



Telomere-related gene risk model for prognosis prediction in colorectal cancer

Hao Chen¹, Yuhao Pan¹, Chenhui Lv¹, Wei He¹, Dingting Wu², Qijia Xuan^{1^}

¹Department of Medical Oncology, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, China; ²Department of Clinical Nutrition, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, China

Contributions: (I) Conception and design: H Chen; (II) Administrative support: Q Xuan; (III) Provision of study materials or patients: D Wu, H Chen; (IV) Collection and assembly of data: Y Pan, C Lv, W He; (V) Data analysis and interpretation: H Chen, Y Pan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Qijia Xuan, PhD. Department of Medical Oncology, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, N1 Shangcheng Road, Yiwu 322000, China. Email: xuanqijia@zju.edu.cn.

Background: Colorectal cancer (CRC) is the third-most prevalent cancer globally. The biological significance of telomeres in CRC carcinogenesis and progression is underscored by accumulating data. Nevertheless, not much is known about how telomere-related genes (TRGs) affect CRC prognosis. Therefore, the aim of this study was to investigate the role of TRGs in CRC prognosis.

Methods: We retrospectively obtained the expression profiles and clinical data of CRC patients from public databases. Utilizing least absolute shrinkage and selection operator (LASSO) regression analysis, we created a telomere-related risk model to predict survival outcomes, identifying ten telomere-related differentially expressed genes (TRDEGs). Based on TRDEGs, we stratified patients from The Cancer Genome Atlas (TCGA) into low- and high-risk subsets. Subsequently, we conducted comprehensive analyses, including survival assessment, immune cell infiltration, drug sensitivity, and prediction of molecular interactions using Kaplan-Meier curves, ESTIMATE, CIBERSORT, OncoPredict, and other approaches.

Results: The model showed exceptional predictive accuracy for survival. Significant differences in survival were observed between the two groups of participants grouped according to the model ($P < 0.001$), and this difference was further confirmed in the external validation set (GSE39582) ($P = 0.004$). Additionally, compared to the low-risk group, the high-risk group exhibited significantly advanced tumor node metastasis (TNM) stages, lower proportions of activated CD4⁺ T cells, effector memory CD4⁺ T cells, and memory B cells, but increased ratios of M2 macrophages and regulatory T cells (Tregs), elevated tumor immune dysfunction and exclusion (TIDE) scores, and diminished sensitivity to dabrafenib, lapatinib, camptothecin, docetaxel, and telomerase inhibitor IX, reflecting the signature's capacity to distinguish clinical pathological characteristics, immune environment, and drug efficacy. Finally, we validated the expression of the ten TRDEGs (*ACACB*, *TPX2*, *SRPX*, *PPARGC1A*, *CD36*, *MMP3*, *NAT2*, *MMP10*, *HIGD1A*, and *MMP1*) through quantitative real-time polymerase chain reaction (qRT-PCR) and found that compared to normal cells, the expression levels of *ACACB*, *HIGD1A*, *NAT2*, *PPARGC1A*, and *TPX2* in CRC cells were elevated, whereas those of *CD36*, *SRPX*, *MMP1*, *MMP3*, and *MMP10* were reduced.

Conclusions: Overall, we constructed a telomere-related biomarker capable of predicting prognosis and treatment response in CRC individuals, offering potential guidance for drug therapy selection and prognosis prediction.

Keywords: Colorectal cancer (CRC); telomere; prognosis; immune infiltration; drug sensitivity

[^] ORCID: 0000-0002-3591-270X.

Submitted Jan 08, 2024. Accepted for publication Jun 02, 2024. Published online Jul 19, 2024.

doi: 10.21037/tcr-24-43

View this article at: <https://dx.doi.org/10.21037/tcr-24-43>

Introduction

Colorectal cancer (CRC) is the third most common malignancy, with 1,880,725 new cases and 915,880 new deaths reported in 2020, according to the International Agency for Research on Cancer (1). Early-stage CRC is associated with a 5-year survival rate of approximately 90%, while advanced CRC confers a markedly lower survival rate of 10–15% (2,3). However, in its early stages, CRC is asymptomatic and difficult to detect. When diagnosed, most individuals are already at an advanced stage. Furthermore, there are wide variations in prognosis and treatment outcomes owing to the significant diversity of CRC. Therefore, the identification of novel prognostic markers to guide tailored therapies and predict outcomes is imperative.

Telomeres, distinctive DNA structures located at chromosome ends, consist of repetitive nucleotide sequences and shelterin proteins, safeguarding chromosome integrity against genomic erosion (4). The telomerase repeat amplification protocol (TRAP) (5), an efficient and sensitive method for detecting telomerase activity, has been pivotal in studies revealing that 85–90% of human malignant tumors express this activity. In contrast, benign tumors or normal tissues are predominately negative or sporadically positive,

rendering telomerase a crucial marker for diagnosing and predicting the prognosis of malignant tumors. Numerous age-related illnesses, including cancer, are associated with shorter telomeres and decreased telomerase activity (6,7). An increasing number of studies have investigated the role of telomeres and telomerase in carcinogenesis. Notably, in prostate cancer, shorter telomeres correlate with higher genomic instability (8). The dynamics of telomere maintenance during hepatocellular carcinoma carcinogenesis vary, aligning with tumor progression and aggressiveness (9). Notably, patients with early-onset CRC (EOCRC) exhibit significantly shorter telomeres than healthy individuals, suggesting a link between telomere shortening and EOCRC susceptibility (10). Nevertheless, conflicting studies exist regarding the association of telomere length (TL) with CRC.

Significant progress has been made in understanding TL in CRC and its effect on CRC prognosis. Kroupa *et al.* conducted a retrospective study on 721 patients with sporadic CRC, finding variations in TL between tumors and adjacent mucosa, among different tumor sites, and that the TL ratio between tumor tissue and adjacent mucosa is associated with patient's prognosis (11). Luu *et al.* performed a large cohort study of ethnic Chinese in Singapore, discovering that longer telomeres are associated with a higher risk of CRC, particularly rectal cancer (12). Previous research has reported on the relationship between telomere-related genes (TRGs) and the prognosis of kidney cancer, pancreatic cancer, breast cancer, lung adenocarcinoma, oral squamous cell carcinoma, and hematological malignancies (13–18). However, no research has been conducted to examine the role of TRGs in the prognosis of CRC. Consequently, our objective was to develop a gene signature utilizing the least absolute shrinkage and selection operator (LASSO) method. This signature aimed to investigate the potential applications of TRGs in predicting survival outcomes, understanding immunological conditions, and assessing therapeutic responses in patients with CRC. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-43/rc>).

Highlight box

Key findings

- This study constructed a risk model for prognosis prediction in colorectal cancer (CRC) based on ten telomere-related genes and developed a clinically relevant prognostic model by integrating risk scores and pathologic stage, illustrating excellent predictive ability.

What is known and what is new?

- Telomere length is closely related to the prognosis of CRC.
- There are fewer research utilizing the telomere-related genes to predict the prognosis of CRC.

What is the implication, and what should change now?

- The risk model developed in this study provides a foundation for comprehending the roles of telomere-related genes and showcasing their clinical significance in CRC.

Methods

Data download

We downloaded the CRC dataset (TCGA-COADREAD) from The Cancer Genome Atlas (TCGA) using the TCGAbiolinks package (19). For analysis, we acquired sample data from 695 CRC cases, including 51 paracancerous normal samples (grouping: control) and 644 CRC tumor samples (grouping: COADREAD). Relevant clinical information was sourced from the UCSC Xena database (20). Tumor mutation burden (TMB) and microsatellite instability (MSI) data were downloaded from the cBioPortal database (21). Furthermore, we examined copy number variation (CNV) and single nucleotide polymorphism (SNP) data for the patients. The Gene Expression Omnibus (GEO) database provided the GSE74602, GSE25070 (22), and GSE9438 (23) datasets. As an independent external validation cohort, we selected the COAD cohort GSE39582. Specifically, GSE74602 comprised 30 CRC samples and 30 correspondingly matched adjacent non-tumor colon tissue samples. GSE25070 included 26 CRC samples and 26 adjacent non-tumor colon tissue samples. GSE9348 comprised 70 tumor samples and 12 samples from healthy controls. The dataset probe names were annotated using the chip GPL platform files. GSE39582 consisted of 566 tumor samples and 19 non-tumoral colorectal mucosa samples. All samples from the GSE74602, GSE25070, GSE9348, and GSE39582 datasets were included in subsequent analyses. We obtained a total of 5,504 TRGs from the GeneCards database (<https://www.genecards.org/>) (24). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Screening of telomere-related differentially expressed genes (TRDEGs) and functional enrichment analysis

We employed the “normalizeBetweenArrays” function from the limma package (25) to standardize the TCGA and three GEO datasets. Subsequently, differential analysis was conducted using the limma package, wherein differentially expressed genes (DEGs) in each dataset were screened based on criteria $|\log(\text{fold change})| > 1$ and $P < 0.01$. The resulting set of common DEGs (CO-DEGs) was then intersected with TRGs to obtain TRDEGs. Furthermore, univariate Cox regression analysis was performed to filter TRDEGs at $P < 0.05$, and the screened TRDEGs were subsequently used for further research.

Enrichment analysis for the TRDEGs was carried out using the R package “clusterProfile r” (26). The analysis encompassed Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (27). The screening criteria for entry into the analysis were set at $P < 0.05$ and false discovery rate (FDR) ($q.\text{value}$) < 0.05 .

Calculating the telomere score (Tscore) for the TCGA-COADREAD dataset

The relative abundance of each gene was measured through a single-sample gene set enrichment analysis (ssGSEA). Utilizing the expression profiles of TRDEGs, we estimated the Tscore for each sample within the TCGA dataset employing the R package “GSVA” (28). Subsequently, we categorized the COADREAD samples into two groups based on the median Tscores. Finally, the distribution of distinct clinical variables was illustrated using stacked histograms.

Constructing weighted gene co-expression network analysis (WGCNA)

WGCNA (29) was used to identify highly synergistic gene sets and candidate biomarker genes or therapeutic targets. We used the R package “WGCNA” (30) for WGCNA analysis. For this study, genes with expression variance in the top 40% among all genes in the TCGA cohort were selected as input. The minimum number of module genes was set to 60, the softpower was set to 6, the modules were merged, the height was set to 0.2, and the minimum distance was set to 0.2. The purpose was to assess the degree of connection between the various modules and Tscore groups. Based on the correlation values, the modules of interest were selected, revealing a strong association between each gene in the modules and telomeres.

Identification of subtypes in the TCGA-COADREAD dataset

According to the expression of TRDEGs identified through WGCNA, we employed the consensus clustering method (31) from the “ConsensusClusterPlus” package (32) to identify different subtypes in the TCGA-COADREAD dataset. During this process, we specified the number of clusters to range from 2 to 8, and repeated the procedure 50 times to extract 80% of the total samples, with $\text{clusterAlg} = \text{“km”}$ and $\text{distance} = \text{“euclidean”}$. A heat map was used

to demonstrate the variation in TRDEGs' expression throughout the clusters. Subsequently, a predictive survival Kaplan-Meier curve was generated after combining clinical data to examine the association between subtypes and survival outcomes.

Construction and validation of TRDEG's prognostic model

To obtain the prognostic model of TRDEGs, we applied the “glmnet” package (33) to conduct LASSO (34) regression analysis. The risk score was calculated using the following formula below:

$$\text{Risk score} = \sum_i \text{Coefficient}(\text{hub gene}_i) * \text{mRNA Expression}(\text{hub gene}_i) \quad [1]$$

The retained TRDEGs in the prognostic model were referred to as prognostic TRDEGs. Utilizing these prognostic TRDEGs, we computed risk scores for all patients in the TCGA cohort. Subsequently, patients were divided into low- and high-risk groups using median values. Next, we displayed the expression differences of prognostic TRDEGs in TCGA-COADREAD and validated them using GEO datasets. The correlation among prognostic TRDEGs was calculated, and the results were visualized using the “corrplot” package. The prognostic risk model was verified using an external validation cohort (GSE39582). Additionally, diagnostic receiver operating characteristic (ROC) curves (35) for the prognostic TRDEGs were presented in the TCGA dataset, assessing their diagnostic capabilities using the “pROC” package.

Gene set enrichment analysis (GSEA)

GSEA (36) was performed using the “clusterProfiler” package. The “c2.cp.all.v2022.1. Hs.symbols.gmt” gene set from the MSigDB database (37) served as the reference gene set. Screening criteria for significant enrichment included $P < 0.05$ and $\text{FDR}(\text{q.value}) < 0.05$.

Construction of a protein-protein interaction (PPI) network and a mRNA-miRNA, mRNA-transcription factors (TFs), and mRNA-RNA binding proteins (RBPs) interaction network

To investigate the relationship among prognostic TRDEGs, we constructed a PPI network using the STRING database (38), and Cytoscape (39) was employed for

visualization. Subsequently, we utilized the miRDB (40) database (<https://mirdb.org/>) to predict miRNAs interacting with prognostic TRDEGs and visualized the mRNA-miRNA interaction network in Cytoscape. In addition, we explored TFs that attached to prognostic TRDEGs using the CHIPBase (<https://rna.sysu.edu.cn/chipbase/>) and hTFtarget databases (41) (<http://bioinfo.life.hust.edu.cn/hTFtarget>). To predict the RBP with prognostic TRDEGs, we utilized the ENCORI (42) database (<https://starbase.sysu.edu.cn/>). Simultaneously, we predicted the protein structure of each predictive TRDEG within the PPI network using the AlphaFold website (43) (<https://alphafold.com/>).

Immune score and immune infiltration analysis

To enhance our comprehension of the predictive significance of the risk score related to immunity and stromal cells, we utilized the “ESTIMATE” package (44). Additionally, correlation dot plots were generated to illustrate the relationship between immune scores and the risk scores of the TRDEG prognostic model. Moreover, we used ssGSEA and the CIBERSORT (45) algorithm to assess the association between prognostic TRDEGs and immune cell infiltration. Boxplots and stacked histograms were used to display variations in immune cell infiltration abundance between the low- and high-risk groups. The correlation between immune cells and prognostic TRDEGs was illustrated using correlation dot plots with the “ggplot2” package.

Tumor immune dysfunction and exclusion (TIDE), TMB, MSI, and drug sensitivity between low- and high-risk groups

We examined differences in tumor immunotherapy responses between the two groups using TIDE, TMB, and MSI analysis results (46). The half-maximal inhibitory concentrations (IC50) of targeted medications, reflecting treatment sensitivity, were estimated based on the degree of gene expression using the R package “oncoPredict” (47).

Construction and validation of a clinical prognostic model

We conducted a univariate Cox regression analysis by combining clinical data with prognostic TRDEGs to assess their clinical significance. Factors with a significance level of $P < 0.1$ were included in the multivariate Cox regression

analysis for constructing a clinical prognostic model. Risk scores for the clinical prognostic model were computed based on the expression of relevant variables. The results of the univariate Cox regression were presented using a forest plot. Subsequently, we generated a nomogram (48) using the “rms” package to show the results of multivariate Cox regression analysis. Additionally, we constructed a calibration curve (49) to assess the accuracy of the nomogram. Furthermore, we evaluated the performance of the clinical prognostic model using decision curve analysis (DCA) (50), visualized with the “ggDCA” package (51).

Validation of the telomere-related signature by relative quantitative real-time PCR (qRT-PCR)

A CRC cell line (SW620, human CRC cells) and a control cell line (NCM460, human colon mucosal epithelial cells) were used to assess the expression levels of genes that comprise the telomere-related signature. All cell lines were cultured in RPMI-1640 (Hyclone, Logan, UT, USA) supplemented with 10% fetal bovine serum (Hyclone, USA). Total RNA was extracted from the cells using the TRIZOL reagent (Invitrogen, Carlsbad, CA, USA), and reverse transcription was performed using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA, USA, K1622). qRT-PCR was performed using a SYBR Green Real-time PCR Kit (Thermo Fisher Scientific, F-415XL) on an ABI 7500 Real-Time PCR System (Thermo Fisher Scientific). The primer sequences are listed in Table S1. The relative expression was computed using the $2^{-\Delta\Delta C_t}$ approach and β -actin was used as the internal control gene.

Protein expression analysis of the telomere-related signature

The immunohistochemical (IHC) data for CRC and colorectal tissues were acquired from the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>) and analyzed using the “HPAanalyze” package (52).

Statistical analysis

Data processing and analysis were conducted using R software (version 4.1.2). For comparing two sets of continuous data, the independent Student’s *t*-test determined the statistical significance of normally distributed variables, while the Mann-Whitney *U* test

assessed differences among non-normally distributed variables. Kaplan-Meier survival curves illustrated survival differences, with the log-rank test employed to assess the significance between the two groups. Correlation coefficients were calculated using Spearman’s analysis unless specified otherwise, with statistical significance set at $P < 0.05$.

Results

The study design is illustrated in Figure 1.

Identification of TRDEGs

First, data normalization was performed from TCGA-COADREAD, GSE74602, GSE25070, and GSE9348, and box plots were generated to compare the expression profiles before and after processing (Figure S1). The figure shows a consistent trend in the expression profiles of the four datasets following normalization. Subsequently, a differential analysis of all genes was conducted for TCGA-COADREAD (Figure 2A), GSE74602 (Figure 2B), GSE25070 (Figure 2C), and GSE9348 (Figure 2D), visualized as volcano maps. A total of 289 CO-DEGs were identified across the four datasets and represented in a Venn diagram (Figure 2E). The CO-DEGs were then intersected with TRGs to obtain 92 TRDEGs (Figure 2F). Finally, univariate Cox regression analysis was employed to screen TRDEGs, applying a screening criterion of $P < 0.05$, resulting in a total of thirteen identified TRDEGs (Figure 2G). The thirteen identified TRDEGs were *ACACB*, *ADH1B*, *CD36*, *CXCL1*, *FABP4*, *HIGD1A*, *MMP1*, *MMP10*, *MMP3*, *NAT2*, *PPARGC1A*, *SRPX*, and *TPX2*.

Functional enrichment analysis of TRDEGs

To examine the potential biological processes (BPs) associated with the TRDEGs, we conducted GO and KEGG analyses on the identified set of thirteen TRDEGs. The analysis revealed significant enrichment in various BPs, including cellular component disassembly, energy homeostasis, response to oxidative stress, regulation of fatty acid (FA) oxidation, and digestion. Molecular function (MF) analysis highlighted the top five enriched activities as serine-type endopeptidase activity, endopeptidase activity, metalloendopeptidase activity, long-chain FA transporter activity, and monocarboxylic acid-binding.

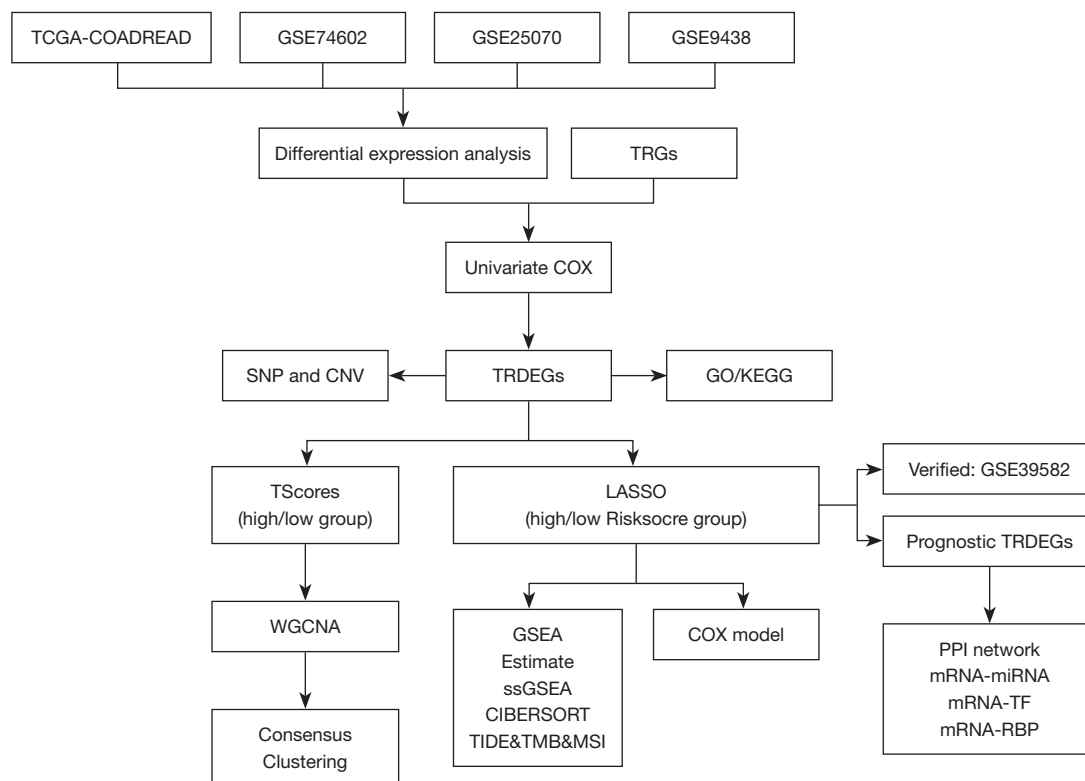


Figure 1 Study design. TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; TRGs, telomere-related genes; SNP, single nucleotide polymorphism; CNV, copy number variation; TRDEGs, telomere-related differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TScore, telomere scores; LASSO, least absolute shrinkage and selection operator; WGCNA, weighted gene co-expression network analysis; GSEA, gene set enrichment analysis; ssGSEA, single-sample gene set enrichment analysis; TIDE, tumor immune dysfunction and exclusion; TMB, tumor mutation burden; MSI, microsatellite instability; PPI, protein-protein interaction; TF, transcription factor; RBP, RNA-binding protein.

Regarding KEGG, the main enriched pathways included the adipocytokine signaling pathway, the PPAR signaling pathway, the interleukin (IL)-17 signaling pathway, insulin resistance, and the AMPK signaling pathway. The results of the GO and KEGG enrichment analyses are displayed in histograms (Figure S2A). In addition, a circular network diagram was generated to visually represent the primary BP (Figure S2B), MF (Figure S2C), and KEGG pathways (Figure S2D) enriched with thirteen TRDEGs.

Mutation analysis of TRDEGs

We tabulated the mutation analysis results for thirteen TRDEGs and visualized them using the “maftools” package. There were seven main mutation types: missense mutations, nonsense mutations, frameshift deletion (DEL) mutations, frameshift insertion (INS) mutations, in-frame

DEL, in-frame INS, and splice sites, of which missense mutations accounted for the majority (Figure 3A). In addition, the predominant variant types were SNPs, with a minority represented by INS and DEL. Notably, C > T was the most common single nucleotide variant, followed by C > A and T > C (Figure 3A). Among the thirteen TRDEGs, *ACACB* exhibited the highest SNP frequency, featuring six primary mutation types, with the missense mutation being the most prevalent. We displayed the mutation information using a waterfall plot (Figure 3B).

Subsequently, we analyzed the CNV for the thirteen TRDEGs. Following the acquisition and merging of CNV data, we subjected the combined dataset to GISTIC 2.0 and visualized the results (Figure 3C, 3D). Our analysis revealed a large number of amplifications and DEL among the thirteen TRDEGs, with *TPX2* exhibiting the highest amplification, followed by *FABP4*, while *NAT2* accounted

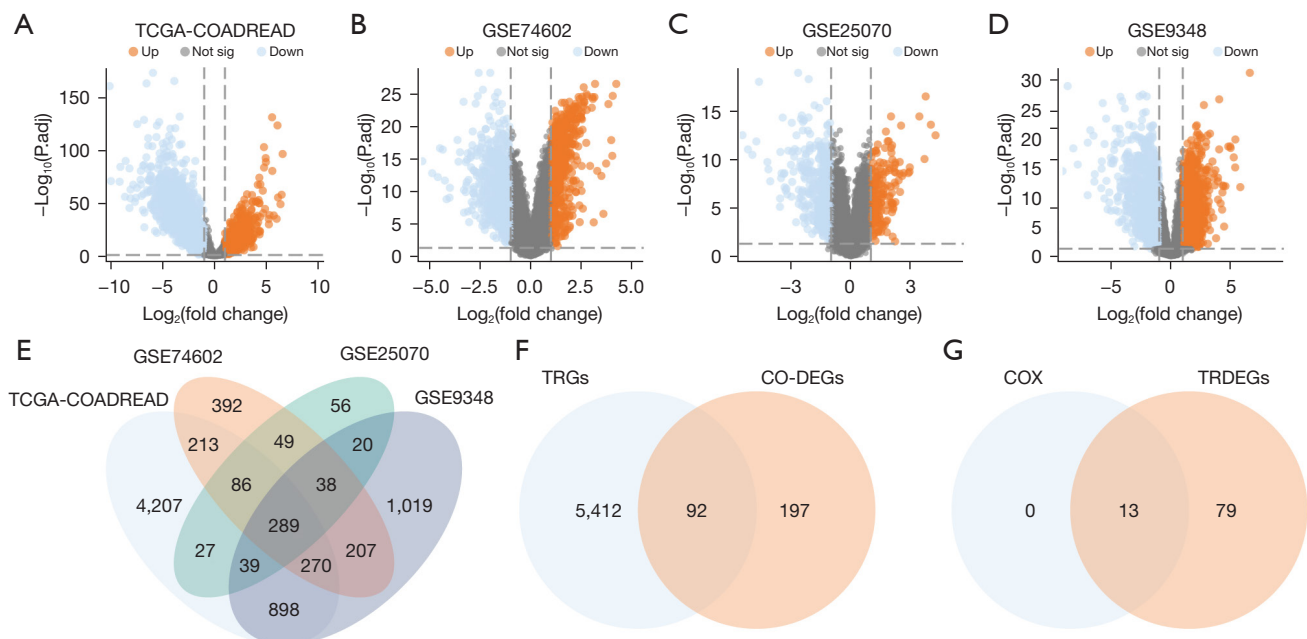


Figure 2 Identification of telomere-related differentially expressed genes. Differential analysis volcano maps among (A) TCGA-COADREAD, (B) GSE74602, (C) GSE25070, (D) GSE9348. (E) Venn diagrams of differentially expressed genes among cancer control groups in the four datasets. (F) Venn diagram of CO-DEGs and TRGs. (G) Venn diagram of TRDEGs after the screening of univariate Cox regression analysis. TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; TRGs, telomere-related genes; CO-DEGs, common differentially expressed genes; TRDEGs, telomere-related differentially expressed genes.

for the majority of DEL.

Construction of Tscore in the TCGA-COADREAD dataset

We applied the ssGSEA algorithm to the thirteen identified TRDEGs in TCGA, deriving the Tscore for each patient, and subsequently categorized them into high- and low-scoring groups based on the median score. To identify the DEGs associated with the Tscore, we conducted a differential analysis using the “limma” package, comparing groups with high and low scores (Figure S3A). A visual representation of the expression patterns of the thirteen TRDEGs in the two groups was generated using the “pheatmap” package (Figure S3B).

Our study comprehensively analyzed clinical factors between the two groups, revealing no discernible differences (Figure S3C-S3F). Additionally, a Laplace correlation graph was generated to illustrate the correlation between the Tscore and TRDEGs (Figure S3G), highlighting strong correlations of *MMP1*, *MMP10*, and *MMP3* with the Tscore.

WGCNA-identified co-expressed modules in the TCGA-COADREAD dataset

The co-expression modules were screened via WGCNA for all genes within the TCGA dataset. During the WGCNA analysis, we initially selected genes with variance in the top 40% of all genes as the input genes and utilized the clustering tree to group genes into high- and low-scoring categories. The screening criteria were set to 0.85 to ascertain the optimal power threshold (Figure 4A), and the genes between the two groups were assigned to the MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, and MEmagenta modules (Figure 4B,4C). The combined cut height of the modules was adjusted to 0.2. Subsequently, modules with a combined cut height below 0.2 were identified, merged, and re-clustered (Figure 4B). The association between the genes and the resulting new modules is depicted (Figure 4C). Based on the expression patterns of the module genes and grouping information, we identified twelve modules (MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, MEmagenta, and MEmagenta).

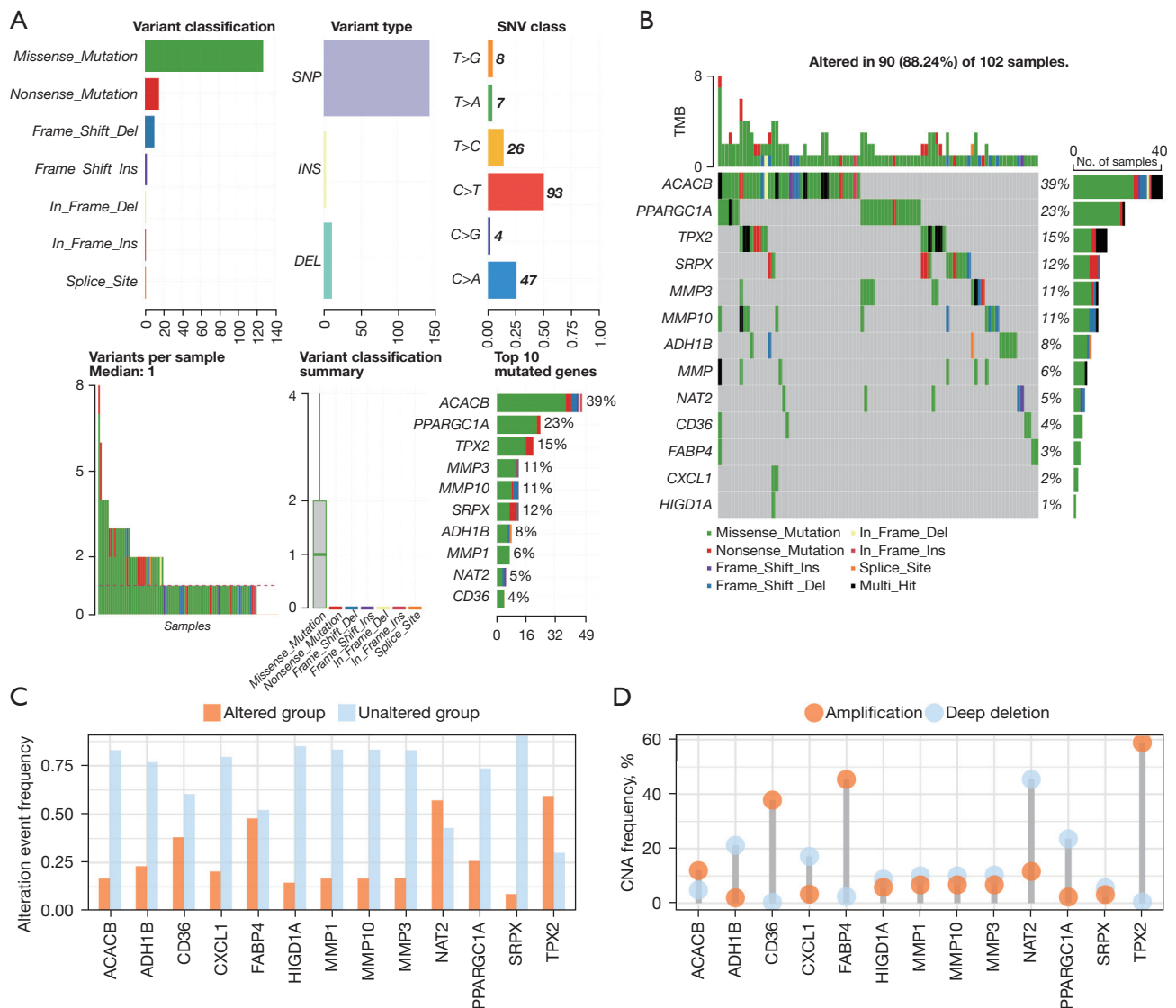


Figure 3 Mutation analysis of TRDEGs. (A) SNP status of TRDEGs in the TCGA-COADREAD dataset. (B) Waterfall plot of the proportion of SNPs in TRDEGs. (C,D) CNV status of TRDEGs. DEL, deletion; INS, insertion; SNP, single nucleotide polymorphism; SNV, single nucleotide variant; TMB, tumor mutation burden; CNA, copy number alteration; TRDEGs, telomere-related differentially expressed genes; TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; CNV, copy number variation.

MEblack, MEBrown, MEpink, METurquoise, MEyellow, MEBLue, MEGreen, MEGreen-yellow, and MEGrey) and determined the correlation between the control and COADREAD groups (Figure 4D).

We included genes from the eleven modules (excluding the MEGray module) with notable changes ($P < 0.05$) for subsequent analysis. After excluding modules with no intersection with TRGs, we determined the intersection of the thirteen TRDEGs with the genes in the MEBLue (Figure 4E),

MEyellow (Figure 4F), MEpink (Figure 4G), and MEBrown modules (Figure 4H) and constructed a Venn diagram to identify the module genes. Ultimately, eight module genes (*MMP1*, *MMP3*, *NAT2*, *MMP10*, *SRPX*, *PPARGC1A*, *HIGD1A*, and *CXCL1*) were identified, as shown.

Subtype identification and prognosis analysis

Based on the expression levels of the eight TRDEGs

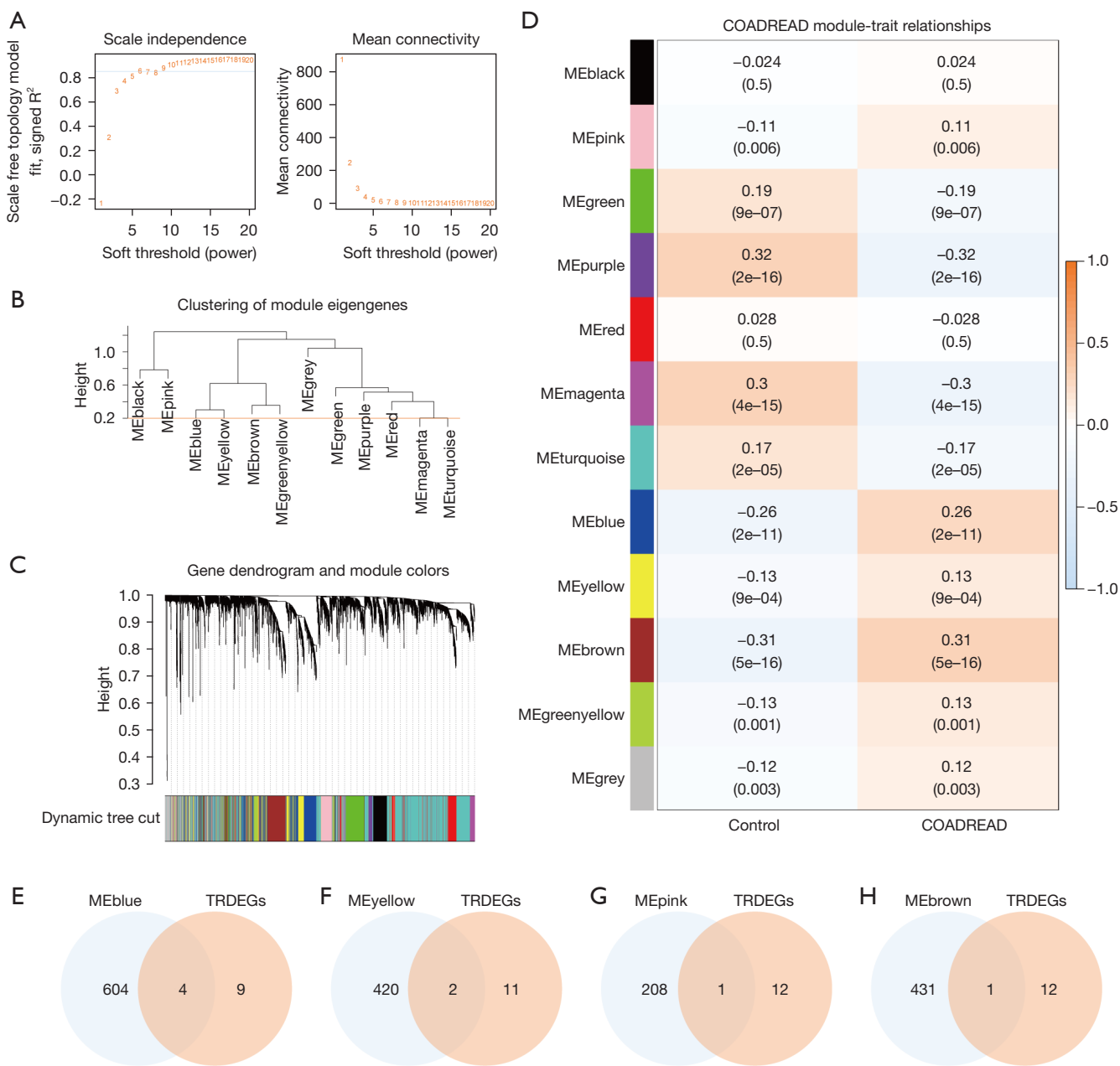


Figure 4 WGCNA identified co-expressed modules in the TCGA-COADREAD dataset. (A) Module screening threshold scale-free network. (B) Display of gene module aggregation results. (C) The corresponding relationship between genes and modules. (D) Correlation analysis between the clustering modules and different groups. (E-H) Venn diagrams of TRDEGs: (E) MEblue module, (F) MEyellow module, (G) MEpink module, (H) MEbrown module. COADREAD, colorectal adenocarcinoma; TRDEGs, telomere-related differentially expressed genes; WGCNA, weighted gene co-expression network analysis; TCGA, The Cancer Genome Atlas.

identified by WGCNA, we conducted a consensus clustering method to delineate distinct subtypes, ultimately identifying two subtypes (Figure 5A). Cluster 1 comprised 317 samples, while cluster 2 contained 327 samples.

Additionally, we present the consensus clustering cumulative distribution function (CDF) graph (Figure 5B) and the delta graph of the area under the CDF curve (Figure 5C) for varying cluster numbers. As shown in the figure, the TCGA-COADREAD dataset exhibited the most consistent clustering results when $k=2$ for unsupervised clustering. Principal component analysis was performed on samples from clusters 1 and 2, revealing significant differences between the two subtypes (Figure 5D). A heat map and a grouping comparison chart were generated to illustrate differences in the expression of the eight TRDEGs between the two subtypes (Figure 5E, 5F). The findings indicated a notable disparity in the expression levels of seven TRDEGs (*MMP1*, *MMP3*, *MMP10*, *SRPX*, *PPARGC1A*, *HIGD1A*, and *CXCL1*) between the two subtypes, with *PPARGC1A* upregulated in cluster 1 and the remaining six upregulated in cluster 2.

Finally, we combined the overall survival (OS) information to assess the disparity in survival rates between the two clusters, visualizing the data using a Kaplan-Meier curve (Figure 5G). The results demonstrated a significant difference in prognosis between the two subtypes ($P=0.001$).

Construction and validation of TRDEG's prognostic model

To assess the prognostic significance of the thirteen TRDEGs, a prognostic model was developed using LASSO regression (Figure 6A). Based on the minimum partial likelihood bias, ten prognostic TRDEGs (*ACACB*, *TPX2*, *SRPX*, *PPARGC1A*, *CD36*, *MMP3*, *NAT2*, *MMP10*, *HIGD1A*, and *MMP1*) were retained as diagnostic markers for CRC. Additionally, the results of the LASSO regression were visualized using a variable trajectory map (Figure 6B). To validate the prognostic significance of the ten selected TRDEGs, a risk score algorithm was generated, categorizing samples into high- and low-risk groups based on the median and depicted in a risk-factor graph (Figure 6C).

$$\begin{aligned} \text{Risk score} = & ACACB * 0.193041902 + TPX2 * (-0.194305398) \\ & + SRPX * 0.026934249 + PPARGC1A * (-0.294030267) \\ & + CD36 * 0.236316955 + MMP3 * (-0.041996492) \quad [2] \\ & + NAT2 * (-0.165504255) + MMP10 * (-0.039274203) \\ & + HIGD1A * (-0.150787603) + MMP1 * (-0.091522228) \end{aligned}$$

Subsequently, a Kaplan-Meier curve was constructed

based on patient survival information (OS) for the two risk groups, revealing a more statistically significant difference ($P<0.001$) in survival prediction than ssGSEA and WGCNA (Figure 6D). Subsequent analyses were carried out based on the LASSO results. The efficacy of this risk model in predicting prognosis was maintained in the validation cohort (Figure 6E). Additionally, differential analysis was conducted to compare the two groups (Figure 6F). A comparative analysis of the clinical characteristics highlighted that the high-risk group exhibited a more advanced tumor node metastasis (TNM) stage compared to the low-risk group (Figure S4A-S4D). Moreover, a correlation Laplace graph was generated between the risk score and TRDEGs (Figure S4E), revealing a significant association between the risk score and the expression levels of *MMP1*, *MMP3*, and *MMP10*.

Differential, correlation, and diagnostic analysis of prognostic TRDEGs

The expression differences of the ten prognostic TRDEGs were examined in TCGA and verified in GSE74602, GSE25070, and GSE9348 (Figure S5A-S5D). In addition, the expression trends of the ten prognostic TRDEGs in the four datasets exhibited consistency: *ACACB*, *CD36*, *HIGD1A*, *NAT2*, *PPARGC1A*, and *SRPX* were downregulated in the cancer group, whereas *MMP1*, *MMP10*, *MMP3*, and *TPX2* were upregulated in the cancer group.

Correlations between the ten prognostic TRDEGs in TCGA (Figure S5E), GSE74602 (Figure S5F), GSE25070 (Figure S5G), and GSE9348 (Figure S5H) were further analyzed. The results indicated that the correlation between prognostic TRDEGs in the four datasets was relatively consistent, with *MMP3* and *MMP1* showing the strongest correlation. Diagnostic ROC curves of the ten TRDEGs are also presented in the TCGA dataset, demonstrating excellent diagnostic performance for all patients (Figure S6A-S6J).

GSEA between high- and low-risk groups

GSEA was conducted to assess differences among subgroups identified by the prognostic TRDEGs model in terms of gene functions and pathways. Criteria for significant enrichment were set at $P<0.05$ and $FDR<0.05$. The results were visually represented using a mountain map (Figure 7A). The analysis demonstrated notable enrichment in genes associated with cytokines and inflammatory response

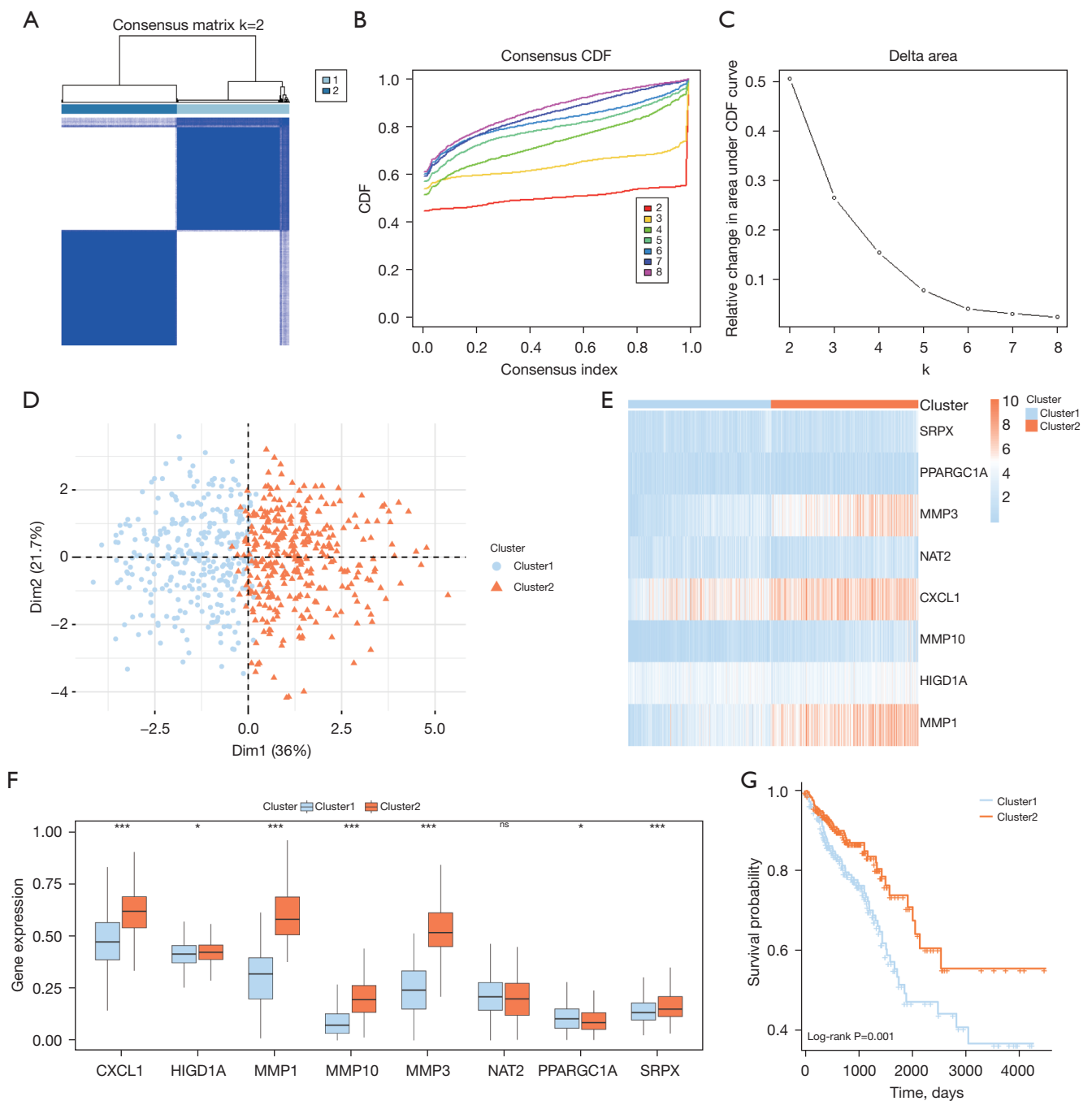


Figure 5 Construction of COADREAD related subtypes. (A) Consensus clustering results (k=2). (B) Consensus clustering CDF plot. (C) Delta plot of the area under the CDF curve for different numbers of clusters in the consensus clustering. (D) PCA between the two clusters. (E) Heat map of TRDEGs between two subtypes. (F) Grouping comparison chart of TRDEGs between two subtypes. (G) The results of the prognostic Kaplan-Meier curve of the two subtypes. ns, not statistically significant; *, P<0.05; ***, P<0.001. CDF, cumulative distribution function; COADREAD, colorectal adenocarcinoma; PCA, principal component analysis; TRDEGs, telomere-related differentially expressed genes.

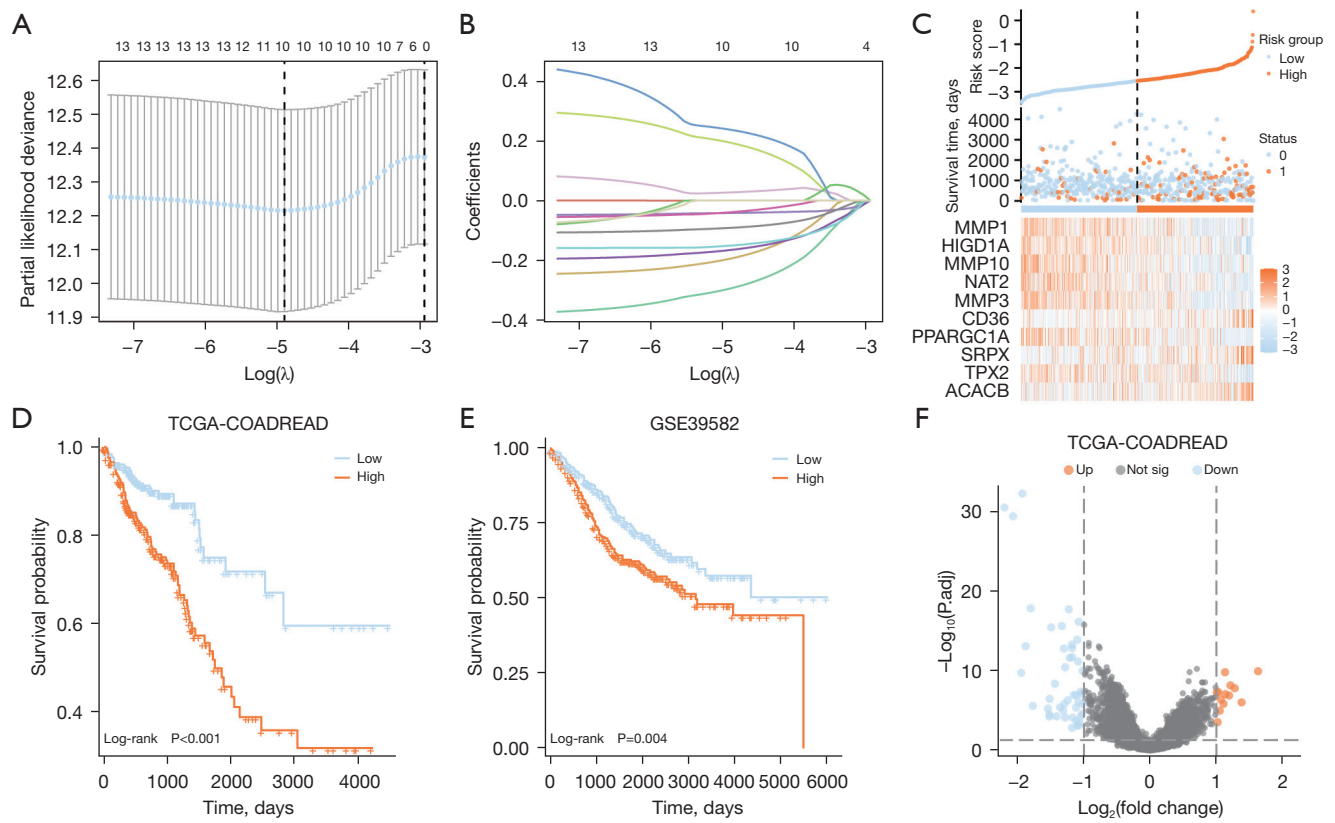


Figure 6 Construction and validation of TRDEGs prognostic model. (A) Ten prognostic TRDEGs were identified by the LASSO regression model. (B) Variable trajectory plot of thirteen TRDEGs. (C) Risk factor plot. (D) The Kaplan-Meier curve between the high- and low-risk groups in TCGA cohort. (E) The Kaplan-Meier curve between the high- and low-risk groups in GSE39582 (validation cohort). (F) Volcano plot of difference analysis between high- and low-risk groups. TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; TRDEGs, telomere-related differentially expressed genes; LASSO, least absolute shrinkage and selection operator.

(Figure 7B), programmed cell death (Figure 7C), IL-23 pathway (Figure 7D), IL-17 pathway (Figure 7E), extension of telomeres (Figure 7F), regulation of TP53 activity through phosphorylation (Figure 7G), and other pathways (Table S2 provides detailed information).

Construction of the PPI, mRNA-miRNA, mRNA-TF, and mRNA-RBP interaction networks

To begin, the STRING database facilitated the construction of the PPI network for TRDEGs (Figure 8A). Subsequently, miRNA data were employed to predict interaction relationships between miRNAs and the ten prognostic TRDEGs, with interaction relationships screened based on “pancancerNum >10”. The resultant mRNA-miRNA interaction network was visualized using Cytoscape software (Figure 8B). Comprising seven mRNAs (*CD36*,

HIGD1A, *MMP1*, *MMP3*, *PPARGC1A*, *SRPX*, and *TPX2*) and 33 miRNA molecules, the network formed 57 pairs of mRNA-miRNA interactions (Table S3). Furthermore, CHIPBase and the hTFtarget databases were used to identify TFs interacting with the ten prognostic TRDEGs recorded in both databases. This yielded 78 interaction relationships involving seven mRNAs (*ACACB*, *HIGD1A*, *MMP1*, *MMP10*, *PPARGC1A*, *SRPX*, and *TPX2*) and 58 TFs, visualized using Cytoscape software (Figure 8C; details in Table S4). The ENCORI database was then employed to predict RBP interacting with prognostic TRDEGs, and Cytoscape was used for visualizing the mRNA-RBP interaction network (Figure 8D). This network comprised 79 pairs of mRNA-RBP interactions involving six TRDEGs and 49 RBP molecules (Table S5).

The AlphaFold Protein Structure Database comprised approximately 350,000 protein structures predicted using

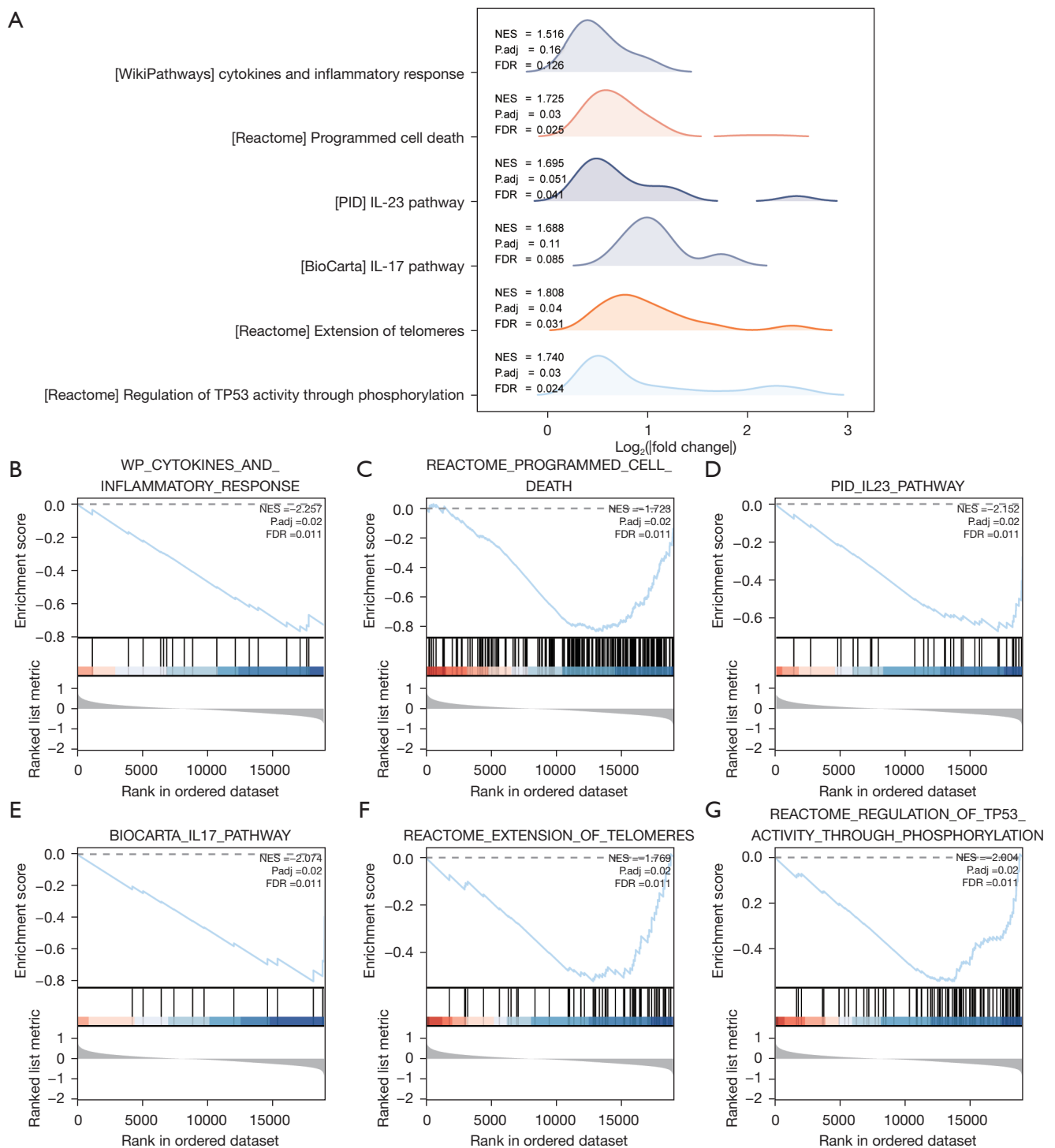


Figure 7 GSEA between high- and low-risk groups. (A) The six most important biological characteristics of GSEA enrichment analysis. Genes in TCGA-COADREAD dataset are significantly enriched in (B) cytokines and inflammatory response, (C) programmed cell death, (D) IL-23 pathway, (E) IL-17 pathway, (F) extension of telomeres, (G) regulation of TP53 activity through phosphorylation. NES, normalized enrichment score; FDR, false discovery rate; GSEA, gene set enrichment analysis; TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; IL, interleukin.

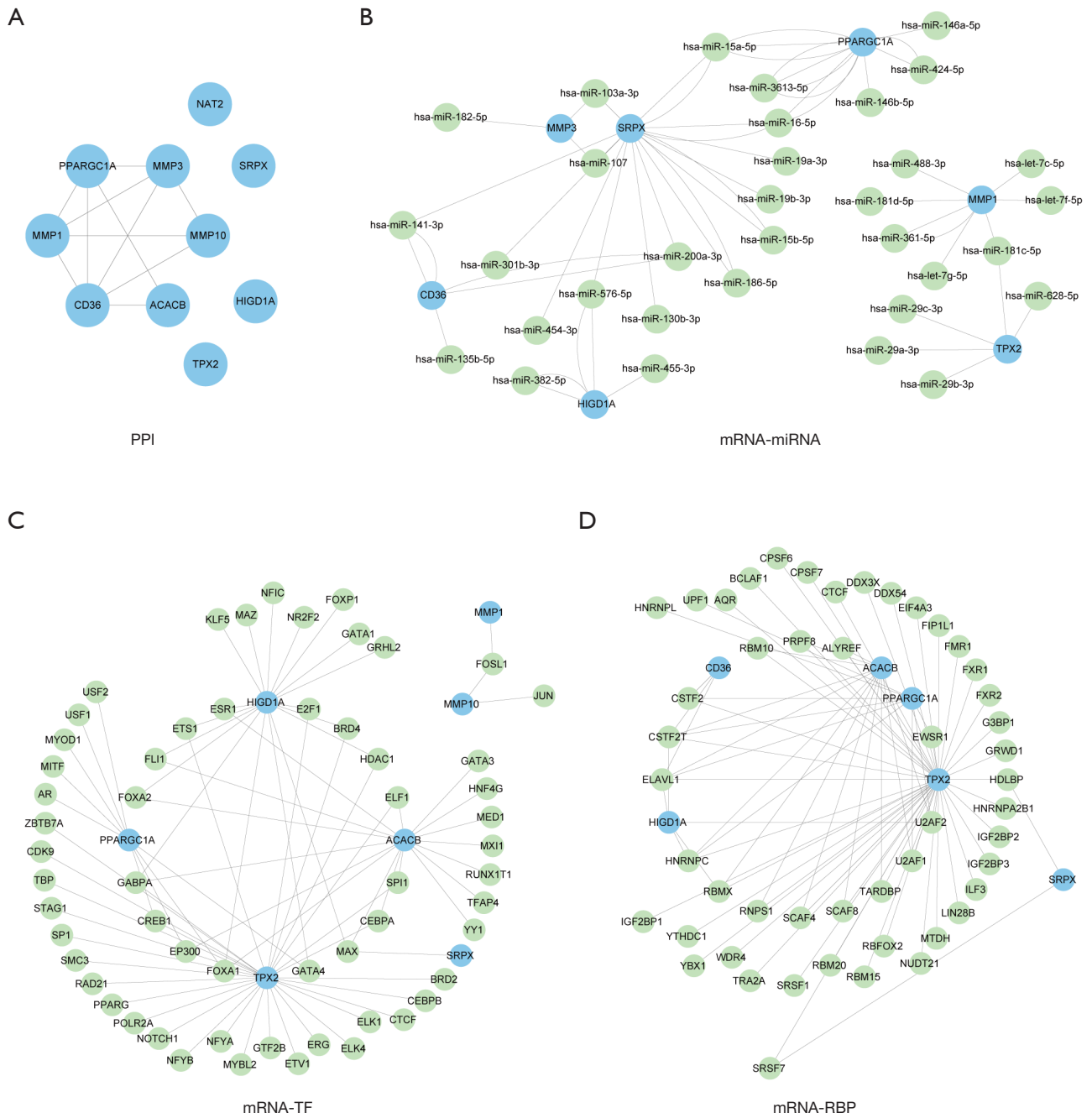


Figure 8 Construction of PPI, mRNA-miRNA, mRNA-TF and mRNA-RBP interaction network. (A) PPI network. (B) mRNA-miRNA, (C) mRNA-TF, (D) mRNA-RBP interaction network. The blue circles represent mRNAs; the green circles represent miRNAs, TFs and RBPs respectively. PPI, protein-protein interaction; TF, transcription factor; RBP, RNA-binding protein.

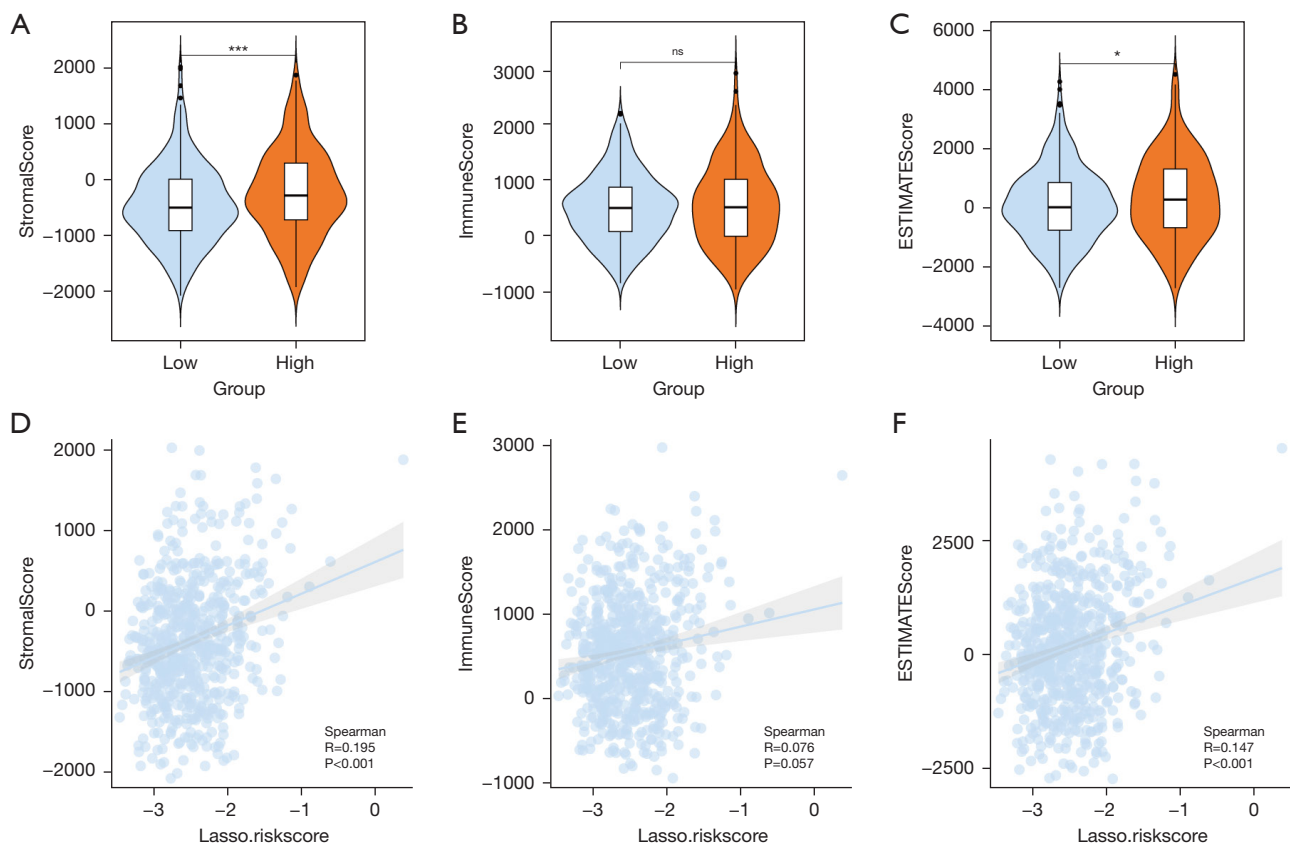


Figure 9 Immune score analysis. (A) Stromal score, (B) immune score, (C) ESTIMATE score violin diagrams between high- and low-risk groups. The scatter diagram showing the correlation between the risk score: (D) stromal score, (E) immune score, (F) ESTIMATE score. ns, not statistically significant; *, $P < 0.05$; ***, $P < 0.001$. The absolute value of the correlation coefficient (R) > 0.8 , strongly correlated; $0.5 < R < 0.8$, moderately correlated; $0.3 < R < 0.5$, weakly correlated; $R < 0.3$, irrelevant. ESTIMATE, Estimation of STromal and Immune cells in Malignant Tumour tissues using Expression data.

the AlphaFold AI system, successfully predicting the structure of 98.5% of the human proteome. The AlphaFold website was utilized to analyze the protein structures of the prognostic TRDEGs (Figure S7).

Immune score, ssGSEA, and CIBERSORT analysis

The “estimate” package was employed to compute several immune and matrix scores. The results were visualized by violin plots (Figure 9) which revealed a significant disparity in stromal (Figure 9A) and ESTIMATE scores (Figure 9C) between the two groups, whereas the immune score (Figure 9B) exhibited no significant difference. In addition, correlation scatter diagrams depicting the relationship between immunity and risk scores were generated (Figure 9D-9F).

To investigate the variance in immune infiltration between the high- and low-risk groups, we assessed the abundances of 28 immune cell types using ssGSEA (Figure 10A). These results indicated significant differences in eleven immune cell types, including activated CD4 T cells, CD56dim natural killer cells, effector memory CD4⁺ T cells, immature dendritic cells, memory B cells, natural killer cells, neutrophils, plasmacytoid dendritic cells, T follicular helper cells, type 17 T helper cells, and type 2 T helper cells. The Spearman statistical algorithm was then employed to calculate the correlation between the infiltration abundance of the eleven immune cell types (Figure 10B, 10C). In the low-risk group, a significant positive association was observed across the majority of the eleven immune cells, with the strongest link found between natural killer cells and T follicular helper cells (Figure 10B).

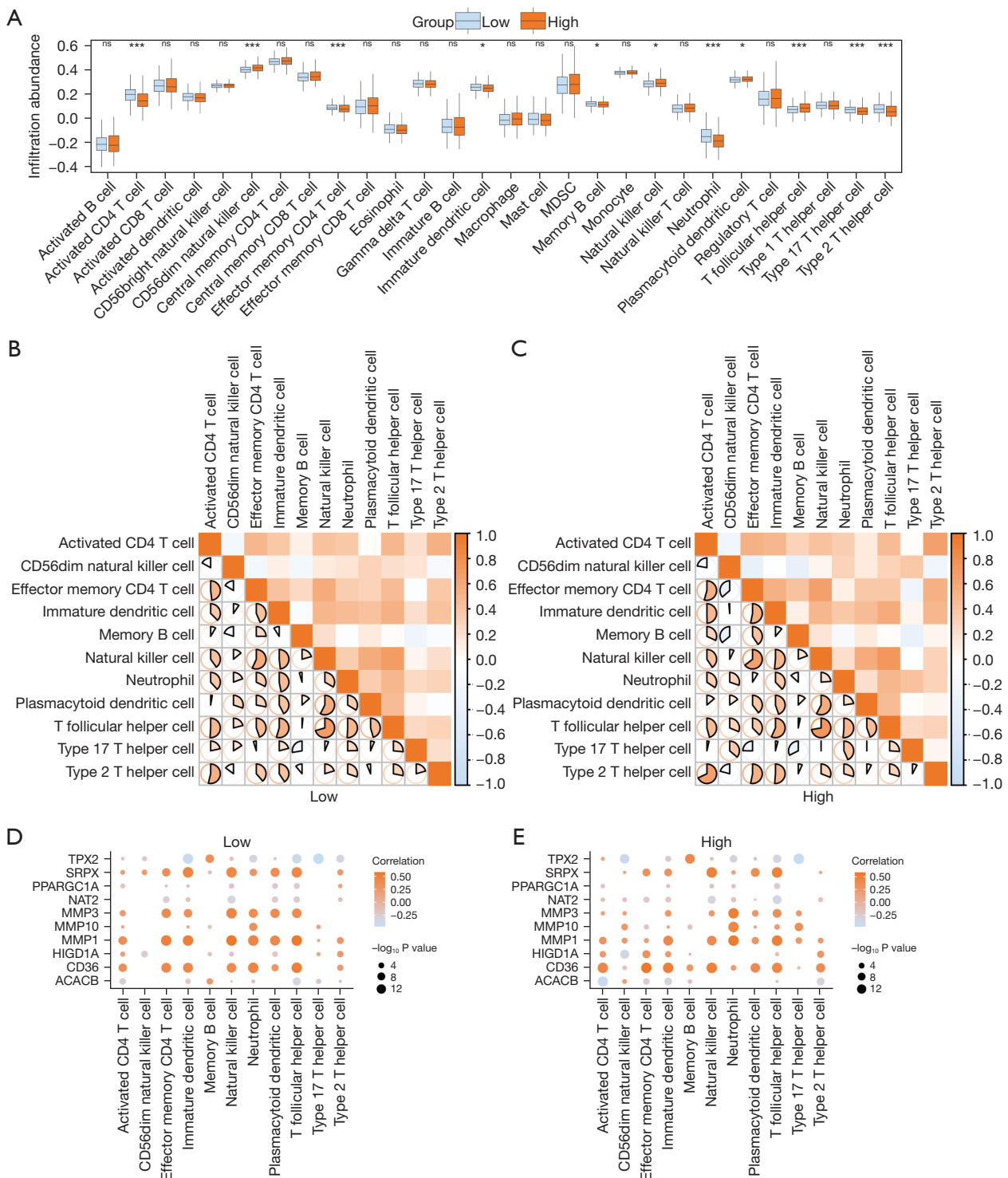


Figure 10 ssGSEA immune infiltration analysis. (A) The group comparison chart of the ssGSEA immune infiltration results between the high- and low-risk groups in the TCGA-COADREAD dataset. Correlation analysis between immune cells in (B) low-risk group and (C) high-risk group. Correlation dot plots of immune cells and prognostic TRDEGs in (D) low-risk group and (E) high-risk group. ns, not statistically significant; *, $P < 0.05$; ***, $P < 0.001$. MDSC, myeloid-derived suppressor cell; ssGSEA, single-sample gene-set enrichment analysis; TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; TRDEGs, telomere-related differentially expressed genes.

Similar results were obtained for the high-risk group (Figure 10C). Furthermore, we analyzed the association between the infiltration of eleven immune cells and the expression patterns of ten TRDEGs (Figure 10D,10E). In the low-risk group, a significant positive correlation between *CD36*, *MMP1*, *MMP3*, and immune cells was observed, with the strongest correlation between natural killer cells and *MMP1* (Figure 10D). In the high-risk group, the number of significant positive and negative correlations between immune cells and genes was similar, with the correlation between effector memory CD4⁺ T cells and *CD36* being the strongest (Figure 10E).

Additionally, the CIBERSORT algorithm was applied to conduct an analysis similar to ssGSEA (Figure 11A). The results demonstrated that twelve immune cell types showed a statistically significant difference (Figure 11B). In the low-risk group, the correlation between neutrophils and *MMP1* was strongest (Figure 11C). In the high-risk group, the correlation between macrophages M2 and *CD36* expression was the most robust (Figure 11D).

Immunotherapy response and drug sensitivity prediction

We evaluated the immunotherapy sensitivity of high- and low-risk groups using the TIDE algorithm (Figure 12A), revealing a considerable disparity in responsiveness to immunotherapy. Analysis of TMB and MSI indicated no statistically significant differences in TMB (Figure 12B) and MSI (Figure 12C). Correlation scatter plots were generated to show the relationship between TIDE (Figure 12D), TMB (Figure 12E), MSI (Figure 12F), and the risk score. As expected, the results were consistent with the group comparison plots.

To examine the correlation between drug sensitivity and risk score, we employed drug response-related datasets, specifically GDSC and CTRP. The findings indicated that the low-risk group exhibited heightened sensitivity to dabrafenib, lapatinib, camptothecin, docetaxel, and telomerase inhibitor IX compared to the high-risk group. Conversely, patients in the high-risk group demonstrated increased sensitivity to paclitaxel and regorafenib. However, the difference in drug sensitivity for 5-fluorouracil, irinotecan, and oxaliplatin between the two groups was not statistically significant (Figure 13).

Construction and validation of a clinical prognostic model

To assess the prognostic significance of the model, we

integrated it with clinical variables for univariate Cox regression analysis (Figure 14A). Variables meeting the criterion of $P < 0.1$ were included in the multivariate Cox regression analysis to develop a clinically relevant prognostic model (refer to Table 1 for clinical baseline data and Table 2 for Cox regression results). A nomogram analysis, based on the multivariate Cox regression results, was performed to determine the prognostic capability of the clinical prognostic model (Figure 14B). The results indicated that the T stage, N stage, M stage, and risk score contributed to the clinical prognostic models, with the T stage demonstrating the highest utility. In addition, prognostic calibration analyses for 1-year (Figure 14C), 3-year (Figure 14D), and 5-year (Figure 14E) periods were conducted, illustrating excellent predictive ability of the clinical prognostic model. To assess the clinical efficacy, DCA was performed at 1 (Figure 14F), 3 (Figure 14G), and 5 years (Figure 14H), respectively.

The mRNA and protein expression levels validate the telomere-related signature

The expression patterns of the ten genes comprising the risk model were validated through qRT-PCR, revealing statistically significant differences. In CRC cell lines, *ACACB*, *HIGD1A*, *NAT2*, *PPARGC1A*, and *TPX2* exhibited elevated expression levels, while *CD36*, *SRPX*, *MMP1*, *MMP3*, and *MMP10* displayed decreased levels (Figure 15A). Furthermore, the HPA database reported IHC images of seven proteins (Figure 15B).

Discussion

CRC is the third most prevalent cancer globally, characterized by high morbidity and mortality rates. TNM classification is a traditionally accurate prognostic tool but unavoidably has its limits. Advances in genomics and precision medicine have catalyzed the development of molecular signatures reliant on gene expression levels to forecast clinical outcomes. Using risk stratification in individuals with CRC enables the precise prediction of survival outcomes with a high degree of certainty (53). Although TNM staging remains the most reliable approach for assessing patient prognosis and directing treatment, improvements to the prognostic model are feasible through an evolving comprehension of the molecular mechanisms governing CRC onset and progression.

To the best of our knowledge, this is the first study to

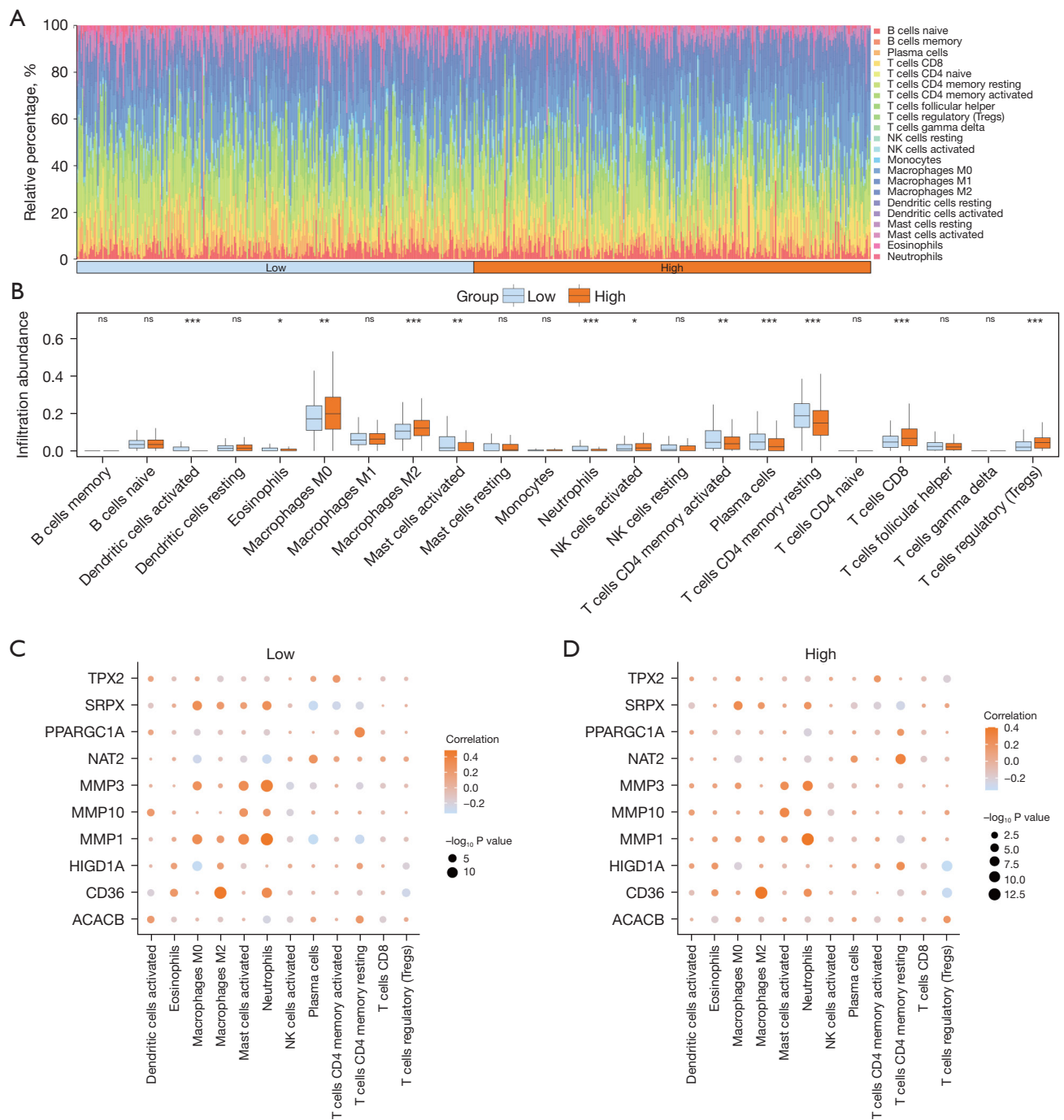


Figure 11 CIBERSORT immune infiltration analysis. The CIBERSORT immune infiltration analysis results of 22 immune cells between the high- and low-risk groups of the TCGA-COADREAD dataset are displayed in (A) a stacked histogram, and (B) a group comparison chart. Correlation dot plots between immune cells and prognostic TRDEGs in the (C) low-risk group and (D) high-risk group. ns, not statistically significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. NK, natural killer; TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; TRDEGs, telomere-related differentially expressed genes.

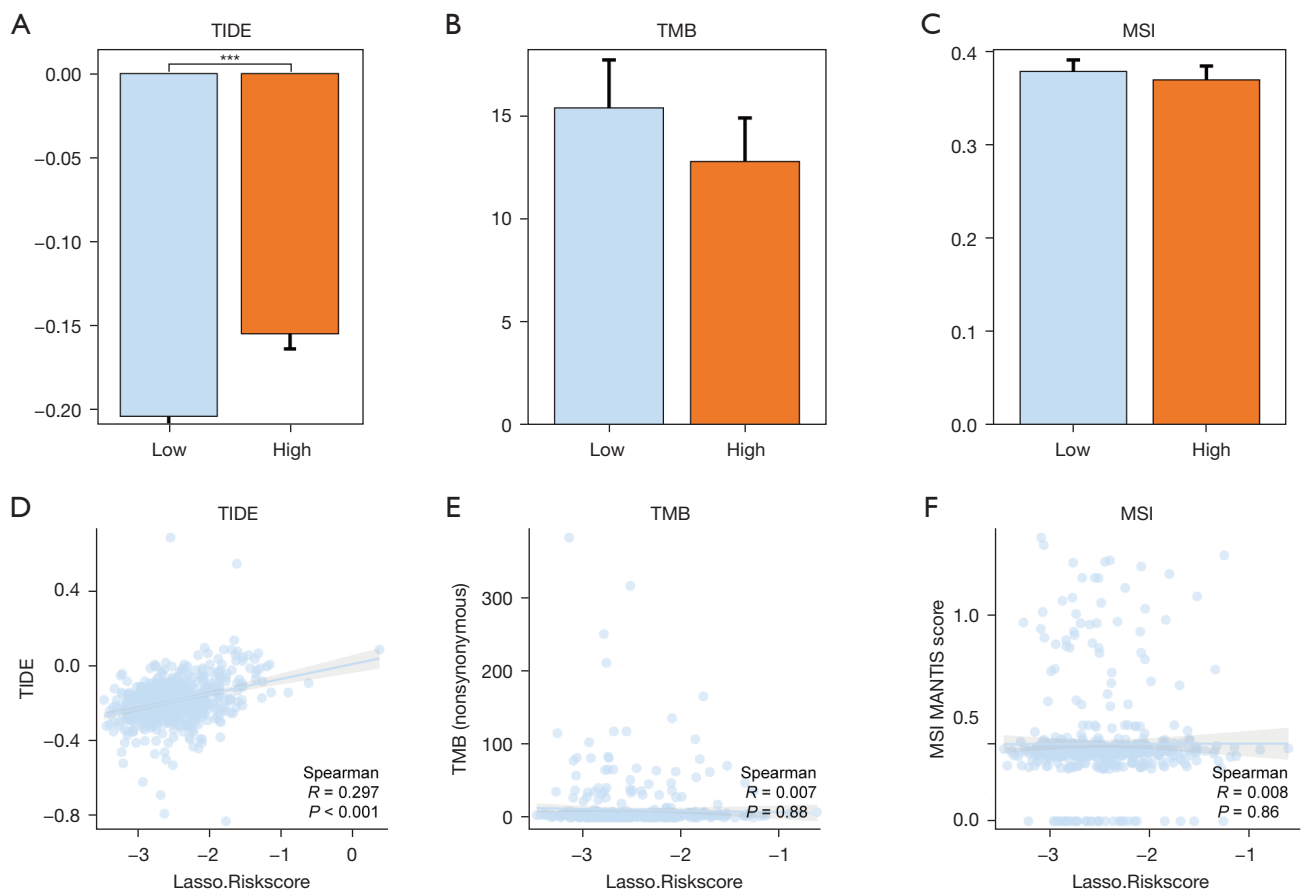


Figure 12 TIDE, TMB, MSI between high- and low-risk groups. The group comparison diagrams of the difference between the high- and low-risk groups of (A) TIDE, (B) TMB, (C) MSI. Correlation scatter plots between (D) TIDE, (E) TMB, (F) MSI and the risk score. ***, $P < 0.001$. The absolute value of the correlation coefficient (R) > 0.8 , strongly correlated; $0.5 < R < 0.8$, moderately correlated; $0.3 < R < 0.5$, weakly correlated; $R < 0.3$, irrelevant. TIDE, tumor immune dysfunction and exclusion; TMB, tumor mutation burden; MSI, microsatellite instability.

investigate telomere-related markers in patients with CRC. A prognostic model, leveraging TRGs, was formulated to predict outcomes for patients categorized into distinct risk groups. This model was employed to predict immune cell infiltration and drug sensitivity in CRC. Finally, this model was integrated with clinical features to construct a composite nomogram, exhibiting a high degree of prognostic prediction accuracy for CRC patients. DCA curves underscored the excellent clinical benefits for patients with CRC.

Previous studies have demonstrated variations in tumor prognosis contingent upon immune status. Our findings revealed that the low-risk group exhibited a higher proportion of activated $CD4^+$ T cells, effector memory $CD4^+$ T cells, and memory B cells, alongside a lower

proportion of macrophage M2 and regulatory T cells (Tregs) compared to the high-risk group. This discrepancy may elucidate the superior outcomes observed in the low-risk group. Furthermore, patients classified in the high-risk group exhibited higher TIDE scores, indicating an enhanced susceptibility to immune evasion. Overall, these findings suggest a potential contributory role of immune status in the prognostic variations between the high- and low-risk groups.

GO and KEGG enrichment analyses revealed prominent enrichment of TRDEGs in the regulation of FA oxidation, long-chain FA transporter activity, the adipocytokine signaling pathway, the PPAR signaling pathway, and the AMPK signaling pathway. Dysregulated lipid metabolism has been established as a significant metabolic aberration

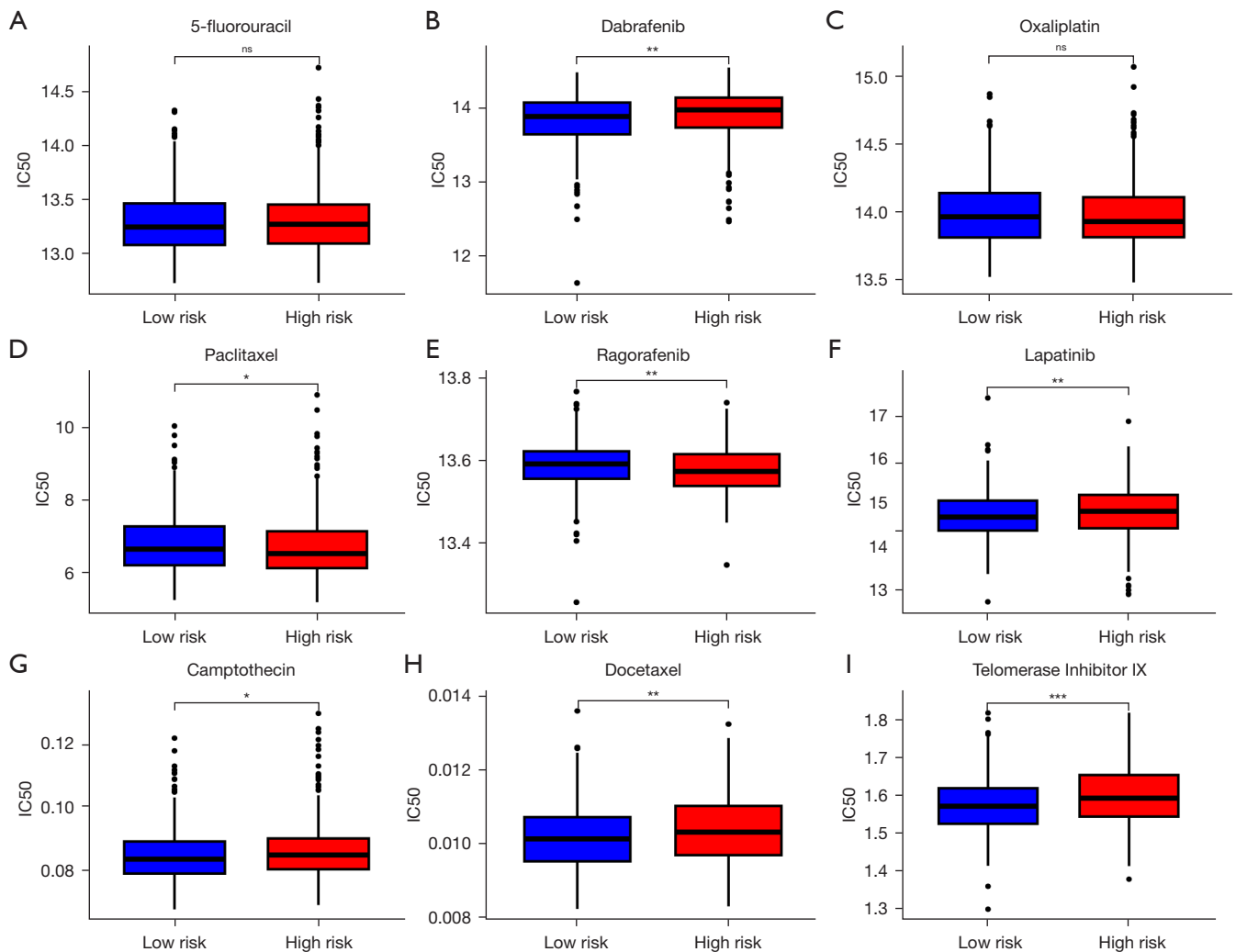


Figure 13 Drugs' sensitivity between high- and low-risk groups. Differential analysis of IC50 for drugs in CTRP (A-E) and GDSC (F-I). ns, not statistically significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. IC50, half maximal inhibitory concentration; CTRP, The Cancer Therapeutics Response Portal; GDSC, Genomics of Drug Sensitivity in Cancer.

in cancer, as demonstrated by previous studies. The dynamic modulation of nutrient availability within the tumor microenvironment necessitates the utilization of lipid metabolism by cancer cells to acquire energy, biofilm constituents, and signaling molecules essential for their proliferation, survival, invasion, metastasis, and responses to both the tumor microenvironment and cancer treatment (54). Moreover, GSEA revealed a notable enrichment of TRDEGs in cytokines and inflammatory response, the IL-17 pathway, and the IL-23 pathway. A previous study confirmed the crucial role of inflammation in tumor progression, with the involvement of the IL-23/IL-17 axis suggested in the pathogenesis of various

inflammatory disorders, including inflammatory bowel disease, recognized as a precancerous lesion in CRC (55). In summary, these findings underscore the significant roles of lipid metabolism and chronic inflammation in CRC pathogenesis.

In this study, we examined the predictive accuracy of TRG signatures in estimating CRC prognosis and their potential impact on drug sensitivity. This was achieved by combining the expression profiles of TRGs with the LASSO analysis method, a proven effective approach to identifying potential biomarkers. Ultimately, a prognostic risk model consisting of 10 genes (*ACACB*, *TPX2*, *SRPX*, *PPARGC1A*, *CD36*, *MMP3*, *NAT2*, *MMP10*, *HIGD1A*,

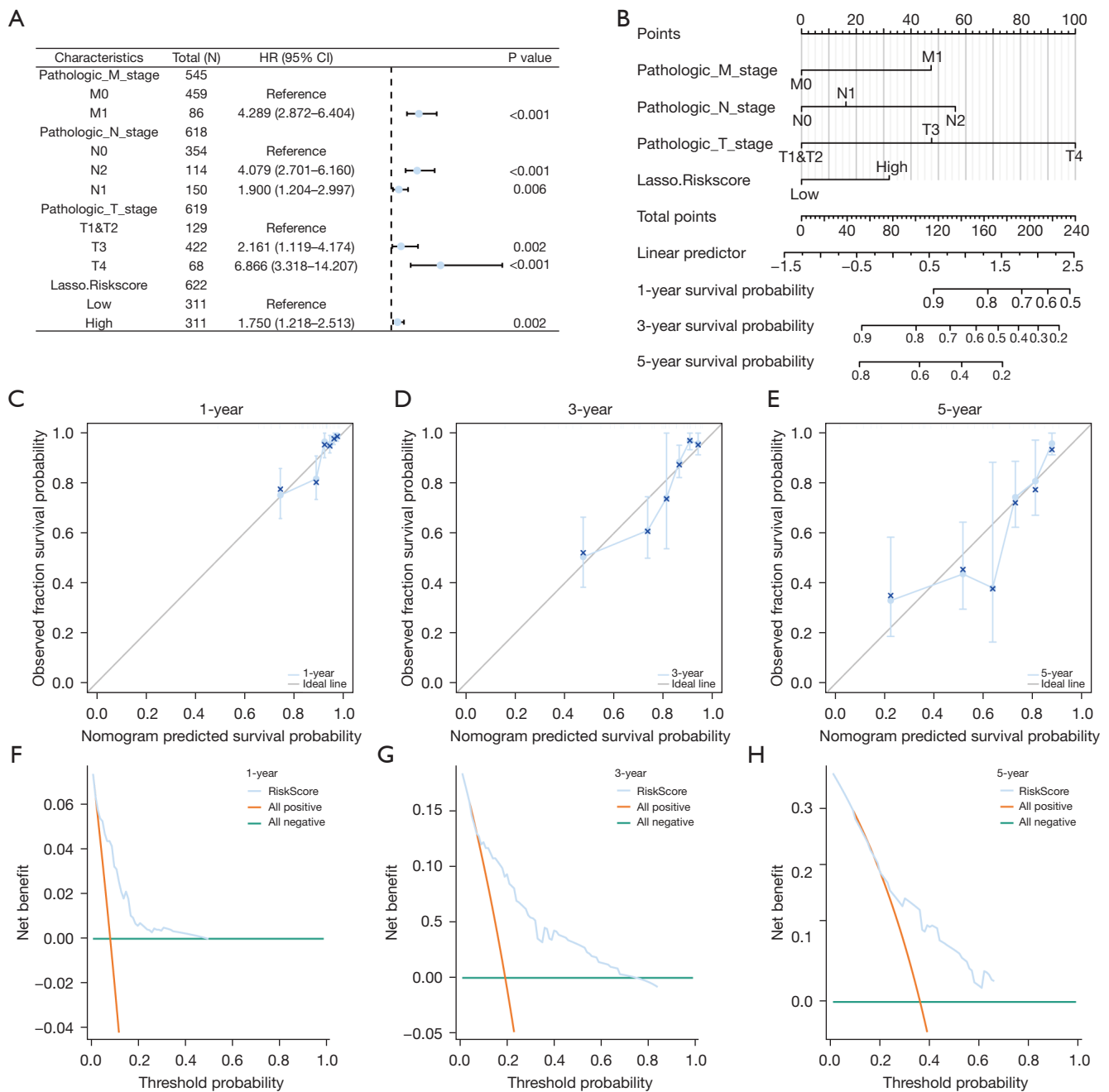


Figure 14 Construction and validation of clinical prognostic model. (A) Forest plot of the univariate Cox regression analysis. (B) Nomogram of clinical prognostic model. (C) 1-year, (D) 3-year, (E) 5-year calibration curves for the nomogram analysis. (F) 1-year, (G) 3-year, (H) 5-year DCA plots of clinical prognostic model. CI, confidence interval; HR, hazard ratio; DCA, decision curve analysis.

and *MMP1*) was constructed. *TPX2*, a pivotal mediator of microtubule biology during cell division and in postmitotic cells, contributes to the response to DNA damage (56). In CRC, *TPX2* is upregulated and plays an oncogenic role in CRC pathogenesis (57), aligning with the findings from our

bioinformatics analysis and PCR experiments. In estrogen receptor-positive breast cancer, *TPX2* has been associated with poor distant metastasis-free survival and is involved in metastasis induction, as evidenced by the inhibitory effects of *TPX2* knockdown on the invasive potential and growth

Table 1 Patient characteristics of TCGA-COADREAD datasets

Characteristics	Overall, n (%)
Pathologic_M_stage (n=545)	
M0	459 (84.2)
M1	86 (15.8)
Pathologic_N_stage (n=618)	
N0	354 (57.3)
N1	150 (24.3)
N2	114 (18.4)
Pathologic_T_stage (n=619)	
T1	20 (3.2)
T2	109 (17.6)
T3	422 (68.2)
T4	68 (11.0)
OS (n=622)	
Dead	125 (20.1)
Alive	497 (79.9)

TCGA, The Cancer Genome Atlas; COADREAD, colorectal adenocarcinoma; OS, overall survival.

of adenocarcinoma cells (58).

SRPX is involved in a hypoxia-related prognostic model of endometrial cancer, and its expression was found to be reduced in tumor tissue compared to normal tissue, consistent with our CRC results (59). In addition, evidence supports the downregulation of this gene in various human tumor cells, classifying it as a tumor suppressor (60). However, it has also been reported to be significantly upregulated in cancer-associated fibroblasts from high-grade serous and clear-cell carcinoma samples in ovarian cancer (OC). Silencing *SRPX* in fibroblasts significantly suppresses the invasive capabilities of OC cells (61), suggesting that opinions on *SRPX* involvement in carcinogenesis might fluctuate depending on the tumor type.

The PCR results confirmed the downregulation of *CD36* expression in CRC. *CD36* plays a critical role in lipid homeostasis, angiogenesis, tumor metastasis, and therapeutic resistance by enhancing lipid uptake and FA oxidation, rendering it a promising candidate for cancer therapy (62,63). Several drugs targeting *CD36* have entered clinical trials; however, most have failed due to severe side effects and unsatisfactory performance. Further investigation is needed to comprehend the regulatory mechanisms and

Table 2 Univariate and multivariate Cox regression analysis for clinical characteristics

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	P value	HR (95% CI)	P value
T	619				
T1&T2	129	Reference		Reference	
T3	422	2.161 (1.119–4.174)	0.02	2.115 (0.894–5.004)	0.09
T4	68	6.866 (3.318–14.207)	<0.001	4.835 (1.867–12.524)	0.001
N	618				
N0	354	Reference		Reference	
N1	150	1.900 (1.204–2.997)	0.006	1.292 (0.753–2.217)	0.35
N2	114	4.079 (2.701–6.160)	<0.001	2.425 (1.434–4.100)	<0.001
M	545				
M0	459	Reference		Reference	
M1	86	4.289 (2.872–6.404)	<0.001	2.108 (1.305–3.404)	0.002
Risk score	622				
Low	311	Reference		Reference	
High	311	1.750 (1.218–2.513)	0.002	1.656 (1.103–2.488)	0.02

HR, hazard ratio; CI, confidence interval.

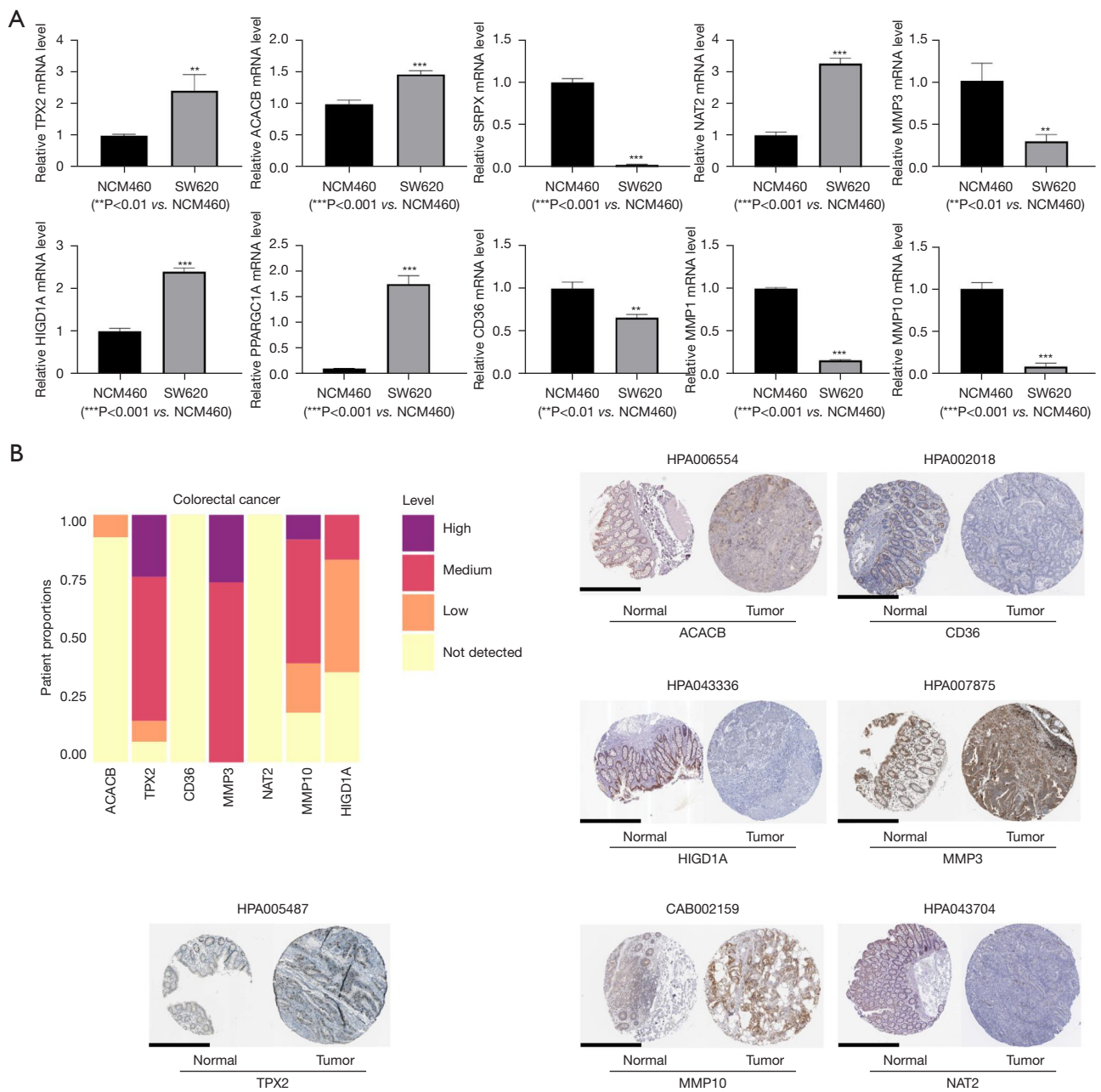


Figure 15 The expression validation of the telomere-related signature. (A) qRT-PCR results of the telomere-related signature in CRC cell lines (SW620) and control cell lines (NCM460). **, $P < 0.01$; ***, $P < 0.001$. (B) The detection of seven proteins by IHC in the HPA database (<https://www.proteinatlas.org/>) (scale bar = 600 μm). The links to the individual normal and tumor tissues of each protein are provided for ACACB (<https://www.proteinatlas.org/ENSG00000076555-ACACB/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000076555-ACACB/pathology/colorectal+cancer#img>), CD36 (<https://www.proteinatlas.org/ENSG00000135218-CD36/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000135218-CD36/pathology/colorectal+cancer#img>), HIGD1A (<https://www.proteinatlas.org/ENSG00000181061-HIGD1A/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000181061-HIGD1A/pathology/colorectal+cancer#img>), MMP3 (<https://www.proteinatlas.org/ENSG00000149968-MMP3/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000149968-MMP3/pathology/colorectal+cancer#img>), MMP10 (<https://www.proteinatlas.org/ENSG00000166670-MMP10/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000166670-MMP10/pathology/colorectal+cancer#img>), NAT2 (<https://www.proteinatlas.org/ENSG00000156006-NAT2/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000156006-NAT2/pathology/colorectal+cancer#img>), and TPX2 (<https://www.proteinatlas.org/ENSG00000088325-TPX2/tissue/colon#img>; <https://www.proteinatlas.org/ENSG00000088325-TPX2/pathology/colorectal+cancer#img>), respectively. qRT-PCR, quantitative real-time polymerase chain reaction; CRC, colorectal cancer; IHC, immunohistochemistry; HPA, the Human Protein Atlas.

downstream signaling pathways of *CD36* to facilitate future clinical applications (64).

Hypoxia-induced gene domain family-1a (*HIGD1A*) is a 10.4 kDa protein attached to the mitochondrial inner membrane (65), belonging to the conserved hypoxia-inducible gene domain family in eukaryotes (66). The relationship between *HIGD1A* and tumor growth is intricate and may depend on other pathways and the expression level of *HIGD1A* (67). While research on *HIGD1A* and CRC is limited, a multi-omics analysis of CRC identified three mitochondrial genes, including *HIGD1A*, that were significantly reduced in CRC and associated with poor outcomes in CRC patients, aligning with our study results (68).

In this study, the transcription levels of *CD36*, *SRPX*, and *TPX2* were consistent with the bioinformatics analysis, while other genes showed the opposite expression status. A common reason for this discrepancy is that PCR analyses and bioassays employ different methods and techniques. PCR, based on DNA replication, typically targets specific genes or sequences for analysis, whereas bioinformatics is a statistical and computational analytical method based on large-scale genomic data. The two methods have distinct limitations and assumptions in the measurement and interpretation of gene expression, potentially yielding inconsistent results. At the protein level, the HPA results were consistent with the bioinformatics analysis. The expression status of genes at the RNA and protein levels with conflicting results should be investigated further.

Although this study could serve as a valuable clinical guide, several limitations require resolution. First, data from the TCGA and GEO databases were insufficient; therefore, additional information must be collected. Second, further *in vivo* and *ex vivo* studies are required to explore the protein expression and biological roles of TRGs in CRC. Third, considering that the transcriptomic profile served as the foundation for our findings, further investigations are required to explore the potential mechanisms, molecular interactions, and clinical applications of prognostic genes in CRC.

Conclusions

In summary, this study underscores the significance of TRDEGs in predicting CRC prognosis. The envisaged impact of this signature is to enhance the accuracy of patient survival predictions, complementing the well-established TNM staging system. Furthermore, this study

revealed the genetic alterations, immune cell infiltration, immunotherapy response, and medication sensitivity that underlie this signature. These findings establish a basis for understanding the functions of TRDEGs and demonstrate their clinical relevance in CRC.

Acknowledgments

We would like to thank Editage (www.editage.cn) for English language editing.

Funding: None.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-43/rc>

Peer Review File: Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-43/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-43/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study complies with the Declaration of Helsinki (as revised 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.

2. Lombardi L, Morelli F, Cinieri S, et al. Adjuvant colon cancer chemotherapy: where we are and where we'll go. *Cancer Treat Rev* 2010;36 Suppl 3:S34-41.
3. Mattiuzzi C, Sanchis-Gomar F, Lippi G. Concise update on colorectal cancer epidemiology. *Ann Transl Med* 2019;7:609.
4. Tsatsakis A, Oikonomopoulou T, Nikolouzakis TK, et al. Role of telomere length in human carcinogenesis (Review). *Int J Oncol* 2023;63:78.
5. Kim NW, Piatyszek MA, Prowse KR, et al. Specific association of human telomerase activity with immortal cells and cancer. *Science* 1994;266:2011-5.
6. Blasco MA. Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet* 2005;6:611-22.
7. Jäger K, Walter M. *Therapeutic Targeting of Telomerase*. Genes (Basel) 2016;7:39.
8. Livingstone J, Shiah YJ, Yamaguchi TN, et al. The telomere length landscape of prostate cancer. *Nat Commun* 2021;12:6893.
9. Ningharhari M, Caruso S, Hirsch TZ, et al. Telomere length is key to hepatocellular carcinoma diversity and telomerase addiction is an actionable therapeutic target. *J Hepatol* 2021;74:1155-66.
10. Martel-Martel A, Corchete LA, Martí M, et al. Telomere Length as a New Risk Marker of Early-Onset Colorectal Cancer. *Int J Mol Sci* 2023;24:3526.
11. Kroupa M, Rachakonda SK, Liska V, et al. Relationship of telomere length in colorectal cancer patients with cancer phenotype and patient prognosis. *Br J Cancer* 2019;121:344-50.
12. Luu HN, Qi M, Wang R, et al. Association Between Leukocyte Telomere Length and Colorectal Cancer Risk in the Singapore Chinese Health Study. *Clin Transl Gastroenterol* 2019;10:1-9.
13. Li SC, Jia ZK, Yang JJ, et al. Telomere-related gene risk model for prognosis and drug treatment efficiency prediction in kidney cancer. *Front Immunol* 2022;13:975057.
14. Chen H, Liang W, Zheng W, et al. A novel telomere-related gene prognostic signature for survival and drug treatment efficiency prediction in lung adenocarcinoma. *Aging (Albany NY)* 2023;15:7956-73.
15. Hu Y, You X, Zhang Z, et al. Telomere-Associated Gene Signatures Correlate with Prognosis, Tumor Microenvironment, and Chemosensitivity in Breast Cancer. *Med Sci Monit* 2023;29:e939921.
16. Chen S, Hu S, Zhou B, et al. Telomere-related prognostic biomarkers for survival assessments in pancreatic cancer. *Sci Rep* 2023;13:10586.
17. Zhao Z, Shen X, Zhao S, et al. A novel telomere-related genes model for predicting prognosis and treatment responsiveness in diffuse large B-cell lymphoma. *Aging (Albany NY)* 2023;15:12927-51.
18. Yue K, Yao X. Prognostic model based on telomere-related genes predicts the risk of oral squamous cell carcinoma. *BMC Oral Health* 2023;23:484.
19. Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;44:e71.
20. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020;38:675-8.
21. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
22. Hinoue T, Weisenberger DJ, Lange CP, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res* 2012;22:271-82.
23. Hong Y, Downey T, Eu KW, et al. A 'metastasis-prone' signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics. *Clin Exp Metastasis* 2010;27:83-90.
24. Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* 2016;54:1.30.1-1.30.33.
25. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
26. Yu G. Gene Ontology Semantic Similarity Analysis Using GOSemSim. *Methods Mol Biol* 2020;2117:207-15.
27. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28:27-30.
28. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
29. Liu W, Li L, Ye H, et al. Weighted gene co-expression network analysis in biomedicine research. *Sheng Wu Gong Cheng Xue Bao* 2017;33:1791-801.
30. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
31. Brière G, Darbo É, Thébault P, et al. Consensus clustering applied to multi-omics disease subtyping. *BMC Bioinformatics* 2021;22:361.
32. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a

- class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;26:1572-3.
33. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics* 2019;11:123.
 34. Cai W, van der Laan M. Nonparametric bootstrap inference for the targeted highly adaptive least absolute shrinkage and selection operator (LASSO) estimator. *Int J Biostat* 2020;/j/ijb.ahead-of-print/ijb-2017-0070/ijb-2017-0070.xml. [Epub ahead of print]. doi: 10.1515/ijb-2017-0070.
 35. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol* 2010;5:1315-6.
 36. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
 37. Liberzon A, Subramanian A, Pinchback R, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739-40.
 38. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607-13.
 39. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-504.
 40. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res* 2020;48:D127-31.
 41. Zhang Q, Liu W, Zhang HM, et al. hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genomics Proteomics Bioinformatics* 2020;18:120-8.
 42. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014;42:D92-7.
 43. Varadi M, Anyango S, Deshpande M, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;50:D439-44.
 44. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.
 45. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453-7.
 46. Patel SS, Lovko VJ, Lockey RF. Red Tide: Overview and Clinical Manifestations. *J Allergy Clin Immunol Pract* 2020;8:1219-23.
 47. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* 2021;22:bbab260.
 48. Park SY. Nomogram: An analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg* 2018;155:1793.
 49. Perkins NJ, Weck J, Mumford SL, et al. Combining Biomarker Calibration Data to Reduce Measurement Error. *Epidemiology* 2019;30 Suppl 2:S3-9.
 50. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* 2018;74:796-804.
 51. Tataranni T, Piccoli C. Dichloroacetate (DCA) and Cancer: An Overview towards Clinical Applications. *Oxid Med Cell Longev* 2019;2019:8201079.
 52. Tran AN, Dussaq AM, Kennell T Jr, et al. HPAanalyze: an R package that facilitates the retrieval and analysis of the Human Protein Atlas data. *BMC Bioinformatics* 2019;20:463.
 53. Tang H, Wang J, Luo X, et al. An Apoptosis-Related Gene Prognostic Index for Colon Cancer. *Front Cell Dev Biol* 2021;9:790878.
 54. Bian X, Liu R, Meng Y, et al. Lipid metabolism and cancer. *J Exp Med* 2021;218:e20201606.
 55. Schinocca C, Rizzo C, Fasano S, et al. Role of the IL-23/IL-17 Pathway in Rheumatic Diseases: An Overview. *Front Immunol* 2021;12:637829.
 56. Neumayer G, Belzil C, Gruss OJ, et al. TPX2: of spindle assembly, DNA damage response, and cancer. *Cell Mol Life Sci* 2014;71:3027-47.
 57. Shaath H, Vishnubalaji R, Elango R, et al. Therapeutic targeting of the TPX2/TTK network in colorectal cancer. *Cell Commun Signal* 2023;21:265.
 58. Hu Y, Wu G, Rusch M, et al. Integrated cross-species transcriptional network analysis of metastatic susceptibility. *Proc Natl Acad Sci U S A* 2012;109:3184-9.
 59. Jiao Y, Geng R, Zhong Z, et al. A Hypoxia Molecular Signature-Based Prognostic Model for Endometrial Cancer Patients. *Int J Mol Sci* 2023;24:1675.
 60. Tambe Y, Hasebe M, Kim CJ, et al. The drs tumor suppressor regulates glucose metabolism via lactate dehydrogenase-B. *Mol Carcinog* 2016;55:52-63.
 61. Liu CL, Pan HW, Torng PL, et al. SRPX and HMCN1 regulate cancer associated fibroblasts to promote

- the invasiveness of ovarian carcinoma. *Oncol Rep* 2019;42:2706-15.
62. Zuo Q, He J, Zhang S, et al. PPAR γ Coactivator-1 α Suppresses Metastasis of Hepatocellular Carcinoma by Inhibiting Warburg Effect by PPAR γ -Dependent WNT/ β -Catenin/Pyruvate Dehydrogenase Kinase Isozyme 1 Axis. *Hepatology* 2021;73:644-60.
 63. DeFilippis RA, Chang H, Dumont N, et al. CD36 repression activates a multicellular stromal program shared by high mammographic density and tumor tissues. *Cancer Discov* 2012;2:826-39.
 64. Wang J, Li Y. CD36 tango in cancer: signaling pathways and functions. *Theranostics* 2019;9:4893-908.
 65. An HJ, Ryu M, Jeong HJ, et al. Higd-1a regulates the proliferation of pancreatic cancer cells through a pERK/p27(KIP1)/pRB pathway. *Cancer Lett* 2019;461:78-89.
 66. Li T, Xian WJ, Gao Y, et al. Higd1a Protects Cells from Lipotoxicity under High-Fat Exposure. *Oxid Med Cell Longev* 2019;2019:6051262.
 67. Zhu JY, Chen M, Mu WJ, et al. The functional role of Higd1a in mitochondrial homeostasis and in multiple disease processes. *Genes Dis* 2022;10:1833-45.
 68. Zhang W, Lin L, Xia L, et al. Multi-omics analyses of human colorectal cancer revealed three mitochondrial genes potentially associated with poor outcomes of patients. *J Transl Med* 2021;19:273.

Cite this article as: Chen H, Pan Y, Lv C, He W, Wu D, Xuan Q. Telomere-related gene risk model for prognosis prediction in colorectal cancer. *Transl Cancer Res* 2024;13(7):3495-3521. doi: 10.21037/tcr-24-43