

# The IRF family can influence tumor immunity and the prognosis of patients with colorectal cancer

**Yan-Jie Chen**

Zhongshan Hospital Fudan University

**Shu-Neng Luo**

Zhongshan Hospital Fudan University

**Ling Dong**

Zhongshan Hospital Fudan University

**Tao-Tao Liu**

Zhongshan Hospital Fudan University

**Xi-Zhong Shen**

Zhongshan Hospital Fudan University

**Ning-Ping Zhang**

Zhongshan Hospital Fudan University

**Li Liang** (✉ [liang.li@zs-hospital.sh.cn](mailto:liang.li@zs-hospital.sh.cn))

Zhongshan Hospital Fudan University <https://orcid.org/0000-0002-1694-380X>

---

## Research Article

**Keywords:** IRF family, colorectal cancer, colon adenocarcinoma/rectum adenocarcinoma, prognosis, immune cell infiltration

**Posted Date:** April 2nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-344747/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** The roles of interferon-regulatory factors (IRFs) in colorectal cancer (CRC) have not been studied through bioinformatics analysis.

**Methods:** We used gene- and microRNA-expression data for patients with somatic mutations and colon adenocarcinoma/rectum adenocarcinoma from The Cancer Genome Atlas Genomic Data Commons as a training dataset. Gene-expression data (accessions GSE17536 and GSE39582) were downloaded from the Gene Expression Omnibus database as the validation dataset. Expressional differences, clinical correlations, disease prognosis, functional enrichment, and immune correlations of IRF genes were analyzed. The results were validated via immunohistochemistry.

**Results:** The mRNA-expression levels of IRF3 and IRF7 differed between tumor and normal tissues and were correlated with patient prognosis. The IRF score was an independent risk factor for overall survival. IRFs recruited inflammatory cells; however, the immune and stromal scores showed inflammatory-cell recruitment only in the tumor stroma; therefore, they did not help eliminate tumor cells. Functional-enrichment analysis and pan-cancer expression analysis revealed that IRFs were differentially expressed in tumor tissues and associated with patient prognosis.

**Conclusions:** IRFs were differentially expressed in tumor tissues and were associated with prognosis in CRC patients. Although IRFs can promote the infiltration of immune cells, the immune and stromal score showed that the infiltrated immune cells mostly stayed in the tumor stroma and cannot directly eliminate the tumor. Our findings can help to improve CRC prognosis and treatment strategies.

## Background

Colorectal cancer (CRC) is one of the leading causes of cancer-related death worldwide [1]. Approximately 1.8 million new CRC cases and >860,000 CRC-related deaths occurred globally in 2018, making CRC the third most frequent cancer worldwide [1, 2]. CRC develops through a multistep process characterized by accumulated genetic and epigenetic abnormalities that cause genomic instability and mutations in tumor-suppressor and oncogenic genes [3]. Most CRC lesions show little sensitivity to immune-checkpoint inhibitor-based therapies, although immunologic parameters may have prognostic value [4]. Therefore, further research on the tumor immunity of CRC will provide a theoretical basis for developing CRC immunotherapeutics.

Interferons were first discovered as antiviral proteins, and subsequently, interferon-regulatory factors (IRFs) were discovered and studied intensively. IRFs are transcription factors participating in interferon gene regulation [5]. The amino termini of IRFs contain a DNA-binding domain (DBD) composed of 115 amino acids (similar to the DBD of Myb) and can bind DNA. The carboxyl termini of IRFs have a variable region that serves various biological functions [6]. Ten IRFs (IRF1–IRF9 and virus IRF) have been discovered. IRFs are found in various tissues and play important roles in cell-cycle regulation, cell differentiation, apoptosis, and tumor-immune regulation [6]. Future studies on IRFs will provide a

theoretical basis for their mechanistic roles in tumor development and tumor immunity and for choosing drug therapies.

The roles of IRFs in CRC have not been investigated through bioinformatics analysis. Here, we used public databases to analyze IRF-expression levels and mutations in patients with CRC to determine distinct prognostic values, study tumor-immunity regulation, and identify potential functions of IRFs in CRC. We verified these results via immunohistochemistry (IHC) analysis with our own cohort of patients with CRC.

## Materials And Methods

### Data Acquisition

Data regarding fragments per kilobase million (FPKM) values and microRNA (miRNA)-expression levels for patients with colon adenocarcinoma/rectum adenocarcinoma (COAD/READ) were downloaded from The Cancer Genome Atlas (TCGA) Genomic Data Commons website (<https://portal.gdc.cancer.gov/>) and used as the training dataset. FPKM values were converted to transcripts per million values and divided into mRNA- and long non-coding RNA (lncRNA)-expression groups. “Masked Somatic Mutation” data were downloaded for patients with COAD/READ, pre-processed using VarScan software, and visualized using the R software package, maftools [7]. The clinicopathological features and prognoses of patients with COAD/READ, such as gender, age, and stage, were downloaded from the UCSC Xena website (<http://xena.ucsc.edu/>). After removing samples with missing clinical information, 597 tumor samples and 51 normal tissue samples were obtained. Table 1 shows clinical information for patients with COAD/READ. The likelihood of each response to immunotherapy was predicted using the Tracking of Indels by DEcomposition (TIDE) algorithm (<http://tide.dfci.harvard.edu>) [8]. Gene-expression data from different organizations and in different cell lines were downloaded from TCGA and the Cell Line Cancer Encyclopedia (CCLE) databases (<https://portals.broadinstitute.org/ccle/about>) to compare IRF-expression levels between tumor and normal tissues.

Gene-expression data in GSE17536 [9] and GSE39582 [10] and clinicopathological patient characteristics were downloaded as validation datasets from the Gene Expression Omnibus (GEO) database. The data were downloaded from Homo Sapiens; this platform is based on the GPL570 [HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array. GSE17536 included 177 COAD tissue samples, and GSE39582 included 566 COAD tissue samples and 19 colon non-tumor tissue samples.

### Genetic Characteristics of the IRF Family and Validation with Clinical Prediction Models

We incorporated the expression levels of IRF-family genes into our model. The random forest package of R [11] was used to develop an IRF-based risk-assessment model for patients with COAD/READ. Patients were divided into high- and low-IRF risk groups, based on the median value.

To assess patient prognosis by combining the IRF risk score with clinicopathological features, univariate and multivariate Cox proportional-hazards analysis were used to analyze the independent predictive power of risk scores for overall survival (OS) and disease-free survival (DFS). Subsequently, a survival-prediction nomogram was constructed for patients in the TCGA dataset and was validated for patients in the GEO dataset. To quantify differentiation performance, Harrell's consistency index (C-index) was measured. A calibration curve was generated to evaluate the performance of the line map by comparing the predicted value of the line map with the observed OS rate.

### **Differentially Expressed Genes (DEGs) and Clinical Correlation Analysis**

Data for patients with COAD/READ were downloaded from TCGA and the GEO databases, and the patients were divided into high- and low-expression groups, according to the IRF score. The DESeq2 package of R [12] was used to analyze DEGs in both groups, where a log fold-change (logFC)>1.0 and  $P<0.05$  were set as threshold values for DEGs.

We compared the expression levels of IRF-family genes at different TNM stages. The Human Protein Atlas (HPA, <https://www.proteinatlas.org>) provides immunohistochemical expression data for nearly 20 different cancers [13] and enables identification of tumor type-specific, differential protein-expression patterns, where protein-expression levels of all IRF-family genes were compared between normal and CRC tissues.

### **Functional-Enrichment Analysis and Gene-Set Enrichment Analysis (GSEA)**

Gene Ontology (GO) analysis is commonly used for large-scale functional-enrichment research of biological processes (BPs), molecular functions, and cellular components. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a widely used database containing information regarding genomes, biological pathways, diseases, and drugs. GO and KEGG pathway-enrichment analyses were performed with signature genes using the clusterProfiler R package [14]. A false-discovery rate of  $<0.05$  was considered statistically significant.

To investigate differences in BPs among different subgroups, GSEA was performed using the gene-expression profiles of patients with COAD/READ. GSEA can be used to identify statistical differences between two groups in a gene set and estimate changes in pathways and BP activities [15]. The gene set "C2.CP.kegg.V6.2.-symbols" [15] was downloaded from the Molecular Signatures Database for GSEA analysis. An adjusted  $P$  value of  $<0.05$  was considered statistically significant.

### **Constructing a Protein–Protein-Interaction (PPI) network and Screening Hub Genes**

We used the Search Tool for Retrieving Interacting Genes (STRING) database [16], which predicts PPIs, to construct PPI networks for the selected genes. Genes with scores of  $>0.4$  were selected to construct a network model, which was visualized with Cytoscape V3.7.2 [17]. In the co-expression network, the maximum clique centrality (MCC) algorithm most effectively located the node in a set. The MCC of each

node was calculated using CytoHubba plugins [18] in Cytoscape, and genes with the highest eight MCC values were selected as hub genes.

### **Constructing a Competing-Endogenous RNA (ceRNA) Network Based on miRNA–mRNA and miRNA–lncRNA interactions**

lncRNA–miRNA-interaction data were downloaded from the miRcode database. and miRNA–mRNA-interaction data were downloaded from the miRTarBase, miRDB, and TargetScan databases. The DESeq2 packet of R [12] was used to analyze miRNA- and lncRNA-expression differences between the high-IRF and low-IRF risk groups.  $\text{LogFC} > 1.0$  and  $P < 0.05$  were set as criteria for a statistically significant difference. Cytoscape (V3.7.2) was used to construct a ceRNA network by analyzing correlations between lncRNA- and mRNA-regulated miRNAs simultaneously.

### **Tumor Immune Estimation Resource (TIMER)-Database Analysis and Comparing Immune-Correlation Scores Between Both Groups**

The TIMER database (<https://cistrome.shinyapps.io/timer/>) enables users to estimate B-cell, CD4+ T-cell, CD8+ T-cell, macrophage, neutrophil, and dendritic-cell infiltration into different tumor types [19]. We used the TIMER database to analyze correlations between the expression levels of different IRF genes and immune cell infiltrations in COAD/READ samples.

The R estimate package [20] quantifies immune cell-infiltration levels in tumor samples, based on gene-expression profiles, and was used to assess the immune-activity and stromal score of each tumor sample. Immune cell-infiltration levels between both groups were compared using the Mann–Whitney U test.

### **Analysis of Anticancer Therapy-Sensitivity**

The Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://www.cancerrxgene.org/>) enables exploration of gene mutations and targeted cancer therapies. We downloaded gene-expression data from cell lines and  $\text{IC}_{50}$  values to analyze correlations between differentially expressed-IRF genes and anticancer drug sensitivities.

### **Calculating Tumor-Mutation Load Fractions and Analyzing Genetic Variation of IRF Family Members in CRC**

The tumor burden (TMB) of each tumor sample was defined as the number of somatic cell mutations identified, excluding silent mutations. Patients with COAD/READ were divided into high-TMB and low-TMB groups according to the median TMB value. The Wilcoxon rank-sum test was used to compare risk scores of IRF-family genes between both groups.

### **Patients and Specimens in the Validation Cohort**

Tumor specimens were obtained from 114 patients with CRC who underwent treatment at Zhongshan Hospital (Fudan University) between 2008 and 2016. The inclusion criteria were as follows: (a) a clear pathological diagnosis of CRC, (b) complete follow-up data until December 2019, (c) suitable formalin-fixed and paraffin-embedded tissues, and (d) agreeing to participate in the study and provide signed informed consent. CRC diagnosis was based on the World Health Organization criteria, and tumor stages were classified according to the 7<sup>th</sup> edition of TNM classification of International Union Contra Cancrum. Ethical approval was obtained from the research ethics committee of Zhongshan Hospital. The clinical characteristics of the 102 patients with follow-up data are presented in Supplemental Table 1.

## **IHC-staining Evaluation**

Cancer and adjacent normal tissues were formalin-fixed, paraffin-embedded, and prepared as tissue microarrays (TMAs) after hematoxylin and eosin staining and histopathology-guided location. Five-micron-thick TMA sections were deparaffinized and rehydrated in 0.1 M citrate buffer (pH 6.0), followed by high-temperature antigen retrieval in a microwave for 15 min. The sections were incubated overnight at 4°C with primary antibodies against IRF3 and IRF7 (Abcam, Cambridge, UK), CD4 (Servicebio Technology, Wuhan, China), CD8 (Servicebio Technology), CD19 (Servicebio Technology), and CD68 (Servicebio Technology). The sections were incubated for 30 min with a secondary antibody at room temperature and immunostained based on avidin–biotin complex formation, using 3,3'-diaminobenzidine. Hematoxylin was used as a counterstain.

Antigen–antibody complexes in whole samples were detected using a panoramic slice scanner (3DHISTECH, Hungary) and viewed with CaseViewer 2.2 (3DHISTECH). H-scores were calculated to evaluate gene-expression levels using Quant Center 2.1 (3DHISTECH):  $H\text{-score} = \sum (PI \times I) = (\% \text{ of weakly stained cells} \times 1) + (\% \text{ of moderately stained cells} \times 2) + (\% \text{ of strongly stained cells} \times 3)$ , where PI is the proportion of the positive area, and I is the staining intensity.

## **Statistical Analysis**

The data were analyzed with R software (V4.0.2). The independent Student t test was used to estimate the statistical significance of normally distributed variables, and the Mann–Whitney U test was used to analyze differences in non-normally distributed variables between two groups of continuous variables. The chi-squared test or Fisher exact test was used to analyze statistical significance between two groups of categorical variables. Correlation coefficients between different genes were calculated by Pearson correlation analysis. The survival package of R was used for survival analysis, Kaplan–Meier survival curves were used to determine survival differences, and the log-rank test was used to evaluate significant differences in survival times between two groups. Univariate and multivariate Cox analyses were used to determine independent prognostic factors. The pROC package of R [21] was used to draw receiver operating-characteristic (ROC) curves, and area under the curve (AUC) values were calculated to assess the accuracy of risk scores for prognosis estimations. All statistical *P* values were bilateral, and *P*<0.05 was considered statistically significant.

# Results

## Transcriptional Levels of IRFs in Patients with COAD/READ and Their Effect on Prognosis

Principal-component analysis (PCA) of TCGA data revealed differentially expressed IRF genes in patients with COAD/READ (Figure 1A). Specifically, IRF3, IRF4, IRF7, and IRF9 showed significant differences between cancer and normal tissues (Figure 1B;  $|\log_{2}FC| > 1.0$  and  $P < 0.05$ ). IRF3, IRF7, and IRF9 mRNAs were significantly upregulated in tumor tissues, whereas IRF4 mRNA was downregulated ( $P < 0.05$ ; Figure 1C). ROC analysis showed that their expression levels had good diagnostic value for COAD/READ (IRF3: AUC=0.908, IRF4: AUC=0.912, IRF7: AUC=0.623, IRF9: AUC=0.660; Figure 1C).

Analyzing protein-expression levels in CRC and normal tissues with the HPA database (Figure 1D, 1F; Supplementary Figure 1) revealed that IRF3 and IRF7 were upregulated in cancer tissues (Figure 1D, 1F). IHC confirmed these results and also that the IRF3 protein was more highly expressed in cancer tissues than that in normal tissues (Figure 1E, 1G).

## Increased IRF3 and IRF7 Expression Worsened the Prognosis of Patients with COAD/READ

IRF3 and IRF7 mRNA-expression levels were significantly correlated with OS ( $P=0.04$  and  $P=0.05$ , respectively; Figure 2A, 2C; Supplementary Figure 2). IHC verified these results at the protein level in 102 patients with recurrent or advanced CRC. IRF3- and IRF7-protein upregulation showed significant negative correlations with OS ( $P=0.026$  and  $0.033$ , respectively, Fig. 2B, D).

## An IRF Risk Model Predicted OS and DFS in Patients with COAD/READ

We compared IRF-expression levels with tumor stages in patients with COAD/READ. IRF1 and IRF6 expression significantly varied (IRF1:  $P < 0.001$ , Figure 3A; IRF6:  $P=0.041$ , Figure 3B), whereas IRF2–5 and IRF7–9 expression did not. IRF6 expression was positively correlated with TNM staging, whereas IRF1 expression was not completely positively correlated. A random-forest model was applied to the GEO dataset, and patients in the TCGA and GEO datasets were divided into high- and low-IRF score groups, based on the median risk score (Figure 3C). Patients in the low-IRF score group showed a better prognosis (TCGA: log-rank  $P < 0.001$ , Figure 3D; GEO: log-rank  $P=0.045$ , Figure 3E).

Univariate and multivariate Cox analysis showed that IRF risk score was an independent risk factor for OS and DFS (Table 2, 3; Figure 3F, 3G). IRF risk scores and clinicopathological features were used to construct a nomogram to predict OS and DFS (Figure 3H, 3J). Based on the C-index, the nomogram showed high discriminability in the TCGA and GEO datasets (OS: TCGA: 0.928 [0.910–0.945]; GEO: 0.610 [0.571–0.649]; DFS: TCGA: 0.940 [0.922–0.958]; GEO: 0.656 [0.616–0.65]). A calibration curve showed good consistency between the nomograms and the recorded 1-, 3-, and 5-year OS and DFS rates (Figure 3I, 3K).

## Relationship Between IRF Scores and Gene-Expression Profiles

Analysis of data for patients in the high and low IRF-score groups identified 126 DEGs ( $|\log_{2}FC| > 1.0$  and  $P < 0.05$ ; Figure 4A, 4B).

GO analysis showed that the DEGs were closely related to BP terms such as gas transport, antimicrobial response, humoral immune response, and sensory organ morphogenesis (Figure 4C, Supplemental Table 2). Differentially expressed IRF genes were associated with enriched KEGG terms such as nitrogen metabolism, JAK-STAT signaling pathway, *Staphylococcus aureus* infection, and cytokine receptor interaction pathways (Figure 4D; Supplemental Table 3).

GSEA showed that the ribosome and cardiac muscle contraction terms were significantly enriched for patients with high IRF scores (Figure 4E; Supplemental Table 4), whereas the terms hematopoietic cell, intestinal immune network for IgA production, and chemokine signaling pathway were significantly underrepresented for patients with high IRF scores (Figure 4E). Figure 4F shows enrichments for the related pathways.

### **IRF-Expression Levels Corresponded with Immune Cell Infiltration**

In patients with COAD/READ, IRF mRNA-expression levels correlated positively, in most cases, with the infiltration levels of different immune cells (Figure 5A, B; Supplemental Figure 2). We also observed positive correlations between IRF3 and IRF7 protein-expression levels and tumor-infiltrating immune cell markers via IHC in 102 patients with CRC. IRF3 expression was positively correlated with CD4 expression, suggesting a correlation with CD4+ T cell-infiltration, whereas IRF7 expression was positively correlated with CD4 and CD68 expression, suggesting correlations with T-cell and macrophage infiltration (Figure 5C, D).

### **Correlations Between IRF Gene-Expression Levels and the Biological Characteristics of Patients with COAD/READ**

Analysis of TCGA and GEO datasets showed that patients in the high-IRF risk group had lower immune and stromal related scores than those in the low-IRF risk group ( $P < 0.05$ ; Figure 6A, 6B). Significant differences in IRF scores were found between patients that benefitted from immune therapy and those that did not ( $P = 0.025$ , Figure 6C), based on TIDE scores. We analyzed the effects of IRF gene-expression levels on sensitivities to different chemotherapeutic drugs or small-molecule inhibitors. In Figure 6D, red font indicates increased drug sensitivity with increased IRF gene-expression levels, and green font indicates negative correlations between drug sensitivity and gene-expression levels. Significant differences in IRF scores were also found between patients with high and low TMBs (Figure 6E). Analysis of TCGA data for IRF gene mutations in patients with COAD/READ showed that the IRF2 gene has the highest mutation rate (Figure 6F).

The STRING database was used to build a PPI network for the DEGs identified in this study (Figure 6G), which was imported into Cytoscape software (Figure 6H). The top eight DEGs were selected from the PPI network as hub genes with CytoHubba plug-ins, using the MCC algorithm (Figure 6I). We also constructed

a ceRNA network based on differentially expressed mRNAs, miRNAs, and lncRNAs in patients with COAD/READ (Figure 6J).

### **Analysis of IRF Gene-Expression Levels in Different Tumors**

The CCLE and TCGA databases were used to analyze mRNA-expression levels of IRFs in different tumor cells (Supplemental Figure 3 and 4, respectively).

## **Discussion**

Differential expression of IRF genes has been reported in many cancers [6], and IRFs play important roles in CRC tumorigenesis and prognosis. However, this study is the first to explore IRF-expression levels, at both the mRNA and protein levels, and the prognostic value, effects on immune cells, and potential molecular pathways of IRFs in CRC.

IRF3 and IRF7 are closely related, and unlike other IRFs, they are considered key for evading innate immune responses to virulence factors [22]; thus, they may play crucial roles in anticancer immunity. IRF3 plays important roles in DNA-damage responses (DDRs) in cancer [23]. During chemotherapy with DDR agents and immunotherapy involving checkpoint blockade, IRF3 expression is upregulated via cGAS–STING pathway activation [24, 25]. IRF3 activation in response to DDR promotes its role in upregulating RAE1 [26], which is the tumor-cell ligand for NKG2D on NK cells. Together, RAE1 and NKG2D stimulate NK cell-effector function. IRF3 overexpression inhibits tumor-cell growth by increasing p53 activity in vitro [27]. Additionally, IRF3 may be involved in STING activity [28]. Increased PD-L1 expression following treatment with DDR inhibitors is mainly IRF3-dependent [25], and tumor-growth inhibition and immune-checkpoint blockade with DDR inhibitors is completely dependent on the cGAS–STING–IRF3 axis. Our current findings further suggest an additional benefit of cGAS–STING–IRF3 axis activation due to increased expression of the CXCL10 and CCL5 chemokines, leading to T-cell tumor infiltration. Previously, we found that IRF3 and IRF7 could mediate the acquisition of new anti-tumor effector functions in macrophages [29]. In the present study, we observed that high IRF3 and IRF7 expression was related to CD4+ T-cell, CD8+ T-cell, B-cell, and macrophage activation, indicating that IRF3 and IRF7 could promote the anticancer effect of immune cells.

Interestingly, among all IRF factors, the mRNA- and protein-expression levels of IRF3 and IRF7 were significantly upregulated in tumor tissues and associated with poor OS in patients with CRC. As IRFs are transcription factors, they may also influence tumor-cell development by regulating the transcription of other oncogenes, although the related mechanisms require further investigation. We further assessed the relationship between IRF risk scores and immune and stromal scores in tumor patients to examine why increased IRF3 and IRF7 expression can promote immune cell recruitment without killing tumors. Using TCGA and GEO datasets, patients in the high-IRF risk group showed lower immune and stromal scores than those in the low-IRF risk group, indicating that the immune cells clustered in the tumor stroma without reaching the tumor cells. This may also explain why CRC is not sensitive to tumor immunotherapy.

In conclusion, we investigated the mRNA and protein levels of IRFs in patients with COAD/READ and their effects on prognosis. IRF3 and IRF7 were significantly upregulated in tumor tissues and associated with poor OS in patients with CRC. Although IRFs can promote immune cell infiltration, the immune and stromal scores showed that infiltrating immune cells mostly stayed in the tumor stroma and did not directly kill the tumor. We also investigated pathways potentially related to IRFs; however, these results require further study and confirmation.

## **Declarations**

### **Funding**

This study was supported by the Outstanding Youth Foundation of Zhongshan Hospital (No. 2019ZSYQ21) and National Natural Science Foundation of China (No. 81900482).

### **Competing interests**

The authors declare that there are no conflicts of interest.

### **Ethics approval and consent to participate**

Ethical approval was obtained from the research ethics committee of Zhongshan Hospital (SK2020-104).

### **Author Contributions**

YJC conceived and designed the experiments; YJC and SNL performed the bioinformatics analysis; LD and TTL performed the immunohistochemical analysis and patient data compilation; XZS followed up the patients; YJC, NPZ and LL wrote and edited the manuscript.

### **Acknowledgements**

The authors would like to express gratitude to the staff of Prof. Xi-Zhong Shen's laboratory for their critical discussion and reading of the manuscript.

### **Availability of data and materials**

All data generated or analyzed during this study are included in this published article and its supplementary information files.

### **Consent for publication**

All authors have approved the manuscript and agree with submission and publication.

## **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 Cancer J Clin.* 2018; 68: 394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; 69: 7-34.
3. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology.* 2015; 149: 1204-1225 e1212.
4. Kroemer G, Galluzzi L, Zitvogel L, Fridman WH. Colorectal cancer: the first neoplasia found to be under immunosurveillance and the last one to respond to immunotherapy? *Oncoimmunology.* 2015; 4: e1058597.
5. Zhang R, Chen K, Peng L, Xiong H. Regulation of T helper cell differentiation by interferon regulatory factor family members. *Immunol Res.* 2012; 54: 169-176.
6. Chen YJ, Li J, Lu N, Shen XZ. Interferon regulatory factors: A key to tumour immunity. *Int Immunopharmacol.* 2017; 49: 1-5.
7. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018; 28: 1747-1756.
8. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018; 24: 1550-1558.
9. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, Lu P, Johnson JC, Schmidt C, Bailey CE, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* 2010; 138: 958-968.
10. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenot D, Ayadi M, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* 2013; 10: e1001453.
11. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods.* 2009; 14: 323-348.
12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550.
13. Asplund A, Edqvist PH, Schwenk JM, Ponten F. Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. *Proteomics.* 2012; 12: 2067-2077.
14. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012; 16: 284-287.
15. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013; 14: 7.
16. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage,

- supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019; 47: D607-D613.
17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13: 2498-2504.
  18. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014; 8 Suppl 4: S11.
  19. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017; 77: e108-e110.
  20. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013; 4: 2612.
  21. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011; 12: 77.
  22. Angeletti M, Hsu WN, Majo N, Moriyama H, Moriyama EN, Zhang L. Adaptations of Interferon Regulatory Factor 3 with Transition from Terrestrial to Aquatic Life. *Sci Rep.* 2020; 10: 4508.
  23. Kim T, Kim TY, Song YH, Min IM, Yim J, Kim TK. Activation of interferon regulatory factor 3 in response to DNA-damaging agents. *J Biol Chem.* 1999; 274: 30686-30689.
  24. Li T, Cheng H, Yuan H, Xu Q, Shu C, Zhang Y, Xu P, Tan J, Rui Y, Li P, Tan X. Antitumor Activity of cGAMP via Stimulation of cGAS-cGAMP-STING-IRF3 Mediated Innate Immune Response. *Sci Rep.* 2016; 6: 19049.
  25. Sen T, Rodriguez BL, Chen L, Corte CMD, Morikawa N, Fujimoto J, Cristea S, Nguyen T, Diao L, Li L, et al. Targeting DNA Damage Response Promotes Antitumor Immunity through STING-Mediated T-cell Activation in Small Cell Lung Cancer. *Cancer Discov.* 2019; 9: 646-661.
  26. Lam AR, Bert NL, Ho SS, Shen YJ, Tang LF, Xiong GM, Croxford JL, Koo CX, Ishii KJ, Akira S, et al. RAE1 ligands for the NKG2D receptor are regulated by STING-dependent DNA sensor pathways in lymphoma. *Cancer Res.* 2014; 74: 2193-2203.
  27. Kim TK, Lee JS, Oh SY, Jin X, Choi YJ, Lee TH, Lee E, Choi YK, You S, Chung YG, et al. Direct transcriptional activation of promyelocytic leukemia protein by IFN regulatory factor 3 induces the p53-dependent growth inhibition of cancer cells. *Cancer Res.* 2007; 67: 11133-11140.
  28. Prabakaran T, Bodda C, Krapp C, Zhang BC, Christensen MH, Sun C, Reinert L, Cai Y, Jensen SB, Skouboe MK, et al. Attenuation of cGAS-STING signaling is mediated by a p62/SQSTM1-dependent autophagy pathway activated by TBK1. *EMBO J.* 2018; 37.
  29. Solis M, Goubau D, Romieu-Mourez R, Genin P, Civas A, Hiscott J. Distinct functions of IRF-3 and IRF-7 in IFN-alpha gene regulation and control of anti-tumor activity in primary macrophages. *Biochem Pharmacol.* 2006; 72: 1469-1476.

## Tables

**Table 1.** GSEA of results for patients with COAD/READ in TCGA

Variable	All patients (n=597)	Low expression (n=298)	High expression (n=299)	P-value
Gender				0.904
Female	277 (46.4%)	139 (46.6%)	138 (46.2%)	
Male	320 (53.6%)	159 (53.4%)	161 (53.8%)	
Age				0.378
<60	170 (28.5%)	80 (26.8%)	90 (30.1%)	
≥60	427 (71.5%)	218 (73.2%)	209 (69.9%)	
Pathologic stage				<0.001***
I	108 (18.1%)	69 (23.1%)	39 (13.1%)	
II	225 (37.7%)	120 (40.3%)	105 (35.1%)	
III	177 (29.6%)	86 (28.9%)	91 (30.4%)	
IV	87 (14.6%)	23 (7.7%)	64 (21.4%)	
T				0.002**
T1	19 (3.2%)	11 (3.7%)	8 (2.7%)	
T2	105 (17.6%)	65 (21.8%)	40 (13.4%)	
T3	408 (68.3%)	201 (67.5%)	207 (69.2%)	
T4	65 (10.9%)	21 (7.0%)	44 (14.7%)	
N				<0.001***
N0	342 (57.3%)	191 (64.1%)	151 (50.6%)	
N1	145 (24.3%)	71 (23.8%)	74 (24.7%)	
N2	110 (18.4%)	36 (12.1%)	74 (24.7%)	
M				<0.001***
M0	453 (75.9%)	249 (83.6%)	204 (68.2%)	
M1	85 (14.2%)	22 (7.4%)	63 (21.1%)	
MX	59 (9.9%)	27 (9.0%)	32 (10.7%)	

Abbreviations: GSEA, Gene-set enrichment analysis; COAD/READ, colon adenocarcinoma/rectum adenocarcinoma

**Table 2.** Univariate and multivariate Cox analysis of OS prediction, based on IRF scores for TCGA patient data

e	Univariate Cox analysis		Multivariate Cox analysis	
	HR (95% CI)	<i>P</i> value	HR (95% CI)	<i>P</i> value
≥60 vs. <60)	1.73 (1.09-2.77)	0.020*	1.42 (0.86-2.33)	0.171
(male vs.	1.08 (0.74-1.56)	0.678	0.69 (0.47-1.04)	0.075
e (T3&T4 vs.	3.08 (1.50-6.33)	0.002**	1.23 (0.59-2.57)	0.584
e (N1&N2 vs.	2.82 (1.93-4.14)	<0.001***	0.64 (0.25-1.62)	0.341
e (M1&MX vs.	2.86 (1.98-4.15)	<0.001***	1.83 (1.18-2.82)	0.006**
III+IV vs. I+II)	3.22 (2.18-4.77)	<0.001***	2.77 (0.99-7.76)	0.052
ore (high vs.	2216.68 (654.36-7509.12)	<0.001***	2401.99 (624.22-9242.82)	<0.001***

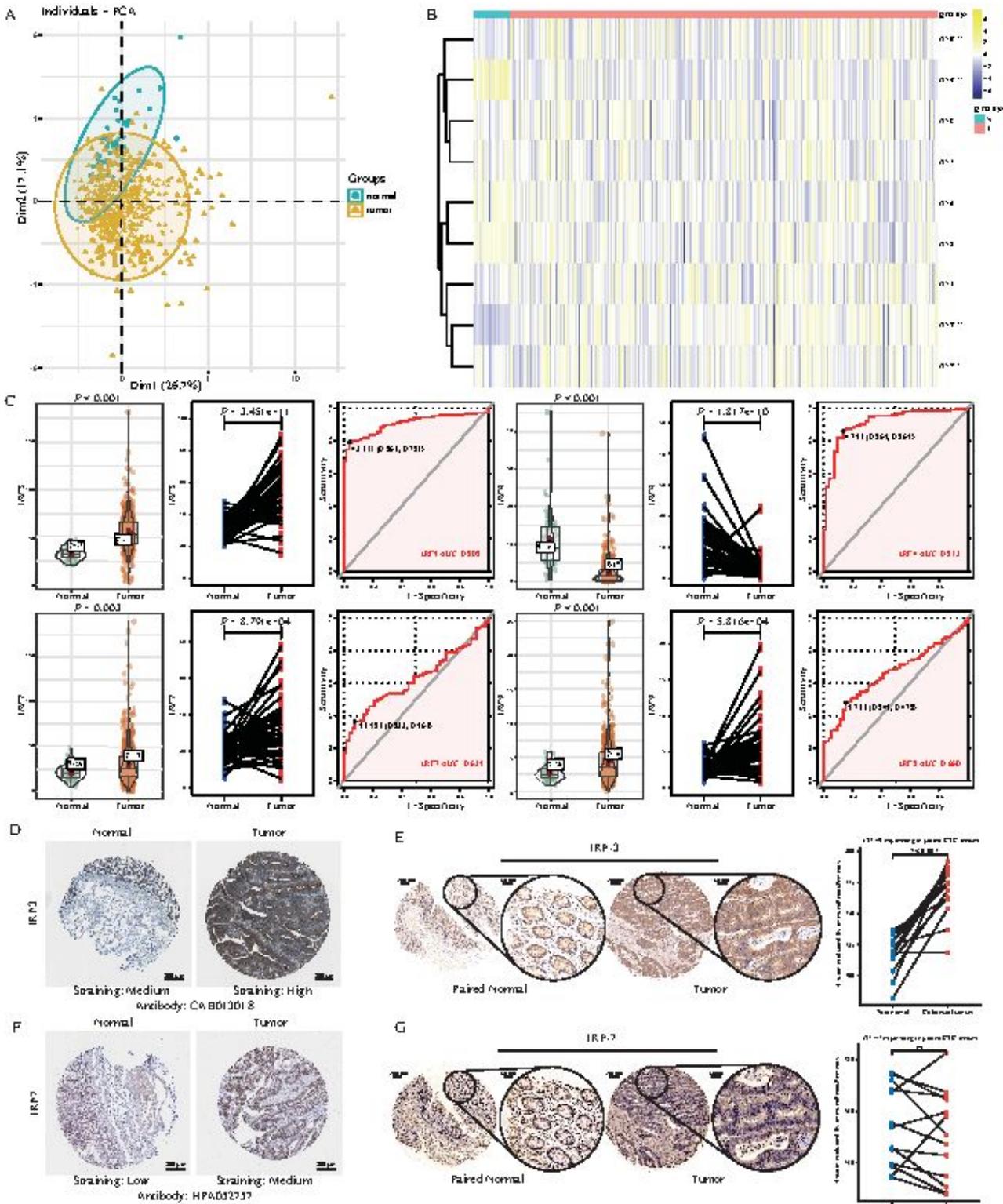
Abbreviations: OS, Overall survival; IRF, Interferon-regulatory factors; TCGA, The Cancer Genome Atlas; CI, Confidence interval; HR, Hazard ratio

**Table 3.** Univariate and multivariate Cox analysis of DFS prediction, based on IRF scores for TCGA data

e	Univariate Cox analysis		Multivariate Cox analysis	
	HR (95% CI)	<i>P</i> -value	HR (95% CI)	<i>P</i> -value
≥60 vs. <60)	1.00 (0.60-1.69)	0.975	1.02 (0.59-1.78)	0.934
(male vs.	1.15 (0.72-1.84)	0.567	0.75 (0.46-1.24)	0.263
e (T3&T4 vs.	8.40 (2.06-34.31)	0.003**	3.41 (0.82-14.17)	0.091
e (N1&N2 vs.	4.77 (2.77-8.24)	<0.001***	0.74 (0.28-1.96)	0.547
e (M1&MX vs.	5.57 (3.45-8.99)	<0.001***	2.99 (1.72-5.18)	0.001**
III+IV vs. I+II)	6.52 (3.57-11.91)	<0.001***	3.32 (1.04-10.61)	0.043*
ore (high vs.	1585.65 (372.03-6758.19)	<0.001***	1855.25 (335.16-10269.55)	<0.001***

Abbreviations: DFS, Disease-free survival; IRF, Interferon-regulatory factors; TCGA, The Cancer Genome Atlas; CI, Confidence interval; HR, Hazard ratio

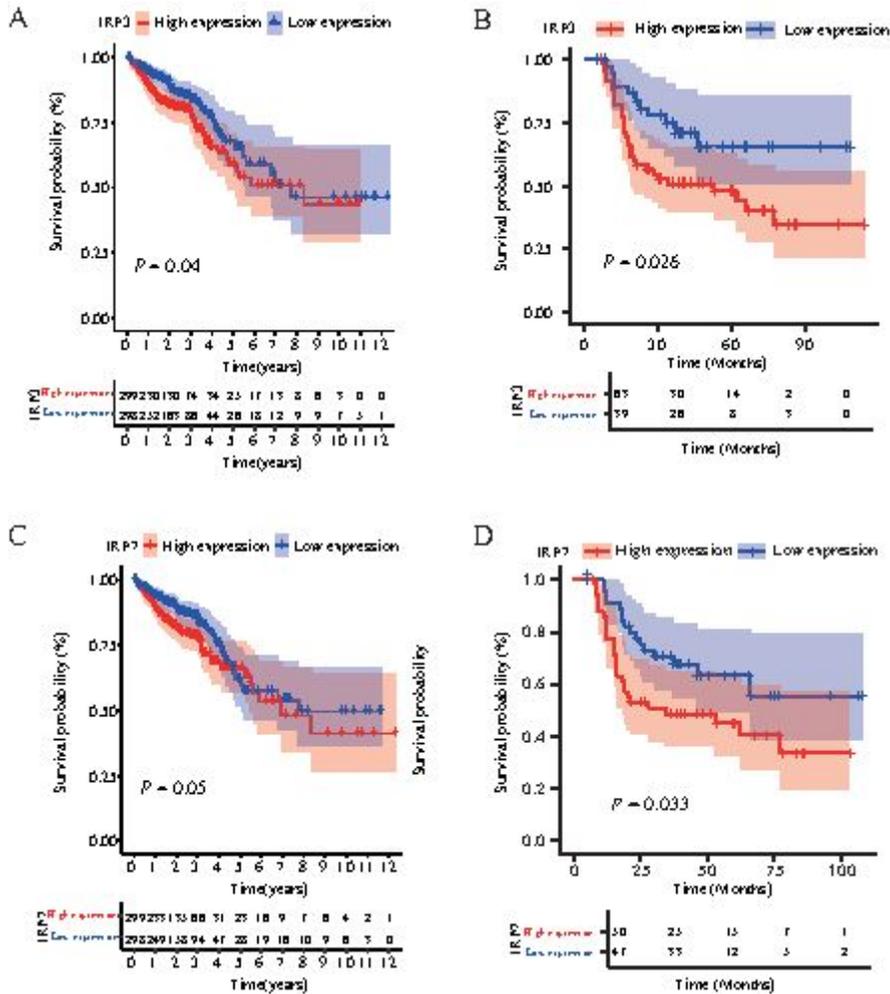
# Figures



**Figure 1**

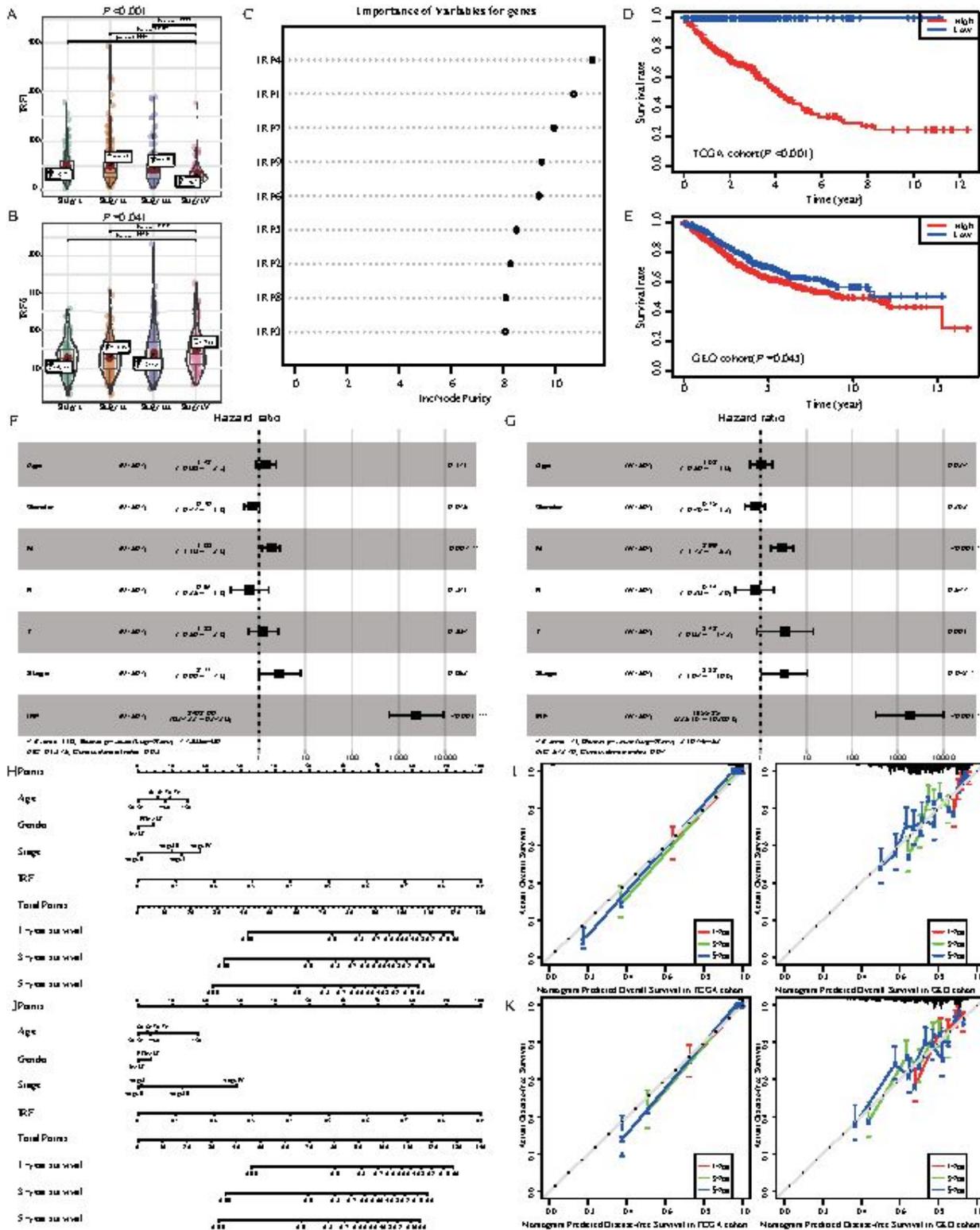
Differential expression of interferon-regulatory factor (IRF) genes in patients with colorectal cancer (CRC). (A) principal-component analysis (PCA) showing that IRF-expression levels differed between normal and tumor tissues in patients with colon adenocarcinoma/rectum adenocarcinoma (COAD/READ). (B) Heat

maps showing IRF-expression levels in cancer and paracancerous normal tissues. (C) Significant differential expression of IRF genes was observed in CRC tissues and paired normal tissues, and receiver operating-characteristic (ROC) curves showed that their expression levels could be used to distinguish between tumor and normal tissues. (D, F) Representative immunohistochemistry (IHC) results from the HPA database for IRF3 and IRF7 in normal intestinal tissue and intestinal cancer tissue are shown. (E, G) IHC analysis of cancer and paracancerous tissues in 12 patients with CRC confirmed the IRF3 and IRF7 protein levels in CRC tissues, revealing that IRF3 was upregulated in CRC tissues, whereas the IRF7 level showed no significant difference between cancer and normal tissues.



**Figure 2**

Prognostic analysis of IRF genes in patients with CRC. (A, C) Survival analysis showed that IRF3 and IRF7 mRNA-expression levels impacted the prognosis. (B, D) IHC showing that IRF3 and IRF7 protein-expression levels were significantly correlated with the overall survival (OS) of 102 patients with CRC.



**Figure 3**

A clinical-prediction model based on IRF gene-expression levels. (A, B) Clinical-correlation analysis showed that IRF1 and IRF6 expression was significantly correlated with different clinical stages (IRF1:  $P < 0.001$ ; IRF6:  $P = 0.041$ ). (C) A random-forest model was constructed based on IRF-expression levels. The weighted values of different genes are shown. (D, E) Survival analysis showing that patients with low-IRF score had a good prognosis with both the The Cancer Genome Atlas (TCGA) and Gene Expression

Omnibus (GEO) datasets. (F) The predictive power of multivariate Cox analysis of IRF scores combined with clinicopathological features for OS. (G) The predictive power of multivariate Cox analysis of IRF scores combined with clinicopathological features for predicting disease-free survival (DFS). (H) A histogram for predicting OS based on IRF scores and clinicopathological features. (I) Calibration curve for an OS nomogram; (J) A histogram for predicting DFS based on IRF scores and clinicopathological features. (K) Calibration curve for a DFS nomogram.

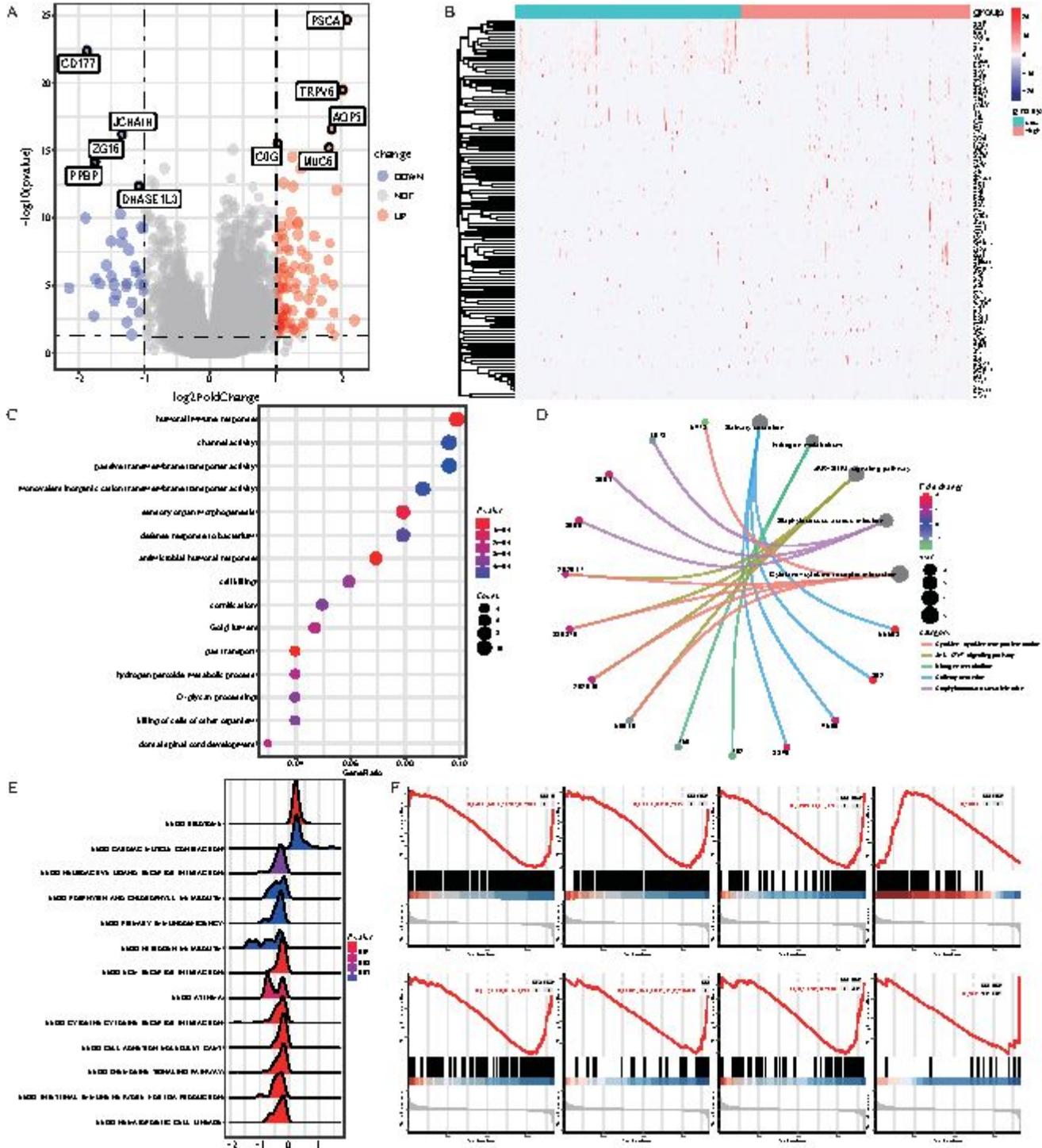
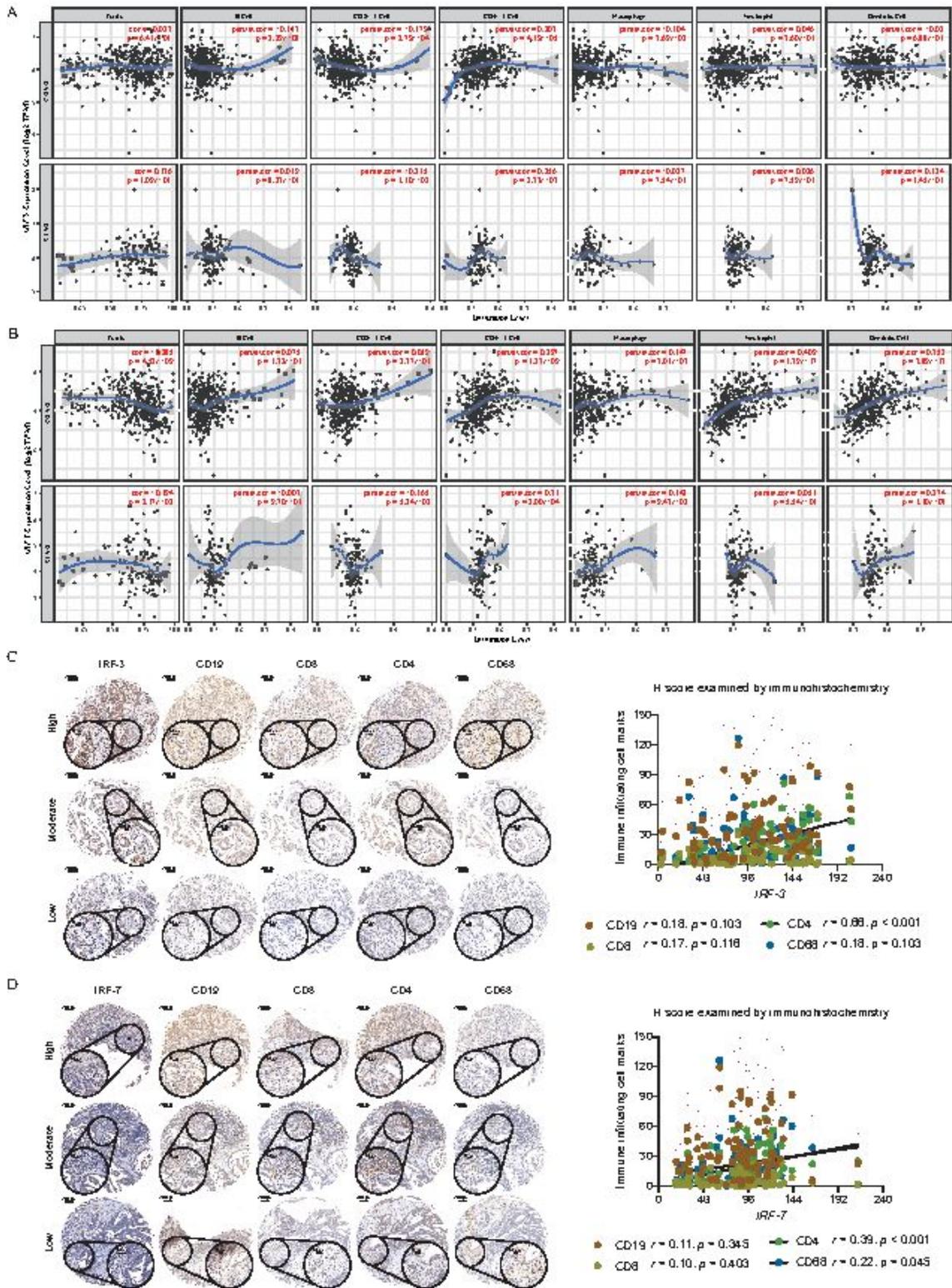


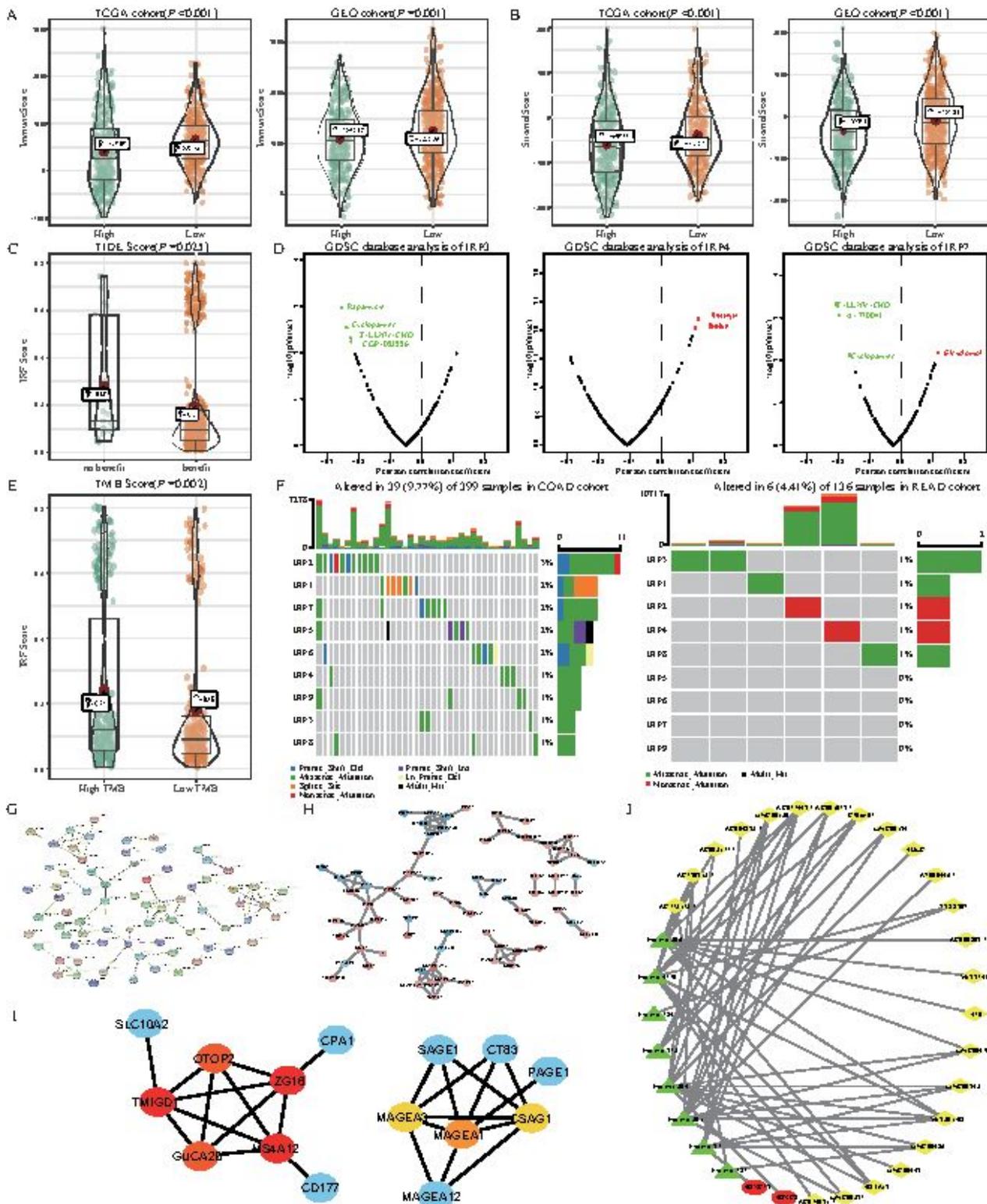
Figure 4

Differentially expressed gene (DEG) and functional-enrichment analysis between high- and low-IRF groups. (A, B) Volcano and heat maps showing DEGs between high- and low-IRF groups. (C) GO analysis suggested that DEGs were closely correlated to the terms gas transport, antimicrobial humoral response, humoral immune response, and sensory organ morphogenesis. (D) Kyoto Encyclopedia of Genes and Genomes analysis showed that these DEGs were enriched for terms such as nitrogen metabolism, the JAK-STAT signaling pathway, Staphylococcus aureus infection, and cytokine-cytokine receptor interaction. (E) Volcanic maps showing gene-set enrichment analysis (GSEA) results for upregulated and downregulated pathways. (F) Patients in the high-IRF group showed correlations with the terms ribosome and cardiac muscle contraction pathways, whereas the terms hematopoietic cell lineage, intestinal immune network for IgA production, and chemokine signaling pathway (among other pathways) were significantly underrepresented for patients in the high-IRF group.



**Figure 5**

Effects of differentially expressed IRFs on immune cell infiltration. (A, B) The correlations between IRF3 and IRF7 mRNA-expression levels and different levels of immune cell infiltration in patients with COAD/READ, analyzed using the tumor immune estimation resource (TIMER) database. (C, D) Relationships between IRF3- and IRF7-protein levels and tumor-infiltrating immune cell markers in 102 patients with CRC, as determined by IHC



**Figure 6**

Effects of IRF-expression levels on different biological phenomena. (A, B) Based on the TCGA and GEO datasets, the immune and stromal scores of patients with CRC in the high-IRF group were significantly lower than those in the low-IRF group ( $P < 0.05$ ). (C) Significant differences in IRF-risk scores were found between the immunotherapy-benefit and non-benefit groups ( $P = 0.025$ ). (D) The GDSC database was used to evaluate correlations between IRF-expression levels and sensitivities to chemotherapeutic drugs. The

green text indicates negative correlations between IRF expression and sensitivity, and the red text indicates positive correlations. (E) Significant differences in IRF scores between the immunotherapy-benefit and non-benefit groups ( $P=0.002$ ). (F) IRF-mutation frequencies in patients with COAD/READ. (G) The Search Tool for Retrieving Interacting Genes (STRING) database was used to analyze a PPI network of DEGs. (H) The STRING results were imported into Cytoscape software. Red text represents upregulated genes and blue text represents downregulated genes, where the color intensity was positively correlated with logFC. (I) The maximum clique centrality algorithm was used to identify core genes in the PPI network, and the red and yellow nodes represented the top eight hub genes. (J) A ceRNA network was constructed based on differentially expressed mRNAs, miRNAs, and lncRNAs, where the yellow diamonds represent lncRNAs, the green triangles represent miRNAs, and the red ovals represent mRNAs.

## Supplementary Files

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

- [SupplementalFig1.eps](#)
- [SupplementalFig2.eps](#)
- [SupplementalFig3.eps](#)
- [SupplementaryFig4.eps](#)