Gastroenterology Report, 9(5), 2021, 451–460

doi: 10.1093/gastro/goab023
Advance Access Publication Date: 25 August 2021
Original Article

ORIGINAL ARTICLE

An 11-gene signature for the prediction of systemic recurrences in colon adenocarcinoma

Jia-Wei Cai[1,2,†], Xiao-Ming Huang[3,†], Xiao-Lan Li[2,4,†], Si Qin[2,5], Yu-Ming Rong[6], Xi Chen[1,2], Jing-Rong Weng[1,2], Yi-Feng Zou[1,2] and Xu-Tao Lin[2,7,8,*]

[1]Department of Colorectal Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China, [2]Guangdong Institute of Gastroenterology, Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China, [3]Department of Hepatobiliary Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China, [4]Department of Reproductive Medicine, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China, [5]Department of Medical Ultrasonics, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China, [6]Department of VIP Region, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, P. R. China, [7]Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China; [8]Department of Colorectal Surgery, The Sixth Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, P. R. China

[†]These authors contributed equally to this study.
*Corresponding author. Department of Gastrointestinal Endoscopy, The Sixth Affiliated Hospital of Sun Yat-sen University, 26 Yuancun Erheng Road, Guangzhou, Guangdong 510655, P. R. China. Tel: +86-020-38254009; Fax: +86-020-38254166;
Email: linxutao@mail2.sysu.edu.cn

Abstract

Background Prognosis varies among patients within the same colon adenocarcinoma (COAD) stage, indicating the need for reliable molecular markers to enable individualized treatment. This study aimed to investigate gene signatures that can be used for better prognostic prediction of COAD.

Methods Gene-expression profiles of COAD patients were obtained from the Gene Expression Omnibus database ($n = 332$) and The Cancer Genome Atlas database ($n = 431$). The relationship between gene signature and relapse-free survival was analysed in the training set ($n = 93$) and validated in the internal validation set ($n = 94$) and external validation sets ($n = 145$ and 431).

Results Overall, 11 genes (N-myc downstream regulated gene 1 [*NDRG1*], fms-like tyrosine kinase 1 [*FLT1*], lipopolysaccharide binding protein [*LBP*], fatty acid binding protein 4 [*FABP4*], adiponectin gene [*ADIPOQ*], angiotensinogen gene [*AGT*], activin A receptor, type II-like kinase 1 [*ACVRL1*], CC chemokine ligand 11 [*CCL11*], cell division cycle 42 [*CDC42*], T-cell receptor alpha variable 9_2 [*TRAV9_2*], and proopiomelanocortin [*POMC*]) were identified by univariable and least absolute shrinkage

Submitted: 22 July 2020; **Revised:** 16 May 2021; **Accepted:** 16 May 2021

and selection operator (LASSO) Cox regression analyses. Based on the risk-score model, the patients were grouped into the high-risk or low-risk groups using the median risk score as the cut-off. The area under the curve (AUC) values for 1-, 3-, and 5-year recurrence were 0.970, 0.849, and 0.859, respectively. Patients in the high-risk group had significantly poorer relapse-free survival than did those in the low-risk group. The predictive accuracy of the 11-gene signature was proven in the validation sets. Our gene signature showed better predictive performance for 1-, 3-, and 5-year recurrence than did the other four models.

Conclusions The 11-gene signature showed good performance in predicting recurrence in COAD. The accuracy of the signature for prognostic classification requires further confirmation.

Key words: colon cancer; gene signature; recurrence; prognosis

Introduction

Colorectal cancer (CRC), as the main type of gastrointestinal tumor, is the third most commonly diagnosed carcinoma worldwide [1, 2]. Radical resection is the primary treatment modality for CRC. However, the 5-year survival rates of the patients after resection are only ~50%. Further, 40% of CRC patients experience recurrence, which is the main cause of cancer-related death [3]. Thus, accurate prediction of the recurrence risk is crucial for making effective personalized therapy decisions and improving prognosis.

Tumor recurrence cannot be accurately predicted only with clinicopathological factors. The prognoses of patients with the same stage of colon adenocarcinoma (COAD) and similar medical conditions are still diverse. There is an urgent need for reliable molecular markers to identify tumor characteristics and predict tumor recurrence. Gene signatures have been recently used for risk stratification in patients with cancer [4]. In CRC, the microsatellite status and *RAS/BRAF* gene mutations have been found to have prognostic and predictive values [5, 6]. Kim *et al.* [7] investigated a novel prognostic predictor based on an 11-gene signature for identifying high-risk CRC and predicting patients who will have the worst response to adjuvant chemotherapy. Dai *et al.* [8] also established a robust mRNA signature that can effectively assess the risk of early relapse in COAD patients.

Given the apparent clinical utility of gene signatures, an increasing number of studies have focused on the relationship between tumor cells and host immune cells. The immune-system components play important roles during tumorigenesis and cancer progression. Some studies have suggested that immune-response evasion is a unique feature of tumor cells [9]. Accordingly, immune-related targets have significant potential in the treatment and prognostic prediction of CRC.

This study aimed to investigate and validate individual prognostic features for predicting COAD recurrence based on immune-related genes. Towards this goal, we used gene-expression data and clinical information from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). We combined immune features with clinical factors to establish a composite prognostic index to improve the prediction of relapse in patients with COAD.

Materials and methods

Microarray data

Microarray-expression data and clinical information of the patients with COAD were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) on 10 September 2019. Data sets were included if they met the following criteria: (i) the sample size was >120 and (ii) complete clinical information on stage, relapse-free survival (RFS) interval, and RFS status was available. Finally, we acquired GSE17538 (238 samples) and GSE17536 (177 samples) for further processing.

For external validation, RNA-seq data of COAD patients were obtained from the TCGA RNA-seq database (https://cancergenome.nih.gov/) on 5 September 2019. Based on the same selection criteria above, we acquired TCGA-COAD (454 samples) for further processing.

Immune-related gene sets were obtained from the ImmPort database (https://immport.niaid.nih.gov) on 15 September 2019 including 1,811 genes after duplicates were removed.

Pre-processing of data sets

Genes whose fragments per kilobase million were <1 in more than half of the cases were excluded.

After pre-processing (Figure 1), GSE17538-processed included 187 samples comprising 1,376 genes. To avoid bias, all samples in GSE17538-processed were randomly divided 100 times in a 1:1 ratio into a training set (GSE17538-train) and an internal validation set (GSE17538-test). The final training set for GSE17538-train included 93 samples and the final validation set for GSE17538-test included 94 samples. Meanwhile, GSE17536 as the independent internal validation set included 145 samples. TCGA-COAD as the independent external validation set included 431 samples. The clinical features for preprocessed data sets are shown in Table 1.

Development of risk score and statistical analysis

Univariate Cox proportional hazards model was used to analyse the relationship between the immune-related genes and RFS. Among the 1,376 genes, 73 genes were significantly related to RFS ($P < 0.05$). Next, the LASSO Cox regression model [10] was used to identify 11 relapse-related genes (Figure 2A and B). We developed the risk-score formula based on these 11 genes (N-myc downstream regulated gene 1 [*NDRG1*], fms-like tyrosine kinase 1 [*FLT1*], lipopolysaccharide binding protein [*LBP*], fatty acid binding protein 4 [*FABP4*], adiponectin gene [*ADIPOQ*], angiotensinogen gene [*AGT*], activin A receptor, type II-like kinase 1 [*ACVRL1*], CC chemokine ligand 11 [*CCL11*], cell division cycle 42 [*CDC42*], T-cell receptor alpha variable 9_2 [*TRAV9_2*], and proopiomelanocortin [*POMC*]) using a multivariable Cox regression model (Table 2). The patients were then divided into the high-risk and low-risk groups using the median risk score in training set as the cut-off point. Differences in RFS between the low-risk and high-risk groups in each validation set were assessed using the Kaplan–Meier estimate. A cumulative/

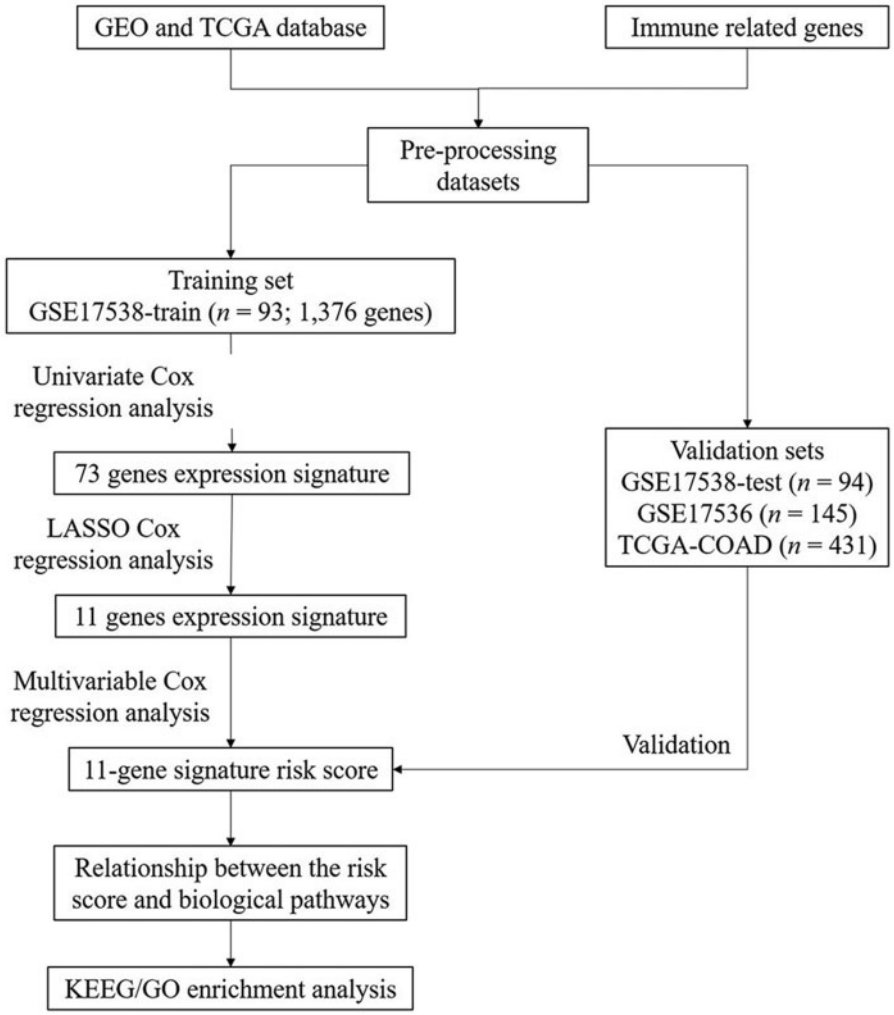

Figure 1. Flow chart of the development process of the 11-gene signature for the prediction of systemic recurrences in colon cancer. GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; COAD, colon adenocarcinoma; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

dynamic receiver-operating characteristic (ROC) $^{C/D}$ (t) analysis was then used to investigate the prognostic value for predicting RFS. All statistical analyses were performed using the R program (version 3.12.0, www.r-project.org). *P*-values of <0.05 were considered statistically significant.

Results

Risk-score formula in the training set

The risk-score formula was established based on the 11 genes as follows: risk score = expression level of *POMC* × (–5.752) + expression level of *AGT* × (–1.497) + expression level of *LBP* × (4.162) + expression level of *CCL11* × (–0.929) + expression level of *ACVRL1* × (–2.563) + expression level of *TRAV9_2* × (–3.124) + expression level of *CDC42* × (–2.958) + expression level of *NDRG1* × (0.828) + expression level of *FABP4* × (0.710) + expression level of *ADIPOQ* × (–0.356) + expression level of *FLT1* × (0.118).

In the high-risk group, the RNA expression level of *NDRG1*, *FLT1*, *LBP*, *FABP4*, and *ADIPOQ* was relatively high, whereas the RNA expression level of *AGT*, *ACVRL1*, *CCL11*, *CDC42*, *TRAV9_2*, and *POMC* was relatively low (Figure 2C).

Relationship between the 11-gene signature risk score and RFS in the data sets

In GSE17538-train, the area under the curve (AUC) values for predicting 1-, 3-, and 5-year recurrence were 0.970, 0.849, and 0.859, respectively (Figure 3A). In GSE17538-test, the AUCs for predicting 1-, 3-, and 5-year recurrence were 0.792, 0.823, and 0.786, respectively (Figure 3B). In both GSE17538-train and GSE17538-test sets, patients in the high-risk group had a significantly lower RFS rate than did those in the low-risk group ($P < 0.001$, Figure 3F and $P = 0.019$, Figure 3G). In GSE17538-processed, the AUCs were 0.879, 0.849, and 0.811, respectively (Figure 3C). In GSE17536, the AUCs were 0.880, 0.788, and 0.816, respectively (Figure 3D). In the TCGA-COAD set, the AUCs were 0.541, 0.672, and 0.704, respectively (Figure 3E). Patients in the high-risk group had a significantly lower RFS rate than did those in the low-risk group in each data set (GSE17538-processed, $P < 0.001$, Figure 3H; GSE17536, $P < 0.001$, Figure 3I; TCGA-COAD, $P = 0.002$, Figure 3J).

Gene-set enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were further used for enrichment analysis of

Table 1. Clinical information of samples from processed data sets

Clinical features	GSE17538-processed	GSE17538-train	GSE17538-test	GSE17536	TCGA-COAD
Recurrence					
Yes	42	23	19	36	87
No	145	70	75	109	344
TNM stage					
I	28	14	14	24	73
II	70	33	37	55	165
III	75	38	37	56	123
IV	14	8	6	10	59
Unknown					11
Sex					
Male	99	47	52	76	232
Female	88	46	42	69	199
Age (years)					
<65	85	45	40	62	167
>65	102	48	54	83	264
T category					
T1					11
T2					75
T3					296
T4					48
TX					1
N category					
N0					254
N1					100
N2					77
M category					
M0					320
M1					59
MX					52
Radiation therapy					
Yes					0
No					33
Unknown					398
Lymphatic invasion					
Yes					150
No					239
Unknown					42
Microsatellite instability					
Yes					11
No					81
Unknown					339
Venous invasion					
Yes					89
No					285
Unknown					57

11 genes (R program cluster profiler, v3.12.0). A total of 1,177 annotations were found in the GO database, including 1,087 for biological processes, 35 for cellular components, and 55 for molecular function. In addition, 58 pathways were enriched from the KEGG database. We identified 10 GO terms and 10 KEGG pathways, which were mainly related to tumor progression (**Figure 4A and B**).

Relationship between the 11-gene signature risk score and biological pathways

The Gene set variation analysis function of the R program was used to calculate the KEGG functional enrichment scores of 187 samples from GSE17538-processed. The correlation between the enrichment score and the risk score was evaluated. We obtained 29 KEGG pathways; of these, 8 and 21 pathways were negatively and positively correlated with the risk score, respectively. Clustering analysis was conducted on the enrichment scores of the most correlated 14 pathways in the GSE17538-train set (**Figure 5**). The highest negative correlation coefficient was –0.545 and the highest positive correlation coefficient was 0.340. The top four KEGG pathways with the strongest correlation with the risk score are shown in **Supplementary Figure S1**.

Comparison of predictive capability between the 11-gene signature risk score and clinical features

To estimate the independent prognostic value of the 11-gene signature risk score, we applied Cox regression analyses of RFS

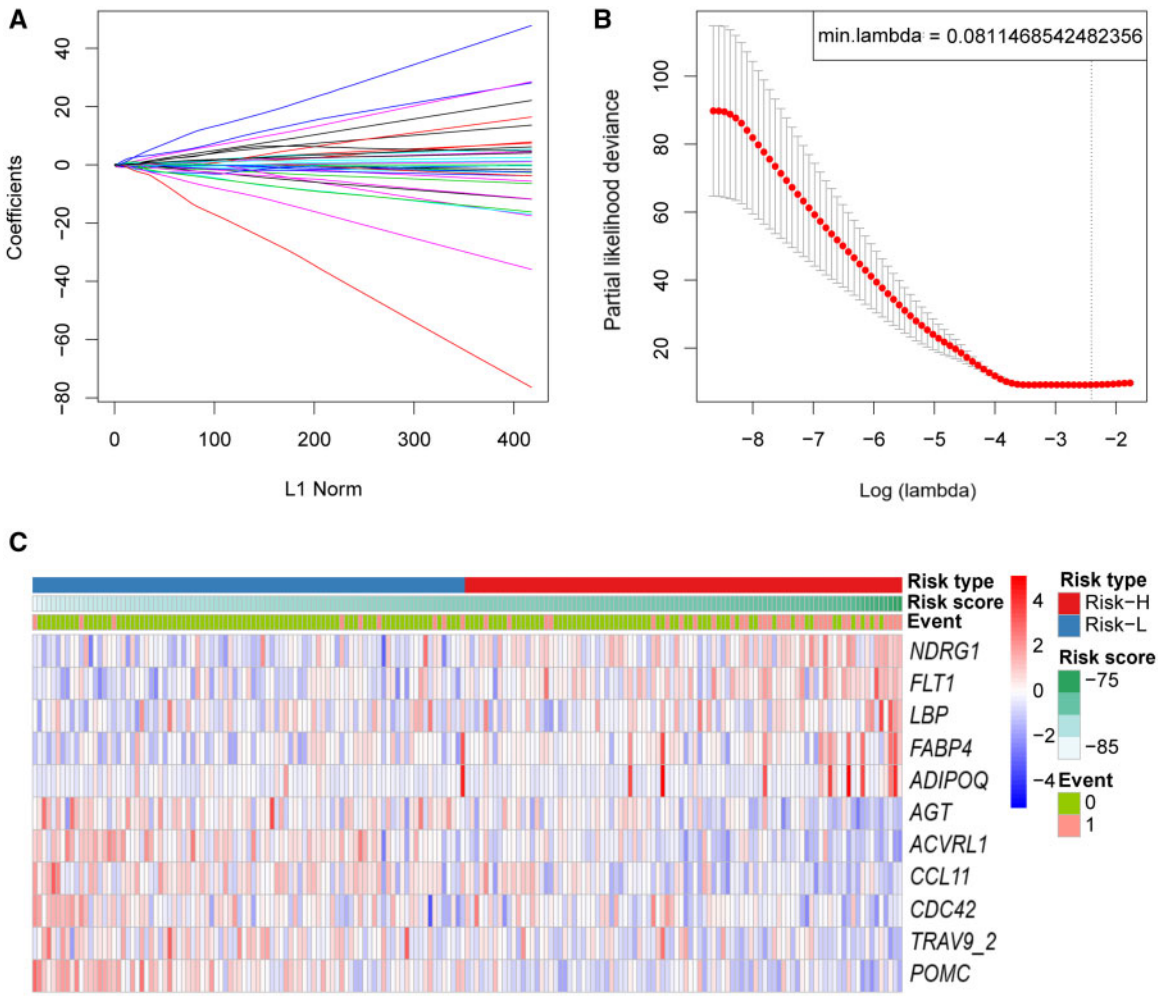

Figure 2. LASSO Cox regression of 11-genes signature. (A) The changing trajectory of each independent variable (the horizontal and vertical axes represent the log value and the coefficient, respectively). (B) The confidence interval for each lambda. (C) RNA expression of 11 genes in the training set. *NDRG1*, N-myc downstream regulated gene 1; *FLT1*, fms-like tyrosine kinase 1; *LBP*, lipopolysaccharide binding protein; *FABP4*, fatty acid binding protein 4; *ADIPOQ*, adiponectin gene; *AGT*, angiotensinogen gene; *ACVRL1*, activin A receptor, type II-like kinase 1; *CCL11*, CC chemokine ligand 11; *CDC42*, cell division cycle 42; *TRAV9_2*, T-cell receptor alpha variable 9_2; *POMC*, proopiomelanocortin.

Table 2. Multivariable Cox regression analysis of 11 genes in colon adenocarcinoma

Variable	HR (95% CI)	*P*-value
ACVRL1	0.077 (4.11e-03–1.444)	0.086
FABP4	2.034 (0.353–11.707)	0.427
CCL11	0.395 (0.143–1.089)	0.073
NDRG1	2.289 (0.675–7.765)	0.184
FLT1	1.125 (0.055–22.877)	0.939
CDC42	0.052 (7.71e-04–3.493)	0.168
ADIPOQ	0.700 (0.246–1.998)	0.505
AGT	0.224 (0.067–0.745)	0.015
TRAV9_2	0.044 (5.24e-04–3.689)	0.167
LBP	64.206 (1.600–2,581.374)	0.027
POMC	0.003 (6.70e-05–0.151)	0.003

HR, hazard ratio; 95% CI, 95% confidence interval; *ACVRL1*, activin A receptor, type II-like kinase 1; *FABP4*, fatty acid binding protein 4; *CCL11*, CC chemokine ligand 11; *NDRG1*, N-myc downstream regulated gene 1; *FLT1*, fms-like tyrosine kinase 1; *CDC42*, cell division cycle 42; *ADIPOQ*, adiponectin gene; *AGT*, angiotensinogen gene; *TRAV9_2*, T-cell receptor alpha variable 9_2; *LBP*, lipopolysaccharide binding protein; *POMC*, proopiomelanocortin.

in patients with COAD from the GSE17538-train set (Table 3). In the univariate analysis, the significant prognostic indicators of COAD recurrence were TNM stage and risk score. In the multivariate analysis, both the TNM stage (hazard ratio [HR] = 2.083, 95% confidence interval [95% CI] = 1.394–3.112, $P < 0.001$) and risk score (HR = 1.480, 95% CI = 1.300–1.690, $P < 0.001$) were found to be independent predictors of recurrence. The forest map is shown in Figure 6B.

We created a nomogram model based on the risk score and clinical features (TNM stage and sex; Figure 6A). Among all the factors in this study, the risk score had the greatest effect on RFS, indicating that the 11-gene signature is a potential predictor of COAD recurrence.

Comparison of predictive capability between the 11-gene signature risk score and other gene signature risk scores

We selected four gene signature-based prognostic risk models to compare with our 11-gene model: 15-gene signature (Xu *et al.* [11]), 15-gene signature (Dai *et al.* [8]), 12-gene signature

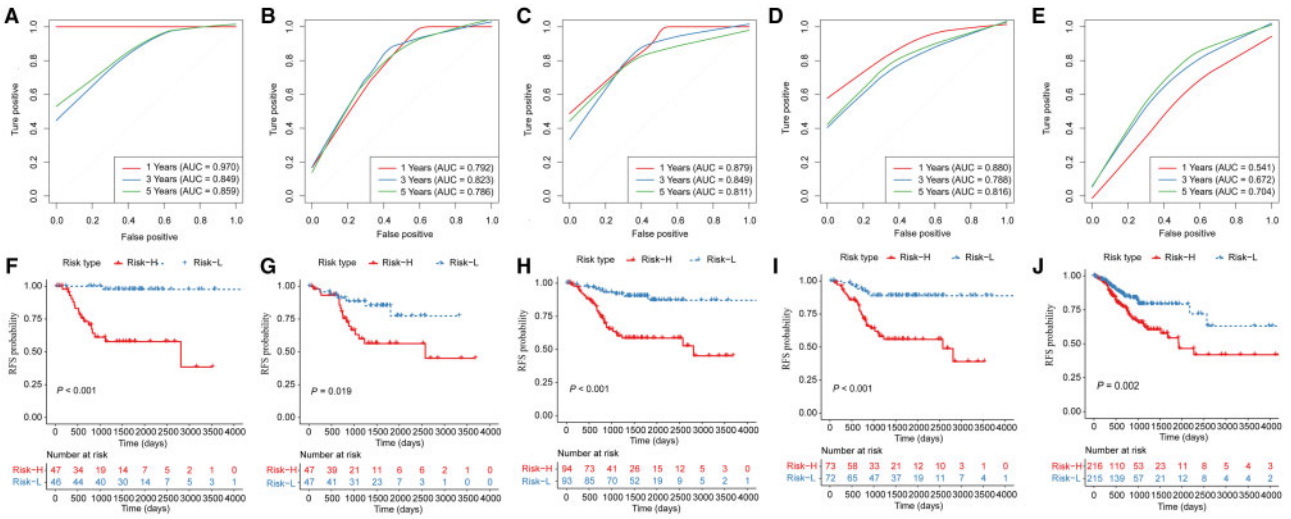

Figure 3. Receiver-operating characteristic and Kaplan–Meier RFS curve of each data set. (A) Receiver-operating characteristic (ROC) of the 11-gene signature in the GSE17538-train set. (B) ROC of the 11-gene signature in the GSE17538-test set. (C) ROC of the 11-gene signature in the GSE17538-processed set. (D) ROC of the 11-gene signature in the GSE17536 set. (E) ROC of the 11-gene signature in the TCGA-COAD set. (F) Kaplan–Meier RFS curve of the GSE17538-train set. (G) Kaplan–Meier RFS curve of the GSE17538-test set. (H) Kaplan–Meier RFS curve of GSE17538-processed set. (I) Kaplan–Meier RFS curve of the GSE17536 set. (J) Kaplan–Meier RFS curve of the TCGA-COAD set. AUC, area under the curve; RFS, relapse-free survival.

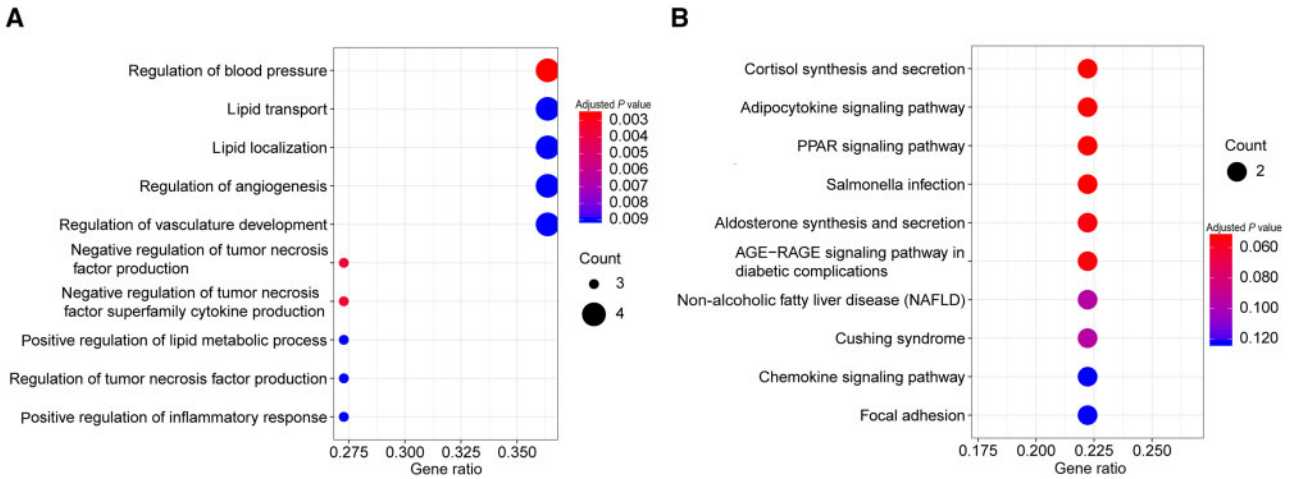

Figure 4. Enrichment of the 11 genes. (A) GO enrichment. (B) KEGG enrichment. PPAR, peroxisome proliferators-activated receptors; AGE, advanced glycation end products; RAGE, receptor for advanced glycation end products; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

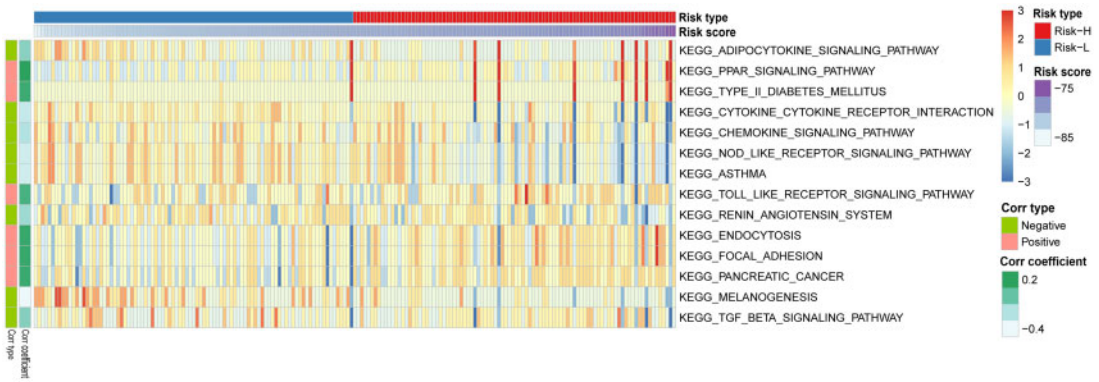

Figure 5. Relationship between the risk score and KEGG pathways in the GSE17538-train set. Corr, correlation; KEGG, Kyoto Encyclopedia of Genes and Genomes.

(Sun *et al*. [12]), and 9-gene signature (Mo *et al*. [13]). We used the same method according to the corresponding gene models to calculate the risk scores of each COAD sample in the GSE7538-processed data set. According to each median risk score, the samples were divided into the high-risk and low-risk groups. We then compared the overall survival (OS) rate between the two groups and evaluated the predictive capability among the risk scores for 1-, 3-, and 5-year OS. In Xu's model, the AUCs were 0.791, 0.708, and 0.727, respectively (Figure 7A). In Dai's model, the AUCs were 0.662, 0.761, and 0.762, respectively (Figure 7B). In Sun's model, the AUCs were 0.578, 0.681, and 0.668, respectively (Figure 7C). In Mo's model, the AUCs were 0.708, 0.610, and 0.582, respectively (Figure 7D). In Xu's, Dai's, and Sun's models, the patients in the high-risk group had significantly lower OS rates than did those in the low-risk group. Meanwhile, there was no significant difference in OS rate between these two groups in Mo's model.

Table 3. Univariate and multivariate Cox regression analysis of prognostic factors in colon adenocarcinoma

Variable	Univariable analysis		Multivariable analysis	
	HR (95% CI)	*P*-value	HR (95% CI)	*P*-value
TNM stage	2.221 (1.501–3.287)	<0.001	2.083 (1.394–3.112)	<0.001
Sex	0.982 (0.536–1.799)	0.952	0.944 (0.509–1.751)	0.855
Risk score	1.491 (1.313–1.693)	<0.001	1.480 (1.300–1.690)	<0.001

HR, hazard ratio; 95% CI, 95% confidence interval.

The predictive performance of our 11-gene signature was compared with that of the other four signatures according to the concordance index (C-index). Among these five models, the 11-gene signature showed the highest C-index (Figure 8A), indicating that our risk-score model has better predictive performance than do the other four models. Furthermore, we estimated the predictive performance of the five models at a certain time point by restricted mean survival (RMS) time. These five RMS curves had intersections at 105 months (Figure 8B). When the follow-up time was <105 months, the 11-gene risk model showed better predictive performance, indicating that our risk-score model is appropriate for prediction within a 5-year period.

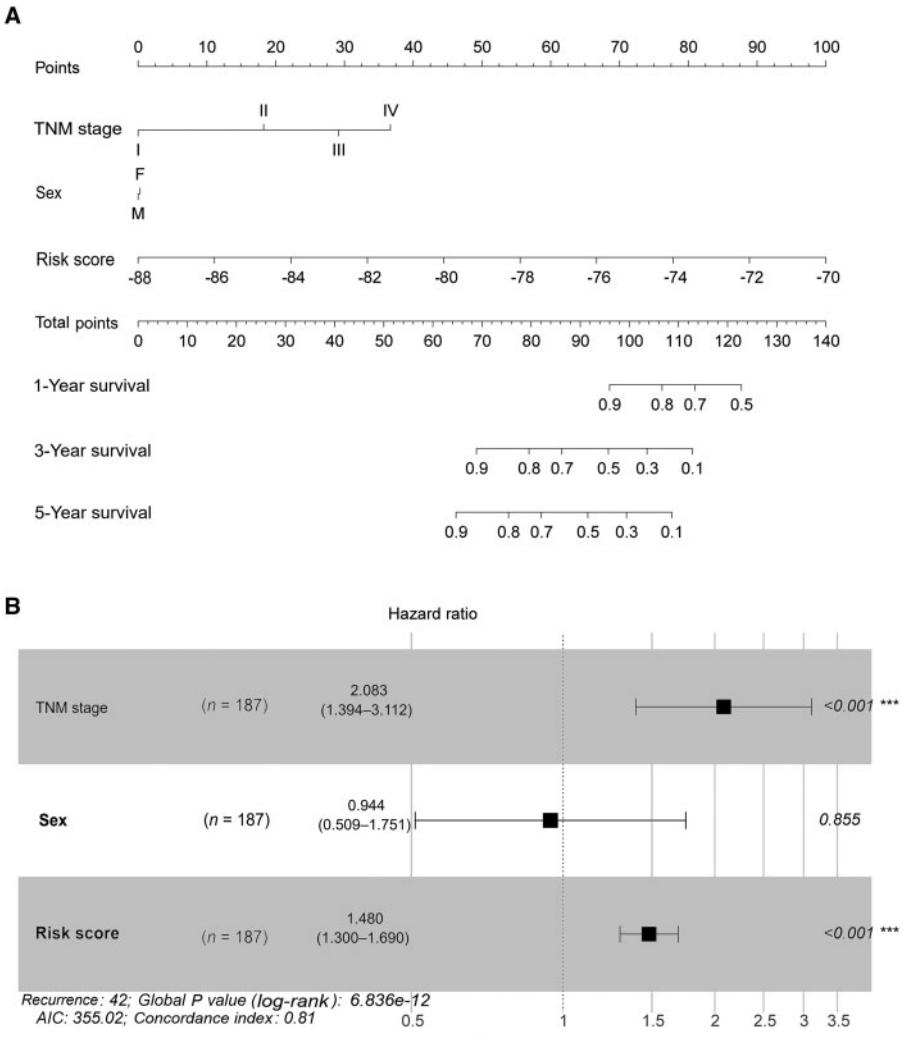

Figure 6. Relationship between risk score and clinical features. (A) Nomogram model based on risk score and clinical features. (B) Forest map based on risk score and clinical features. F, female; M, male; AIC, Akaike information criterion.

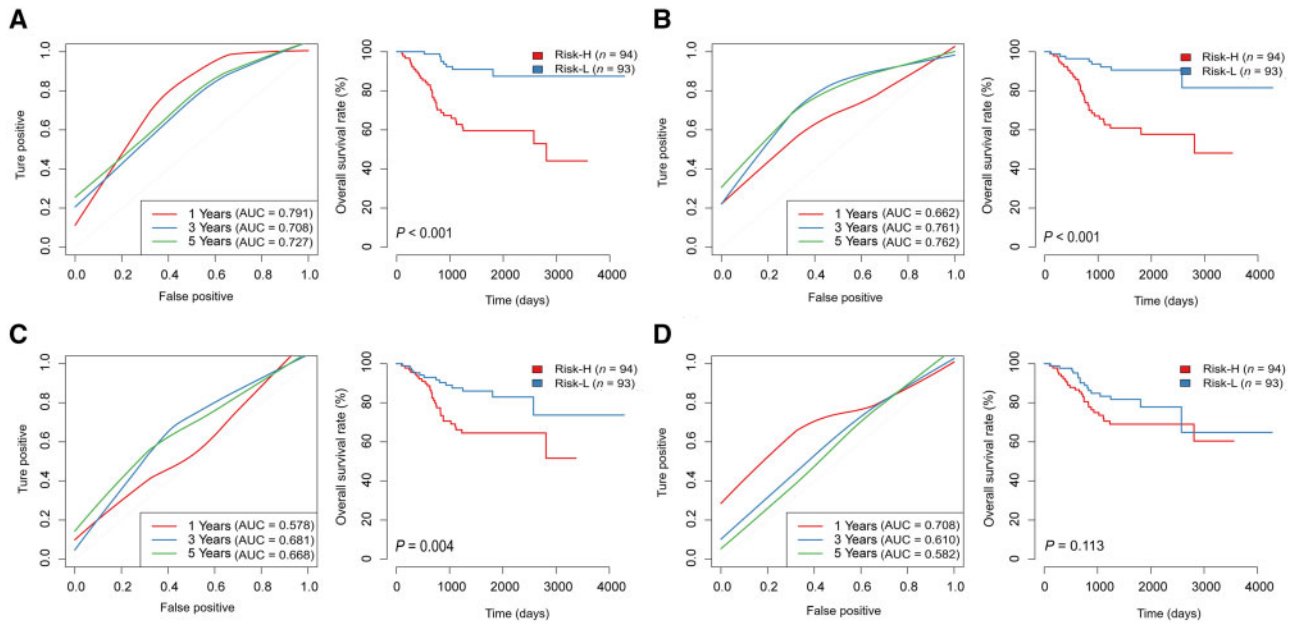

Figure 7. Receiver-operating characteristic and Kaplan–Meier curve of the gene signatures. (A) Xu's model (15 genes); (B) Dai's model (15 genes); (C) Sun's model (12 genes); (D) Mo's model (9 genes). AUC, area under the curve.

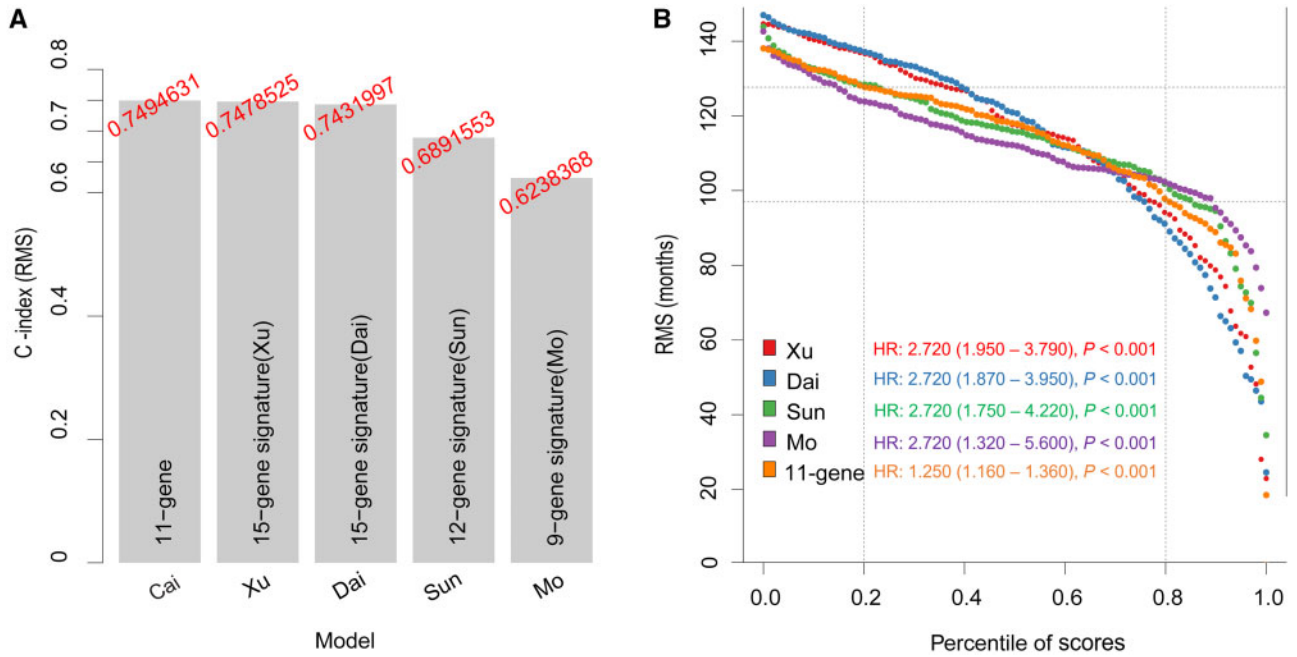

Figure 8. Comparison among five gene signatures. (A) Concordance indexes of five gene signatures. (B) Restricted mean survival curves of five gene signatures. C-index, concordance index; RMS, restricted mean survival; HR, hazard ratio.

Discussion

After radical resection for COAD, most patients still suffer a relapse or die [14, 15]. Further, the prognosis varies among patients in the same TNM stage owing to the difference in molecular subtypes [16]. Thus, accurate biomarkers predictions of treatment response and survival are needed. Several studies explored reliable biomarkers for better prediction of COAD prognosis and achievement of individualized treatment [17–22]. However, few reliable biomarkers have been identified.

We established a risk score based on an 11-gene signature to improve the prognostic prediction for COAD after surgical resection. The risk-score formula was established based on the GSE17538-train set and the cut-off was set as the median risk score. This cut-off was then used to divide the patients into the high-risk and low-risk groups. The predictive accuracy of the 11-gene signature was validated by the GSE17538-test, GSE17536, and TCGA-COAD data sets.

Through gene-set enrichment analysis, we found 29 KEGG pathways related to the 11-gene signature; of these, 8 and 21

pathways were negatively and positively correlated with the risk score, respectively. Many reports verified some of our findings about the relations between these pathways and gastrointestinal cancer. Yang *et al.* [23] found that nucleotide-binding and oligomerization domain-containing receptors (NOD-like receptors) can inhibit the proliferation of CRC cells. Chen *et al.* [24] also suggested that the NOD1 receptor plays an indispensable role in inflammatory bowel disease-related colon cancer. Amiri *et al.* [25] showed an association between high expression of NOD2 receptor and gastric-cancer progression. Focal adhesion kinase is a cytoplasmic protein tyrosine kinase that was recently found to be related to cell-cycle regulation, proliferation, apoptosis, migration, invasion, metastasis, and angiogenesis. Focal adhesion kinase overexpression has been reported in poorly differentiated colon and breast cancer tissues [26,27].

Our research had some limitations. First, it was based on publicly available data sets without verification in a clinical trial. Second, the cut-off of the risk score should be further optimized instead of simply using the median risk score. Finally, bioinformatics of the identified immune-related genes was unclear, requiring further investigations. Future validation studies should be conducted for practical clinical applications.

In summary, we identified an 11-immune-related-gene signature to improve the prediction of relapse in COAD patients. Patients in the high-risk group had significantly poorer RFS than did those in the low-risk group. The 11-gene signature had good performance for predicting the 1-, 3-, and 5-year recurrence in COAD.

Supplementary Data

Supplementary data is available at *Gastroenterology Report* online.

Authors' Contributions

J.W.C., X.M.H., X.L.L., and X.T.L. contributed to the study design; acquisition, analysis, and interpretation of the data; and drafting of the manuscript. S.Q., Y.M.R., X.C., Y.F.Z., and J.R.W. contributed to the majority of the data analysis. Y.F.Z. and X.T.L. supervised the study. All authors read and approved the final manuscript.

Funding

This work was supported by National Key Clinical Discipline, the Fundamental Research Funds for the young teacher training program of Sun Yat-sen University [grant number 18ykpy02], the "5010 Clinical Research Program" of Sun Yat-sen University [grant number 2010012], the Natural Science Foundation of Guangdong Province, China [grant number 2020A1515010428], and the Medical Science Research Grant from the Health Department of Guangdong Province [grant number A2018007].

Acknowledgements

The authors are grateful for the contribution of the team of the Guangdong Provincial Key Laboratory of Colorectal and Pelvic Floor Diseases.

Conflict of Interest

The authors declare that there is no conflict of interests in this study.

References

[1]. Siegel R, Miller K, Jemal A. Cancer statistics 2016. *CA Cancer J Clin* 2016;**66**:1.
[2]. Xie Y, Shi L, He X *et al.* Gastrointestinal cancers in China, the USA, and Europe. *Gastroenterol Rep (Oxf)* 2021; **9**:91–104.
[3]. Renouf DJ, Woods R, Speers C *et al.* Improvements in 5-year outcomes of stage II/III rectal cancer relative to colon cancer. *Am J Clin Oncol* 2013;**36**:558–64.
[4]. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med* 2011;**364**: 340–50.
[5]. Calistri D, Rengucci C, Seymour I *et al.* Mutation analysis of p53, K-ras, and BRAF genes in colorectal cancer progression. *J Cell Physiol* 2005;**204**:484–8.
[6]. Li ZN, Zhao L, Yu LF *et al.* BRAF and KRAS mutations in metastatic colorectal cancer: future perspectives for personalized therapy. *Gastroenterol Rep (Oxf)* 2020;**8**:192–205.
[7]. Kim SK, Kim SY, Kim CW *et al.* A prognostic index based on an eleven gene signature to predict systemic recurrences in colorectal cancer. *Exp Mol Med* 2019;**51**:1–12.
[8]. Dai W, Li Y, Mo S *et al.* A robust gene signature for the prediction of early relapse in stage I-III colon cancer. *Mol Oncol* 2018;**12**:463–75.
[9]. Gabrilovich D, Pisarev V. Tumor escape from immune response: mechanisms and targets of activity. *Curr Drug Targets* 2003;**4**:525–36.
[10]. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc B* 1996;**58**:267–88.
[11]. Xu G, Zhang M, Zhu H *et al.* A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* 2017;**604**:33–40.
[12]. Sun D, Jing C, Liu L *et al.* Establishment of a 12-gene expression signature to predict colon cancer prognosis. *PeerJ* 2018; **6**:e4942.
[13]. Mo S, Dai W, Xiang W *et al.* Prognostic and predictive value of an autophagy-related signature for early relapse in stages I-III colon cancer. *Carcinogenesis* 2019;**40**:861–70.
[14]. Steinert R, Hantschick M, Vieth M *et al.* Influence of subclinical tumor spreading on survival after curative surgery for colorectal cancer. *Arch Surg* 2008;**143**:122–8.
[15]. Zhang XW, Yang HY, Fan P *et al.* Detection of micrometastasis in peripheral blood by multi-sampling in patients with colorectal cancer. *World J Gastroenterol* 2005;**11**: 436–8.
[16]. Bathe OF, Farshidfar F. From genotype to functional phenotype: unraveling the metabolomic features of colorectal cancer. *Genes (Basel)* 2014;**5**:536–60.
[17]. Marisa L, de Reyniès A, Duval A *et al.* Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013;**10**: e1001453.
[18]. Sadanandam A, Lyssiotis CA, Homicsko K *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* 2013;**19**: 619–25.
[19]. De Sousa E Melo F, Wang X, Jansen M *et al.* Poor-prognosis colon cancer is defined by a molecularly distinct subtype

and develops from serrated precursor lesions. *Nat Med* 2013; **19**:614–8.
[20]. Tsai HL, Chu KS, Huang YH *et al*. Predictive factors of early relapse in UICC stage I-III colorectal cancer patients after curative resection. *J Surg Oncol* 2009;**100**: 736–43.
[21]. Cho WC. MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int J Biochem Cell Biol* 2010;**42**:1273–81.
[22]. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;**487**:330–7.
[23]. Yang LZ, Tang Z, Zhang HQ *et al*. PSMA7 directly interacts with NOD1 and regulates its function. *Cell Physiol Biochem* 2013;**31**:952–9.
[24]. Chen GY, Shaw MH, Redondo G *et al*. The innate immune receptor Nod1 protects the intestine from inflammation-induced tumorigenesis. *Cancer Res* 2008;**68**:10060–7.
[25]. Amiri RM, Tehrani M, Taghi-Zadeh S *et al*. Association of nucleotide-binding oligomerization domain receptors with peptic ulcer and gastric cancer. *Iran J Allergy Asthma Immunol* 2016;**15**:355–62.
[26]. Planas-Silva MD, Bruggeman RD, Grenko RT *et al*. Role of c-Src and focal adhesion kinase in progression and metastasis of estrogen receptor-positive breast cancer. *Biochem Biophys Res Commun* 2006;**341**:73–81.
[27]. Brunton VG, Ozanne BW, Paraskeva C et al. A role for epidermal growth factor receptor, c-Src and focal adhesion kinase in an in vitro model for the progression of colon cancer. Oncogene 1997;**14**:283–93.