



# TSPAN1, TMPRSS4, SDR16C5, and CTSE as Novel Panel for Pancreatic Cancer: A Bioinformatics Analysis and Experiments Validation

Hua Ye<sup>1</sup>, Tiandong Li<sup>1,2,3</sup>, Hua Wang<sup>1,3</sup>, Jinyu Wu<sup>1,3</sup>, Chuncheng Yi<sup>1,3</sup>, Jianxiang Shi<sup>3,4</sup>, Peng Wang<sup>1,3</sup>, Chunhua Song<sup>1,3</sup>, Liping Dai<sup>3,4</sup>, Guozhong Jiang<sup>5</sup>, Yuxin Huang<sup>6</sup>, Yongwei Yu<sup>7\*</sup> and Jitian Li<sup>2,3\*</sup>

<sup>1</sup> College of Public Health, Zhengzhou University, Zhengzhou, China, <sup>2</sup> Laboratory of Molecular Biology, Henan Luoyang Orthopedic Hospital (Henan Provincial Orthopedic Hospital), Zhengzhou, China, <sup>3</sup> Henan Key Laboratory of Tumor Epidemiology, Zhengzhou, China, <sup>4</sup> Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou, China, <sup>5</sup> Department of Pathology, First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, <sup>6</sup> Program in Public Health, University of California, Irvine, Irvine, CA, United States, <sup>7</sup> Department of Pathology, Second Military Medical University, Shanghai, China

## OPEN ACCESS

### Edited by:

Bilian Jin,  
Dalian Medical University, China

### Reviewed by:

Fan Xu,  
Nanchong Central Hospital, China  
Yanli Li,  
The Second Affiliated Hospital of  
Guangzhou Medical University, China

### \*Correspondence:

Jitian Li  
jitianlee@hotmail.com  
Yongwei Yu  
yongweiyu@126.com

### Specialty section:

This article was submitted to  
Cancer Immunity and Immunotherapy,  
a section of the journal  
Frontiers in Immunology

**Received:** 05 January 2021

**Accepted:** 23 February 2021

**Published:** 18 March 2021

### Citation:

Ye H, Li T, Wang H, Wu J, Yi C, Shi J,  
Wang P, Song C, Dai L, Jiang G,  
Huang Y, Yu Y and Li J (2021)  
TSPAN1, TMPRSS4, SDR16C5, and  
CTSE as Novel Panel for Pancreatic  
Cancer: A Bioinformatics Analysis and  
Experiments Validation.  
Front. Immunol. 12:649551.  
doi: 10.3389/fimmu.2021.649551

Pancreatic cancer is a lethal malignancy with a poor prognosis. This study aims to identify pancreatic cancer-related genes and develop a robust diagnostic model to detect this disease. Weighted gene co-expression network analysis (WGCNA) was used to determine potential hub genes for pancreatic cancer. Their mRNA and protein expression levels were validated through reverse transcription PCR (RT-PCR) and immunohistochemical (IHC). Diagnostic models were developed by eight machine learning algorithms and ten-fold cross-validation. Four hub genes (*TSPAN1*, *TMPRSS4*, *SDR16C5*, and *CTSE*) were identified based on bioinformatics. RT-PCR showed that the four hub genes were expressed at medium to high levels, IHC revealed that their protein expression levels were higher in pancreatic cancer tissues. For the panel of these four genes, eight models performed with 0.87–0.92 area under the curve value (AUC), 0.91–0.94 sensitivity, and 0.84–0.86 specificity in the validation cohort. In the external validation set, these models also showed good performance (0.86–0.98 AUC, 0.84–1.00 sensitivity, and 0.86–1.00 specificity). In conclusion, this study has identified four hub genes that might be closely related to pancreatic cancer: *TSPAN1*, *TMPRSS4*, *SDR16C5*, and *CTSE*. Four-gene panels might provide a theoretical basis for the diagnosis of pancreatic cancer.

**Keywords:** pancreatic cancer, WGCNA, diagnostic model, machine learning, bioinformatics, panel

## INTRODUCTION

Pancreatic cancer is the seventh leading cause of cancer-related deaths worldwide, and the mortality rate closely parallels the incidence (1). In recent years, deaths associated with pancreatic cancer are gradually increasing and it is predicted to be the second leading cause of cancer-related death by 2030 (2). In the United States, it is estimated that there will be approximately 56,770 new pancreatic cancer cases diagnosed, and 45,750 estimated deaths occurring among these new cases (3). From 2003 to 2015 statistics from China show that the age-standardized 5-year relative survival rate for

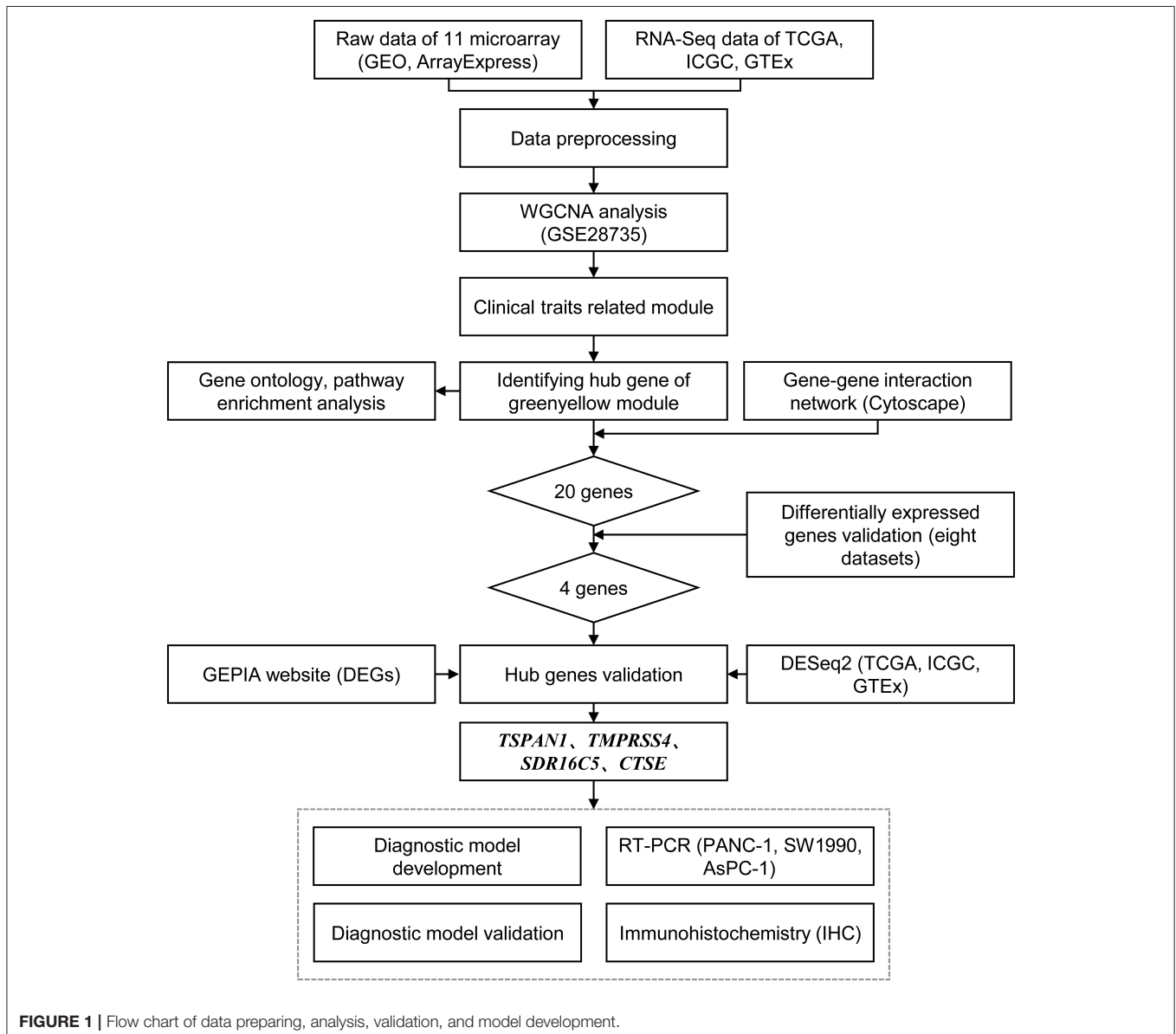
pancreatic cancer was only 7.2%(4). Despite advances in pancreatic cancer treatment strategies, the prognosis remains poor, largely due to the lack of early diagnostic approaches (5). Additionally, carbohydrate antigen 19-9 is widely used for the diagnosis of pancreatic cancer, but its sensitivity and specificity are only 0.80 (95% CI: 0.77-0.82) and 0.80 (95% CI: 0.77-0.82), respectively (6, 7). Therefore, the identification of new biomarkers or a panel with high specificity and sensitivity for diagnosing pancreatic cancer are important.

In recent years, with the development of microarray and high-throughput sequencing technologies, gene expression profiles

**Abbreviations:** WGCNA, Weighted gene co-expression network analysis; DEGs, differentially expressed genes; TOM, topological overlap matrix; GS, gene significance; MM, module membership; TCGA, the Cancer Genome Atlas; ICGC, International Cancer Genome Consortium; GTEx, Genotype-Tissue Expression; ROC, receiver operating characteristic; Se, Sensitivity; Sp, Specificity.

have become an effective source of biomarkers discovery. Weighted gene expression network analysis (WGCNA) has been widely used to reveal the phenotype-related genes by constructing scale-free gene co-expression networks, especially in cancers, including lung (8), bladder (9), breast (10), and pancreatic cancer (11). In developing prediction models, satisfying the sensitivity and specificity requirements are the most interesting and challenging tasks for tumor biomarker screening. Previous studies have shown that machine learning method can improve the accuracy of disease diagnosis or prognosis (12, 13), and cancer models with higher accuracy have been developed by applying those methods (14–17).

Therefore, this study was designed to explore novel biomarkers with high performance using bioinformatics. Potential genes, screened by bioinformatics, will be validated using RT-PCR and IHC experiments. Diagnostic models will



**FIGURE 1** | Flow chart of data preparing, analysis, validation, and model development.

be constructed using different machine learning methods and ten-fold cross-validation.

## MATERIALS AND METHODS

### Data Collection and Preprocessing

The study design is shown in **Figure 1**. A systematic search on two electronic databases (Gene Expression Omnibus and ArrayExpress) was performed for potential datasets before 1 June 2019. Datasets with a sample size  $>20$  were included. Eleven pancreatic cancer microarray datasets from three platforms were downloaded (Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Human Gene 1.0 ST Array, and Affymetrix Human Genome U219 Array). The raw data were pre-processed with the “oligo” package and the “affy” packages. The Robust Multichip Average (RMA) function was used for background correction and normalization. In this study, GSE28735 was used to construct a weighted gene co-expression network because

it contained the most balanced case and control samples, nine datasets (E-MEXP-2780, GSE15471, GSE16515, GSE32688, GSE71989, GSE106189, GSE62452, E-MTAB-6134, GSE62165) were combined to develop diagnostic models with a total of 818 samples, and the GSE32676 dataset with 32 samples was chosen to externally validate the model's performance. The ComBat algorithm was used to adjust the expression data from nine datasets for batch effects using the “sva” package (18). The characteristics of all microarray datasets are summarized in **Supplementary Table 1**.

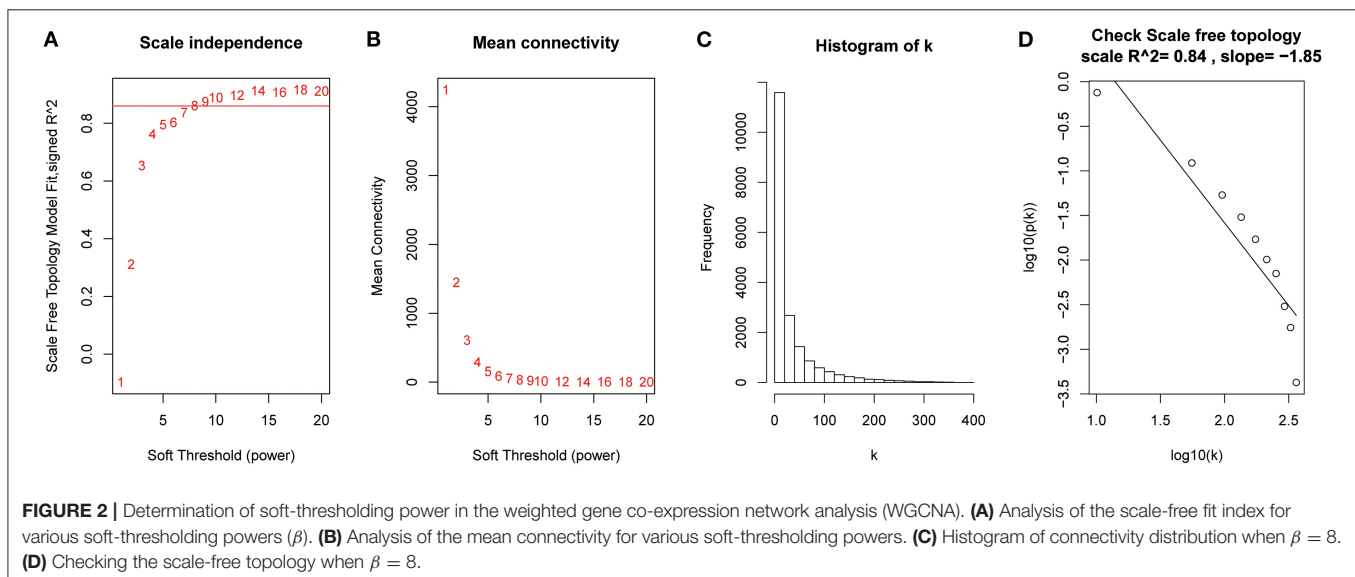
Besides, TCGA data on RNA-sequencing (RNA-Seq) was downloaded using the “TCGAbiolinks” package (19), the ICGC data on RNA-Seq was download from Data Portal (<https://dcc.icgc.org/releases/current/Projects>), normal sample data was download from GTEx Portal (<https://gtexportal.org/home/datasets>). In total, RNA-Seq data were collected from 598 samples, including 270 cases of carcinoma and 328 cases of normal tissue.

### Weighted Gene Co-expression Network Analysis

The theoretical framework of the WGCNA algorithm has been described (20). The “WGCNA” package was used to construct the co-expression network (21). Firstly, the quality of samples and genes was checked. Then, outlier samples were removed by cluster analysis using the average linkage method. When constructing a weight co-expression network, the soft threshold power is an important parameter that can emphasize strong and reduce weak correlations between genes. The power of  $\beta = 8$  (scale-free  $R^2 = 0.86$ ) was selected to ensure a scale-free network. Then, the adjacency was transformed into a topological overlap matrix (TOM), and the topological overlap dissimilarity (1-TOM) was used as hierarchical clustering input. Next, gene modules were identified using a dynamic hybrid branch cutting method with a minimum size of 30 for the gene

**TABLE 1** | Primers sequences of hub genes and internal reference genes.

Gene name	Primers sequences	Amplified fragment size
<i>TSPAN1</i>	Forward 5':TGGGCTGCTATGGTCTAAG Reverse 5':GGCACTACCAGCAACGTCAG	154 bp
<i>TMPRSS4</i>	Forward 5':GGGAAGTCACCGAGAAGA Reverse 5':ATGCCACTGGTCAGATTG	107 bp
<i>CTSE</i>	Forward 5':CTATACCCTCAGCCCACTG Reverse 5':GTTATTCCACGGTCAAAGAC	169 bp
<i>SDR16C5</i>	Forward 5':AATGGGCTGGCAGATTACTG Reverse 5':CACAATCGTGGTTTTGATCC	111 bp
<i>GAPDH</i>	Forward 5':TGACTTCAACAGCGACACCCA Reverse 5':CACCCGTGTGCTGTAGCCAAA	121 bp

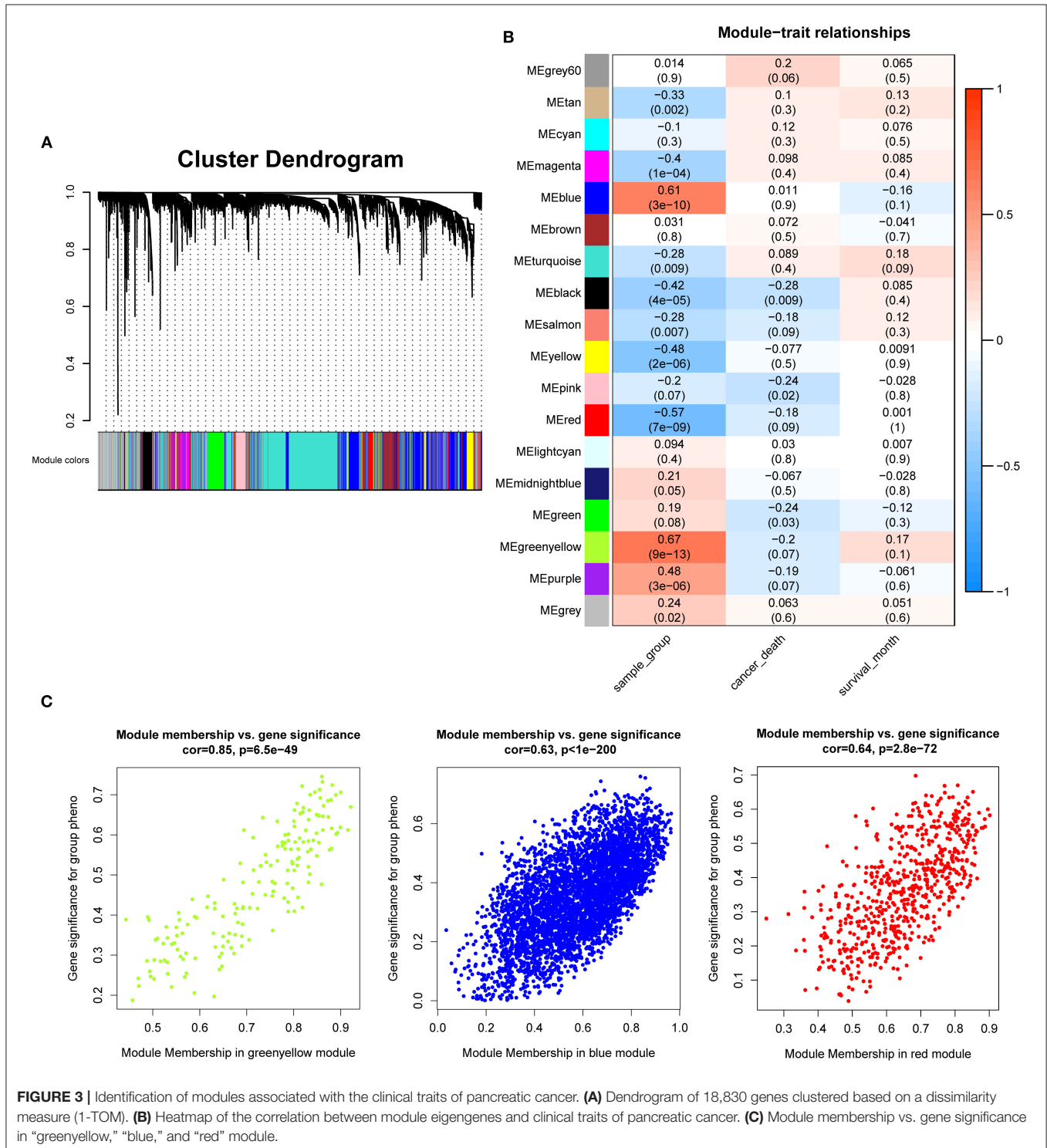


dendrogram (22), and gene modules with a height of <0.25 were combined. An important goal of WGCNA is to detect the gene module subsets that are closely related to clinical traits. Genes within an identified module may have great biological significance. To this end, gene significance (GS) and module significance (MS) were calculated. Also, module membership

(MM) was defined to select highly corrected modules with certain clinical traits.

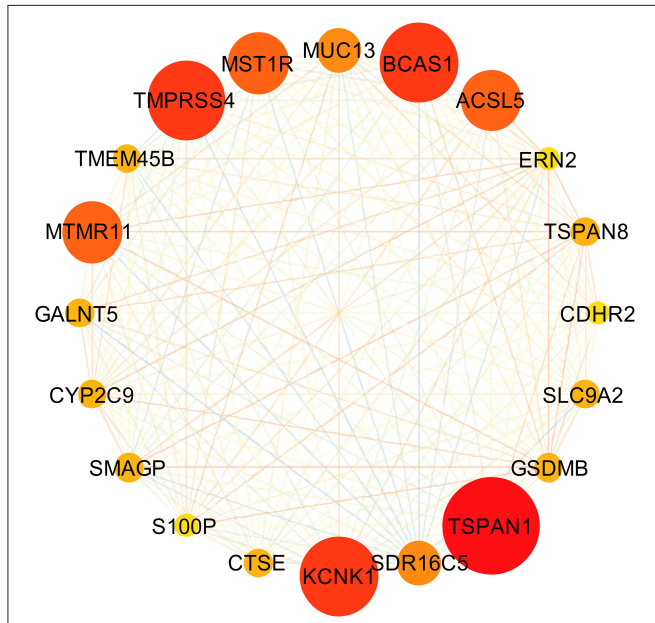
### Identification of the Hub Genes

In gene networks, genes that have many interactions with other genes are defined as hub genes. Hub genes usually play



an important role in a biological system (23). All genes in the significant module were included to construct a gene-gene interaction network using the “cytoHubba” Cytoscape plugin

(24, 25). The top 20 genes, ranked by degrees of interactions, were selected. These genes may play important roles in the pathogenesis of pancreatic cancer. Then, differentially expressed genes (DEGs) were identified for GSE15471, GSE28735, GSE62165, GSE32688, GSE71989, GSE62452, GSE62165, and GSE32676 datasets, respectively. The “limma” (26) package was used to identify DEGs, false discovery rate (FDR) < 0.05 and  $|\log_2 \text{fold change (FC)}| > 1$  were set as the cut-offs. Overall, the hub genes were determined by the intersection of the top 20 genes and the results of the eight DEGs analyses.



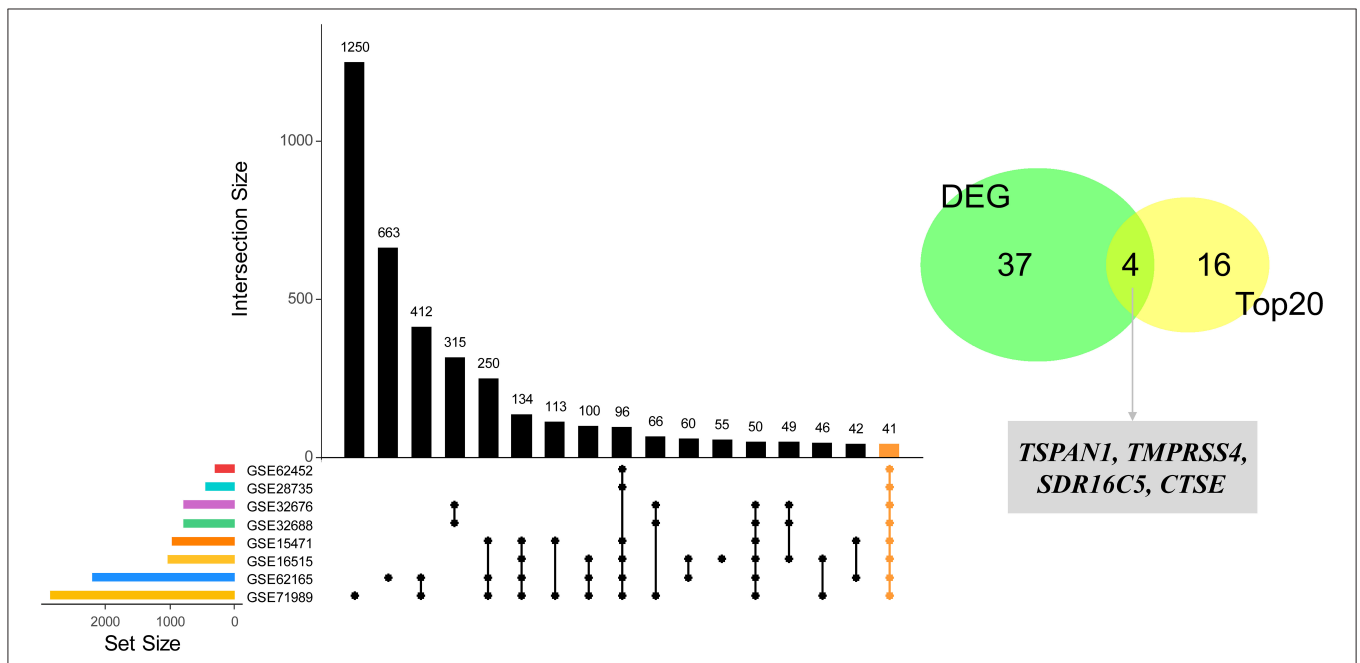
**FIGURE 4 |** Gene-gene interaction network of the top 20 genes. Through constructing a gene-gene interaction network by using 171 genes obtained from WGCNA analysis, the top 20 genes, ranked by degrees of interactions, were identified.

### Validation of the Hub Genes

To validate hub gene expression in pancreatic cancer and normal tissues, the GEPIA tool (<http://gepia.cancer-pku.cn/>) was firstly applied using the RNA-Seq data (27). It is worth emphasizing that the GEPIA website included the TCGA and GTEx datasets (19, 28). And the transcripts per million (TPM) algorithm was used to measure RNA expression (29). Using the “DESeq2” package, further validation was performed based on the negative binomial distribution model using the raw counts of TCGA, ICGC, and GTEx data (30).

### Reverse Transcription PCR (RT-PCR)

cDNA was synthesized using 1  $\mu\text{g}$  of total RNA isolated from three pancreatic cancer cell lines (PANC-1, GCC-PA0001RT; SW1990, GCC-PA0004RT; and AsPC-1, GCC-PA0006RT) and RT-PCR was performed using 400 ng cDNA per 12  $\mu\text{l}$  reactions. The primer sequences used in RT-PCR are described in **Table 1**. Relative expression abundance was determined by  $\Delta\text{Ct} = \text{Ct}(\text{hub gene}) - (\text{GAPDH})$ .  $\Delta\text{Ct} \leq 12$ ,  $12 < \Delta\text{Ct} < 16$  and  $\Delta\text{Ct} \geq 16$  were considered to be a high expression



**FIGURE 5 |** Identification of four hub genes by eight datasets validation. Forty-one DEGs were identified through the intersection of the DEGs of 8 GEO datasets (GSE15471, GSE28735, GSE62165, GSE32688, GSE71989, GSE62452, GSE62165, and GSE32676), and then four hub genes were identified by an intersection with the top 20 genes.

abundance, moderate expression abundance and low expression abundance, respectively.

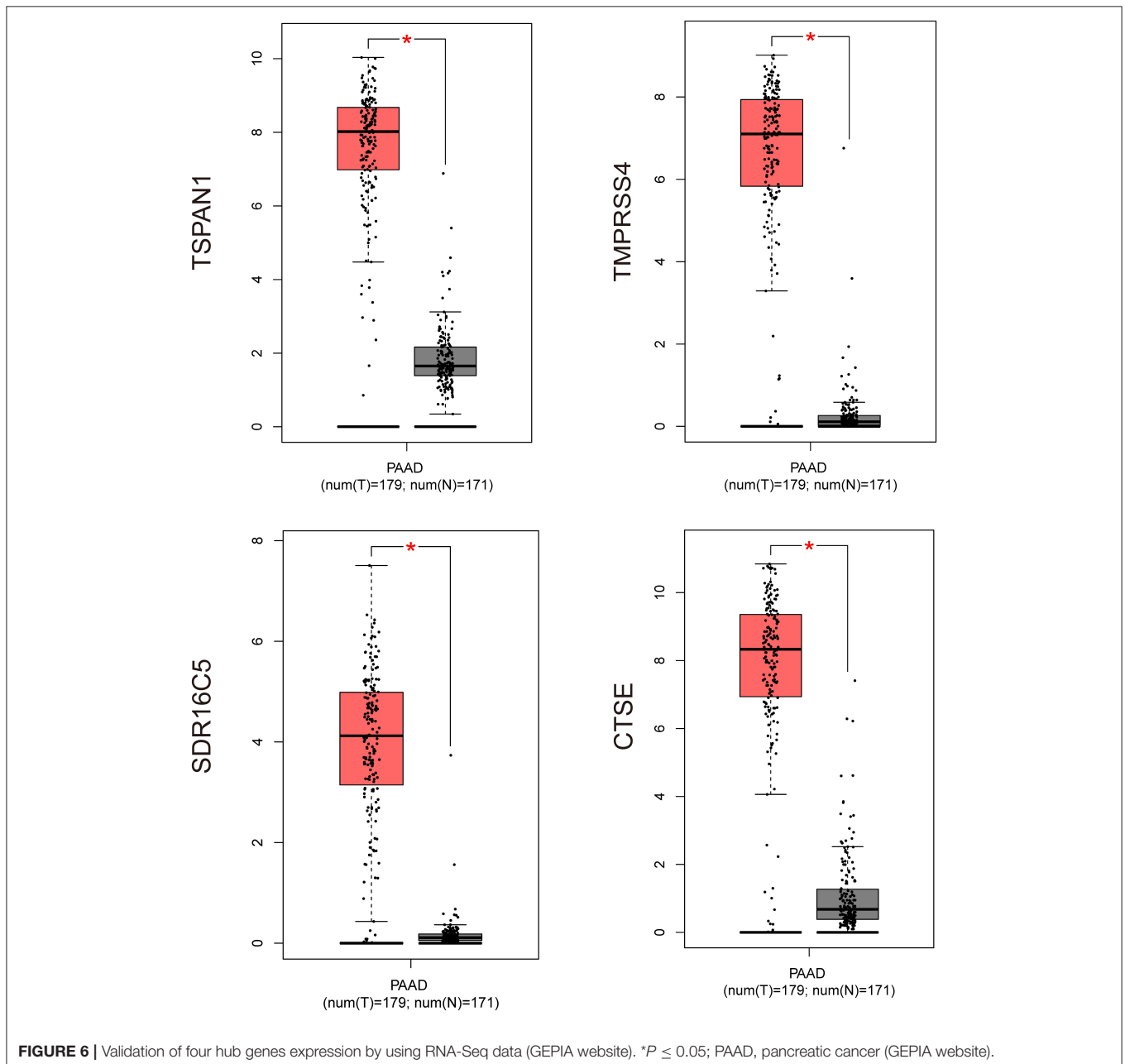
## Immunohistochemistry (IHC)

Specimens of 70 pancreatic cancer tissues and 70 adjacent tissues were deparaffinized and rehydrated. The sections were incubated with polyclonal anti-TSPAN1 antibody (1:1000 dilution) (SANTA CRUZ BIOTECHNOLOGY, sc-376551), anti-TMPRSS4 antibody (1:500 dilution) (proteintech, 11283-1-AP), anti-SDR16C5 antibody (1:300 dilution) (Thermo Fisher, PA5-55229), and anti-CTSE antibody (1:1000 dilution) (SANTA CRUZ BIOTECHNOLOGY, sc-166500). Two independent

pathologists evaluated and scored the IHC, and the log<sub>2</sub> (H-score) described the semi-quantitative expression of the four proteins.

## Diagnostic Model Development and Validation

In this analysis, the merged dataset was used to construct models of pancreatic cancer using four hub genes. A total of 818 samples were randomly assigned into a training cohort and a validation cohort at 7:3 ratios. The GSE32676 dataset was used as the external validation cohort. The support vector machine, random forest, Naive Bayes, neural network, linear



discriminant analysis, mixture discriminant analysis, flexible discriminant analysis, and logistic regression were used to classify pancreatic cancer and normal tissues. To strengthen the robustness of the prediction with these genes, 10-fold cross-validation was also applied reiteratively 100 times. The receiver operating characteristic (ROC) curve was drawn to estimate the diagnostic performance of each model, and the sensitivity and specificity were determined. All statistical analyses were conducted using R 3.5.3.

## RESULTS

### Gene Co-expression Network Construction and Key Modules Identification

After the quality assessment for the GSE28735 dataset, GSM711915 and GSM711957 samples were removed. Eventually, a total of 18,830 genes and 88 samples were included to construct a gene co-expression network using the “WGCNA” package. In the current study, to ensure a scale-free network,  $\beta = 8$  (scale-free  $R^2 = 0.86$ ) was selected (Figures 2A,B), and scale-free topology ( $R^2 = 0.84$ , slope =  $-1.85$ ) was obtained (Figures 2C,D). Through the obtained scale-free topology, 18,830 genes were classified as 18 modules (Figure 3A). Three modules were acquired that were significantly related to the sample category (greenyellow:  $r = 0.67$ ,  $P = 9e-13$ ; blue:  $r = 0.61$ ,  $P = 3e-10$ ; and red:  $r = -0.57$ ,  $P = 7e-9$ ; Figure 3B). The greenyellow module showed the highest correlation with clinical information ( $cor = 0.85$ ,  $P = 6.5e-49$ , Figure 3C). Therefore, the 171 genes of the greenyellow module were used for subsequent analyses.

### Hub Gene Identification and Validation

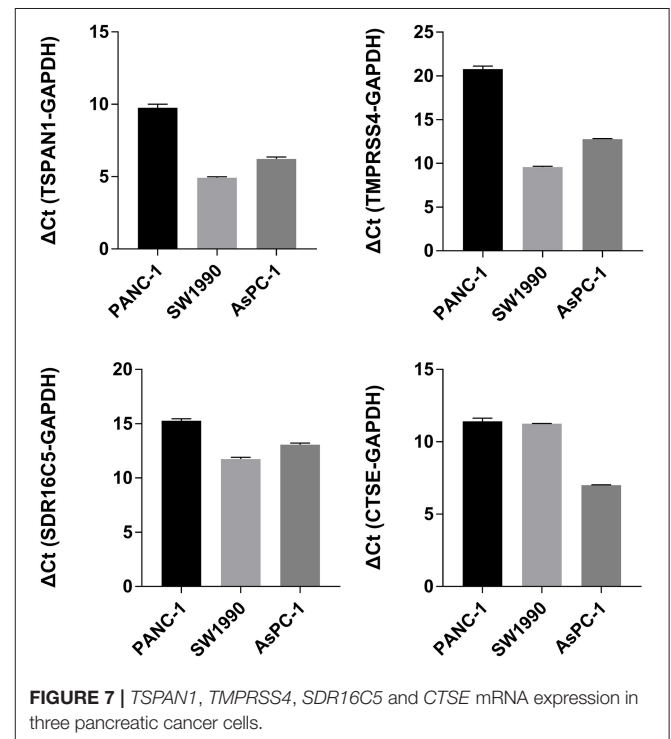
Based on the interaction parameters of the 171 genes obtained from WGCNA analysis, the top 20 genes were identified (Figure 4). DEGs analysis of eight gene datasets revealed a total of 41 genes, so four hub genes, *TSPAN1*, *TMPRSS4*, *SDR16C5*, and *CTSE* was identified (Figure 5). The validation results showed that four hub genes derived from the GEPIA tool were differentially expressed in cancer and normal tissues (Figure 6), as was the result of DESeq2 analysis (Supplementary Table 2). The details of the four genes are shown in Table 2.

### RT-PCR and IHC

The expression of the four hub genes in three cell lines showed that *TSPAN1* and *CTSE* were expressed at high levels, *TMPRSS4* and *SDR16C5* were expressed at medium expression levels (Figure 7). IHC staining results are shown in Figure 8. The expression levels in pancreatic cancer tissues and adjacent tissues showed as follows:  $7.27 \pm 0.31$  and  $6.88 \pm 0.14$ ;  $7.16 \pm 0.24$  and  $7.02 \pm 0.13$ ;  $7.15 \pm 0.24$  and  $6.99 \pm 0.14$ ;  $7.00 \pm 0.26$  and  $6.76 \pm 0.09$ . Higher levels of *TSPAN1*, *TMPRSS4*, *SDR16C5* and *CTSE* expression were observed in pancreatic cancer than in normal pancreatic tissue (paired *t*-test,  $P < 0.0001$ ).

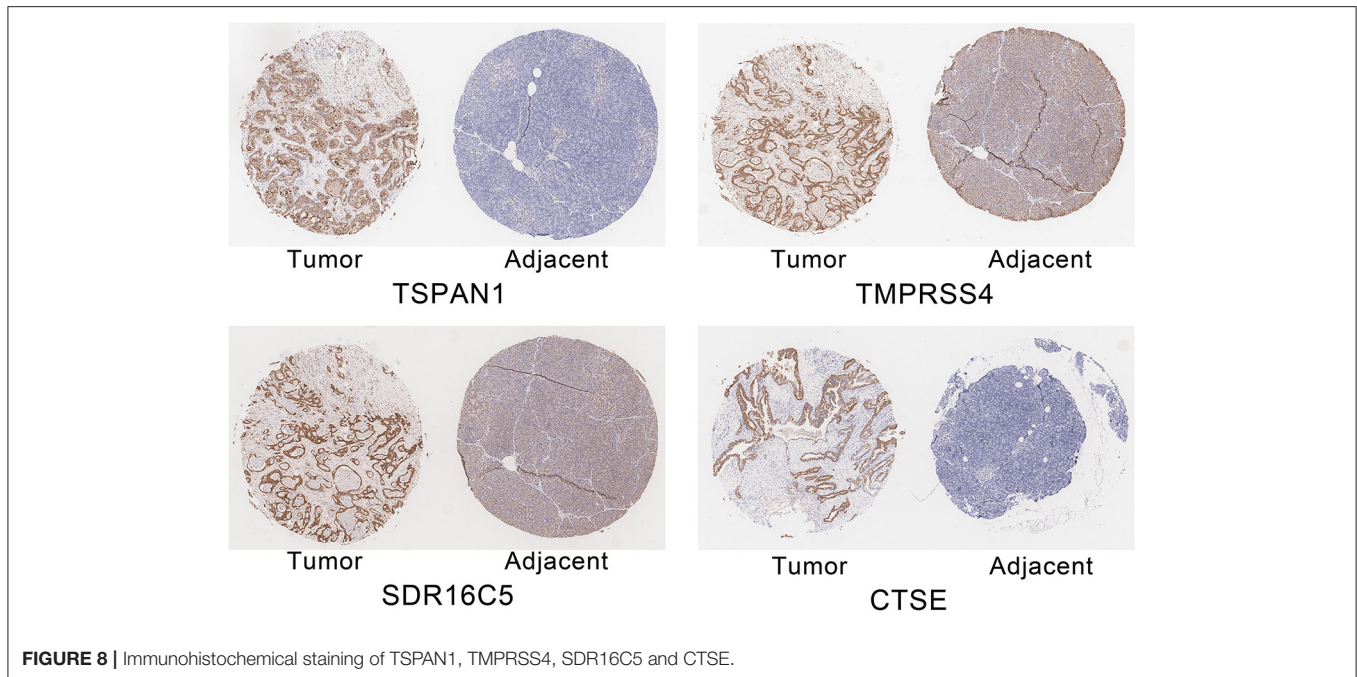
**TABLE 2** | Summary of four hub genes identified by weighted gene co-expression network analysis.

Gene ID	Official full name	Description	References
<i>TSPAN1</i>	Tetraspanin 1	Cell development, activation, growth, and motility	(31, 32)
<i>TMPRSS4</i>	Transmembrane serine protease 4	Integral component of membrane; regulation of gene expression; scavenger receptor activity	(33, 34)
<i>CTSE</i>	Cathepsin E	Antigen processing and presentation of exogenous peptide antigen via MHC class II; protein autoprocessing; protein catabolic process	(35, 36)
<i>SDR16C5</i>	Short chain dehydrogenase/reductase family 16C member 5	Activating transcription factor binding; keratinocyte proliferation; oxidation-reduction process	NA



### Diagnostic Model Development and Validation

In the validation cohort, the AUC of the eight models constructed by machine learning ranged from 0.87 to 0.92, sensitivity ranged from 0.91 to 0.94, and specificity ranged from 0.84 to 0.86. In the external validation cohort, the AUC of the eight models ranged from 0.86 to 0.98, sensitivity ranged from 0.84 to 1.00, and specificity ranged from 0.86 to 1.00 (Table 3).



**FIGURE 8** | Immunohistochemical staining of TSPAN1, TMPRSS4, SDR16C5 and CTSE.

**TABLE 3** | Diagnostic performance of eight machine learning methods for pancreatic cancer.

Methods	Validation (30%)			External validation (GSE32676)		
	AUC	Se	Sp	AUC	Se	Sp
Support vector machine	0.87 (0.79–0.95)	0.92	0.84	0.90 (0.73–1.00)	0.96	0.86
Random forest	0.91 (0.86–0.97)	0.91	0.86	0.94 (0.83–1.00)	0.96	0.86
Naive Bayes	0.91 (0.86–0.96)	0.93	0.84	0.92 (0.77–1.00)	0.96	0.86
Neural network	0.91 (0.86–0.97)	0.94	0.84	0.97 (0.91–1.00)	0.84	1.00
Linear discriminant analysis	0.91 (0.86–0.96)	0.93	0.84	0.95 (0.86–1.00)	1.00	0.86
Mixture discriminant analysis	0.91 (0.87–0.96)	0.91	0.84	0.98 (0.93–1.00)	1.00	0.86
Flexible discriminant analysis	0.91 (0.85–0.96)	0.92	0.84	0.86 (0.71–1.00)	0.84	0.86
Logistic regression	0.92 (0.87–0.97)	0.93	0.84	0.97 (0.90–1.00)	0.96	0.86

AUC, receiver operating characteristic area under the curve value; Se, Sensitivity; Sp, Specificity.

## DISCUSSION

There is an urgent need for a relatively reliable, clinically easy to use, cost-effective biomarker panel for the diagnosis of pancreatic cancer. This study identified four hub genes through bioinformatics, DEGs analysis in multiple datasets, experimental verification of mRNA and protein levels. Using machine learning methods, the expression of four hub genes was utilized to construct models with satisfactory diagnostic value.

Pancreatic cancer is a polygenic and highly heterogeneous disease, the diagnosis of which is challenging (37). A single biomarker may not be sufficient for accurate diagnosis, and a panel consisting of multiple biomarkers might be more beneficial and accurate (38). In the study of pancreatic cancer, some diagnostic models have been developed (39–41). However, most models are not cost-effective for patients, because multiple biomarkers are difficult to routinely screen and/or identify

clinically. Most importantly, a recent study demonstrated that a three-miRNA panel can be as effective as the panel of 1800 miRNAs (42). It is necessary to weigh the number of biomarkers in clinical application and their predictive abilities. Therefore, the focus of this study is to screen hub genes and explore a diagnostic model with cost-effective performance.

With the development of next-generation sequencing, bioinformatics has been used in many ways of research, such as biomarker screening, molecular mechanism exploration. Currently, WGCNA was widely applied to screen hub genes in various cancers (9). This approach can identify critical cancer driver genes that may be a significant therapeutic target or diagnostic marker (43). In recent years, several biomarkers have been identified in the field of cancer research using WGCNA (44–47). However, most studies only used DEGs or the first 25% variation genes to construct a weighted gene co-expression network, which may result in a loss of genetic



diversity. Moreover, some studies only used the feature selection method to select biomarkers (17, 43, 48). Although this method can reduce the dimensionality of data, these genes that play important roles in the cancer process may be lost.

In this study, transcriptome data related to pancreatic cancer were systematically retrieved and its raw data were preprocessed. During the WGCNA analysis process, all genes were included in the construction of a co-expression network to find diagnostic biomarkers, which enhanced the diversity of genes. After using WGCNA to identify a set of genes highly correlated with pancreatic cancer, hub genes were identified through gene-gene interaction network analysis and DEGs analysis in independent eight datasets. It is important to emphasize the interactions between these genes, it can provide deeper insight into the mechanism of cancer (9, 49–51). To increase the credibility of the selected hub genes. DEGs validation was firstly applied using the RNA-Seq data. And then their gene and protein expression levels were verified through experimental methods, including RT-PCR and IHC methods.

In recent years, many studies have suggested that machine learning can provide promising tools for diagnosis in the cancer domain (13). For example, Pu et al. (52) identified a diagnostic model based on five hyper-methylated CpG sites with 0.82% accuracy using the support vector machine method. It is more practical to explore an optimal panel with few biomarkers and high diagnostic performance. Therefore, this study used the four hub gene expression profiles of 818 samples to construct the diagnostic models through machine learning. After internal verification and external verification, the results showed that panels of the four hub genes had a better diagnostic performance for pancreatic cancer.

Four hub genes were identified by bioinformatics in this study. *TSPAN1* (31, 32), *TMPRSS4* (33, 34), and *CTSE* (35, 36) have previously been studied in pancreatic cancer. Among them, *TMPRSS4* was overexpressed in, and identified as a biomarker of, pancreatic carcinoma (33), *TSPAN1*, *TMPRSS4*, and *CTSE* are potential diagnostic or prognostic markers for pancreatic ductal adenocarcinoma (31, 33, 35), and most of these genes are associated with metastasis and proliferation and in pancreatic cancer. Although *SDR16C5* has not been reported in pancreatic cancer, a study showed that it is involved in the regulation of triple-negative breast cancer (53). Its potential as diagnostic marker warrants further functional investigations on its roles in the development of pancreatic cancer.

Certain important strengths of this study should be emphasized. First, the data used in this study are very comprehensive, and the sample size is the largest in the current study of pancreatic cancer. Second, multiple validations of hub genes expression were executed using eight microarray data sets and RNA-Seq data sets, and the RT-PCR and IHC methods were used to validate their expression at the gene and protein level. Those validations can maximize the reliability of the selected hub genes. Third, logistic regression and several machine learning methods were applied to evaluate the diagnostic ability of our

panels. Iterative ten-fold cross-validation repeated 100 times was also used to obtain a robust evaluation of the prediction ability using these genes. There are also some limitations in this study. First, the research samples included in this study were from diverse populations from the USA, France, and Japan. There may exist some differences in gene expression profiles among various ethnic groups. Next, our prediction models will be improved with further validation using independent experimental data.

In conclusion, four hub genes were identified using bioinformatics and experimental verification approaches. More importantly, the four-gene panels can accurately predict pancreatic cancer. Our findings encourage future clinical research to validate the robustness of the diagnostic model and additional functional research.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Life Science Ethics Review Committee of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

TL conceived the project. HW, JW, and CY collected the datasets. TL participated in the pre-processing of the datasets and performed the computational analysis. TL and HY drafted the manuscript. JS, PW, CS, LD, GJ, YH, JL, and YY thoroughly revised the manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was funded by the National Science and Technology Major Project of China (2018ZX10302205), Zhengzhou Major Project for Collaborative Innovation (Zhengzhou university) (18XTZX12007), the Major Project of Science and Technology in Henan Province (161100311400), the Program of Natural Science Foundation of Henan Province (182300410009) and the Key Project of Tackling Key Problems in Science and Technology of Henan Province (192102310308).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.649551/full#supplementary-material>

## REFERENCES

- Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. *Lancet*. (2016) 388:73–85. doi: 10.1016/s0140-6736(16)00141-0
- Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res*. (2014) 74:2913–21. doi: 10.1158/0008-5472.Can-14-0155
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69:7–34. doi: 10.3322/caac.21551
- Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. *Lancet Global Health*. (2018) 6:e555–e67. doi: 10.1016/s2214-109x(18)30127-x
- Idachaba S, Dada O, Abimbola O, Olayinka O, Uma A, Olunu E, et al. A review of pancreatic cancer: epidemiology, genetics, screening, and management. *Open Access Maced J Med Sci*. (2019) 7:663–71. doi: 10.3889/oamjms.2019.104
- Zhang Y, Jiang L, Song L. Meta-analysis of diagnostic value of serum Carbohydrate antigen 199 in pancreatic cancer. *Minerva Med*. (2016) 107:62–9.
- Loosen SH, Neumann UP, Trautwein C, Roderburg C, Luedde T. Current and future biomarkers for pancreatic adenocarcinoma. *Tumour Biol*. (2017) 39:1010428317692231. doi: 10.1177/1010428317692231
- Udyavar AR, Hoeksema MD, Clark JE, Zou Y, Tang Z, Li Z, et al. Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC Syst Biol*. (2013) 7 Suppl. 5:S1. doi: 10.1186/1752-0509-7-s5-s1
- Giulietti M, Occhipinti G, Righetti A, Bracci M, Conti A, Ruzzo A, et al. Emerging biomarkers in bladder cancer identified by network analysis of transcriptomic data. *Front Oncol*. (2018) 8:450. doi: 10.3389/fonc.2018.00450
- Guo X, Xiao H, Guo S, Dong L, Chen J. Identification of breast cancer mechanism based on weighted gene coexpression network analysis. *Cancer Gene Ther*. (2017) 24:333–41. doi: 10.1038/cgt.2017.23
- Giulietti M, Righetti A, Principato G, Piva F. LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer. *Carcinogenesis*. (2018) 39:1016–25. doi: 10.1093/carcin/bgy069
- Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics*. (2005) 21:631–43. doi: 10.1093/bioinformatics/bti033
- Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. (2015) 13:8–17. doi: 10.1016/j.csbj.2014.11.005
- Keutgen XM, Filicori F, Crowley MJ, Wang Y, Scognamiglio T, Hoda R, et al. A panel of four miRNAs accurately differentiates malignant from benign indeterminate thyroid lesions on fine needle aspiration. *Clin Cancer Res*. (2012) 18:2032–8. doi: 10.1158/1078-0432.Ccr-11-2487
- Blanco-Prieto S, De Chiara L, Rodriguez-Gironde M, Vazquez-Iglesias L, Rodriguez-Berrolcal FJ, Fernandez-Villar A, et al. Highly sensitive marker panel for guidance in lung cancer rapid diagnostic units. *Sci Rep*. (2017) 7:41151. doi: 10.1038/srep41151
- Elias KM, Fendler W, Stawiski K, Fiascone SJ, Vitonis AF, Berkowitz RS, et al. Diagnostic potential for a serum miRNA neural network for detection of ovarian cancer. *Elife*. (2017) 6:e28932. doi: 10.7554/eLife.28932
- Liu WT, Wang Y, Zhang J, Ye F, Huang XH, Li B, et al. A novel strategy of integrated microarray analysis identifies CENPA, CDK1 and CDC20 as a cluster of diagnostic biomarkers in lung adenocarcinoma. *Cancer Lett*. (2018) 425:43–53. doi: 10.1016/j.canlet.2018.03.043
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. (2007) 8:118–27. doi: 10.1093/biostatistics/kxj037
- Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiobioinformatics: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. (2016) 44:e71. doi: 10.1093/nar/gkv1507
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. (2005) 4:17. doi: 10.2202/1544-6115.1128
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. (2008) 9:559. doi: 10.1186/1471-2105-9-559
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science*. (2002) 297:1551. doi: 10.1126/science.1073374
- Yu D, Lim J, Wang X, Liang F, Xiao G. Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinform*. (2017) 18:186. doi: 10.1186/s12859-017-1576-1
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. (2003) 13:2498–504. doi: 10.1101/gr.1239303
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. (2014) 8 Suppl. 4:S11. doi: 10.1186/1752-0509-8-s4-s11
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. (2015) 43:e47. doi: 10.1093/nar/gkv007
- Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. (2017) 45:W98–102. doi: 10.1093/nar/gkx247
- Consortium G. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. (2013) 45:580–5. doi: 10.1038/ng.2653
- Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. (2012) 131:281–5. doi: 10.1007/s12064-012-0162-3
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. (2014) 15:550. doi: 10.1186/s13059-014-0550-8
- Hou FQ, Lei XF, Yao JL, Wang YJ, Zhang W. Tetraspanin 1 is involved in survival, proliferation and carcinogenesis of pancreatic cancer. *Oncol Rep*. (2015) 34:3068–76. doi: 10.3892/or.2015.4272
- Zhang X, Shi G, Gao F, Liu P, Wang H, Tan X. TSPAN1 upregulates MMP2 to promote pancreatic cancer cell migration and invasion via PLCgamma. *Oncol Rep*. (2019) 41:2117–25. doi: 10.3892/or.2019.6989
- Bhasin MK, Ndebele K, Bucur O, Yee EU, Otu HH, Plati J, et al. Meta-analysis of transcriptome data identifies a novel 5-gene pancreatic adenocarcinoma classifier. *Oncotarget*. (2016) 7:23263–81. doi: 10.18632/oncotarget.8139
- Cheng Y, Wang K, Geng L, Sun J, Xu W, Liu D, et al. Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma. *EBioMedicine*. (2019) 40:382–93. doi: 10.1016/j.ebiom.2019.01.003
- Azuma T, Hirai M, Ito S, Yamamoto K, Taggart RT, Matsuba T, et al. Expression of cathepsin E in pancreas: a possible tumor marker for pancreas, a preliminary report. *Int J Cancer*. (1996) 67:492–7.
- Cruz-Monserrate Z, Abd-Elgaliel WR, Grote T, Deng D, Ji B, Arumugam T, et al. Detection of pancreatic cancer tumours and precursor lesions by cathepsin E activity in mouse models. *Gut*. (2012) 61:1315–22. doi: 10.1136/gutjnl-2011-300544
- Chudasama P, Mughal SS, Sanders MA, Hubschmann D, Chung I, Deeg KI, et al. Integrative genomic and transcriptomic analysis of leiomyosarcoma. *Nat Commun*. (2018) 9:144. doi: 10.1038/s41467-017-02602-0
- Russell MR, Graham C, D'Amato A, Gentry-Maharaj A, Ryan A, Kalsi JK, et al. A combined biomarker panel shows improved sensitivity for the early detection of ovarian cancer allowing the identification of the most aggressive type II tumours. *Br J Cancer*. (2017) 117:666–74. doi: 10.1038/bjc.2017.199
- Zhou X, Huang Z, Xu L, Zhu M, Zhang L, Zhang H, et al. A panel of 13-miRNA signature as a potential biomarker for predicting survival in pancreatic cancer. *Oncotarget*. (2016) 7:69616–24. doi: 10.18632/oncotarget.11903
- Klett H, Fuellgraf H, Levit-Zerdoun E, Hussung S, Kowar S, Kusters S, et al. Identification and validation of a diagnostic and prognostic multi-gene biomarker panel for pancreatic ductal adenocarcinoma. *Front Genet*. (2018) 9:108. doi: 10.3389/fgene.2018.00108
- Mellby LD, Nyberg AP, Johansen JS, Wingren C, Nordestgaard BG, Bojesen SE, et al. Serum biomarker signature-based liquid biopsy for diagnosis of early-stage pancreatic cancer. *J Clin Oncol*. (2018) 36:2887–94. doi: 10.1200/jco.2017.77.6658

42. Ferroni P, Zanzotto FM, Riondino S, Scarpato N, Guadagni F, Roselli M. Breast cancer prognosis using a machine learning approach. *Cancers*. (2019) 11:3. doi: 10.3390/cancers11030328
43. Fan Z, Xue W, Li L, Zhang C, Lu J, Zhai Y, et al. Identification of an early diagnostic biomarker of lung adenocarcinoma based on co-expression similarity and construction of a diagnostic model. *J Transl Med*. (2018) 16:205. doi: 10.1186/s12967-018-1577-5
44. Chen L, Yuan L, Wang Y, Wang G, Zhu Y, Cao R, et al. Co-expression network analysis identified FCER1G in association with progression and prognosis in human clear cell renal cell carcinoma. *Int J Biol Sci*. (2017) 13:1361–72. doi: 10.7150/ijbs.21657
45. Chen L, Yuan L, Qian K, Qian G, Zhu Y, Wu CL, et al. Identification of biomarkers associated with pathological stage and prognosis of clear cell renal cell carcinoma by co-expression network analysis. *Front Physiol*. (2018) 9:399. doi: 10.3389/fphys.2018.00399
46. Tang J, Kong D, Cui Q, Wang K, Zhang D, Gong Y, et al. Prognostic genes of breast cancer identified by gene co-expression network analysis. *Front Oncol*. (2018) 8:374. doi: 10.3389/fonc.2018.00374
47. Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, et al. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int J Biol Sci*. (2018) 14:124–36. doi: 10.7150/ijbs.22619
48. Zheng B, Liu J, Gu J, Lu Y, Zhang W, Li M, et al. A three-gene panel that distinguishes benign from malignant thyroid nodules. *Int J Cancer*. (2015) 136:1646–54. doi: 10.1002/ijc.29172
49. Zeng T, Sun SY, Wang Y, Zhu H, Chen L. Network biomarkers reveal dysfunctional gene regulations during disease progression. *Febs j*. (2013) 280:5682–95. doi: 10.1111/febs.12536
50. Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat Commun*. (2014) 5:3231. doi: 10.1038/ncomms4231
51. Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, et al. ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. *Cancer Manag Res*. (2018) 10:663–70. doi: 10.2147/cmar.S162813
52. Pu W, Wang C, Chen S, Zhao D, Zhou Y, Ma Y, et al. Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC). *Clin Epigenetics*. (2017) 9:129. doi: 10.1186/s13148-017-0430-7
53. Qi F, Qin WX, Zang YS. Molecular mechanism of triple-negative breast cancer-associated BRCA1 and the identification of signaling pathways. *Oncol Lett*. (2019) 17:2905–14. doi: 10.3892/ol.2019.9884

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ye, Li, Wang, Wu, Yi, Shi, Wang, Song, Dai, Jiang, Huang, Yu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.