Candidate Genes Identified by Constructing a Weighted Gene Co-Expression Network and Key Prognostic Factors of Pancreatic Adenocarcinoma

Peng Xu  
The Central Hospital of Jiading, Shanghai, China  https://orcid.org/0000-0002-5076-5533

Shanjuan Wang (✉ 18930862558@163.com)  
The Central Hospital of JiaDing affiliated to Shanghai Health Medical College  https://orcid.org/0000-0002-7507-3891

Yongchen Wang  
The Blood Station of Jiading, Shanghai, China

Research

Keywords: Bioinformatics, Pancreatic adenocarcinoma, Weighted gene co-expression network analysis, Candidate genes, Overall survival, Pathologic T stage, Neoplasm histologic grade

DOI: https://doi.org/10.21203/rs.3.rs-138900/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Pancreatic adenocarcinoma (PAAD) is a kind of terribly malignant tumor of the digestive tract with high mortality globally, however, the pathogenic molecular mechanisms of which remain unclear currently. Therefore, it is very important to explore the underlying mechanisms and screen for novel prognostic markers and treatment targets. We performed a weighted gene co-expression network analysis (WGCNA) to identify hub genes associated with clinical parameters of interest using The Cancer Genome Atlas (TCGA) data. In addition, Gene ontology (GO) enrichments and survival analyses were conducted.

Methods

The 14 datasets including gene expression profiles and clinical phenotypes for PAAD patients in TCGA were analyzed by WGCNA in R. After constructing the scale-free network of gene co-expression, modules with clinical significance were distinguished, and hub genes were validated, which also key genes regulated relevant traits. GO enrichments were performed on the modules. Furthermore, the significance of hub genes was confirmed via survival analysis.

Results

We discovered 12 candidate genes in total which had significant effects on the overall survival of PAAD patients. The expression of DUSP26, TCEAL2, CELF3, CDK5R2, NAP1L2, INA had strong and positive influences on the pathological T stage, while the expression of CEP55, HJURP, KIF23, CDC45, SAC3D1, ATP6L had great negative impacts on the histological stage. Meanwhile, we also found that the tumor primary location and age of onset were key factors affecting the prognosis of PAAD.

Conclusion

These candidate genes probably play important roles in proliferation, differentiation and metastasis of PAAD cells, which may be acted as novel prognostic markers and effective therapeutic targets. Besides, most of them are first reported in PAAD and deserved in-depth research.

Introduction

Siegel et al. (2019) released the latest estimates of new cancer cases and deaths in the United States. In 2019, PAAD would be expected to strike approximately 56,770 people and to kill 45,750 others, making it the second most frequently diagnosed cancer in the digestive tract just behind colon tumor and the second highest mortality rate among all types of tumors. In contrast to the steady increase in survival for most cancer types, improvements have been so slow for PAAD, in large part because greater than a half of cases are diagnosed at a advanced stage since its special anatomical structure and location, meanwhile, patients do not have unique symptoms (Modolell et al., 1999). In recent years, screening is mainly based on annual magnetic resonance imaging (MRI) and endoscopic ultrasound (EUS), but it is only be used for high-risk individuals, and the acceptability and cost limit its spread (Diane et al., 2019). Given there are no standard protocol for screening currently (Kamisawa et al., 2016; He et al., 2014), chemoresistance and impaired drug access (Schober et al., 2014), and more than 80% of patients have locally advanced or metastatic tumors that are unresectable when diagnosed (Higuera et al., 2016), it is extremely vital to discover novel candidate genes, which play important roles in tumorigenesis and could be new targets for screening and treatment.

Development of high-throughput sequencing had brought light on the research of screening and molecular mechanisms of tumors. TCGA database, which provided numerous publicly available genomic and clinical data of many cancer types, could help researchers to further study and understand the molecular level pathogenesis of each cancer. To explore the underlying mechanisms and identify novel prognostic markers and treatment targets, we performed WGCNA on genomic and clinical data of PAAD downloaded from TCGA to identify hub genes within interested phenotypes and analyze functional enrichment pathways which located in to find early biomarkers and provide new targets for future molecular level treatments, ultimately benefits the patients.

Materials And Methods

Acquiring and preparing genetic and clinical data

The 182 standardized gene expression RNAseq and 222 phenotype data of the 14 datasets for PAAD patients were downloaded from UNIVERSITY OF CALIFORNIA SANTA CRUZ (UCSC Xena) (https://xenabrowser.net/). Due to the absence of some clinical phenotypic data, the sample TCGA-HZ-ABP0-01A was deleted, ultimately 181 gene expression RNAseq samples with corresponding clinical phenotypic data were selected to study. The gene expression level was measured as fragments per kilobase of transcript per million mapped reads (FPKM), the matching annotation information was from the original file downloaded by Illumina. Phenotype data contained the pathologic TNM stage, survival information, histologic grade, tumor growth location, history of exposure to hazardous factors and so on, items of incomplete information were not included.
R was used to preprocess the sequencing data. The *genefilter* package was applied to remove the low expression value usually represent noise, and the *limma* package was served to normalize and adjust the data between row groups (*Fig. S2*). The median absolute deviation (MAD) of the first 5,000 genes were selected.

### Constructing gene co-expression network

Gene co-expression network was constructed by the *WGCNA* package in R. After detecting whether there were missing value and outlier samples (*Fig. S3*), squared Euclidean distance of each sample was calculated by function *adjacency*, and whole sample network connectivity according to distance was standardized by function *scale*. Function *pickSoftThreshold* was used to calculate scale-free topology fitting indices $R^2$ corresponding to different soft thresholding powers $\beta$ (*Fig. S1*). The $\beta$ value was chosen as long as $R^2$ reached 0.85. After that, the gene expression matrix was transformed into an adjacency matrix and a Topological Overlap Matrix (TOM), and then the corresponding dissimilarity of TOM (dissTOM) was calculated. For module detection, hierarchical clustering was used to produce a hierarchical clustering tree (dendrogram) of genes by function *hclust* based on dissTOM. The Dynamic Tree Cut method was performed for branch cutting to obtain different gene modules, represented by the branches of the cluster tree and different colors (*Fig. 1*). A relatively large minimum module size of minClusterSize = 30 was chosen to avoid generating too many small modules. The module eigengene (ME), which can be considered as a representative of the gene expression profiles of a module, was defined as the first principal component of a given module. It was calculated by function *moduleEigengenes*. Modules would be merged if their correlation of MEs was greater than 0.75, which means they have similar expression profiles. Co-expression networks of key genes were visualized by *Cytoscape* (*Fig. 7*).

### Identifying the genes associated with interested clinical phenotype

The correlations between MEs and clinical phenotypes of interest were evaluated by Pearson's correlation tests and $p < 0.05$ was considered to be significantly correlated, subsequently the relationships between gene correlation with trait associated modules and gene significance for trait were plotted (*Fig. 3*). Genes that were significantly related to both a trait and a model may be very important.

### Screening for candidate genes with clinical significance

Module membership (MM) of genes represents the membership with respect to the module, the closer its value is to 1, the more relevant it is to the module. Gene signicancer (GS) reflects the correlation between gene expression and phenotypic data. In our study, at least MM-value $> 0.8$ & GS-value $> 0.2$ were selected as key genes, which had higher degree of connectivities than other genes in the module.

GO enrichment and survival analysis

To explore the potential biological themes of genes in the modules, the *clusterProfiler* package in R was used to annotate and visualize GO terms (*Yu et al., 2012*). In order to validate our conclusion, the R package *survival* was used to carry out log-rank tests, risk form and plot Kaplan–Meier survival curves to confirm significant genes. Besides, we also analyzed the survival of patients affected by different levels of influencing factors, including age, gender, alcohol intake, tumor growth location, smoking index and surgical procedures.

### Results

Gene co-expression network of PAAD

The 5,000 most variant genes according to MAD value were selected for further analysis from the original over 18,000 protein-coding genes. When the value of soft thresholding power $\beta$ was 7, the connectivity between genes met a unexceptionable scale-free network distribution (*Fig. S1*). Thirteen modules were identified by hierarchical clustering and the Dynamic branch Cutting (*Fig. 1*). Each module was assigned a unique color as an identifier. The ME of each module was calculated (*Table S1*). The number of genes in different modules ranged from 38 to 819 (*Table 1*), the name of genes were also listed (*Table S2*), the grey module represented a gene set that was not assigned to any of the modules.

| Module number | 0* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---------------|----|---|---|---|---|---|---|---|---|---|----|----|----|----|
| Module colour | grey* | turquoise | blue | brown | yellow | green | red | black | pink | Magenta | purple | greenyellow | tan | salmon |
| Gene amount   | 379 | 819 | 666 | 644 | 509 | 438 | 423 | 402 | 248 | 230 | 77  | 74  | 53  | 38  |

**Notes.**
The grey module genes with module number 0 were unclassified.

**Clustering dendrogram of genes.**

The hierarchical clustering tree was produced by hierarchical clustering based on dissTOM of genes. Thirteen modules were identified by Dynamic Tree Cutting method with a medium sensitivity (deepSplit = 2) to branch splitting. Each module was assigned a color as an identifier. The colored rows below the dendrogram represented the gene similarity expression modules.

**Gene ontology enrichment analysis of each module**

4,621 genes in 13 modules were mapped to GO database, be in order to obtain their potential functions in the following three aspects, biological process (BP), cellular component (CC) and molecular function (MF) (Tables S3–S5).

**Notes.**

There were no statistically significant results in GO enrichment analysis of purple module.

**Functional enrichment results of each module (took the first three terms of the minimum FDP value ).**

The length of bars represented the number of genes, the color of bars corresponded to module color according to legend, on the left side of the label represented functional enrichment results and corresponding GO encoding.

**Modules associated with the clinical phenotypes of interest**

We analyzed the correlation and significance between MEs and clinical phenotypes, including pathologic stage, history of diabetes, tumor diameter, histologic grade, age at initial pathologic diagnosis and so on (Fig. 3). Meaningful and interesting clinical phenotypes such as pathologic stage, tumor diameter and histologic grade were selected for in-depth analysis.

**Module - trait associations were evaluated via correlations and p-value between MEs and clinical traits of interest.**

Each row and column corresponds to a module eigengene and a trait, respectively. Each cell contains the corresponding correlation (first line) and p-value (second line). The table is color-coded by correlation according to the color legend. Pink, blue and salmon modules negatively correlated to maximum_tumor_dimension (p<0.05). Magenta and black modules positively and red module negatively correlated to neoplasm_histologic_grade (p<0.01). Tan module positively and red module negatively correlated to pathologic_T (p<0.05).

Although the correlation between traits and modules had been found, the module itself still contained many genes, and which needed to be further studied were the most important genes. The genes highly correlated with trait associated modules were also greatly significant for the trait (Fig. 4).

**The relationships between gene correlation with trait associated modules and gene significance for trait.**

Module membership in red module vs. gene significance for pathologic_T (cor=0.61, p=1.8e-44). Pink and blue module memberships vs. gene significance for maximum_tumor_dimension (cor=0.3, p=1.5e-06; cor=0.21, p=4.5e-08). Red, Magenta and black module memberships vs. gene significance for neoplasm_histologic_grade (cor=0.57, p=8.2e-38; cor=0.77, p=2.2e-46; cor=0.48, p=1.5e-24). Correlation between tan module and pathologic_T, salmon module and maximum_tumor_dimension were excluded according to Module membership vs. gene significance (p>0.05). We searched for the hub genes in trait associated modules, which were also the key genes to regulate trait.

**Screened and verified candidate genes of interested clinical phenotypes**

A total of three interested clinical phenotypes and their five corresponding modules were selected to study, we extracted the key genes from modules with significant correlation with phenotypes (Fig. 4, Table 2), which were core regulatory genes in the gene regulatory networks, then used NCBI database (https://www.ncbi.nlm.nih.gov/gene/) to query their functions for further screening. We tried our best to select higher MM & GS value in order to obtain more accurate target genes. Overall, 22 candidate genes were identified that played regulatory or pivotal roles in the pathological T stage, histological stages and tumor diameter of PAAD, respectively.

**Table 2 The meaningful candidate genes related to the interested phenotype filtered out in the module.**
| Module | Relevant phenotype | Hub genes symbol | MM & GS threshold* (greater than) |
|--------|--------------------|------------------|----------------------------------|
| Red    | pathologic_T       | DUSP26 TCEAL2 CELF3 CDK5R2 NAP1L2 INA | 0.9 & 0.25 |
| Pink   | (-0.23,0.001)     | CTRC SLC30A2 C6  | 0.8 & 0.2 |
| Blue   | maximum_tumor_dimension | CFTR CYS1 GRB14 | 0.2 & 0.85 |
| Magenta| (-0.22,0.003)     | MOB2 GDF11       | 0.9 & 0.3 |
| Black  | maximum_tumor_dimension | CEP55 HJURP KIF23 | 0.25 & 0.85 |
|        | (0.17, 0.02)     | CDC5 | |
|        | neoplasm_histologic_grade | SAC3D1 ATP6L ZEB2 | |
|        | (0.29, 9e-5)     | MAML1 | |
|        | (0.25,6e-04)     | | |

Notes.

* MM: the full name is module membership, the closer its absolute value is to 1, the more correlated it is with the module. GS, the abounding name is gene signigicancer, reflects gene expression and phenotypic data correlation.

Survival analysis was used to verify and screening for more accurate and efficient candidate genes in the modules (Table 2). We set the critical expression values of candidate genes based on their median. According to the p value of log-rank test in Kaplan-Meier curve (Fig. 5), we found that the high expression groups for Dual Specificity Phosphatase 26 (DUSP26), Transcription Elongation Factor A Like 2 (TCEAL2), CUGBP Elav-like Family Member 3 (CELF3), Cyclin Dependent Kinase 5 Regulatory Subunit 2 (CDK5R2), Nucleosome Assembly Protein 1 Like 2 (NAP1L2), Internexin Neuronal Intermediate Filament Protein Alpha (INA) and low expression groups for Centrosomal Protein 55 (CEP55), Holliday Junction Recognition Protein (HJURP), Kinesin Family Member 23 (KIF23), Cell Division Cycle Associated 5 (CDC5), SAC3 Domain Containing 1 (SAC3D1), ATPase H+ Transporting V0 Subunit C (ATP6L) had significantly higher overall survival (p<0.05). The results of survival analysis on other genes with no remarkable differences were shown in supplemental materials (Fig. S5).

Survival analysis for twelve candidate genes based on Kaplan Meier plotter.

Reduction in the number of patients at risk represented death or closure. For pathologic_T, the overall survival was significantly increased when comparing higher vs. lower expression group (p<0.05), (i) DUSP26; (ii) TCEAL2; (iii) CELF3; (iv) CDK5R2; (V) NAP1L2; (VI) INA, whereas for neoplasm_histologic_grade, the overall survival was significantly decreased (p<0.05), (i) CEP55; (ii) HJURP; (iii) KIF23; (iv) CDC5; (v) SAC3D1; (vi) ATP6L.

We also analyzed the effects of different exposure factors in the phenotypic data on survival and discovered that the overall survival varies with episode age and tumor growth site, divided by onset at age 67 years, the overall survival was obviously increased when comparing younger vs. older group (p<0.01), pancreatic body tumors had the highest overall survival, followed by the tail, the head was the lowest (p<0.01). Alcohol consumption, smoking index, gender and different surgical types had no significant effect on survival.

Survival analysis for six exposure factors based on Kaplan Meier plotter.

Reduction in the number of patients at risk represented death or closure. (A) Alcohol consumption frequency; (B) Disparate ages of onset; (C) Gender; (D) Tumor growth sites; (E) Smoking index; (F) Different surgical approaches.

Cytoscape of the final selected candidate genes

Mutations in human genes occur at every moment. The key genes have essentially greater connectivities and play crucial pivotal roles in genetic network. Many studies had shown that the human gene networks conform to scale-free characteristics, which means that it can still work as long as the key genes are not abnormal.

The first 30 weighted neighbor genes of each candidate genes

Nodes and edges represented genes and correlations between candidate genes and weighted neighbor genes. Light red nodes were the hub genes of the network.
Discussion

WGCNA is a systems biology approach used to describe patterns of genetic association between different samples. It can be used to identify highly synergistically changing sets of genes. The candidate biomarker genes or therapeutic targets were authenticated according to the endogeneity of gene sets and the association between gene sets and phenotypes. Instead of focusing only on differentially expressed genes, WGCNA used information from thousands or nearly 10,000 of the most varied genes, or from all genes, to identify the sets of genes and perform a significant association analysis with the phenotypes of interest. On the one hand, it made full use of information, on the other hand, it transformed the association between thousands of genes and phenotypes into the association between several gene sets and phenotypes, which avoided the problem of multiple hypothesis testing and correction. Hence, it had been generally and successfully applied in various biological contexts (Amin et al., 2016).

We constructed gene co-expression network, divided the modules according to the similarity of gene expression, performed GO enrichment analysis and then established the figure of module-trait relationships. After verifying that the genes significantly correlated with trait were also remarkably related to trait associated module, we get candidate genes that were meaningful for traits. In order to fully screen and validate the accuracy and effectiveness of candidate genes, cox regression analysis was used with the total survival time and endpoint events corresponding to each sample in the phenotypic data, we finally obtained 12 candidate genes (Fig. 5). The use of differentially expressed genes in WGCNA analysis is not recommended because it completely invalidates the scale-free topology assumption.

In our study, we used MAD to filter out the first 5,000 genes with the most variation. We identified modules and candidate genes associated with pathologic T stage and histologic grade. The expression levels of which were associated with overall survival of PAAD patients, and they could be prognostic biomarkers and effective therapeutic targets for PAAD.

The enrichment analysis of genes in red module revealed that they could encode proteins including various growth factors functioning as regulation of hormone levels, localization, ion migration, biological process and negative regulation of response to wounding, etc, which were a series of positive neuro-endocrine regulation of pancreatic function, while maintaining the acid-base balance and stability of microenvironment. It had inhibitory effects on the destruction of normal organ structure in pathological stage T, organ dysfunction, and the acid-base imbalance caused by high metabolism in anoxic microenvironment. DUSP26/NEAP caused down-regulation of EGFR (Wang et al., 2008), many types of tumour produce far more EGFR than normal cells do, and one effect of that is to stimulate the proliferation of tumour cells. TCEAL2 encoded a member of the transcription elongation factor A (SII)-like (TCEAL) gene family, the family members contain TFA domains and may function as nuclear phosphoproteins that modulate transcription. Pillutla et al. (1999) found the expression of TCEAL1 was the lowest in hematopoietic cells of tumor origin and affected survival of patients. TCEAL2 may promote the encoding of tumor suppressor genes, its mechanism deserves further study. CELF3 played a role in the carcinogenesis and progression of colorectal cancer (Zhou et al., 2018), but our analysis proved that the high expression of CELF3 had a negative regulatory effect on pathological T and significantly prolonged the overall survival of PAAD patients. Genes play diverse roles in different tumors, which needed to be further proved by follow-up experiments. CDK5R2 was involved in positive regulation of cyclin-dependent protein serine/threonine kinase activity, the main ligands were TGF-βs, including TGF-β1~TGF-β5, these members had similar structures and functions that inhibited cell proliferation. It had been demonstrated that decreased expression of CDK5R2 was correlated with poor prognosis of hepatocellular carcinoma (Lu et al., 2011), and Pérez-Morales et al. (2018) indicated CDK5R2 could be a potential lung cancer grading and staging biomarker, it may have similar effects on PAAD patients. NAP1L2 was a kind of nucleosome assembly protein, regulated cellular differentiation and tissue maturation. It may control pathologic T by means of promoting the differentiation and maturation of PAAD cells. INA encoded internexin neuronal intermediate filament protein alpha, loss of the expression of alpha-internexin was closely related to aggressiveness of tumors and patient's adverse prognosis (Wang et al., 2018), and alpha-internexin could be independent prognostic biomarker of pancreatic neuroendocrine tumors (Song et al., 2017). The expression of module genes could significantly inhibit the increase of tumor volume and the expansion of involvement range of adjacent tissues, providing new target sites for future neoadjuvant chemotherapy.

The enrichment analysis of magenta module showed it was associated with cell proliferation, such as cell cycle process, cell division, mitotic nuclear division, etc. CEP55, upregulated and associated with the Janus kinase/signal transducer, activator of transcription signaling pathway and cell proliferation in clear cell renal cell carcinoma (ccRCC), was significantly associated with survival time, and overexpression of CEP55 could promote migration, invasion and EMT of ccRCC cells via the activation of PI3K/AKT/mTOR pathway (Chen et al., 2019). Tumor recurrence and metastasis is the primary cause of cancer-associated mortality in patients, therefore, identification of efficient diagnostic and prognostic molecular markers may improve survival times. In advanced-stage serous ovarian carcinoma, high expression level of HJURP was significantly associated with lymph node metastases (p=0.018) and lower overall survival, which was identified as an independent prognostic biomarker (Li et al., 2018). Montes de Oca et al. (2015) also found HJURP was associated with local and metastatic relapse and poor outcome in breast cancer. KIF23, a nuclear protein and a key regulator of cellular cytokinesis, was associated with glioma malignancy and conferred a worse survival time in glioma, which suggested KIF23 was a new novel prognostic biomarker with potential therapeutic implications (Sun et al., 2016). The expression of CDC5 was associated with decreased survival in HCC and CDC5 knockout arrested cell cycle in the G2/M phase (Tian et al., 2018). Wang et al. (2018) also found the expression of CDC5 was significantly higher in HCC tumour tissues and predicted poor overall survival. These basically explained why the magenta module was related to histological staging of PAAD patients.

Page 6/15
GO enrichment analysis of the black module indicated that it was related to regulation of cell migration, cell motility, cell differentiation, vasculature development, etc. Han et al. (2018) demonstrated SAC3D1 was overexpressed in the hepatocellular carcinoma tissues when compared with matched normal tissues. ATP6L was markedly overexpressed in the poorly differentiated colorectal cancer (CRC) tissues evidently located in invasive front and metastatic foci via the EMT program (Wang et al., 2020), meanwhile, ATP6L enhanced metastatic capacity in prostate cancer cells.

Six kinds of phenotypes were selected for survival analysis. We found that when 67 years old was chosen as the cut-off point of onset age, survival probability of patients was distinctly decreased when comparing older vs. younger group, the risk increased by 1.66 times, suggesting that the age of onset was able to seriously affect the prognosis, which was probably related to cancer consumption and multi-organ functional disturbance. The pancreatic body tumors had a comparatively preferable prognosis, followed by the pancreatic tail tumors. Compared with the body and tail tumors of pancreas, the risk of death from the pancreatic head tumors was 2.87-fold and 1.32-fold, respectively, which possibly related to the fact that the head was more likely to cause obstructive jaundice and secondary suppurative cholangitis and other lethal complications. However, previous drinking level, smoking index, gender, and different surgical methods had no significant impact on the life cycle of patients, which provided references for us to use patient information to judge the prognosis in clinical practice.

We also found the modules related to disease types, which suggested that tumor is a complex regulatory process of multiple genes, and the expression of different genes could develop into various pathological types. In the next step, we will collect relevant cases from the clinic, detect gene expression levels and verify them, so as to provide corresponding markers and therapeutic targets for different disease types and references for pathological classification.

The development of tumor is a multifactor and multistep complex process. By comprehensively analyzing the gene expression profile of patients with pancreatic cancer and their corresponding clinical phenotypes, we had identified the genes that played key roles, some of which is the first to be discovered. They probably serve as novel prognostic markers and effective therapeutic targets.

**Abbreviations**

PAAD: pancreatic adenocarcinoma

WGCNA: weighted gene co-expression network analysis

TCGA: The Cancer Genome Atlas

GO: gene ontology

MRI: magnetic resonance imaging

EUS: endoscopic ultrasound

UCSC Xena: UNIVERSITY OF CALIFORNIA SANTA CRUZ

FPKM: fragments per kilobase of transcript per million mapped reads

MAD: median absolute deviation

Dendrogram: hierarchical clustering tree

dissTOM: dissimilarity of Topological Overlap Matrix

ME: module eigengene

MM: module membership

GS: gene signigicancer

BP: biological process

CC: cellular component

MF: molecular function

ccRCC: clear cell renal cell carcinoma

EGFR: epithelial growth factor receptor

EMT: epithelial-mesenchymal transition
Declarations

-Ethical Approval and Consent to participate

It was an original study that introduced new work that had not been submitted to or accepted by any other journal before. Since we used R software to analyze and study the data in public databases, we had received ethical approval and written informed consent.

-Consent for publication

All our authors have agreed to be published.

-Availability of supporting data

The 182 standardized gene expression RNAseq and 222 phenotype data of the 14 datasets for PAAD patients were downloaded from UNIVERSITY OF CALIFORNIA SANTA CRUZ (UCSC Xena) (https://xenabrowser.net/), the accuracy and availability of supporting data are guaranteed.

-Competing Interests

The authors declare there are no competing interests.

-Funding

This work was supported by the Shanghai Jiading District Science And Technology Development Fund/Key Discipline Funded Project Of Jiading District (Digestive Tumor Direction) (2017ZD01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

-Author Contributions

Peng Xu conceived and designed the experiments, analyzed and reviewed the data, contributed analysis tools, authored drafts of the paper, approved the final draft.

Shanjuan Wang conceived and designed the experiments, reviewed drafts of the paper, approved the final draft.

Yongchen Wang prepared figures and/or tables, approved the final draft.

-Acknowledgements

Thanks to the TCGA database, which provides publicly available genomic and clinical data on many cancer types, as well as the financial supports by the Shanghai Jiading District Science And Technology Development Fund. Useful suggestions given by our hospital scientific research department director Chang Qing, Cai Yiting and other colleagues are also acknowledged. -Authors' information

Peng Xu:

Technical post: attending physician

Work Unit: Jia'ding District Center Hospital Of Shanghai, China.

Educational background:

September 2004 - July 2009 weifang medical college full-time undergraduate

Major: clinical medicine

Sep., 2009-jul., 2012 soochow university full-time academic postgraduate

Journal articles published in the recently two years:

Xu P, wang YF, zhang LH, wang SJ. clinical efficacy and safety of anticoagulant therapy in patients with cirrhotic ascites [J]. Journal of clinical pharmacotherapy, 2019,17 (9) : 46-49.

Xu P, wang YF, zhang LH, wang SJ. Comparative analysis of the risks and benefits of endoscopic treatment for colorectal polyps in inpatient, inpatient, inpatient and outpatient patients [J]. Clinical meta-analysis, 2019,34 (9) : 837-842.

Xu P, wang YF, zhang LH, wang SJ. screening of colorectal cancer in 3798 patients with positive fecal occult blood [J]. International journal of digestive diseases, 2020,40 (1) : 47-52.
References

Siegel et al. (2019) Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34. doi: 10.3322/caac.21551.

Modolell et al. (1999) Modolell I, Guamer L, Malagelada JR. Vagaries of clinical presentation of pancreatic and biliary tract cancer. Ann Oncol. 1999;10(Suppl 4):82–4.

Diane et al. (2019) Diane L, Vinciane R, Frédérique M, Maxime P, Jean-Michel G, Marie-Pierre V, Alain A, Pascal H, Philippe L, Louis DM. Role of endoscopic ultrasound in the screening and follow-up of high-risk individuals for familial pancreatic cancer. World J Gastroenterol. 2019; 25(34): 5082–5096. doi: 10.3748/wjg.v25.i34.5082.

Kamisawa et al. (2016) Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet. 2016;388(10039):73–85. doi: 10.1016/S0140-6736(16)00141-0.

He et al. (2014) He XY, Yuan YZ. Advances in pancreatic cancer research: moving towards early detection. World journal of gastroenterology. 2014;20(32):11241–11248. doi: 10.3748/wjg.v20.i32.11241.

Schober et al. (2014) Schober M, Jesenofsky R, Faissner R, Weidenauer C, Hagmann W, Michl P, Heuchel RL, Haas SL, Löhr J-M. Desmoplasia and chemoresistance in pancreatic cancer. Cancers (Basel). 2014;6(4):2137–2154. doi: 10.3390/cancers6042137.

Higuera et al. (2016) Higuera O, Ghanem I, Nasimi R, Prieto I, Koren L, Feliu J. Management of pancreatic cancer in the elderly. World journal of gastroenterology. 2016;22(2):764–775. doi: 10.3748/wjg.v22.i2.764.

Yu et al. (2012) Yu G, Wang LG, Han Y, He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology. 2012;16:284–287. doi: 10.1089/omi.2011.0118.

Amin et al. (2016) Amin P, DeWitt JP, Mookhtiar AK, Florez M, Ouchida AT, Fan JB, Pasparakis M, Kelliher MA, Ravits J, Yuan J. RIPK1 mediates axonal degeneration by promoting inflammation and necroptosis in ALS. Science. 2016;353:603–608. doi: 10.1126/science.aaf6803.

Wang et al. (2008) Wang JY, Yang CH, Yeh CL, Lin CH, Chen YR. NEAP causes down-regulation of EGFR, subsequently induces the suppression of NGF-induced differentiation in PC12 cells. J Neurochem. 2008;107(6):1544-55. doi: 10.1111/j.1471-4159.2008.05714.x.

Pillutla et al. (1999) Pillutla RC, Shimamoto A, Furuichi Y, Shatkin AJ. Genomic structure and chromosomal localization of TCEAL1, a human gene encoding the nuclear phosphoprotein p21/SIIR. Genomics. 1999;56(2):217-20.

Zhou et al. (2018) Zhou B, Guo R. Integrative analysis of significant RNA-binding proteins in colorectal cancer metastasis. J Cell Biochem. 2018;119(12):9730-9741. doi: 10.1002/jcb.27290.

Lu et al. (2011) Lu JW, Chang JG, Yeh KT, Chen RM, Tsai JJ, Hu RM. Decreased expression of p39 is associated with a poor prognosis in human hepatocellular carcinoma. Med Oncol. 2011; Suppl 1:S239-45. doi: 10.1007/s12032-010-9707-9.

Pérez-Morales et al. (2018) Pérez-Morales J, Mejías-Morales D, Rivera-Rivera S, González-Flores J, González-Loperena M, Cordero-Báez FY, Pedreira-García WM, Charond-Clón C, Cabán-Rivera J, Cress WD, Gordian ER, Muñoz-Antonio T, Cabrera-Ríos M, Isidro A, Coppola D, et al. Hyper-phosphorylation of Rb S249 together with CDK5R2/p39 overexpression are associated with impaired cell adhesion and epithelial-to-mesenchymal transition: implications as a potential lung cancer grading and staging biomarker. PLoS One. 2018;13(11):e0207483. doi: 10.1371/journal.pone.0207483.

Song et al. (2017) Song YL, Yu R, Qiao XW, Bai CM, Lu CM, Xiao Y, Zhong DR, Chen J, Zhao YP, Zhang TP, Song TT, Gao HL, Wan YH, Shen L, Chen J, Lv B, Hao JJ, Zhang Y, Tang L, Chen YJ. Prognostic relevance of UCH-L1 and α-internexin in pancreatic neuroendocrine tumors. Sci Rep. 2017;7(1):2205. doi: 10.1038/s41598-017-02051-1.

Chen et al. (2019) Chen H, Zhu D, Zheng Z, Cai Y, Chen Z, Xie W. CEP55 promotes epithelial-mesenchymal transition in renal cell carcinoma through PI3K/AKT/mTOR pathway. Clin Transl Oncol. 2019;21(7):939-949. doi: 10.1007/s12094-018-02012-8.

Li et al. (2018) Li L, Li X, Meng Q, Khan AQ, Chen X. Increased Expression of Holiday Junction-Recognizing Protein (HJURP) as an Independent Prognostic Biomarker in Advanced-Stage Serous Ovarian Carcinoma. Med Sci Monit. 2018 May 10;24:3050-3055. doi: 10.12659/MSM.906647.

Montes de Oca et al. (2015) Montes de Oca R, Gurard-Levin ZA, Berger F, Rehman H, Martel E, Martel E, Corpet A, Koning LD, Vassias I, Wilson LOW, Meseure D, Reyal F, et al. The histone chaperone HJURP is a new independent prognostic marker for luminal A breast carcinoma. Mol
Oncol. 2015;9(3):657-674. doi: 10.1016/j.molonc.2014.11.002.

Sun et al. (2016) Sun L, Zhang C, Yang Z, Wu Y, Wang H, Bao Z, Jiang T. KIF23 is an independent prognostic biomarker in glioma, transcriptionally regulated by TCF-4. Oncotarget. 2016 Apr 26;7(17):24646-55. doi: 10.18632/oncotarget.8261.

Tian et al. (2018) Tian Y, Wu J, Chagas C, Du Y, Lyu H, He Y, Qi S, Peng Y, Hu J. CDCA5 overexpression is an Indicator of poor prognosis in patients with hepatocellular carcinoma (HCC). BMC Cancer. 2018;18(1):1187. doi: 10.1186/s12885-018-5072-4.

Han et al. (2018) Han ME, Kim JY, Kim GH, Park SY, Kim YH, Oh SO. SAC3D1: a novel prognostic marker in hepatocellular carcinoma. Sci Rep. 2018;8(1):15608. doi: 10.1038/s41598-018-34129-9.

Wang et al. (2020) Wang J, Chen D, Song W, Liu Z, Ma W, Li X, Zhang C, Wang X, Wang Y, Yang Y, Cao W, Qi L. ATP6L promotes metastasis of colorectal cancer by inducing epithelial-mesenchymal transition. Cancer Sci. 2020;111(2):477-488. doi: 10.1111/cas.14283.

**Figures**

Figure 1

Clustering dendrogram of genes.
**Figure 2**

Functional enrichment results of each module (took the first three terms of the minimum FDP value).

**Module-trait relationships**

**Figure 3**

Module-trait associations were evaluated via correlations and p-value between MEs and clinical traits of interest.
Figure 4

The relationships between gene correlation with trait associated modules and gene significance for trait.
Figure 5

Survival analysis for twelve candidate genes based on Kaplan Meier plotter.
Figure 6

Survival analysis for six exposure factors based on Kaplan Meier plotter.

Figure 7

The first 30 weighted neighbor genes of each candidate genes.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalInformation1.csv
- SupplementalInformation2.csv
- FigureS1.doc
- FigureS2.doc
- FigureS3.doc
- FigureS4.doc
- FigureS5.doc
- TableS1.csv
- TableS2.xls
- TableS3.xls
- TableS4.xls
- TableS5.xls
- TableS6.xls