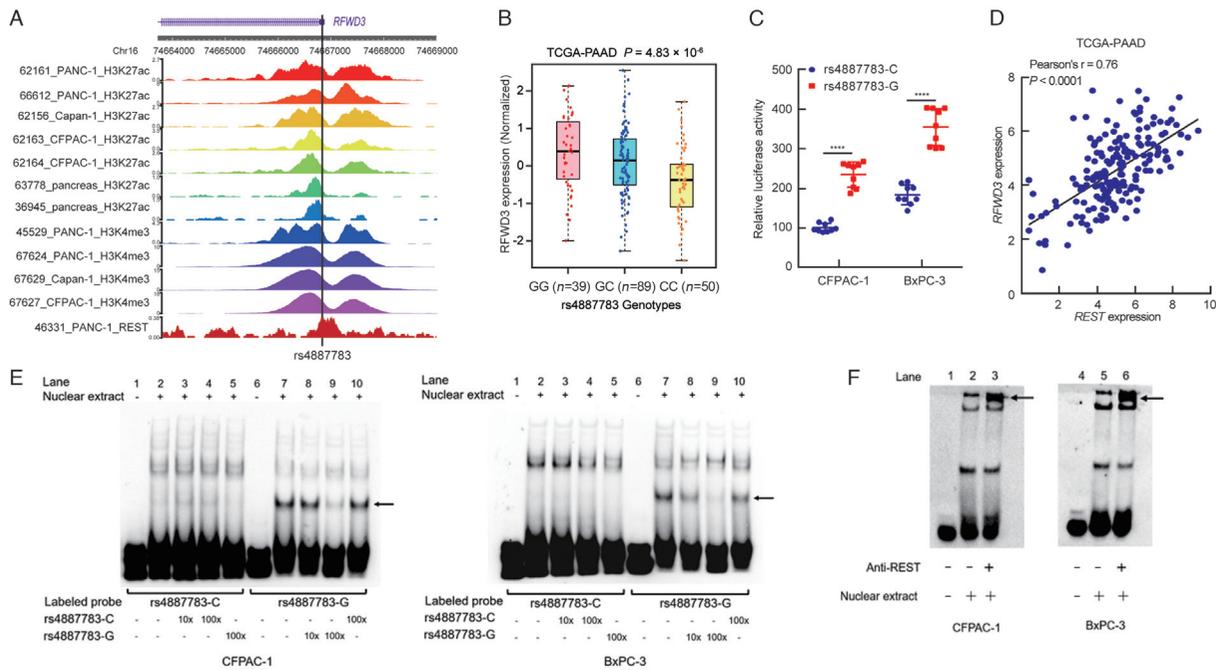


Figure 3: rs4887783 variation increased *RFWD3* expression through enhancing the binding of REST. (A) rs4887783 is located in the active promoter region represented by the histone modification markers (H3K27ac and H3K4me3) and transcription factor *RFWD3* binding regions in pancreas tissues and pancreatic cancer cell lines. (B) Different genotypes of rs4887783 are associated with the *RFWD3* expression in pancreatic cancers from TCGA. Data were shown as the mean ± standard deviation and *P* values were calculated by linear regression analysis. (C) Relative reporter gene activity from constructs bearing promoter fragments of *RFWD3* with different alleles of the rs4887783 in the CFPAC-1 and BxPC-3 cell lines. Data represent mean ± standard error from three independent transfection experiments with assays conducted in three replications. * represents for *P* < 0.0001 of student's *t* test. (D) The strong positive correlation of genes expression between *RFWD3* and *REST* in PAAD from TCGA. (E–F) EMSAs and REST super-shift EMSAs with biotin-labeled probes containing rs4887783 in CFPAC-1 and BxPC-3 cells. Arrows indicate allele-specific bands that interact with nuclear protein in the cells (E) and the super-shifted band (F). In addition, 10× and 100×, respectively, represent ten-fold and 100-fold excess amounts of an unlabeled probe compared with the amount of the labeled probe. “+” and “-” indicate added and not added, respectively. EMSA: Electrophoretic mobility shift assay; PAAD: Pancreatic adenocarcinoma; *RFWD3*: Ring finger and WD repeat domain 3; TCGA: The Cancer Genome Atlas.

***RFWD3* was highly expressed in PAAD and promoted cancer cells migration**

Through analyzing the gene expression data from GEO datasets (GSE62452), we discovered that the *RFWD3* expressed higher in tumorous compared to normal pancreatic tissues [Figure 4A]. Additional assessment with TCGA resources presented that the expression levels of *RFWD3* were also significantly higher in a number of other digestive tumors than normal tissues [Figure 4B]. These findings suggested the important role of *RFWD3* in tumor development and progression. To further verify the effects of *RFWD3* on PAAD development, we conducted a series of phenotypic experiments after artificial alteration of this gene expression in cells. As a result, transient knock-down of *RFWD3* with siRNAs significantly inhibited pancreatic cancer cells migration, while over-expression with gene expression vector promoted them [Figure 4C–E]. It is to be noted that we did not observe the cell proliferation rate being modulated by *RFWD3* knockout or overexpression [Figure 4F].

***RFWD3* might orchestrate the genes in DNA repair process**

With the intention to preliminarily elucidate the molecular mechanism behind the promotion effects of *RFWD3* on PAAD progression, we analyzed the genes expression in TCGA-PAAD, and found a total of 539 genes correlated with *RFWD3* (|Pearson correlation coefficient >0.5, FDR <0.05). Moreover, these co-expressed genes were enriched

in the cell cycle and DNA repair related biological process [Figure 5], indicating that *RFWD3* might promote PAAD progression and impair patients' survival through affecting the cell cycle progression and DNA damage repair (DDR) pathway.

Discussion

PAAD is a highly malignant disease with unimproved dismal prognosis over the past decades. Identifying the functional genetic variants associated with PAAD prognosis is beneficial but challenging. In the current study, firstly, we systematically discovered 128 potentially prognostic genes for PAAD through a meta-analysis. Next, the eQTLs for 12 prognostic genes were ascertained via a comprehensive tissue-specific investigation. Then, rs4887783, the eQTL of *RFWD3*, was identified to be associated with adverse outcome of PAAD in a two-stage population study. Finally, a series of functional studies demonstrated that rs4887783-G enhances the TF binding of REST to the *RFWD3* promoter, thus elevating *RFWD3* expression and promoting PAAD progression.

Applying a meta-analysis to the public dataset is a bright spot of this study. Meta-analysis is an ideal approach to enhance statistical power and improve reproducibility, which is widely used in numerous studies in regard to discovery of genetic biomarkers.^[20–22] Former studies leveraged high-throughput genomic technologies to investigate the association between PAAD prognosis and gene

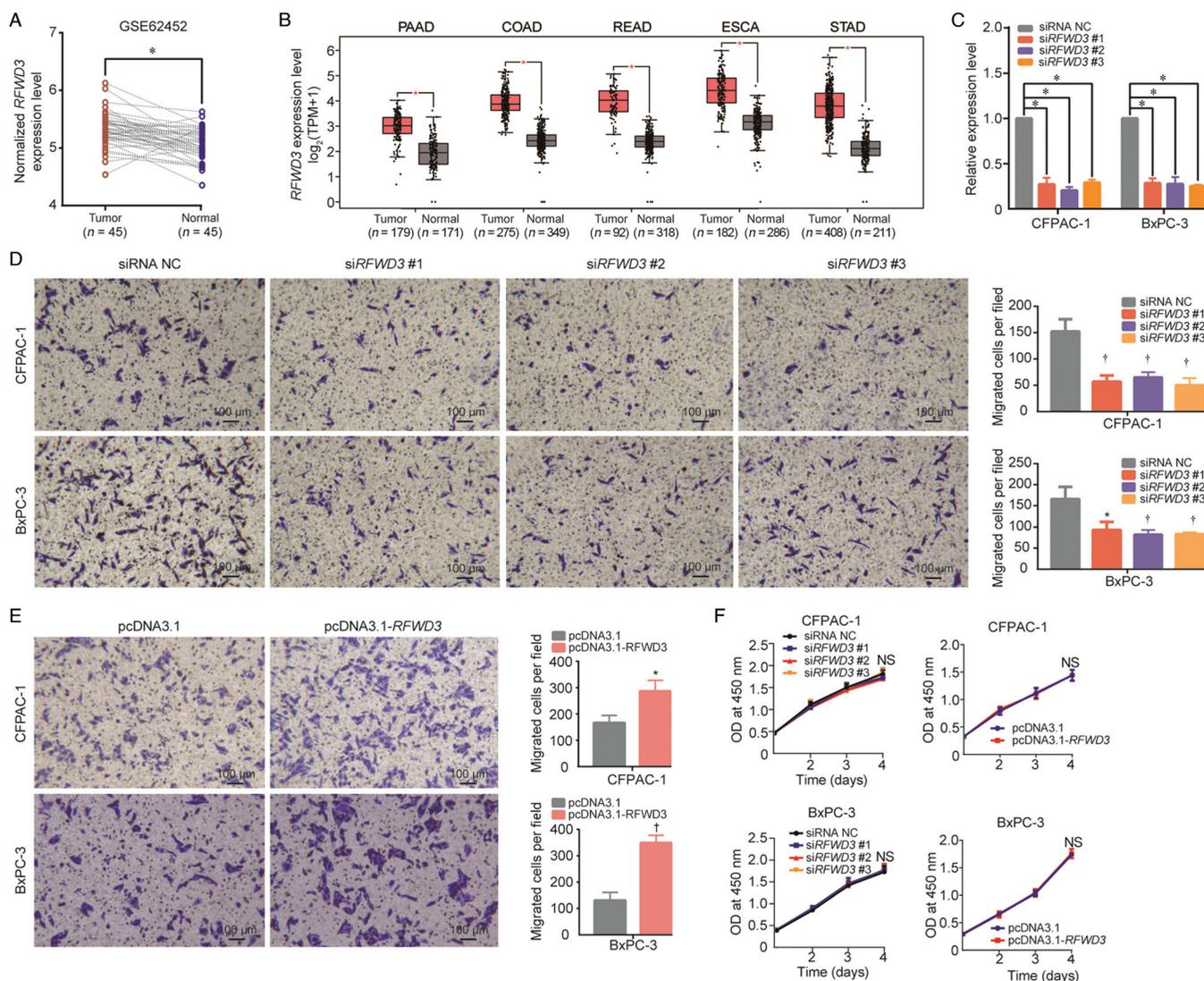


Figure 4: *RFWD3* is highly expressed in PAAD and promotes cancer cells migration. (A) The expression of *RFWD3* is higher in PAAD tumor tissues than paired normal tissues in GSE62452 dataset. * represents for $P < 0.0001$ of paired *t* test. (B) The differential expression of *CAV2* between normal and the digestive tumor tissues in TCGA. Red asterisk represents for $P < 0.0001$ of student's *t* test. (C) Significant knockdown of *RFWD3* by three siRNAs is demonstrated at mRNA level. * represents for $P < 0.0001$ of student's *t* test. (D–E) Transwell assays demonstrate that *RFWD3* knockdown significantly suppresses cell migration, while overexpression promotes cell migration. Representative photographs of tumor cell migrating through and fixed on the lower surface of the transwell membrane are shown for each group (under microscope, original magnification $\times 100$). The numbers of migrating cells are counted in four random fields (bar plot). * Represents for $P < 0.05$, † Represents for $P < 0.01$ of student's *t* test. (F) CCK-8 assays indicate no significant difference of proliferation between the *RFWD3* knockdown or overexpression cells and negative control cells over a period of 4 days. CCK-8: Cell counting kit-8; NS: Not significant; PAAD: Pancreatic adenocarcinoma; *RFWD3*: Ring finger and WD repeat domain 3; TCGA: The Cancer Genome Atlas.

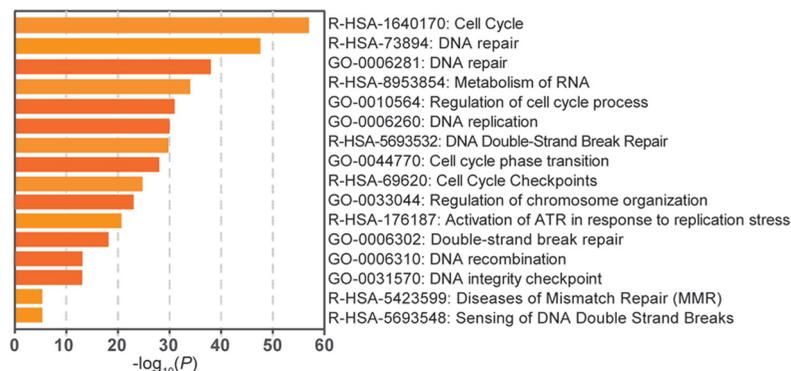


Figure 5: *RFWD3* might orchestrate the genes in DNA repair process. *RFWD3* co-expressed genes enriched in the Gene Ontology terms or Reactome pathways. *RFWD3*: Ring finger and WD repeat domain 3.

expression, but the results were inconsistent.^[11,12,23] Here, by comprehensively searching the TCGA, ICGC, and GEO databases, we gleaned the qualified gene expression profile and clinical information of 812 PAAD patients from seven independent datasets, and conducted a meta-analysis to systemically identify the prognostic genes. The result is more reliable and of greater translational impact than that acquired from existing single dataset.

RFWD3 is a protein coding gene of the E3 Ubiquitin-Protein Ligase *RFWD3* which plays a crucial part in DNA damage response. Under the conditions of replication stress and fork stalling, the E3 ligase *RFWD3* mediates the multisite ubiquitination of the replication protein A complex and contributes to both the restart of stalled replication forks and homologous recombination (HR) at stalled forks.^[24-26] *RFWD3* is also identified as a positive regulator of p53 stability even in the presence of high levels of MDM2 in the late response to DNA damage.^[27] Conversely, the functional impairment of *RFWD3* is reported to result in unrepaired DNA damage and connect to the initiation of Fanconi anemia, lung cancer and multiple myeloma.^[28-30] Here, our study added to the knowledge of *RFWD3* and cancer progression and outcome. We clarified that high expression of *RFWD3* promoted PAAD cell migration and correlated with adverse prognosis. Through co-expression analysis, we found *RFWD3* orchestrated with genes in DDR process. Since DDR is vital for maintaining the genomic stability and cell survival, tumor cells might withstand and survive from the DNA damage caused by the radiochemotherapy and other clinical treatment by dint of the *RFWD3* enhanced DNA repair ability, mechanically. In accordance with the existing awareness that genetic variants in the DDR related genes can affect PAAD survival,^[31-33] our findings presented and proved a new prognostic variant, rs4887783, that affected survival through transcriptionally regulating *RFWD3* expression. These findings need to be further confirmed *in vivo* in future studies.

In summary, by systematical investigation on the regulatory variants of the prognostic genes for PAAD, we found for the first time that the SNP rs4887783 could predict poor survival through increasing expression of the DDR related gene *RFWD3*. What we found might facilitate the discovery of biomarkers and therapeutic targets for PAAD, although validation with larger sample size and deeper mechanistic interpretation is still warranted in the future.

Funding

This work was supported by grants from the Youth Program of National Natural Science Foundation of China (No. NSFC-82003547), China Postdoctoral Science Foundation (No. 2019M662644), and the Fundamental Research Funds for the Central Universities (No. WHU 2042022kf1031) for Ying Zhu, National Science Fund for Distinguished Young Scholars of China (No. NSFC-81925032) and Natural Science Foundation of Hubei Province (No. 2019CFA009) for Xiaoping Miao, National Program for Support of Top-notch Young Professionals and the Young Elite Scientist Sponsorship Program by CAST (No. 2018QNRC001) for Jiang Chang,

Youth Program of National Natural Science Foundation of China (No. NSFC-82103929) and the Fundamental Research Funds for the Central Universities (No. WHU 2042022kf1205) for Jianbo Tian.

Conflicts of interest

None.

References

- Cao W, Chen HD, Yu YW, Li N, Chen WQ. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chin Med J* 2021;134:783–791. doi: 10.1097/CM9.0000000000001474.
- Gentiluomo M, Canzian F, Nicolini A, Gemignani F, Landi S, Campa D. Germline genetic variability in pancreatic cancer risk and prognosis. *Semin Cancer Biol* 2020;79:105–131. doi: 10.1016/j.semcancer.2020.08.003.
- Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, et al. Pancreatic cancer. *Nat Rev Dis Primers* 2016;2:16022. doi: 10.1038/nrdp.2016.22.
- Wu C, Kraft P, Stolzenberg-Solomon R, Steplowski E, Brotzman M, Xu M, et al. Genome-wide association study of survival in patients with pancreatic adenocarcinoma. *Gut* 2014;63:152–160. doi: 10.1136/gutjnl-2012-303477.
- Tang H, Wei P, Chang P, Li Y, Yan D, Liu C, et al. Genetic polymorphisms associated with pancreatic cancer survival: a genome-wide association study. *Int J Cancer* 2017;141:678–686. doi: 10.1002/ijc.30762.
- Dimitrakopoulos C, Vrugt B, Flury R, Schraml P, Knippschild U, Wild P, et al. Identification and validation of a biomarker signature in patients with resectable pancreatic cancer via genome-wide screening for functional genetic variants. *JAMA Surg* 2019;154:e190484. doi: 10.1001/jamasurg.2019.0484.
- Willis JA, Olson SH, Orlov I, Mukherjee S, McWilliams RR, Kurtz RC, et al. A replication study and genome-wide scan of single-nucleotide polymorphisms associated with pancreatic cancer risk and overall survival. *Clin Cancer Res* 2012;18:3942–3951. doi: 10.1158/1078-0432.CCR-11-2856.
- Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* 2015;16:197–212. doi: 10.1038/nrg3891.
- Zou D, Lou J, Ke J, Mei S, Li J, Gong Y, et al. Integrative expression quantitative trait locus-based analysis of colorectal cancer identified a functional polymorphism regulating *SLC22A5* expression. *Eur J Cancer* 2018;93:1–9. doi: 10.1016/j.ejca.2018.01.065.
- Tian J, Lou J, Cai Y, Rao M, Lu Z, Zhu Y, et al. Risk SNP-mediated enhancer-promoter interaction drives colorectal cancer through both *FADS2* and *AP002754.2*. *Cancer Res* 2020;80:1804–1818. doi: 10.1158/0008-5472.CAN-19-2389.
- Stratford JK, Bentrem DJ, Anderson JM, Fan C, Volmar KA, Marron JS, et al. A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. *PLoS Med* 2010;7:e1000307. doi: 10.1371/journal.pmed.1000307.
- Chen DT, Davis-Yadley AH, Huang PY, Husain K, Centeno BA, Permuth-Wey J, et al. Prognostic fifteen-gene signature for early stage pancreatic ductal adenocarcinoma. *PLoS One* 2015;10:e0133562. doi: 10.1371/journal.pone.0133562.
- Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, et al. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res* 2018;46:D971–D976. doi: 10.1093/nar/gkx861.
- Wu C, Miao X, Huang L, Che X, Jiang G, Yu D, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. *Nat Genet* 2012;44:62–66. doi: 10.1038/ng.1020.
- Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–D934. doi: 10.1093/nar/gkr917.
- Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–1797. doi: 10.1101/gr.137323.112.

17. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, *et al.* Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res* 2019;47:D729–D735. doi: 10.1093/nar/gky1094.
18. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523. doi: 10.1038/s41467-019-09234-6.
19. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 2011;39:W316–W322. doi: 10.1093/nar/gkr483.
20. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015;21:938–945. doi: 10.1038/nm.3909.
21. Thakurta A, Ortiz M, Bleuca P, Towfic F, Corre J, Serbina NV, *et al.* High subclonal fraction of 17p deletion is associated with poor prognosis in multiple myeloma. *Blood* 2019;133:1217–1221. doi: 10.1182/blood-2018-10-880831.
22. Fisher OM, Lord SJ, Falkenback D, Clemons NJ, Eslick GD, Lord RV. The prognostic value of TP53 mutations in oesophageal adenocarcinoma: a systematic review and meta-analysis. *Gut* 2017;66:399–410. doi: 10.1136/gutjnl-2015-310888.
23. Zhang G, Schetter A, He P, Funamizu N, Gaedcke J, Ghadimi BM, *et al.* DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One* 2012;7:e31507. doi: 10.1371/journal.pone.0031507.
24. Elia AE, Wang DC, Willis NA, Boardman AP, Hajdu I, Adeyemi RO, *et al.* RFWD3-dependent ubiquitination of RPA regulates repair at stalled replication forks. *Mol Cell* 2015;60:280–293. doi: 10.1016/j.molcel.2015.09.011.
25. Feeney L, Munoz IM, Lachaud C, Toth R, Appleton PL, Schindler D, *et al.* RPA-mediated recruitment of the E3 ligase RFWD3 is vital for interstrand crosslink repair and human health. *Mol Cell* 2017;66:610.e4–621.e4. doi: 10.1016/j.molcel.2017.04.021.
26. Dubois JC, Yates M, Gaudreau-Lapierre A, Clement G, Cappadocia L, Gaudreau L, *et al.* A phosphorylation-and-ubiquitylation circuitry driving ATR activation and homologous recombination. *Nucleic Acids Res* 2017;45:8859–8872. doi: 10.1093/nar/gkx571.
27. Fu X, Yucer N, Liu S, Li M, Yi P, Mu JJ, *et al.* RFWD3-Mdm2 ubiquitin ligase complex positively regulates p53 stability in response to DNA damage. *Proc Natl Acad Sci U S A* 2010;107:4579–4584. doi: 10.1073/pnas.0912094107.
28. Zhang Y, Zhao X, Zhou Y, Wang M, Zhou G. Identification of an E3 ligase-encoding gene RFWD3 in non-small cell lung cancer. *Front Med* 2020;14:318–326. doi: 10.1007/s11684-019-0708-6.
29. Knies K, Inano S, Ramirez MJ, Ishiai M, Surralles J, Takata M, *et al.* Biallelic mutations in the ubiquitin ligase RFWD3 cause Fanconi anemia. *J Clin Invest* 2017;127:3013–3027. doi: 10.1172/JCI92069.
30. Mitchell JS, Li N, Weinhold N, Forsti A, Ali M, van Duin M, *et al.* Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun* 2016;7:12050. doi: 10.1038/ncomms12050.
31. Sehdev A, Gbolahan O, Hancock BA, Stanley M, Shahda S, Wan J, *et al.* Germline and somatic DNA damage repair gene mutations and overall survival in metastatic pancreatic adenocarcinoma patients treated with FOLFIRINOX. *Clin Cancer Res* 2018;24:6204–6211. doi: 10.1158/1078-0432.CCR-18-1472.
32. Li D, Frazier M, Evans DB, Hess KR, Crane CH, Jiao L, *et al.* Single nucleotide polymorphisms of RecQ1, RAD54L, and ATM genes are associated with reduced survival of pancreatic cancer. *J Clin Oncol* 2006;24:1720–1728. doi: 10.1200/JCO.2005.04.4206.
33. Okazaki T, Jiao L, Chang P, Evans DB, Abbruzzese JL, Li D. Single-nucleotide polymorphisms of DNA damage response genes are associated with overall survival in patients with pancreatic cancer. *Clin Cancer Res* 2008;14:2042–2048. doi: 10.1158/1078-0432.CCR-07-1520.

How to cite this article: Zhu Y, Peng X, Wang X, Ying P, Wang H, Li B, Li Y, Zhang M, Cai Y, Lu Z, Niu S, Yang N, Zhong R, Tian J, Chang J, Miao X. Systematic analysis on expression quantitative trait loci identifies a novel regulatory variant in ring finger and WD repeat domain 3 associated with prognosis of pancreatic cancer. *Chin Med J* 2022;135:1348–1357. doi: 10.1097/CM9.0000000000002180