Article

J. Biol. Regul. Homeost. Agents. 2024; 38(9): 5859–5875
https://doi.org/10.23812/j.biol.regul.homeost.agents.20243809.467

High Erythroferrone Expression as an Independent Predictor of Poor Prognosis in Patients with Colorectal Cancer

Anlong Zhu[1,2], Bo Zhang[1,2], Linfeng Yu[1,2], Jiaqi Gao[1,2], Yanwei Xing[1,2], Ye Song[1,2], Chang Yan[2,3], Jian Zhang[1,2,*], Yuekun Zhu[1,2,*]

[1]Department of General Surgery, The First Affiliated Hospital of Harbin Medical University, 150001 Harbin, Heilongjiang, China
[2]Key Laboratory of Hepatosplenic Surgery, Ministry of Education, The First Affiliated Hospital of Harbin Medical University, 150001 Harbin, Heilongjiang, China
[3]Department of Anesthesiology, The First Affiliated Hospital of Harbin Medical University, 150001 Harbin, Heilongjiang, China
*Correspondence: zhangjian0737@sina.com (Jian Zhang); zhuyuekun2011@126.com (Yuekun Zhu)
Submitted: 4 December 2022 Revised: 28 December 2022 Accepted: 23 January 2023 Published: 30 August 2024

Background: Colorectal cancer, an invasive tumor originating in the mucosal lining of the large intestine and rectum, represents a prevalent form of gastrointestinal malignancy. Although extensive investigations have been conducted on colorectal cancer, the precise molecular mechanisms underlying this neoplasm remain uncharacterized. Alterations in the genetic material have been linked to colorectal cancer progression. To analyze the prognosis of individuals with colorectal cancer, it is crucial to identify new biomarkers. Using integrated bioinformatics analysis, this study successfully identified and confirmed the fundamental gene linked to colorectal cancer.

Methods: Bioinformatics software tools were employed to assess the mRNA expression level of Erythroferrone (*ERFE*) by analyzing the Cancer Genome Atlas (TCGA) dataset, which included 647 tumor samples and 51 control samples. In order to verify the findings, a comparison was conducted with data available on the Gene Expression Omnibus (GEO). In order to determine the clinical significance and expression level of *ERFE*, a bioinformatics investigation was carried out using logistic regression analysis. To assess survival rates in both high- and low-expression *ERFE* groups, univariate and multivariate Cox proportional hazards model (Cox) regression analyses were conducted. To validate the expression of *ERFE* at both the gene and protein levels in colorectal cancer (CRC) cells (HCT116 and Lovo) and normal epithelial cells, reverse transcription quantitative polymerase chain reaction and western blot experiments were performed. The knock-down efficiency of *ERFE* in HCT116 and LoVo cell lines was assessed using western blot experiments. The effect of *ERFE* gene function was compared and analyzed through various assays such as cell counting kit-8 (CCK-8), Transwell, and scratch tests before and after *ERFE* knock-down in HCT116 and LoVo cell lines.

Results: An analysis of data obtained from TCGA and GEO databases unveiled a remarkable increase in the expression of *ERFE* mRNA in CRC tissue compared to both normal and paracancerous tissues ($p < 0.001$). Furthermore, it was discovered that the overexpression of *ERFE* was linked to an unfavorable prognosis and had the potential to act as an independent prognostic indicator for predicting overall survival (OS), progression-free interval (PFI), and disease-specific survival (DSS) among colorectal cancer patients. Additionally, a positive correlation was found between heightened *ERFE* expression, infiltration of numerous immune cells, and levels of immune checkpoint molecules. Experimental evaluations, including the CCK-8 assay, scratch assay, and transwell assay, provided compelling evidence demonstrating a significant reduction in the proliferative, migratory, and invasive capabilities of colorectal cancer cells upon knocking down *ERFE*.

Conclusions: In summary, *ERFE*, functioning as a pro-oncogene in CRC, is linked to the initiation and advancement of cancer, and can serve as a standalone marker for unfavorable prognosis among CRC patients.

Keywords: Erythroferrone; colorectal cancer; prognosis; the Cancer Genome Atlas; biomarkers

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

Introduction

Colorectal cancer is shortened to CRC, is a highly prevalent form of cancer worldwide, ranking as the third most common type. Each year, over 1.4 million people are diagnosed with CRC, and the number of deaths exceeds 700,000 [1]. In the United States (US), CRC is the third leading cause of cancer-related mortality for both men and women, surpassing the incidence rates in other nations. Projections indicate a startling 60% surge in CRC cases by 2030 [2,3]. The initial line of treatment for CRC, including chemotherapy and radiotherapy, often leads to dose-limiting side effects and the emergence of drug-resistant cancer cells. Consequently, it is imperative to develop additional targeted therapies to enhance patient prognosis and elevate their quality of life [4]. Despite the prevalence of CRC, there exist only a limited number of molecular techniques exclusively designed for its detection. The role of molecular biomarkers is becoming an increasingly important one in diagnosing, treating, as well as predicting the prognosis of a disease. The low availability of biomarkers for CRC is a cause for concern. Hence, molecular biomarkers with a non-invasive nature and high sensitivity and specificity to CRC need to be fabricated [5].

Erythrocytes manufacture a hormone known as Erythroferrone (*ERFE*) in response to hypoxia, hemorrhage, or other pre-erythrocytic stimuli. The erythroid cells in the bone marrow and spleen tissues release *ERFE* through stimulation by erythropoietin (EPO). In the liver cells, *ERFE* performs as an inhibitory agent in the transcription machinery of hepcidin, an important hormone that controls the regulation of iron levels in the body. To date, most studies related to *ERFE* have concentrated on the physiological role of *ERFE* in regulating iron homeostasis and its role in the pathogenesis of iron-related disorders [6]. Persistent secretion by erythroid cells causes high levels of *ERFE* to attenuate the hepcidin's response to the iron load, leading to iron overload in the system, as evidenced by the high iron concentrations and storage in the plasma and liver tissue, respectively [7]. Epidemiological data suggest that iron levels are associated with CRC risk. Red meat contains high amounts of heme iron. Individuals who consume higher amounts of red meat and those with iron overload have an increased risk of CRC [8,9]. Although the association of *ERFE* with blood disorders has been extensively studied, the significance of *ERFE* in the prognostic or therapeutic aspects of CRC remains less studied. The differential expression of *ERFE* in pan-cancer was observed in the results of this study, revealing its crucial role in promoting CRC development. In addition, the predictive and immune-related roles of *ERFE* in CRC were investigated using publicly available databases. These analyses were primarily based on bioinformatics analysis of the Cancer Genome Atlas (TCGA), an extensive project in cancer genomics that covers more than 20,000 primary cancer tissues, along with their corresponding healthy tissues, across 33 different cancer categories at a molecular level. The creation of this cancer genomics program and the growing accessibility of high-throughput sequencing data indicate that cancer-related genomic datasets are being shared more frequently nowadays [10,11]. Thanks to the advancements in bioinformatics, we now have a powerful tool at our disposal that greatly aids in understanding the underlying causes and pathogenesis of colon cancer. Moreover, this technology enables us to identify previously undiscovered biomarkers that hold significant prognostic value.

Material and Methods

Data Download

The TCGA0-Colon adenocarcinoma/Rectum adenocarcinoma (COADREAD) (https://portal.gdc.cancer.gov/) project contains RNA sequencing (RNA-seq) data and clinical information for 51 paracancerous and 647 tumor tissues. The TCGA's COADREAD was utilized to obtain the data of the unpaired samples (cancerous samples), and the data of the corresponding normal tissue was accessed at the Genotype-Tissue Expression (GTEx) (https://gtexportal.org/home/datasets). The gene expression information of unpaired samples (pan-cancer) contained normal tissue, paracancerous tissue, and tumor data of pan-cancer. The paired samples were derived from RNA-seq data from TCGA. Additionally, RNA-seq information and clinical data were accessed at the University Of Cingifornia Sisha Cruz (UCSC) Xena link (http://www.genome.ucsc.edu/index.html). The dataset numbered Gene Expression Omnibus Series (GSE) 4107 was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo). The GeneChip U133-Plus 2.0 array (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570) was used to analyze the RNA information extracted from mucosal tissue of the colon of healthy controls (10 samples) and CRC patients (12 samples), as well as blood samples from the dataset numbered GSE164191 of 59 CRC patients and 62 healthy controls. Subsequently, the whole blood gene expression profiling was performed using Affymetrix microarray hybridization. First, CRC queues with survival time and survival status: GSE72969, GSE87211, GSE17536 were downloaded to verify the prognostic performance of the *ERFE* gene. Then, immunotherapeutic-related cohort GSE179351 was downloaded. There were three groups of CRC patients in this cohort, which were before-treatment, Immune Checkpoints Blockade (ICB) treatment (ICB-treatment), and Immune Checkpoints Blockade treatment combined with radiation treatment (ICB + radiation).

Differential Expression Analysis

Perform log2 conversion of Transcripts Per Million (TPM) format RNA seq data. R package all downloaded from https://www.bioconductor.org/ website. First of all,

ggplot2 3.3.3 (https://www.bioconductor.org/help/search/index.html?search-bar=ggplot2/) is used for statistical analysis and visualization of differentially expressed genes.

Receiver Operating Characteristic Curves and Survival Analysis

To visualize the information, we utilized the R package (version 3.6.3) (https://mirrors.tuna.tsinghua.edu.cn/CRAN/) and the survminer package (version 0.4.9) (https://cran.r-project.org/). Furthermore, we performed statistical analysis on the survival data using the survival package (version 3.2.10) (https://mirrors.tuna.tsinghua.edu.cn/CRAN/) (install.packages ("survival")) to generate plots depicting progression-free interval (PFI), overall survival (OS), and disease-specific survival (DSS) related to *ERFE*. To classify the cohort into groups based on expression levels, we employed the expression median as the chosen threshold. To assess the relationship between clinical information and *ERFE* expression, logistic regression analysis was conducted. We utilized both univariate and multivariate Cox proportional hazards model (Cox) analyses to identify independent prognostic variables. In order to evaluate the predictive prognostic value of *ERFE* in CRC patients, we analyzed the data and visualized the Receiver Operating Characteristic (ROC) curves. This analysis was carried out using the pROC package (version 1.17.0.1) (https://mirrors.tuna.tsinghua.edu.cn/CRAN/) (install.packages ("pROC")) and "ggplot2" package (version 3.3.3) (https://www.bioconductor.org/help/search/index.html?search-bar=ggplot2/). By using the median expression level of the *ERFE* gene, we divided CRC patients into high and low expression groups. Initially, Kaplan-Meier (KM) survival analysis was conducted in three validation cohorts using the survival package, and the resulting survival curves were plotted accordingly. Subsequently, survivalROC analysis was performed to assess 1, 3, and 5-year survival rates using the survivalROC package (version 1.0.3.1) (https://mirrors.tuna.tsinghua.edu.cn/CRAN/) (install.packages ("survivalROC")). Finally, the results were presented in line plots.

GeneMANIA Database

Protein-Protein Interaction (PPI) network analysis was performed through GeneMANIA (http://genemania.org/), and the proteins in the network were subjected to functional annotation.

Functional Enrichment Analysis

To thoroughly examine and describe diverse gene functionalities, encompassing biological processes (BP), cellular components (CC) and molecular functions (MF), an extensive investigation was carried out using Gene Ontology (GO) for functional enrichment analysis. Subsequently, intricate interactions and network relationships at the molecular level were observed through the utilization of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. In order to establish connections, a two-by-two correlation analysis of *ERFE* gene expression levels and other molecular constituents was performed using the STAT package (version 3.6.3) (https://mirrors.tuna.tsinghua.edu.cn/CRAN/) (install.packages ("STAT")). For screening purposes, a threshold of pSpearman <0.05 was employed, resulting in the selection of the top 200 genes based on coSpearman values in descending order. Additionally, to facilitate ID conversion, the "org.Hs.eg.db" package (version 3.10.0) (https://www.bioconductor.org/help/search/index.html?search-bar=org.Hs.eg.db/) in R was utilized. Enrichment analysis was conducted using the "clusterProfiler" package (version 3.14.3) (https://www.bioconductor.org/help/search/index.html?search-bar=clusterProfiler/), while the filtering and visualization of BP, CC, MF, and KEGG pathways that met the adjusted p-value < 0.1 and q-value < 0.2 criteria were accomplished using ggplot2 (version 3.3.3) (https://www.bioconductor.org/help/search/index.html?search-bar=ggplot2/).

Immune Cell Infiltration

Based on mRNA expression profile data, the single-sample gene set enrichment analysis (ssGSEA) algorithm was utilized to compute enrichment scores for 24 immune cells in the tumor microenvironment (TME). This algorithm was employed to assess the levels of infiltration by 24 immune cells in CRC samples from the TCGA database. Furthermore, the Estimation algorithm estimated the immune infiltration score, stromal score, and estimated score in the TME using mRNA expression. Higher scores implied greater presence of stromal components and immune infiltration in the internal environment. In addition, the correlation of expression trends between *ERFE* genes and immune cell markers was visualized by a heat map utilizing the R package "ggplot2" 3.3.3.

Susceptibility to Immunotherapy

RNA-seq data (level 3) and relevant clinical data for colorectal cancer (CRC) were gathered from the ICGC website (https://docs.icgc-argo.org/docs/data-access/icgc-25k-data). The subsequent eight genes, namely *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *PDCD1LG2*, *TIGIT*, and *SIGLEC15*, are linked to immune checkpoints. Consequently, an analysis was performed to examine the expression of these genes concerning immunological checkpoints. To achieve this, a uniformly normalized dataset was acquired from the UCSC website (https://xenabrowser.net/). The expression profile of *ERFE* genes from COADREAD samples was then extracted, with emphasis on samples originating from Primary Blood-Derived Cancer-Peripheral Blood and Primary Tumors. Furthermore, the Simple Nucleotide Variation dataset (level 4) of TCGA-COADREAD, processed using MuTect2 software, was accessed from the GDC website (https://portal.gdc.can

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

cer.gov/). The Tumor Mutation Burden (TMB) score was calculated using the "tmb" function found in the R package maftools (version 2.8.05) (https://www.bioconductor.org/help/search/index.html?search-bar=+maftools/) for a more robust analysis. To enhance the analysis further, the TMB, microsatellite instability (MSI) scores, neoantigen (NEO) data, and gene expression data from previous COADREAD studies were integrated. To compare TMB, MSI, and NEO scores, the *ERFE* gene expression data was divided in accordance with the median value into high and low groups, and R package ggplot2 3.3.3 was utilized for its visualization. To check for plagiarism in the papers, the Genomics of Drug Sensitivity in Cancer (GDSC) website (https://www.cancerrxgene.org/) was utilized for each TCGA-COADREAD sample. The prediction of chemotherapy response to 5-Fluorouracil (5-FU) was conducted using the pRRophetic R package (version 0.5) (https://mirrors.tuna.tsinghua.edu.cn/CRAN/) (install.packages ("pRRophetic")). The ridge regression method was employed to estimate the sample's half-maximum inhibitory concentration (IC_{50}), with all parameters set at their default values (e.g., batch effect: COMBAT; tissue type: ALL). Duplicate gene expression was represented by calculating the mean. Initially, the pheatmap package was used to perform hierarchical clustering of *ERFE* and ICB-related genes in the GSE179351 cohort, which is associated with immunotherapy. The expression levels were standardized using the z-score method and presented visually as a heatmap. Next, the Pearson method was utilized to calculate the transcriptional-level correlation between *ERFE* and ICB-related genes at various treatment stages. The results were then visualized using scatter plots.

Construction of the Risk Prediction Model

The risk prediction model was developed based on TCGA data. We developed a nomogram incorporating gene expression and clinical information to predict 1-, 3-, and 5-years OS for people suffering from CRC. For risk assessment, the status of each clinical factor and the expression level of *ERFE* were first determined. Then, the associated survival rates over 1, 3, and 5 years were obtained by a straight line drawn on the risk axis. Subsequently, the fit between the actual and model-predicted survival rates under different scenarios was visualized in a calibration plot to determine the model's predictive accuracy. In the end, the validation of the model's predictive capability involved the illustration of *ERFE* risk factors based on the risk score. This was accomplished by plotting time-dependent ROC curves for 1-, 3-, and 5-year durations.

Cell Line Screening

The mRNA expression levels of *ERFE* genes in various CRC cell lines were investigated utilizing the website Cancer Cell Line Encyclopedia (CCLE) (https://sites.broadinstitute.org/ccle) (**Supplementary Fig. 1**). Two cell lines showing increased *ERFE* values were chosen for this study HCT116 (FH0027, Fuheng Bio, Shanghai, China) and LoVo cells (FH0023, Fuheng Bio, Shanghai, China). In addition, the control cells used in this experiment were human FHC cells (FH1283, Fuheng Bio, Shanghai, China). Cells used in this study was authenticated by STR and no cross-contamination between cells was identified through mycoplasma testing.

RNA Extraction and Quantitative Reverse Transcription-Polymerase Chain Reaction

The two selected untreated cell lines (HCT116 and LoVo) were transfected with the empty plasmid normal control (NC) and siRNA-*ERFE* (si-*ERFE*). Afterward, the RNA level of these cells was analyzed using Quantitative real time polymerase chain reaction (qRT-PCR) with the internal reference glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). The experimental protocol provided by the manufacturer's instructions was followed for using the kit. The qRT-PCR parameters used were as follows: 2 µL of RNA (template), 12.5 µL of SYBR® PrimeScript master mix (2x, RR036A, Takara, Tokyo, Japan), 0.5 µL of forward primer (20 µM), 0.5 µL of reverse primer (20 µM), 11.5 µL of ddH2O, and a total volume of 25 µL. *GAPDH* served as the internal control. The qRT-PCR was conducted with the following parameters: 45 °C for 15 min, 95 °C for 5 min, 95 °C for 20 s, 60 °C for 20 s, and 72 °C for 30 s (40 cycles in total). The following primers were used: upstream CCGGAGCCAGGGTTGATTC and downstream GCACTCCATGAGAACATGAAGAG for human *ERFE*, and upstream ACAAGCCTCAAGATCATCAGC and downstream GCCATCAGCCAAGTTTCC for human *GAPDH*. Data analysis was performed using the $2^{-\Delta\Delta Ct}$ method.

Small Interfering RNA Transfection

Guangzhou RiboBio Co., Ltd. (Guangzhou, China) designed and produced the small interfering RNA (siRNA) to target *ERFE*. Following the manufacturer's protocol, Lipofectamine 2000 (cat. no. 11668027, Thermo Fisher Scientific, Waltham, MA, USA) was utilized for transfection of cells. The siRNA sequences were as follows: si-*ERFE*: GGAGCACAGAUCUAGACAATT, si-NC: UUCUCCGAACGUGUCACGUTT. The transfection efficiency of gene silencing was verified by western blot assays.

Western Blot

The empty plasmid NC and si-*ERFE* were used to transfect the two selected untreated cell lines (HCT116, and LoVo). Subsequently, western blot (WB) assays were performed on these cells at the protein level. The internal reference used for this study was the β-Tubulin. We employed WB techniques following standard procedures to extract and analyze the total proteins. Gel electrophoresis tech-

nique, specifically utilizing a 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis, was applied to separate the proteins. Subsequently, the proteins were transferred onto a nitrocellulose membrane (cat. no. HATF00010, Millipore, Billerica, MA, USA). The Tris buffered saline Tween (TBST) solution containing 5% skim milk was used to seal the nitrocellulose membrane (cat. no. 28360, Thermo Fisher Scientific, Inc., Waltham, MA, USA) for a duration of 2 hours. Subsequently, the monoclonal primary antibody was introduced to the nitrocellulose membranes. Following this, the nitrocellulose membranes were placed in a refrigerator at 4 °C overnight. The specific antibodies utilized were anti-ERFE (Polyclonal; cat. no. YN3093; 1:1000; ImmunoWay Biotechnology; Plano, TX, USA) and anti-β-Tubulin (Monoclonal; cat. no. YM3030; 1:2000; ImmunoWay Biotechnolog, Plano, TX, USA). Each membrane was then exposed to either peroxidase-conjugated goat anti-rabbit Immunoglobin G (IgG) (cat. no. ZB-2301; 1:10,000; ZSGB-BIO; Beijing; China) or peroxidase-conjugated goat anti-mouse IgG (cat. no. ZB-2305; ZSGB-BIO; Beijing; China), depending on the specific secondary antibody. This incubation process lasted for one and a half hours at room temperature. Upon immersion of the nitrocellulose membranes in an enhanced chemiluminescence reagent, specific bands corresponding to the target proteins became visible within one to three minutes in a dark room. The protein signals were detected using a Bio-Rad gel imaging system (GelDoc Go 12009077, Bio-Rad Laboratories, Inc., Hercules, CA, USA) equipped with enhanced chemiluminescence detection technology. Image Lab 2.0.1 software (Bio-Rad Laboratories) (https://www.bio-rad.com) was utilized to analyze the generated blots.

Cell Viability

The presence of viable cells was determined by assessing cell proliferation using cell counting kit-8 (CCK-8) (cat. no. HY-K0301, MedChemExpress, Monmouth Junction, NJ, USA). Two cell lines, namely HCT116 and LoVo cells, were seeded in 96-well plates at a density of 1500 cells per well. To investigate the effect of *interference*, si-*ERFE* and NC were added for different durations (0, 24, 48, and 72 hours). Following the designated time intervals, the supernatant was discarded, and the mixture was incubated at 37 °C for 4 hours. Subsequently, 100 µL of Dulbecco's Modified Eagle's Medium (DMEM) (cat. no. D5796, Thermo Fisher Scientific, Inc., Waltham, MA, USA) containing 10 µL of CCK-8 was added to the mixture. The absorbance at 450 nm was then measured.

Scratch Assay

The scratch assay samples were prepared using the following procedure. To achieve an 80% single-cell layer, 5×10^4 cells were seeded per well in a 6-well plate. After reaching the desired cell density, a scratch was made in the single-cell layer using a 100 µL plastic pipette tip. Subsequently, the well plates were incubated at 37 °C with 5% carbon dioxide and left to incubate. At both 0 and 48 hours, the morphology of the scratch was captured using a microscope (DP73; Olympus Corporation, Tokyo, Japan) and photographed.

Transwell Assay

Transwell chambers, purchased from BD Biosciences (cat. no. 353097, Franklin Lakes, NJ, USA), were placed in 24-well plates. Dilution substrate gel (1:8, cat. no. 356234, BD Biosciences, Franklin Lakes, NJ, USA) of 60 µL was added and incubated for a duration of 4–5 hours. Following the solidification of the gel in the upper incubator, any remaining liquid was removed, and 100 µL of serum-free DMEM was added. The sample was then incubated once more and hydrated for 20 minutes. To prepare a cell suspension, the cells were digested and resuspended in serum-free DMEM, and a hemocytometer was used to count the cells. The upper chamber was inoculated with the aforementioned 1×10^4 cells, while the lower chamber was inoculated with a mixture of 10% Fetal Bovine Serum (FBS) and DMEM, which were then incubated at a temperature of 37 °C with 5% carbon dioxide. After 24 hours had elapsed, the medium was aspirated from the plates, and a fixative agent consisting of 4% paraformaldehyde (cat. no. BL539A, Biosharp Co., Ltd., Hefei, China) was added for a duration of 15 minutes. The cells were stained with crystal violet dye (cat. no. C0121-100 mL, Beyotime Institute of Biotechnology, Shanghai, China) for 20 minutes, and any remaining solution was washed away using Phosphate Buffer Saline (PBS) multiple times. A microscope was utilized to capture images of the cells.

Statistical Analysis

The analysis of gene differential expression between two subgroups was conducted using the Wilcoxon rank-sum test. Categorical data were examined using Fisher's test. Moreover, Spearman's correlation analysis was employed for conducting correlation analyses. The assessment of survival was performed through Kaplan-Meier survival analysis and tested using the log-rank method. All statistical analysis was conducted utilizing R 3.6.3, and statistically significant differences were considered at a p value less than 0.05 (**, $p < 0.01$; ***, $p < 0.001$).

Results

Differential Expression of ERFE between Pan-Cancer and CRC

The mRNA expression profile of *ERFE* in paracancerous tissues showed a significant decrease in unpaired samples compared to the CRC tumor group ($p < 0.001$, Fig. 1A). Likewise, both normal and paracancerous tissues exhibited a significant reduction in *ERFE* mRNA expression compared to the tumor group ($p < 0.001$, Fig. 1B).

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

Analysis of paired samples indicated a notable decrease in gene expression in the paracancerous group compared to the CRC tumor group ($p < 0.001$, Fig. 1C). Pan-cancer analysis illustrated distinct variations in *ERFE* mRNA expression among different tumor types, including Colon adenocarcinoma (COAD) and Rectum adenocarcinoma (READ), as well as adjacent tissues (or GTEx). These findings suggest that *ERFE* holds promise as a cancer marker (Fig. 1D–G). Specifically, it can be seen that there are differences of *ERFE* mRNA expression in Fig. 1D among BLCA, BRCA, CESC, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, OV, PCPG, READ, STAD, THCA and UCEC. In Fig. 1E, *ERFE* mRNA is differentially expressed in ACC, BLCA, BRCA, CESC, CHOL, COAD, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LIHC, LUAD, LUSC, OV, PAAD, PCPG, READ, STAD, TGCT, THCA, UCEC and UCS. In Fig. 1F, ERFE is differentially expressed in BLCA, BRCA, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, READ, STAD, THCA and UCEC. ERFE is differentially expressed in BLCA, BRCA, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, READ, STAD, THCA and UCEC in G-diagram.

High ERFE Expression Indicates Poor Prognosis of Patients with CRC

The assessment of the diagnostic significance of *ERFE* gene expression in CRC was determined by analyzing ROC curves. The resulting ROC curve Area Under Curve (AUC) was 0.933, as depicted in Fig. 2A. Through the examination of datasets GSE4107 and GSE164191, it was observed that individuals diagnosed with CRC displayed significantly higher levels of *ERFE* mRNA expression compared to the general population, as shown in Fig. 2B,C. The patients in these datasets were then categorized based on the median *ERFE* expression, resulting in the formation of low-expression and high-expression groups. Furthermore, survival curves were generated using the Kaplan-Meier method to identify any variations between these groups using the log-rank test. The findings revealed a strong association between increased levels of *ERFE* expression and an unfavorable prognosis for the patients. Moreover, these patients experienced considerably shorter overall survival (OS), disease-specific survival (DSS), and progression-free interval (PFI), as illustrated in Fig. 2D–F. Specifically, patients at stages T3 and T4 who displayed high *ERFE* expression had a significantly worse prognosis compared to those with low expression levels. Similarly, these patients also exhibited significantly shorter OS, DSS, and PFI, as depicted in Fig. 2G–I. Through the examination of clinical data from 644 patients (as shown in **Supplementary Table 1**), a deeper understanding of the role of *ERFE* expression in the progression of CRC was achieved. These patients were categorized into high-expression (n = 322) and low-expression (n = 322) groups based on their levels of *ERFE* expression. The analysis of clinical data using statistical methods revealed a significant correlation between elevated *ERFE* levels and advanced age, diverse treatment outcomes, and a tendency to be associated with high T-stage. However, there were no notable differences in other clinical characteristics (**Supplementary Table 2**). Univariate analysis demonstrated a connection between *ERFE* expression and T-stage as well as advanced age (**Supplementary Table 3**). To assess the risk ratios of various variables in CRC patients, Cox regression models were utilized for both univariate and multivariate calculations. These analyses highlighted that the expression profile of *ERFE* acted as an independent factor impacting the prognosis of CRC patients. Moreover, the association between *ERFE* expression level and DSS, OS, and PFI in these patients indicated that *ERFE* expression served as a reliable prognostic indicator for individuals with CRC (**Supplementary Tables 4,5,6**). The prognostic capabilities of the *ERFE* gene were further validated using the GEO database. KM survival analysis yielded results showing a significantly lower survival probability in the high *ERFE* expression group compared to the low *ERFE* expression group in the three validation cohorts. The ROC curve indicated that the AUC value for 1-year survival time performance exceeded that for 3-year and 5-year survival time performance (**Supplementary Fig. 2**).

Nomogram Constructed Based on ERFE and Clinicopathological Variables

ERFE and clinicopathological variables were taken into account when fabricating a nomogram, which can be used to obtain the total score and estimate the 1-, 3-, and 5-year survival rate of patients with CRC, making the prediction method more intuitive (Fig. 3A). Furthermore, 1-, 3-, and 5-year calibration curves illustrated the satisfactory performance of the constructed nomogram, with a C-index of 0.789 (Fig. 3B). Additionally, the 1-, 3-, and 5-years AUCs were 0.787, 0.858, and 0.733, respectively, indicating a better and more accurate prediction ability (Fig. 3C). Moreover, the distribution of *ERFE* expression, the survival status of affected individuals, and the risk scores of the high- and low-*ERFE* expression groups are shown in Fig. 3D.

GO and KEGG Analyses of Proteins and Related Genes Interacting with ERFE

ERFE forms a complex network with a variety of proteins. To identify these proteins, a PPI network of *ERFE* proteins was generated using an online web analysis tool. Additionally, we explored 20 genes that may interact with *ERFE*, including *TMEM25*, *GAL*, *STC1*, *LRP1*, *EPHA6*, *DUSP13*, *NPPB*, *DNAJA4*, *PKNOX1*, *C1QL4*, *CBLN4*, *CBLN2*, *CLCN5*, *C1QTNF12*, *C1QTNF4*, *TGFB1*, *CBLN3*, *WDR60*, *CBLN1*, and *ZNF91*, which were primarily associated with cell junction maintenance, regulation of gluconeogenesis, the positive regulation of cell-based response

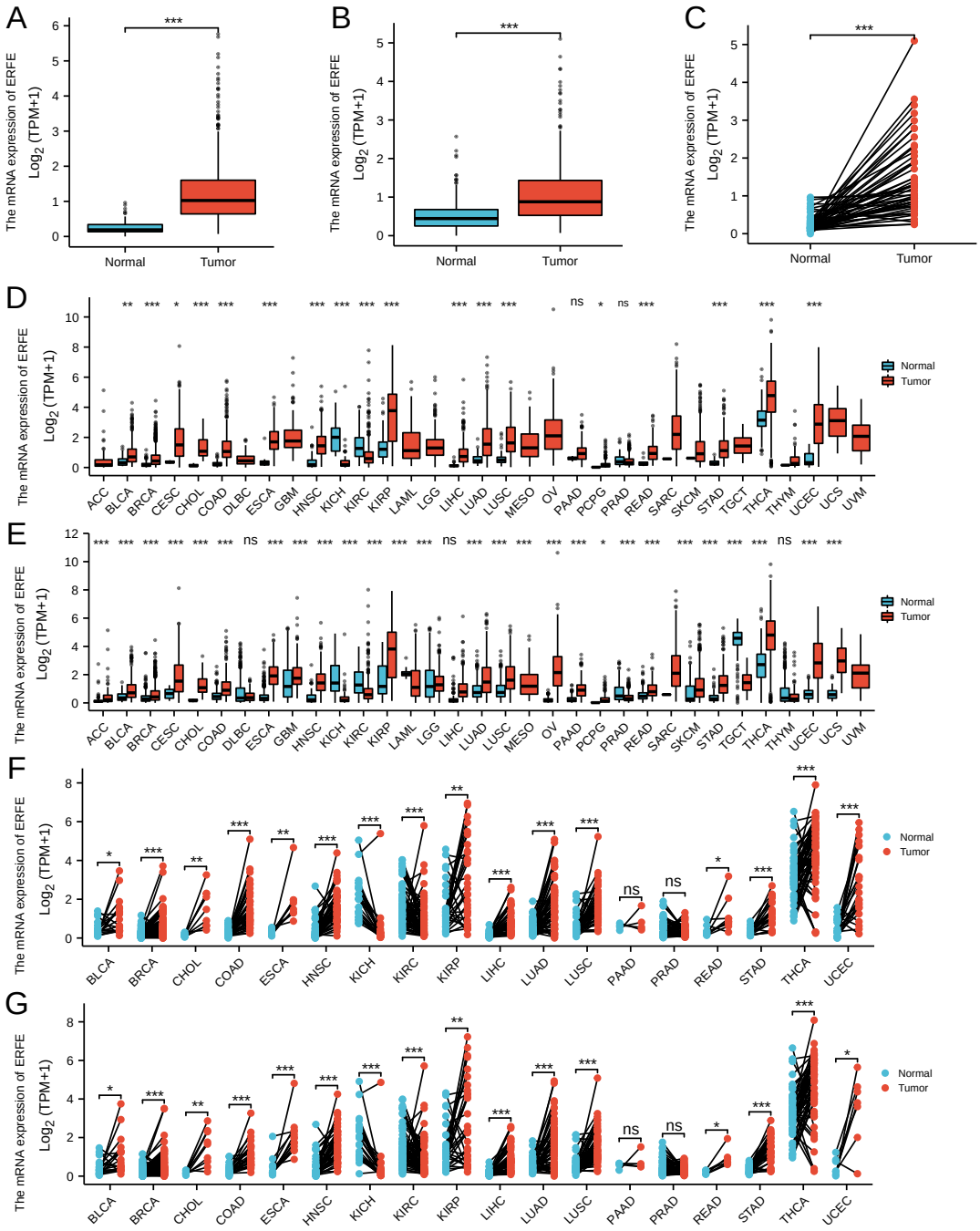

Fig. 1. Detection of *ERFE* expression discrepancies in various cancer types and CRC. (A) Comparative analysis using boxplots to visualize the distinction in *ERFE* mRNA expression levels between normal tissues and CRC in TCGA COADREAD dataset. (B) Comparative analysis using boxplots to visualize the distinction in *ERFE* mRNA expression levels between normal tissues and CRC in TCGA-GTEx-COADREAD dataset. (C) Comparative analysis using pairwise boxplots to visualize the distinction in *ERFE* mRNA expression levels between normal tissues and CRC in TCGA dataset. (D) Evaluation of *ERFE* mRNA expression levels in multiple cancer types based on TCGA database. (E) Evaluation of *ERFE* mRNA expression levels in multiple cancer types based on XENA-TCGA-GTEx database. (F) Comparative analysis using pairwise boxplots to visualize the distinction in *ERFE* mRNA expression levels among multiple cancer types using the TCGA database. (G) Comparative analysis using pairwise boxplots to visualize the distinction in *ERFE* mRNA expression levels among multiple cancer types using the XENA-TCGA database. (ns, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). *ERFE*, Erythroferrone; TCGA, the Cancer Genome Atlas; CRC, colorectal cancer; COADREAD, Colon adenocarcinoma/Rectum adenocarcinoma; GTEx, Genotype-Tissue Expression; TPM, Transcripts Per Million; ns, no significance.

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

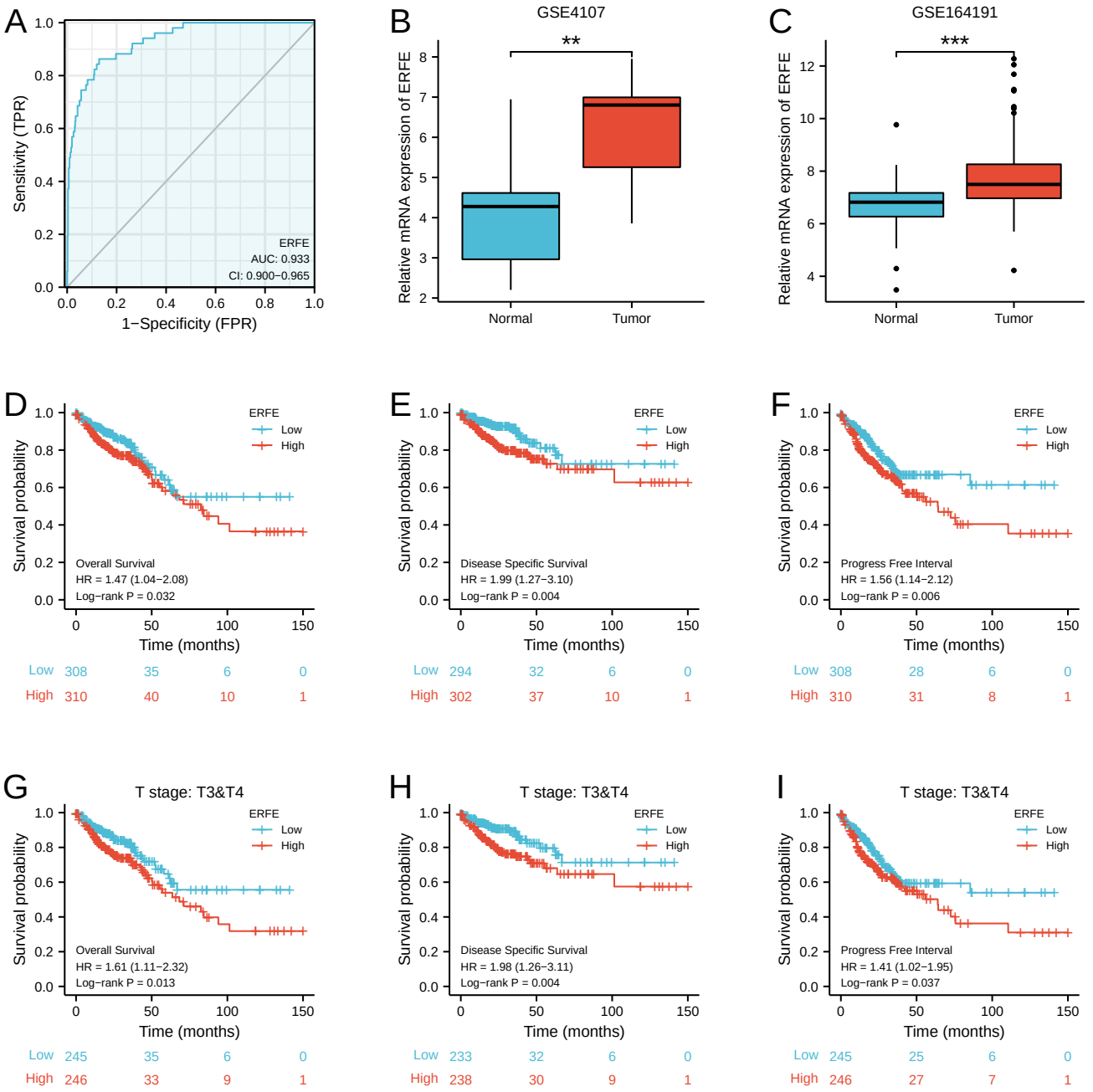

Fig. 2. ROC analysis and survival analysis of *ERFE*. (A) The diagnostic value of the *ERFE* gene level was assessed using the TCGA data set. (B) The relative mRNA expression level of the *ERFE* gene was analyzed in the GSE4107 data set. (C) The relative mRNA expression level of the *ERFE* gene was further analyzed in the GSE4107 data set. (D) Kaplan-Meier curves were used to demonstrate the overall survival of CRC patients based on their *ERFE* expression. (E) Kaplan-Meier curves were used to show the disease-specific survival of CRC patients based on their *ERFE* expression. (F) Kaplan-Meier curves were used to demonstrate the progression-free interval of CRC patients based on their *ERFE* expression. (G) Kaplan-Meier curves were used to assess the overall survival of CRC patients in T3&T4 stage, stratified by *ERFE* expression. (H) Kaplan-Meier curves were used to analyze the disease-specific survival of CRC patients in T3&T4 stage, stratified by *ERFE* expression. (I) Kaplan-Meier curves were used to assess the progression-free interval of CRC patients in T3&T4 stage, stratified by *ERFE* expression. (**, $p < 0.01$; ***, $p < 0.001$). GSE, Gene Expression Omnibus Series; ROC, Receiver Operating Characteristic; FPR, False Positive Rate; AUC, Area Under Curve; HR, Hazard ratio; CI, Confidence interval; TPR, True positive rate.

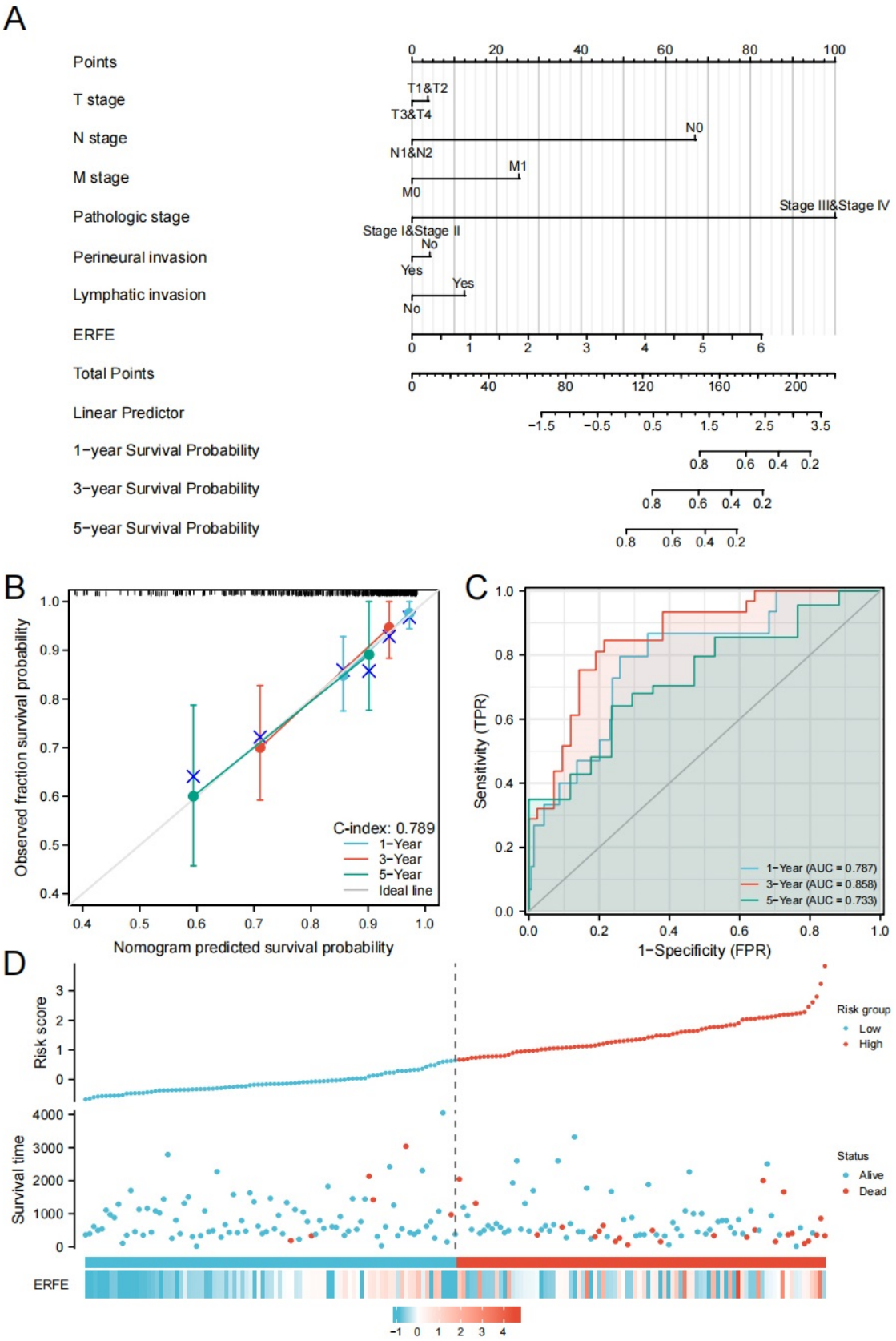

Fig. 3. Evaluation of prognostic value of *ERFE* in CRC patients. (A) Nomogram construction based on *ERFE* and clinicopathologic variable. (B) Calibration curves of 1-, 3-, and 5-year. (C) ROC curves and its AUCs for 1-, 3-, and 5-year survival of *ERFE.* (D) From top to bottom: the risk scores, the distribution of patients' living conditions and the expression heat map of *ERFE* in the *ERFE* low-expression group and the *ERFE* high-expression group. AUC, Area Under Curve; TPR, True positive rate.

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

to insulin stimulus, and the negative regulation of synaptic transmission, carbohydrate metabolic process, cell migration, and excretion (Fig. 4A). The potential biological functions of *ERFE* were studied through the KEGG and GO pathway analysis. GO terms included MF, BP, and CC. The GO enrichment analysis results suggested that *ERFE* was primarily enriched in chemokine activity, in the structural constituent of the extracellular matrix, which confers tensile strength, cytokine activity, receptor-ligand activity, tertiary granule membrane, collagen-containing extracellular matrix, secretory granule membrane, positive regulation of secretion, neutrophil-mediated immunity, positive regulation of cytokine production, and extracellular matrix organization (Fig. 4B). The enrichment analysis of KEGG suggested that *ERFE* was primarily enriched in signaling pathways such as Mitogen-activated protein kinase (MAPK), nuclear factor (NF)-Kappa B, Toll-like receptor, Tumor necrotic factor (TNF), and Nucleotide oligomerization domain like receptor (NLR) signaling pathway. *ERFE* is also enriched in transcriptional misregulation in cancer, hematopoietic cell lineage, proteoglycans in cancer, leukocyte transendothelial migration, and Cytokine-cytokine receptor interaction (Fig. 4C).

Correlation of ERFE with Immune Characteristics of Patients with CRC

The immune scores and degree of immune cell infiltration in the TME of CRC were determined using the ssGSEA algorithm. The expression profile of the *ERFE* gene and the immune cell infiltration was shown to be somewhat correlated. Positive correlation of *ERFE* with helper T cell 1 (Th1) cells, macrophages, neutrophils, Follicular helper T cell (Tfh), Dendritic cell (DC), aDC, cytotoxic cells, iDC, Tem, Natural Killer (NK) cells, T cells, Regulatory cells (Treg), Mast cells, Th2 cells, Tgd, pDC, cytotoxic T cell ($CD8^+$ T cells), B cells, Tfh, NK, CD56dim cells, T helper cells, and Eosinophils, were found but were positively linked to Tcm and Th17 cells (Fig. 5A). The application of the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) algorithm elucidated that the immune score, interstitial score, and ESTIMATE score experienced notable enhancement in the high-*ERFE* expression group compared to the low-*ERFE* expression group (as illustrated in Fig. 5B–D). Furthermore, the enrichment scores showcasing immune cell infiltration exhibited significant variations between the high and low expression groups, with the high expression group displaying higher values for a diverse array of cells such as aDC, $CD8^+$ T cells, cytotoxic cells, DCs, iDCs, macrophages, Mast cells, neutrophils, NK cells, pDCs, T helper cells, Tem, Tfh, Tgd, Th1 cells, Th2 cells, and Treg (as portrayed in Fig. 5E).

Susceptibility to Immunotherapy

Bioinformatics tools were utilized to analyze the TCGA patient data in order to determine the expression profile of vital immune checkpoints. Furthermore, an exploration was conducted to establish the connection between the expression of the *ERFE* gene and the aforementioned immune checkpoints. The findings indicated a notable rise in the mRNA expression of specific immune checkpoints among all patients exhibiting high *ERFE* expression, include *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *PDCDILG2*, *TIGIT* and *SIGLEC15* (Fig. 6A). The significant link between the *ERFE* gene and the expression of multiple immune cell gene markers was demonstrated by a heat map (Fig. 6B). A histogram was employed to visualize the correlation analysis results of *ERFE* gene expression with MSI, TMB, and NEO. This allowed for the examination of the relationship between these factors. As shown in (Fig. 6C–E), MSI, NEO, and TMB were significantly higher when patients had high expression of *ERFE* ($p < 0.001$). In addition, IC_{50} values were calculated to analyze the susceptibility of patients to molecular targeting drugs. The data obtained suggested that patients with increased *ERFE* expression had a significantly lower IC_{50} value for 5-FU (Fig. 6F). Based on the heat map analysis, it is evident that the *ERFE* gene and three ICB-related genes display similar expression patterns. Their expression levels are low in the pre-treatment group, but significantly increase in the ICB+ radiation group (**Supplementary Fig. 3A**). The scatter plots represent the expression levels of these genes in the before-treatment, ICB-treatment, and ICB+ radiation groups, with the horizontal and vertical axes indicating the expression levels of *ERFE* genes and ICB-related genes, respectively. According to the correlation coefficient and significance test *p* value, there is no significant correlation between *ERFE* gene and ICB-related genes before treatment, but with the implementation of immunotherapy, *ERFE* gene expression in ICB+ radiation group has a significant positive correlation with CD274 and PDCD1 (**Supplementary Fig. 3B**).

Vitro Assays

The expression of *ERFE* in the chosen CRC cell lines (HCT116 and LOVO) was examined using qRT-PCR and WB methods. According to the qRT-PCR results, the RNA level of *ERFE* in CRC cells (HCT116 and LOVO) was significantly higher compared to FHC cells ($p < 0.05$, Fig. 7A). Subsequently, WB analysis was conducted to determine the protein level expression of *ERFE*, and the results showed a significant increase in *ERFE* expression in CRC cell lines compared to FHC cells at the protein level (Fig. 7B). By analyzing the protein images, it can be visually observed that the protein level of *ERFE* in Lovo and HCT116 cells is considerably higher than that in normal colon epithelial cells ($p < 0.01$, Fig. 7C). WB analysis also revealed a decrease in *ERFE* expression in CRC cell lines

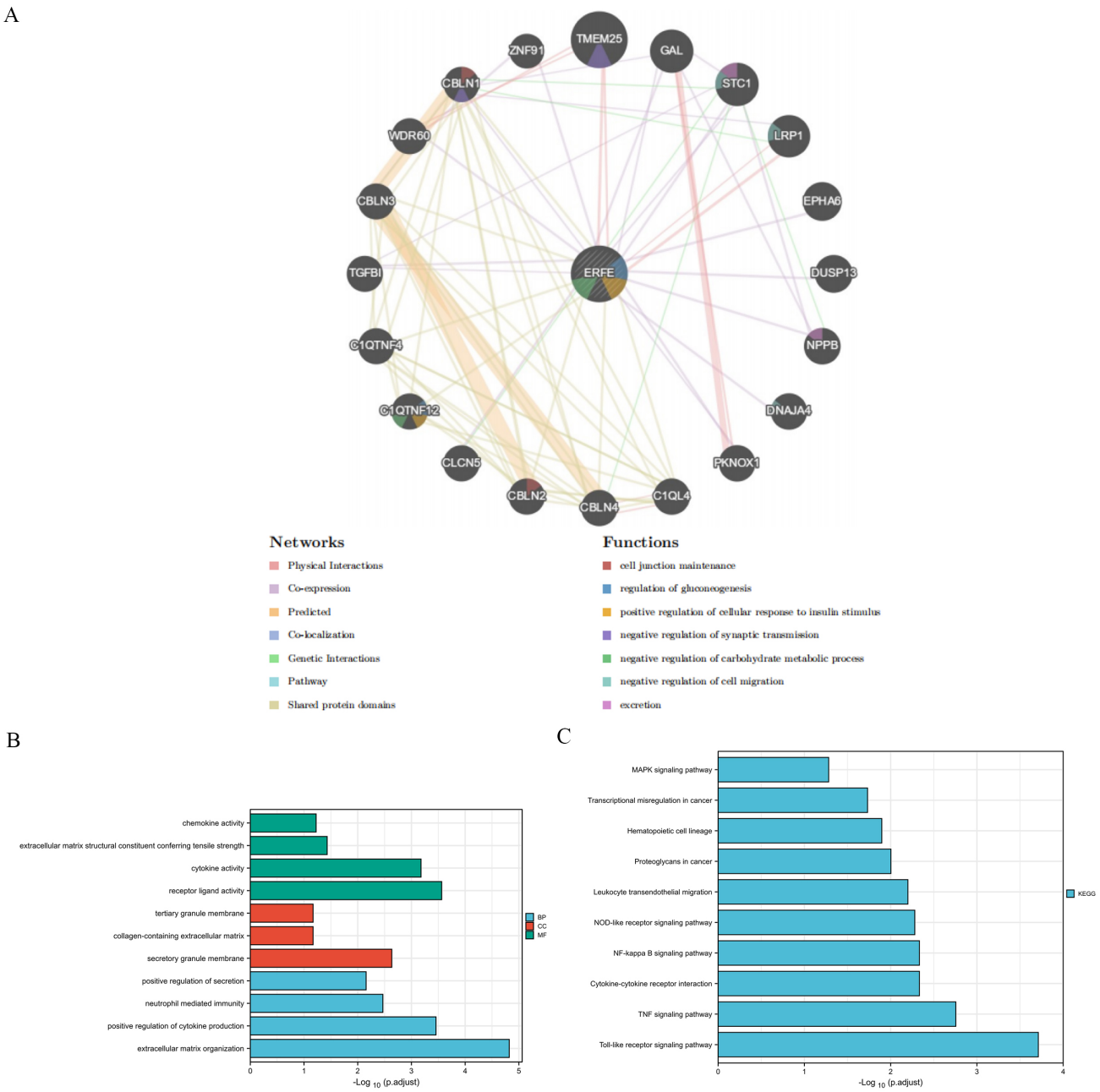

Fig. 4. GO and KEGG analysis of *ERFE* gene. (A) Using GeneMANIA database to build Protein-Protein Interaction (PPI) network of *ERFE.* (B) Histogram (BP, MF and CC) of GO function enrichment analysis of *ERFE*-related genes. (C) Histogram of the first 10 KEGG pathways of *ERFE*-related genes. BP, biological processes; MF, molecular functions; CC, cellular components; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

after transfection with siRNA (Fig. 7D). Furthermore, the protein images before and after knocking down the *ERFE* gene were quantified. It is evident that in HCT116 cells, knocking down *ERFE* led to a significant reduction in protein level expression ($p < 0.01$). Similarly, in Lovo cells, knocking down *ERFE* resulted in a significant decrease in protein expression ($p < 0.05$) (Fig. 7E). The CCK-8 assay demonstrated a weaker proliferative ability in CRC cells transfected with si-*ERFE* compared to control cells ($p < 0.05$, Fig. 7F,G), while the transwell assay indicated a significantly reduced invasive capacity of CRC cells in the si-*ERFE* group compared to the control group (Fig. 7H). The scratch assay revealed a substantial attenuation in the migratory ability of CRC cells transfected with si-*ERFE* compared to control cells 48 hours after scratching (Fig. 7I).

Discussion

A bioinformatics approach was utilized to conduct a comprehensive pan-cancer transcriptional analysis in this

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

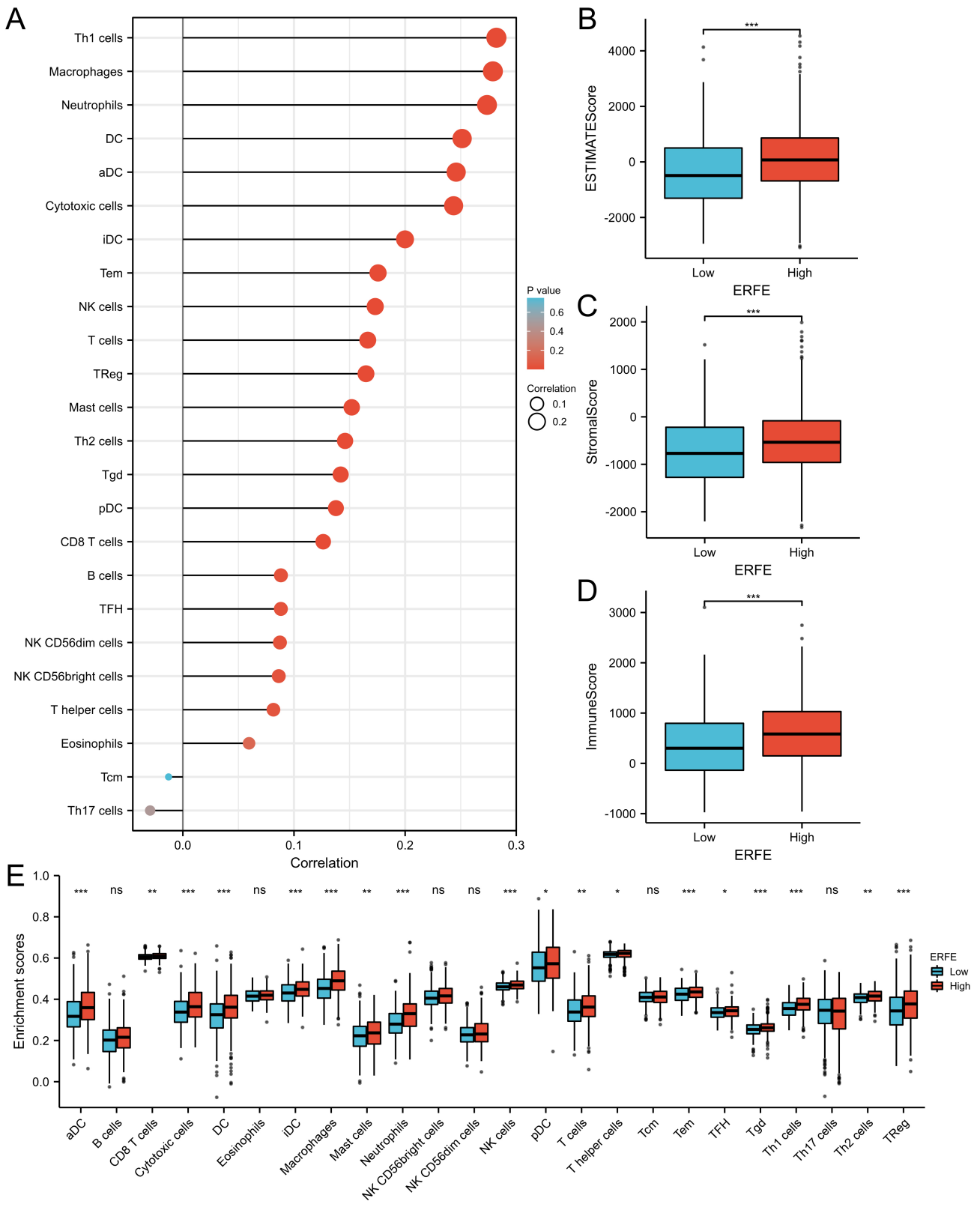

Fig. 5. The correlation between the expression level of *ERFE* and the infiltration level of immune cells in CRC. (A) Examining the relationship between the expression level of *ERFE* and the presence of tumor infiltrating lymphocytes. (B) Assessing the estimated scores in high and low expression groups of *ERFE* for comparison. (C) Comparing the stromal scores in high and low expression groups of *ERFE*. (D) Analyzing the immune scores in high and low expression groups of *ERFE* for comparison. (E) Evaluating the enrichment scores of 24 different immune cell types in high and low expression groups of *ERFE*. (ns, $p > 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

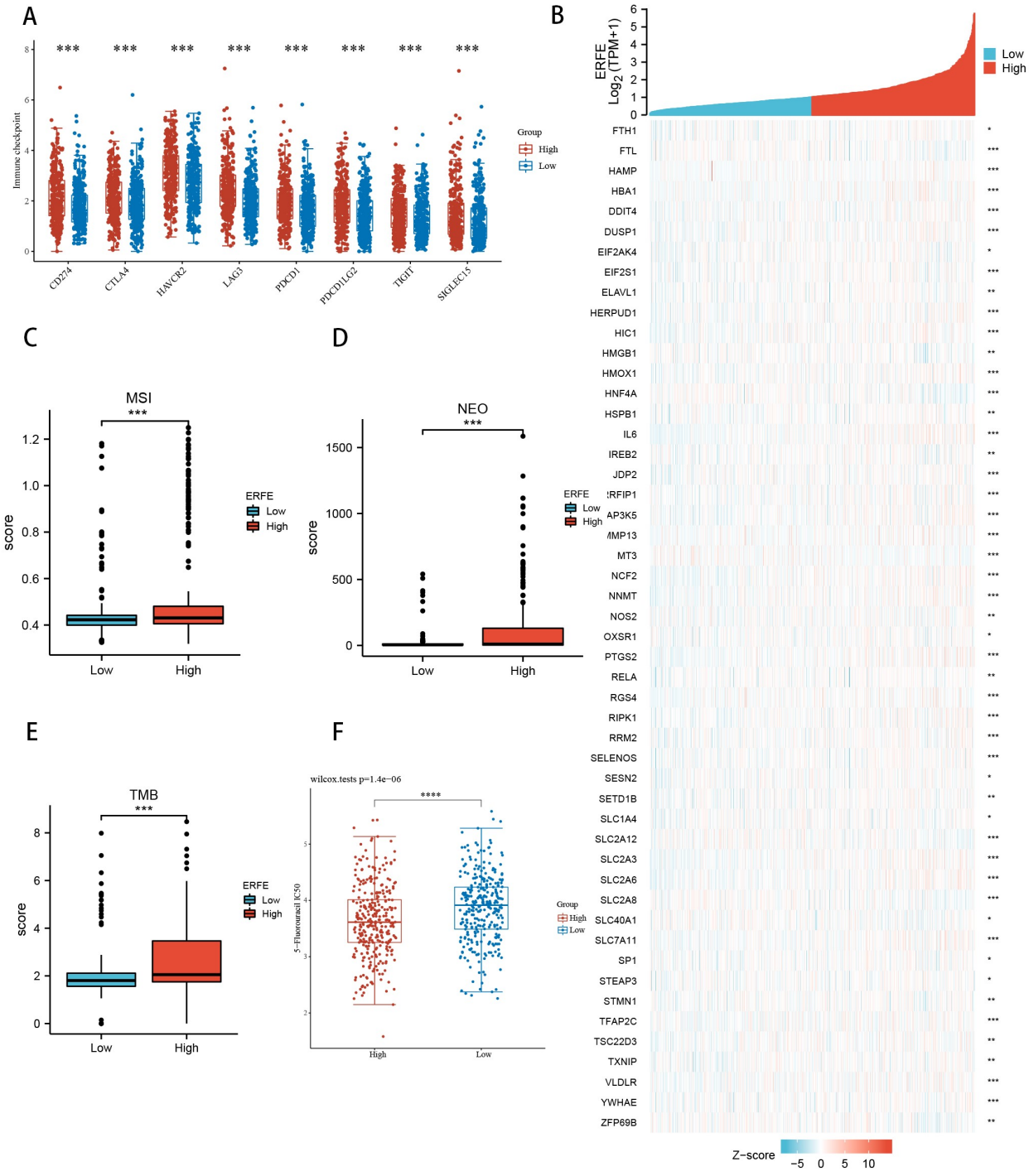

Fig. 6. Sensitivity analysis of immunotherapy. (A) Comparison of the expression levels at immune checkpoints between the high and low *ERFE* expression groups. (B) Correlation analysis between *ERFE* and immuno-infiltrating cell-related gene markers. (C) Comparison of microsatellite instability (MSI) fractional expression in the *ERFE* high-low expression group. (D) Comparison of neoantigen (NEO) fractional expression in high and low *ERFE* expression groups. (E) Comparison of Tumor Mutation Burden (TMB) fractional expression in high and low *ERFE* expression groups. (F) Comparison of half-maximum inhibitory concentration (IC_{50}) values of 5-Fluorouracil (5-FU) between the high and low *ERFE* expression groups. (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).

study. *ERFE* exhibits diverse functions across various types of cancer and is notably upregulated in multiple tumor types, including thyroid, breast, pancreatic, and CRC. Through analyzing data from TCGA and GEO databases, a significantly higher expression of *ERFE* was observed in CRC tissues compared to healthy tissues. Detecting

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

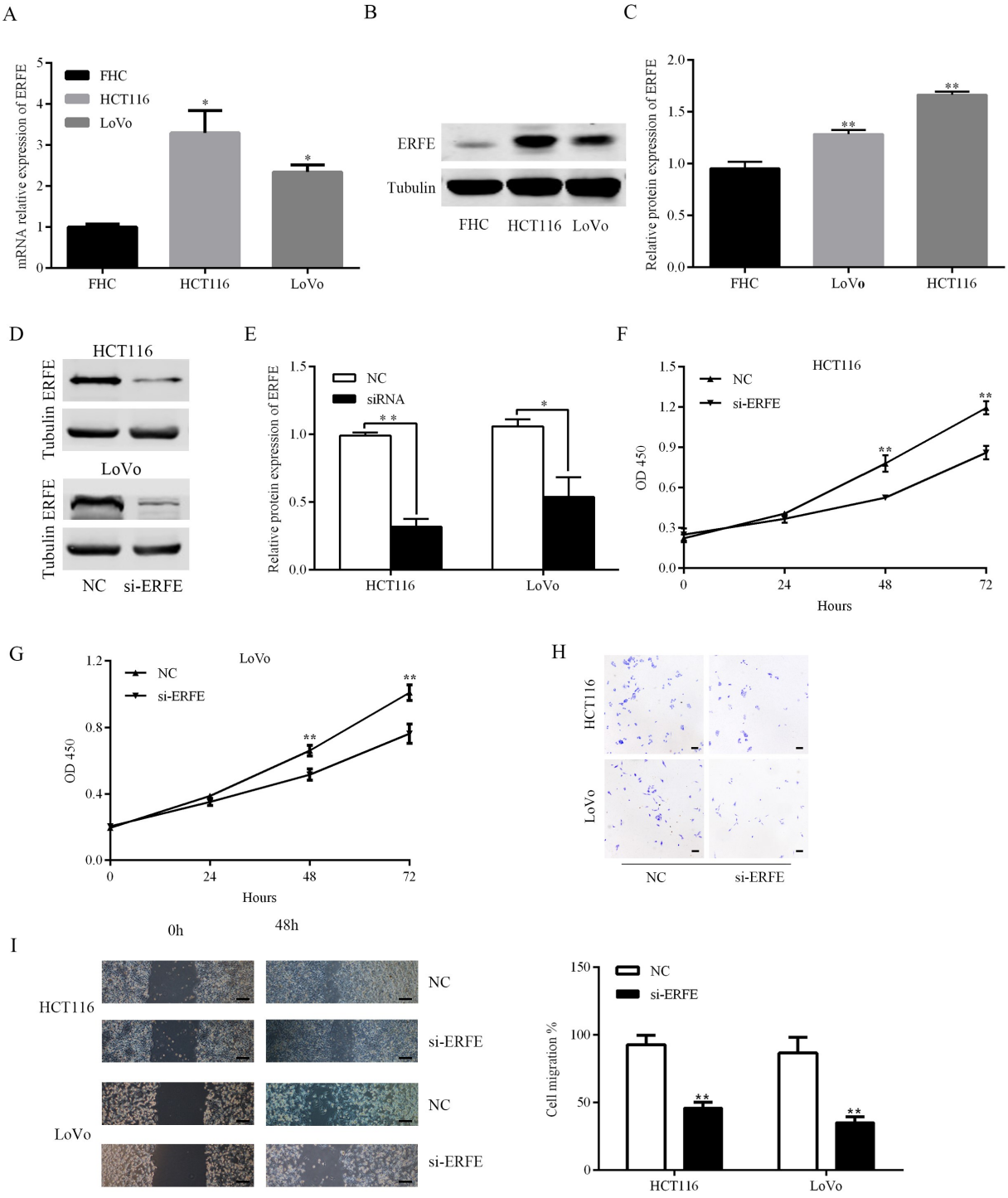

Fig. 7. Presents the results of cellular assays and immunohistochemical staining. (A) The mRNA expression of *ERFE* was assessed using Quantitative real time polymerase chain reaction (qRT-PCR). (B) The protein expression of *ERFE* was evaluated by performing western blot (WB). (C) The relative protein expression of *ERFE* was determined. (D) Upon transfection with siRNA-*ERFE* (si-*ERFE*), a significant decrease in the expression of *ERFE* protein was observed. (E) The relative protein expression of *ERFE* was measured before and after transfection. (F,G) Transfection of si-*ERFE* resulted in a considerable reduction in the proliferation ability of HCT116 and LoVo cells. (H) The invasive ability of HCT116 and LoVo cells was significantly reduced after being transfected with si-*ERFE. (*scale bar represents 50 μm). (I) Transfection of si-*ERFE* significantly weakened the migration ability of HCT116 and LoVo cells (Note: The original magnification is ×4, and the black scale bar represents 100 μm). (*, $p < 0.05$; **, $p < 0.01$). NC, Normal control; OD, Optical Density.

CRC early through accurate diagnosis and appropriate treatment can enhance the prognosis for individuals diagnosed with this condition. Non-localized, minimally invasive, safe, and rapid assays for tumor marker detection offer several advantages, including precise diagnosis, effective treatment, and improved prognosis prediction for CRC. While numerous existing molecules currently contribute to guiding CRC treatment and improving prognosis, complete cure remains challenging for most patients with advanced-stage CRC. Thus, the identification of new biomarkers holds immense clinical significance by facilitating the screening of new drugs and personalized therapy. This study was conducted to investigate the diagnostic value of *ERFE* in CRC. The data from the ROC curve yielded an AUC value of 0.933, indicating the potential of *ERFE* as a biomarker for CRC diagnosis. In order to evaluate the association between *ERFE* and patient prognosis, we conducted analyses on OS, DSS, and PFI, considering the notable expression of *ERFE* in CRC. The unfavorable prognosis of patients with an upregulated *ERFE* gene indicates a correlation between increased gene expression and poorer prognosis. A subgroup analysis was conducted, which revealed that high *ERFE* expression predicted an adverse prognosis specifically for mid- to late-stage patients (stages T3 and T4). To identify the factors that endanger the prognosis of CRC patients, univariate and multivariate Cox analyses were performed. As demonstrated in **Supplementary Tables 4,5,6**, there was a strong correlation between *ERFE* expression levels and OS, DSS, and PFI. This compelling evidence suggests that *ERFE* can independently serve as a predictor and risk factor for CRC. Nevertheless, it should be noted that there was no significant correlation found between *ERFE* expression and the clinical and pathological features of CRC, potentially due to various underlying causes influencing features such as clinical stage and pathological grade.

Despite the fact that the role of *ERFE* in promoting tumors in CRC is well established, the specific mechanism through which *ERFE* influences CRC progression remains uncertain. In order to shed light on the potential involvement of *ERFE* in CRC, an investigation was conducted using TCGA data to analyze the co-expression of genes related to *ERFE*, employing both GO and KEGG methodologies. The results of the KEGG analysis indicated that these *ERFE*-associated genes are mainly involved in signaling pathways related to the progression of cancer and inflammation. Additionally, GO enrichment analysis revealed that these genes primarily regulate the immune microenvironment by influencing cytokines, components of the extracellular matrix, and immunity mediated by neutrophils. Furthermore, molecular functional analysis demonstrated that these genes primarily modulate cytokines, receptor-ligand activity, and components of the extracellular matrix, with the ability to interact with various structures, including chemokines. The chemokine signaling system participates significantly in either suppressing or promoting growth, as well as tumor proliferation, angiogenesis, and metastasis [12]. Moreover, the analysis of cellular composition revealed that these genes are highly concentrated in diverse membranes, such as secretory granule membranes, tertiary granule membranes, and components of the extracellular matrix. It was postulated that *ERFE* might play a role in the development and progression of CRC by being expressed on membranes and contributing to dysregulated transcription, hematopoietic processes, altered extracellular matrix components, inflammatory response, and immune response in cancer. The existing literature solely focuses on *ERFE* and CRC, presenting a model that independently predicts the prognosis of patients with right-sided colon cancer by considering four genes: Erythroblastosis virus E26 oncogene homolog (*ERG*), *ERFE*, Growth factor independent 1 RASL10B: Ras-like protein family member 10B (*GFI1*), and *RASL10B*. The data obtained from the analysis of functional enrichment suggest a connection between right-sided colon cancer and antigen presentation and processing pathways, natural killer cell-mediated cytotoxicity, immunoglobulin production within the intestinal immune network, and type I diabetes [13]. Additionally, speculations regarding its potential functions in CRC align well with the findings of the current study.

The formation of a complete functional unit involves the tumor cell and the tumor microenvironment (TME), where the cell acts as the seed and the microenvironment acts as the soil [14]. The TME encompasses the surrounding blood vessels, various cells including immune cells, fibroblasts, and bone marrow-derived inflammatory cells, as well as signaling molecules and the extracellular matrix. The interaction and co-evolution between tumor cells and the microenvironment play a crucial role in promoting tumorigenesis. A particular focus is given to tumor-infiltrating immune cells, which serve as markers for studying the response of cancer cells to immunotherapy. This study investigates the correlation between the expression of *ERFE* and the infiltration status of immune cells in CRC. Yoshihara *et al*. [15] developed the Estimation algorithm, capable of analyzing gene expression data to estimate the levels of immune cell infiltration and stromal cell presence. Their findings demonstrate a positive correlation between increased *ERFE* expression levels and higher interstitial, immune, and assessment scores. Additionally, the correlation between genetic markers of various immune cells and the expression of *ERFE* indicates a significant role of *ERFE* in the regulation of the tumor microenvironment (TME). However, the certainty of utilizing immune-related treatments for individuals with colorectal cancer (CRC) and high *ERFE* expression remains uncertain. Therefore, further exploration was conducted to determine whether *ERFE* is associated with immune checkpoints, such as PD-L1, which plays a critical role in predicting the efficacy of immunotherapy in CRC. Genes as-

Journal of
BIOLOGICAL REGULATORS
and Homeostatic Agents

sociated with immune checkpoints include *CD274* (*PD-L1*), *CTLA-4*, *HAVCR2* (*TIM3*), *LAG-3*, *PDCD1* (*PD-1*), *PDCD1LG2* (*PD-L2*), *TIGIT*, and *SIGLEC15*. Interestingly, patients with increased *ERFE* expression exhibited significant upregulation of almost all these genes, suggesting that *ERFE* is involved in the immune checkpoint pathway and may serve as a potential biomarker for CRC immunotherapy. Mismatch repair is a mechanism that accurately detects and corrects errors in DNA replication or recombination, specifically base mismatches. Microsatellite instability (MSI) refers to changes in the length of microsatellite sequences during DNA replication, which can result from insertion or deletion mutations due to faulty mismatch repair. This condition can serve as a prognostic factor for patients with CRC, predicting their outcomes [16]. Consequently, the assessment of MSI can effectively determine the efficacy of anti-Programmed death 1 (PD1) immunotherapy in the treatment of CRC [17]. Immune checkpoint inhibitors (ICIs) have shown significant improvements in patients with advanced CRC who have high levels of MSI. Besides ICIs, NEO, a class of antigens specific to tumors, exhibits tumor specificity, making it an excellent target for anti-tumor immunotherapy. NEO has been extensively utilized in both basic and clinical studies on immunotherapy for CRC treatment [18]. The TMB score, a crucial biomarker in various cancers, is frequently employed to predict the effectiveness of immunotherapy and identify individuals who are more likely to benefit from it. The TMB score aids in identifying patients with worse prognoses in the high-*ERFE* expression group, particularly those who stand to gain the most from immunotherapy [19]. In this study, MSI, TMB, and NEO were found to be significantly higher in the high-*ERFE* expression group compared to the low-*ERFE* expression group. Therefore, *ERFE* may hold potential as an immunotherapeutic target for CRC.

Clinical diagnosis and treatment can benefit greatly from the utilization of bioinformatics. Although this study has yielded numerous valuable findings, there are still gaps that require filling. This study aimed to investigate the specific role of *ERFE* in CRC. By manipulating the expression of *ERFE* in CRC cell lines, it was observed that a significant reduction in *ERFE* expression effectively impeded the migratory, invading, and proliferative capabilities of CRC cells. These findings constitute the initial evidence of *ERFE* acting as a pro-oncogene in CRC. However, further research is necessary to elucidate the specific mechanism of action. In summary, this research stands as the pioneer in identifying and validating the distinct expression of *ERFE* and its clinical significance in CRC. The discoveries made in this study have immense diagnostic and predictive value in terms of response to immunotherapy and molecular targeting therapy in CRC. Overall, our findings shed light on the involvement of *ERFE* in CRC initiation and progression, serving as a guide for future personalized and precise treatments.

Conclusions

In summary, *ERFE*, functioning as a pro-oncogene in CRC, is linked to the initiation and advancement of cancer, and can serve as a standalone marker for unfavorable prognosis among CRC patients.

Availability of Data and Materials

All experimental data included in this study can be obtained by contacting the first author if needed.

Author Contributions

Conception and design: AZ, JZ, YZ; Administrative support: JZ; Provision of experiment design: YS, YX; Collection and assembly of data: LY, BZ, JG; Data analysis and interpretation: AZ, JG, CY; Manuscript writing: All authors; Final approval of manuscript: All authors. All authors contributed to important editorial changes in the manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This study was supported by The Natural Science Foundation of Heilongjiang Province (YQ2019H018).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.23812/j.biol.regul.homeost.agents.20243809.467.

References

[1] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, *et al*. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer. 2015; 136: E359–E386.

[2] Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, *et al*. Colorectal cancer statistics, 2020. CA: a Cancer Journal for Clinicians. 2020; 70: 145–164.

[3] Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017; 66: 683–691.

[4] Carlsen L, Huntington KE, El-Deiry WS. Immunotherapy for Colorectal Cancer: Mechanisms and Predictive Biomarkers. Cancers. 2022; 14: 1028.
[5] Kamel F, Eltarhoni K, Nisar P, Soloviev M. Colorectal Cancer Diagnosis: The Obstacles We Face in Determining a Non-Invasive Test and Current Advances in Biomarker Detection. Cancers. 2022; 14: 1889.
[6] Appleby S, Chew-Harris J, Troughton RW, Richards AM, Pemberton CJ. Analytical and biological assessment of circulating human erythroferrone. Clinical Biochemistry. 2020; 79: 41–47.
[7] Coffey R, Jung G, Olivera JD, Karin G, Pereira RC, Nemeth E, *et al*. Erythroid overproduction of erythroferrone causes iron overload and developmental abnormalities in mice. Blood. 2022; 139: 439–451.
[8] Brookes MJ, Boult J, Roberts K, Cooper BT, Hotchin NA, Matthews G, *et al*. A role for iron in Wnt signalling. Oncogene. 2008; 27: 966–975.
[9] Le NTV, Richardson DR. The role of iron in cell cycle progression and the proliferation of neoplastic cells. Biochimica et Biophysica Acta. 2002; 1603: 31–46.
[10] Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncology. 2015; 19: A68–A77.
[11] Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, *et al*. The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics. 2013; 45: 1113–1120.
[12] Do HTT, Lee CH, Cho J. Chemokines and their Receptors: Multifaceted Roles in Cancer Progression and Potential Value as Cancer Prognostic Markers. Cancers. 2020; 12: 287.
[13] Liang L, Zeng JH, Qin XG, Chen JQ, Luo DZ, Chen G. Distinguishable Prognostic Signatures of Left- and Right-Sided Colon Cancer: a Study Based on Sequencing Data. Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology. 2018; 48: 475–490.
[14] Wang L, Ding K, Zheng C, Xiao H, Liu X, Sun L, *et al*. Detachable Nanoparticle-Enhanced Chemoimmunotherapy Based on Precise Killing of Tumor Seeds and Normalizing the Growing Soil Strategy. Nano Letters. 2020; 20: 6272–6280.
[15] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, *et al*. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications. 2013; 4: 2612.
[16] Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology. 2005; 23: 609–618.
[17] Peng L, Chang J, Liu X, Lu S, Ren H, Zhou X, *et al*. *MC1R* Is a Prognostic Marker and Its Expression Is Correlated with MSI in Colorectal Cancer. Current Issues in Molecular Biology. 2021; 43: 1529–1547.
[18] Zheng Y, Fu Y, Wang PP, Ding ZY. Neoantigen: A Promising Target for the Immunotherapy of Colorectal Cancer. Disease Markers. 2022; 2022: 8270305.
[19] Tang Z, Wu Y, Sun D, Xue X, Qin L. A novel prognostic immunoscore based on The Cancer Genome Atlas to predict overall survival in colorectal cancer patients. Bioscience Reports. 2021; 41: BSR20210039.