

Research Paper

Development and validation of a DNA repair gene signature for prognosis prediction in Colon Cancer

Xin Wang^{1,2,3*}, Cong Tan^{1,2,3*}, Min Ye^{1,2,3*}, Xu Wang^{1,2,3}, Weiwei Weng^{1,2,3}, Meng Zhang^{1,2,3}, Shujuan Ni^{1,2,3}, Lei Wang^{1,2,3}, Dan Huang^{1,2,3}, Zhaohui Huang⁴, Midie Xu^{1,2,3}✉ and Weiqi Sheng^{1,2,3}✉

1. Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, 200032, China.
2. Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, 200032, China.
3. Institute of Pathology, Fudan University, Shanghai 200032, China.
4. Wuxi Cancer Institute, Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, China.

*These authors contributed equally to this work.

✉ Corresponding authors: Weiqi Sheng, PhD, Department of Pathology, Fudan University Shanghai Cancer Center, 270 Dong'an Road, Shanghai 200032, China. E-mail: shengweiqi2006@163.com; Tel.: 86-21-64175590; Fax: 86-21-64174774; and, Midie Xu, PhD, Department of Pathology, Fudan University Shanghai Cancer Center, 270 Dong'an Road, Shanghai 200032, People's Republic of China. E-mail: xumid27202003@sina.com; Tel.: +86-21-64175590; Fax: +86-21-64174774.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2020.03.25; Accepted: 2020.07.31; Published: 2020.08.12

Abstract

Aberrant expression of DNA repair genes (DRGs) can be related to tumor progression and clinical outcomes in colon cancer. Here, we aimed to establish a DRGs signature to identify the vital prognostic DRGs in colon cancer. Firstly, gene set enrichment analysis (GSEA) was performed to demonstrate the association between abnormal expression level of DRGs and tumorigenesis. Then, a total of 476 DRGs were obtained for detecting candidate biomarkers in randomly selected 295 cases from The Cancer Genome Atlas (TCGA) colon cancer cohort. Eleven genes were screened by LASSO Cox regression analyses to develop the prognostic model. Then, the prognostic model and the expression levels of the eleven genes were validated using the internal validation dataset (the rest 125 cases in TCGA cohort) and an external validation dataset (obtained from Gene Expression Omnibus dataset). Further analysis revealed the independent prognostic capacity of the prognostic model in relation to other clinical characteristics. The receiver operating characteristic (ROC) curve analysis confirmed the good performance of the prognostic model. Furthermore, we provided a nomogram for interpreting the clinical application of the 11-DRG signature. In conclusion, we propose a newly developed 11-DRG signature as a practical prognostic predictor for patients with colon cancer, which can facilitate the individualized counselling and treatment.

Key words: colon cancer, DNA repair, prognosis, TCGA, GEO

Introduction

Colon cancer is one of the most common and lethal malignancies with an estimated 1,096,601 new cases and 551,269 deaths reported in Globocan 2018 [1]. A large amount of patients benefits from early detection and comprehensive treatment. However, clinical outcomes of patients occasionally present unpredictable diversity even with similar treatment strategies [2, 3]. Although much effort has been made to identifying pathological or biochemical markers for prognosis prognostication, results from single factor is unsatisfying. Recently, with advancement of the next generation sequencing technology, the new

insight into the molecular network of cancer hallmarks and the availability of high-throughput genome and transcriptome data offer us a richer theoretical and practical basis for prognosis prediction of colon cancer [4, 5].

DNA repair occurs all the time in human cells to identify and correct damage to the DNA molecules that encode its genome [6]. Complex signaling pathways, including base excision repair, nucleotide excision repair, and mismatch repair, work to deal with varieties of DNA damage from endogenous or exogenous sources. DNA damage and repair is a

widely concerned hallmark in tumor biology [7]. Researchers have reported the association between the aberrant expression of DNA repair genes (DRGs) with the tumor generation [8-10]. Disorders in DNA repair process has been recognized as one of the critical characteristics in tumor generation and progress in molecular biological behavior in colon cancer [11, 12]. In view of the important role that DRGs play in colon cancer, attentions should be paid to understand their connections with prognosis.

In this study, we aim to identify the activity of DRGs in colon cancer and establish a prognostic modal with DRGs. Firstly, we noted the differences in the transcriptional profile of epithelium cells, and predicted the actions of DRGs during the malignant transformation of colon by functional analysis. Then, we applied LASSO Cox regression analysis to establish a prognostic model for patients with colon cancer. Furthermore, we performed internal and external data validation to assess the robustness of the classifier. Finally, we analyzed the relationship between the modal and clinical information. The results demonstrated the independency of our 11-DRG signature in predicting clinical outcomes for patients with colon cancer.

Material and Methods

Data download and processing

Transcriptome profiling and clinical data of TCGA-colon adenocarcinoma (COAD) cohort were downloaded from the GDC Data Portal (<https://portal.gdc.cancer.gov/>). Only samples with complete prognostic data were included. The pathologic stages of 420 cases were reconfirmed according to the seventh edition of American Joint Committee on Cancer staging system [13]. TCGA-COAD counts were normalized with R package "DESeq2". Four colon cancer datasets, including GSE21510, GSE24514, GSE32323 and GSE39582 (Table S1), were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). All the GEO datasets were normalized with log₂ transformation and probe annotation by using "limma" package, if necessary. For DRGs identification, we choose "GO_DNA_REPAIR" gene list, containing 544 DRGs, downloaded from "GO biological process" gene set, Gene Set Enrichment Analysis (GSEA), to obtain sufficient DRGs. Finally, 476 DRGs were annotated and detected in above datasets (Table S2).

Gene Set Enrichment Analysis (GSEA)

Samples from three GEO datasets, GSE21510, GSE24514 and GSE32323, were divided into normal colon and tumor group depended on pathologic

diagnosis. Functional analysis was conducted by "GSEA" function using "clusterProfiler" package. Hallmarks gene set "h.all.v7.0.entrez.gmt" was set as the reference gene database, and cutoff criteria was set to $p < 0.05$ and $q < 0.25$. Common hallmarks of GSEA were presented in Venn diagram.

Lasso-Cox regression modal

The Cox proportional hazards model with LASSO penalty could establish the relationship between predictor genes and survival time, and at the same time, reduce the coefficients of correlated predictors while making pick or discard decision [14]. Here, we tried to establish a DRGs prognostic modal based on LASSO Cox regression, with "glmnet" package in R. A total of 476 DRGs were included for detecting candidate biomarkers. The optimal penalty parameter λ (lambda) was selected by cross-validation method, "lambda.min" was determined as the best selection, which is the values of λ that gives minimum mean cross-validated error. Finally, the modal output a risk score of each patient, which was calculated by summing relative expression levels of the vital DRGs (Exp_i) and the corresponding LASSO coefficient (L_i): Risk Score = $\sum_{i=1}^n Exp_i \times L_i$.

To predict the prognosis, patients were divided into high-risk or low-risk groups, according to the risk score. The best cutoff of risk score was determined when optimal Area Under Curve (AUC) in Receiver operating characteristic (ROC) curve achieved for predicting 5-year survival in the training set, by utilizing the assessment of Youden's index [15, 16]. Kaplan-Meier survival analysis was used to predict clinical outcomes for patients in each cohort. ROC curve was conducted to calculate AUC for one-year, three-year and five-year overall survival in training, test, and validation set, to estimate the accuracy of the modal.

Statistical assessment of the modal and visualization

Boxplot was used to display the distribution of Risk Score in patient with various clinic characteristics. *T* test or one-way ANOVA was applied to evaluate differences between groups. Relationship between risk group of patients and clinic parameters was assessed by chi-square test or fisher exact test, as appropriate. Univariate and multivariate Cox regression of the modal was used to test independency of the modal. Nomogram and calibration analysis were plotted by R package "rms". All statistical analysis was conducted in R software (R 3.5.3, <https://www.r-project.org/>). Significant level of *P* value was set as < 0.05 .

Results

Aberrant expression level of DNA repair genes involves in tumorigenesis of colon cancer

First, we collected transcriptome sequencing data of 231 human tissue samples from three independent GEO datasets (GSE21510, GSE24514 and GSE32323). Heat maps showed that DNA repair genes displayed different expression patterns in normal colon mucosa and colon cancer samples (**Figure 1A**). Moreover, we performed GSEA for hallmarks gene sets in the three datasets. Reports demonstrated that DNA repair pathways were significantly activated in all datasets during the transformation from normal mucosa to colon cancer (**Figure 1B-C**). The normalized enrichment score (NES) of HALLMARK_DNA_REPAIR was 1.7158 (adjust $P = 0.0052$) in GSE21510, 2.2097 (adjust $P = 0.0051$) in GSE24514, and 1.7122 (adjust $P = 0.0058$) in GSE32323, respectively (**Figure 1D**). The results suggested that DRGs play important roles in tumorigenesis in colon cancer and the recognition of key prognostic DRGs might be of great significance in understanding tumor pathological physiology.

Identification of survival-related DRGs and establishment of the eleven-gene prognostic signature TCGA-COAD cohort

A total of 420 patients diagnosed as colon adenocarcinoma from TCGA-COAD data set was enrolled for establishment of the DRG signature, and 2/3 of them ($n = 295$) were randomly assigned to the training set. The baseline information of the patients in the training and internal validation cohort was listed in **Table S3-S4**. LASSO Cox regression was conducted to screen survival-related DRGs to predict OS and calculated the risk score (**Figure S1**). An eleven-DRGs-based prognosis risk score was established, among which six risk and five protective genes were identified for indicating clinical outcomes. The formula of risk factor was as follow:

$$\begin{aligned} \text{Risk Score} = & \text{ACTR8} * (-0.23337) + \text{DMC1} * (-0.16365) + \\ & \text{MAGEF1} * 0.042617 + \text{MC1R} * 0.108694 + \text{POLG} * \\ & 0.05497 + \text{RBM17} * 0.418946 + \text{SFPQ} * (-0.22302) + \\ & \text{TERF2IP} * 0.135403 + \text{TP53BP1} * 0.092702 + \text{UIMC1} * \\ & (-0.20938) + \text{USP7} * (-0.04848). \end{aligned}$$

According to the risk factor, patients could be divided into the high risk and the low risk groups, and the cutoff value, obtained by utilizing the assessment of Youden's index, was set at 0.18 for best prediction accuracy (**Figure 2A**). In the training set, Kaplan-Meier survival analysis suggested prognosis of patient in the low risk groups is significantly better than those in high risk groups ($P < 0.001$, **Figure 2B**).

ROC curve displayed that the accuracy of the prognostic DRG signature for 1-year, 3-year and 5-year survival was 0.773, 0.775 and 0.751, respectively (**Figure 2C**). Distribution of the risk score, hazard ratio and gene expression of the 11 DRGs in the training dataset were showed in **Figure 2D**. Moreover, analysis of the correlation of risk stratification with clinicopathological data for patients with colon cancer showed that pathological T, pathological N and TNM stage were significant associated with risk stratification in the training set (**Table 1**).

Internal validation of 11-DRG signature in TCGA-COAD cohort

To validate the predictive capability of the DRG signature as a prognostic indicator, the rest 125 samples of TCGA-COAD cohort were used for internal validation, which showed significant differences of overall survival between distinct risk groups ($P < 0.0001$, **Figure 3A-B**). Result of ROC analysis showed the AUC for OS prediction was 0.910 for 1-year, 0.599 for 3-year and 0.827 for 5-year in the internal validation set, respectively (**Figure 3C**). Distribution of the risk score, hazard ratio and gene expression of the 11 DRGs in the internal validation dataset were showed in **Figure 3D**. However, there was no significant difference between clinical characteristics and risk stratification in the internal validation set (**Table 1**). Distribution of various clinical parameters at low or high risk group was plotted in **Figure S2**. Furthermore, the stratified analyses of patients with different clinicopathological characteristics of the whole TCGA-COAD cohort showed that OS of the patients could be significantly distinguished in all subgroups (**Figure 4**).

External validation of 11-DRG signature in independent GEO colon cancer cohort

Another GEO dataset (GSE39582), which containing 550 patients with definite diagnosis of colon cancer, was served as the external validation set (**Table S5**). Significant differences of overall survival between distinct risk groups were observed in external validation set ($P < 0.001$, **Figure 5A-B**). In dataset GSE39582, the AUC was 0.663, 0.610 and 0.622 for 1-year, 3-year, 5-year survival, respectively (**Figure 5C**). Distribution of the risk score, hazard ratio and gene expression of the 11 DRGs in the external validation dataset were showed in **Figure 5D**. Moreover, analysis of the correlation of risk stratification with clinicopathological data for patients with colon cancer also showed that pathological T, pathological N and TNM stage were significant associated with risk stratification in the external

validation set (Table 1). The stratified analyses of patients with different clinicopathological characteristics from the training set showed that OS of the patients could be significantly distinguished in most

of subgroups except patients younger than 65-year-old, at T4 stage, N1-2 stage, or M1-x stage (Figure S3).

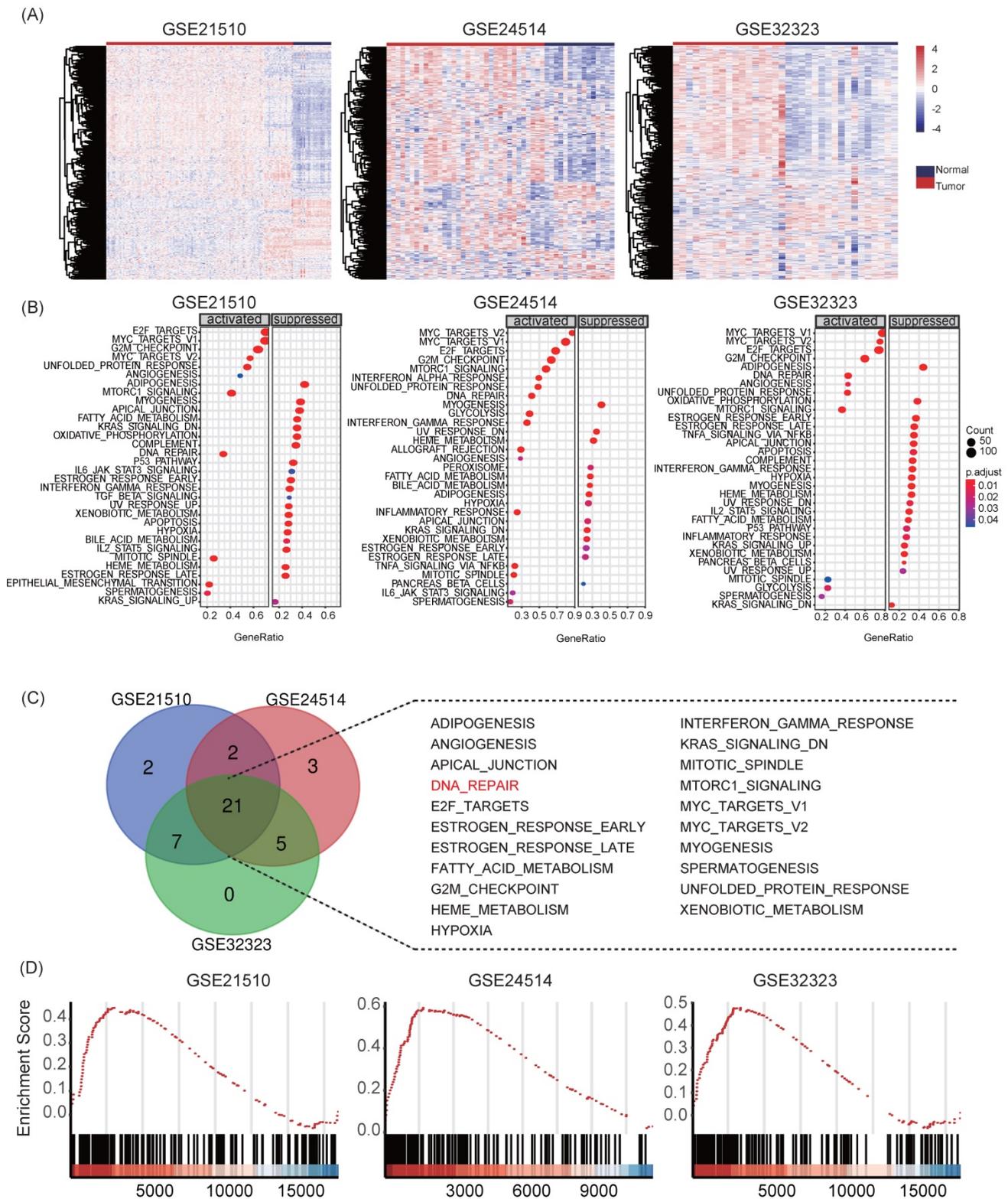


Figure 1. The DNA repair signaling pathway is up-regulated in development in colon cancer. **(A)** Heatmaps of GSE21510, GSE24514 and GSE32323 data sets. Different patterns of transcriptional expression were observed in tumor and non-tumor samples. **(B-D)** Enriched pathways of GSEA, HALLMARK_DNA_REPAIR was activated in all three data sets.

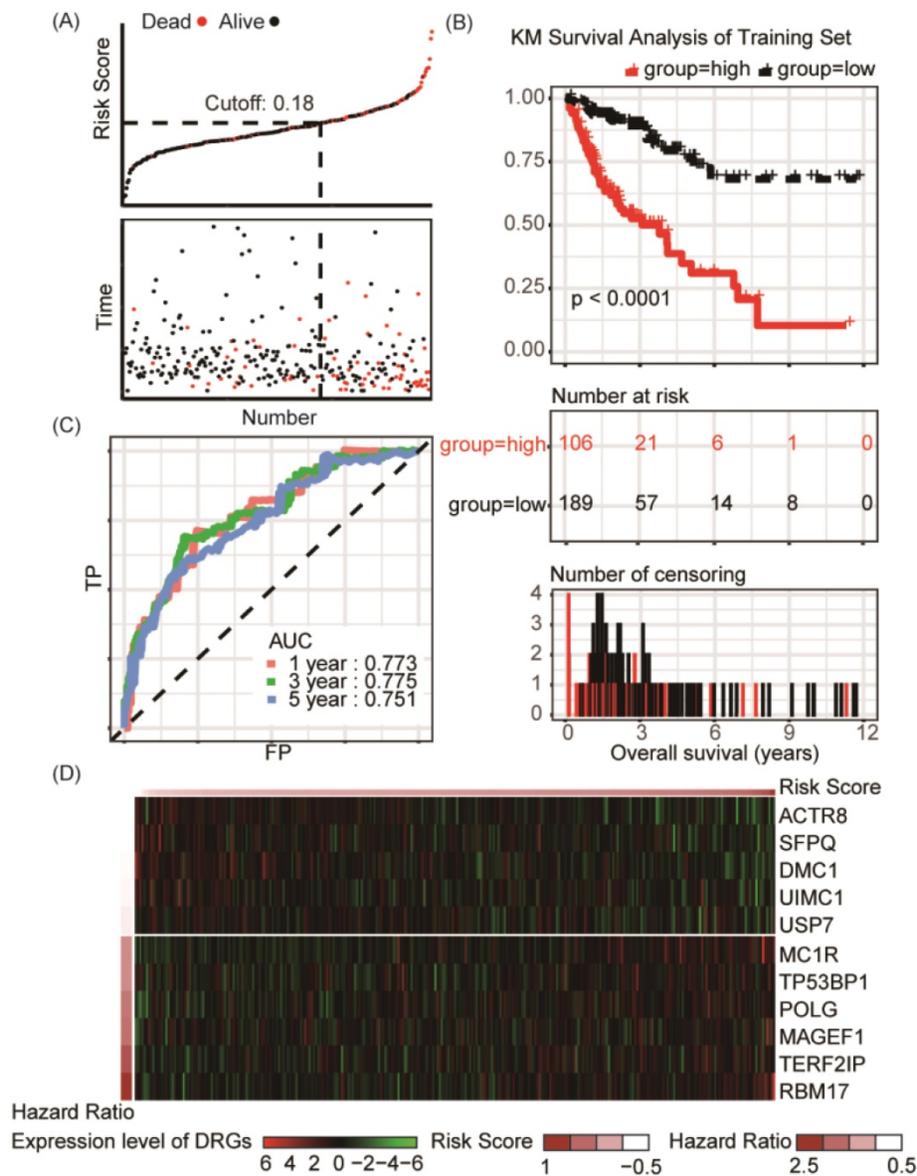


Figure 2. Risk group identified by the DRG classifier, KM survival analysis and ROC curve of TCGA-COAD training set. **(A)** The relationship between alive/dead status with Risk Score and survival time (years). The cutoff of Risk Score was set at 0.18. **(B)** KM survival analysis of overall survival for high-risk or low-risk group patients. **(C)** ROC analysis of the eleven-DRG prognostic signature. The AUC for 1-year, 3-year, 5-year predicting were 0.773, 0.775 and 0.751, respectively. **(D)** Heatmap displayed the expression level of eleven DRGs.

Evaluation of independent prognostic factors in colon cancer

Prognostic factors of overall survival for colon cancer in the TCGA-COAD set were identified using univariate and multivariate Cox regression analyses. The results revealed that Risk Score and pathologic M stage were independent indicators (Table 2). This suggested that the 11-DRG signature has good independence in clinical application. In ROC curve comparing multi-factors to predict survival probability of 1 year, 3 years and 5 years, the DRG signature showed better capability of prognostic prediction than pathologic M stage in 1-year and 3-year OS (Figure 6).

To establish a clinically applicable method for predicting the survival probability of patients with colon cancer, we developed a nomogram to predict the probability of the 1-, 3- and 5-year OS in the whole TCGA-COAD cohort. The predictors of the nomogram included both independent prognostic factors (pathologic M stage and 11-DRG signature; Table 2). Finally, we drew a nomogram for predicting the survival probability of patients considering the feasibility of clinical practice (Figure 7A). Calibration plots demonstrated stability of the nomogram in predicting 3- or 5- year overall survival (Figure 7B). These findings demonstrated that the nomogram is a robust model for predicting survival for patients with colon cancer, which might facilitate patient

counselling, decision-making and follow-up scheduling.

Discussion

Increasing evidences has elucidated the roles of DNA damage and repair deficiencies in malignancies, including colon cancer [7]. Different expression patterns of DRGs are frequently detected in normal mucosa and cancer tissues, which were closely related to patients' prognosis [17, 18]. However, large variations in the mechanisms of tumorigenesis and heterogeneity of tumors suggest the obviously limited prognostic value of individual gene detection. Establish a multi-gene prognostic panel instead of a single gene biomarker provides potentially more optimal feasibility in predicting clinical outcome for patients with colon cancer [19, 20]. Up to now, in colon cancer, no applicable research on transcriptional patterns of DRGs and its prospective prognostic value has been reported. Thus, we developed an 11-DRG signature predicting the survival outcomes of patients with colon cancer. In the present study, by screened the candidate prognostic signature in TCGA and GEO colon cancer datasets, a 11-DRG signature was identified to be able to stratified the survival risk of patients with colon cancer, and the overall survival of high-risk group was significantly worse than that of low-risk group. The signature also performed well in internal and external validation cohorts. Univariate and multivariate Cox regression determined that Risk Score was an independent prognosis factor for OS in

patients with colon cancer. Furthermore, a nomogram was developed and validated to predict OS based on the age of patients. The nomogram provided favorable discrimination and calibration plots. The results of this nomogram may help to optimize the preoperative management of patients with colon cancer.

In the current study, transcriptome sequencing data and clinical information was obtained from public database, TCGA and GEO. To confirm the involvement of DRGs in development of colon cancer, we conducted analysis of expression profiles for patients from three GEO datasets, discovering higher transcriptional activity of DRGs in the most neoplastic epithelium. Furthermore, all of the datasets dramatically achieved "HALLMARK_DNA_REPAIR" enrichment in GSEA, which suggested that the abundance of DRGs might be a candidate indicator of malignant transformation proceedings in colon cancer and, to some extent, high levels of DRGs expression might be protective factors for prognostic prediction. Given the fact above, we established a novel prognostic DRG signature consist of 11 genes, which was validated as an independent predictor of patient's survival for colon cancer. The Risk Score assessment could successfully classify the patients into high or low risk group with significant differences in overall survival. In addition, further validation was conducted in an internal data set and an independent external set, which reflected the good accuracy and reproducibility of the modal.

Table 1. Associations with risk group and clinical characteristics in the training and validation sets

	TCGA-COAD training set			TCGA-COAD validation set			GSE39582		
	High risk	Low risk	P	High risk	Low risk	P	High risk	Low risk	P
Gender			0.929			0.265			0.957
Female	48	88		24	33		142	105	
Male	58	101		21	47		176	127	
Age	0.979					0.744			0.988
<65	40	73		17	34		122	88	
≥65	66	116		28	46		196	144	
T stage			0.000			0.805			0.000
T1	3	3		2	3		1	14	
T2	7	46		7	13		23	20	
T3	76	121		31	59		207	148	
T4	20	19		5	5		75	42	
N stage			0.001			0.179			0.016
N0	46	124		25	49		153	139	
N1	32	41		8	20		83	51	
N2	28	24		12	11		70	34	
M stage	0.070					0.173			0.368
M0	69	150		30	61		266	203	
M1	20	22		10	8		38	21	
Mx	13	16		5	11		2	0	
Stage			0.001			0.321			0.003
I	7	42		7	14		12	24	
II	37	74		15	32		140	116	
III	38	47		12	24		129	71	
IV	20	22		10	8		37	21	

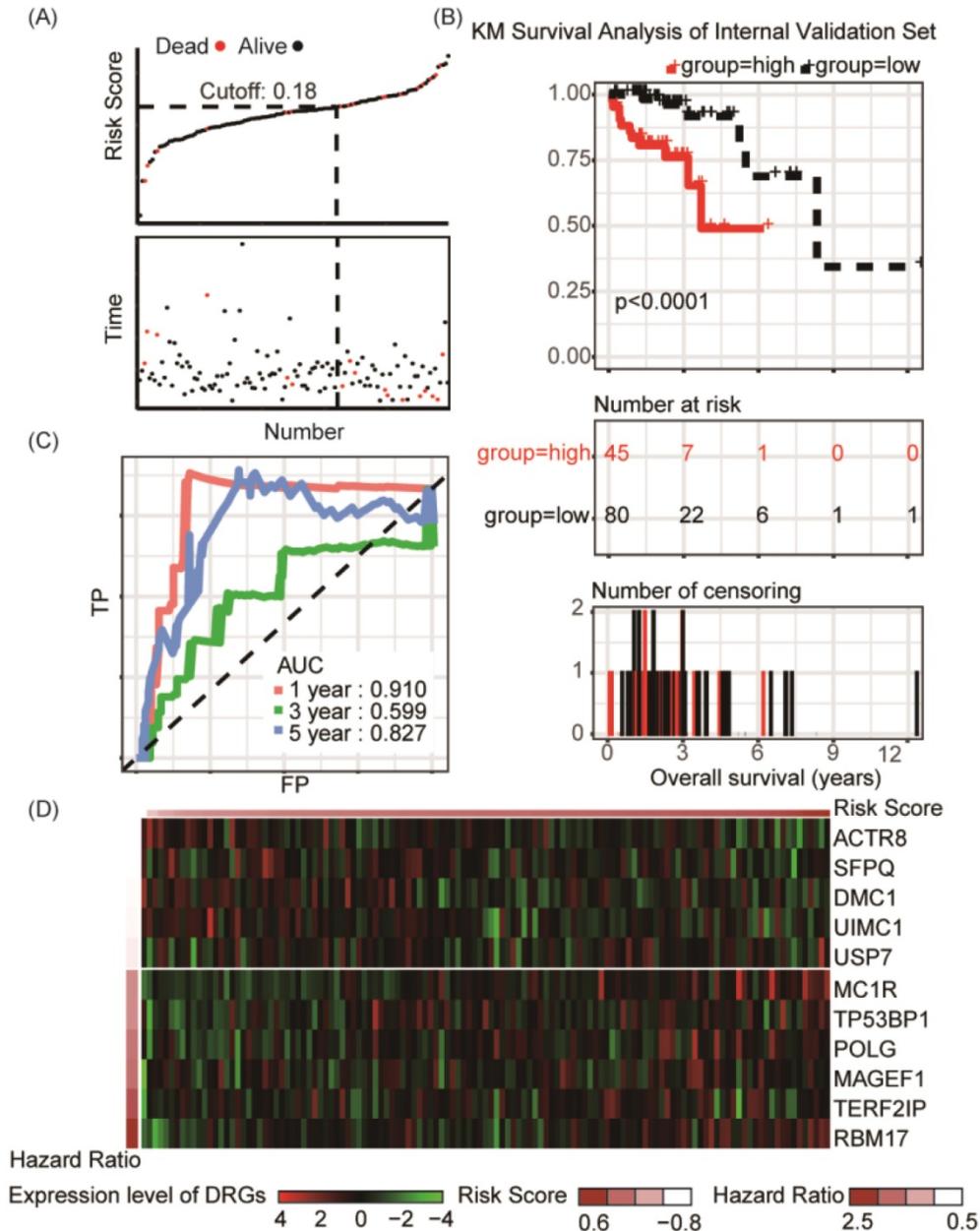


Figure 3. Risk group identified by the DRG classifier, KM survival analysis and ROC curve of TCGA-COAD internal validation set. **(A)** KM survival analysis of overall survival for high-risk or low-risk group patients. **(B)** The relationship between alive/dead status with Risk Score and survival time (years). The cutoff of Risk Score was set at 0.18. **(C)** ROC analysis of the eleven-DRG prognostic signature. The AUC for 1-year, 3-year, 5-year predicting were 0.910, 0.599 and 0.827, respectively. **(D)** Heatmap displayed the expression level of eleven DRGs.

Table 2. Univariate and multivariate Cox analysis of Risk Score and clinical characteristics on overall survival in TCGA-COAD

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
Age	1.02 (1.00-1.04)	0.064		
Gender	1.19 (0.79-1.81)	0.405		
Stage	2.40 (1.87-3.09)	<0.001	1.58 (0.63-0.99)	0.058
T	2.95 (1.95-4.46)	<0.001	1.42 (0.76-2.64)	0.400
N	2.16 (1.69-2.76)	<0.001	0.66 (0.26-1.71)	0.606
M	3.80 (2.48-5.83)	<0.001	2.37 (1.43-3.90)	0.047
Risk Score	24.27 (11.21-52.55)	<0.001	14.65 (6.53-32.83)	<0.001

In our study, we identified 11 DRGs (ACTR8,

DMC1, MAGEF1, MC1R, POLG, RBM17, SFPQ, TERF2IP, TP53BP1, UIMC1, USP7), whose altered expression level was closely related to the prognosis of colon cancer patients. TP53BP1, also known as 53BP1, encodes a kind of chromatin-binding protein, which is a key component in DNA double-strand break signaling in response to DNA damage by promoting non-homologous and joining mediated repair [21, 22]. The functions of TP53BP1 in chromatin stability determines its critical role in cancer and reports emergence in breast, lung, prostate and colorectal cancer [23-27]. Bi *et al.* [28] demonstrated

that deficiency of TP53BP1 significantly inhibited apoptosis of tumor cells and prompted proliferation and S phase accumulation in cell cycle. UIMC1 (also known as RAP80) is a member of BRCA1-A complex, and engages in checkpoint arrest in cell cycle and was lately recognized as a regulator in tumor cell apoptosis and epithelium mesenchymal transition [29, 30]. SFPQ (also named as PSF) translates protein binding to a nuclear receptor PPAR γ and modulate growth of colon cancer cells [31]. Although TP53BP1 and UIMC1 have been identified as tumor suppressor [28-30] and SFPQ has been identified as oncogene [31], TP53BP1 was ranked as risk factor, while UIMC1 and SFPQ are protect factor in our 11-gene signature, which revealed the complication of molecular

signaling, specific gene may play totally opposite role during tumorigenesis and development. This phenomenon is an important research point and needs for further more attention. Besides, roles of some of the candidate genes that play in colon cancer have not been revealed. However, our analysis suggested these DRGs might provide promising value in colon cancer progression.

In conclusion, the current study proposed a newly developed eleven-DRG signature as a practical prognostic predictor for patients with colon cancer, which can contribute independent value in identifying clinical outcomes that complements the TNM system in colon cancer.

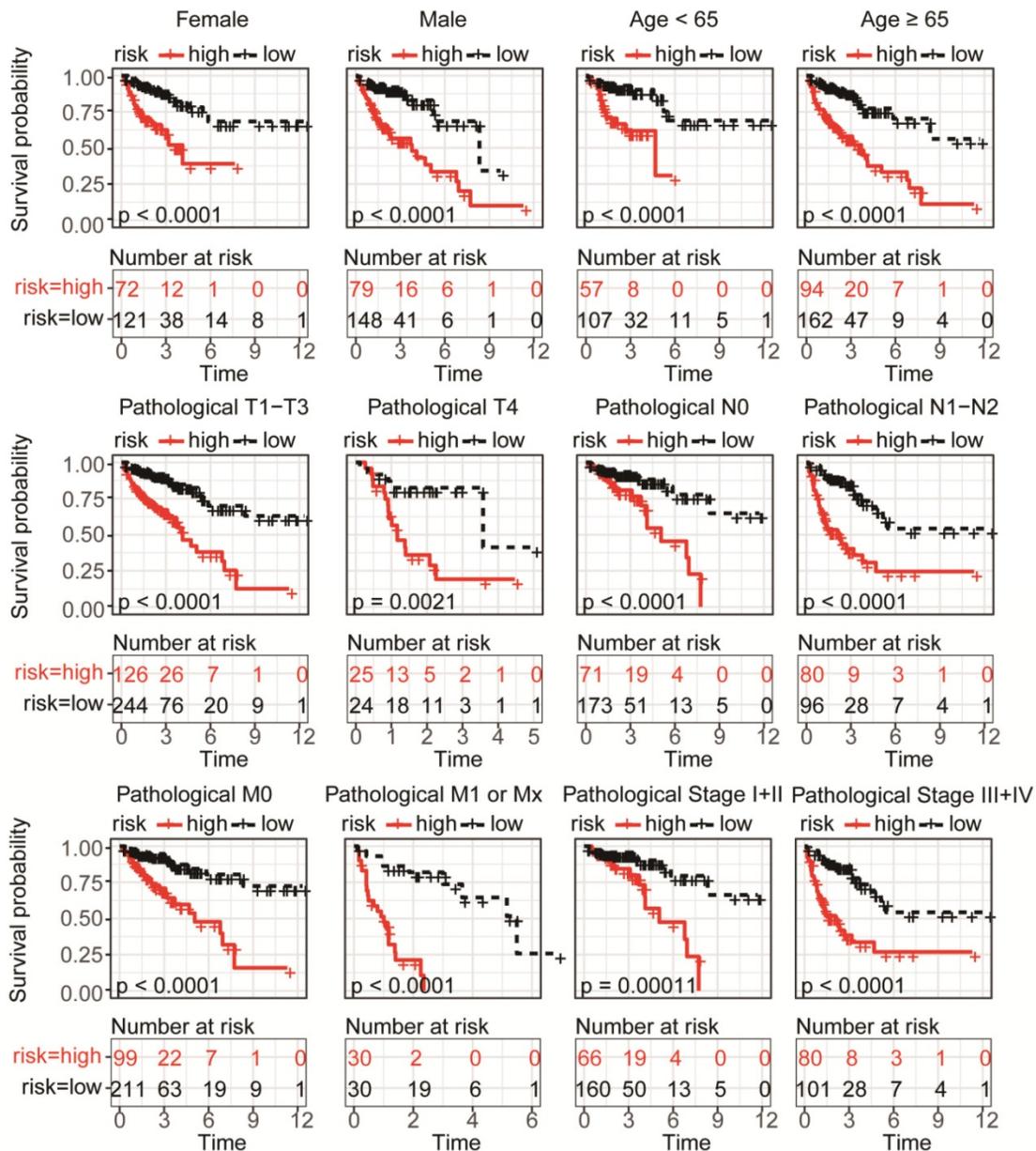


Figure 4. Subgroup KM analysis in high or low risk group patients of TCGA-COAD according to clinical characteristics. Significance differences of overall survival was detected in all subgroup analysis, including distinct gender, age, pathological T, N, M and stage.

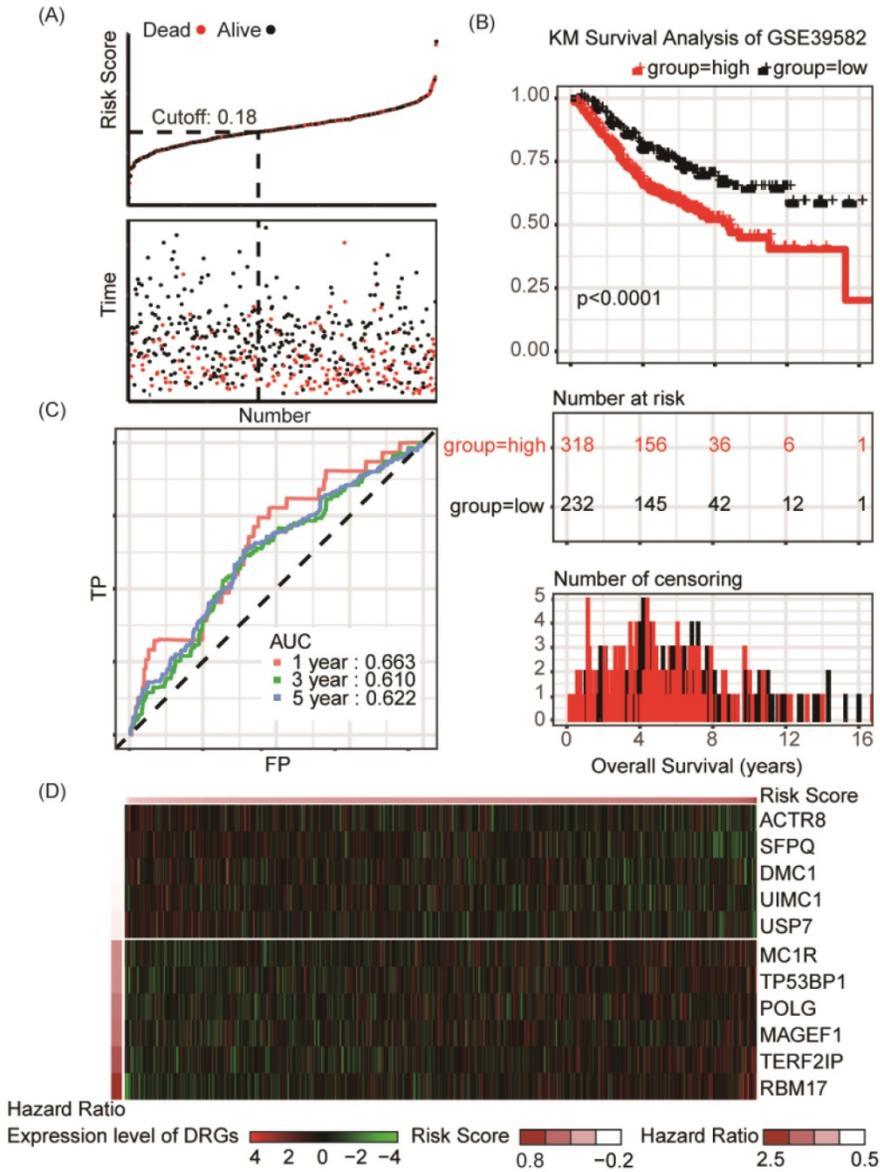


Figure 5. Risk group identified by the DRG classifier, KM survival analysis and ROC curve of GSE39582 dataset. **(A)** The relationship between alive/dead status with Risk Score and survival time (years). The cutoff of Risk Score is set at 0.18. **(B)** KM survival analysis of overall survival for high-risk or low-risk group patients. **(C)** ROC analysis of the eleven-DRG prognostic signature. The AUC for 1-year, 3-year, 5-year predicting were 0.663, 0.610 and 0.622, respectively. **(D)** Heatmap displayed the expression level of eleven DRGs.

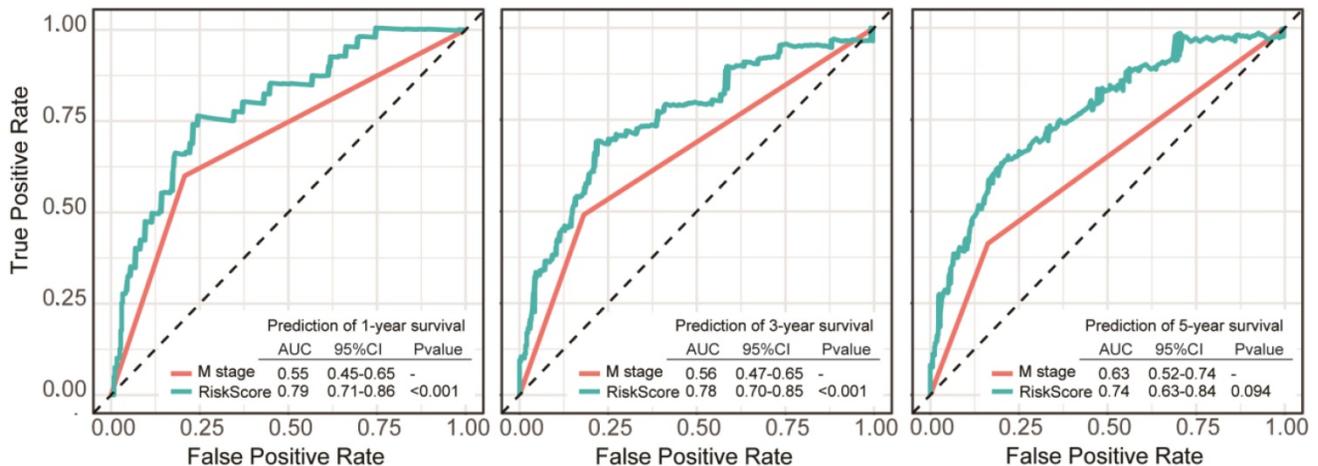


Figure 6. Comparisons of sensitivity and specificity for survival prediction by DRG signature and pathologic M stage as independent factors. The eleven-DRG signature showed a better capability for survival prediction than pathologic M stage. Significant differences reached at 1-year and 3-year prediction.

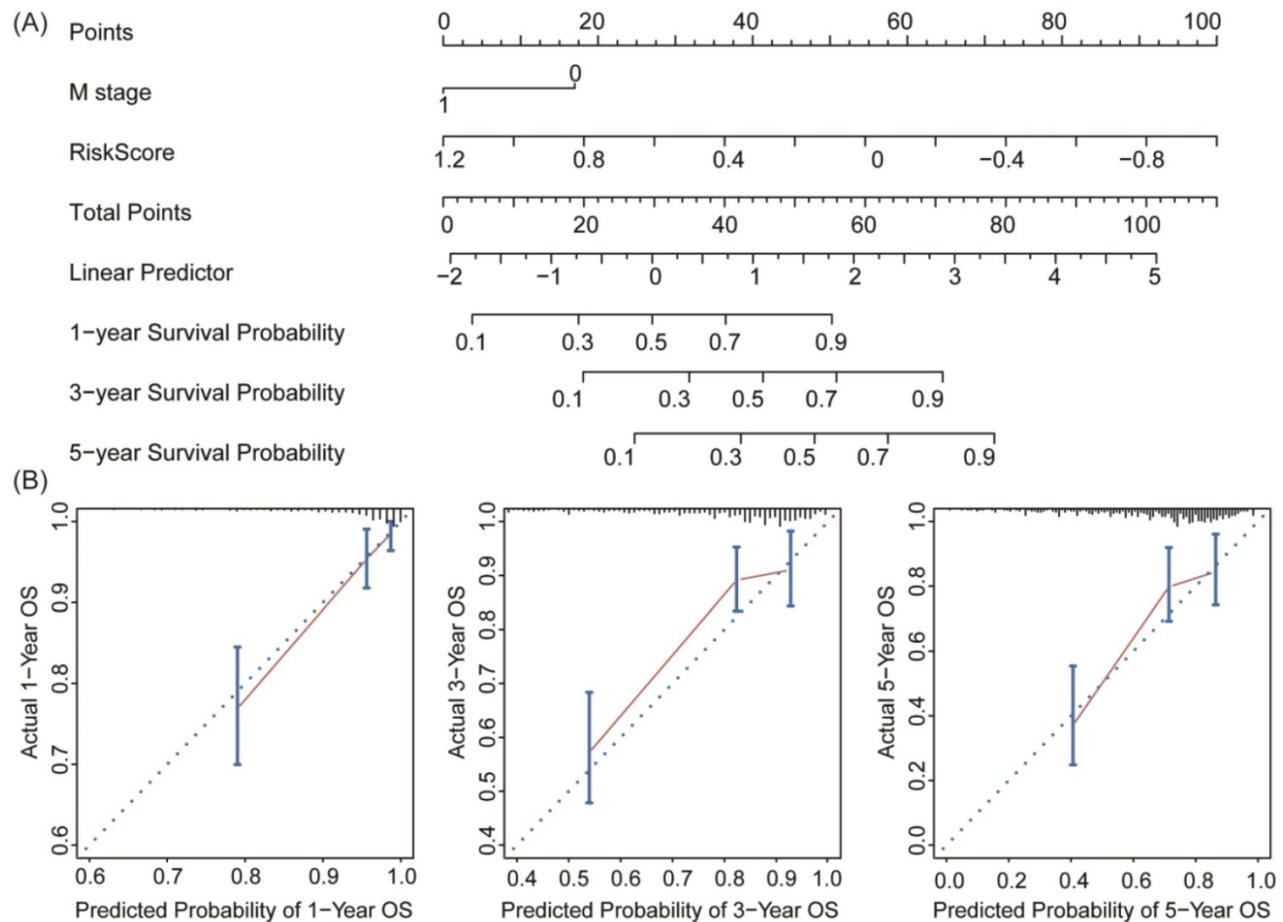


Figure 7. Nomogram and calibration analysis for the DRG prognostic signature. **(A)** Nomogram plotted by the independent factors of patients' survival. The probability of long-term survival can be calculated by adding the corresponding points of M stage and Risk Score in the nomogram. **(B)** Calibration plots displayed the relationship between actual and the nomogram-predicted survival, which indicated a powerful predicting capability of the nomogram.

Supplementary Material

Supplementary figures and tables.

<http://www.jcancer.org/v11p5918s1.pdf>

Acknowledgements

The authors would like to thank all researchers contributed to the TCGA and GEO data sets included.

Author Contributions

WS and MX designed the study. XW conducted the data process, modal establishment and visualization of analysis. MX, CT, MY and WW did the data analysis and interpretation. SN and MZ performed statistical analysis. CT and LW, WS and MX planned and supervised the project, performed data analysis and wrote the manuscript. CT and DH revised the manuscript. All authors have read and approved the final manuscript.

Funding

The study was supported by National Natural Science Foundation of China (81972249, 81802367,

81802361), Shanghai Science and Technology Development Fund (18ZR1408000, 17ZR1406500 and 15ZR1407400), Shanghai Science and technology development fund (19MC1911000), Clinical Research Project of Shanghai Municipal Health Committee (20194Y0348), National Key R&D Program of China (2017YFC1311004), National Human Genetic Resources Sharing Service Platform (2005DKA21300), and Hospital Foundation of Fudan University Shanghai Cancer Center (YJMS201907, YJQN201906, YJ201704).

Availability of data and materials

The datasets generated and analyzed during the current study are available in the TCGA repository (<https://portal.gdc.cancer.gov/>) and the GEO repository (<https://www.ncbi.nlm.nih.gov/geo/>).

Competing Interests

The authors have declared that no competing interest exists.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018; 68: 394-424.
2. Pages F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet (London, England)*. 2018; 391: 2128-39.
3. Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, et al. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. *Cancer immunology, immunotherapy : CII*. 2019; 68: 433-42.
4. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhorji G, et al. A pathology atlas of the human cancer transcriptome. *Science (New York, NY)*. 2017; 357.
5. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature*. 2016; 531: 47-52.
6. Lans H, Hoeijmakers JHJ, Vermeulen W, Marteijn JA. The DNA damage response to transcription stress. *Nature reviews Molecular cell biology*. 2019; 20: 766-84.
7. Nickoloff JA, Jones D, Lee SH, Williamson EA, Hromas R. Drugging the Cancers Addicted to DNA Repair. *Journal of the National Cancer Institute*. 2017; 109.
8. Laurini E, Marson D, Fermeglia A, Aulic S, Fermeglia M, Priel S. Role of Rad51 and DNA repair in cancer: A molecular perspective. *Pharmacology & therapeutics*. 2020; 208: 107492.
9. Baretta M, Le DT. DNA mismatch repair in cancer. *Pharmacology & therapeutics*. 2018; 189: 45-62.
10. Mateo J, Boysen G, Barbieri CE, Bryant HE, Castro E, Nelson PS, et al. DNA Repair in Prostate Cancer: Biology and Clinical Implications. *European urology*. 2017; 71: 417-25.
11. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology*. 2015; 149: 1177-90.e3.
12. Reilly NM, Novara L, Di Nicolantonio F, Bardelli A. Exploiting DNA repair defects in colorectal cancer. *Molecular Oncology*. 2019; 13: 681-700.
13. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of surgical oncology*. 2010; 17: 1471-4.
14. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*. 2011; 39: 1-13.
15. Nies M, Everett A. Biomarkers in Care of Congenital Heart Disease Patients in the Intensive Care Unit. In: Ungerleider RM, Meliones JN, Nelson McMillan K, Cooper DS, Jacobs JP, editors. *Critical Heart Disease in Infants and Children (Third Edition)*. Philadelphia: Elsevier. 2019. p. 480-7.e2.
16. Kallner A. Interpretation of the Elements of the ROC Analysis. In: Kallner A, editor. *Laboratory Statistics (Second Edition)*: Elsevier; 2018. p. 1-140.
17. Jinjia C, Xiaoyu W, Hui S, Wenhua L, Zhe Z, Xiaodong Z, et al. The use of DNA repair genes as prognostic indicators of gastric cancer. *Journal of Cancer*. 2019; 10: 4866-75.
18. Kun S, Duan Q, Liu G, Lu JM. Prognostic value of DNA repair genes based on stratification of glioblastomas. *Oncotarget*. 2017; 8: 58222-30.
19. Deng CX, Wang RH. Roles of BRCA1 in DNA damage repair: a link between development and cancer. *Human molecular genetics*. 2003; 12 Spec No 1: R113-23.
20. Fang M, Xia F, Mahalingam M, Virbasius CM, Wajapeyee N, Green MR. MEN1 is a melanoma tumor suppressor that preserves genomic integrity by stimulating transcription of genes that promote homologous recombination-directed DNA repair. *Molecular and cellular biology*. 2013; 33: 2635-47.
21. Noordermeer SM, Adam S, Setiawati D, Barazas M, Pettitt SJ, Ling AK, et al. The shieldin complex mediates 53BP1-dependent DNA repair. *Nature*. 2018; 560: 117-21.
22. Panier S, Boulton SJ. Double-strand break repair: 53BP1 comes into focus. *Nature Reviews Molecular Cell Biology*. 2014; 15: 7-18.
23. Du R, Zheng L, Huang WT, Zhang HZ, Jiang ZM. Correlation of 53BP1 gene mutation with prostatic adenocarcinoma. *Chinese Journal of Pathology*. 2011; 40: 449-53.
24. Lai TC, Chow KC, Lin TY, Chiang IP, Fang HY, Chen CY, et al. Expression of 53BP1 as a cisplatin-resistant marker in patients with lung adenocarcinomas. *Oncology Reports*. 2010; 24: 321-8.
25. Li X, Kong X, Kong X, Wang Y, Yan S, Yang Q. 53BP1 Sensitizes Breast Cancer Cells to 5-Fluorouracil. *PLOS ONE*. 2013; 8: e74928.
26. Xiao Y, Zheng X, Huang A, Liu T, Zhang T, Ma H. Deficiency of 53BP1 inhibits the radiosensitivity of colorectal cancer. *International Journal of Oncology*. 2016; 49: 1600-8.
27. Mirza-Aghazadeh-Attari M, Mohammadzadeh A, Yousefi B, Mihanfar A, Karimian A, Majidinia M. 53BP1: A key player of DNA damage response with critical functions in cancer. *DNA Repair*. 2019; 73: 110-9.
28. Bi J, Huang A, Liu T, Zhang T, Ma H. Expression of DNA damage checkpoint 53BP1 is correlated with prognosis, cell proliferation and apoptosis in colorectal cancer. *International Journal of Clinical and Experimental Pathology*. 2015; 8: 6070-82.
29. Park SY, Korm S, Chung HJ, Choi SJ, Jang JJ, Cho S, et al. RAP80 regulates epithelial-mesenchymal transition related with metastasis and malignancy of cancer. *Cancer science*. 2016; 107: 267-73.
30. Jin G, Mao X, Qiao Z, Chen B, Jin F. RAP80 expression in breast cancer and its relationship with apoptosis in breast cancer cells. *OncoTargets and therapy*. 2019; 12: 625-34.
31. Tsukahara T, Haniu H, Matsuda Y. PTB-associated splicing factor (PSF) is a PPARgamma-binding protein and growth regulator of colon cancer cells. *PLoS One*. 2013; 8: e58749.