

Integrated analysis of multi-omics data to identification of prognostic genes for pancreatic cancer

Feng Jiang (✉ jiangfeng@wmu.edu.cn)

The First Affiliated Hospital of Wenzhou Medical University <https://orcid.org/0000-0001-6953-5204>

Fan Zhang

The First Affiliated Hospital of Wenzhou Medical University

Jian Chen

The First Affiliated Hospital of Wenzhou Medical University

Junjun Wang

The First Affiliated Hospital of Wenzhou Medical University

Lijuan Hu

The First Affiliated Hospital of Wenzhou Medical University

Jie Chen

The First Affiliated Hospital of Wenzhou Medical University

Yumin Wang

The First Affiliated Hospital of Wenzhou Medical University

Primary research

Keywords: multi-omics, prognosis, pancreatic cancer, biomarker

Posted Date: December 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-125359/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

We aim to develop core modules related to pancreatic cancer (PC) to predict the prognosis of PC patients and explore their tumor microenvironment.

Method:

We merged 175 pancreatic cancer samples in the TCGA database with gene mutation expression, methylation level distribution, mRNA expression and pancreatic cancer-related genes into a public database, and then through weighted correlation network analysis (WGCNA), Two expression modules associated with pancreatic cancer are combined. Then, by integrating these selected genes identified from the first 10 genes of the two co-expression modules, a model risk score is established, and patients are divided into high-risk and low-risk subgroups. Kaplan-Meier survival analysis method is used to analyze differences, analyze the correlation of survival between subgroups, and analyze prognostic models. These selected core genes can divide early pancreatic cancer into two subgroups, compare the prognosis of these two groups, and screen for differentially expressed genes. Use GO and KEGG enrichment analysis to predict the function of differentially expressed genes. The differential expression level and immune cell infiltration level of these selected core genes were further analysis.

Results

Our findings shown nine core genes (MST1R, TMPRSS4, PTK6, KLF5, CGN, ABHD17C, MUC1, CAPN8, B3GNT3) were prognostic biomarkers of pancreatic cancer. These 9 genes could divide early pancreatic cancer into two subgroups, and the two subgroups had significant differences in prognosis, and were mainly different in functions such as digestion and extracellular cell adhesion. Further analysis revealed that the expression of these 9 genes were expressed at high levels in pancreatic cancer tissues. In addition, we validated pancreatic cancer cells and pancreatic epithelial cells through quantitative real-time PCR (qRT-PCR), suggesting that the MST1R, PTK6, ABHD17C and CGN expressed higher in PC cells. CIBERSORT analysis indicated that these genes expression were closely correlated with B-cell naive, CD8⁺ T cells, Macrophages M0 cells, suggesting that these genes may play a carcinogenic role in the preservation of immune-dominant status for tumor microenvironment.

Conclusions

Our research identified 9 key genes which may enhance our understanding of the molecular mechanisms associated with pancreatic cancer.

Background

Pancreatic cancer(PC) is the most invasive malignant tumor in the digestive system[1]. Its incidence is low, but due to early metastasis to local lymph nodes and distant organs, the 5-year survival rate of patients with PC is very low. The only treatment that may cure PC is surgical resection. However, about 80% of tumors are unresectable at the time of diagnosis[1–3]. For patients with advanced PC, chemotherapy is the treatment of choice, but chemotherapy has a wide range of side effects[2, 3]. Therefore, early detection of PC is essential in order to provide patients with an optimal treatment.

Based on the analysis of single omics data, researchers had found many factors related to PC from various aspects[4, 5]. As a complex regulatory system, the occurrence of diseases usually involved genetic mutations, epigenetic changes, and abnormal gene expression. Therefore, it was very meaningful and important to identify the prognostic biomarkers of PC through the integrated analysis of multi-omics data.

Material And Methods

TCGA data download

Pancreatic cancer mutation data was downloaded using the R package TCGAbiolinks[6] (<https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html>). Screen the cancer type as PAAD from <http://firebrowse.org/>, download the SNP6 Copy Number segment data of PC samples and the methylation chip data of PC samples (platform was illumina 450K chip). Both mRNA and miRNA expression profiles and sample clinical data were downloaded from the TCGA official website (<https://portal.gdc.cancer.gov/>). Finally, the data of 175 patients who included 5 data sets (mutation, CNV, methylation, mRNA, miRNA) were analyzed. The clinical data of these 175 patients with PC were shown in Table 1.

Mutation (SNV) analysis

MutSigCV was used to analyze high-frequency mutation genes in tumors. By screening genes with higher mutation frequencies, more mutations, and mutations that occur more frequently in conservative sites. The analysis was performed by the corresponding MutSigCV module in the online analysis tool GenePattern[7] (<https://cloud.genepattern.org/gp/pages/index.jsf>) developed by the Broad Institute. Various types of mutations can occur in cancer, including six basic mutations: C>A, C>G, C> T, T> A, T> C, T>G. In order to study different types of mutations, the researchers proposed the concept of mutational signatures[8]. First, based on six basic types of mutations, and then consider 1 base upstream and 1 base downstream of the mutation site. There were four cases of A, T, C, and G, so a total of 96 mutation types ($4 * 6 * 4$) may be obtained. The frequency of these 96 mutation types was different in different cancers. Non-negative matrix factorization was performed on the frequency of 96 mutation types in each sample to obtain mutation signatures for PC. Here we use the R package maftools[9] (<https://bioconductor.org/packages/release/bioc/html/maftools.html>), somatic signatures[10]

(https://bioconductor.org/packages/release/bioc/html/Somatic_Signatures.html) for mutation signature analysis of tumor samples. Then unsupervised hierarchical clustering was performed on the samples based on the contribution of each label, and the clinical characteristics of the sample subgroups with different mutation characteristic labels were observed.

Difference analysis

The limma[11] package in R was used for differential methylation site analysis, and gene annotation was performed based on the position information of the site. Similarly, the limma package was used to analyze the mRNA and miRNA expression profile data of cancer samples and control samples to screen for differential mRNA and miRNA.

Identify candidate gene sets

The differentially expressed genes were searched together with pancreatic cancer-related genes searched in the GENE database, the OMIM database, and the KEGG database, and genes that appeared at least once were selected as the object of investigation. Then, the genes annotated in the differential methylation site analysis and the markedly mutated genes obtained by the MutSigCV analysis were added to further expand the scope of the investigation and finally obtain a candidate gene set. The expression profile data of candidate gene sets in TCGA in PC samples were selected for analysis of weighted co-expression networks in the next step.

Weight co-expression network analysis and cluster analysis

WGCNA[12] is a systems biology method using gene expression data to construct a scale-free network. Using the WGCNA package of R, a weighted co-expression network is constructed on the expression profile data of the candidate gene set obtained in the previous step. Screen for modules related to PC, construct an interaction subnet for the co-expressed gene sets within the module, and screen candidate core genes for PC based on the network node degree. All clustering heatmaps were completed using the R package pheatmap. The clustering method selected is ward.D.

Survival analysis

According to the clinical data downloaded and compiled by TCGA, Kaplan-Meier analysis was performed on candidate core genes of PC to find genes that had a significant effect on patient prognosis ($P < 0.05$).

Function and path enrichment analysis

GO and KEGG enrichment analysis was performed on the gene set of interest using R package cluster profiler (https://bioconductor.org/packages/release/bioc/html/cluster_Profiler.html)[13].

Immune correlation

CIBERSORT analysis tool was applied for estimate the abundance of immune cells in 175 tumor samples, and the correlation between nine core genes and immune infiltration level.

Quantitative real-time PCR

Expression levels of MST1R, PTK6, ABHD17C and CGN were detected by ABI 7500 Real-time PCR System. Relative expression levels were normalized to β -actin which is internal control.

Statistical analysis

We used R software (3.5.3) for statistical analysis. The T test was used to compare subgroups. We carried out a Chi-square test to describe clinical parameters and compare patient characteristics. Pearson's correlation was applied to analyze the correlation between genes and the level of immune infiltration. K-M survival curves were used to analyze OS between different groups. Enrichment analysis was accomplished by using the hypergeometric test.

Results

Sample and clinical data

The study included omics data and clinical information of 175 patients in total, including mutation results and CNV results of 175 samples, methylation expression profile data including 175 tumor samples and 10 control samples, and mRNA expression profile data including 175 of the tumor samples and 4 control samples, the miRNA expression profile data included 175 tumor samples and 4 control samples. The analysis process of our study is shown in Fig. 1.

Table1 TCGA pancreatic cancer patient characteristics

characteristics	Number(n)
Age(years)	79
<65	
>=65	96
Gender	
Male	96
Female	79
Histological type	
Pancreas-Adenocarcinoma-Other Subtype	25
Pancreas-Adenocarcinoma Ductal Type	144
Pancreas-Colloid (mucinous non-cystic) Carcinoma	4
Pancreas-Undifferentiated Carcinoma	1
Discrepancy	1
Status	
Alive	84
Dead	91
T	
T1	7
T2	24
T3	139
T4	3
TX	1
Not Available	1
N	
N0	49
N1	117
N1b	4
NX	4
Not Available	1

M	
M0	78
M1	4
MX	93
Stage	
Stage I	1
Stage IA	5
Stage IB	15
Stage IIA	28
Stage IIB	116
Stage III	3
Stage IV	4
Not Available	3
Grade	
G1	30
G2	94
G3	47
G4	2
GX	2

Mutant genes and mutation characteristics

Many somatic mutations occur during the development of cancer. A few gene mutations directly promote tumorigenesis and development. These genes were called driver genes. MutSigCV was used to predict mutations in 175 tumor samples. When the significance threshold $P < 0.05$, a total of 991 candidate driver genes were screened, including SMAD4, TP53 and other genes related to cancer development.

Table 2 Results of correlation between gene module characteristics and phenotype

Module	Cor	P value
turquoise	-0.292345297	7.16E-05
black	0.226804486	0.002265162
green	0.127502617	0.088967995
magenta	0.095603005	0.203004528

DNA methylation

The limma package was used to screen the methylation chip data of 175 cancer samples and 10 control samples for differential methylation sites. When the significance threshold was 0.05 and $|\logFC| > 0.3$, a total of 1106 differential methylations were screened. The methylation level of 351 methylation sites decreased, and the methylation level of 755 methylation sites increased.

Analysis of mRNA differentially expressed genes and mutations

Differential gene analysis was performed on the mRNA expression profiles of 175 cancer samples and 4 control samples. When the significance threshold was 0.05 and $|\logFC| > 1$, a total of 246 differentially expressed genes were screened. The effect of each mutation site on genes was different. We counted the effects of mutations on genes in all samples of each gene. These effects could be divided into the following categories, including frame_shift_del, frame_shift_ins, in_frame_del, missense_mutation, nonsense_mutation, nonstop_mutation, splice_site, translation_start_site. statistical analysis by t test found that the distribution of mutation types was significantly different in frame_shift_del ($P=1.56e-12$), in_frame_ins ($P=3.95e-11$), nonstop_mutation ($P=6.22e-05$) and splice_site ($P=0.005$), which indicated that SNV has a certain effect on gene expression. Further enrichment analysis was performed on the differential genes and the significant mutation gene set analyzed by MutSigCV. The hypergeometric distribution test showed that the differential genes were significantly enriched into the significant mutation gene set analyzed by MutSigCV and p value was 0.037.

Candidate mRNA gene set and construction of weighted co-expression networks for candidate gene sets

The differentially expressed genes were searched in the pancreatic cancer-related genes searched in the GENE database, the OMIM database, and the KEGG database. The relationship between the genes and differential genes in the database was shown in the Fig. 2a, of which 216 gene were identified in the differential analysis, and these genes might be related to the onset of PC. In order to further expand the scope of the investigation, the 501 genes annotated in the differential methylation site analysis and the 991 significant mutant genes obtained by MutSigCV analysis were added to finally obtain a candidate gene set with a total of 3284 genes. The PC sample expression data of 3284 genes in TCGA were selected for the next analysis of weighted co-expression networks.

The WGCNA software package of R was used to construct a weighted co-expression network for candidate gene sets. To ensure that the network was scale-free, we chose the optimal $\beta=5$ (Fig. 2b). We used average-linkage hierarchical clustering method to cluster genes, and obtained 13 modules in total.

In order to determine the correlation between the genetic module and the disease, we calculated the Pearson correlation coefficient of each module and the sample characteristics (cancer or normal) (the higher the module was more important) and the significance P value of the corresponding correlation. The results of the correlation coefficient between gene module characteristics and phenotypes were shown in Table 1. Subsequently, the gene significance (GS) value of each gene module was calculated (Fig. 2c, d). The larger the GS value, the more relevant the module was to the disease. Through correlation analysis and GS value calculation, two modules related to PC were finally selected. These two modules were turquoise and black, respectively.

Based on the expression relationship of the genes in the two co-expression modules analyzed above, we constructed a co-expression network for each module separately. In the end, we selected the top 10 genes as the key genes in the co-expression network, that is, the key genes related to PC.

Survival analysis

In order to determine whether the 20 key genes associated with PC obtained in the previous analysis were significantly correlated with prognosis, we performed Kaplan-Meier survival analysis based on the clinical data of these 175 PC samples in TCGA and the expression of these genes in the samples. When the significance threshold was set to 0.05, a total of 9 module core genes, including MST1R(P=0.047), TMPRSS4(P=0.0077), PTK6(P=0.0027), KLF5 (P=0.00019), CGN(P=0.014), ABHD17C(P=0.026), MUC1(P=0.039), CAPN8 (P=0.024), B3GNT3 (P=4e-04) were related to prognosis (Fig. 3).

Further multi-factor Cox regression analysis was performed on these 9 genes, and a regression model $\text{Score} = -0.0877 * \text{MST1R} - 0.0325 * \text{TMPRSS4} + 0.0693 * \text{PTK6} + 0.1893 * \text{KLF5} - 0.0634 * \text{CGN} - 0.1143 * \text{ABHD17C} + 0.1589 * \text{MUC1} - 0.2122 * \text{CAPN8} + 0.3480 * \text{B3GNT3}$. According to the R package maxstat, then determined the classification's best score threshold point was 2.110527 and divided the sample into high score and low score groups according to 2.110527 (Fig. 4). Then Kaplan-Meier analysis was performed based on the survival time and status of the samples, and the analysis results showed that there were significant differences in the prognosis of the samples with high and low groups (P=0.00035).

Driver genes mediate molecular typing of early pancreatic cancer

We selected 150 early PC samples (Stage I, Stage II) from TCGA, and used the expression of the 9 prognostic-related genes (MST1R, TMPRSS4, PTK6, KLF5, CGN, ABHD17C, MUC1, CAPN8, B3GNT3) in the early cancer samples analyzed in the previous step to perform unsupervised clustering analysis (Fig. 5a). As shown in the Fig. 5b, all PC samples could be divided into two categories, and the two types of samples had significantly different prognosis.

In order to further analyze the differences of these key genes in different subgroups, the average expression value of each gene in each type of sample was used as the expression value of the gene in that category. After t-test analysis, it was found that these 9 genes related to prognosis genes were differentially expressed in two subclasses (Fig. 6).

Functional differences in early pancreatic cancer subgroups

The limma package in R was used for differential gene analysis of the two types of PC samples obtained in the previous step. When the significance threshold was 0.05 and $|\log_{2}FC| > 1$, a total of 563 differentially expressed mRNAs were screened (Fig. 7). Perform functional and pathway enrichment analysis of these genes (Fig. 8). These differential gene results show that these genes mainly were related to various digestion and absorption and cell adhesion functions. The KEGG pathway shown that these genes were related to pancreatic secretory function. It was inferred that the process of PC was related to various digestive disorders and changes in extracellular cell connections.

The analysis of these genes combined with clinical characteristics and the proportion of tumor infiltrating leukocytes (TILs)

These genes of expression level in PC tissues were higher in those of paracancerous tissues ($P < 0.05$) (Fig. 9a). These genes showed high expression levels in stage III/IV of PC, but it is not statistically significant (Fig. 9b). It may be because of the sample size was not large enough to pick up a statistical difference. In order to further confirm the correlation between the expression of these nine genes and the immune microenvironment, CIBERSORT algorithm was used to analyze the correlation of 22 kinds of TILs proportion with these 9 genes expression. Among them, B-cell naive, CD8⁺ T cells, and Macrophages M0 cells were correlated with these 9 genes expression (Fig. 10). These results support that the expression levels of these nine genes affect the immune activity of tumor microenvironment. Later, according to the drug-target interactions recorded in the drug-bank database, MUC1, PTK6 were the target genes of 2 drugs (Potassium nitrate, Vandetanib).

MST1R, PTK6, ABHD17C and CGN is highly expressed in PC cells

We next investigated the expression level of four of these genes in cancer cell lines and pancreatic epithelial HPDE lines, MST1R, PTK6, ABHD17C and CGN were expressed relatively high in BxPC-3, Panc-1 and PATU-8988 (Fig. 11, Table 3).

Table 3 Primers designed for the qRT-PCR validation of candidate gene mRNA levels and β -actin

Symbol	Forward primer (5' to 3')	Reverse primer (5' to 3')
MST1R	CATACCGCCACATTGACCCT	CTGCATCACTTGGTACAGAGAAT
PTK6	GCTATGTGCCCCACAACCTACC	CCTGCAGAGCGTGAACCTCC
CGN	CGAGGAAAAAGTCTCACGGCT	GTCTTCAGGTCCTTGTCTGTCTCT
ABHD17C	GCCGTTCGAGGTCTTCTTCTC	GACTCACGCCATACCGGG
β -actin	CCTGGCACCCAGCACAAT	GCTGATCCACATCTGCTGGAA

Discussion

A study by Mishra et al. used multi-omics data to explore genes related to survival. This study is the first analysis of pancreatic ductal adenocarcinoma methylome data based on TCGA[14]. Recently, Chen et al. used multiple datasets to screen DEGs related to DNA methylation. SULT1E1, IGF2BP3, MAP4K4 were potential prognostic biomarkers of PC and selected to construct a prognostic model[15]. In this study, we integrated multi-omics data of 175 cases of PC from TCGA database to generate the feature matrix of genes, which include genomic mutation, methylation and pancreatic cancer-related genes. These four kinds of data comprehensively provide 3284 genes. By WGCNA analysis, we identified two PC related modules. Eventually, by comparing the selected genes in two modules with prognostic values, 9 genes (MST1R, TMPRSS4, PTK6, KLF5, CGN, ABHD17C, MUC1, CAPN8 and B3GNT3) were identified as hub genes. Studies have found that a 15 gene set was associated with the prognosis of PC, and the high expression of the CAPN8 gene in this gene set was associated with a poor prognosis[16]. MST1R kinase was overexpressed in more than 80% of PC, and its effect on epithelial cells and macrophages which can accelerate the progression of PC[17]. However, some studies have found that although RON (also known as MST1R) was involved in the progress of PC in experimental models and may constitute a therapeutic target, its expression was not related to the prognosis of patients with PC who have undergone surgical resection[18, 19]. The prognostic effect of MST1R on early PC has not been found in the literature. PTK6 promotes tumor migration and invasion in PC cells that rely on the ERK signaling pathway[20], and no prognostic effect of PTK6 on early PC has been reported in the literature. Studies have analyzed the expression profile data of PC in GEO and found that TMPRSS4 can be combined with other genes as candidate markers for the prognosis of PC[16]. ABHD17C and CGN have not been reported to be related to the prognosis of PC. MUC1 was associated with tumor invasion and metastasis, and was highly expressed in PC, and its expression was related to overall prognosis[21]. KLF5 expression was increased in pancreatic ductal adenocarcinoma, which can promote mouse cell proliferation, acinar-ductal metaplasia, pancreatic epithelial neoplasia and tumor growth in mice[22]. Other studies reported patients with high expression of KLF5 had shorter overall and tumor-free survival time, and KLF5 promoted cell cycle progression of PC cells[23]. The high expression of glucosyltransferases B3GNT3 plays an important role in the self-renewal of PC stem cells[24]. The absence of B3GNT3 leads to increased proliferation and invasion of PC cells[25], and it was related to prognosis[26]. This indicates that the unreported genes (including MST1R, PTK6, ABHD17C and CGN) can be used as potential driver genes for

PC. Using qPCR, we elucidated its expression in cell lines. Four hub genes were expressed relatively high in PC cell lines. At the same time, we analyzed the potential links between these two modules and 68 differentially expressed miRNAs and found that four differential miRNAs were associated with these two modules. Among them, hsa-miR-375 is related to KLF5.

Afterwards, we constructed a multi-factor Cox regression analysis of these 9 genes and divided the sample into high score and low score groups. Our study found that there were differences between high and low score group in the prognosis. We further confirmed these 9 genes were differentially expressed in two subclasses. Enrichment analyses demonstrated that various digestion, absorption and cell adhesion functions were important pathways. In particular, KEGG pathway demonstrated that these differential gene were closely associated with pancreatic secretory function. It may infer that potential connection existed between the process of PC with various digestive disorders and changes in extracellular cell.

In addition, we confirmed gene expression in PC tissues, and performed CIBERSORT analysis revealed that these 9 genes were negatively correlated with immune infiltrating level of B cell, and CD8⁺ T cell. The high expression of T cell and B cell is related to the prolonged overall survival rate, which can be predicted in many types of tumors including PC[27]. Our study also found that MUC1, PTK6 were the target genes of 2 drugs (Potassium nitrate, Vandetanib). Vandetanib was a multi-target oral small molecule inhibitor that inhibited tumor cells by inhibiting different intracellular signaling pathways during rearrangement during VEGF, EGFR, RET tyrosine kinase transfection, growth and angiogenesis[28]. There are some limitations in current research. First, the number of normal samples of patients is too small. Second, these genetic screenings are based on public data sets. In order to further confirm that these genes are related to prognosis, clinical cohorts need to confirm.

In conclusion, we constructed a co-expression network by using WGCNA and obtained 9 core genes, which can provide theoretical basis for studying dearly PC. In future research, we need to use molecular biology methods to verify our findings. Moreover, functional studies on key genes are needed.

Abbreviations

PC: pancreatic cancer; FC: Fold change; RT-qPCR: Real time-quantitative polymerase chain reaction; GSEA: Gene set enrichment analysis; KEGG: Kyoto Encyclopaedia of Genes and Genomes; GO: Gene ontology; NA: Not available; Adj: Adjusted.

Declarations

Acknowledgments

We thank the First affiliated hospital of wenzhou medical university for providing support on research conditions in this study.

Authors' contributions

All authors contributed to this manuscript, including conception and design (JC, YMW, FJ), acquisition of data (FZ, JC), analysis and interpretation of data (JJW, LJH), material support (FJ), study supervision (YMW). All authors read and approved the final manuscript.

Funding

National Natural Science Foundation of China (81672088), Zhejiang Provincial Natural Science Foundation (LY19H200002), and the Wenzhou Municipal Science and Technology Bureau of China (Y20170718), Zhejiang Provincial Research Center for Cancer Intelligent Diagnosis and Molecular Technology [JBZX-202003].

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the ethics committee of the affiliated hospital of wenzhou medical university, and the written informed consent was obtained from all patients

Consent for publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Author details

¹Department of Laboratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China

²Department of Intensive Care Unit, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China

Feng Jiang, Fan Zhang, Jian Chen contributed equally to this work.

References

1. Rainone M, Singh I, Salo-Mullen EE, Stadler ZK, O'Reilly EM: **An Emerging Paradigm for Germline Testing in Pancreatic Ductal Adenocarcinoma and Immediate Implications for Clinical Practice: A Review.** *JAMA oncology* 2020.

2. Buckley AM, Lynam-Lennon N, O'Neill H, O'Sullivan J: **Targeting hallmarks of cancer to enhance radiosensitivity in gastrointestinal cancers.** *Nature reviews Gastroenterology & hepatology* 2020.
3. Buscail L, Bournet B, Cordelier P: **Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer.** *Nature reviews Gastroenterology & hepatology* 2020.
4. Nagashima T, Yamaguchi K, Urakami K, Shimoda Y, Ohnami S, Ohshima K, Tanabe T, Naruoka A, Kamada F, Serizawa M *et al.*: **Japanese version of The Cancer Genome Atlas, JCGA, established using fresh frozen tumors obtained from 5143 cancer patients.** *Cancer science* 2020, **111**(2):687-699.
5. Long NP, Jung KH, Anh NH, Yan HH, Nghi TD, Park S, Yoon SJ, Min JE, Kim HM, Lim JH *et al.*: **An Integrative Data Mining and Omics-Based Translational Model for the Identification and Validation of Oncogenic Biomarkers of Pancreatic Cancer.** *Cancers* 2019, **11**(2).
6. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I *et al.*: **TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data.** *Nucleic acids research* 2016, **44**(8):e71.
7. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP: **GenePattern 2.0.** *Nature genetics* 2006, **38**(5):500-501.
8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL *et al.*: **Signatures of mutational processes in human cancer.** *Nature* 2013, **500**(7463):415-421.
9. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP: **Maftools: efficient and comprehensive analysis of somatic variants in cancer.** *Genome research* 2018, **28**(11):1747-1756.
10. Gehring JS, Fischer B, Lawrence M, Huber W: **SomaticSignatures: inferring mutational signatures from single-nucleotide variants.** *Bioinformatics* 2015, **31**(22):3673-3675.
11. !!! INVALID CITATION !!! [11].
12. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Statistical applications in genetics and molecular biology* 2005, **4**:Article17.
13. 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**(5):284-287.
14. Mishra NK, Southekal S, Guda C: **Survival Analysis of Multi-Omics Data Identifies Potential Prognostic Markers of Pancreatic Ductal Adenocarcinoma.** *Frontiers in genetics* 2019, **10**:624.
15. Chen H, Kong Y, Yao Q, Zhang X, Fu Y, Li J, Liu C, Wang Z: **Three hypomethylated genes were associated with poor overall survival in pancreatic cancer patients.** *Aging* 2019, **11**(3):885-897.
16. Cheng Y, Wang K, Geng L, Sun J, Xu W, Liu D, Gong S, Zhu Y: **Identification of candidate diagnostic and prognostic biomarkers for pancreatic carcinoma.** *EBioMedicine* 2019, **40**:382-393.
17. Babicky ML, Harper MM, Chakedis J, Cazes A, Mose ES, Jaquish DV, French RP, Childers B, Alakus H, Schmid MC: **MST1R kinase accelerates pancreatic cancer progression via effects on both epithelial cells and macrophages.** 2019, **38**(28):5599-5611.

18. Tactacan CM, Chang DK, Cowley MJ, Humphrey ES, Wu J, Gill AJ, Chou A, Nones K, Grimmond SM, Sutherland RL *et al*: **RON is not a prognostic marker for resectable pancreatic cancer.** *BMC cancer* 2012, **12**:395.
19. Wang W, Xing H, Huang C, Pan H, Li D: **Identification of pancreatic cancer type related factors by Weighted Gene Co-Expression Network Analysis.** *Medical oncology (Northwood, London, England)* 2020, **37**(4):33.
20. Ono H, Basson MD, Ito H: **PTK6 promotes cancer migration and invasion in pancreatic cancer cells dependent on ERK signaling.** *PloS one* 2014, **9**(5):e96060.
21. Hinoda Y, Ikematsu Y, Horinouchi M, Sato S, Yamamoto K, Nakano T, Fukui M, Suehiro Y, Hamanaka Y, Nishikawa Y *et al*: **Increased expression of MUC1 in advanced pancreatic cancer.** *Journal of gastroenterology* 2003, **38**(12):1162-1166.
22. He P, Yang JW, Yang VW, Bialkowska AB: **Kruppel-like Factor 5, Increased in Pancreatic Ductal Adenocarcinoma, Promotes Proliferation, Acinar-to-Ductal Metaplasia, Pancreatic Intraepithelial Neoplasia, and Tumor Growth in Mice.** *Gastroenterology* 2018, **154**(5):1494-1508.e1413.
23. Li Y, Kong R, Chen H, Zhao Z, Li L, Li J, Hu J, Zhang G, Pan S, Wang Y *et al*: **Overexpression of KLF5 is associated with poor survival and G1/S progression in pancreatic cancer.** *Aging* 2019, **11**(14):5035-5057.
24. Barkeer S, Chugh S, Karmakar S, Kaushik G, Rauth S, Rachagani S, Batra SK, Ponnusamy MP: **Novel role of O-glycosyltransferases GALNT3 and B3GNT3 in the self-renewal of pancreatic cancer stem cells.** *BMC cancer* 2018, **18**(1):1157.
25. Gupta R, Leon F, Thompson CM, Nimmakayala R, Karmakar S, Nallasamy P, Chugh S, Prajapati DR, Rachagani S, Kumar S *et al*: **Global analysis of human glycosyltransferases reveals novel targets for pancreatic cancer pathogenesis.** *British journal of cancer* 2020, **122**(11):1661-1672.
26. Tian G, Li G, Liu P, Wang Z, Li N: **Glycolysis-Based Genes Associated with the Clinical Outcome of Pancreatic Ductal Adenocarcinoma Identified by The Cancer Genome Atlas Data Analysis.** *DNA and cell biology* 2020, **39**(3):417-427.
27. Iglesia MD, Parker JS, Hoadley KA, Serody JS, Perou CM, Vincent BG: **Genomic Analysis of Immune Cell Infiltrates Across 11 Tumor Types.** *Journal of the National Cancer Institute* 2016, **108**(11).
28. Lee SH, Lee JK, Ahn MJ, Kim DW, Sun JM, Keam B, Kim TM, Heo DS, Ahn JS, Choi YL *et al*: **Vandetanib in pretreated patients with advanced non-small cell lung cancer-harboring RET rearrangement: a phase II clinical trial.** *Annals of oncology : official journal of the European Society for Medical Oncology* 2017, **28**(2):292-297.

Figures

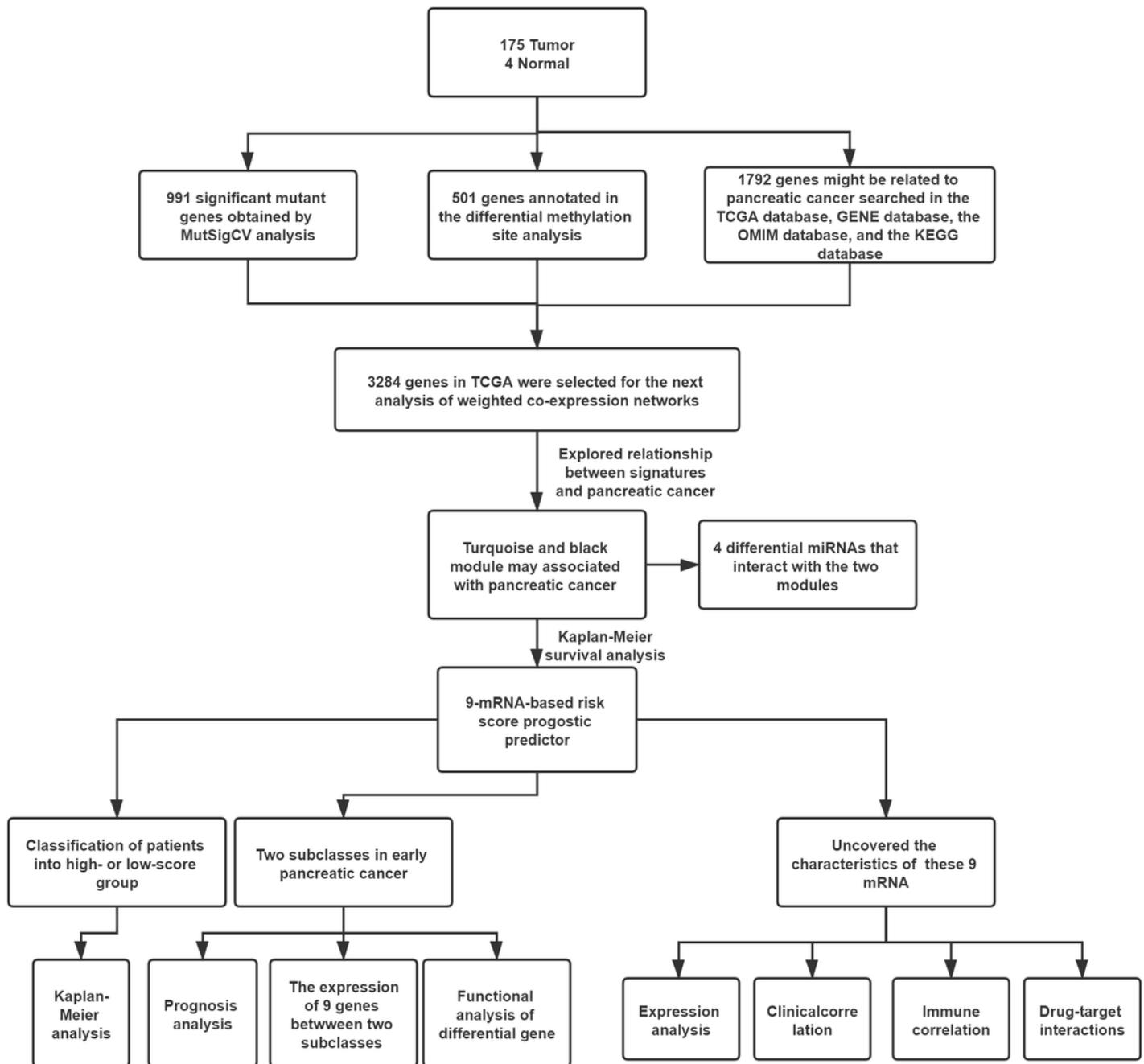


Figure 1

Analysis workflow of this study

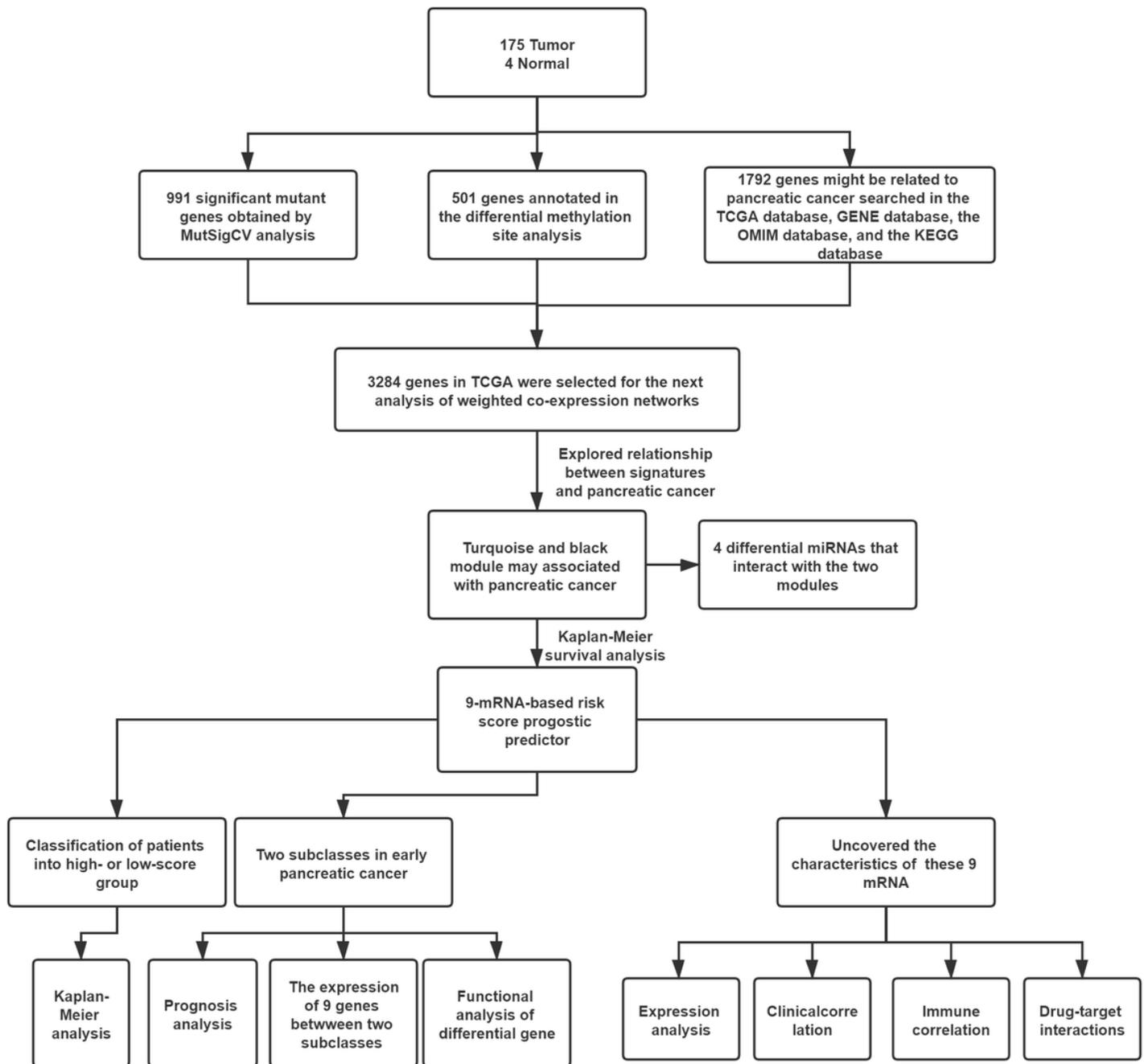


Figure 1

Analysis workflow of this study

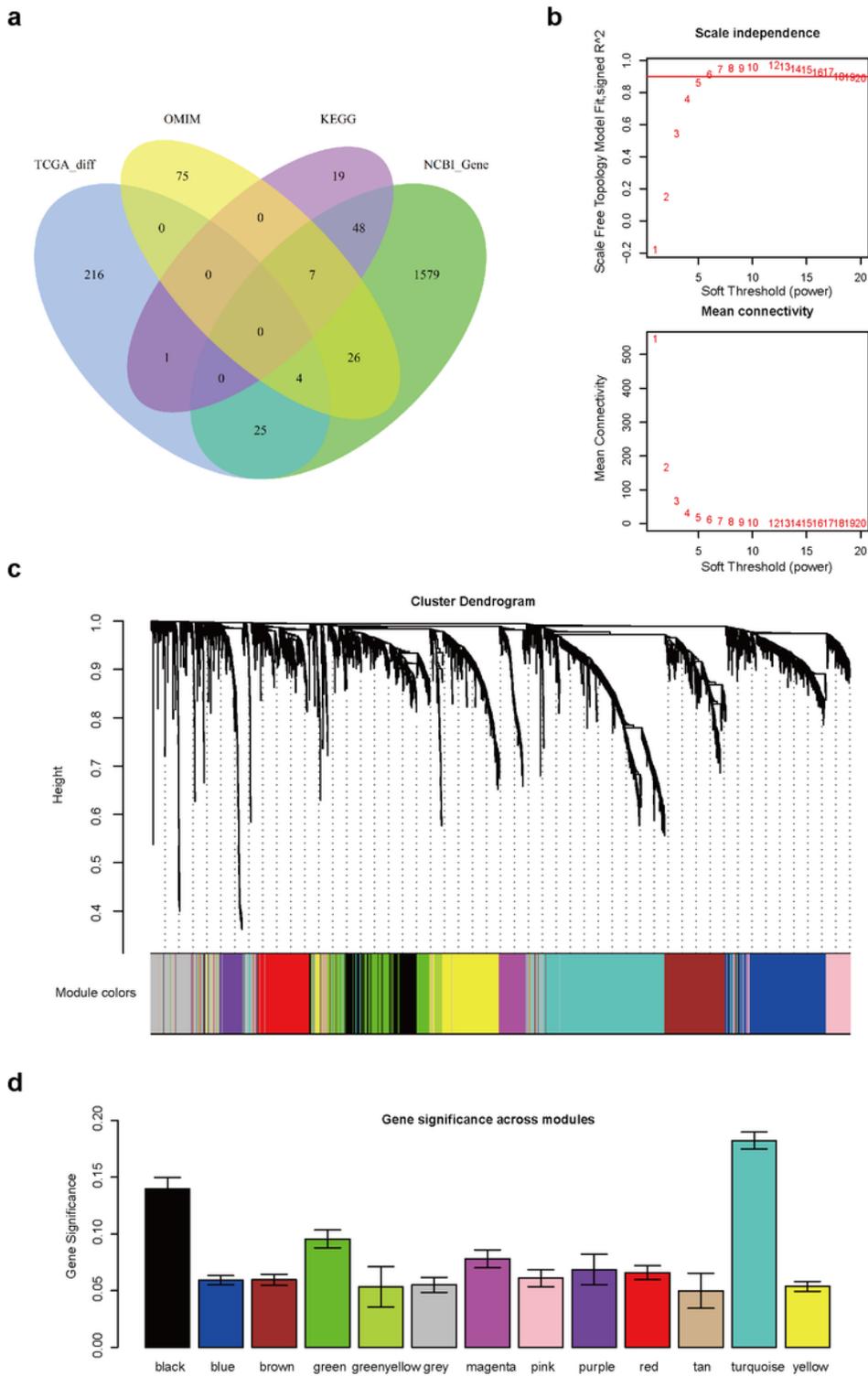


Figure 2

Hub module selection. a Genes related to pancreatic cancer from different platforms. b Network soft threshold β determination. c co-expression module. d Mean distribution of GS in different gene modules

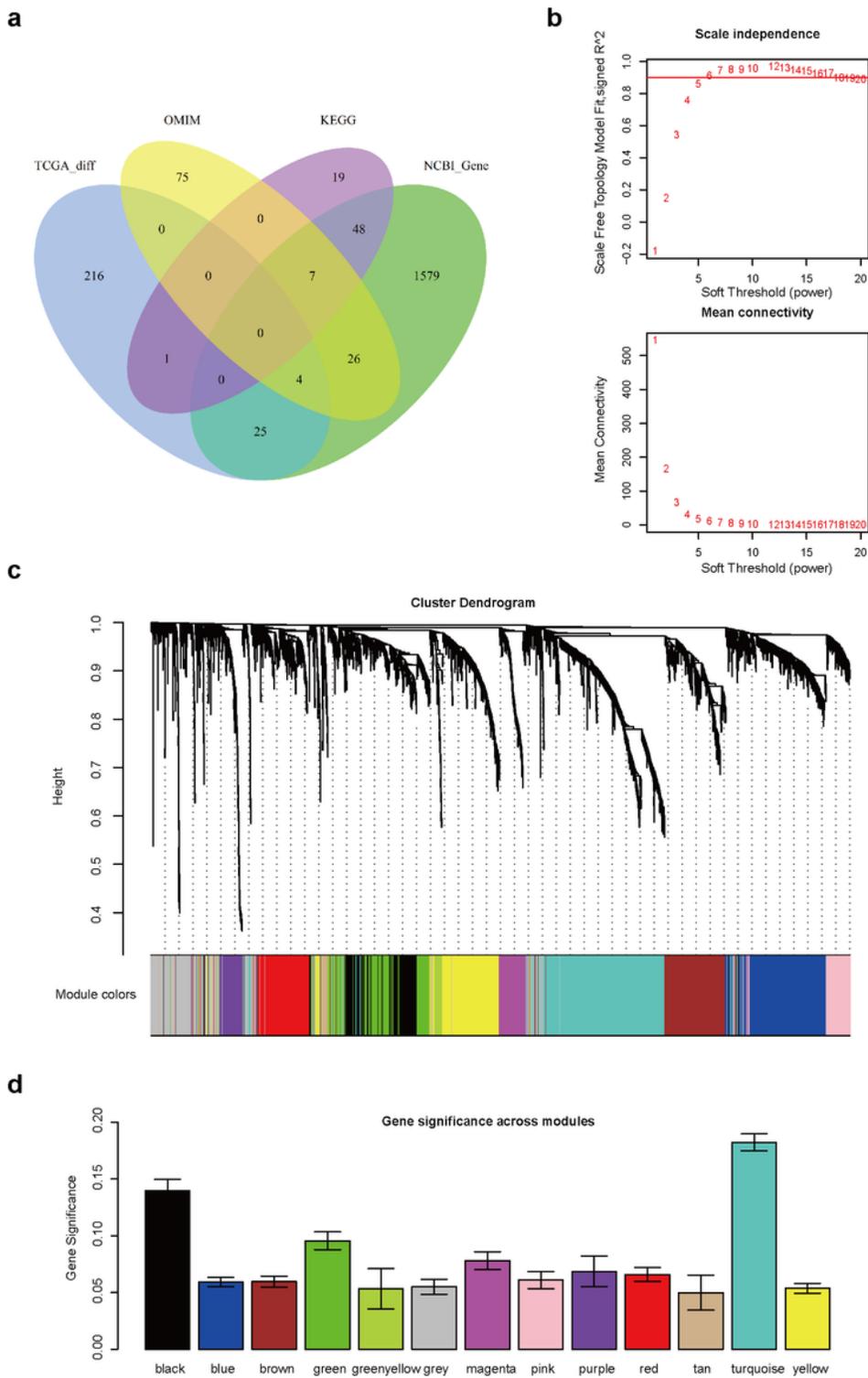


Figure 2

Hub module selection. a Genes related to pancreatic cancer from different platforms. b Network soft threshold β determination. c co-expression module. d Mean distribution of GS in different gene modules

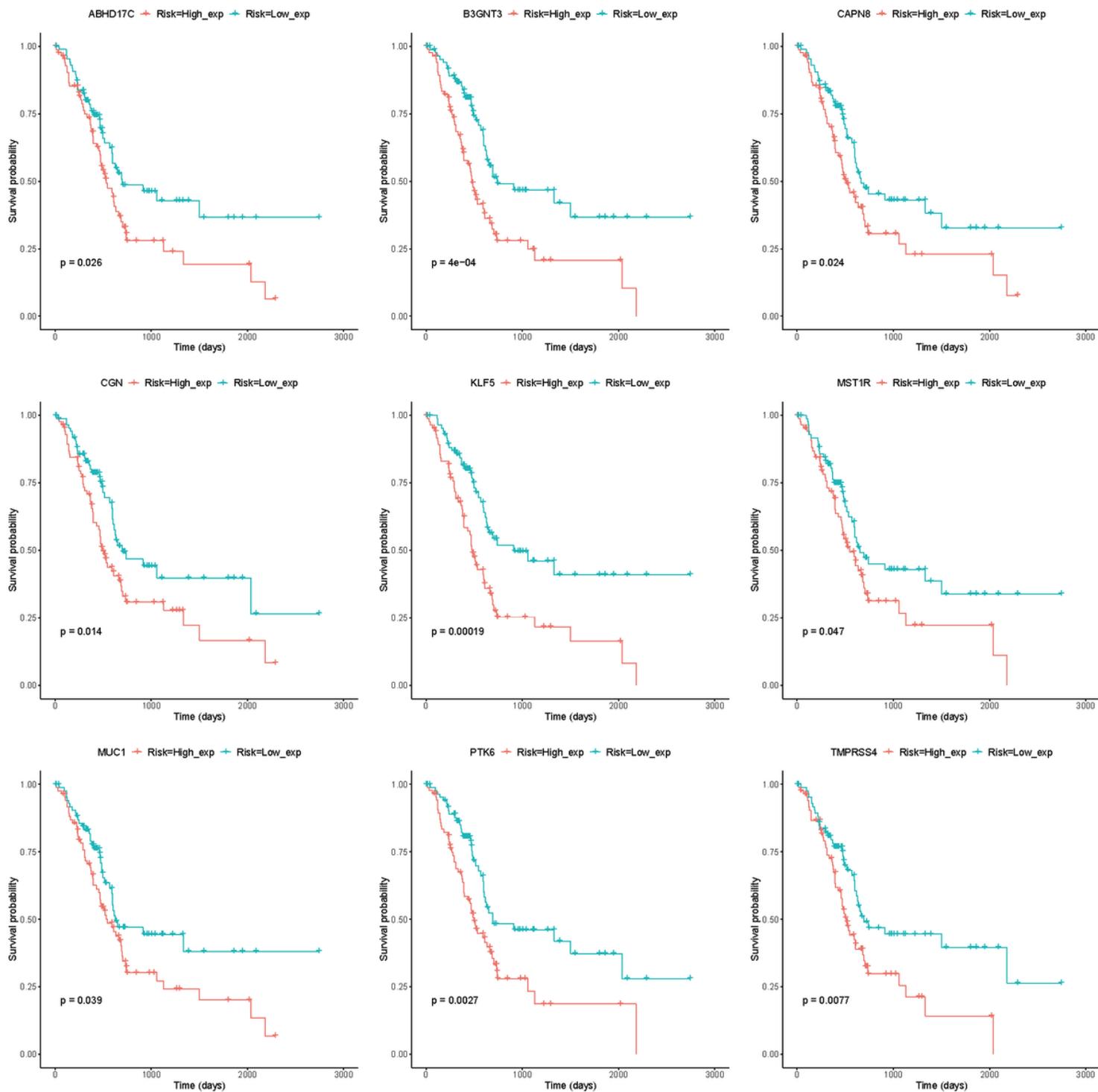


Figure 3

gene expression survival curves

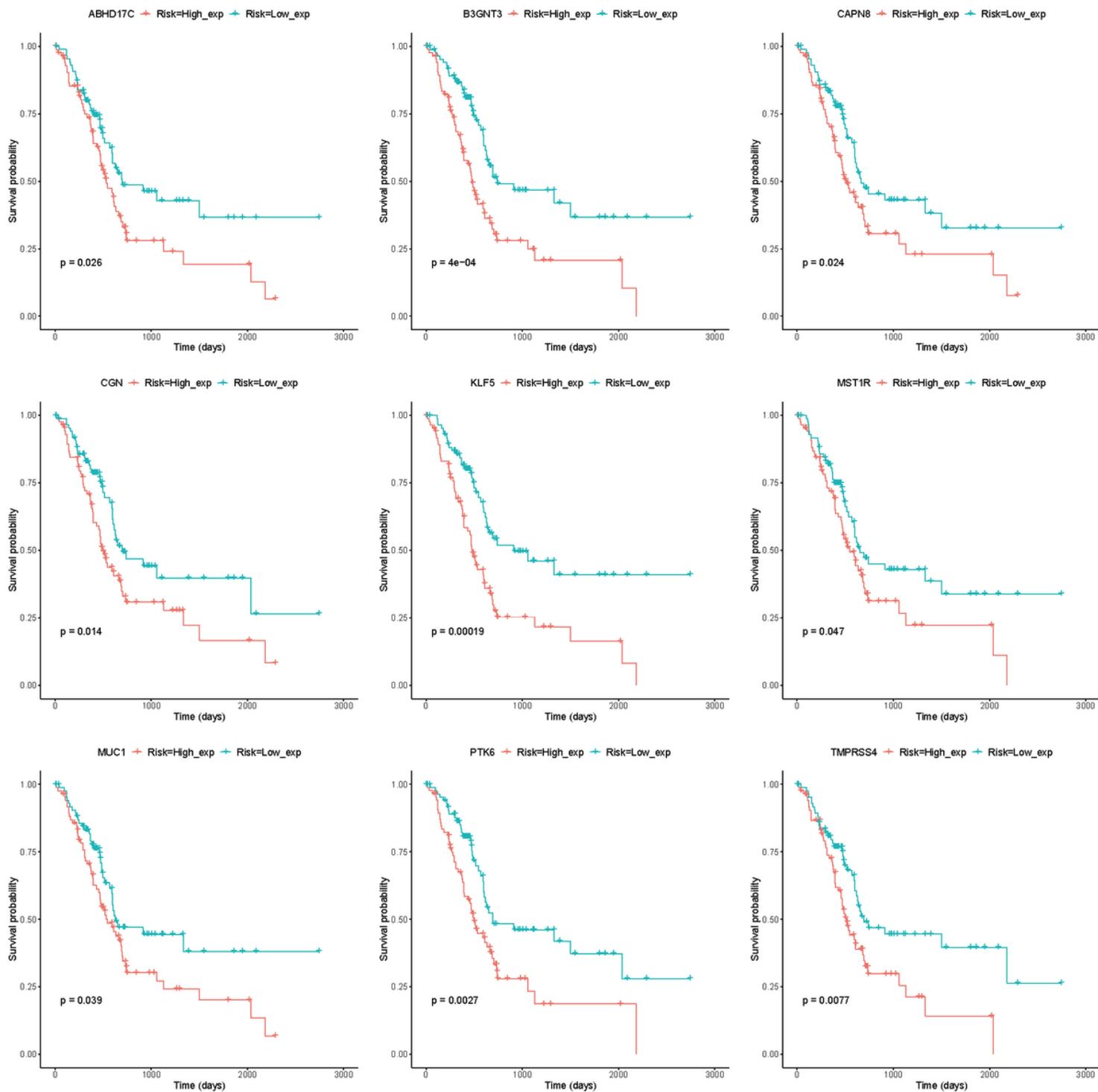


Figure 3

gene expression survival curves

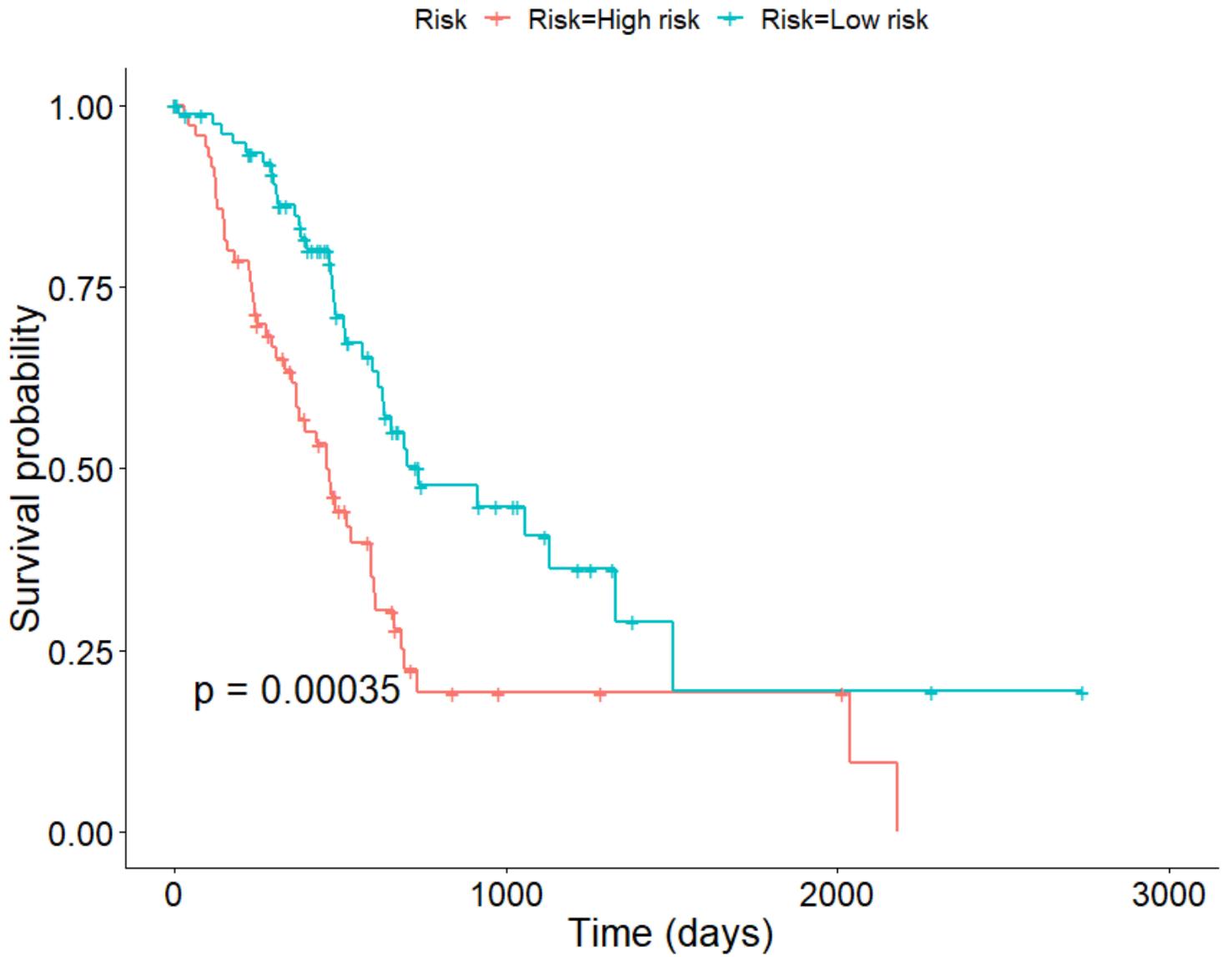


Figure 4

Survival curves of two types of samples based on the comprehensive score of 9 gene expressions

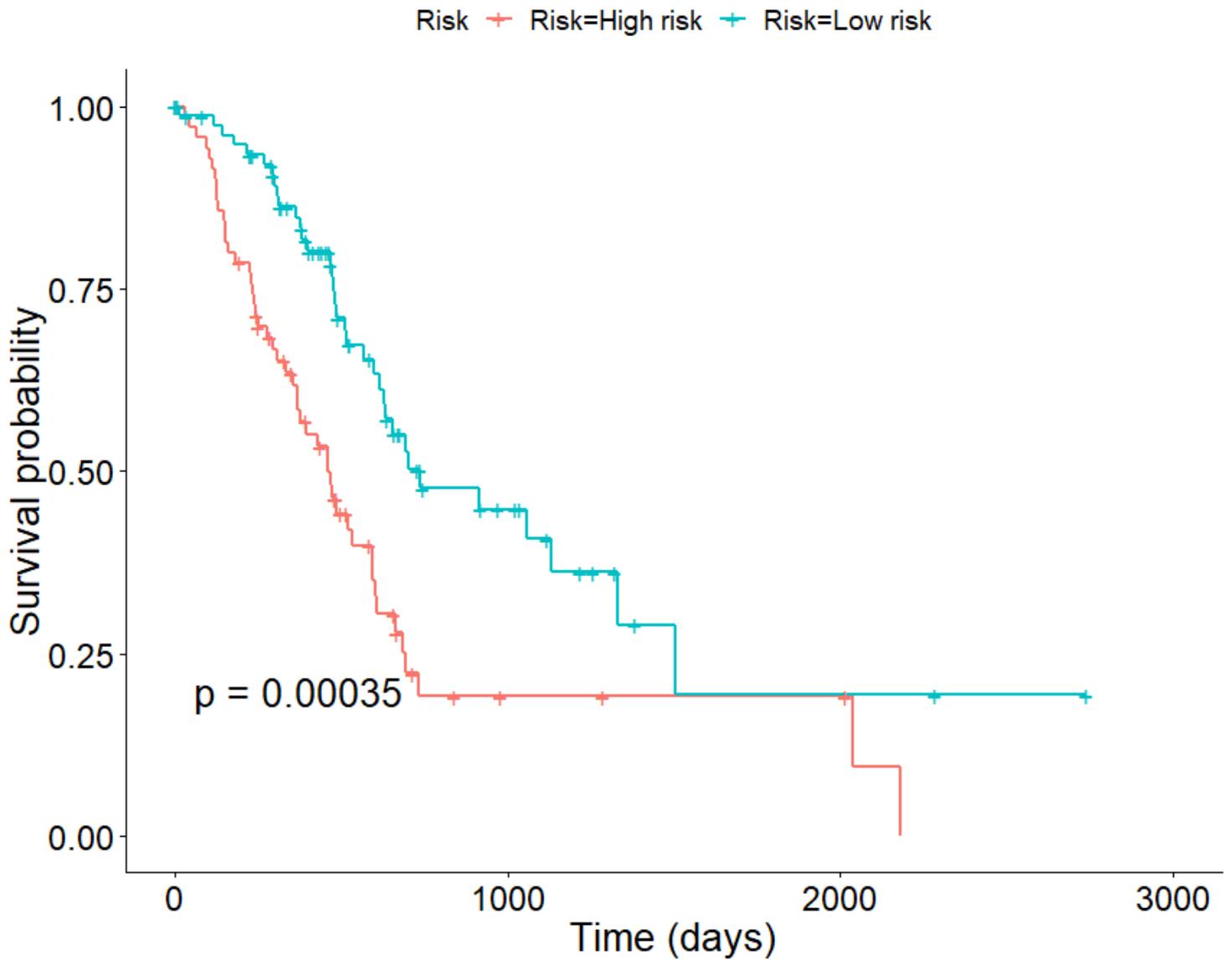


Figure 4

Survival curves of two types of samples based on the comprehensive score of 9 gene expressions

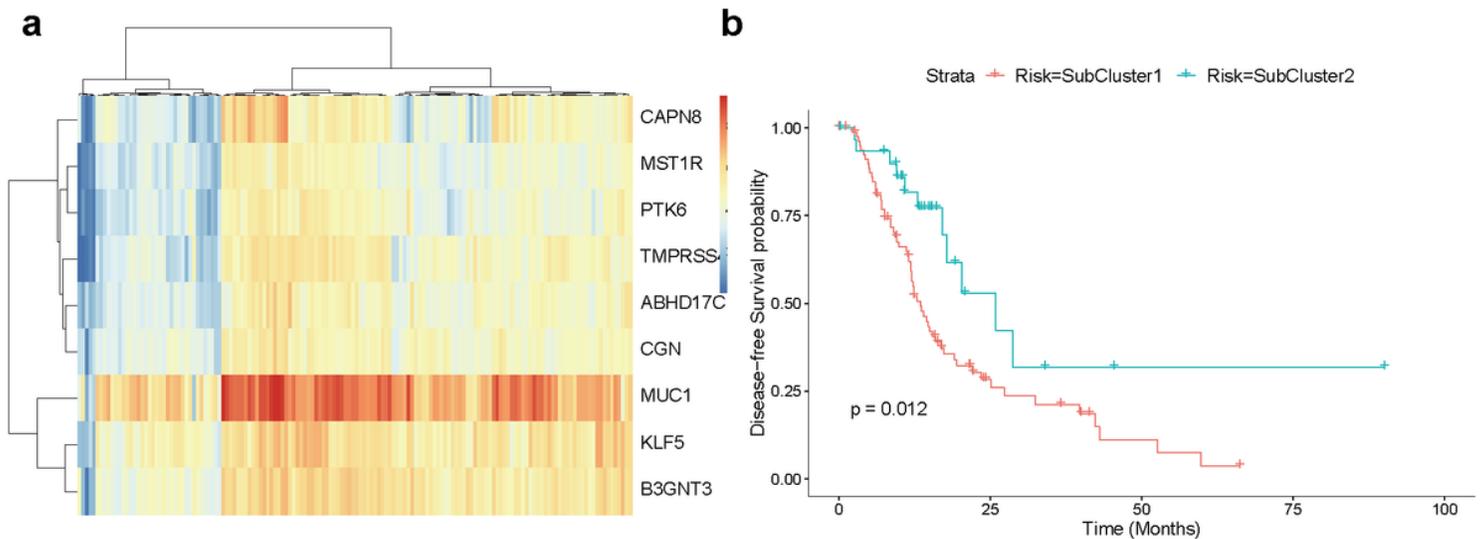


Figure 5

prognostic-related genes. a Expression of prognostic genes in early samples. b Prognostic difference between two subgroups

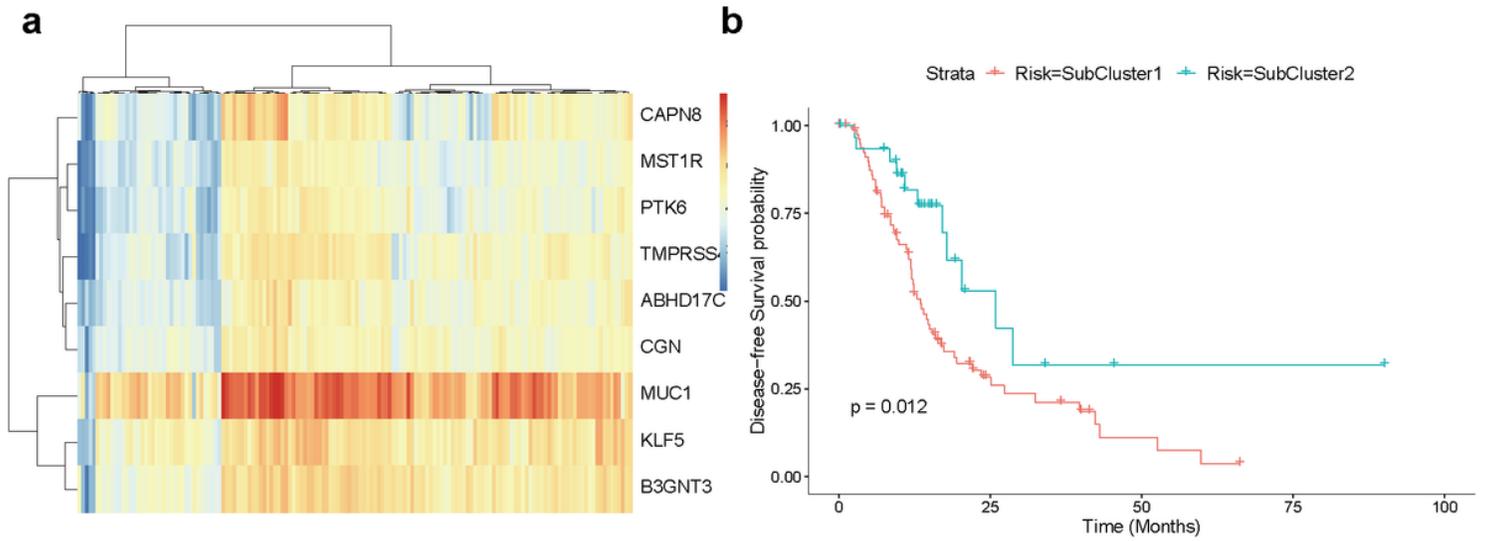


Figure 5

prognostic-related genes. a Expression of prognostic genes in early samples. b Prognostic difference between two subgroups

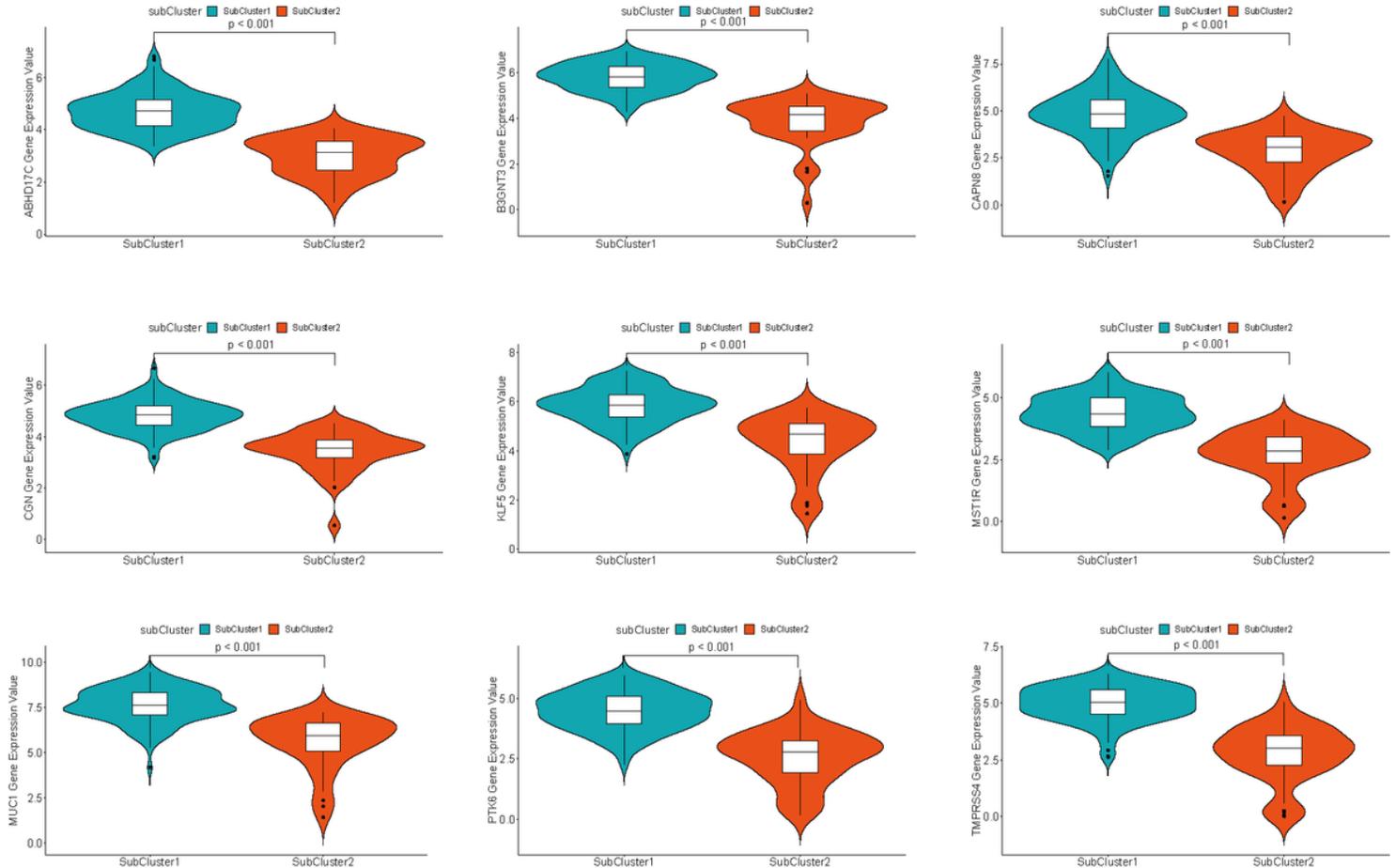


Figure 6

genes differentially expressed in two subgroups.

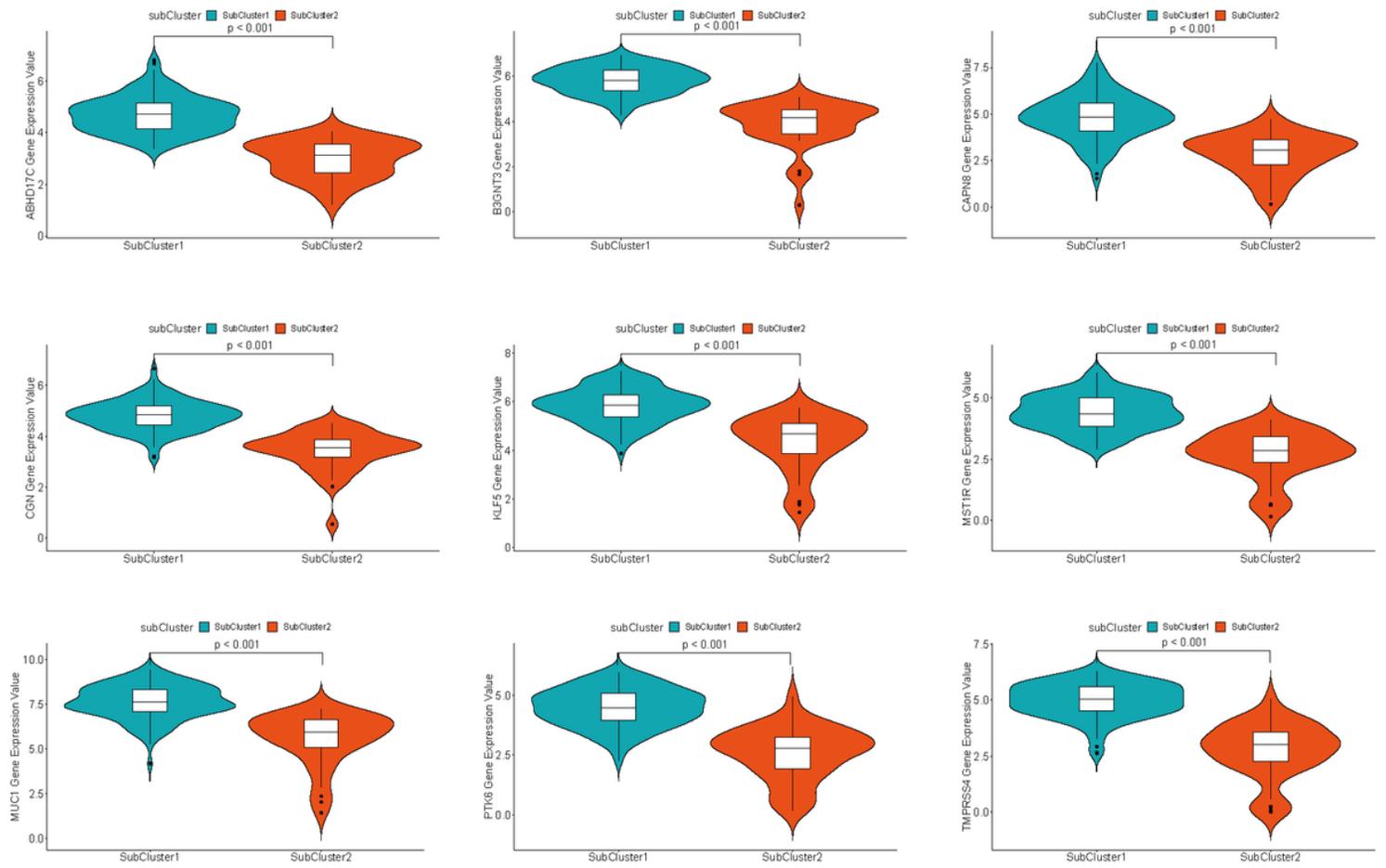


Figure 6

genes differentially expressed in two subgroups.

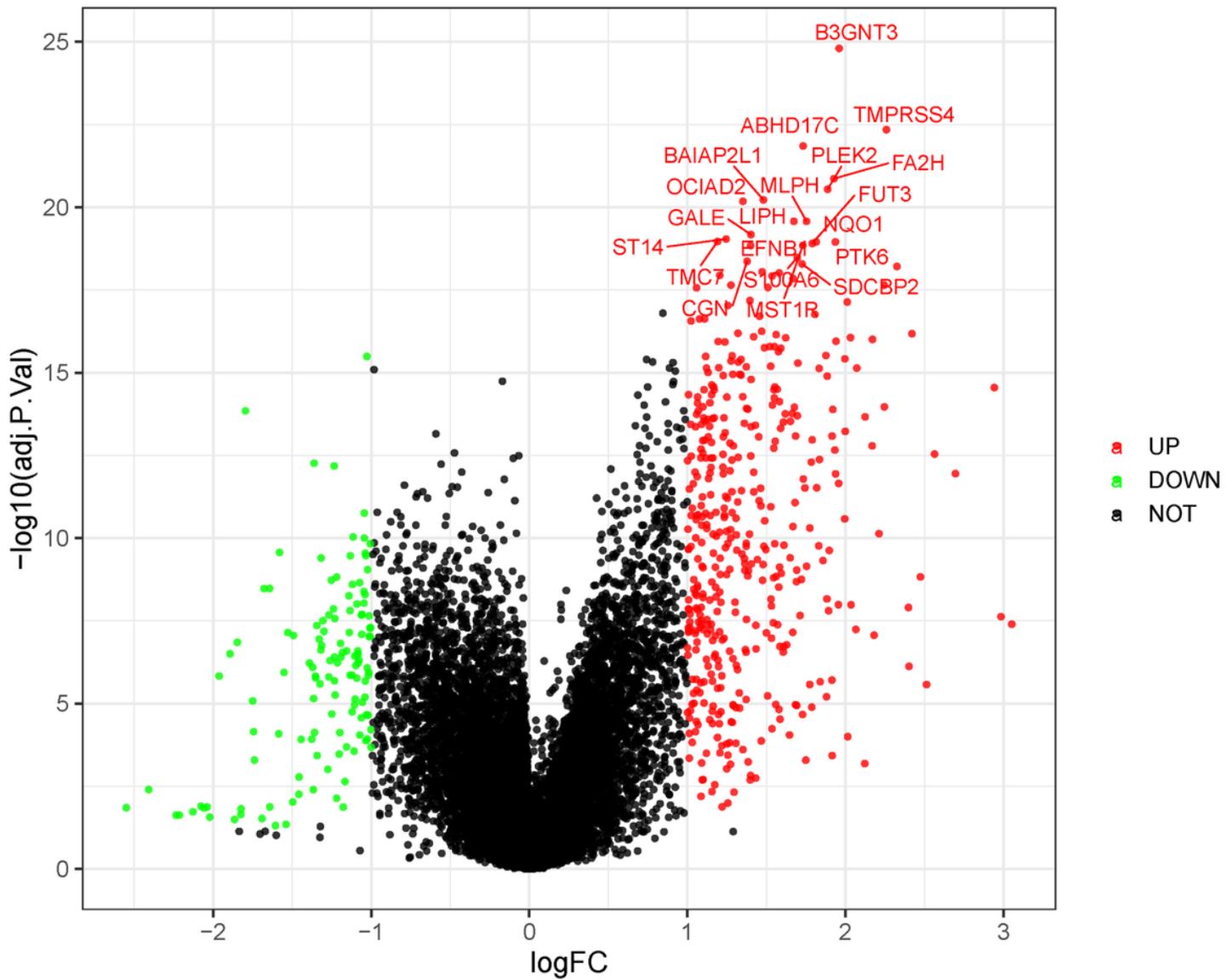


Figure 7

Volcano map of differentially expressed genes in two types of early pancreatic cancer

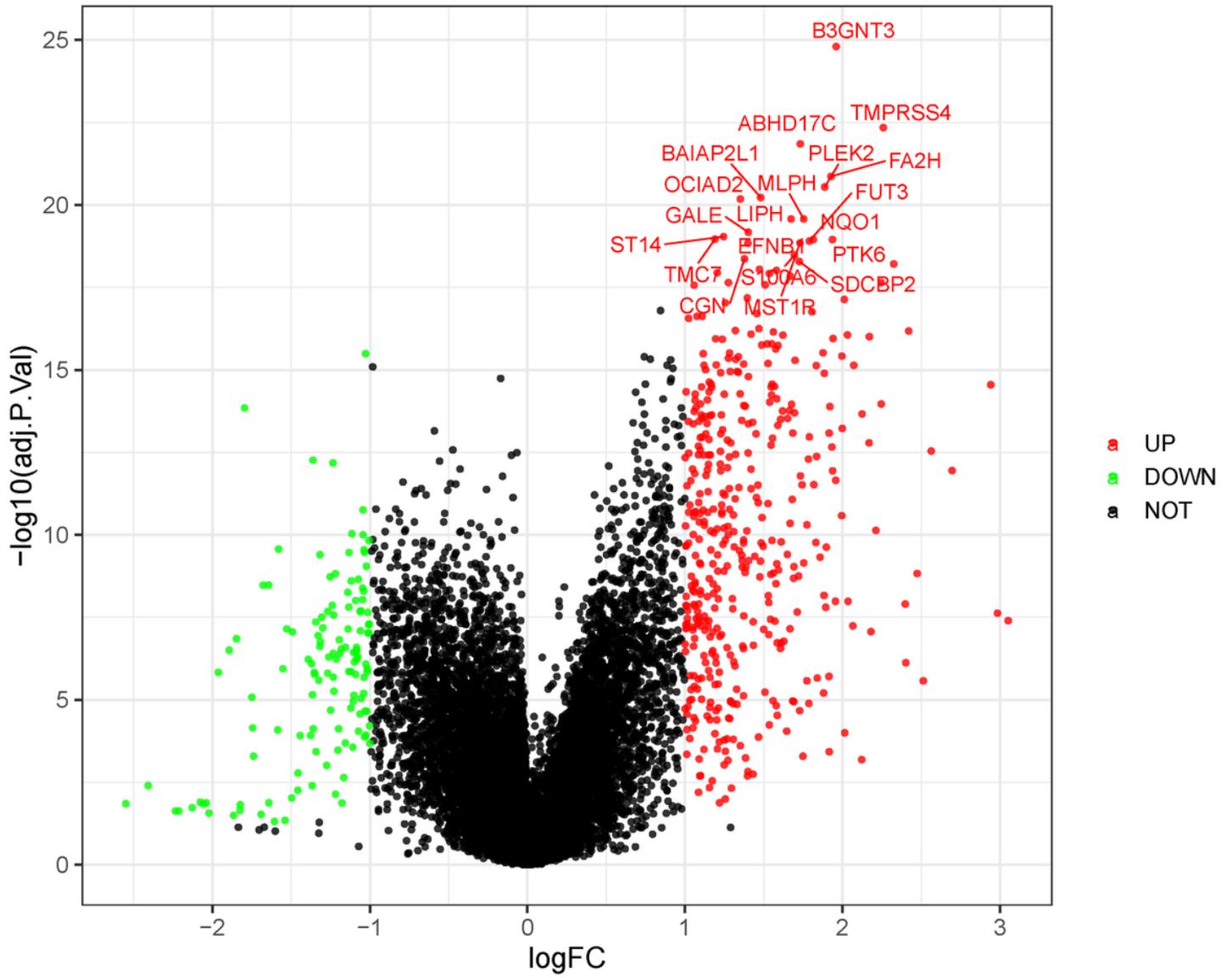


Figure 7

Volcano map of differentially expressed genes in two types of early pancreatic cancer

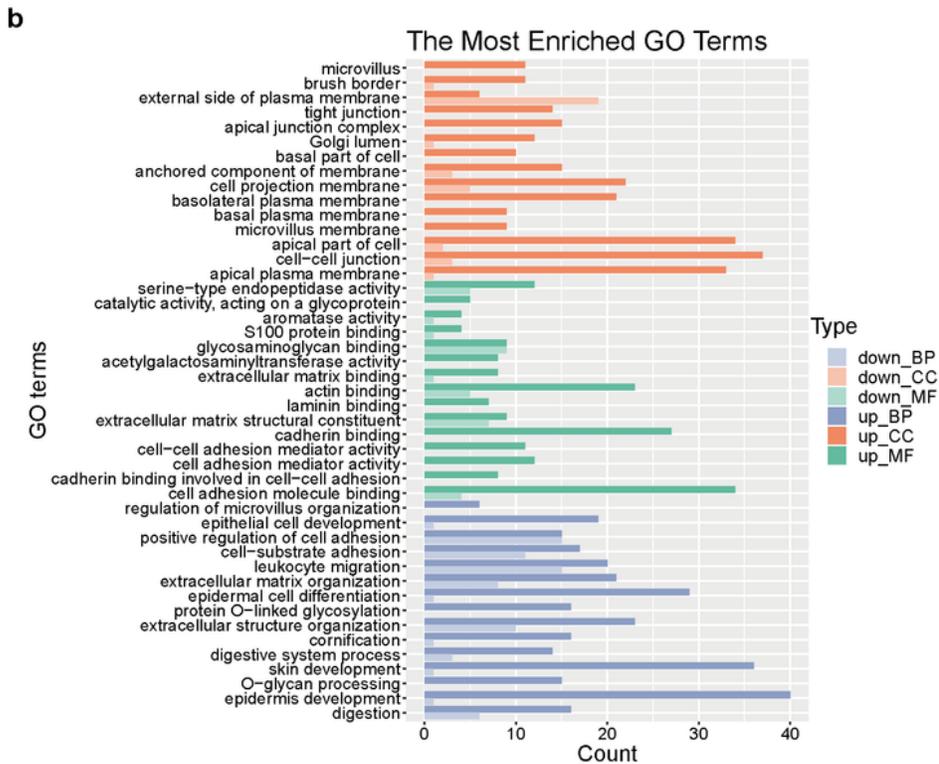
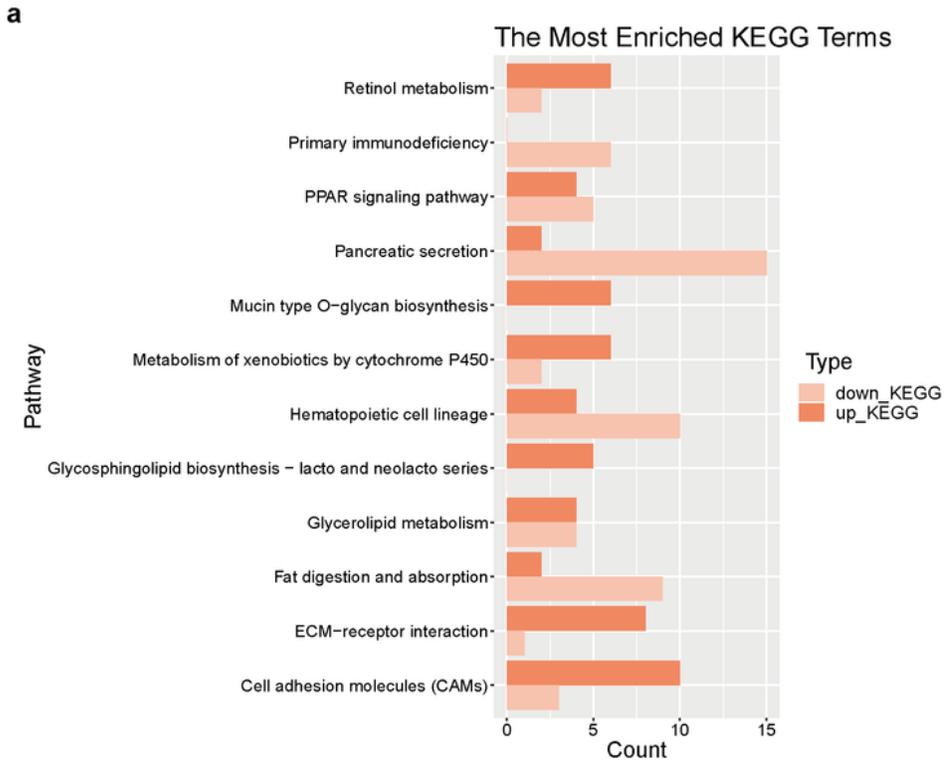


Figure 8

Enrichment analysis. a KEGG enrichment analysis. b GO enrichment analysis

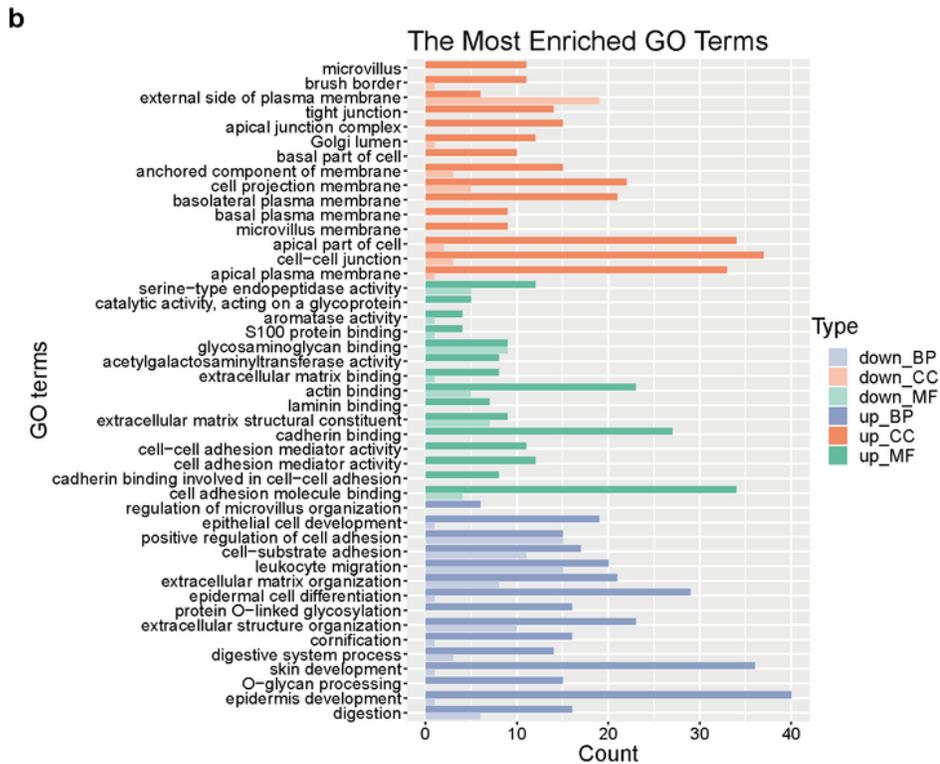
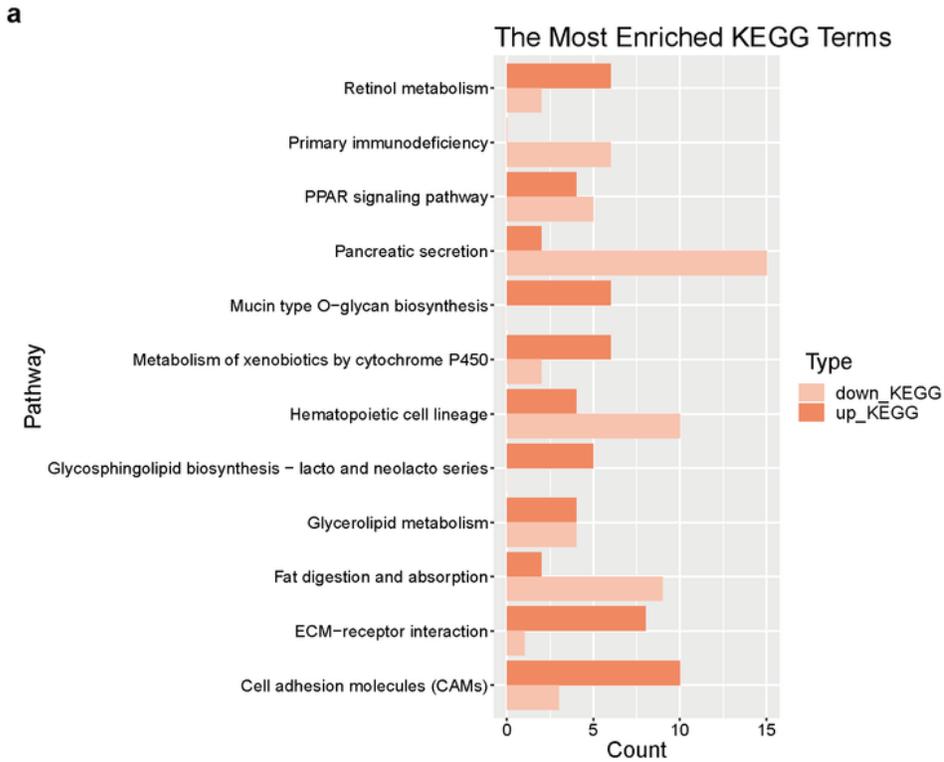


Figure 8

Enrichment analysis. a KEGG enrichment analysis. b GO enrichment analysis

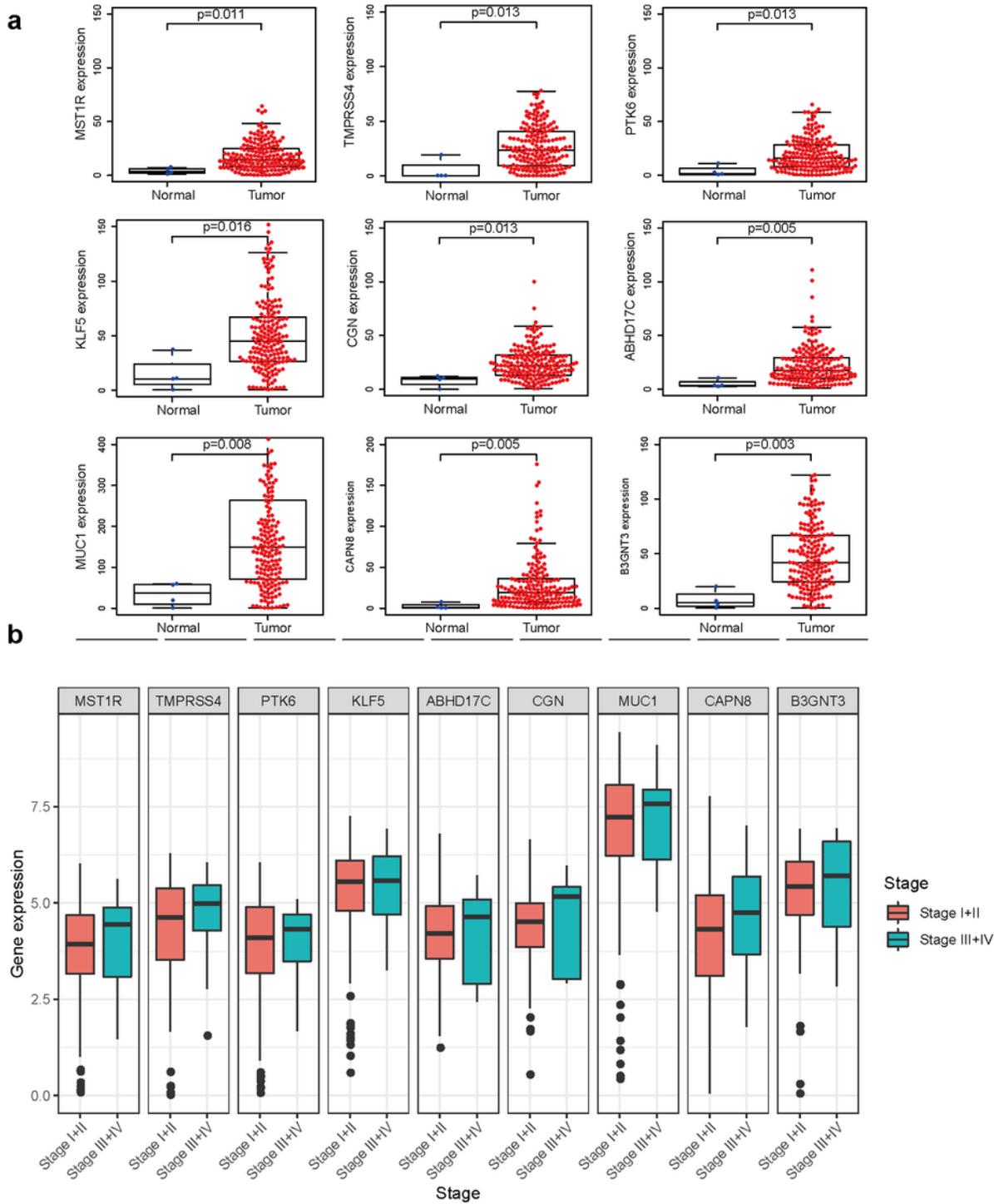


Figure 9

The expression level of 9 key genes. a Relationships between the 9 genes and the cancer status among PC patients. b Expression of four genes by grade in PC patients.

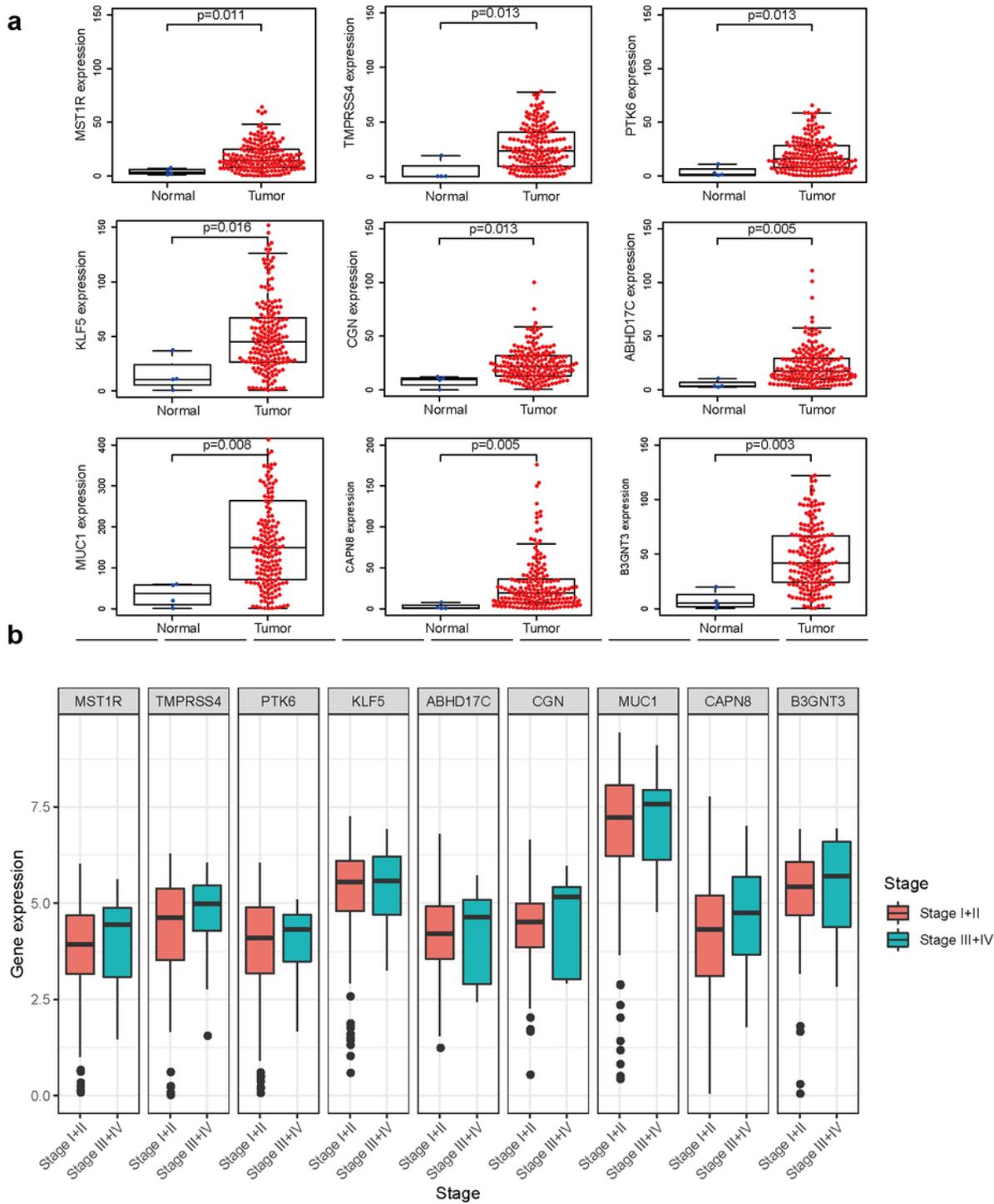


Figure 9

The expression level of 9 key genes. a Relationships between the 9 genes and the cancer status among PC patients. b Expression of four genes by grade in PC patients.

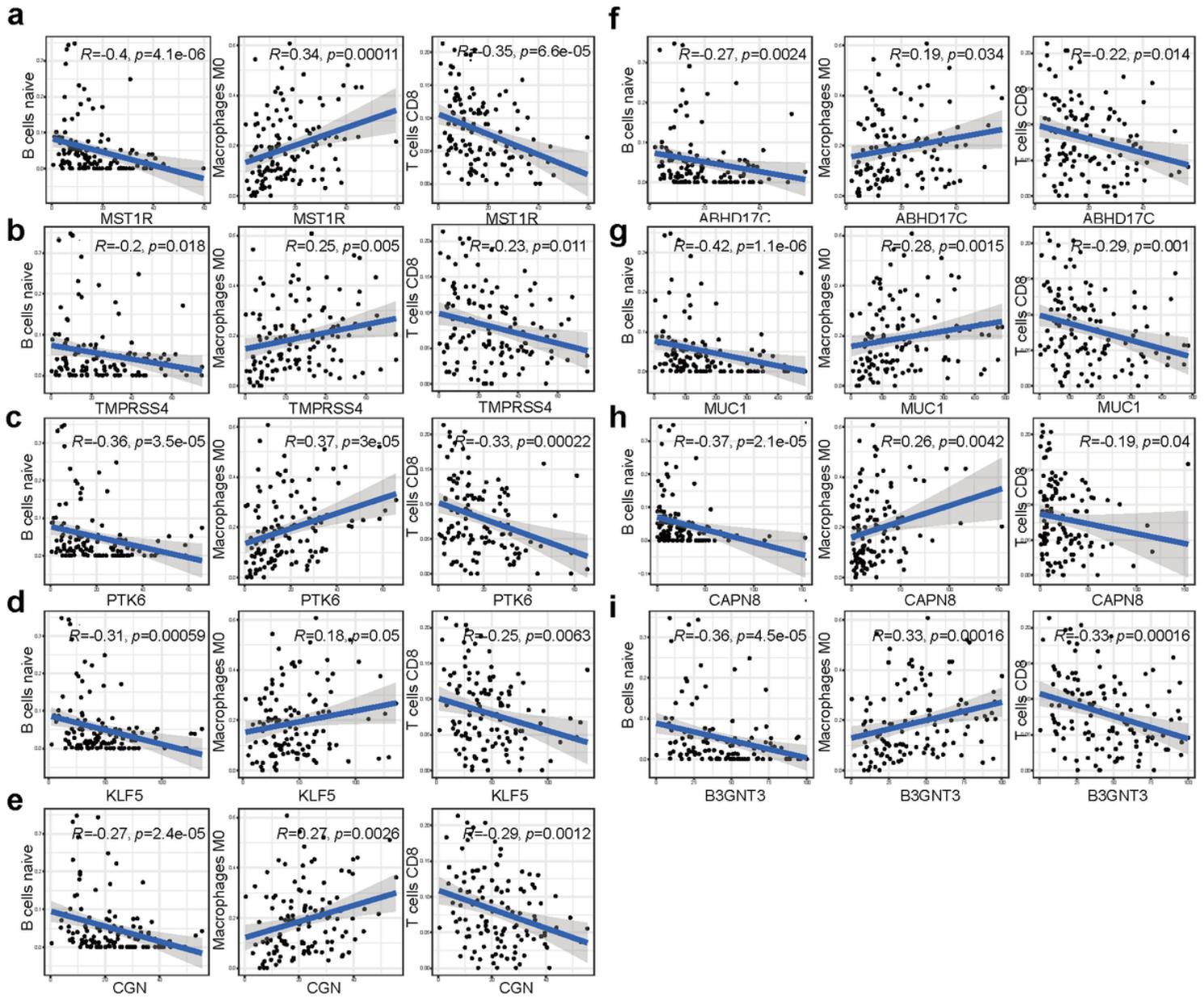


Figure 10

Immune infiltration of 9 genes

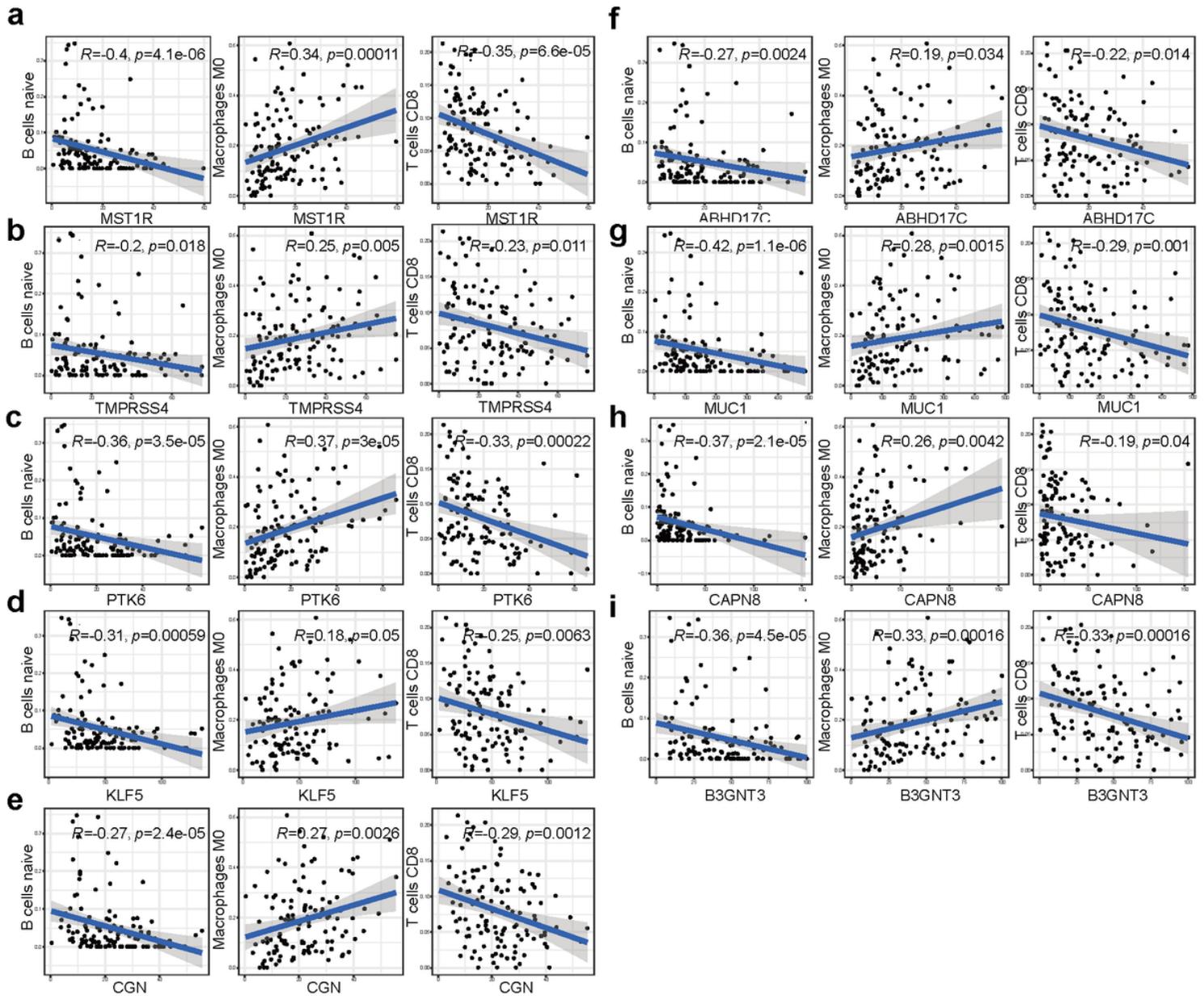


Figure 10

Immune infiltration of 9 genes

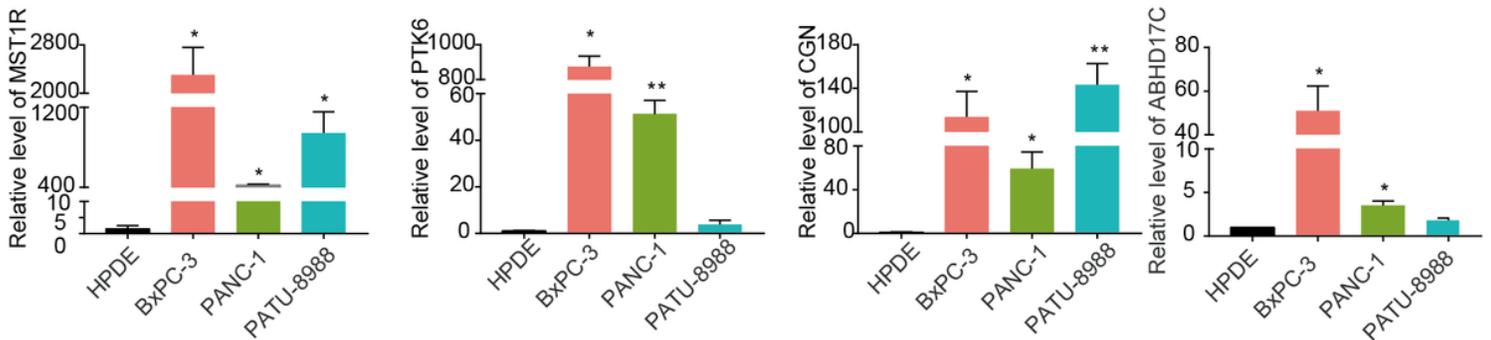


Figure 11

The expression level of MST1R, PTK6, ABHD17C and CGN in cancer cell lines and pancreatic epithelial HPDE lines

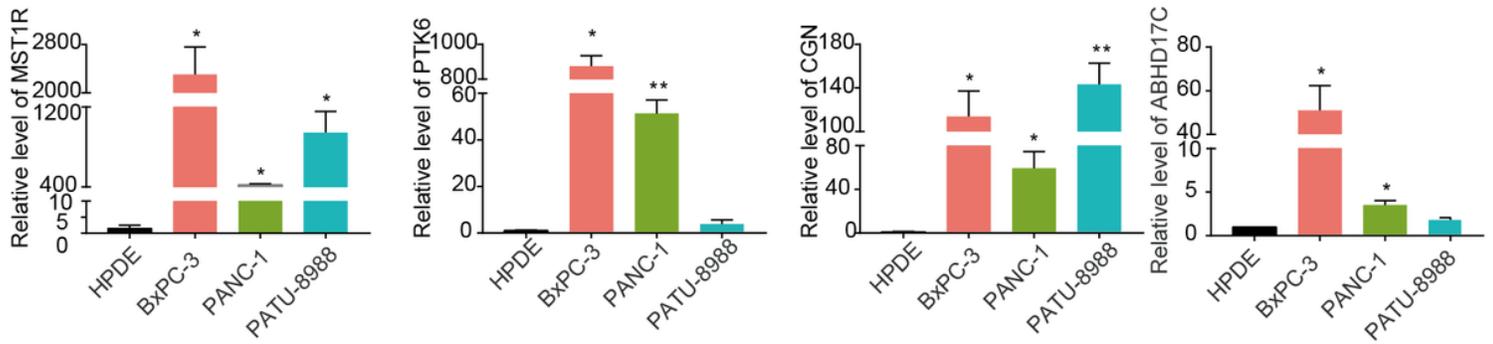


Figure 11

The expression level of MST1R, PTK6, ABHD17C and CGN in cancer cell lines and pancreatic epithelial HPDE lines