Received: 3 December 2018 | Revised: 6 March 2019 | Accepted: 7 April 2019

DOI: 10.1111/cas.14017

ORIGINAL ARTICLE

WILEY Cancer Science

Three novel genetic variants in NRF2 signaling pathway genes are associated with pancreatic cancer risk

Wenjun Yang[1,2] | Hongliang Liu[3,4] | Bensong Duan[3,4] | Xinyuan Xu[3,4] | Dennis Carmody[3,4] | Sheng Luo[5] | Kyle M. Walsh[3,6] | James L. Abbruzzese[3,7] | Xuefeng Zhang[3,8] | Xiaoxin Chen[2] | Qingyi Wei[3,4,7]

[1]Key Laboratory of Fertility Preservation and Maintenance, School of Basic Medicine and the General Hospital, Ningxia Medical University, Yinchuan, China

[2]Cancer Research Program, Julius L. Chambers Biomedical Biotechnology Research Institute, North Carolina Central University, Durham

[3]Duke Cancer Institute, Duke University Medical Center, Durham

[4]Department of Population Health Sciences, Duke University School of Medicine, Durham

[5]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham

[6]Department of Neurosurgery, Duke University School of Medicine, Durham

[7]Department of Medicine, Population Health Sciences, Duke University School of Medicine, Durham

[8]Department of Pathology, Duke University School of Medicine, Durham

Correspondence
Xiaoxin Chen, Cancer Research Program, Julius L. Chambers Biomedical Biotechnology Research Institute, North Carolina Central University, Durham, NC.
Email: lchen@nccu.edu
and
Qingyi Wei, Duke Cancer Institute, Duke University Medical Center and Department of Medicine, Duke School of Medicine, Durham, NC.
Email: qingyi.wei@duke.edu

Funding information
The North Carolina Central University, Grant/Award Number: NIH MD012392; Duke University Medical Center, Grant/Award Number: NIH CA014236; US National Institutes of Health (NIH), Grant/Award Number: HHSN261200800001E; NIH/NCI, Grant/Award Number: K07 CA140790; American Society of Clinical Oncology Conquer Cancer Foundation; Howard Hughes Medical Institute

Pancreatic cancer (PanC) is one of the most lethal solid malignancies, and metastatic PanC is often present at the time of diagnosis. Although several high- and low-penetrance genes have been implicated in PanC, their roles in carcinogenesis remain only partially elucidated. Because the nuclear factor erythroid2-related factor2 (NRF2) signaling pathway is involved in human cancers, we hypothesize that genetic variants in NRF2 pathway genes are associated with PanC risk. To test this hypothesis, we assessed associations between 31 583 common single nucleotide polymorphisms (SNP) in 164 NRF2-related genes and PanC risk using three published genome-wide association study (GWAS) datasets, which included 8474 cases and 6944 controls of European descent. We also carried out expression quantitative trait loci (eQTL) analysis to assess the genotype-phenotype correlation of the identified significant SNP using publicly available data in the 1000 Genomes Project. We found that three novel SNP (ie, rs3124761, rs17458086 and rs1630747) were significantly associated with PanC risk ($P = 5.17 \times 10^{-7}$, 5.61×10^{-4} and 5.52×10^{-4}, respectively). Combined analysis using the number of unfavorable genotypes (NUG) of these three SNP suggested that carriers of two to three NUG had an increased risk of PanC ($P < 0.0001$), compared with those carrying zero to one NUG. Furthermore, eQTL analysis showed that

Abbreviations: CI, confidence interval; EAF, effect allele frequency; eQTL, expression quantitative trait loci; FDR, false discovery rate; GLUT, glucose transporter; GWAS, genome-wide association study; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; MAF, minor allele frequency; MMR, mismatch repair; NRF2, nuclear factor erythroid2-related factor2; NUG, number of unfavorable genotypes; OR, odds ratio; PanC, pancreatic cancer; SLC, solute carrier; SNP, single nucleotide polymorphism; TCGA, The Cancer Genome Atlas; TF, transcriptional factor.

Yang and Liu contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2019 The Authors. *Cancer Science* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association.

both rs3124761 T and rs17458086 C alleles were associated with increased mRNA expression levels of *SLC2A6* and *SLC2A13*, respectively ($P < 0.05$). In conclusion, genetic variants in *NRF2* pathway genes could play a role in susceptibility to PanC, and further functional exploration of the underlying molecular mechanisms is warranted.

KEYWORDS

genome-wide association study, NRF2, pancreatic cancer susceptibility, pathway analysis, single nucleotide polymorphism

1 | INTRODUCTION

Pancreatic cancer (PanC) is one of the most lethal solid malignancies, and PanC is the fourth most common cause of cancer deaths in the USA,[1] responsible for an estimated 44 330 deaths in 2018.[2,3] The poor prognosis of pancreatic ductal adenocarcinoma relates to the advanced disease stage at the time of diagnosis and its profound resistance to therapies,[4,5] because metastatic PanC is commonly present by the time of initial diagnosis as a result of a lack of effective screening tests. To reduce the enormous death toll related to this cancer, an enhanced screening method among the at-risk populations is urgently needed. However, identification of the at-risk populations requires highly effective biomarkers that predict PanC risk.[6]

Genetic and environmental factors for the etiology of PanC remains only partially elucidated. Risk factors for PanC cancer identified in epidemiological studies include cigarette smoking, increased body mass index, heavy alcohol consumption, and a diagnosis of diabetes mellitus.[7] Recently, investigations have reported a number of genetic factors or susceptibility genes or loci for PanC risk. PanC high-penetrance genes include *BRCA1*, *BRCA2*, *TP53*, *CDKN2A*, *APC*, *STK11* and *MMR* genes, whereas low-penetrance risk loci include chromosome 9q34 (in the *ABO* blood group gene), 1q32.1 (in *NR5A2*), 5p15.33 (in the *CLPTM1L-TERT* gene region), 16q23.1 (*BCAR1*), 13q12.2 (*PDX1*), 22q12.1 (*ZNRF3*), 1p36.33 (*NOC2L*), and 22q13.1 (*PDGFB*).[7-11] However, these established genes or loci could explain only 20%-30% of PanC risk, including 5%-10% of the familial aggregation of PanC. In the remaining 80%-85% of sporadic patients, there has been limited success in resolving the genetic architecture of PanC.[8]

The nuclear factor E2-related factor 2 (NFE2L2 or NRF2) pathway is one of the major signaling cascades involved in cell defense and survival against endogenous and exogenous stress.[12] *NRF2* mediates the expression of more than 100 oxidative stress-related genes. These cytoprotective genes all contain the *cis*-regulatory element sequence antioxidant response element (5′-GTGACnnnGC-3′) in their promoter regulatory regions for NRF2 binding.[12] It has been shown that constitutive activation of *NRF2* in cancer cells increases the expression of cytoprotective genes and, consequently, enhances proliferation via metabolic reprogramming and inhibition of apoptosis.[13] A growing number of studies have shown that aberrant activation of the transcription factor NRF2 promotes PanC tumorigenesis, likely by regulating the expression of a vast array of genes.[14-16] A high rate of somatic mutations in NRF2-KEAP1 pathway genes has been observed in several types of carcinoma in TCGA database,[17] which highlights the roles of genes in the *NRF2* pathway as cancer driver genes with potential clinical ramifications. Several polymorphisms in NRF2 pathway genes have recently been identified to be associated with cancer risk, including lung adenocarcinoma and breast cancer.[18,19] Therefore, a complete understanding of the germline changes in PanC is essential to identify potential susceptibility. In the present study, we aimed to identify novel susceptibility loci in NRF2 pathway genes for their associations with PanC risk by using a pathway-based approach to leverage the published PanC GWAS datasets.

2 | MATERIALS AND METHODS

2.1 | Study participants

Of the available GWAS datasets, there were 15 423 participants in the case-control study of PanC comprising 8477 cases and 6946 controls, who were from the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case-Control Association Study. We first used 4755 cases and 3446 controls from the PanScan GWAS dataset derived from three phases of 17 cohort and 11 case-control studies, including PanScan I, PanScan II, and PanScan III with 1760 cases and 1780 controls, 1457 cases and 1666 controls, and 1538 cases and 0 controls, respectively. PanScan II and PanScan III were merged into one dataset of PanScan II/III in the analysis due to lack of controls in PanScan III. We then used another Pancreatic Cancer Case-Control Association Study from the Pancreatic Cancer Case-Control Consortium (PanC4) that included individuals from the USA, Europe, and Australia (3722 cases and 3500 controls) (Figure S1). Details of cases (individuals with pancreatic ductal adenocarcinoma) and controls have been previously described.[11,20,21] Distributions of demographic characteristics between pancreatic cancer cases and controls are shown in Table S1. Both the PanScan and PanC4 GWAS datasets are available through dbGAP with permission from NCI of NIH to use the datasets with the accession numbers of phs000206.v5.p3 and phs000648.v1.p1, respectively.

Each study obtained written informed consent from study participants and approval from its Institutional Review Board (IRB) including IRB certification permitting data-sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH Supported or

Conducted GWAS. The present study protocol, which had been laid down in accordance with the ethical standards in the 1964 Declaration of Helsinki and its later amendments, was approved by the Duke University Health System Institutional Review Board and strictly followed.

2.2 | Gene selection, genotyping, and imputation

The term "*NRF2* pathway" was searched in GeneCards: The Human Gene Database (https://www.genecards.org/). Overall, 164 genes located on autosomal chromosomes were selected (details are presented in Table S2). GWAS genotyping was done in accordance with the approach described previously.[8]

Genotyping data for SNP located in these genes and their ± 500-kb flanking regions were extracted for imputation with IMPUTE2 software. For quality control, variants were excluded if: (i) completion rate <98%; (ii) minor allele frequency (MAF) <0.01; (iii) Hardy-Weinberg proportion $P < 1 \times 10^{-5}$; and (iv) low quality imputation score (IMPUTE 2 INFO score <0.8). The imputed SNP with an information score >0.8 were qualified for further analysis. After quality control, there were 33 566 SNP, 32 450 SNP, and 33 170 SNP within 5.0 kb up- and downstream of *NRF2* signaling pathway genes for PanScan I, PanScan II/III, and panC4, respectively. The final meta-analysis contained 31 583 SNP that met the inclusion criteria for all three studies.

2.3 | Association analysis

In the single-locus analysis, we used a logistic regression model with adjustment for age, gender, and the top five principal components, which were selected from unconditional logistic regression analysis with the top 20 principal components from all three studies (Table S3). An OR and its 95% CI were estimated by unconditional logistic regression analysis with PLINK 1.9[22] with a score test for the log additive genetic effect. A meta-analysis was further used on the results of a log-additive model of 31 583 SNP using the fixed-effects inverse-variance method based on β estimates and standard errors using Stata software (v.12; Stata Corp., College Station, TX, USA). Cochran's Q statistics and I^2 were used to assess heterogeneity (Q-test $P < 0.10$ or $I^2 > 50\%$).[23]

The FDR approach with a cut-off value of 0.15 was applied to control for multiple testing and to reduce the probability of false-positive findings.[24,25] The association between each SNP and PanC risk was evaluated with an additive genetic model.

The multivariable stepwise logistic regression model with a cut-off value of 0.05 was used to identify independent SNP. NUG of SNP with independent effects was used to assess the classification performance of the model. All individuals were also divided into a low-risk group (0-1 NUG) and a high-risk group (2-3 NUG) for additional analysis.

Moreover, Haploview v.4.2 was used to produce the Manhattan plot, and LocusZoom was used to construct the regional association plots by using the 1000 Genomes Project CEU dataset (phase I integrated release 3, March 2012), which were the same as previously described.[8] Linear regression analysis was applied to analyze correlations between SNP and the corresponding gene expression levels. All statistical analyses were carried out with SAS software (version 9.4; SAS Institute, Cary, NC, USA) if not otherwise specified.

2.4 | In silico functional prediction and validation

We used four in silico tools of F-SNP (http://compbio.cs.queensu.ca/F-SNP),[26] SNPinfo (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm),[27] RegulomeDB (http://regulomedb.org/),[28] and HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php)[29] to predict potential functions of the significant SNP. We carried out an eQTL analysis to estimate the associations between SNP and mRNA expression levels of the corresponding gene by using mRNA expression data from the lymphoblastoid cell lines of 373 Europeans available in the 1000 Genomes Project[30] and 127 tumor tissues in TCGA[31] as well as the eQTL results from the Genotype-Tissue Expression (GTEx) project.[32] In addition, we also compared the mRNA expression levels of targeted genes between tumor and adjacent normal tissues available in the Oncomine database (https://www.oncomine.org/).[33] Possible allelic effects of these variants on TF-binding motifs were determined using PrEdict Regulatory Functional Effect of SNP by Approximate *P*-value Estimation (PERFECTOS-APE; http://opera.autosome.ru/perfectosape/), which determines the probability of a TF motif (using position weight matrices, from HOCOMOCO-10, JASPAR, HTSELEX, SwissRegulon, and HOMER databases) in the DNA sequence overlapping each variant. Fold change in the probability of a TF binding site present for each allele of a variant was then calculated.[34]

3 | RESULTS

3.1 | Single-locus analysis

Workflow of the present analysis is shown in Figure 1. First, we carried out the single-locus analysis to estimate the associations between selected SNP (MAF ≥0.01) and PanC risk for each of the three European-ancestry populations by using logistic regression analysis with adjustment for the five principal components. Of those SNP included for PanScan I, PanScan II/III and PanC4, we identified 1673, 2193 and 1415 SNP with a nominal $P < 0.05$, respectively (Figure S2). In the subsequent meta-analysis of the three populations, 1073 SNP remained associated with PanC risk at $P < 0.05$ in an additive genetic model, of which 13 SNP on *SLC2A6*, *PDGFB*, *SLC5A3*, *SLC2A13* and *MAPK8* passed multiple testing corrections with an FDR <0.15 (Figure 2A; Table 1). Although some SNP in some chromosome regions (ie, 22q13.1-*PGDFB* and 10q11.22-*MAPK8*) have been reported by GWAS or pathway-based analyses,[8,35] three SNP (ie, *SLC2A6* rs3124761 at 9q34, *SLC2A13* rs17458086 at 12q12, and *SLC5A3* rs1630747 at 21q22.11) are novel findings (Figure 2B-D), for which we carried out additional in silico analysis for their functional relevance. The results of these three SNP in each of the GWAS datasets

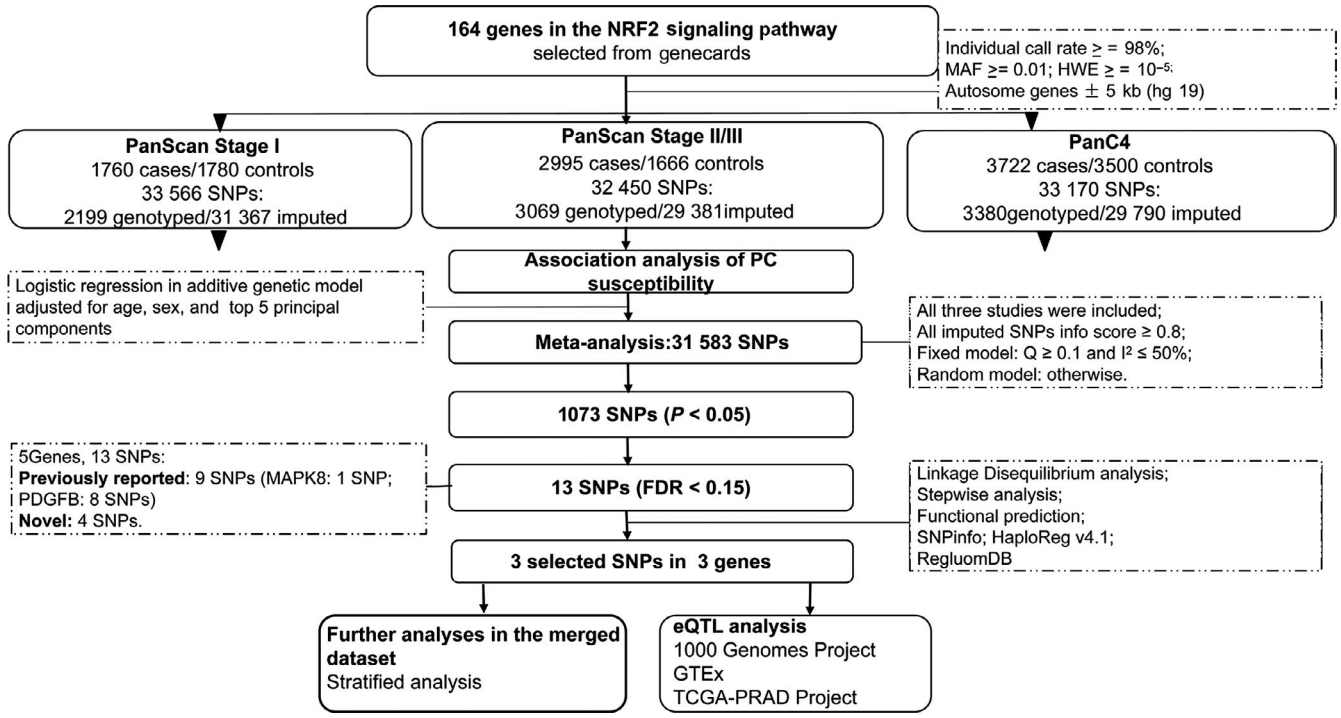

FIGURE 1 Flowchart of the present study. eQTL, expression quantitative trait loci; FDR, false discovery rate; GTEx, Genotype-Tissue Expression; HWE, Hardy-Weinberg equilibrium; MAF, minor allele frequency; NRF2, nuclear factor erythroid2-related factor2; PC, pancreatic cancer; SNP, single nucleotide polymorphism; TCGA-PRAD, The Cancer Genome Atlas Prostate Adenocarcinoma

and the final meta-analysis are summarized in Figure S3. All three SNP were from imputation and showed a relatively low heterogeneity among the three GWAS datasets (all Q-test $P > 0.1$ and $I^2 < 25.0\%$).

3.2 | Genotype effect and the joint-effect of the three significant SNP

We then assessed the independent genotype effects and their joint effects of the three identified SNP in the presence of age, gender and the top five principal components in a multivariate stepwise logistic regression model. All three SNP remained significantly and independently associated with PanC risk (Table S4). Specifically, the genotypes of *SLC2A6* rs3124761 C/T, *SLC2A13* rs17458086 T/C, and *SLC5A3* rs1630747 A/C were significantly associated with PanC risk in the additive models ($P < 0.0001$, $P = 0.004$, and $P = 0.002$, respectively, Table 2). In dominant models, both the rs3124761 T allele and rs17458086 C allele carriers were at increased risk of PanC (OR = 1.19, 95% CI = 1.11-1.27, and $P < 0.0001$; OR = 1.36, 95% CI = 1.14-1.64, and $P < 0.001$), whereas the rs1630747 C allele was associated with reduced risk (OR = 0.89, 95% CI = 0.84-0.96, $P = 0.001$), compared with their corresponding wild-type alleles (Table 2). In recessive models, the rs3124761 TT genotype carriers were at increased risk of PanC (OR = 1.28, 95% CI = 1.04-1.57, and $P = 0.0191$), whereas the rs1630747 CC genotype was associated with reduced risk (OR = 0.88, 95% CI = 0.77-0.99, $P = 0.0433$) compared with their corresponding wild-type genotypes (Table 2).

We then combined the risk genotypes of rs3124761 CT+TT, rs17458086 TC+CC, and rs1630747 AA into a single genetic score NUG. The trend test indicated a significant association between an increased NUG and an increased risk of PanC ($P < 0.0001$, Table 3). We also divided all individuals into a low-risk group (0-1 NUG) and a high-risk group (2-3 NUG) and found that PanC risk in high-risk individuals was greater than that among the low-risk group (OR = 1.26, 95% CI = 1.16-1.37, $P < 0.0001$, Table 3). As the difference in the distribution of age existed in each dataset (Table S1) and age is a known risk factor for PanC, we carried out subgroup analysis by age group (ie, <60, 60-70 and >70 years) and gender. It was found that the risk associated with high-risk NUG was most evident in the <60 years group (OR = 1.34, 95% CI = 1.15-1.56, $P = 0.0002$). There was no obvious difference between males and females. Additionally, there was no evidence for an interaction among and between these strata ($P > 0.05$, Table S5).

3.3 | Genotype and phenotype correlation analysis

Finally, we carried out in silico prediction for potential effects of the three SNP on mRNA expression levels preliminarily through the online tools (Table S6). All three SNP are located in intronic regions, but are also located in the enhancer region of histone H3 mono methyl K4 (H3k4me1) that marks active/poised enhancers and/or in binding sites for DNase or transcription factors (Figure 3A,F, Figure S4A).

We then assessed the effects of the three variants on the predicted TF-binding sites. It is notable that the three SNP are predicted to alter the ability to bind with some motifs, among which are rs3124761 T, rs17458086 C and rs1630747 C alleles that are predicted to disrupt the TF-binding motifs for the RARG, BCL2, and

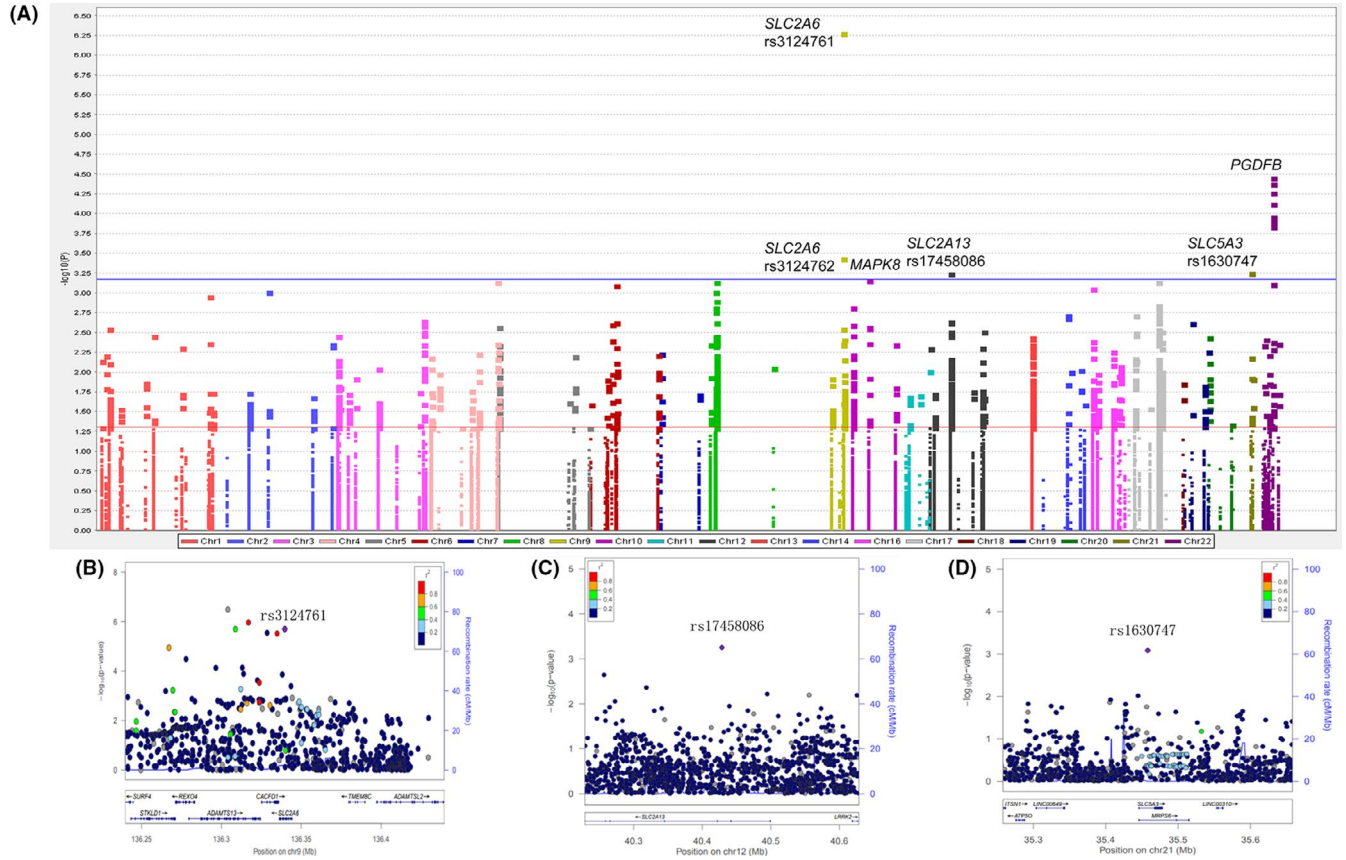

FIGURE 2 Screening for pancreatic cancer (PanC) risk-associated single nucleotide polymorphisms (SNP). A, Manhattan plot of the association results of 31 583 SNP in 164 nuclear factor erythroid2-related factor2 (NRF2) signaling pathway genes and PanC risk in the meta-analysis of three genome-wide association study (GWAS) datasets. Red horizontal line indicates $P = 0.05$ and blue horizontal line indicates false discovery rate (FDR) = 0.15. B-D, Each panel shows the regional association results for the meta-analysis of PanScan I, PanScan II + III, and PanC4 (purple diamonds). Also shown are results for chromosomes 9q34 (B), 12q12 (C), and 21q22.11 (D)

ZN784 proteins, respectively (Figure 3B,G, Figure S4B). This suggests that RARG, BCL2, and ZN784 binding motifs can be altered by rs3124761, rs17458086, and rs1630747, respectively, and thus change *SLC2A6*, *SLC2A13*, and *SLC5A3* mRNA expression levels.

To substantiate the associations between the identified SNP and PanC risk, we evaluated correlations between SNP and mRNA expression levels of the corresponding genes in normal lymphoblastoid cell lines from 373 Europeans available from the 1000 Genomes Project. By using Student's *t* test or linear regression analysis of the logarithm transformed expression values (log2), we showed that the rs3124761 T allele (risk) was correlated with increased mRNA expression levels of *SLC2A6* in either additive or dominant models ($P = 0.0357$ and 0.0129, respectively, Figure 3C,D). We also found that the rs17458086 C (risk) allele was significantly correlated with higher mRNA expression levels of *SLC2A13* compared with the T (protective) allele ($P = 0.026$ in both the additive model and the dominant model due to lack of CC homozygote (Figure 3H,I). eQTL results for SNP rs1630747 were not significant in these lymphoblastoid cell lines in either of the genetic models (Figure S4C,D).

We attempted to use data from the Genotype-Tissue Expression Project (GTEx) database (http://www.gtexportal.org/home/) and the data from 127 Europeans in the TCGA-PDAC Project to query the eQTL results and assessed the correlations. Imputation of the genotype of the three SNP based on the current quality control of these databases was not successful. In addition, we assessed the differences in mRNA expression levels of *SLC2A6*, *SLC2A13*, and *SLC5A3* between adjacent normal pancreatic tissues and pancreatic tumor tissues from the Oncomine database. We found that compared with the expression in normal pancreatic tissues, both *SLC2A6* and *SLC2A13* mRNA levels in tumor tissues were significantly increased ($P = 5.43 \times 10^{-4}$ and $P = 0.027$) (Figure 3E,J), whereas no statistical difference in *SLC5A3* mRNA expression levels was found ($P = 0.954$) (Figure S4E).

4 | DISCUSSION

In the present study of the NRF2-pathway-based approach analysis of published GWAS datasets, we identified 13 loci to be associated with PanC risk, including three novel loci (ie, *SLC5A3* rs1630747, *SLC2A13* rs17458086 and *SLC2A6* rs3124761), which merit additional follow-up investigations for their functional mechanisms underlying the observed associations.

The first genetic variant rs3124761 to be addressed is located in the intron of *SLC2A6* on Chr9q34. The rs3124761 T allele was

TABLE 1 Associations between 13 SNP in the NRF2 signaling pathway and PanC risk with FDR <0.15

SNP rs#	Locus	Position	Encode gene	Location	Allele[a]	EAF1[b]	EAF2[b]	EAF3[b]	OR (95% CI)[c]	*P*-value	FDR	Reason for exclusion[d]
rs3124761	9q34.2	136339755	SLC2A6	Intron	C/T	0.147	0.138	0.154	1.17 (1.10-1.25)	5.17×10^{-7}	0.0002	–
rs1800818	22q13.1	39640703	PGDFB	Intron	T/C	0.371	0.364	0.370	1.10 (1.05-1.16)	3.48×10^{-5}	0.011	2
rs5757575	22q13.1	39636928	PGDFB	Intron	G/A	0.362	0.356	0.363	1.10 (1.05-1.16)	4.13×10^{-5}	0.013	2
rs3985946	22q13.1	39635858	PGDFB	Intron	GCA/G	0.367	0.361	0.368	1.10 (1.05-1.16)	5.34×10^{-5}	0.017	2
rs5757572	22q13.1	39632920	PGDFB	Intron	G/C	0.367	0.362	0.368	1.10 (1.05-1.16)	7.35×10^{-5}	0.023	2
rs5757576	22q13.1	39636930	PGDFB	Intron	G/A	0.366	0.362	0.368	1.10 (1.04-1.15)	1.11×10^{-4}	0.033	2
rs6001512	22q13.1	39632523	PGDFB	Intron	G/A	0.067	0.074	0.073	1.18 (1.08-1.29)	1.29×10^{-4}	0.036	1
rs71319025	22q13.1	39634444	PGDFB	Intron	C/A	0.067	0.074	0.073	1.18 (1.08-1.29)	1.29×10^{-4}	0.036	1
rs56180415	22q13.1	39631963	PGDFB	Intron	G/T	0.067	0.074	0.073	1.18 (1.08-1.29)	1.44×10^{-4}	0.039	1
rs3124762	9q34.2	136338187	SLC2A6	Intron	A/C	0.164	0.160	0.157	0.89 (0.83-0.95)	3.62×10^{-4}	0.088	–
rs1630747	21q22.11	35457991	SLC5A3	Intron	A/C	0.271	0.267	0.264	0.91 (0.86-0.96)	5.52×10^{-4}	0.127	–
rs17458086	12q12	40428639	SLC2A13	Intron	T/C	0.014	0.010	0.016	1.38 (1.14-1.68)	5.61×10^{-4}	0.127	–
rs138947508	10q11.22	49566274	MAPK8	Intron	A/G	0.024	0.023	0.020	0.75 (0.63-0.89)	6.76×10^{-4}	0.147	2

–, SNP included; CI, confidence interval; EAF, effect allele frequency; FDR, false discovery rate; NRF2, nuclear factor erythroid2-related factor2; OR, odds ratio; PanC, pancreatic cancer; SNP, single nucleotide polymorphism.

[a]Referring to "reference allele/effect allele."

[b]EAF1 was EAF in PanScan I controls; EAF2 was EAF in PanScan II/III controls; EAF3 was EAF in PanC4 controls.

[c]Obtained from the meta-analysis of the three studies. Fixed effects models were used when no heterogeneity was found between studies (Q test $P > 0.10$ or $I^2 < 25.0\%$); otherwise, random effects models were used.

[d]1. Genetic variant was previously reported to be associated with PanC risk or 2. Existing linkage disequilibrium ($r^2 > 0.1$, $D' = 1$) with previous reported SNP associated with PanC risk according to data obtained from PubMed.

TABLE 2 Analysis of associations between pancreatic cancer risk and the three SNP in the merged dataset of PanScan and PanC4 studies

SNP rs# and genetic model	Group				
	Genotype	Case (%)	Control (%)	OR (95% CI)[a]	*P*-value
rs3124761					
Additive	CC	5851 (68.61)	5168 (72.59)	1.00	–
	CT	2435 (28.55)	1798 (25.26)	1.17 (1.09-1.26)	<0.0001
	TT	242 (2.84)	153 (2.15)	1.34 (1.09-1.65)	0.0058
	Trend test				<0.0001
Dominant	CT+TT	2677 (31.39)	1951 (27.41)	1.19 (1.11-1.27)	<0.0001
Recessive	CC+CT	8286 (97.16)	6966 (97.85)	1.00	–
	TT	242 (2.84)	153 (2.15)	1.28 (1.04-1.57)	0.0191
Rs17458086[b]					
Additive	TT	8167 (96.38)	6752 (97.24)	1.00	–
	TC	304 (3.59)	190 (2.73)	1.37 (1.14-1.65)	<0.001
	CC	3 (0.03)	2 (0.03)	1.29 (0.21-7.75)	0.785
	Trend test				**0.004**
Dominant	TC+CC	307 (3.62)	192 (2.76)	1.36 (1.14-1.64)	<0.001
rs1630747					
Additive	AA	4757 (56.14)	3725 (53.64)	1.00	–
	AC	3182 (37.55)	2732 (39.35)	0.91 (0.85-0.97)	0.004
	CC	535 (6.31)	487 (7.01)	0.85 (0.75-0.97)	0.018
	Trend test				**0.002**
Dominant	AC+CC	3717 (43.86)	3219 (46.36)	0.89 (0.84-0.96)	0.001
Recessive	AA+AC	7939 (93.69)	6457 (92.99)	1.00	–
	CC	535 (6.31)	487 (7.01)	0.88 (0.77-0.99)	0.0433

–, reference; CI, confidence interval; OR, odds ratio; SNP, single nucleotide polymorphism.
[a]Obtained from logistic regression models with adjustment for age, gender, and the top 5 significant principal components. [b]Recessive model was not shown as a result of small size of homozygote CC.

TABLE 3 Associations between the combined genetic score (NUG) and risk of pancreatic cancer

NUG	Group		OR (95% CI)[b]	*P*-value
	Case (%)	Control (%)		
0	2479 (29.08)	2327 (32.69)	1.00	–
1	4384 (51.42)	3652 (51.30)	0.89 (0.83-0.96)	0.0013
2	1605 (18.82)	1110 (15.59)	1.19 (1.09-1.31)	$<10^{-4}$
3	58 (0.68)	30 (0.42)	1.61 (1.103-2.51)	0.0361
Trend test				$<10^{-4}$
Dichotomized				
0-1	6863 (80.49)	5797 (83.99)	1.00	–
2-3	1163 (19.51)	1140 (16.01)	1.26 (1.16-1.37)	$<10^{-4}$

–, reference; CI, confidence interval; NUG, number of unfavorable genotypes; OR, odds ratio.
[a]Risk genotypes were rs3124761 CT+TT, rs17458086 TC+CC, and rs1630747 AA. [b]Adjusted for age, gender, and the top 5 significant principal components.

identified to be associated with an increased PanC risk compared with the corresponding wild-type C allele. The base change from C to T is predicted to reduce the DNA-binding ability with the RARG motif. Therefore, SLC2A6 expression should be activated when the RARG motif with a suppressing role loses its binding ability to the C locus. eQTL data from 373 lymphocyte cells further highlighted that the substitution of rs3124761 C to T significantly enhanced mRNA expression of *SLC2A6*. The 9q34 region is one of the most frequently altered regions in human cancers.[10,20] A series of SNP located in the *ABO* gene tagged by rs505922 and rs630014 in this

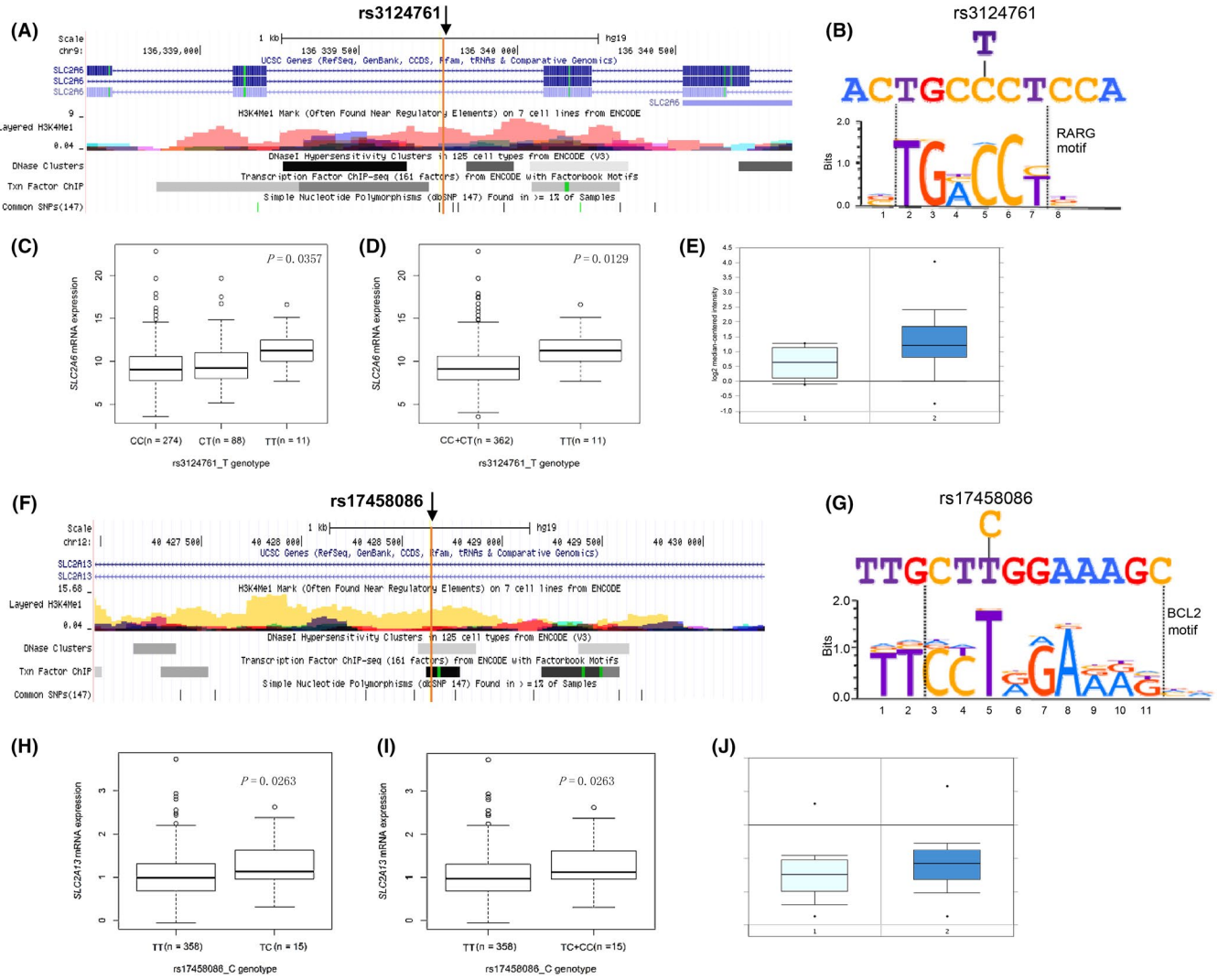

FIGURE 3 Functional analysis of *SLC2A6* rs3124761 and *SLC2A13* rs17458086. A,F, rs3124761 at 9q34 is shown as well as overlapping RefSeq genes on chr9: 136,338,500-136,341,000 (NCBI GRCh37/Hg19) (A). rs17458086 at 12q12 is shown as well as overlapping RefSeq genes on chr12: 40,427,000-40,430,500 (NCBI GRCh37/Hg19) (F). ENCODE data for histone modification marks (H3K4me1) are indicated by colored density plots. DNase clusters and binding of transcription factors (TF ChIP) are indicated by horizontal bars. Numbers next to each bar indicate the number of different transcription factors bound across all tested cell lines. The panel is adapted from the UCSC Genome Browser. B, Analysis of the effects of rs3124761 on transcriptional factor motifs: The rs3124761 risk allele T may alter a predicted DNA-binding motif for RARG. C,D, Expression quantitative trait loci (eQTL) analyses of rs3124761 in 373 Europeans from the 1000 Genomes Project: additive model, $P = 0.0357$ (C); dominant model, $P = 0.0129$ (D). E, mRNA expression of *SLC2A6* in normal tissues (1, n = 16) and tumor tissues (2, n = 26): $P = 5.43 \times 10^{-4}$. G, Analysis of the effects of the rs17458086 on transcriptional factor motifs: The rs17458086 risk allele C may alter a predicted DNA-binding motif for BCL2. H,I, eQTL analyses of rs17458086 in 373 Europeans from the 1000 Genomes Project: additive model, $P = 0.026$; dominant model, $P = 0.026$. J, mRNA expression of *SLC2A13* in normal tissues (1, n = 39) and tumor tissues (2, n = 39): $P = 0.027$

region have been reported by a GWAS to be related to PanC susceptibility.[20] Another recently published associated SNP located in the *ABO* gene was rs687289, which was identified by GTEx functional prediction.[21] However, our linkage disequilibrium (LD) analysis showed that *SLC2A6* rs3124761 was in low LD with any of the three *ABO* SNP previously reported.

The second genetic variant rs17458086 is located in the intron of *SLC2A13* on Chr12q12. We found, for the first time, that carriers of the *SLC2A13* rs17458086 C allele had a 1.38-fold increased PanC risk compared with TT carriers. The change from T allele to C allele is predicted to reduce the DNA-binding ability with the BCL2 motif. Therefore, *SLC2A13* expression should be increased when the BCL2 motif, which is predicted to suppress expression of the gene, binds to *SLC2A13* rs17458086 C instead of the T allele. eQTL data further demonstrated that the substitution of rs17458086 T to C significantly enhanced the mRNA expression of *SLC2A13*. In addition, we noticed that the frequency of the rare allele C of *SLC2A13* rs17458086 was just above the level of the inclusive criteria (1.0%-1.6%), which reinforces that the functions of this class of subpolymorphic risk alleles (ie, those with rare risk allele frequencies

<1%) have not been thoroughly investigated.[36] Both *SLC2A6* and *SLC2A13* belong to SLC family members and encode GLUT6 and GLUT13, respectively. GLUT6, a hexose transporter in liposomes, is expressed predominantly in the brain, spleen and peripheral leukocytes, whereas GLUT13 is an H^+/myoinositol cotransporter that is stimulated by a decrease in the extracellular pH is expressed primarily in the brain.[37] SLC proteins, which are primarily involved in the uptake of small molecules into cells, belong to a superfamily of transporters.

There are 395 membrane-spanning *SLC* transporters that are organized into 52 families in humans. More than 80 SLC transporters have been implicated in monogenic disorders, and many genetic variants in the SLC transporter genes associated with common diseases have been identified through genotype analysis of candidate genes, or from GWAS.[38] The 14 human GLUT proteins, encoded by the *SLC2* gene, have various substrate specificities and are involved in the transport of several hexoses in addition to myoinositol.[39] GLUT proteins in the digestive system serve as important mediators in maintaining normal functions, including the absorption of nutrients and ions, excretion of bile acids, and metabolism of toxins.[40-42] Dysregulation of the *SLC2* gene is likely to be associated with carcinogenesis, tumor progression, metastasis, and chemoresistance. Changes in expression and regulation of the GLUT family proteins, including GLUT13/HMIT, were observed in neoplasms of the digestive system and in breast cancer cells.[43,44] Thus far, there are no detailed reports of biological functions of SLC2A6 in human cells. As for SLC2A13, it is an H^+/myoinositol symporter, whereas all of the other members of the GLUT family are facilitative transporters. There are neither reported data about glucose transport activity for GLUT13/HMIT, nor any available information about the expression of the facilitative glucose transporter family in cancer.[37] Genetic alterations of *SLC2A13* were observed to be prevalent in the early stage of lung cancer in Serbians, suggesting that structural changes of *SLC2A13* could play a role in the development of non-small cell lung cancer.[45]

Additionally, several SNP in *SLC2A13* were found to be closely related to Parkinson's disease independently or dependently at a genome-wide significance level.[46,47] *SLC5A3* encodes a sodium myoinositol transporter involved in the response to hyperosmotic stress, and the sequence of *SLC5A3* lies completely within that of *MRPS6* that encodes a subunit of the mitochondrial ribosome.[48] *SLC5A3/MRPS6* rs9982601 was previously identified to have a genome-wide association with early-onset myocardial infarction.[49] This SNP was also found to be associated with mRNA expression of *MRPS6* in blood, with the risk C allele correlated with an increased mRNA expression.[48] We observed, for the first time, that the *SLC5A3* rs1630747 C allele was associated with a reduced PanC risk compared with the corresponding wild-type allele. However, the eQTL data did not show any difference in mRNA expression levels between the two alleles, and we did not find any other available dataset from either GTXe or TCGA databases for further analysis. Therefore, the functional relevance of this SNP remains unknown.

However, we failed to identify any variants in the *NRF2* gene to be associated with PanC risk. All three genes identified in the present study were members of the *SLC* family, and their encoded proteins were reported to be responsible for transporting hexose, H^+/myoinositol, or sodium myoinositol. All three identified variants are located in the intron regions of these genes instead of in the promoter sites coupled with antioxidant response element for NRF2 binding. These regions were found with a potential enhancer region with histone modification marker (H3k4me1) and/or the predicted binding sites for DNase or transcription factors. Although we do not have direct evidence to show that NRF2 regulates mRNA expression of these genes, it is known that *NRF2* regulates the expression of *BCL2*,[50] and that retinoic acid inhibits NRF2 expression.[51] Therefore, it is possible that NRF2-BCL2 interaction and RARG-NRF2 interaction may indirectly modify the PanC risk. Whether and how NRF2 indirectly affects the binding of the motifs of RARG, BCL2 and ZN84 proteins warrant additional mechanistic studies.

Nevertheless, there are some limitations in the present study. First, there were no control data in the PanScan III GWAS which might have caused the merged PanScan II/III GWAS datasets to be somewhat heterogeneous. Second, we had no access to other risk factors, such as family history, smoking history, alcohol-drinking history as well as other clinical data in publicly available datasets, which could have biased PanC risk assessment without adequate adjustment for these covariates in the risk evaluation model. Finally, our analysis was limited to evaluate whether a particular SNP had biological functions only by using the available online tools and the in silico-based eQTL. Additional mechanistic investigations are warranted to provide direct functional evidence to support our findings.

ACKNOWLEDGMENTS

The authors would like to thank Jie Ge, Ning Gu and Danwen Qian for their technical support and insightful discussions. Qingyi Wei was partly supported by the Duke Cancer Institute as part of the P30 Cancer Center Support Grant (Grant ID: NIH CA014236) and by a start-up fund from Duke Cancer Institute, Duke University Medical Center. Xiaoxin Chen was partly supported by the RCMI Center for Health Disparities Research at The North Carolina Central University (Grant ID: NIH MD012392). Wenjun Yang was partly supported by West China First-Class Discipline Construction Project in Basic Medicine funded by Ningxia Medical University. The authors would like to thank all the participants of the PanScan GWAS Study and Pancreatic Cancer Case Control Association Study and acknowledge dbGaP repository for providing cancer genotyping datasets described as follows.

PanScan: The PanScan project was funded in whole or in part with federal funds from the NCI, US NIH under contract number HHSN261200800001E. Additional support was received from NIH/NCI K07 CA140790, the American Society of Clinical Oncology Conquer Cancer Foundation, the Howard Hughes Medical Institute, the Lustgarten Foundation, the Robert T. and Judith B. Hale Fund for Pancreatic Cancer Research and Promises for Purple. A full list

of acknowledgments for each participating study is provided in the Supplementary Note of the manuscript with PubMed ID: 25086665. The dbGaP accession number for this study used in this article is phs000206.v5.p3.

PanC4: The patients and controls for this study were derived from the following PANC4 studies: Johns Hopkins National Familial Pancreas Tumor Registry, Mayo Clinic Biospecimen Resource for Pancreas Research, Ontario Pancreas Cancer Study (OPCS), Yale University, MD Anderson Case Control Study, Queensland Pancreatic Cancer Study, University of California San Francisco Molecular Epidemiology of Pancreatic Cancer Study, International Agency of Cancer Research and Memorial Sloan Kettering Cancer Center. This work is supported by NCI R01CA154823. Genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN2682011000111. The dbGaP accession number for this study used in this article is phs000648.v1.p1.

TCGA: The results published herein are in whole or part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and National Human Genome Research Institute (NHGRI). Information about TCGA and the investigators and institutions who constitute TCGA Research Network can be found at "http://cancergenome.nih.gov". The TCGA SNP data analyzed here are requested through dbGAP (accession#: phs000178.v1.p1).

CONFLICTS OF INTEREST

Authors declare no conflicts of interest for this article.

ORCID

Qingyi Wei https://orcid.org/0000-0002-3845-9445

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68:7-30.
2. National Cancer Institute. NIH Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Pancreatic Cancer; 2018. https://seer.cancer.gov/statfacts/html/pancreas.html
3. Luo G, Zhang Y, Guo P, Ji H, Xiao Y, Li K. Global patterns and trends in pancreatic cancer incidence: age, period, and birth cohort analysis. *Pancreas*. 2019;48(2):199-208.
4. Hidalgo M. Pancreatic cancer. *N Engl J Med*. 2010;362:1605-1617.
5. Zhou B, Xu JW, Cheng YG, et al. Early detection of pancreatic cancer: where are we now and where are we going? *Int J Cancer*. 2017;141:231-241.
6. Barone E, Corrado A, Gemignani F, Landi S. Environmental risk factors for pancreatic cancer: an update. *Arch Toxicol*. 2016;90:2617-2642.
7. Grant RC, Denroche RE, Borgida A, et al. Exome-wide association study of pancreatic cancer risk. *Gastroenterology*. 2018;154:719-722.
8. Duan B, Hu J, Liu H, et al. Genetic variants in the platelet-derived growth factor subunit B gene associated with pancreatic cancer risk. *Int J Cancer*. 2018;142:1322-1331.
9. Klein AP, Wolpin BM, Risch HA, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun*. 2018;9:556.
10. Amundadottir LT. Pancreatic cancer genetics. *Int J Biol Sci*. 2016;12:314-325.
11. Zhang M, Wang Z, Obazee O, et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget*. 2016;7:66328-66343.
12. Lu M, Ji J, Jiang Z, You Q. The Keap1-Nrf2-ARE pathway as a potential preventive and therapeutic target: an update. *Med Res Rev*. 2016;36:924-963.
13. Leinonen HM, Kansanen E, Pölönen P, Heinäniemi M, Levonen AL. Dysregulation of the Keap1-Nrf2 pathway in cancer. *Biochem Soc Trans*. 2015;43:645-649.
14. Todoric J, Antonucci L, Di Caro G, et al. Stress-activated NRF2-MDM2 cascade controls neoplastic progression in pancreas. *Cancer Cell*. 2017;32:824-839.
15. Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends Biochem Sci*. 2014;39:199-218.
16. DeNicola GM, Karreth FA, Humpton TJ, et al. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. *Nature*. 2011;475:106-109.
17. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502:333-339.
18. Cho HY, Marzec J, Kleeberger SR. Functional polymorphisms in Nrf2: implications for human disease. *Free Radic Biol Med*. 2015;88:362-372.
19. Hartikainen JM, Tengström M, Winqvist R, et al. KEAP1 genetic polymorphisms associate with breast cancer risk and survival outcomes. *Clin Cancer Res*. 2015;21:1591-1601.
20. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet*. 2009;41:986-990.
21. Wolpin BM, Rizzato C, Kraft P, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet*. 2014;46:994-1000.
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.
23. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560.
24. Millstein J, Volfson D. Computationally efficient permutation-based confidence interval estimation for tail-area FDR. *Front Genet*. 2013;4:179.
25. Hafner M, Niepel M, Sorger PK. Alternative drug sensitivity metrics improve preclinical cancer pharmacogenomics. *Nat Biotechnol*. 2017;35:500-502.
26. Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res*. 2008;36:820-824.
27. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37:600-605.
28. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790-1797.
29. 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;40:930-934.

30. Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501:506-511.
31. Network CGAR. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543-550.
32. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648-660.
33. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6:1-6.
34. Kulakovskiy IV, Vorontsov IE, Makeev V. PERFECTOS-APE-predicting regulatory functional effect of SNPs by approximate P-value estimation, vol. 1 Lisbon, Portugal: Conference: 6th International Conference on Bioinformatics Models, Methods and Algorithms, Bioinformatics. 2015: 102-108.
35. Li D, Duell EJ, Yu K, et al. Pathway analysis of genome-wide association study data highlights pancreatic development genes as susceptibility factors for pancreatic cancer. *Carcinogenesis*. 2012;33:1384-1390.
36. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017;17:692-704.
37. Barron CC, Bilan PJ, Tsakiridis T, Tsiani E. Facilitative glucose transporters: implications for cancer detection, prognosis and treatment. *Metabolism*. 2016;65:124-139.
38. Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov*. 2015;14:543-560.
39. Mueckler M, Thorens B. The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med*. 2013;34:121-138.
40. Chan K, Busque SM, Sailer M, et al. Loss of function mutation of the Slc38a3 glutamine transporter reveals its critical role for amino acid metabolism in the liver, brain, and kidney. *Pflugers Arch*. 2016;468:213-227.
41. Gulec S, Anderson GJ, Collins JF. Mechanistic and regulatory aspects of intestinal iron absorption. *Am J Physiol Gastrointest Liver Physiol*. 2014;307:397-409.
42. Suga T, Yamaguchi H, Sato T, Maekawa M, Goto J, Mano N. Preference of conjugated bile acids over unconjugated bile acids as substrates for OATP1B1 and OATP1B3. *PLoS ONE*. 2017;12:e0169719.
43. Xie J, Zhu XY, Liu LM, Meng ZQ. Solute carrier transporters: potential targets for digestive system neoplasms. *Cancer Manag Res*. 2018;10:153-166.
44. Lee DG, Lee JH, Choi BK, et al. H+-myo-Inositol transporter SLC2A13 as a potential marker for cancer stem cells in an oral squamous cell carcinoma. *Curr Cancer Drug Targets*. 2011;11:966-975.
45. Bankovic J, Stojsic J, Jovanovic D, et al. Identification of genes associated with non-small-cell lung cancer promotion and progression. *Lung Cancer*. 2010;67:151-159.
46. Satake W, Nakabayashi Y, Mizuta I, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet*. 2009;41:1303-1307.
47. Gao J, Nalls MA, Shi M, et al. An exploratory analysis on gene-environment interactions for Parkinson disease. *Neurobiol Aging*. 2012;33:2528.
48. Beaney KE, Smith AJP, Folkersen L, et al. Functional analysis of the coronary heart disease risk locus on chromosome 21q22. *Dis Markers*. 2017;2017:1096916.
49. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet*. 2009;41:334-341.
50. Niture SK, Jaiswal AK. Nrf2 protein up-regulates antiapoptotic protein Bcl-2 and prevents cellular apoptosis. *J Biol Chem*. 2012;287(13):9873-9886.
51. Wang XJ, Hayes JD, Henderson CJ, Wolf CR. Identification of retinoic acid as an inhibitor of transcription factor Nrf2 through activation of retinoic acid receptor alpha. *Proc Natl Acad Sci USA*. 2007;104(49):19589-19594.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Yang W, Liu H, Duan B, et al. Three novel genetic variants in NRF2 signaling pathway genes are associated with pancreatic cancer risk. *Cancer Sci*. 2019;110:2022-2032. https://doi.org/10.1111/cas.14017