The identification of a potential prognostic model of colon cancer using integrated bioinformatics analysis

Zhengyu Fang
the first affiliated Hospital of Zhejiang Chinese Medical University
Sumei Xu (xsmdoctor@163.com)
The First Affiliated Hospital of Zhejiang Chinese Medical University https://orcid.org/0000-0003-2805-6295
Yiwen Xie
The first affiliated hospital of Zhejiang Chinese Medical University
Wenxi Yan
The first affiliated hospital of Zhejiang Chinese Medical University

Research article

Keywords: weighed gene co-expression network analysis, Meta-analysis, prognostic model, colon cancer

Posted Date: July 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-39410/v1

Abstract

Background

This study aimed to construct prognostic model by screening prognostic gene signature of colon cancer.

Methods

The gene expression profile data of colon cancer were obtained from The Cancer Genome Atlas (TCGA) and gene expression omnibus (GEO) and differently expressed genes (DEGs) between tumor and control samples were identified. Prognosis-associated genes were then identified and used for the construction of prognostic model. The independent factors that associated with the prognosis of colon in the TCGA cohort was identified.

Results

Totally, 1153 consistent DEGs were screened out between tumor and normal tissues in the TCGA cohort, GSE44861 and GSE44076 datasets. Among these genes, 12 DEGs were related to the prognosis of colon cancer and were used for constructing the prognostic model. This model presented a high predictive power for the prognosis of colon cancer both in the training dataset and in the validation datasets (AUC > 0.8). Statistical analysis showed that age, pathological T, tumor recurrence, and model status were the independent factors for prognosis of patients with colon cancer in TCGA.

Conclusions

The 12-gene signature prognostic model had a high predictive power for colon cancer prognosis.

Introduction

As one of the most common gastrointestinal malignant diseases, colon cancer is the world-wide leading cause of mortality (1). Currently, the standard therapeutic method for colon cancer is the combination of surgery and adjuvant chemotherapy or radiation therapy (2). Additionally, the early diagnosis for primary or recurrent colon cancer is also a critical factor for the prognosis of patients (3). Nowadays, studies have reported the intensive progress in the diagnosis and treatment of colon cancer, such as endoscopic diagnosis, tumor markers, and molecular targeted therapy; However, due to complex pathogenesis and higher metastasis, unsatisfactory diagnosis and poor prognosis still exist (2, 4). Therefore, identification of novel diagnostic, prognostic biomarkers and therapeutic targets, as well as investigation of the underlying molecular mechanism of colon cancer is required.

Nowadays, the revolution of sequencing technologies facilitates to the identification of more novel biomarkers related to diseases by bioinformatics analysis, which contributes to the early recognition and targeted treatment for diseases (5, 6). Dalerba P *et al.* (7) have emphasized that the lack of transcription factor CDX2 is associated with the poor prognosis in patients with colon cancer at stage II or stage III by bioinformatics approach. Another research group has shown that higher tumoral LC3B and p62 expression indicate an unfavorable prognosis by regulating autophagy in colon cancer (8). Demirkol S *et al.* (9) have identified two novel biomarkers, ULBP2 and SEMA5A, to predict the prognosis colon cancer based on gene expression omnibus (GEO) database. Additionally, it has been reported that vascular endothelial growth factor (VEGF)-D and SMAD7 are potential biomarkers for the chemotherapeutic outcome and prognosis of colon cancer (10). Notably, Yang et al. (11) performed a bioinformatics analysis on the basis of gene expression profile of GSE44076 about colon cancer. Consequently, there were 20 hub genes, such as TIMP1, GNG4, CXCL5 and COL1A1, were considered as diagnosis and treatment targets for colon cancer. However, few studies have undertaken the prognostic model based on an integrated analysis with different colon cancer microarray profiles.

In the current work, an integrated bioinformatics analysis based on The Cancer Genome Atlas (TCGA) and four gene expression profiles from GEO database were performed to screen the differentially expressed genes (DEGs) related to prognosis of colon cancer using MetaDE. Meanwhile, gene modules related to colon cancer were constructed by weighed gene co-expression network analysis (WGCNA) algorithm. Afterwards, a prognostic model of colon cancer was constructed. Meanwhile, the related prognostic clinical factors, and pathways associated with prognostic gene were analyzed.

Materials And Methods

Data extraction

Based on the search words “colon cancer”, the public expression profiles data were preliminarily extracted from the GEO repository (https://www.ncbi.nlm.nih.gov/geo/). Then, the eligible data were included in this study as the following inclusion criteria: (1) gene expression profiles data; (2) the samples in datasets were solid tissues of colon cancer; (3) human expression profiles; (4) the datasets contained control tissues; (5) the total number of samples was not less than 100; and (6) the datasets contained prognosis information of samples. As a result, the two eligible datasets (GSE44861 and GSE44076) that met inclusion criteria 1 to 5 was obtained, and contained 1111 samples (56 tumor samples and 55 normal samples) and 246 samples (98 tumor samples and 148 normal samples), respectively, based on the platform of Affymetrix-GPL96 and Affymetrix-GPL13667. GSE44861 and GSE44076 were utilized for the following the WGCNA and MetaDE analysis. Meanwhile, the two eligible datasets (GSE17538 and GSE38832) that met inclusion criteria 1 to 6 was obtained, and contained 244 and 122 colon cancer tumor samples, respectively, based on the platform of Affymetrix-GPL570. GSE17538 and GSE38832 were used for the construction of prognostic prediction model.

In addition, transcriptome RNA expression data of colon cancer were downloaded from TCGA (https://gdc-portal.nci.nih.gov/) and contained 512 colon cancer samples, based on the platform of Illumina HiSeq 2000 RNA Sequencing. Then corresponding to clinical information downloaded at the same time, 495 samples were reserved containing 454 tumor samples and 41 normal samples.

Screening of colon cancer related gene module

WGCNA has widely applied into identifying the gene module associated with diseases and extracting potential therapeutic targets (12). Based on TCGA, GSE44861, and GSE44076 datasets, WGCNA software (version 1.61; https://cran.r-project.org/web/packages/WGCNA/index.html) (13) in R3.4.1 was used to screen the stable gene module related to colon cancer. The TCGA data was utilized as the training set, while GSE44861 and GSE44076 was set as the validation set. WGCNA network was constructed according to the following steps: calculating the expression correlation between three datasets; defining adjacent function; dividing gene module and assessing module stability. The thresholds of module division were the number of genes in each module ≥ 150 and cutHeight = 0.99.

DEGs identification by meta-analysis

The consistently DEGs were extracted from TCGA, GSE44861, and GSE44076 datasets using MetaDE.ES in MetaDE package (https://cran.r-project.org/web/packages/MetaDE/) (14, 15). Briefly, the heterogeneity test of expression value of each gene from different platform was first conducted according to the statistics such as tau2, Q value and Q pval. To be specific, if tau^2 was 0, there was no bias among different subjects; if Q value complied with chi-square test with K-1 freedom and Q pval was more than 0.05, the study subjects would be homogeneous without bias. Then gene expression difference from different cases in the integrated dataset was also evaluated and corresponding P value and false discovery rate (FDR) were computed. FDR < 0.05 represented the remarkable difference. Finally, the fold change (FC) of each dataset was determined. Herein, the cutoffs of consistently DEGs identification were set as $tau^2 = 0$, Q pval > 0.05, FDR < 0.05 and |logFC| with consistent differential direction among the three datasets.

Construction and evaluation of prognostic risk model

Firstly, the intersected genes between genes of gene module by WGCNA and consistently DEGs byMetaDE were obtained. Then, functional analyses of these intersected genes, including the Gene Ontology (GO) functional annotation in terms of biological process category (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, were performed by DAVID (version 6.8; https://david.ncifcrf.gov/) (16, 17) using P value < 0.05 as the threshold of significant enrichment. Combined with the clinical prognostic information in the training set (TCGA) and the intersected genes, the independent prognostic genes were identified using univariate and multivariate cox regression analysis of survival package (version 2.4, https://cran.r-project.org/web/packages/survival/index.html) (18) in R3.4.1 based on the threshold of log-rank p value < 0.05. Afterwards, the optimized set of prognostic gene signature was identified using the Cox-Proportional Hazards (Cox-PH) model (19), which based on the L1-penalized regulatization regreesion algotithm of the penalized package (version 0.9–50,

http://bioconductor.org/packages/penalized/) (20) in R3.4.1. The optimized parameter lambda in this model was obtained by 1000 cycles calculation of cross-validation likelihood (cvl) algorithm. Subsequently, the prognostic score for each sample was calculated as follows: prognostic score = $\sum \beta_{genes} \times Exp_{genes}$. The β_{genes} represented prognostic regression coefficient and Exp_{genes} was defined as the expression value of corresponding gene. According to the median value of PS, all samples in the training set were divided into high risk and low risk groups. The Kaplan-Meier (K-M) curve analysis based on survival package (version 2.41-1) in R 3.4.1 was used to assess the association between high and low risk groupings and actual survival prognostic information. Meanwhile, the association between high and low risk groupings and actual survival prognostic information was further verified in the validation set the prognostic model (GSE17538 and GSE38832) using the above methods.

Further analysis of the prognostic clinical factors

The independent prognostic clinical factors based on the clinical information of tumor samples in the training set were analyzed using univariate and multivariate cox regression analysis of survival package (version 2.41-1) in R3.4.1 according to the threshold of log-rank p value < 0.05. In order to further study the relationship between independent prognostic clinical factors and risk grouping, the samples were divided into different groups based on hierarchical analysis of clinical prognostic factors, and the correlation analysis by prognostic risk model was performed in these different groups. Based on these independent prognostic clinical factors, the nomogram of 3- and 5-year survival prediction models were constructed using rms package (version 5.1-2, https://cran.r-project.org/web/packages/rms/index.html) (21) in R3.4.0.

Screening of genes related to prognostic risk grouping and functional analysis

The samples in the training set were divided into high risk and low risk groups according to prognostic score. Next, DEGs between high risk and low risk groups were screened using limma package (Version 3.34.7, https://bioconductor.org/packages/release/bioc/html/limma.html) (22) with the thresholds of false discovery rate (FDR) < 0.05 and |log fold change (FC)| >0.5. Following this, GO-BP analysis and KEGG pathway enrichment analysis were carried out using DAVID. P value < 0.05 was considered as the cutoffs for significantly statistical difference in functional analyses. A workflow of this study is shown in Fig. 1.

Results

Extraction of gene module related to psoriasis based on WGCNA algorithm

The correlation of expression level and node connection in TCGA, GSE44861, and GSE44076 datasets was evaluated, and the results suggested that there was a significant positive correlation and good comparability among these datasets (Figure S1A). Firstly, the scale-free network distribution was

determined based on the parameter of power = 7 as displayed in Figure S1B. A total of 8 modules related to colon cancer were established in the training datasets on the basis of clustering analysis (Fig. 2A). Then, the same module division was performed in other two validation datasets (GSE44861 and GSE44076) as showed in Fig. 2A. The module stability was also assessed according to the preservation Z score, and generally, Z value > 5 indicated a better module stability. Finally, 5 robust modules (blue, brown, green, red, and yellow) were obtained and showed markedly positive correlation with colon cancer. A total of 381 DEGs in blue module, 205 genes in brown module, 195 genes in green module, 184 genes in red module, and 195 genes in yellow module, were obtained (Table 1). Furthermore, the correlation of gene modules and clinical factors, including age, gender, history of colon polyps, lymphatic invasion, microsatellite instability, radiation therapy, death, tumor recurrence, pathologic M, pathologic N, pathologic T, and pathologic stage, were revealed as shown in Fig. 2B.

Table 1
Gene modules related to colon cancer based on weighed gene co-expression network analysis (WGCNA) algorithm

ID	Color	Module Size	Preservation	
			Z-score	Pvalue
Module 1	black	133	1.9913	1.40E-01
Module 2	**blue**	381	8.7017	4.50E-06
Module 3	**brown**	205	10.4907	4.00E-03
Module 4	**green**	195	8.2073	5.10E-03
Module 5	grey	2469	0.3400	2.30E-05
Module 6	**red**	184	10.9777	1.00E-03
Module 7	turquoise	649	4.0049	1.30E-02
Module 8	**yellow**	195	5.6788	2.00E-05

Identification of consistently DEGs based on Meta analysis

Totally, 1153 consistently DEGs were identified according to the thresholds mentioned in methods, including 724 consistently down-regulated DEGs and 429 consistently up-regulated DEGs, which showed consistent difference degree and dysregulation direction among TCGA, GSE44861, and GSE44076 datasets as shown in Fig. 3.

Construction and verification of prognostic model

Firstly, a total of 556 intersected genes between genes of gene module by WGCNA and consistently DEGs by MetaDE were obtained (Fig. 4A). The functional enrichment analyses of these intersected genes indicated that they were primarily participated in 24 significant GO-BP terms such as immune response

and the defense response (Fig. 4B), as well as closely associated with 8 KEGG pathways such as cytokine-cytokine receptor interaction and focal adhesion (Fig. 4B). Based on univariate Cox regression analysis in the training dataset, a total of 84 DEGs were significantly associated with the prognosis of patient with colon cancer. Then 14 independent prognostic DEGs in patients with colon cancer were obtained by multivariate Cox regression analysis. Afterwards, the optimized set of prognostic gene signature (including 12 DEGs, *ADORA3*, *CPA3*, *CPM*, *EDN3*, *FCRL2*, *MFNG*, *NAT1*, *PCSK5*, *PPARGC1A*, *PRRX2*, *TNFRSF17*, and *WDR78*) was identified using the Cox-PH model (Table 2). Based on the prognostic model, all samples in the training and validation dataset were divided into high risk and low risk groups, and the results showed significant correlation of risk grouping and actual survival prognostic information (Fig. 5).

Table 2
The optimized set of prognostic gene signature was identified using the Cox-Proportional Hazards (Cox-PH) model.

Symbol	Univariate Cox regression analysis			LASSO coefficient
	HR	95%CI	P value	
ADORA3	1.570	1.067–2.549	3.40E-02	0.44262
CPA3	0.810	0.679–0.965	9.50E-03	-0.35894
CPM	0.748	0.561–0.995	2.30E-02	-0.26349
EDN3	0.830	0.670–1.028	4.40E-02	-0.12557
FCRL2	2.465	1.298–4.682	2.90E-03	1.38523
MFNG	1.456	1.127–1.879	2.00E-03	0.35734
NAT1	0.514	0.368–0.717	4.55E-05	-0.42755
PCSK5	1.477	1.021–2.138	1.95E-02	0.30206
PPARGC1A	0.579	0.399–0.842	2.10E-03	-0.34355
PRRX2	1.260	1.017–1.559	1.70E-02	0.04376
TNFRSF17	0.780	0.597–0.919	3.45E-02	-0.21594
WDR78	0.334	0.158–0.707	2.05E-03	-0.07166

The independent prognostic clinical factors analysis

Univariate and multivariate cox regression analysis showed that age [hazard ratio (HR) = 1.018, 95% confidence interval (CI) 1.001–1.035, $P = 3.408 \times 10^{-2}$ and HR = 1.047, 95%CI 1.021–1.073, $P = 3.510 \times 10^{-4}$], pathological T (HR = 2.658, 95%CI 1.775–3.979, $P = 1.116 \times 10^{-6}$ and HR = 3.561, 95%CI 1.781–7.121, $P = 3.280 \times 10^{-4}$), tumor recurrence (HR = 2.567, 95%CI 1.636–4.029, $P = 2.113 \times 10^{-5}$ and HR = 1.881, 95%CI 1.050–3.369, $P = 3.363 \times 10^{-2}$), and prognostic model status (HR = 3.287, 95%CI 2.082–

5.189, P = 4.096×10^{-8} and HR = 2.737, 95%CI 1.447–5.178, P = 1.970×10^{-3}) were considered as the independent prognostic factors in patients with colon cancer (Table 3). In addition, K-M survival analysis also showed that lower age, lower pathological T and no tumor recurrence were associated with better prognosis of patients with colon cancer (Fig. 6A, left). Meanwhile, when samples were divided into based on hierarchical analysis of clinical prognostic factors, the results of prognostic model was consistent with actual survival prognostic information in these different groups (Fig. 6A, middle and right). Furthermore, the nomogram of 3- and 5-year survival prediction models of these independent prognostic factors were constructed as Fig. 6B. The nomogram of 3- and 5-year survival prediction showed compliance to actual 3- and 5-year survival (Fig. 6C).

Table 3
Univariate and multivariate cox regression analysis of colon cancer tumor samples.

Clinical characteristics	TCGA (N = 432)	Uni-variable cox		Multi-variable cox	
		HR (95% CI)	P value	HR (95% CI)	P value
Age (years, mean ± sd)	66.78 ± 12.88	1.018[1.001–1.035]	3.408E-02	1.047[1.021–1.073]	3.510E-04
Gender (Male/Female)	230/202	1.077[0.719–1.610]	7.189E-01	-	-
Pathologic M (M0/M1/-)	319/59/54	4.536[2.851–7.218]	2.649E-12	1.501[0.373–6.036]	5.671E-01
Pathologic N (N0/N1/N2)	254/101/77	2.088[1.648–2.644]	1.342E-10	1.614[0.839–3.103]	1.514E-01
Pathologic T(T1/T2/T3/T4)	11/75/296/50	2.658[1.775–3.979]	1.116E-06	3.561[1.781–7.121]	3.280E-04
Pathologic stage (I/II/III/IV/-)	73/167/123/59/10	2.181[1.719–2.767]	3.376E-11	1.123[0.373–3.378]	8.362E-01
Colon polyps history (Yes/No/-)	128/239/65	0.731[0.426–1.255]	2.537E-01	-	-
Lymphatic invasion (Yes/No/-)	150/241/41	2.150[1.392–3.320]	4.125E-04	0.922[0.489–1.737]	8.024E-01
Recurrence (Yes/No)	78/292/62	2.567[1.636–4.029]	2.113E-05	1.881[1.050–3.369]	3.363E-02
PS model status (High/Low)	216/216	3.287[2.082–5.189]	4.096E-08	2.737[1.447–5.178]	1.970E-03
Vital status (Dead/Alive)	96/336	-	-	-	-
Overall survival time (months,mean ± sd)	29.44 ± 25.43	-	-	-	-

Functional enrichment analysis of DEGs related to prognostic risk grouping

Based on the selective criteria, a total of 514 DEGs were identified between high risk and low risk groups, including 102 down-regulated and 412 up-regulated genes (Fig. 7A). Then, clustering analysis for these DEGs was conducted, indicating that these identified DEGs could significantly distinct high risk from and low risk groups (Fig. 7B). To further identify the functional characteristics of DEGs, the functional enrichment analyses of genes were conducted with DAVID. Consequently, the GO analysis of DEGs revealed 23 significant enriched terms that primarily concentrated on ion transport, cell-cell signaling, regulation of cyclic nucleotide metabolic process (Fig. 7C). In addition, the KEGG pathway analysis implied that these DEGs were responsible for 7 KEGG pathways, such as neuroactive ligand-receptor interaction, and calcium signaling pathway (Fig. 7C).

Discussion

In the present study, 5 significantly stable gene modules related to colon cancer were constructed by WGCNA algorithm. Then, 1153 consistently DEGs were identified between colon cancer tumor and normal tissues samples based on the TCGA, GSE44861 and GSE44076 datasets. Furthermore, based on the intersected genes between genes of gene module by WGCNA and consistently DEGs by MetaDE, 12 DEGs (*ADORA3*, *CPA3*, *CPM*, *EDN3*, *FCRL2*, *MFNG*, *NAT1*, *PCSK5*, *PPARGC1A*, *PRRX2*, *TNFRSF17*, and *WDR78*) related to prognosis of colon cancer were further isolated as the optimized prognostic gene signature, and a prognostic model was constructed based these 12 DEGs, which presented a relative highly forecast ability for the prognosis of colon cancer both in the training dataset and validation datasets. In addition, age, pathological T, tumor recurrence, and prognostic model status were identified as the independent prognostic factors in patients with colon cancer based on TCGA. Furthermore, based on the prognostic model, 514 DEGs related to prognosis of colon cancer were further identified, which were closely associated with ion transport, cell-cell signaling, regulation of cyclic nucleotide metabolic process, neuroactive ligand-receptor interaction, and calcium signaling pathway.

The mining of a large amount of genetic data in various diseases have been enhanced due to the rapid technological advances in high-throughput sequencing and bioinformatics (23). TCGA, as a public and available cancer genomic datasets, provides the comprehensive data of cancers, including mRNA expression data, miRNA expression data, copy number variation, DNA methylation, and clinical information (24). The data from TCGA have been effectively applied to improve diagnostic and therapeutic methods of cancers, as well as finally cancer prevention (24). Thus, this study was performed based on the gene expression profile data and clinical information of BC form TCGA and GEO database. Gene expression profiles have been reported to predict the prognosis outcome of cancers (25–27). Computationally, univariate and multivariate Cox regression were the most common method to construct the prognostic models and screen prognostic factors (28). In this study, the Cox regression model based on the LASSO, a semi-parametric proportional hazards model, was applied. The availability of this model

in survival analysis have been confirmed in recent studies (29, 30). Similarly, in this study, the prognostic model constructed by LASSO Cox regression model showed a higher predictive ability both in training and validation sets. In addition, this study showed that age pathological T, and tumor recurrence were independent prognostic factors in patients with colon cancer. Consistent with our results, previous studies have also demonstrated that advanced age, higher pathological T and tumor recurrence are associated with poor prognosis in patients with colon cancer (31–33). Notably, this study revealed that the results of the prognostic model were consistent with actual survival prognostic information in different groups based on hierarchical analysis of age, higher pathological T and tumor recurrence. Meanwhile, the model status was also been considered as an independent prognostic factor in patients with colon cancer. These results further showed that prognostic model had a significant predictive ability for the prognosis of colon cancer.

In this study, the prognostic model was constructed based on the 12-prognostic gene signature (including 12 DEGs, *ADORA3*, *CPA3*, *CPM*, *EDN3*, *FCRL2*, *MFNG*, *NAT1*, *PCSK5*, *PPARGC1A*, *PRRX2*, *TNFRSF17*, and *WDR78*). Specifically, adenosine receptor A3 (ADORA3) protein encoded by *ADORA3* gene is G-protein-coupled receptor that are implicated in inflammatory and immunological responses as well as cancer growth in various diseases through influencing nucleotide metabolic process (34–36). Increasing evidence has proved that ADORA3 is overexpressed in several cancers, including breast cancer (37), thyroid cancer (38), bladder cancer (39), and colon cancer (40) and functions as a tumor promoter (41). Carboxypeptidase A3 (CPA3) as a member of the CPA family of zinc metalloproteases is released by mast cells and may be involved in the inactivation of venom-associated peptides and the degradation of endogenous proteins (42). Previous study has shown elevated expression of CPA3 in asthma (43) and anaphylactic shock (44); however, few studies have investigated the role of CPA3 in cancers. CPM is also an arginine/lysine CP and exerts important roles in angiogenesis, proliferation, and apoptosis through modulating chemokines or kinins in cancer cells (45). Notably, recent study reports that CPM/Src-FAK pathway is involved in the cell migration and invasion in colon cancer (46). Endothelin 3 (END3) is reported to participate in the progression of several cancers, such as malignant melanoma (47), cervical cancer (48), and colon cancer (49). Fc Receptor Like 2 (FCRL2) is a member of the immunoglobulin receptor superfamily that is involved in the development of lymphoblastic leukemia by immunomodulators of B cell function (50–52). Inherited polymorphism in the acetyltransferase 1 gene (NAT1) increases the risk of colorectal adenocarcinoma (53). Manic fringe (MFNG) is reported to exhibit anti-tumor effects in lung cancer (54). Peroxisome proliferator-activated receptor-γ coactivator 1-α (PPARGC1A) can contribute to tumor growth and metastasis in several cancers (55, 56). In addition, studies have suggested that both paired related homeobox 2 (PRRX2) (57, 58) and tumor necrosis factor receptor superfamily member 17 (TNFRSF17) (59, 60) are associated with several cancers, while proprotein convertase subtilisin/kexin type 5 (PCSK5) and WD repeat domain 78 (WDR78) have not been reported to be involved in cancers. Thus, the functions of these genes in colon cancer should be further investigated.

Conclusions

In conlcusion, the prognostic model based on the prognostic 12-gene signature exhibited a relatively satisfactory predictive potential for colon cancer. Howerer, the prognostic significance of 12-gene signature in colon cancer should be futher confirmed in clinical study.

Abbreviations

Cox-PH, Cox-Proportional Hazards; DEGs, differently expressed genes; FC, fold change; FDR, false discovery rate; GEO, gene expression omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The Cancer Genome Atlas; WGCNA, weighed gene co-expression network analysis.

Declarations

Acknowledgements

None.

Conflict of interest

The authors declare that they have no competing interests.

Funding

This study was supported by the Chinese Medicine Science and Technology Plan of Zhejiang Province (2020ZA054; 2020ZB065) and the Medicine and Health Science and Technology Plan Projects in Zhejiang province (2019RC057).

Availability of data and material

GSE17538, GSE38832, GSE44861 and GSE44076 were preliminarily extracted from the NCBI GEO repository (https://www.ncbi.nlm.nih.gov/geo/). Additional transcriptome RNA expression data of colon cancer were downloaded from TCGA (https://gdc-portal.nci.nih.gov/). All data generated or analyzed during this study are included in this published article.

Author's contribution

Conception and design of the research: Zhengyu Fang and Sumei Xu. Acquisition, analysis and interpretation of data: Yiwen Xie and Wenxi Yan. Drafting the manuscript: Zhengyu Fang. Manuscript review: Sumei Xu, Yiwen Xie, Wenxi Yan Obtaining funding: Zhengyu Fang and Sumei Xu. All authors approved the final revision.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2019. CA: a cancer journal for clinicians. 2019;69(1):7–34.
2. Hashiguchi Y, Muro K, Saito Y, Ito Y, Ajioka Y, Hamaguchi T, et al. Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2019 for the treatment of colorectal cancer. International journal of clinical oncology. 2019:1–42.
3. Vatandoost N, Ghanbari J, Mojaver M, Avan A, Ghayour-Mobarhan M, Nedaeinia R, et al. Early detection of colorectal cancer: from conventional methods to novel biomarkers. J Cancer Res Clin Oncol. 2016;142(2):341–51.
4. Sanoff HK, Sargent DJ, Campbell ME, Morton RF, Fuchs CS, Ramanathan RK, et al. Five-year data and prognostic factor analysis of oxaliplatin and irinotecan combinations for advanced colorectal cancer: N9741. J Clin Oncol. 2008;26(35):5721.
5. Ramos CJB, Cortés GA, Pulido AJP. Use of bioinformatics tools to find new genes involved in rare diseases. Biosaia: Revista de los másteres de Biotecnología Sanitaria y Biotecnología Ambiental, Industrial y Alimentaria. 2019(8).
6. Forman MR, Greene SM, Avis NE, Taplin SH, Courtney P, Schad PA, et al. Bioinformatics: Tools to accelerate population science and disease control research. Am J Prev Med. 2010;38(6):646–51.
7. Dalerba P, Sahoo D, Paik S, Guo X, Yothers G, Song N, et al. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. N Engl J Med. 2016;374(3):211–22.
8. Niklaus M, Adams O, Berezowska S, Zlobec I, Graber F, Slotta-Huspenina J, et al. Expression analysis of LC3B and p62 indicates intact activated autophagy is associated with an unfavorable prognosis in colon cancer. Oncotarget. 2017;8(33):54604.
9. Demirkol S, Gomceli I, Isbilen M, Dayanc BE, Tez M, Bostanci EB, et al. A combined ULBP2 and SEMA5A expression signature as a prognostic and predictive biomarker for colon cancer. J Cancer. 2017;8(7):1113.
10. Su F, Li X, You K, Chen M, Xiao J, Zhang Y, et al. Expression of VEGF-D, SMAD4, and SMAD7 and their relationship with lymphangiogenesis and prognosis in colon cancer. J Gastrointest Surg. 2016;20(12):2074–82.
11. Yang W, Ma J, Zhou W, Li Z, Zhou X, Cao B, et al. Identification of hub genes and outcome in colon cancer based on bioinformatics analysis. Cancer management research. 2019;11:323.
12. Zhai X, Xue Q, Liu Q, Guo Y, Chen Z. Colon cancer recurrence–associated genes revealed by WGCNA co–expression network analysis. Mol Med Rep. 2017;16(5):6499–505.

13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9(1):559—.
14. Qi C, Hong L, Cheng Z, Yin Q. Identification of metastasis-associated genes in colorectal cancer using metaDE and survival analysis. Oncology letters. 2016;11(1):568–74.
15. Wang X, Kang DD, Shen K, Song C, Lu S, Chang L-C, et al. An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. Bioinformatics. 2012;28(19):2534–6.
16. Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44–57.
17. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2008;37(1):1–13.
18. Wang P, Wang Y, Hang B, Zou X, Mao J-H. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget. 2016;7(34):55343.
19. Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in medicine. 1997;16(4):385–95.
20. Goeman JJ. L1 penalized estimation in the Cox proportional hazards model. Biometrical journal. 2010;52(1):70–84.
21. Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. Oncotarget. 2015;6(34):36308.
22. 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 research. 2015;43(7):e47-e.
23. Wee TT, Cheng-yu LE. Beginners Guide to Bioinformatics for High Throughput Sequencing: World Scientific; 2018.
24. Hutter C, Zenklusen JC. The cancer genome atlas: creating lasting value beyond its data. Cell. 2018;173(2):283–5.
25. Kessous R, Octeau D, Klein K, Tonin PN, Greenwood CM, Pelmus M, et al. Distinct homologous recombination gene expression profiles after neoadjuvant chemotherapy associated with clinical outcome in patients with ovarian cancer. Gynecol Oncol. 2018;148(3):553–8.
26. O'Mara TA, Zhao M, Spurdle AB. Meta-analysis of gene expression studies in endometrial cancer identifies gene expression profiles associated with aggressive disease and patient outcome. Scientific reports. 2016;6:36677.
27. McConkey DJ, Choi W, Shen Y, Lee I-L, Porten S, Matin SF, et al. A prognostic gene expression signature in the molecular classification of chemotherapy-naive urothelial cancer is predictive of clinical outcomes from neoadjuvant chemotherapy: a phase 2 trial of dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin with bevacizumab in urothelial cancer. European urology. 2016;69(5):855–62.

28. Bao Z, Zhang W, Dong D. A potential prognostic lncRNA signature for predicting survival in patients with bladder urothelial carcinoma. Oncotarget. 2017;8(6):10485.
29. Ching T, Zhu X, Garmire LX. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. PLoS Comput Biol. 2018;14(4):e1006076.
30. Liang R, Wang M, Zheng G, Zhu H, Zhi Y, Sun Z. A comprehensive analysis of prognosis prediction models based on pathway–level, gene–level and clinical information for glioblastoma. Int J Mol Med. 2018;42(4):1837–46.
31. Di Fabio F, Nascimbeni R, Villanacci V, Baronchelli C, Bianchi D, Fabbretti G, et al. Prognostic variables for cancer-related survival in node-negative colorectal carcinomas. Dig Surg. 2004;21(2):128–33.
32. De Leon MP, Sant M, Micheli A, Sacchetti C, Gregorio CD, Fante R, et al. Clinical and pathologic prognostic indicators in colorectal cancer. A population-based study. Cancer. 1992;69(3):626–35.
33. Roth AD, Delorenzi M, Tejpar S, Yan P, Klingbiel D, Fiocca R, et al. Integrated analysis of molecular and clinical prognostic factors in stage II/III colon cancer. J Natl Cancer Inst. 2012;104(21):1635–46.
34. Jacobson KA, Merighi S, Varani K, Borea PA, Baraldi S, Aghazadeh Tabrizi M, et al. A3 adenosine receptors as modulators of inflammation: from medicinal chemistry to therapy. Medicinal research reviews. 2018;38(4):1031–72.
35. Cohen S, Fishman P. Targeting the A3 adenosine receptor to treat cytokine release syndrome in cancer immunotherapy. Drug Des Devel Ther. 2019;13:491.
36. Gessi S, Merighi S, Borea PA, Cohen S, Fishman P. Adenosine Receptors and Current Opportunities to Treat Cancer. The Adenosine Receptors: Springer; 2018. pp. 543–55.
37. Jafari SM, Panjehpour M, Aghaei M, Joshaghani HR, Enderami SE. A3 adenosine receptor agonist inhibited survival of breast cancer stem cells via GLI-1 and ERK1/2 pathway. Journal of cellular biochemistry. 2017;118(9):2909–20.
38. Morello S, Petrella A, Festa M, Popolo A, Monaco M, Vuttariello E, et al. Cl-IB-MECA inhibits human thyroid cancer cell proliferation independently of A3 adenosine receptor activation. Cancer Biol Ther. 2008;7(2):278–84.
39. Cao H-L, Liu Z-J, Chang Z. Cordycepin induces apoptosis in human bladder cancer cells via activation of A3 adenosine receptors. Tumor Biology. 2017;39(7):1010428317706915.
40. Gessi S, Cattabriga E, Avitabile A, Lanza G, Cavazzini L, Bianchi N, et al. Elevated expression of A3 adenosine receptors in human colorectal cancer is reflected in peripheral blood cells. Clin Cancer Res. 2004;10(17):5895–901.
41. Marucci G, Santinelli C, Buccioni M, Navia AM, Lambertucci C, Zhurina A, et al. Anticancer activity study of A3 adenosine receptor agonists. Life sciences. 2018;205:155–63.
42. Springman EB. Mast cell carboxypeptidase. Handbook of Proteolytic Enzymes: Elsevier; 2004. p. 828–30.

43. Abadalkareem R, Lau LC, Abdelmotelb A, Zhou X, Eren E, Walls AF. Mast cell tryptase and carboxypeptidase A3 (CPA3) as markers for predicting susceptibility to severe allergic drug reactions. Journal of Allergy Clinical Immunology. 2017;139(2):AB39.
44. Yang K, Guo X, Yan X, Gao C. Changes of prostaglandin D2, carboxypeptidase A3 and platelet activating factor in guinea pig in anaphylactic shock. Fa yi xue za zhi. 2012;28(3):175–8.
45. Denis CJ, Lambeir A-M. The potential of carboxypeptidase M as a therapeutic target in cancer. Expert Opin Ther Targets. 2013;17(3):265–79.
46. Lu D, Yao Q, Zhan C, Le-Meng Z, Liu H, Cai Y, et al. MicroRNA-146a promote cell migration and invasion in human colorectal cancer via carboxypeptidase M/src-FAK pathway. Oncotarget. 2017;8(14):22674.
47. Tang L, Su M, Zhang Y, Ip W, Martinka M, Huang C, et al. Endothelin-3 is produced by metastatic melanoma cells and promotes melanoma cell survival. J Cutan Med Surg. 2008;12(2):64–70.
48. Sun DJ, Liu Y, Lu DC, Kim W, Lee JH, Maynard J, et al. Endothelin-3 growth factor levels decreased in cervical cancer compared with normal cervical epithelial cells. Human pathology. 2007;38(7):1047–56.
49. Olender J, Nowakowska-Zajdel E, Kruszniewska-Rajs C, Orchel J, Mazurek U, Wierzgoń A, et al. Epigenetic silencing of endothelin-3 in colorectal cancer. Int J ImmunoPathol Pharmacol. 2016;29(2):333–40.
50. Ehrhardt GR, Leu C-M, Zhang S, Aksu G, Jackson T, Haga C, et al. Fc Receptor–like Proteins (FCRL): Immunomodulators of B Cell Function. Mechanisms of Lymphocyte Activation and Immune Regulation XI: Springer; 2007. p. 155 – 62.
51. Kazemi T, Asgarian-Omran H, Memarian A, Shabani M, Sharifian RA, Vossough P, et al. Low representation of Fc receptor-like 1–5 molecules in leukemic cells from Iranian patients with acute lymphoblastic leukemia. Cancer immunology immunotherapy. 2009;58(6):989.
52. Kazemi T, Asgarian-Omran H, Hojjat-Farsangi M, Shabani M, Memarian A, Sharifian RA, et al. Fc receptor-like 1–5 molecules are similarly expressed in progressive and indolent clinical subtypes of B-cell chronic lymphocytic leukemia. International journal of cancer. 2008;123(9):2113–9.
53. Katoh T, Boissy R, Nagata N, Kitagawa K, Kuroda Y, Itoh H, et al. Inherited polymorphism in the N-acetyltransferase 1 (NAT1) and 2 (NAT2) genes and susceptibility to gastric and colorectal adenocarcinoma. International journal of cancer. 2000;85(1):46–9.
54. Yi F, Amarasinghe B, Dang TP. Manic fringe inhibits tumor growth by suppressing Notch3 degradation in lung cancer. American journal of cancer research. 2013;3(5):490.
55. Andrzejewski S, Klimcakova E, Johnson RM, Tabariès S, Annis MG, McGuirk S, et al. PGC-1α promotes breast cancer metastasis and confers bioenergetic flexibility against metabolic drugs. Cell Metabol. 2017;26(5):778–87. e5.
56. Li Y, Xu S, Li J, Zheng L, Feng M, Wang X, et al. SIRT1 facilitates hepatocellular carcinoma metastasis by promoting PGC-1α-mediated mitochondrial biogenesis. Oncotarget. 2016;7(20):29255.

57. Juang YL, Jeng YM, Chen CL, Lien HC. PRRX2 as a novel TGF-β-induced factor enhances invasion and migration in mammary epithelial cell and correlates with poor prognosis in breast cancer. Molecular carcinogenesis. 2016;55(12):2247–59.
58. Wang Q, Chen D-l, Zhang L-f, Bian H. Promoting cell viability and migration of gastric cancer cells by PRRX2 via activation of Wnt/β-catenin signaling pathway. Chinese Journal of Pathophysiology. 2018;34(3):410–6.
59. Castanas E, Kampa M, Pelekanou V, Notas G, Athanasouli P, Alexakis K, et al. BCMA (TNFRSF17) induces APRIL and BAFF mediated breast cancer cell stemness. Frontiers in oncology. 2018;8:301.
60. Chae S-C, Yu J-l, Uhm T-B, Lee S-Y, Kang D-B, Lee J-K, et al. The haplotypes of TNFRSF17 polymorphisms are associated with colon cancer in a Korean population. Int J Colorectal Dis. 2012;27(6):701–7.

Figures

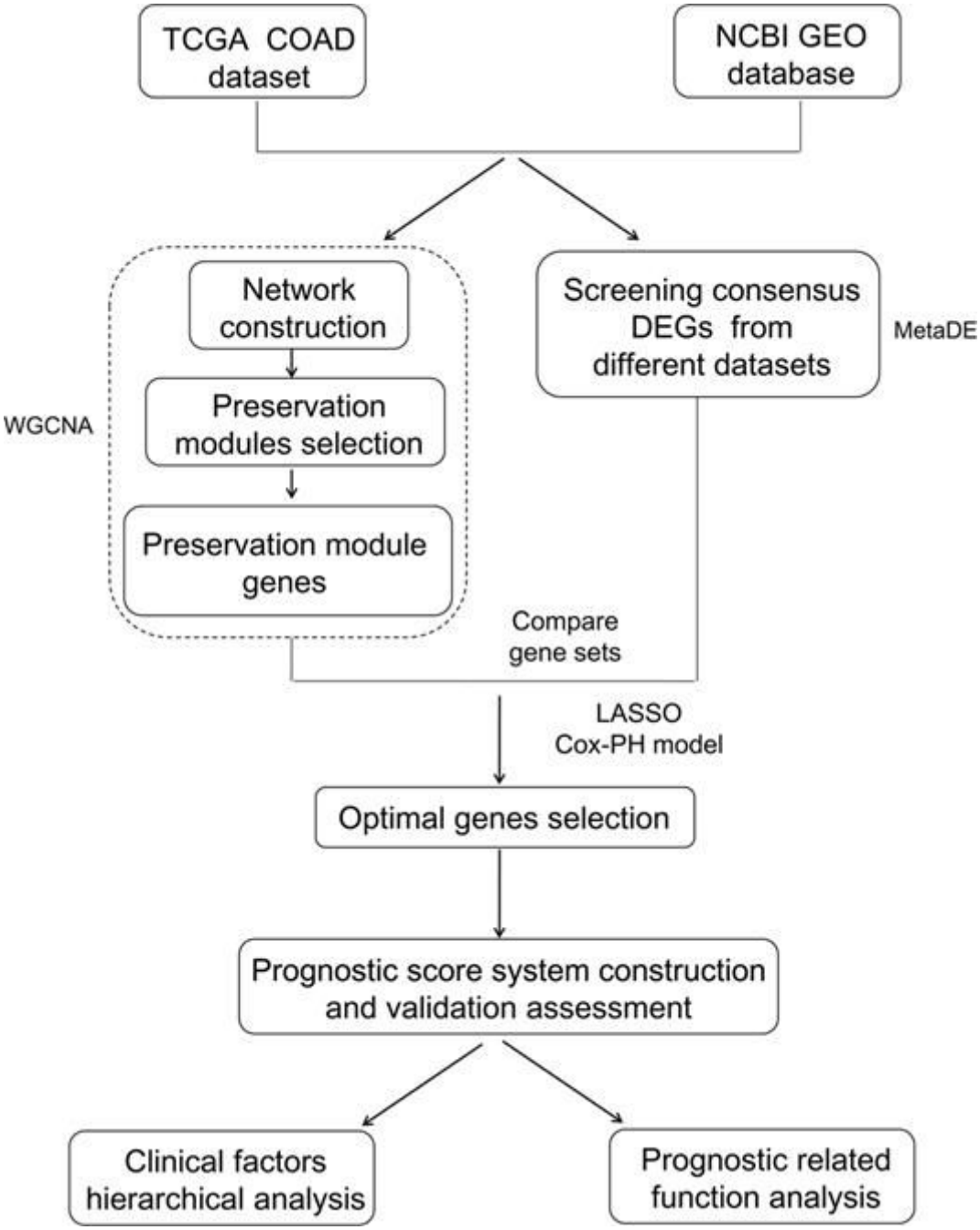

Figure 1

Workflow of this study.

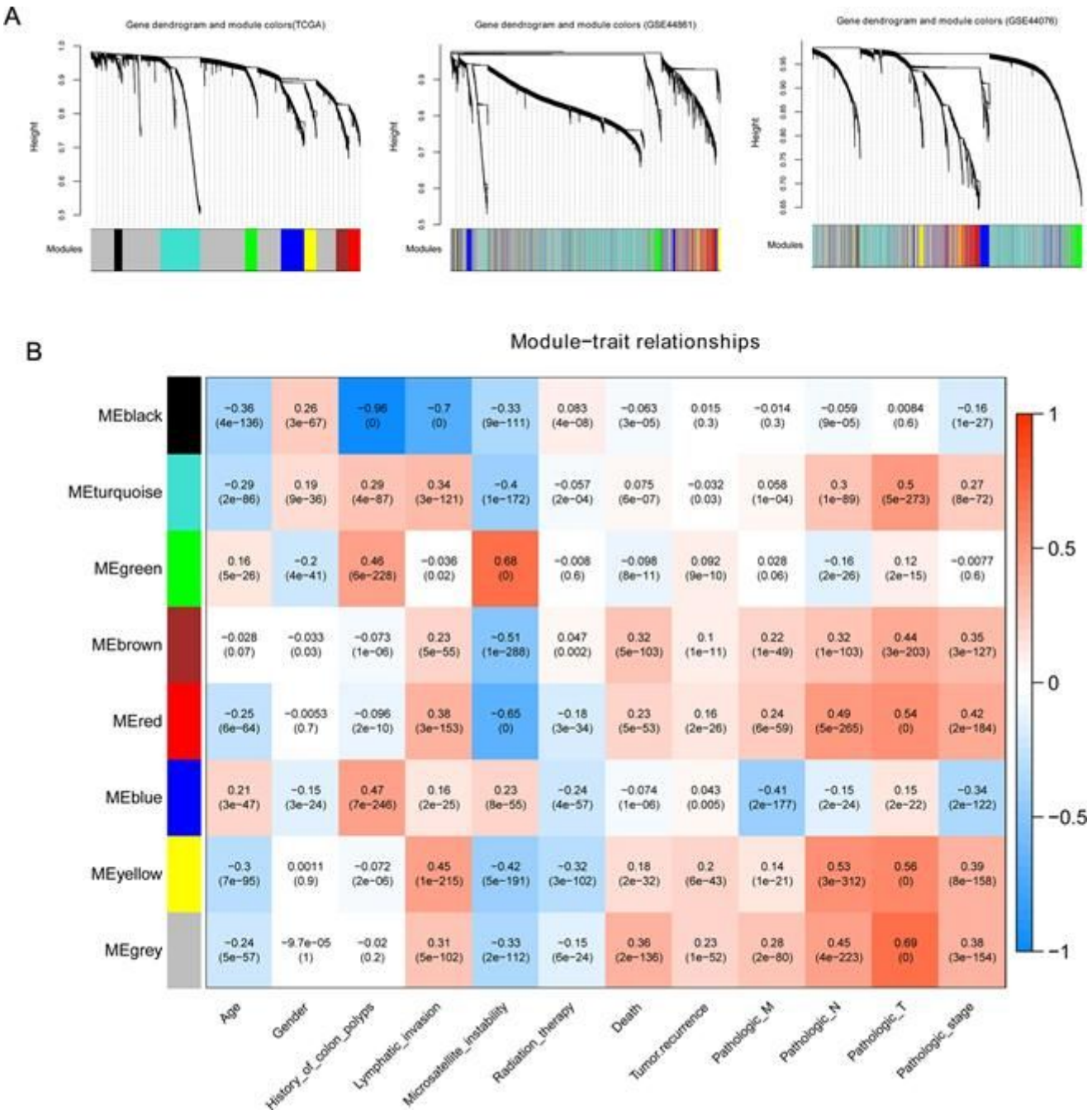

Figure 2

The partition of gene module related to colon cancer based on weighed gene co-expression network analysis algorithm. (A) The module partition results of TCGA, GSE44861 and GSE44076 datasets. The different colors represent the different that modules. (B) The correlation heatmap of gene module and clinical factors of colon cancer. The horizontal axis represents clinical factors, and the vertical axis represents gene modules of different colors. The color changed from blue to red indicates the change process from negative correlation to positive correlation. The numbers in the boxes indicate the correlation coefficients, and the numbers in parentheses indicate the p-values.

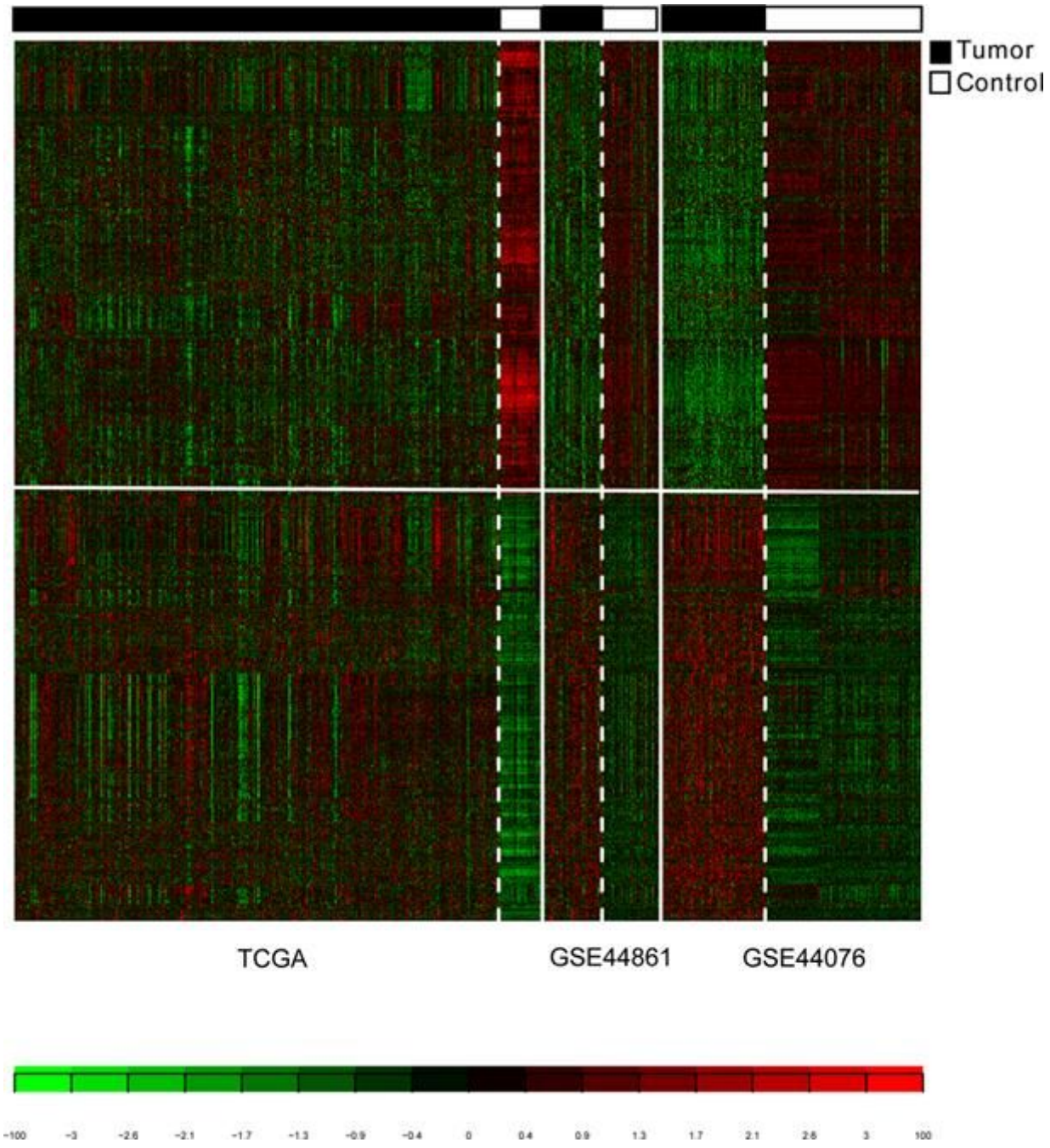

Figure 3

The heat-map of consistently differentially expressed genes based on MetaDE analysis.

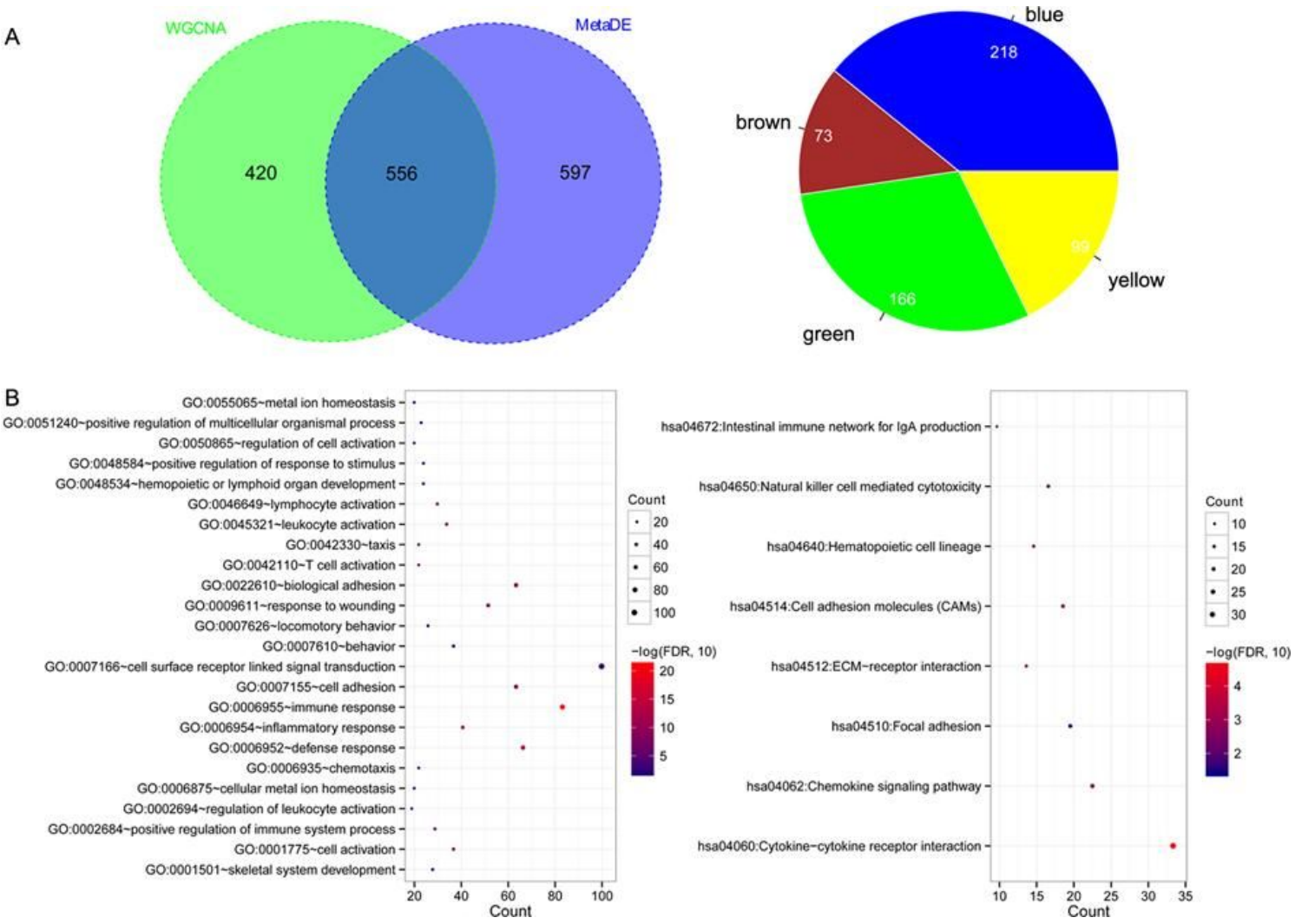

Figure 4

Analysis of intersected genes. (A) The intersected genes between genes of gene module by WGCNA and consistently DEGs by MetaDE. Left: Venn diagram of genes between genes of gene module by WGCNA and consistently DEGs by MetaDE; Right: Pie chart of overlapping genes in WGCNA modules. (B) GO-BP (left) terms and KEGG pathways (right) analyses of the intersected genes. Horizontal axis and vertical axis represent gene number and term, respectively; the color and size of the bots indicate the significant P value, and the closer the color is to red, the higher the significance.

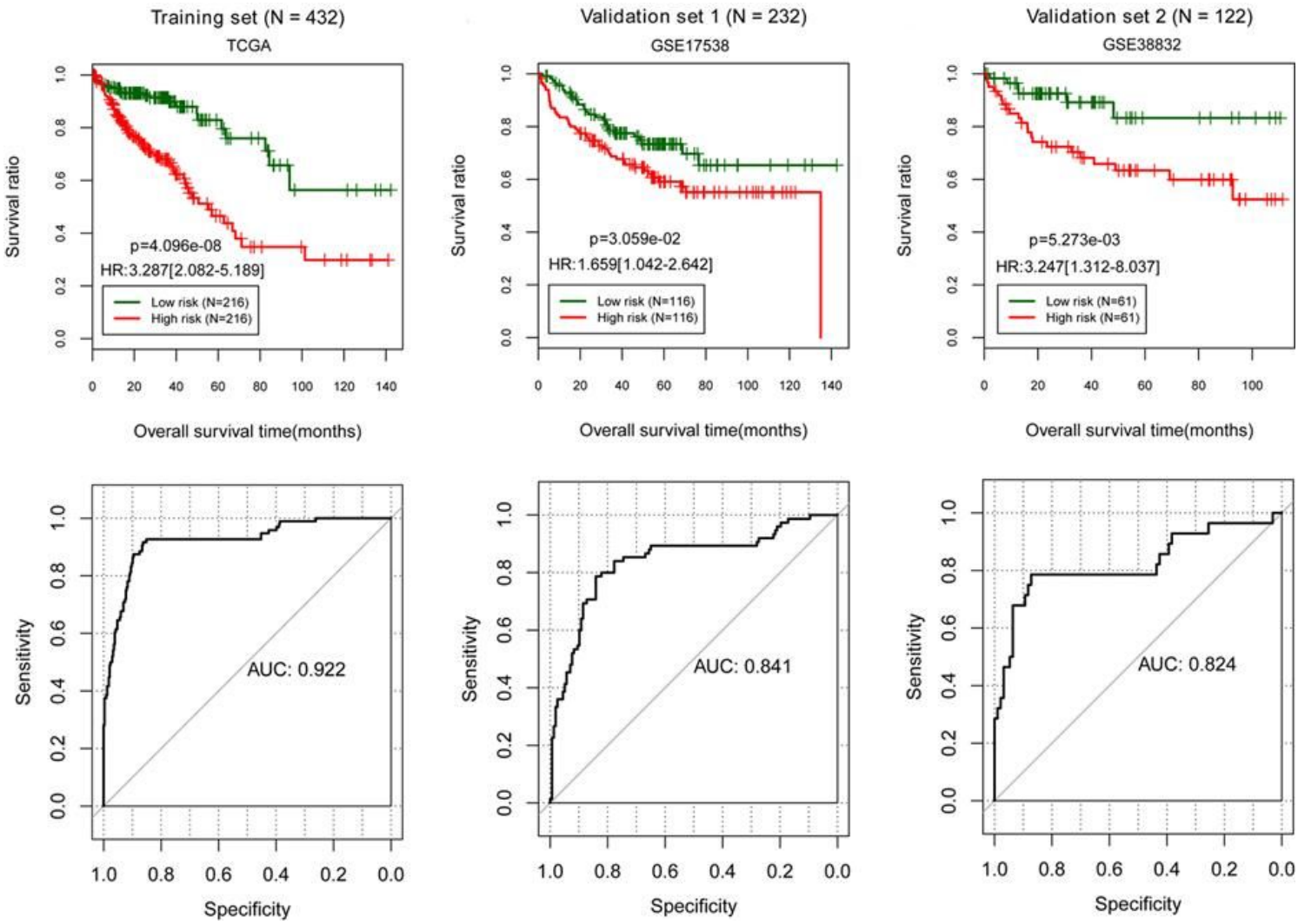

Figure 5

K-M survial analysis. The K-M survial analysis of low- and high- risk groups (upper) and ROC curve analysis of prognostic model (lower) in the training (TCGA) and validation (GSE44861 and GSE44076) datasets. HR represents hazard ratio, and the number in parentheses indicates 95% confidence interval (CI).

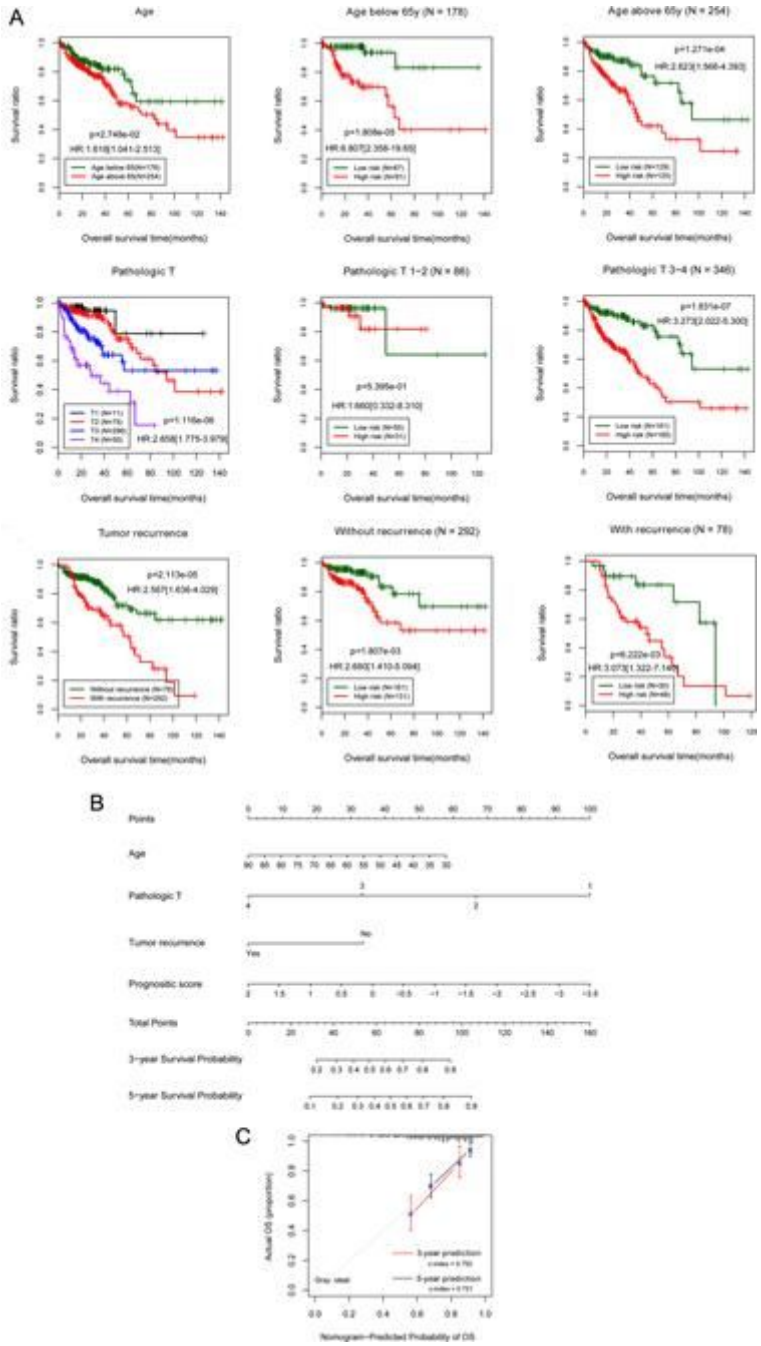

Figure 6

Independent prognostic clinical factors analysis. (A) The K-M survival analysis of age, pathological T and tumor recurrence in all samples (left), as well as different groups based on hierarchical analysis of clinical prognostic factors (middle and right). HR represents hazard ratio, and the number in parentheses indicates 95% confidence interval (CI). (B) The nomogram of 3- and 5-year survival prediction models for these independent prognostic factors. (C) The nomogram of 3- and 5-year survival prediction compliant to actual 3- and 5-year survival.

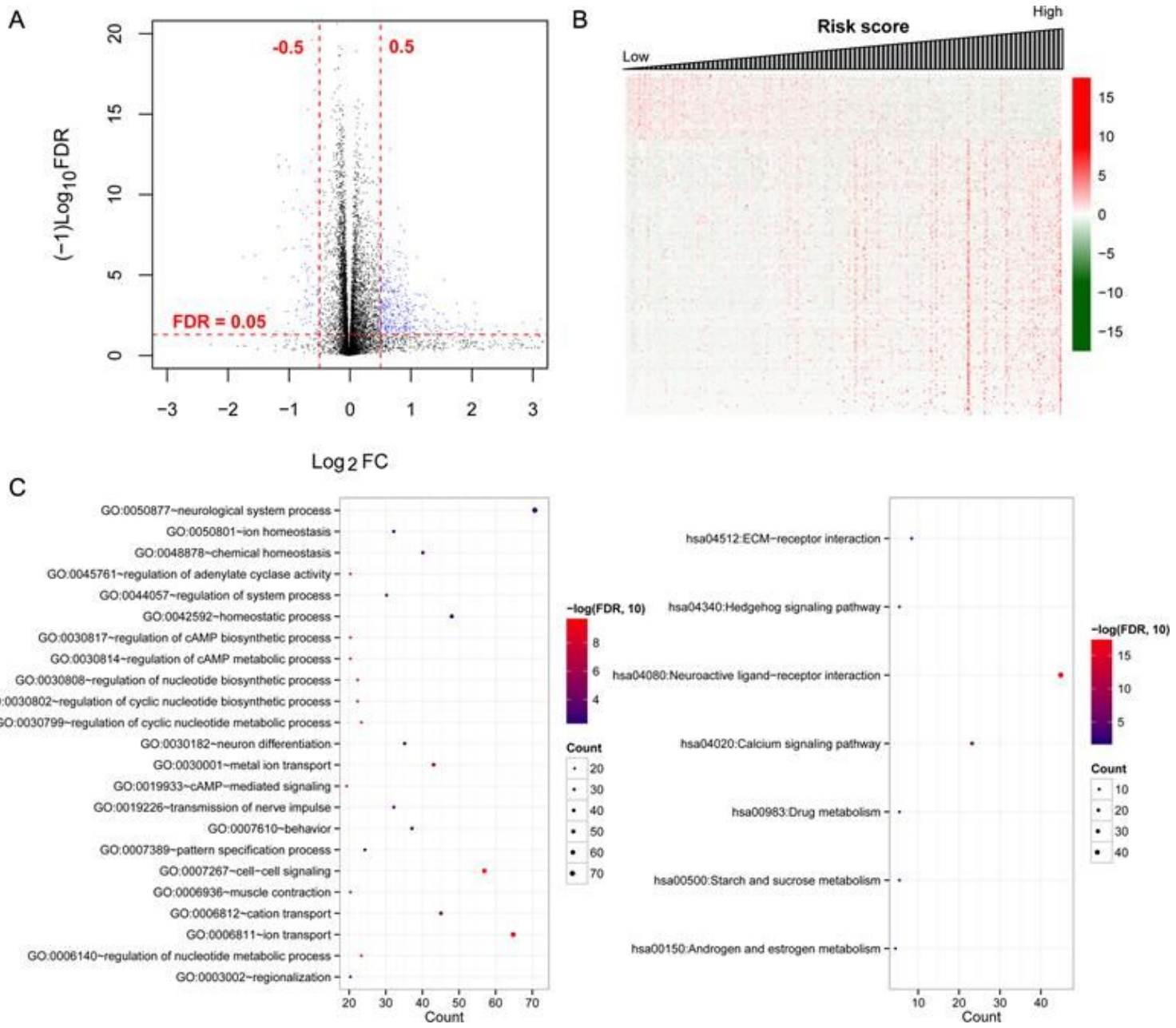

Figure 7

Screening of genes related to prognostic risk grouping and functional analysis. (A) Scatter plot of 514 differently expressed mRNAs (DEGs) between high risk and low risk groups. Red, green triangle, and blue indicate genes are up-regulated, down-regulated, and non-significant differentially expressed mRNAs, respectively. (B) Heat map of 514 DEGs between high risk and low risk groups. (C) GO and KEGG enrichment analyses of DEGs. Horizontal axis and vertical axis represent gene number and term, respectively; the color of the bar indicates the significant P value, and the closer the color is to red, the higher the significance.

Supplementary Files

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

- FigureS1.jpg