



ORIGINAL ARTICLE

A novel prognostic model based on multi-omics features predicts the prognosis of colon cancer patients

Haojie Yang¹ | Wei Jin¹ | Hua Liu¹  | Xiaoxue Wang² | Jiong Wu¹ | Dan Gan¹ | Can Cui¹ | Yilin Han¹ | Changpeng Han¹ | Zhenyi Wang¹ 

¹Department of Coloproctology, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

²Department of Coloproctology, The Sixth Affiliated Hospital of Sun Yat-sen University (Gastrointestinal & Anal Hospital of Sun Yat-sen University), Guangzhou, China

Correspondence

Changpeng Han and Zhenyi Wang, Department of Coloproctology, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 200437, China.

Emails: 13918314533@163.com; yyygchangke@163.com

Funding information

This work was supported by the National Natural Science Foundation of China project (No. 81874468, No. 81403399), three-year plan of action to further accelerate the development of traditional Chinese medicine in Shanghai under Grant ZY(2018-2020)-CCCX-2004-09 and Peak Research Team Project in Shanghai University of Traditional Chinese Medicine.

Abstract

Background: As a common malignant tumor in the colon, colon cancer (CC) has high incidence and recurrence rates. This study is designed to build a prognostic model for CC.

Methods: The gene expression dataset, microRNA-seq dataset, copy number variation (CNV) dataset, DNA methylation dataset, and transcription factor (TF) dataset of CC were downloaded from UCSC Xena database. Using limma package, the differentially methylated genes (DMGs), and differentially expressed genes (DEGs) and miRNAs (DEMs) were identified. Based on random forest method, prognostic model for each omics dataset were constructed. After the omics features related to prognosis were selected using logrank test, the prognostic model based on multi-omics features was built. Finally, the clinical phenotypes correlated with prognosis were screened using Kaplan–Meier survival analysis, and the nomogram model was established.

Results: There were 1625 DEGs, 268 DEMs, and 386 DMGs between the tumor and normal samples. A total of 105, 29, 159, five, and six genes/sites significantly correlated with prognosis were identified in the gene expression dataset (*GABRD*), miRNA-seq dataset (miR-1271), CNV dataset (*RN7SKP247*), DNA methylation dataset (cg09170112 methylation site [located in *SFSWAP*]), and TF dataset (*SIX5*), respectively. The prognostic model based on multi-omics features was more effective than those based on single omics dataset. The number of lymph nodes, pathologic_M stage, and pathologic_T stage were the clinical phenotypes correlated with prognosis, based on which the nomogram model was constructed.

Conclusion: The prognostic model based on multi-omics features and the nomogram model might be valuable for the prognostic prediction of CC.

KEYWORDS

bioinformatics, colon cancer, multi-omics, nomogram model, prognostic model

Haojie Yang, Wei Jin, and Hua Liu contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. Molecular Genetics & Genomic Medicine published by Wiley Periodicals LLC

1 | INTRODUCTION

Colon cancer (CC) is a common malignant tumor that is mainly caused by low-fiber and high-fat diets (Labianca et al., 2010). CC includes adenocarcinoma, mucinous adenocarcinoma, and undifferentiated carcinoma, which has the signs of hematochezia, bellyache, the change of bowel habits, and anemia (Mody & Bekaiisab, 2018). CC is the third most common cancer among gastrointestinal cancers, and its incidence is higher in men compared with that in women (Siegel, Miller, & Jemal, 2018). Due to recurrence and metastasis after surgery, the 5-year survival rate of CC patients receiving surgery is only 60%–70% (Mcguire, 2016; Siegel et al., 2017). Therefore, the mechanisms of CC should be further studied to improve its treatment and prognosis.

Through mediating focal adhesion-related pathways, overexpressed collagen type VIII alpha 1 chain (*COL8A1*) functions as an independent prognostic factor in colon adenocarcinoma (COAD; Shang et al., 2018). MiR-191-5p is inversely correlated with the expression of programmed cell-death 1 ligand 1 (*PD-L1*) and has influences on the overall survival (OS) of COAD patients (Chen et al., 2018). p16INK4a (cyclin-dependent kinase inhibitor 2A) and p14ARF are two cell cycle regulators encoded by *INK4a/ARF* locus, and their simultaneous hypermethylation has great prognostic value in sporadic colorectal cancer (CRC; Sung & Cho, 2006). The expression of the transcription factor (TF) E2F transcription factor 1 (*E2F-1*) is negatively related to tumor growth, indicating that *E2F-1* acts as a tumor suppressor gene in COAD (Bramis et al., 2004; Panayotis et al., 2004). The LIM and senescent cell antigen-like domain 2 (*PINCH-2*) copy number variation (CNV) presents mRNA up-regulation and copy number amplification in the non-relapsed samples in comparison with the systemic relapse samples, which is involved in reduced systemic recurrence of CC via competitively mediating integrin-linked kinase, PINCH, and parvin (IPP complex) formation with *PINCH-1* (Park et al., 2015). Despite of these reports, the pathogenesis of CC has not been thoroughly understood.

The underlying mechanisms of human disease are often complex, and multi-omics approach is valuable for revealing the pathogenic factors in diseases (Hu et al., 2018; Liu, Wang, Genchev, & Lu, 2017). In this study, differential analysis was conducted, and omics features correlated with prognosis were selected from different omics datasets (the gene expression dataset, microRNA [miRNA]-seq dataset, CNV dataset, DNA methylation dataset, and TF dataset) of CC. Furthermore, prognostic model of each omics dataset, prognostic model of multi-omics data and nomogram model respectively were constructed. This study might promote the prognostic prediction of CC patients.

2 | METHODS

2.1 | Data acquisition and integration

From UCSC Xena database (<http://xena.ucsc.edu/>; Goldman, Craft, Zhu, & Haussler, 2017), The Cancer Genome Atlas (TCGA) dataset of CC (RNA-Seq by expectation-maximization expression values; platform: IlluminaHiSeq; including 288 tumor samples and 41 normal samples) was extracted. Besides, the CNV data of 451 CC samples, the methylation data of 337 samples (platform: Methylation450k; including 299 tumor samples and 38 normal samples), the miRNA-seq data of 261 samples (platform: Illumina HiSeq; including 253 CC samples and 8 normal samples), and the TF data of 264 CC samples (platform: HiSeqV2) were obtained. Meanwhile, the corresponding clinical data of the TCGA dataset was downloaded, including age, sex, TNM stage, survival time, and survival status (downloaded in May 2019). A total of 225 tumor samples among the samples of the six datasets have effective prognostic information, which would be used for the construction and evaluation of prognostic models.

2.2 | Differential analysis

For the gene expression dataset, miRNA-seq dataset, and DNA methylation dataset, the R package limma (Ritchie et al., 2015; version 3.10.3, <https://bioconductor.org/packages/release/bioc/html/limma.html>) was applied for performing the differential analysis between tumor samples and normal samples. The *p*-values were adjusted by Benjamini & Hochberg method (Abbas, Kong, Liu, Jing, & Gao, 2013). Based on the $|\log$ fold change (FC)| values and the *p*-values, the differentially expressed genes (DEGs; thresholds: $|\log$ FCI > 2, adjusted *p* < .05), differentially expressed miRNAs (DEMs; thresholds: $|\log$ FCI > 1, adjusted *p* < .05), and differentially methylated genes (DMGs; thresholds: $|\log$ FCI > 0.5, adjusted *p* < .05) were screened.

2.3 | Prognostic model and feature screening of each omics dataset

For each omics data (including the TF dataset, gene expression dataset, miRNA-seq dataset, DNA methylation dataset, and CNV dataset) of the 225 tumor samples, the random forest method (Ellis et al., 2014; parameter: *n_estimators* = 100, *min_samples_leaf* = 3, *min_samples_split* = 3) in Sklearn library (<https://scikit-learn.org/>) was used to construct prognostic model. Parameters were adjusted using the GridSearchCV package (Pontes, Amorim, Balestrassi, Paiva, & Ferreira, 2016), and the prognostic model was trained to predict the 5-year survival rate of the samples. The preprocessing scale method (Tang, Lu,

Cai, Han, & Wang, 2008) was used to homogenize the samples, and then randomly divide the samples into training set and test set with a ratio of 7:3. The threefold cross-validation (Arlot & Celisse, 2010) in the training set was performed to train the model, and the parameter `class_weight = "balanced"` was added during the training to eliminate the influence of category imbalance. Finally, the accuracy and area under the curve (AUC) were used as the criteria for model evaluation on the test set. Meanwhile, the random forest model (Ellis et al., 2014) was used to screen the most relevant features for prognosis.

2.4 | Identification of the omics features correlated with prognosis

Based on the downloaded clinical data, the prognostic information (including OS and OS status) of corresponding patients were organized. Combined with the omics features (TF information, differential mRNA/miRNA/methylation sites, gene CNV, etc) of different omics datasets and the sample prognosis information, Kaplan–Meier (KM) survival analysis (Luo, Zhang, Wang, Zhu, & Jia, 2018) was adopted to divide the patients into high expression group and low expression group according to the eigen values. Meanwhile, logrank test (Lin & León, 2017) was used to calculate the significant p -values, and the omics features related to prognosis were selected. The $p < .01$ was set as the significant threshold.

2.5 | Prognostic model of multi-omics data

The 225 tumor samples were randomly classified into training set and test set with a ratio of 7:3. For the training set, the multi-omics genetic features correlated with prognosis were comprehensively considered, and the risk score model based on random forest (Ellis et al., 2014) was trained.

For each sample in the test set, risk score was calculated using the trained risk score model. With the median of the risk scores as the criterion, the samples were divided into high-risk group and low-risk group. Afterward, KM survival analysis (Luo et al., 2018) was conducted to check whether there was statistical difference. In addition, risk score was used as the standard to predict the 5-year survival rate of the patients in the test set, and the receiver operating characteristic (ROC) curve (Peterson, Papeş, & Soberón, 2017) and AUC value were used for model evaluation.

2.6 | Nomogram model construction

Through corresponding different factors to points and adding the points to get total points, the survival rate can finally

be obtained. Based on the risk score model and the clinical phenotypes correlated with prognosis, the nomogram model (Gotto, Yu, Bernstein, Eastham, & Kattan, 2014) was constructed and visualized. To further verify the prediction ability of the nomogram model, the consistency index (C-index) of each independent prognostic factor and the combined model for the Cox Proportional Hazards (Cox PH) model (Sy & Taylor, 2015) were calculated. In addition, the statistical analysis was performed by resampling technology (Minaei-Bidgoli, Parvin, Alinejad-Rokny, Alizadeh, & Punch, 2014), and the significant p -value were computed.

3 | RESULTS

3.1 | Differential analysis

A total of 20,530 genes, 1,402 miRNAs, and 485,577 methylation sites were identified respectively from the gene expression dataset, miRNA-seq dataset, and DNA methylation dataset. There were 1625 DEGs (1,265 up-regulated genes and 360 down-regulated genes), 268 DEMs (157 up-regulated miRNAs and 111 down-regulated miRNAs), and 386 DMGs (54 hypermethylation genes and 332 hypomethylation genes) between the tumor and normal samples. The volcano plots and clustering heatmaps showing the results of differential analysis are presented in Figure 1.

3.2 | Prognostic model and feature screening of each omics dataset

For each omics data of the 225 tumor samples, prognostic model was constructed and trained to obtain the predicted 5-year survival rate of the samples in test set (Figure 2a). Meanwhile, the more important features in each prognostic model are also shown in Figure 2. The results showed that the prognostic models based on the TF dataset (AUC value = 0.731) and the DNA methylation dataset (AUC value = 0.735) had better efficacies. However, the prognostic model based on the CNV dataset (AUC value = 0.550) had the worst efficacy (Table 1).

3.3 | Identification of the omics features correlated with prognosis

The prognostic information of the corresponding patients of the 225 tumor samples was extracted. Combined with the expression values of each omics characteristic in different samples and the sample prognosis information, KM survival curves were drawn. There were 304 features, including 105 DEGs, 29 DEMs, 159 CNVs, 5 DMGs, and 6 TFs that were

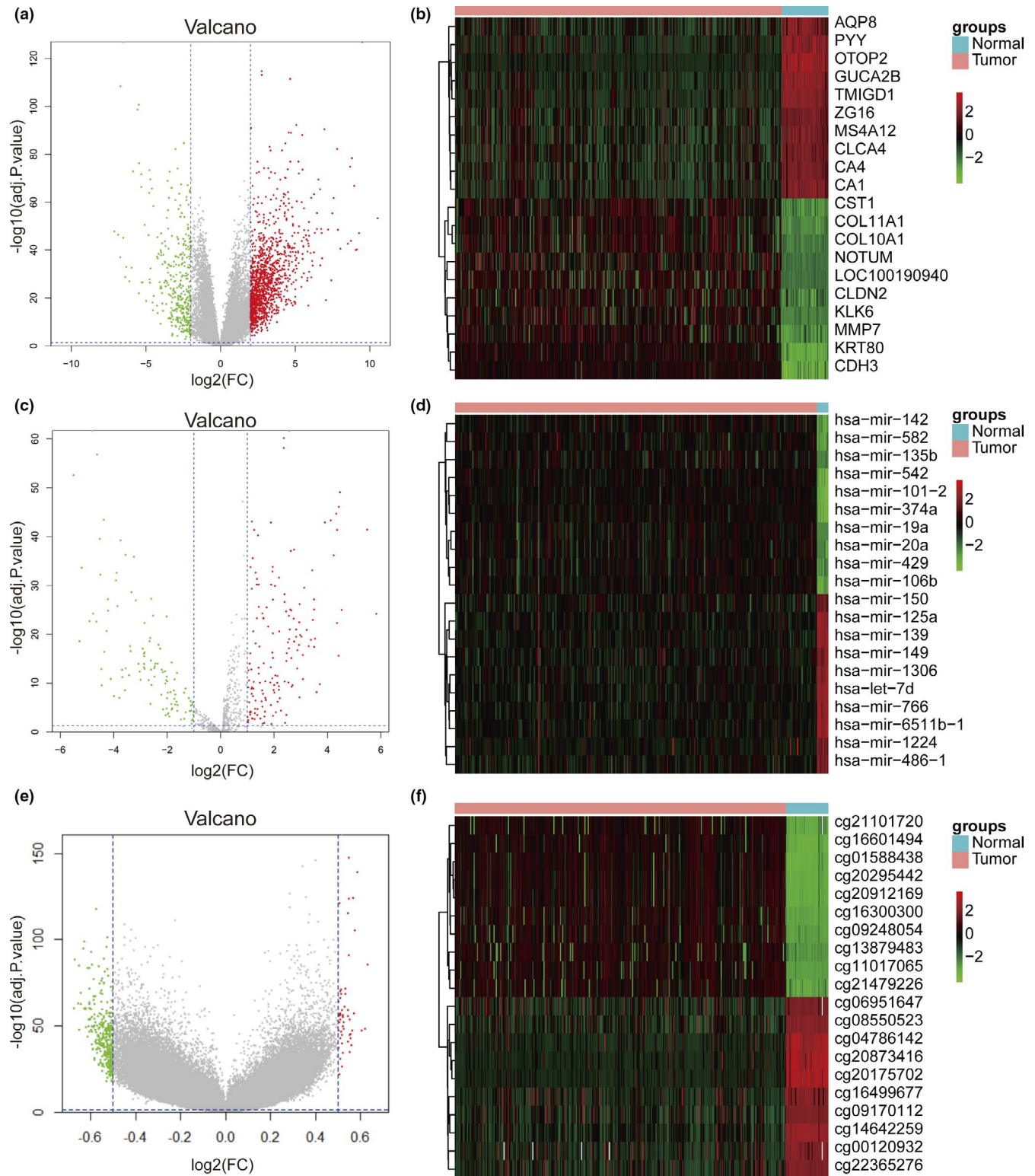


FIGURE 1 Volcano plots and clustering heatmaps. (a) The volcano plot for the differentially expressed genes (DEGs); (b) The clustering heatmap for the DEGs; (c) The volcano plot for the differentially expressed miRNAs (DEMs); (d) The clustering heatmap for the DEMs; (e) The volcano plot for the differentially methylated genes (DMGs); (f) The clustering heatmap for the DMGs. In the volcano plots, red and green respectively represent up-regulation and down-regulation. The clustering heatmaps were drawn based on the top 10 results (up-regulation or down-regulation) of differential analysis. In clustering heatmaps, red and blue sample strips respectively represent tumor samples and normal samples

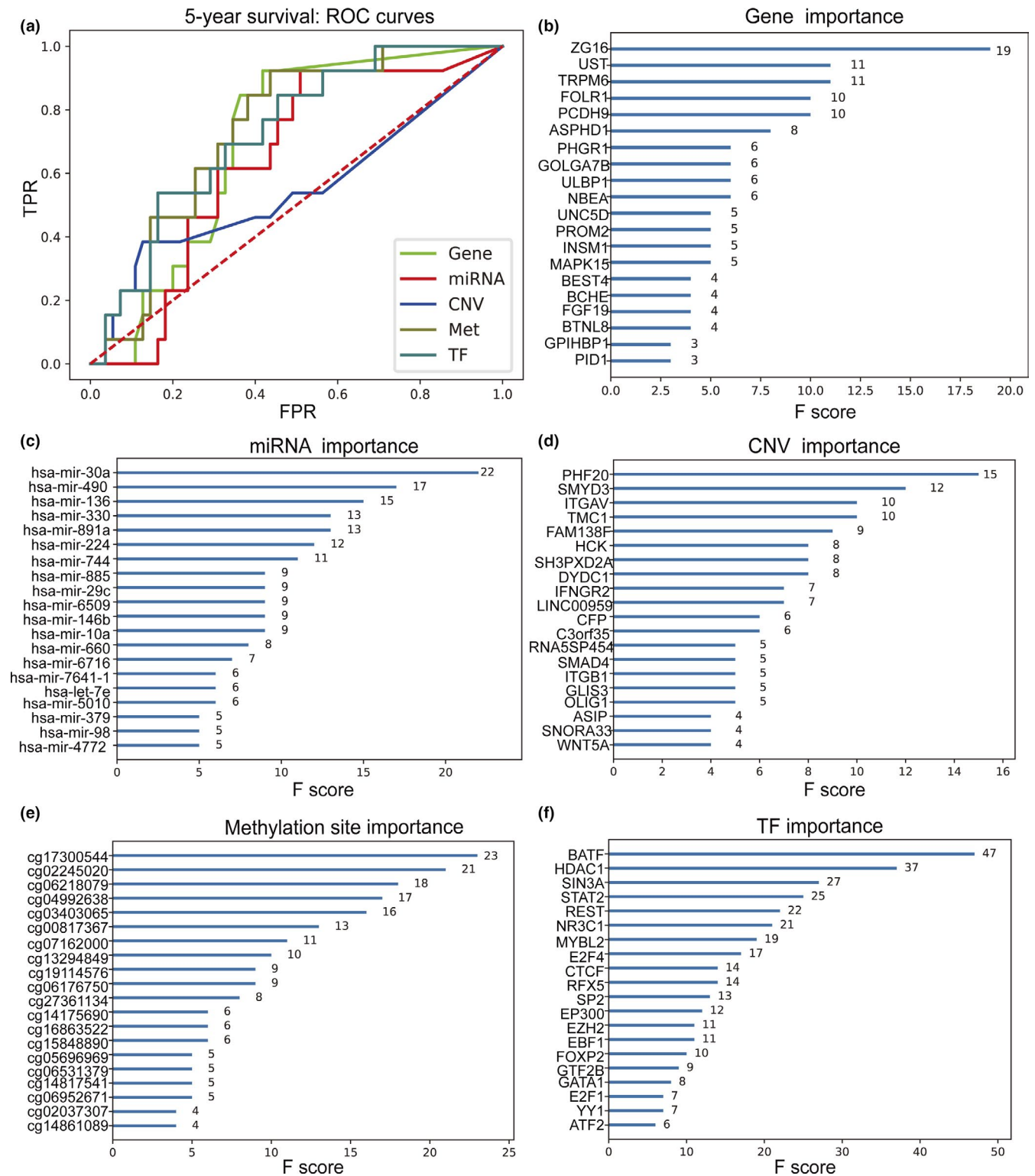


FIGURE 2 Receiver operating characteristic (ROC) curves and the more important features in each prognostic model based on single omics dataset. (a) The ROC curves showing the 5-year survival rates predicted by the prognostic models base on single omics dataset; (b) The top 20 more important features in the prognostic model based on the gene expression dataset; (c) The top 20 more important features in the prognostic model based on the miRNA-seq dataset; (d) The top 20 more important features in the prognostic model based on the CNV dataset; (e) The top 20 more important features in the prognostic model based on the DNA methylation dataset; (f) The top 20 more important features in the prognostic model based on the TF dataset. The greater the area under the curve value, the better the model effect was. CNV, copy number variation; FPR, false positive rate; TF, transcription factor; TPR, true positive rate

significantly correlated with prognosis (Table S1). The gene/site with the highest prognostic significance in the gene expression dataset (gamma-aminobutyric acid type A receptor delta subunit, *GABRD*), miRNA-seq dataset (miR-1271), CNV dataset (RN7SK pseudogene 247, *RN7SKP247*), DNA methylation dataset {cg09170112 methylation site (located in splicing factor SWAP [*SFSWAP*])}, and TF dataset (SIX homeobox 5, *SIX5*) are shown in Figure 3.

3.4 | Prognostic model of multi-omics data

The 225 tumor samples were randomly classified into training set and test set. Combined with the prognosis-correlated features in multi-omics data, the risk score model in the training set was trained. Based on the trained risk score model, the risk score for each sample in the test set was calculated. The survival curves showed that risk score was significantly correlated with prognosis and the survival time of patients in the low-risk group was significantly longer than that in the high-risk group ($p = .012$; Figure 4a). The AUC values of the 1-, 3-, and 5-year survival rates predicted by this risk score model were 0.621, 0.604, and 0.782, respectively (Figure 4b).

TABLE 1 Evaluation of the prognostic models based on each omics dataset

Dataset	Accuracy	AUC
Gene	0.721	0.699
miRNA	0.706	0.641
CNV	0.613	0.550
methylation	0.779	0.735
TF	0.801	0.731

Abbreviations: AUC, area under the curve; CNV, copy number variation; TF, transcription factor.

Compared with the prognostic model based on single omics dataset, the AUC value of the prognostic model based on multi-omics features was significantly higher. Therefore, the integration of these omics features was effective for the construction of prognostic model.

3.5 | Clinical phenotypes and survival analysis

The downloaded clinical data of the samples in the TCGA dataset contained 551 clinical phenotypes, and the survival time and survival status of these samples were extracted. In combination with sex, age, histological type, number of lymph nodes, TNM stage, presence of colon polyps and prognostic information, a total of 244 valid samples were selected. KM survival curves showed that the number of lymph nodes ($p = .0067$), pathologic_M stage ($p = .00011$), and pathologic_T stage ($p < .0001$) were significantly correlated with the survival time of the patients (Figure 5).

3.6 | Nomogram model construction

In the test set, 63 samples have the clinical phenotype information including the number of lymph nodes and TNM stage. Based on the risk scores, the number of lymph nodes, pathologic_M stage, and pathologic_T stage, a nomogram model was established (Figure 6). The 1-year survival rate, 3-year survival rate and 5-year survival rate could be predicted based on the risk score, number of lymph node, pathologic_M and pathologic_T from the nomogram model. Each feature (risk score, number of lymph node, pathologic_M and pathologic_T) corresponded to a single point at the top scale-plate of “points.” The 1-year survival rate, 3-year survival rate

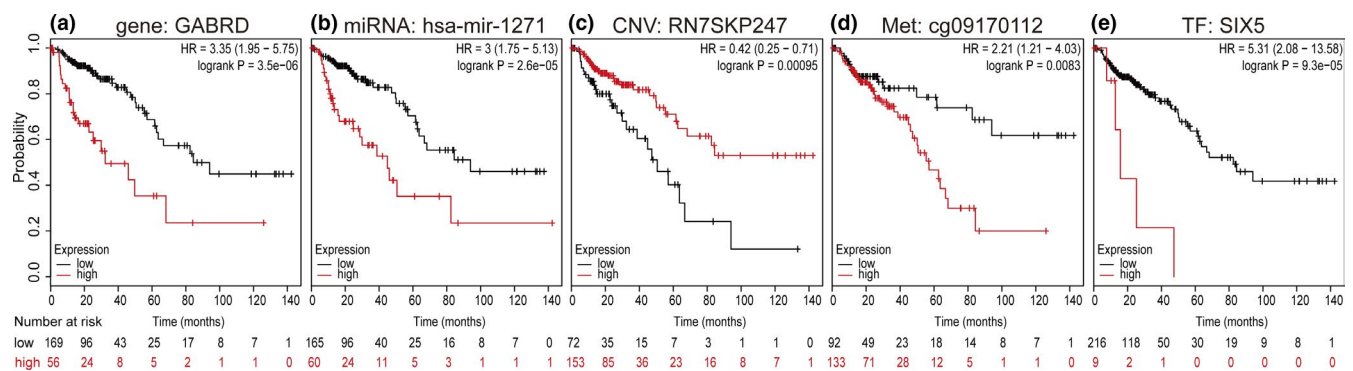


FIGURE 3 Kaplan–Meier (KM) curves for the gene/site with the highest prognostic significance in each omics dataset. (a) The KM curve for the gene with the highest prognostic significance in the gene expression dataset; (b) The KM curve for the miRNA with the highest prognostic significance in the miRNA-seq dataset; (c) The KM curve for the gene with the highest prognostic significance in the copy number variation (CNV) dataset; (d) The KM curve for the methylation gene with the highest prognostic significance in the DNA methylation dataset; (e) The KM curve for the transcription factor (TF) with the highest prognostic significance in the TF dataset. Black and red curves represent low expression group and high expression group, respectively. HR, hazard ratio

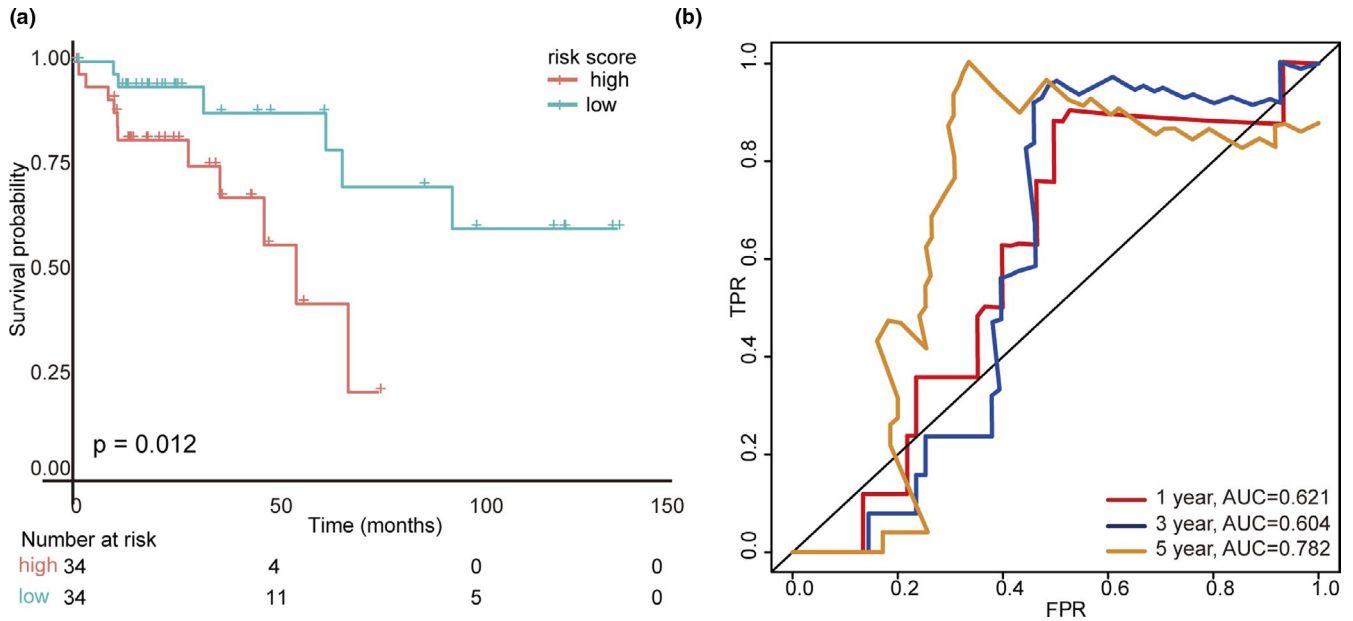


FIGURE 4 Kaplan–Meier (KM) curves and receiver operating characteristic (ROC) curves. (a) The KM curves for the high- and low-risk groups divided by the risk score model based on multi-omics features (green and red curves respectively represent low- and high-risk groups); (b) The 1-, 3-, and 5-year ROC curves predicted by the risk score model based on multi-omics features (red, blue, and orange, respectively, represent 1-, 3-, and 5-year ROC curves). AUC, area under the curve; FPR, false positive rate; TPR, true positive rate

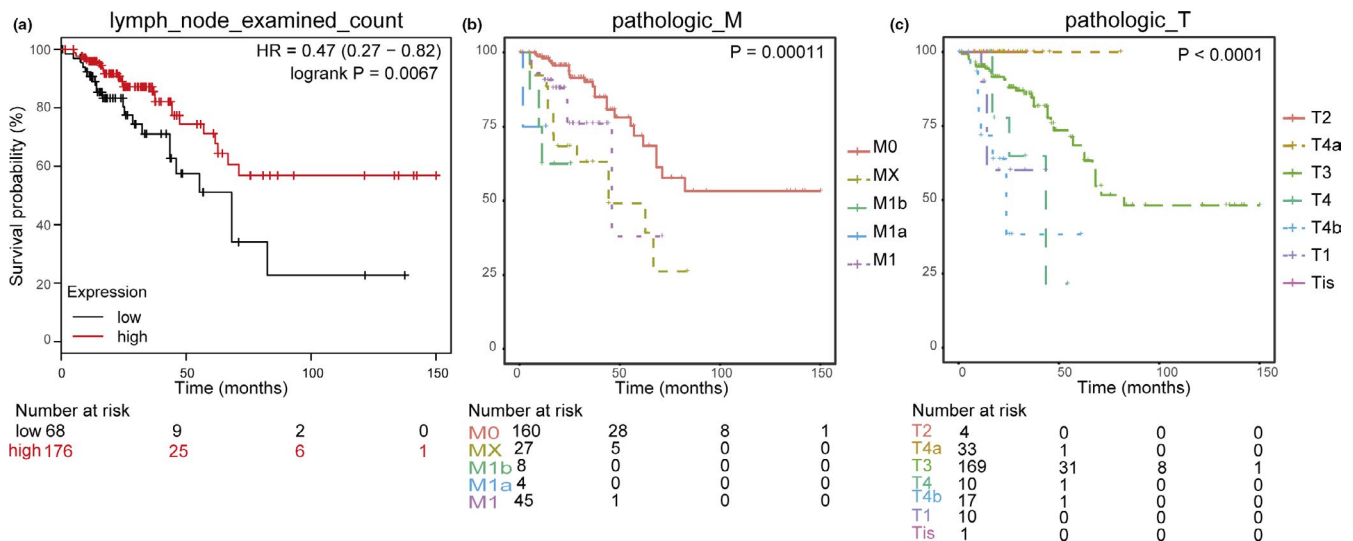


FIGURE 5 Kaplan–Meier (KM) curves showing the correlations between the number of lymph nodes/pathologic_M stage/pathologic_T stage and the survival time of the patients. (a) The KM curves for the number of lymph nodes; (b) The KM curves for pathologic_M stage; (c) The KM curves for pathologic_T stage. HR, hazard ratio

and 5-year survival rate were calculated by the total points combining the point of each feature. Moreover the C-index of each independent prognostic factor and the combined model for the Cox PH model and the significant p -values were calculated. The results showed that the C-index values of the combined model (C-index value = 0.9) and riskscore factor (C-index value = 0.74) were higher, and thus their fitting degrees for Cox PH model were relatively high. Among the clinical factors, pathologic_M ($p = .009$) and pathologic_T ($p = .0447$) had significant test results (Table 2).

4 | DISCUSSION

A total of 1625 DEGs (1,265 up-regulated genes and 360 down-regulated genes), 268 DEMs (157 up-regulated miRNAs and 111 down-regulated miRNAs), and 386 DMGs (54 hypermethylation genes and 332 hypomethylation genes) were identified between the tumor and normal samples. Five methylation sites in the DNA methylation dataset (such as cg09170112 methylation site [located in *SFSWAP*]), six TFs in the TF dataset (such as *SIX5*), 105 genes in the gene

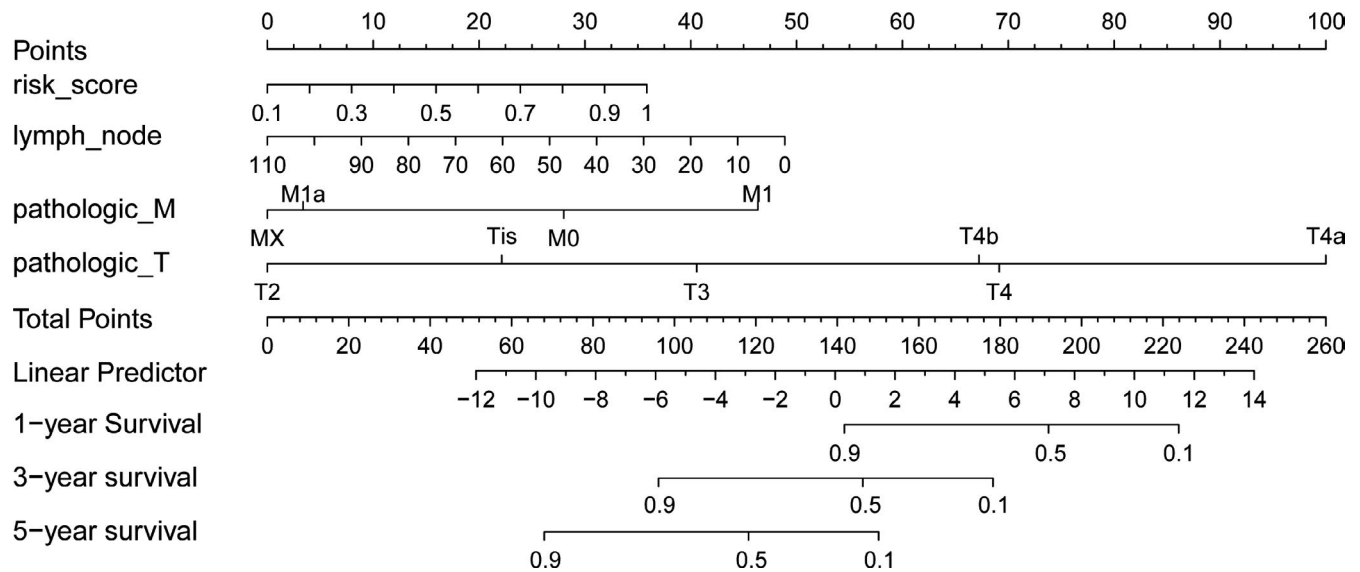


FIGURE 6 Nomogram model based on the risk scores, the number of lymph nodes, pathologic_M stage, and pathologic_T stage

TABLE 2 Fitting degrees of each independent prognostic factor and the combined model for Cox Proportional Hazards model

Factor	C-index	p-Value
Combined model	0.9	0
risk_score	0.74	.00425
lymph_node	0.527	.797
pathologic_M	0.7	.009
pathologic_T	0.686	.0447

expression dataset (such as *GABRD*), 29 miRNAs in the miRNA-seq dataset (such as miR-1271), and 159 genes in the CNV dataset (such as *RN7SKP247*) were significantly correlated with prognosis. Compared with the prognostic model based on single omics dataset, the prognostic model based on multi-omics features was more effective. After the number of lymph nodes, pathologic_M stage, and pathologic_T stage were selected as the clinical phenotypes correlated with prognosis, the nomogram model was established.

GABRD is strongly up-regulated and most other GABA receptors are down-regulated in cancerous cells, and the alteration of GABA receptors may have an influence on the differentiation of tumor cells (Gross, Kreisberg, & Ideker, 2015). *GABRD* is a stage-IV specific gene and is continuously up-regulated across different tumor stages, which may be a diagnostic and therapeutic target for hepatocellular carcinoma (Sarathi & Palaniappan, 2019). MiR-1271 suppresses cell proliferation and cell invasion and causes cell cycle arrest in CRC cells through negatively mediating metadherin (*MTDH*), and thus miR-1271 may serve as a potential therapeutic target for the tumor (Li et al., 2018; Sun, Zhai, Chen, Kong, & Zhang, 2018). Five lncRNAs, eight miRNAs (including miR-1271), and five mRNAs are found to be potential

prognostic markers, based on which three prognostic models are built for CC (Huang & Pan, 2019). Therefore, *GABRD* and miR-1271 might be implicated in the prognosis of CC patients.

The co-occurrence of IQ motif containing GTPase activating protein 2 (*IQGAP2*) genomic alterations and the deletion in small nucleolar RNA, H/ACA box 50A (*SNORA50A*), *SNORA50C*, and RNA component of 7SK nuclear ribonucleoprotein (*RN7SK*) genes is correlated with the reductions in disease-free survival, therefore, *IQGAP2* plays a role in the progression of prostate cancer (Xie, Zheng, & Tao, 2019). *RN7SK* paralogs promotes the activation of RNA polymerase II and pre-mRNA processing via negatively regulating positive transcription elongation factor b, which is correlated with the development and prognosis of renal cell carcinoma (Zhang et al., 2011). As a putative splicing factor, *SFSWAP* can exert specific influences on inner ear development and Notch pathway members (Yalda et al., 2014). Notch signaling is related to the activation of macrophages and the functions of relevant effector, which may be applied for the immune intervention in treating tumors (Palaga, Wongchana, & Kueanjinda, 2018). Through maintaining the balance between cell proliferation and cell apoptosis, Notch signaling is involved in tumor immunity and multidrug resistance and can be used for improving tumor treatment (Majidinia, Alizadeh, Yousefi, Akbarzadeh, & Zarghami, 2016). These suggested that *RN7SKP247* and *SFSWAP* might also be correlated with the survival of CC patients.

Dysregulation of Dach, Eya, and Six genes occurs in multiple tumors, indicating that targeting these genes may be effective for inhibiting tumor growth and progression (Christensen, Patrick, McCoy, & Ford, 2008). *SIX1*, *SIX2*, *SIX3*, *SIX4*, *SIX5*, and *SIX6* all belong to the sine oculis homeobox homolog (*SIX*) family, which play distinct roles in

the progression and prognosis of breast cancer (Xu et al., 2016). *SIX5* functions in promoting epithelial differentiation in female reproductive tract, and may be a biomarker of epithelial differentiation in borderline epithelial ovarian tumors (Winchester, Robertson, MacLeod, Johnson, & Thomas, 2000). Thus, *SIX5* might play a role in the prognosis of CC patients.

In conclusion, 1625 DEGs, 268 DEMs, and 386 DMGs were selected for the tumor and normal samples. Besides, the prognostic model based on multi-omics features and the nomogram model might be applied for the prognostic prediction of CC patients. However, proper experiments still needed to be designed and implemented to confirm these results.

ACKNOWLEDGMENTS

The authors thank Huaping Liu for critical reading of this manuscript and providing helpful suggestions.

CONFLICT OF INTEREST

All the authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Funding acquisition, Changpeng Han and Zhenyi Wang; Investigation, Haojie Yang, Changpeng Han; Methodology, Haojie Yang, Wei Jin, Hua Liu; Software, Haojie Yang; Discussion, Xiaoxue Wang, Jiong Wu, Yilin Han, Dan Gan, Can Cui; Validation, Haojie Yang; Writing-original draft, Haojie Yang, Wei Jin, Changpeng Han and Zhenyi Wang.

DATA AVAILABILITY STATEMENT

All data included in this study are available upon request by contact with the corresponding author.

ORCID

Hua Liu  <https://orcid.org/0000-0003-2985-1821>

Zhenyi Wang  <https://orcid.org/0000-0001-5653-9636>

REFERENCES

- Abbas, A., Kong, X.-B., Liu, Z., Jing, B.-Y., & Gao, X. (2013). Automatic peak selection by a Benjamini-Hochberg-based algorithm. *PLoS ONE*, *8*(1), e53112. <https://doi.org/10.1371/journal.pone.0053112>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*(2010), 40–79. <https://doi.org/10.1214/09-SS054>
- Bramis, J., Zacharatos, P., Papaconstantinou, I., Kotsinas, A., Sigala, F., Korkolis, D. P., ... Gorgoulis, V. G. (2004). E2F-1 transcription factor immunoreexpression is inversely associated with tumor growth in colon adenocarcinomas. *Anticancer Research*, *24*(5A), 3041–3047.
- Chen, X.-Y., Zhang, J., Hou, L.-D., Zhang, R., Chen, W., Fan, H.-N., ... Zhu, J.-S. (2018). Upregulation of PD-L1 predicts poor prognosis and is associated with miR-191-5p dysregulation in colon adenocarcinoma. *International Journal of Immunopathology and Pharmacology*, *32*, 2058738418790318.
- Christensen, K. L., Patrick, A. N., McCoy, E. L., & Ford, H. L. (2008). Chapter 5 the six family of homeobox genes in development and cancer. *Advances in Cancer Research*, *101*(8), 93–126.
- Ellis, K., Kerr, J., Godbole, S., Lanckriet, G., Wing, D., & Marshall, S. (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiological Measurement*, *35*(11), 2191. <https://doi.org/10.1088/0967-3334/35/11/2191>
- Goldman, M., Craft, B., Zhu, J., & Haussler, D. (2017). Abstract 2584: The UCSC Xena system for cancer genomics data visualization and interpretation. *Cancer Research*, *77*(13 Suppl), 2584.
- Gotto, G. T., Yu, C., Bernstein, M., Eastham, J. A., & Kattan, M. W. (2014). Development of a nomogram model predicting current bone scan positivity in patients treated with androgen-deprivation therapy for prostate cancer. *Frontiers in Oncology*, *4*(5), 296. <https://doi.org/10.3389/fonc.2014.00296>
- Gross, A. M., Kreisberg, J. F., & Ideker, T. (2015). Analysis of matched tumor and normal profiles reveals common transcriptional and epigenetic signals shared across cancer types. *PLoS ONE*, *10*(11), e0142618. <https://doi.org/10.1371/journal.pone.0142618>
- Hu, W., Yang, Y., Li, X., Huang, M., Xu, F., Ge, W., ... Zheng, S. (2018). Multi-omics approach reveals distinct differences in left- and right-sided colon cancer. *Molecular Cancer Research*, *16*(3), 476–485. <https://doi.org/10.1158/1541-7786.MCR-17-0483>
- Huang, Q. R., & Pan, X. B. (2019). Prognostic lncRNAs, miRNAs, and mRNAs form a competing endogenous RNA network in colon cancer. *Frontiers in Oncology*, *9*, 712. <https://doi.org/10.3389/fonc.2019.00712>
- Labianca, R., Beretta, G. D., Kildani, B., Milesi, L., Merlin, F., Mosconi, S., ... Wils, J. (2010). Colon cancer. *Critical Reviews in Oncology Hematology*, *74*(2), 106–133. <https://doi.org/10.1016/j.critrevonc.2010.01.010>
- Li, J., Xu, J., Yan, X., Jin, K., Li, W., & Zhang, R. (2018). Suppression of Capn4 by microRNA-1271 impedes the proliferation and invasion of colorectal cancer cells. *Biomedicine & Pharmacotherapy*, *99*, 162–168. <https://doi.org/10.1016/j.biopha.2017.12.107>
- Lin, R. S., & León, L. F. (2017). Estimation of treatment effects in weighted log-rank tests. *Contemporary clinical trials. Communications*, *8*, 147–155. <https://doi.org/10.1016/j.conctc.2017.09.004>
- Liu, C., Wang, X., Genchev, G. Z., & Lu, H. (2017). Multi-omics facilitated variable selection in Cox-regression model for cancer prognosis prediction. *Methods*, *124*, 100. <https://doi.org/10.1016/j.ymeth.2017.06.010>
- Luo, H., Zhang, C., Wang, J., Zhu, J., & Jia, X. (2018). Heart failure with preserved ejection fraction trials have heterogeneous control groups: A comparison of Kaplan–Meier curves. *Cardiovascular Drugs and Therapy*, *32*(6), 577–580. <https://doi.org/10.1007/s10557-018-6827-5>
- Majidinia, M., Alizadeh, E., Yousefi, B., Akbarzadeh, M., & Zarghami, N. (2016). Downregulation of notch signaling pathway as an effective chemosensitizer for cancer treatment. *Drug Research*, *66*(11), 571–579. <https://doi.org/10.1055/s-0042-111821>
- Mcguire, S. (2016). World cancer report (2014). Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Advances in Nutrition*, *7*(2), 418.
- Minaei-Bidgoli, B., Parvin, H., Alinejad-Rokny, H., Alizadeh, H., & Punch, W. F. (2014). Effects of resampling method and adaptation on clustering ensemble efficacy. *Artificial Intelligence Review*, *41*(1), 27–48. <https://doi.org/10.1007/s10462-011-9295-x>

- Mody, K., & Bekaiisab, T. (2018). Clinical trials and progress in metastatic colon cancer. *Surgical Oncology Clinics of North America*, 27(2), 349–365. <https://doi.org/10.1016/j.soc.2017.11.008>
- Palaga, T., Wongchana, W., & Kueanjinda, P. (2018). Notch signaling in macrophages in the context of cancer immunity. *Frontiers in Immunology*, 9, 652. <https://doi.org/10.3389/fimmu.2018.00652>
- Panayotis, Z., Kotsinas, A., Evangelou, K., Karakaidos, P., Vassiliou, L. V., Rezaei, N. ... Gorgoulis, V. G. (2004). Distinct expression patterns of the transcription factor E2F-1 in relation to tumour growth parameters in common human carcinomas. *Journal of Pathology*, 203(3), 744–753.
- Park, C. H., Rha, S. Y., Ahn, J. B., Shin, S. J., Kwon, W. S., Kim, T. S., ... Chung, H. C. (2015). PINCH-2 presents functional copy number variation and suppresses migration of colon cancer cells by paracrine activity. *International Journal of Cancer*, 136(10), 2273–2283.
- Peterson, A. T., Papeş, M., & Soberón, J. (2017). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, 213(1), 63–72.
- Pontes, F. J., Amorim, G. F., Balestrassi, P. P., Paiva, A. P., & Ferreira, J. R. (2016). Design of experiments and focused grid search for neural network parameter optimization. *Neurocomputing*, 186, 22–34. <https://doi.org/10.1016/j.neucom.2015.12.061>
- Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47. <https://doi.org/10.1093/nar/gkv007>
- Sarathi, A., & Palaniappan, A. (2019). Novel significant stage-specific differentially expressed genes in hepatocellular carcinoma. *BMC Cancer*, 19(1), 663. <https://doi.org/10.1186/s12885-019-5838-3>
- Shang, J., Wang, F., Chen, P., Wang, X., Ding, F., Liu, S., & Zhao, Q. (2018). Co-expression network analysis identified COL8A1 is associated with the progression and prognosis in human colon adenocarcinoma. *Digestive Diseases & Sciences*, 63(5), 1219–1228. <https://doi.org/10.1007/s10620-018-4996-5>
- Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G. S., Barzi, A., & Jemal, A. (2017). Colorectal cancer statistics. *CA: A Cancer Journal for Clinicians*, 67(3), 104–117. <https://doi.org/10.3322/caac.21395>
- Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *Ca A Cancer Journal for Clinicians*, 68(1), 11. <https://doi.org/10.3322/caac.21442>
- Sun, X., Zhai, H., Chen, X. I., Kong, R., & Zhang, X. (2018). MicroRNA-1271 suppresses the proliferation and invasion of colorectal cancer cells by regulating metadherin/Wnt signaling. *Journal of Biochemical and Molecular Toxicology*, 32(2), e22028. <https://doi.org/10.1002/jbt.22028>
- Sung, S. H., & Cho, M. S. (2006). Prognostic value of p16INK4a and p14ARF gene hypermethylation in human colon cancer. *Pathology Research & Practice*, 202(6), 415–424.
- Sy, J. P., & Taylor, J. M. (2015). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56(1), 227–236.
- Tang, C., Lu, W., Cai, Y., Han, L., & Wang, G. (2008). Nearly pre-processing-free method for skeletonization of gray-scale electronic speckle pattern interferometry fringe patterns via partial differential equations. *Optics Letters*, 33(2), 183–185. <https://doi.org/10.1364/OL.33.000183>
- Winchester, C., Robertson, S., MacLeod, T., Johnson, K., & Thomas, M. (2000). Expression of a homeobox gene (SIX5) in borderline ovarian tumours. *Journal of Clinical Pathology*, 53(3), 212–217.
- Xie, Y., Zheng, L., & Tao, L. (2019). Downregulation of IQGAP2 correlates with prostate cancer recurrence and metastasis. *Translational Oncology*, 12(2), 236–244. <https://doi.org/10.1016/j.tranon.2018.10.009>
- Xu, H. X., Wu, K. J., Tian, Y. J., Liu, Q., Han, N., He, X. L., ... Wu, K. M. (2016). Expression profile of SIX family members correlates with clinic-pathological features and prognosis of breast cancer: A systematic review and meta-analysis. *Medicine*, 95(27), e4085.
- Yalda, M., Basch, M. L., Pacheco, N. L., Gao, S. S., Wang, R., Harrison, W., ... Groves, A. K. (2014). The candidate splicing factor Sfsmap regulates growth and patterning of inner ear sensory organs. *PLoS Genetics*, 10(1): e1004055.
- Zhang, X., Hildebrandt, M. A., Horikawa, Y., Pu, X., Wood, C. G., & Gu, J. (2011). Abstract 1176: The expression of RN7SK paralogs in tumors of renal cell carcinoma is associated with patient survival. *Cancer Research*, 71(8 Suppl), 1176.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Yang H, Jin W, Liu H, et al.

A novel prognostic model based on multi-omics features predicts the prognosis of colon cancer patients. *Mol Genet Genomic Med*. 2020;8:e1255.

<https://doi.org/10.1002/mgg3.1255>