Original Article

Development of a prognostic index of colon adenocarcinoma based on immunogenomic landscape analysis

Weijie Qiang[1,2], Yifei Dai[3], Guibo Sun[1,2], Xiaoyan Xing[1,2], Xiaobo Sun[1,2]

[1]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China; [2]Key Laboratory of New Drug Discovery Based on Classic Chinese Medicine Prescription, Chinese Academy of Medical Sciences, Beijing 100193, China; [3]School of Medicine, Tsinghua University, Beijing 100084, China

Contributions: (I) Conception and design: W Qiang, Y Dai, X Xing; (II) Administrative support: G Sun, X Sun; (III) Provision of study materials or patients: Y Dai; (IV) Collection and assembly of data: Y Dai, W Qiang; (V) Data analysis and interpretation: W Qiang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Xiaobo Sun; Xiaoyan Xing. Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China. Email: sun-xiaobo@163.com; lilyclxy@163.com.

Background: Colon adenocarcinoma (COAD) is one of the most commonly diagnosed cancers, and it is closely related to the immune microenvironment. Considering that immunotherapy is not effective for all COAD patients, it is necessary to identify the effective population before administering treatment. In this study, we established an independent prognostic index based on immune-related genes (IRGs), in order to evaluate the clinical outcome of COAD.

Methods: The gene expression profiles and IRGs taken from The Cancer Genome Atlas (TCGA) and Immunology Database and Analysis Portal (ImmPort), respectively, were integrated in order to identify the differentially expressed IRGs. Functional enrichment analysis was conducted and the prognostic value of survival-related IRGs was determined. Based on Cox regression analysis, the IRG-based prognostic index (IRGPI) was established, and the model was evaluated and applied.

Results: A total of 51 differentially expressed survival-related IRGs were identified. The most significant signaling pathway was "cytokine-cytokine receptor interaction". The index established herein was based on 12 survival-related IRGs, and it was highly accurate in monitoring prognosis. Moreover, the IRGPI was significantly correlated with multiple clinicopathologic factors, as well as with the infiltration of immune cells.

Conclusions: An independent IRGPI was established in order to assess the immune status and tumor prognosis in COAD patients. This index can serve as a robust biomarker in clinical prognosis applications, including cancer immunotherapy.

Keywords: Colon adenocarcinoma (COAD); immune-related genes (IRGs); bioinformatics; prognostic index

Introduction

According to the Global Cancer Report released at the end of 2018, colon adenocarcinoma (COAD) is one of the most commonly diagnosed cancers, and one of the top three malignant tumors in terms of morbidity and mortality (1). Considering that the early symptoms are not visible, most patients diagnosed with COAD are in the middle or late stages of the disease, which impedes treatment and timely surgical intervention. Therefore, despite the increasing number of targeted drugs (2,3), the 5-year average survival rate of COAD patients in China is only 57%. To improve this rate, the prognosis status of patients should be monitored closely, and the treatment plans should be

Page 2 of 16

Qiang et al. An immune-related prognostic signature of COAD

timely adjusted. This necessitates the development of an independent prognostic index that allows for the evaluation of COAD clinical outcomes.

Evidence of the close relation between the immune microenvironment and tumor development has substantially increased over the past few years (4-7). Consequently, tumor immunotherapy has emerged as a promising technique of treatment, and it has garnered the attention of many researchers (8,9). This technique relies on the activation of immune cells that can trigger anti-tumor immune responses, leading to the elimination of tumor residual lesions, the inhibition of tumor growth, and the breaking of immune tolerance (10). Unfortunately, immunotherapy is not effective for all tumor patients. Therefore, clinical research efforts are currently focused on the identification of patients who can be efficiently treated using immunotherapy techniques. This may be accomplished by determining the type and proportion of immune cells and factors in the tumor immune microenvironment, which are in turn controlled by gene expression. The recent progress in gene sequencing technology has allowed for the design of new and sensitive gene-based biomarkers (11,12) that may be used to improve the prognosis of COAD. The comprehensive analysis of the relevance between the immune-related genes (IRGs) and survival is conducive to elucidating the potential prognostic value of these genes. For example, an immuno-genomic study of tumor immune microenvironment has previously shown that mRNA-seq is effective in predicting clinical response (5).

In this study, we identified the differentially expressed IRGs in COAD patients by relating the gene expression profiles to the IRGs from different online databases. The functional enrichment analysis and prognostic value of the survival-related IRGs were also determined. The prognostic index was finally established based on Cox regression analysis, followed by model evaluation and application. In theory, the robust prognostic biomarker reported herein can be used to predict the clinical outcomes of therapeutic methods, including cancer immunotherapy.

Methods

Acquisition of gene expression data and IRGs

The RNA-seq data of COAD samples was taken from The Cancer Genome Atlas (TCGA) database (13) that is published in the Genomic Data Commons (GDC) portal (https://portal.gdc.cancer.gov/) (14), and the clinical information of these samples was downloaded and extracted by Perl. The collected dataset is comprised of 473 COAD samples and 41 adjacent normal samples. In addition, a comprehensive list of IRGs (1,811 in total) was downloaded from the Immunology Database and Analysis Portal (ImmPort) (https://www.immport.org/home) (15). These genes were divided into 17 categories based on their molecular functions (MFs) (e.g., cytokine, interleukin, and natural killer cell cytotoxicity).

Identification of differentially expressed genes (DEGs)

The DEGs were identified by comparing the COAD and adjacent normal samples using R package "limma" (16). Specifically, the genes exhibiting $|\log_2(\text{fold change})| >1$ and false discovery rate (FDR) <0.05 were considered as DEGs. Afterwards, the differentially expressed IRGs were extracted from the list of identified DEGs. Heat maps and volcano plots of DEGs and differentially expressed IRGs were generated by R package pheatmap and ggplot2, respectively (17,18).

Screening of survival-related IRGs and evaluation of their prognostic value

A survival analysis of all of the differentially expressed IRGs was performed using the R package "survival" (19), and the survival-related IRGs were determined by univariate Cox analysis (P<0.05). In order to explore the potential molecular mechanisms of differentially expressed IRGs and survival-related IRGs, functional enrichment analysis were conducted on the basis of Kyoto encyclopedia of genes and genomes (KEGG) and gene ontology (GO) database, including biological process (BP), MF and cellular component (CC) (20,21), using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) 6.8 (22). The GO terms and KEGG signaling pathways were enriched with an adjusted P value of <0.05. Moreover, the top 10 KEGG signaling pathways were drawn into a bubble chart using the R package "ggplot2" (18). To determine the prognostic value of survival-related IRGs, a forest plot of these genes was constructed using the hazard ratio (HR) as an indicator, based on univariate Cox analysis.

Development of the IRG-based prognostic index (IRGPI)

Survival-related IRGs were submitted for multivariate analysis to develop the IRGPI, and integrated IRGs was

kept as independent prognostic indicators. Specifically, the IRGPI was established by multiplying the expression values with the Cox regression coefficient (23). It should be noted that the survival-related IRGs included in the IRGPI are referred to as hub IRGs.

Assessment of IRGPI and genetic alteration analysis

Patients were divided into high- and low-risk groups, depending on whether their IRGPI values are greater or less than the median value, respectively. The corresponding Kaplan-Meier (K-M) survival curves were constructed to demonstrate the overall survival (OS) within different risk groups. The ROC curve was also established in order to assess the sensitivity and specificity of the model. Additionally, univariate and multivariate analyses of survival were carried out for both IRGPI and clinicopathologic factors. These analyses were performed using the R package "survival" (19). We also explored the correlation between the expression of hub IRGs and the clinicopathologic factors. Moreover, genetic alteration analysis of hub IRGs was conducted on cBioPortal (https://www.cbioportal.org/) (24).

Evaluation of immune cell infiltration in tumors

The Tumor Immune Estimation Resource (TIMER) was used to systematically analyze immune cell infiltration in 23 types of cancer listed in the TCGA (25). Statistical methods were adopted to estimate the abundances of six immune cell infiltrates, including B cells, CD4+ T cells, CD8+ T cells, neutrophils, macrophages, and dendritic cells. The obtained results were validated by pathological estimations. Subsequently, the immune cell infiltration levels in COAD samples were downloaded, and the correlation between IRGPI and immune cell infiltration was calculated based on Pearson correlation analysis.

Construction of the transcription factor (TF)-mediated regulatory network

TFs play an essential role in controlling gene expressions, including IRG, and thus, they can affect the immune function of the body. Consequently, it is necessary to identify the particular TFs that can potentially regulate hub IRGs. For this purpose, the list of TFs was first downloaded from the Cistrome Cancer database (http://cistrome.org/CistromeCancer/) (26). Then, the differentially expressed TFs were extracted from all DEGs. Subsequently, we carried out the correlation analysis of differentially expressed TFs and hub IRGs using the R software. The correlativity was considered reliable if |correlation value| >0.6 and P<0.05. As a result, we submitted them to build a TF-mediated regulatory network via Cytoscape software. To screen for the most important TF in this network, topological analysis was performed using the Network Analyzer plug-in (27).

Statistical analysis

To evaluate IRGPI effectiveness, the R package "survival" was used to calculate the area under curve (AUC) of the receiver operating characteristic (ROC) curve (28), and the R package "ggplot2" was used to generate boxplots (18). An independent t-test was conducted to test the differences among diverse clinical parameters. These differences were considered to be statistically significant if P<0.05.

Results

Acquisition of differentially expressed IRGs

In order to identify the DEGs, the genomic data of 473 COAD and 41 adjacent normal samples were compared using R package "limma". Based on the cut-off criteria of $|\log_2(\text{fold change})|$ >1 and FDR <0.05, a total of 6,477 DEGs were detected, including 4,561 up-regulated genes and 1,916 down-regulated genes (*Figure 1A,B*). With the list of IRGs, 467 differentially expressed IRGs were extracted from all DEGs, including 179 up-regulated IRGs and 288 down-regulated IRGs (*Figure 1C,D*).

Functional enrichment analysis of differentially expressed IRGs

To explore the biological functions of differentially expressed IRGs, functional enrichment analysis was conducted using the DAVID 6.8. The results illustrated in *Figure 2A* show that "immune response", "extracellular region", and "antigen binding" are the most relevant BP, CC, and MF of genes, respectively. Furthermore, the KEGG enrichment analysis presented in *Figure 2B* indicates that the differentially expressed IRGs are mostly correlated with "cytokine-cytokine receptor interaction".

Functional enrichment analysis of survival-related IRGs

As key prognostic indicators in applications of clinical

Page 4 of 16

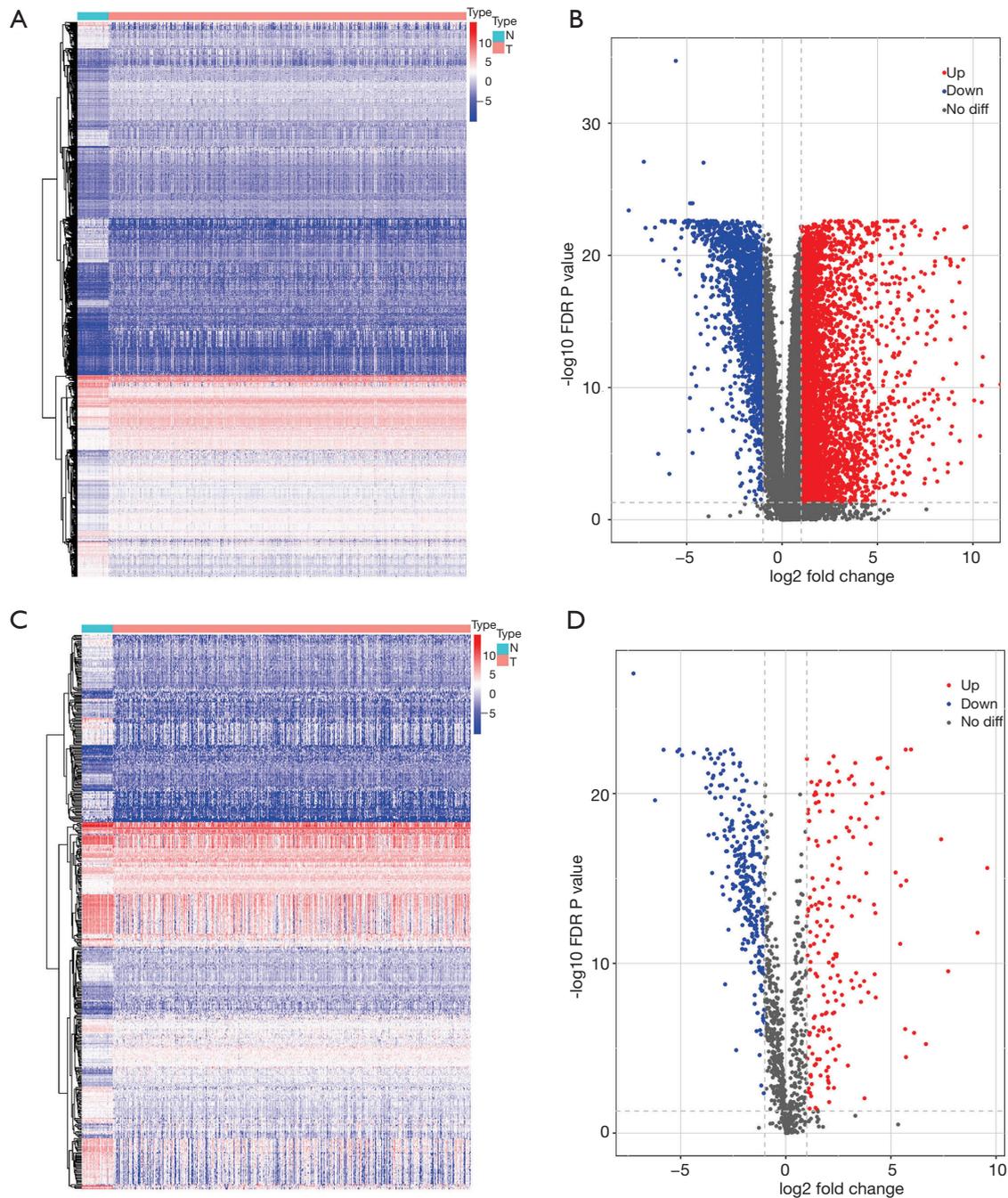Qiang et al. An immune-related prognostic signature of COAD

Figure 1 Differentially expressed IRGs. The heatmap (A) and volcano plot (B) of genes that are differentially expressed in COAD samples, compared to adjacent normal samples; red and blue dots represent up-regulated and down-regulated DEGs, respectively; the heatmap (C) and volcano plot (D) of differentially expressed IRGs. Red and blue dots represent up-regulated and down-regulated differentially expressed IRGs, respectively. IRGs, immune-related genes; COAD, colon adenocarcinoma; DEGs, differentially expressed genes; FDR, false discovery rate.
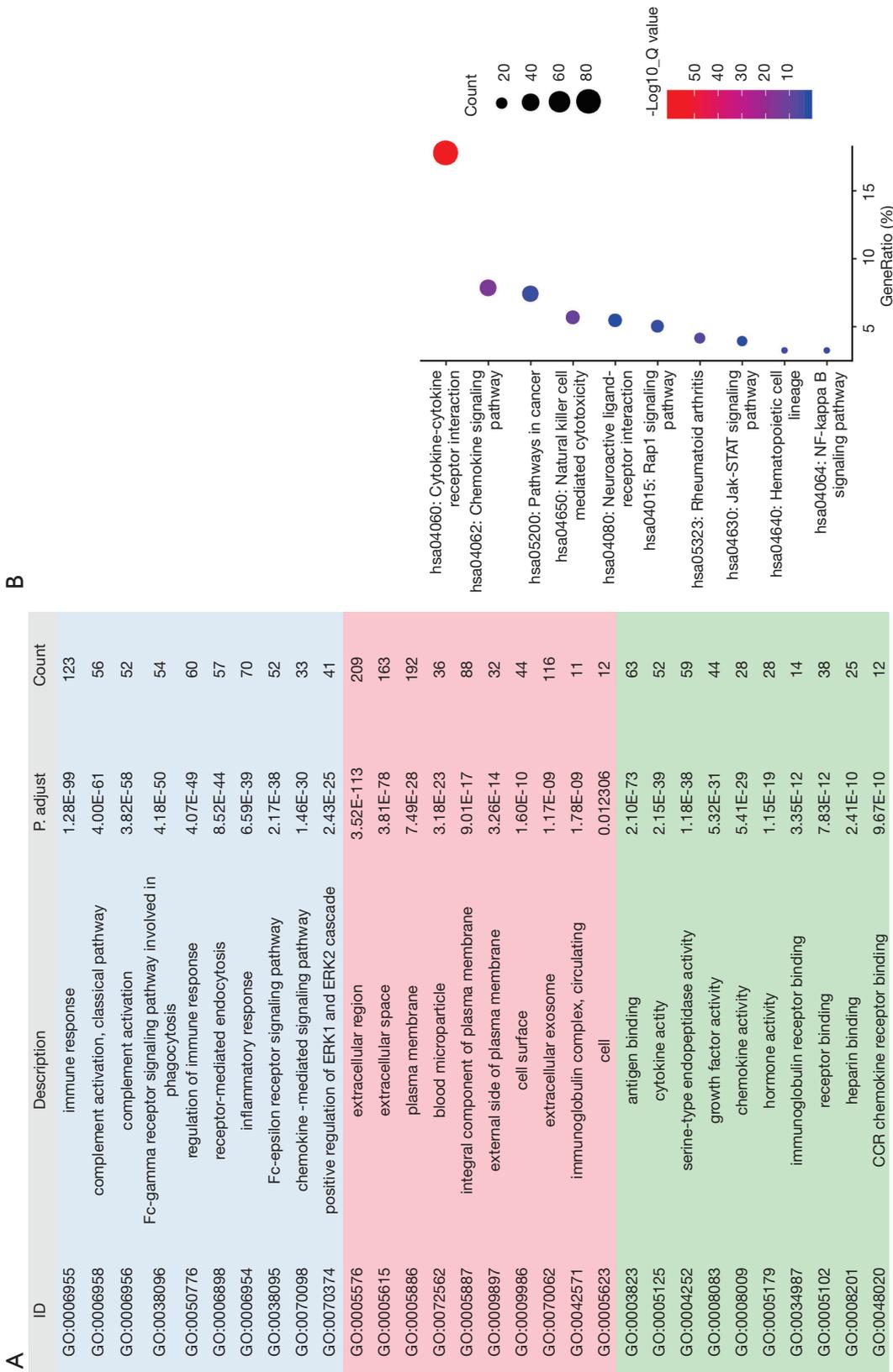
A

ID	Description	P. adjust	Count
GO:0006955	immune response	1.28E-99	123
GO:0006958	complement activation, classical pathway	4.00E-61	56
GO:0006956	complement activation	3.82E-58	52
GO:0038096	Fc-gamma receptor signaling pathway involved in phagocytosis	4.18E-50	54
GO:0050776	regulation of immune response	4.07E-49	60
GO:0006898	receptor-mediated endocytosis	8.52E-44	57
GO:0006954	inflammatory response	6.59E-39	70
GO:0038095	Fc-epsilon receptor signaling pathway	2.17E-38	52
GO:0070098	chemokine-mediated signaling pathway	1.46E-30	33
GO:0070374	positive regulation of ERK1 and ERK2 cascade	2.43E-25	41
GO:0005576	extracellular region	3.52E-113	209
GO:0005615	extracellular space	3.81E-78	163
GO:0005886	plasma membrane	7.49E-28	192
GO:0072562	blood microparticle	3.18E-23	36
GO:0005887	integral component of plasma membrane	9.01E-17	88
GO:0009897	external side of plasma membrane	3.26E-14	32
GO:0009986	cell surface	1.60E-10	44
GO:0070062	extracellular exosome	1.17E-09	116
GO:0042571	immunoglobulin complex, circulating	1.78E-09	11
GO:0005623	cell	0.012306	12
GO:0003823	antigen binding	2.10E-73	63
GO:0005125	cytokine actity	2.15E-39	52
GO:0004252	serine-type endopeptidase activity	1.18E-38	59
GO:0008083	growth factor activity	5.32E-31	44
GO:0008009	chemokine activity	5.41E-29	28
GO:0005179	hormone activity	1.15E-19	28
GO:0034987	immunoglobulin receptor binding	3.35E-12	14
GO:0005102	receptor binding	7.83E-12	38
GO:0008201	heparin binding	2.41E-10	25
GO:0048020	CCR chemokine receptor binding	9.67E-10	12

B

Figure 2 Functional enrichment analysis of differentially expressed IRGs. (A) GO enrichment analysis. Blue, red, and green parts represent BP, CC, and MF, respectively; (B) the top 10 KEGG signaling pathways. IRGs, immune-related genes; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto encyclopedia of genes and genomes.

Page 6 of 16

Qiang et al. An immune-related prognostic signature of COAD

tumor surveillance, molecular biomarkers related to IRGs could be used to evaluate the therapeutic efficiency of cancer immunotherapy. Based on univariate Cox analysis, 51 survival-related IRGs were distinguished (P<0.05). As shown in *Figure 3A*, the most relevant BP, CC, and MF terms of these genes are "regulation of response to stimulus", "extracellular space", and "receptor ligand activity", respectively. Again, "cytokine-cytokine receptor interaction" was found to be the most significant signaling pathway (*Figure 3B*).

Prognostic values of survival-related IRGs

In order to better define the characteristics of survival-related IRGs, the prognostic values of these genes were calculated. The resulting forest plot presented in *Figure 4* demonstrates that 40 genes are characterized by HR >1, whereas 11 genes have HR <1. This suggests that most survival-related IRGs were risk factors affecting the prognosis of COAD.

Development of IRGPI

In view of the results of the multivariate Cox regression analysis, the IRGPI was developed. This index was based on the following formula: risk score = [expression level of *CD1B* × (–2.09883)] + [expression level of *SLC10A2* × 0.718209] + [expression level of *FGF2* × 0.354982] + [expression level of *CCL28* × (–0.1162)] + [expression level of *IGHV4-31* × 0.00944] + [expression level of *IGKV1-6* × 0.006782] + [expression level of *ESM1* × 0.199876] + [expression level of *TNFSF12* × 0.09248] + [expression level of *UCN* × 0.376204] + [expression level of *VIP* × 0.066404] + [expression level of *GLP2R* × (–4.08525)] + [expression level of *IL1RL2* × 0.156916]. Depending on the scores and on the survival time, the patients were divided into high- (IRGPI > median value) and low-risk (IRGPI < median value) groups. As shown in *Figure 5*, the high-risk group exhibits more deaths than the low-risk one.

Evaluation of IRGPI

The calculated IRGPI values were used to determine the survival probability of COAD patients. As shown in

Figure 6A, the OS of high-risk group patients was found to be less than that of the low-risk group patients (P<0.001). To assess the predictive accuracy of the IRGPI established herein, the ROC curve was generated, and the AUC was calculated (*Figure 6B*). This area was found to be 0.855, which indicates that the proposed IRGPI has excellent potential in monitoring prognosis. Moreover, the results of both univariate and multivariate analyses (*Table 1*) demonstrate that IRGPI is significantly correlated with survival, and thus, it can be used as an independent indicator when multiple clinicopathologic factors are taken into account.

To further assess the clinical value of IRGPI, the relationships between the hub IRGs implicated in IRGPI and the clinicopathologic factors were analyzed. As shown in the results of *Table 2*, tumor status and M stage involved the expression of more genes. Besides, as an independent indicator, IRGPI had statistically significant difference in pathological stage (*Figure 7A*), T stage (*Figure 7B*), N stage (*Figure 7C*), and M stage (*Figure 7D*), which suggested that IRGPI could make an accurate prediction in different pathological stages of COAD tumors. However, IRGPI showed no significant difference in age (*Figure 7E*), gender (*Figure 7F*) and tumor status (*Figure 7G*).

Mutation landscape of hub IRGs included in IRGPI

Considering the significant clinical value of hub IRGs, their molecular characteristics were comprehensively explored. Specifically, genetic alteration analysis of these genes was carried out using the cBioPortal. The obtained results indicate that the most frequent type of genetic alteration is the elevated expression of mRNA (*Figure 8A*). For most IRGs (75%), the genetic alteration rate was found to be ≥5%, with the *ESM1* and *IL1RL2* genes showing the highest alteration frequencies (*Figure 8B*).

Clinical application of IRGPI

In order to estimate whether the IRGPI could accurately reflect the status of tumor immune microenvironments, we analyzed the relationship between IRGPI and the abundance of immune cell infiltration. The obtained results demonstrate that IRGPI is significantly correlated with 5
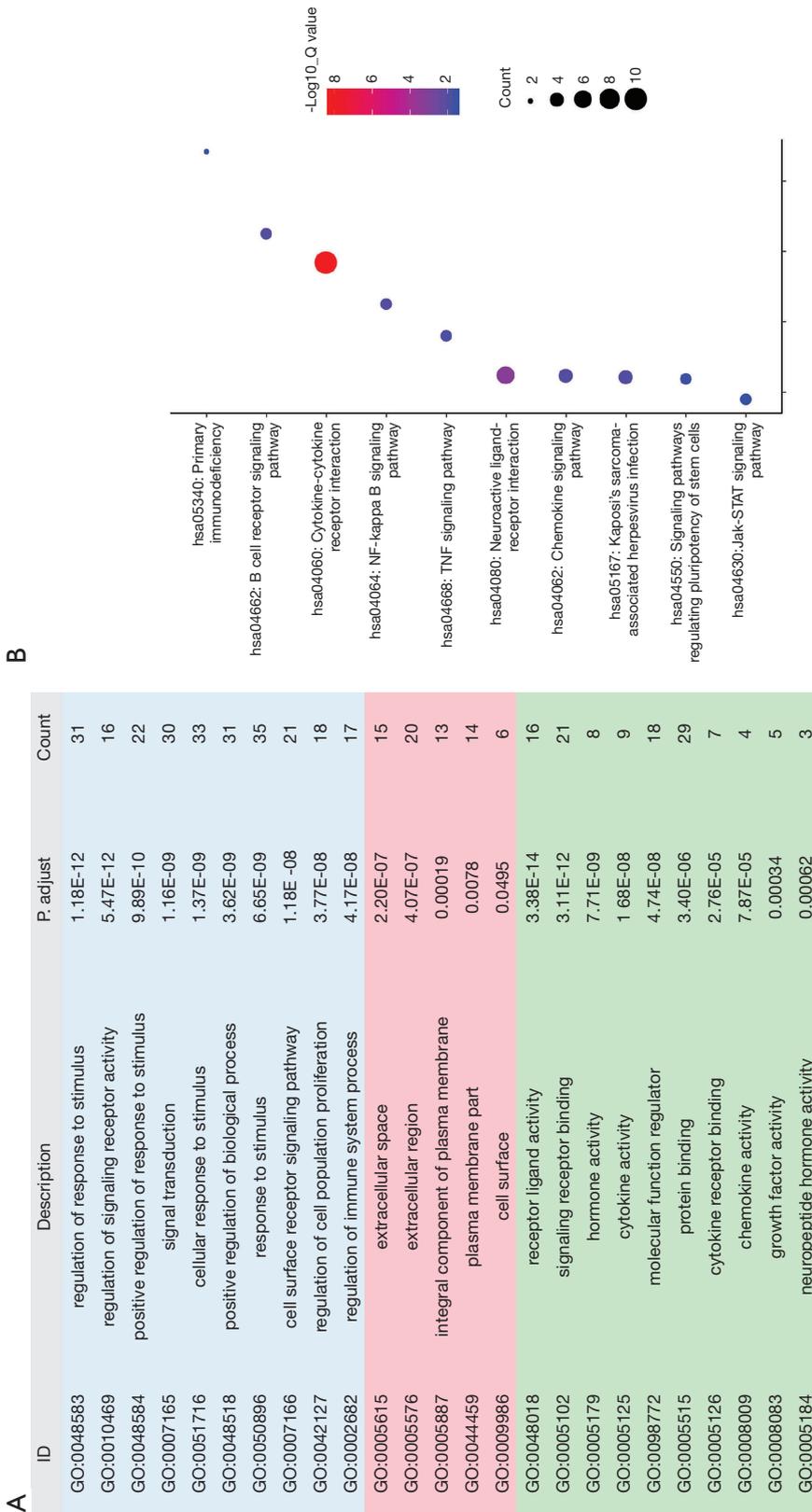
A

ID	Description	P. adjust	Count
GO:0048583	regulation of response to stimulus	1.18E-12	31
GO:0010469	regulation of signaling receptor activity	5.47E-12	16
GO:0048584	positive regulation of response to stimulus	9.89E-10	22
GO:0007165	signal transduction	1.16E-09	30
GO:0051716	cellular response to stimulus	1.37E-09	33
GO:0048518	positive regulation of biological process	3.62E-09	31
GO:0050896	response to stimulus	6.65E-09	35
GO:0007166	cell surface receptor signaling pathway	1.18E-08	21
GO:0042127	regulation of cell population proliferation	3.77E-08	18
GO:0002682	regulation of immune system process	4.17E-08	17
GO:0005615	extracellular space	2.20E-07	15
GO:0005576	extracellular region	4.07E-07	20
GO:0005887	integral component of plasma membrane	0.00019	13
GO:0044459	plasma membrane part	0.0078	14
GO:0009986	cell surface	0.0495	6
GO:0048018	receptor ligand activity	3.38E-14	16
GO:0005102	signaling receptor binding	3.11E-12	21
GO:0005179	hormone activity	7.71E-09	8
GO:0005125	cytokine activity	1 68E-08	9
GO:0098772	molecular function regulator	4.74E-08	18
GO:0005515	protein binding	3.40E-06	29
GO:0005126	cytokine receptor binding	2.76E-05	7
GO:0008009	chemokine activity	7.87E-05	4
GO:0008083	growth factor activity	0.00034	5
GO:0005184	neuropeptide hormone activity	0.00062	3

B

Figure 3 Functional enrichment analysis of survival-related IRGs. (A) GO enrichment analysis. Blue, red, and green parts represent BP, CC, and MF, respectively; (B) the top 10 KEGG signaling pathways. IRGs, immune-related genes; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto encyclopedia of genes and genomes.

Page 8 of 16

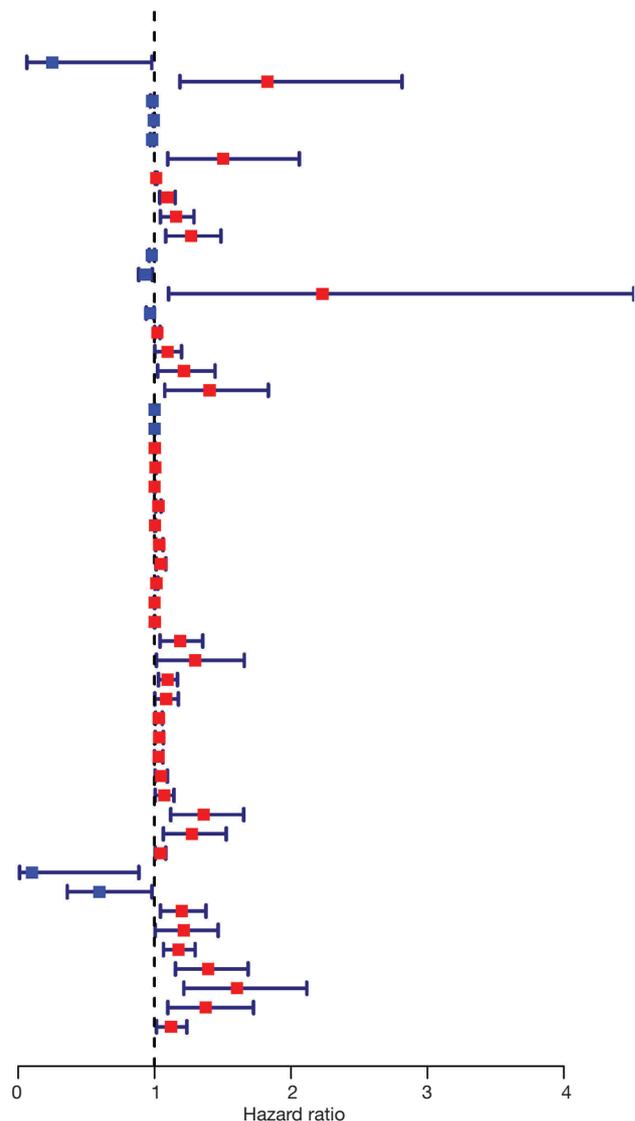Qiang et al. An immune-related prognostic signature of COAD

Figure 4 Prognostic values of survival-related IRGs. P values <0.05 are considered to be statistically significant. IRGs, immune-related genes.

TF-mediated regulatory network

In order to explore the molecular mechanisms of hub IRGs, we studied their regulatory mechanism. Knowing that TFs play a key role in regulating the molecular network, the expression profiles of 318 TFs were examined, and 68 differentially expressed TFs were identified (*Figure 10A,B*). As a result, a TF-mediated regulatory network was constructed based on 15 TFs and 8 hub IRGs (*Figure 10C*). Afterwards, we performed a topological analysis on these TFs based on three centrality algorithms, including degree centrality, betweenness centrality and closeness centrality. As shown in *Table 3*, *MAF*, *MYH11*, and *NR3C1* ranked the top 5 TFs in all three centrality analyses, and thus, they were considered as key TFs in the TF-mediated regulatory network.
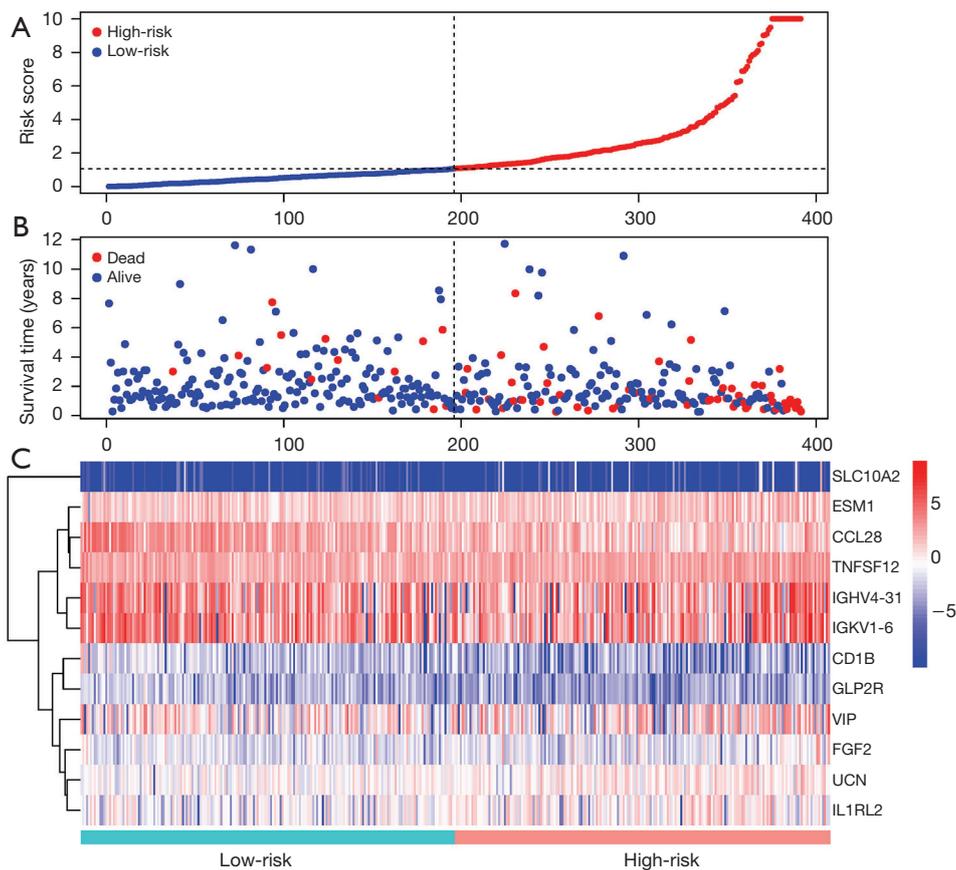
types of immune cells, including B cells (*Figure 9A*), CD4 T cells (*Figure 9B*), dendritic cells (*Figure 9C*), macrophages (*Figure 9D*), and neutrophils (*Figure 9E*). Moreover, there was no significant correlation between IRGPI and CD8 T cells (*Figure 9F*).

Figure 5 Development of the IRGPI. (A) Risk scores of patients in different groups; (B) survival status of patients in different groups; (C) heat map of hub IRG expression profiles. IRGPI, immune-related genes-based prognostic index.
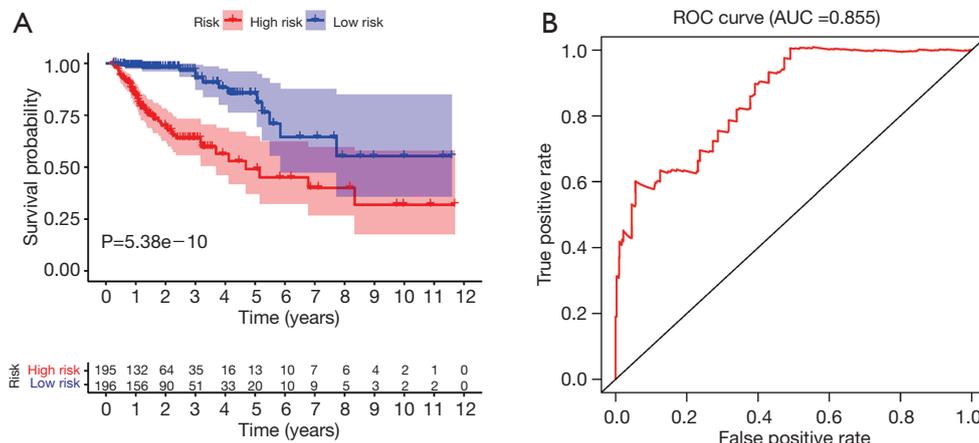

Figure 6 The evaluation of the IRGPI. (A) The K-M survival analysis of the IRGPI. Patients with high-risk scores showed poor OS, compared to those with low-risk scores; (B) survival-dependent ROC curve validation of the prognostic value of the IRGPI. The X-axis represents the false positive rate, while Y-axis represents the true positive rate. IRGPI, immune-related genes-based prognostic index; K-M, Kaplan-Meier; OS, overall survival; ROC, receiver operating characteristic; AUC, area under curve.

Page 10 of 16

Qiang et al. An immune-related prognostic signature of COAD

Table 1 Univariate and multivariate analysis of COAD

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
Age	1.024 (0.994–1.055)	0.117	1.029 (0.996–1.063)	0.090
Gender (male/female)	1.020 (0.524–1.987)	0.953	1.102 (0.528–2.299)	0.795
Tumor status (with tumor/tumor free)	2.673 (1.153–6.198)	0.022	1.662 (0.673–4.105)	0.271
Pathological stage	2.686 (1.804–3.998)	<0.001	0.727 (0.197–2.687)	0.633
T stage	3.637 (1.890–7.000)	<0.001	2.612 (1.157–5.895)	0.021
M stage	6.103 (3.110–11.976)	<0.001	5.443 (0.858–34.516)	0.072
N stage	2.143 (1.430–3.211)	<0.001	1.408 (0.689–2.879)	0.348
IRGPI	1.339 (1.226–1.463)	<0.001	1.354 (1.212–1.513)	<0.001

COAD, colon adenocarcinoma; HR, hazard ratio; IRGPI, immune-related genes-based prognostic index.

Discussion

The development of targeted drugs and cancer immunotherapy has shifted COAD treatment from surgery and chemoradiotherapy to precise and individualized therapies (29). However, due to the limitations of immunotherapy treatments, it is necessary to identify the effective population and predict the clinical outcome beforehand. This may be accomplished by analyzing the immune genes and prognostic status of patients. In this study, we established an IRGPI that may be used to evaluate immune response in COAD patients, and thus, serves as a new biomarker for cancer immunotherapy.

Our results demonstrate that the most relevant BP of differentially expressed IRGs is "immune response", and that the survival-related IRGs are highly associated with "the regulation of response to stimulus". Based on KEGG enrichment analysis, the survival-related IRGs were significantly correlated with "cytokine-cytokine receptor interaction", which suggested that the response of the immune system to stimuli is related to the expression of IRGs and patients' survival. Furthermore, assessments of the prognostic efficiency of survival-related IRGs showed that most genes are characterized by high HR values, and thus, they affect the prognosis of COAD patients. This explains the variations in the efficacy of the same treatment between different patients.

Currently, the methods used to identify potential mechanisms and prognostic biomarkers are based on different types of bioinformatic analyses. For example,

Dalerba *et al.* used the bioinformatics approach in combination with clinical-grade diagnostic assays to show that a subgroup of patients benefits from adjuvant chemotherapy (30). Other researchers used gene expression signatures based on mRNA, miRNA, or lncRNA expression profiles to predict the prognosis of COAD (31-33). In this study, we propose a biomarker based on IRGs, which is the most suitable way to reflect the immune status and tumor prognosis in patients. To establish this biomarker, the expression levels of survival-related IRGs were integrated using Cox regression analysis. Ultimately, 12 of 51 extracted genes were used to construct the predictive model that was used to calculate the risk scores of COAD patients. Based on these scores, the patients in the analyzed sample were divided into high- and low-risk groups.

The model was also used to establish an IRGPI. The reliability and efficiency of this index in terms of survival prediction was confirmed by ROC curve analysis. Univariate and multivariate analyses further prove the significant correlation between IRGPI and survival. Although the univariate analysis yields 5 independent indicators, only one (IRGPI) could be used in multivariate analysis. To test the feasibility of the IRGPI, the relationship between hub IRGs and multiple clinicopathological factors was also analyzed. The results indicate that as a comprehensive index, IRGPI exhibits stronger correlation with multiple pathological stages than single genes. Furthermore, the index is capable of accurate survival prediction in the pathological, T, N, and M stages.

Considering that cancer and immunity are closely associated, clinical treatments must take tumor immune microenvironments into account. Assessments of the correlation between the IRGPI and immune cell infiltration can be used to reflect the status of such microenvironments. Herein, we show that the IRGPI is significantly correlated with the infiltration levels of B cells, CD4[+] T cells, dendritic cells, macrophages, and neutrophils (P<0.05). These immune cells play a very important role in COAD prognosis. For example, Gu *et al.* have recently shown that the elimination of pathogenic CD4[+] T-cells and the induction of antitumor CD8[+] T-cell activity can suppress colon carcinogenesis (34). Also, Vo *et al.* demonstrate that dendritic cell and lenalidomide combination therapy can effectively enhance antitumor immunity in a mouse COAD model (35). Moreover, according to Vinnakota *et al.*, M2-like macrophages induce COAD cell invasion via matrix metalloproteinases (36).

To explore the molecular mechanisms underlying hub IRGs activity, a TF-mediated regulatory network was established. This network is comprised of several TFs that can simultaneously regulate the expressions of multiple IRGs, such as *MAF*, *MYH11*, and *NR3C1*. The available research indicates that *c-MAF* is capable of controlling immune responses by regulating disease-specific gene networks and repressing IL-2 in CD4[+] T cells (37). As for *MYH11*, it exhibits significantly different expressions in tumor samples, compared to normal tissues, and thus, it is considered as a potential tumor biomarker (38). Finally, the *NR3C1* gene expressed during CD8[+] T cell differentiation regulates the formation of memory-precursor cell fates (39). Other studies have also investigated the role of TFs in COAD prognosis. For instance, Wei *et al.* established a regulatory network of TF-miRNA-target genes to analyze disease progression and optimize treatment strategies (40). However, Mullany *et al.* showed that TF expression and related miRNAs influence the survival of patients diagnosed with colorectal cancer (41). Based on these studies and on our own results, we believe that TFs may play an important role in regulating the IRG expressions in COAD patients. Further investigation is needed to confirm this hypothesis.

Considering that the IRGPI established herein may be applied to accurately assess the immune status and monitor COAD prognosis in patients, it is actually an outstanding biomarker of the disease. This robust biomarker can be used in clinical prognosis applications, including cancer immunotherapy.
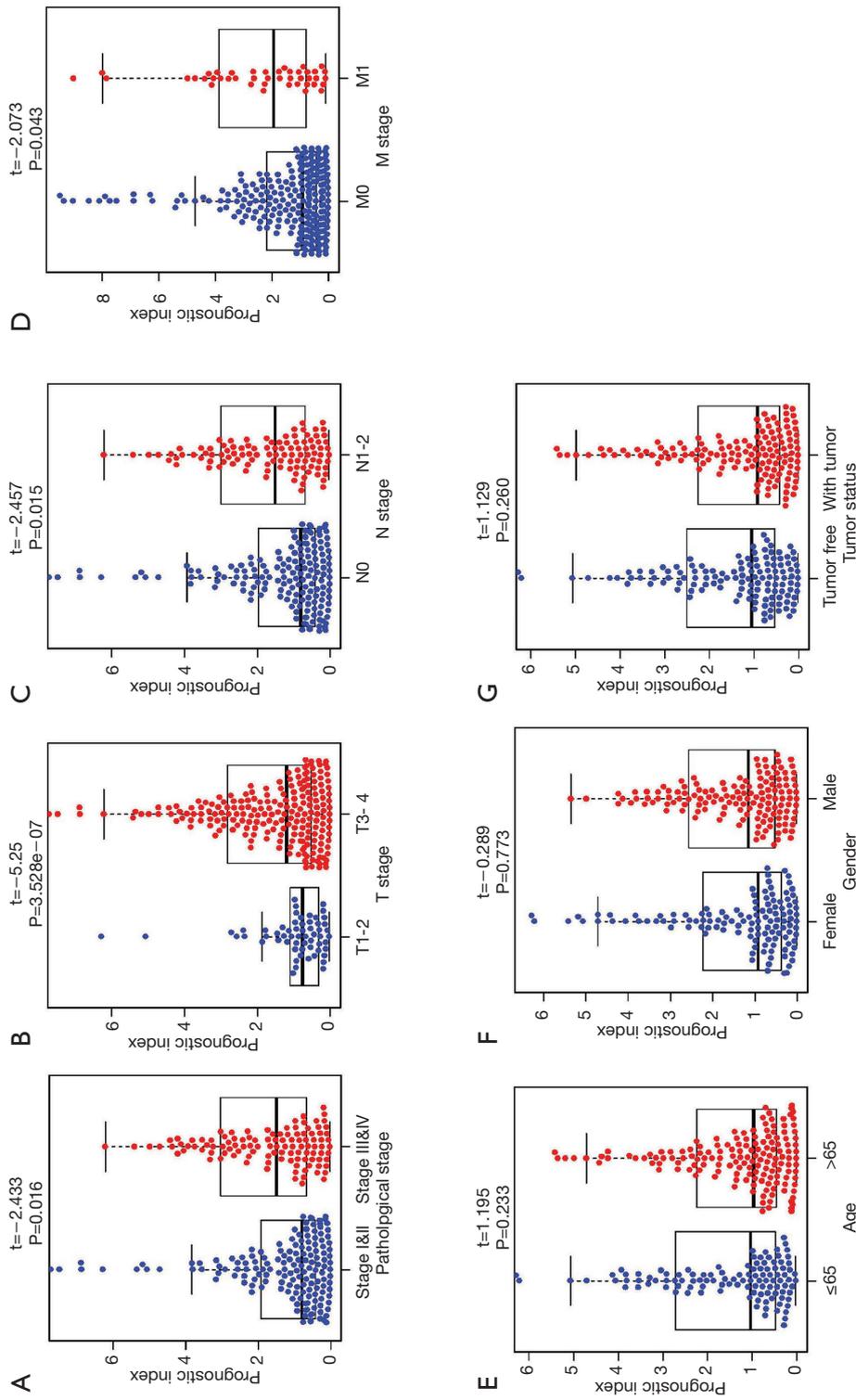
Table 2 The correlation between the expression of the IRGs and the clinical factors in COAD

Genes	Age (>65/≤65)		Gender (male/female)		Tumor status (with tumor/tumor free)		Pathological stage (III&IV/I&II)		T stage (T3–4/T1–2)		M stage (M1/M0)		N stage (N1–3/N0)	
	t	P	t	P	t	P	t	P	t	P	t	P	t	P
CD1B	-0.155	0.877	1.717	0.087	1.519	0.13	2.849	0.005	1.579	0.118	4.658	8.85E-06	2.653	0.008
SLC10A2	0.775	0.439	1.711	0.089	0.859	0.391	-0.193	0.847	-0.871	0.385	2.211	0.028	-0.256	0.798
FGF2	-1.014	0.311	-1.2	0.231	4.287	2.98E-05	0.449	0.654	0.354	0.725	2.743	0.007	0.152	0.879
CCL28	-2.463	0.014	0.09	0.928	-2.198	0.029	2.192	0.029	0.612	0.542	3.013	0.003	2.167	0.031
IGHV4-31	-1.508	0.133	-2.096	0.037	-2.717	0.007	1.675	0.095	0.573	0.568	2.366	0.02	1.451	0.148
IGKV1-6	-0.071	0.944	-0.478	0.633	-0.706	0.481	0.459	0.646	-0.899	0.37	1.955	0.054	0.353	0.725
ESM1	1.269	0.205	0.033	0.973	2.314	0.021	-1.638	0.103	-1.866	0.065	-1.132	0.263	-1.843	0.067
TNFSF12	1.917	0.056	0.864	0.388	1.455	0.147	-0.761	0.447	-1.383	0.169	0.469	0.641	-1.002	0.317
UCN	-1.765	0.079	0.298	0.766	-5.221	3.80E-07	-2.052	0.041	-0.941	0.349	-1.135	0.262	-1.866	0.064
VIP	0.993	0.321	0.907	0.365	2.48	0.014	-1.672	0.096	-2.215	0.029	-0.586	0.56	-2.071	0.04
GLP2R	-0.88	0.379	0.408	0.684	-0.402	0.688	-1.293	0.197	0.363	0.718	0.318	0.751	-1.323	0.187
IL1RL2	-0.956	0.34	-0.078	0.938	0.722	0.471	-1.045	0.297	-0.257	0.798	-1.622	0.111	-0.481	0.631
IRGPI	1.195	0.233	-0.289	0.773	1.129	0.26	-2.433	0.016	-5.25	3.53E-07	-2.073	0.043	-2.457	0.015

IRGs, immune-related genes; COAD, colon adenocarcinoma; IRGPI, immune-related genes-based prognostic index.

Figure 7 The relationships between the prognostic index and (A) pathological stage, (B) T stage, (C) N stage, (D) M stage, (E) age, (F) gender, and (G) tumor status.
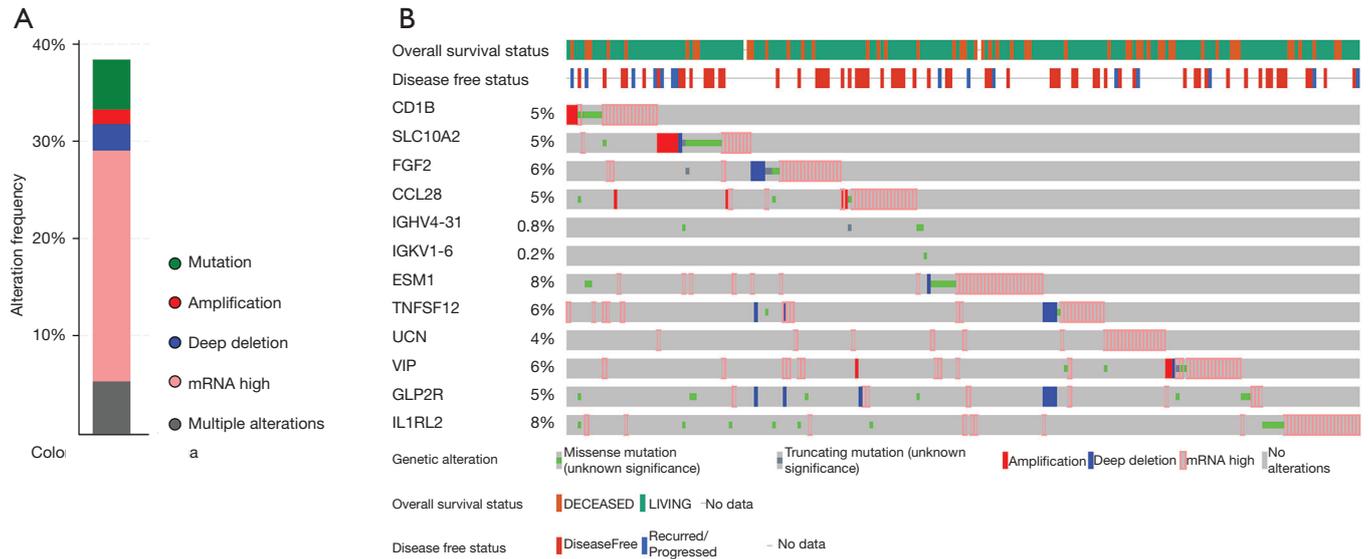
Figure 8 Genetic alterations associated with hub IRGs. (A) The total alteration frequency of hub IRGs; (B) the landscape of genetic alterations of hub IRGs in COAD patient. IRGs, immune-related genes; COAD, colon adenocarcinoma.
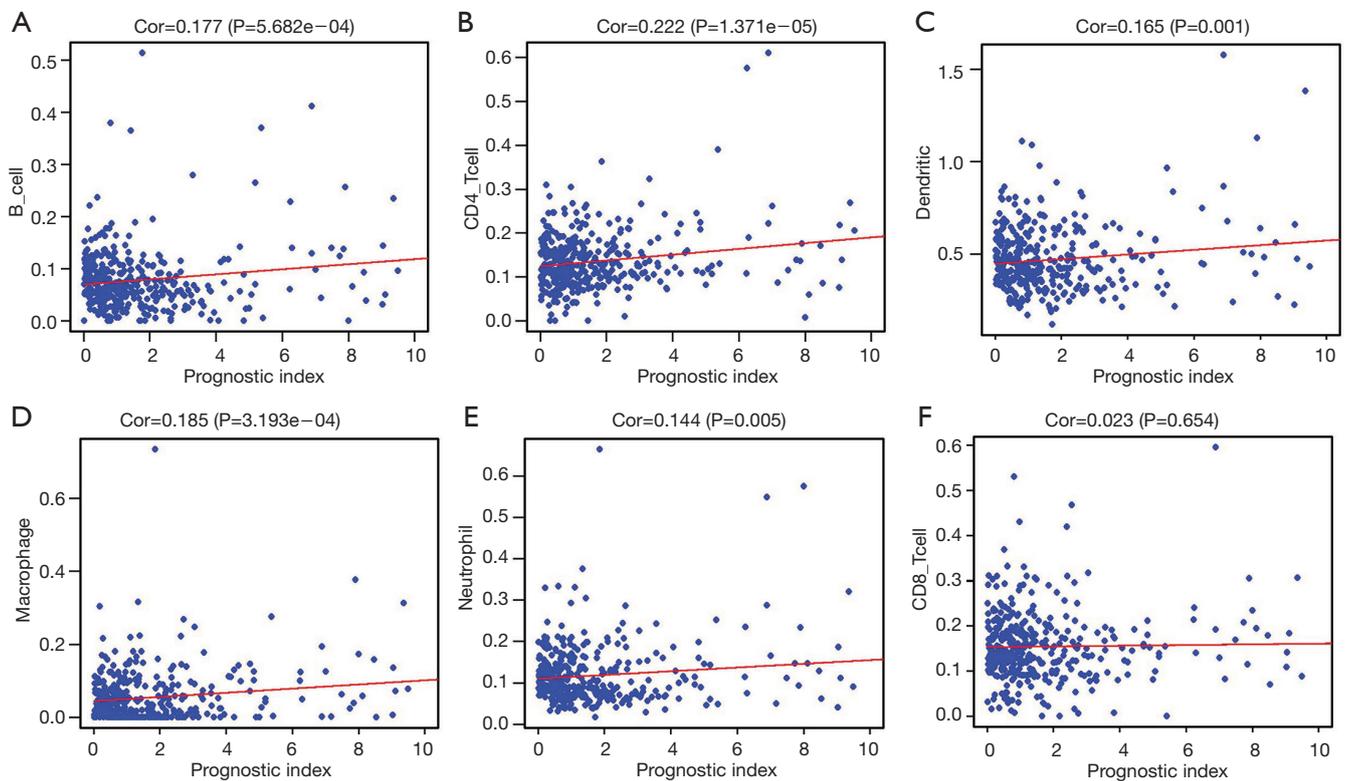

Figure 9 Relationships between the prognostic index and the abundance of immune cell infiltration. (A) B cells, (B) CD4 T cells, (C) dendritic cells, (D) macrophages, (E) neutrophils, and (F) CD8 T cells.
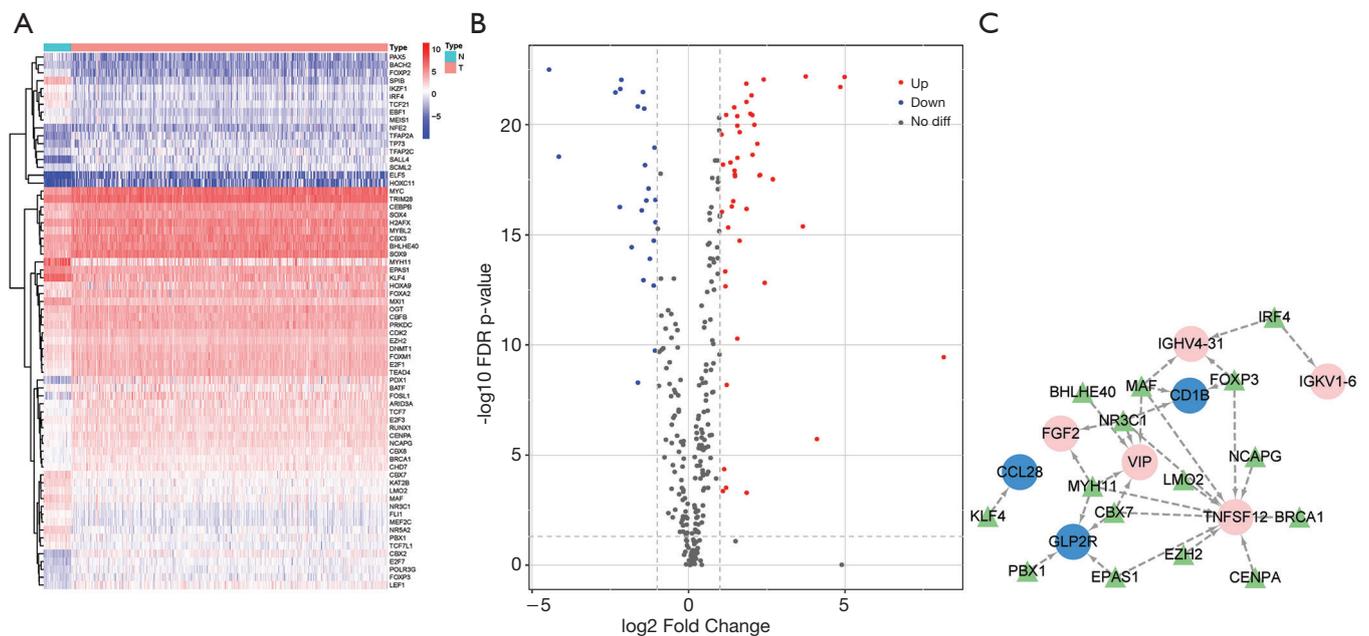
Figure 10 TF-mediated regulatory network. The heatmap (A) and volcano plot (B) of differentially expressed TFs. Red and blue dots represent up-regulated and down-regulated differentially expressed TFs, respectively; (C) the regulatory network. Green triangles represent TFs, whereas red and blue circles represent high-risk and low-risk IRGs, respectively. IRGs, immune-related genes; TF, transcription factor.

Table 3 The topological analysis of TFs in regulating the hub IRGs

TF name	Degree centrality	TF name	Betweenness centrality	TF name	Closeness centrality
MAF	4	MAF	0.202086	KLF4	1
MYH11	4	MYH11	0.129743	MAF	0.487805
NR3C1	4	FOXP3	0.114229	MYH11	0.465116
CBX7	3	IRF4	0.1	FOXP3	0.444444
FOXP3	3	NR3C1	0.094054	NR3C1	0.444444
EPAS1	2	CBX7	0.079211	CBX7	0.444444
IRF4	2	EPAS1	0.038571	EPAS1	0.408163
BHLHE40	1	BHLHE40	0	BRCA1	0.377358
BRCA1	1	BRCA1	0	CENPA	0.377358
CENPA	1	CENPA	0	EZH2	0.377358
EZH2	1	EZH2	0	LMO2	0.377358
KLF4	1	KLF4	0	NCAPG	0.377358
LMO2	1	LMO2	0	BHLHE40	0.307692
NCAPG	1	NCAPG	0	IRF4	0.28169
PBX1	1	PBX1	0	PBX1	0.266667

TF, transcription factor; IRGs, immune-related genes.

Acknowledgments

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
2. Banerjee A, Pathak S, Subramanium VD, et al. Strategies for targeted drug delivery in treatment of colon cancer: current trends and future perspectives. Drug Discov Today 2017;22:1224-32.
3. Bonetti A, Giuliani J, Muggia F. Targeted agents and oxaliplatin-containing regimens for the treatment of colon cancer. Anticancer Res 2014;34:423-34.
4. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 2013;14:1014-22.
5. Chen YP, Zhang Y, Lv JW, et al. Genomic analysis of tumor microenvironment immune types across 14 solid cancer types: immunotherapeutic implications. Theranostics 2017;7:3585-94.
6. Sokratous G, Polyzoidis S, Ashkan K. Immune infiltration of tumor microenvironment following immunotherapy for glioblastoma multiforme. Hum Vaccin Immunother 2017;13:2575-82.
7. Locy H, de Mey S, de Mey W, et al. Immunomodulation of the tumor microenvironment: turn foe into friend. Front Immunol 2018;9:2909.
8. Newick K, O'Brien S, Moon E, et al. CAR T cell therapy for solid tumors. Annu Rev Med 2017;68:139-52.
9. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discov 2019;18:197-218.
10. Makkouk A, Weiner GJ. Cancer immunotherapy and breaking immune tolerance: New approaches to an old challenge. Cancer Res 2015;75:5-10.
11. Roychowdhury S, Chinnaiyan AM. Translating cancer genomes and transcriptomes for precision oncology. CA Cancer J Clin 2016;66:75-88.
12. Li T, Kung HJ, Mack PC, et al. Genotyping and genomic profiling of non-small-cell lung cancer: implications for current and future therapies. J Clin Oncol 2013;31:1039-49.
13. Wang Z, Jensen MA, Zenklusen JC. A practical guide to The Cancer Genome Atlas (TCGA). Methods Mol Biol 2016;1418:111-41.
14. Jensen MA, Ferretti V, Grossman RL, et al. The NCI Genomic Data Commons as an engine for precision medicine. Blood 2017;130:453-9.
15. Bhattacharya S, Andorf S, Gomes L, et al. ImmPort: disseminating data to the public for the future of immunology. Immunol Res 2014;58:234-9.
16. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
17. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847-9.
18. Ito K, Murphy D. Application of ggplot2 to pharmacometric graphics. CPT Pharmacometrics Syst Pharmacol 2013;2:e79.
19. Rizvi AA, Karaesmen E, Morgan M, et al. gwasurvivr: an R package for genome-wide survival analysis. Bioinformatics 2019;35:1968-70.
20. Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017;45:D353-61.

21. Vesztrocy AW, Dessimoz C. A gene ontology tutorial in python. Methods Mol Biol 2017;1446:221-9.

22. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.

23. Zhang Z, Reinikainen J, Adeleke KA, et al. Time-varying covariates and coefficients in Cox regression models. Ann Transl Med 2018;6:121.

24. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.

25. Li T, Fan J, Wang B, et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res 2017;77:e108-10.

26. Tang Q, Chen Y, Meyer C, et al. A comprehensive view of nuclear receptor cancer cistromes. Cancer Res 2011;71:6940-7.

27. Assenov Y, Ramírez F, Schelhorn SE. et al. Computing topological parameters of biological networks. Bioinformatics 2008;24:282-4.

28. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Med Res Methodol 2017;17:53.

29. Dienstmann R, Salazar R, Tabernero J. Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. J Clin Oncol 2015;33:1787-96.

30. Dalerba P, Sahoo D, Paik S, et al. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. N Engl J Med 2016;374:211-22.

31. Sun D, Chen J, Liu L, et al. Establishment of a 12-gene expression signature to predict colon cancer prognosis. PeerJ 2018;6:e4942.

32. Jacob H, Stanisavljevic L, Storli KE, et al. Identification of a sixteen-microRNA signature as prognostic biomarker for stage II and III colon cancer. Oncotarget 2017;8:87837-47.

33. Xing Y, Zhao Z, Zhu Y, et al. Comprehensive analysis of differential expression profiles of mRNAs and lncRNAs and identification of a 14-lncRNA prognostic signature for patients with colon adenocarcinoma. Oncol Rep 2018;39:2365-75.

34. Gu T, De Jesus M, Gallagher HC, et al. Oral IL-10 suppresses colon carcinogenesis via elimination of pathogenicCD4+ T-cells and induction of antitumor CD8+ T-cell activity. Oncoimmunology 2017;6:e1319027.

35. Vo MC, Nguyen-Pham TN, Lee HJ, et al. Combination therapy with dendritic cells and lenalidomide is an effective approach to enhance antitumor immunity in a mouse colon cancer model. Oncotarget 2017;8:27252-62.

36. Vinnakota K, Zhang Y, Selvanesan BC, et al. M2-like macrophages induce colon cancer cell invasion via matrix metalloproteinases. J Cell Physiol 2017;232:3468-80.

37. Gabryšová L, Alvarez-Martinez M, Luisier R, et al. C-Maf controls immune responses by regulating disease-specific gene networks and repressing IL-2 in CD4+ T cells article. Nat Immunol 2018;19:497-507.

38. Lee WS, Seo G, Shin HJ, et al. Identification of differentially expressed genes in microsatellite stable HNPCC and sporadic colon cancer. J Surg Res 2008;144:29-35.

39. Yu B, Zhang K, Milner JJ, et al. Epigenetic landscapes reveal transcription factors that regulate CD8+ T cell differentiation. Nat Immunol 2017;18:573-82.

40. Wei S, Chen J, Huang Y, et al. Identification of hub genes and construction of transcriptional regulatory network for the progression of colon adenocarcinoma hub genes and TF regulatory network of colon adenocarcinoma. J Cell Physiol 2020;235:2037-48.

41. Mullany LE, Herrick JS, Wolff RK, et al. Transcription factor-microRNA associations and their impact on colorectal cancer survival. Mol Carcinog 2017;56:2512-26.