Discover Oncology

Analysis

SUMOylation-regulated genes in colon cancer: expression patterns and clinical implications

Wen Peng[1] · Zirong Yang[2] · Ru Yan[1] · Lan Mu[1] · Lan Li[3] · Shan Jin[1] · Shisheng Tan[1]

Received: 3 February 2025 / Accepted: 7 May 2025
Published online: 20 May 2025

Abstract

Colon cancer (CRC) demonstrates significant heterogeneity, and identifying effective biomarkers can advance the development of precision therapies. Emerging evidence implicates SUMOylation-regulated genes as pivotal regulators of cancer-associated pathways, yet their prognostic potential and therapeutic implications in CRC remain unexplored. A comprehensive analysis of SUMOylation-regulated gene expression, clinical and prognostic value in CRC was performed using transcriptomic data from TCGA-COAD and GEO datasets. We identified 46 differentially expressed SUMOylation-regulated genes (33 upregulated, 13 downregulated) in CRC tumors versus normal tissues. Unsupervised clustering based on 216 SUMOylation-related genes stratified CRC patients into two distinct subtypes: SUMO Cluster 1 (aggressive phenotype, poor prognosis) and SUMO Cluster 2 (favorable prognosis). Cluster 1 exhibited advanced tumor stages (N-stage, $p < 0.05$) and may present an immunosuppressive microenvironment marked by reduced HLA/immune checkpoint gene expression, while Cluster 2 showed enhanced anti-tumor immunity (activated dendritic cells, γδ T cells). A five-gene SUMOylation-based prognostic signature (MC1R, LRRC4C, SAGE1, GJB6, HOXC5) was developed, and patients were divided into high Riskscore and low Riskscore groups with significant survival differences (log-rank $p < 0.05$). The nomogram integrating risk score, age, and stage demonstrated robust predictive accuracy (C-index = 0.763, AUC = 0.728–0.785). Nomoscore-high patients exhibited resistance to AMG.706 and ABT.888, suggesting therapeutic vulnerabilities. These findings highlight SUMOylation plays a critical role in CRC heterogeneity, immune modulation, and prognosis, offering a novel biomarker system for risk stratification and personalized therapy.

Keywords SUMOylation · Colon cancer · Tumor microenvironment · Prognostic signature · Chemotherapy sensitivity

Wen Peng, Zirong Yang and Ru Yan contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02614-z.

✉ Shisheng Tan, tssh18018@126.com; Wen Peng, pengwen8160@163.com; Zirong Yang, yangzr799@126.com; Ru Yan, yanru1203dl@126.com; Lan Mu, mulannice@163.com; Lan Li, 328931207@qq.com; Shan Jin, 18985199851@163.com | [1]Department of Oncology, The People's Hospital of Guizhou Province, Guiyang, China. [2]Institute of Life Science, eBond Pharmaceutical Technology Ltd, Chengdu, China. [3]Department of General Medicine, The People's Hospital of Guizhou Province, Guiyang, China.

1 Introduction

Colon cancer (CRC) is a major global health challenge, ranking third in diagnosed cancers and second in terms of cancer mortality. In 2020, more than 555,477 new cases and 286,162 deaths were reported by the CRC. Both incidence and mortality rates for CRC have shown persistent growth over the last few decades [1]. In China, individuals diagnosed with CRC have a 5-year survival rate of approximately 62%, highlighting the importance of improving early detection, treatment, and public health efforts [2]. Over the past few years, immunotherapy has been increasingly utilized as a treatment for colon cancer, and NCCN guidelines recommend its use in patients with Microsatellite Instability (MSI-H) or Mismatch Repair (dMMR). Nevertheless, this cohort constitutes merely 5% of the total patient population. Hence, the identification of new prognostic biomarkers and the exploration of alternative therapeutic targets are imperative [3, 4].

Post-translational modifications (PTMs) are covalent modifications that alter the structure and function of proteins after they are synthesized. These modifications involve forming or breaking covalent bonds on the protein backbone or amino acid side chains. These modifications can significantly influence protein activity, stability, subcellular localization, and overall functionality. Owing to their close association with disease mechanisms, PTMs hold great potential as therapeutic targets for disease diagnosis, treatment, and intervention [5]. SUMO (Small Ubiquitin-like Modifier) is a type of PTM analogous to ubiquitination, but it involves the conjugation of SUMO proteins to target substrates.

SUMOylation involves covalent attachment of evolutionarily conserved ~ 12 kDa polypeptides (structurally analogous to ubiquitin) to target lysine residues. The human genome encodes three canonical paralogs (SUMO1-3) and the less characterized SUMO4, all phylogenetically conserved across vertebrates. Pathological SUMOylation dysregulation manifests as either hyperconjugation or hypoconjugation, contributing to oncogenic signaling rewiring and proteotoxic stress in malignancies. SUMO modification regulates cellular processes, including protein localization, interactions, and activity. It plays a critical role in various diseases, such as cancer and neurodegenerative disorders, and represents a potential therapeutic target for treatment [6].

Numerous clinical studies have revealed that proteins associated with SUMOylation are markedly overexpressed in solid tumor tissues and are closely linked to tumor progression [7, 8]. Recent research highlights the role of SUMOylation in influencing the tumor microenvironment, particularly in modulating immune responses such as macrophage activation and CD8 + T cell infiltration. Its impact on the immune microenvironment in bladder and prostate cancers presents a promising opportunity to improve the efficacy of immunotherapy [9, 10]. However, the interplay between SUMOylation, CRC progression, and immune-related characteristics remains insufficiently elucidated.

In this study, we investigated the correlation between the expression of SUMOylation-regulated genes and the progression of CRC, providing valuable data for the development of personalized treatment strategies.

2 Materials and methods

2.1 Data source and processing

Transcriptomic profiles and clinically annotated metadata for colon adenocarcinoma (COAD) were retrieved from The Cancer Genome Atlas (TCGA) portal. The dataset comprised 480 tumor specimens alongside 41 matched normal tissues, with 427 cases having complete demographic and clinicopathological variables (including survival status, age, gender, TNM staging, and smoking history) for subsequent analyses.

2.2 Identification of differentially expressed SUMOylation genes

216 SUMOylation-regulated genes were identified from the MsigDB database (http://www.broadinstitute.org/gsea/msigdb/index.jsp) (Table S1). The"limma"package" [11] (https://bioconductor.org/packages/release/bioc/html/limma.html) was used to identify differentially expressed genes (DEGs) between tumor and normal tissues, with criteria set at |log2(Fold Change)|> 1 and FDR < 0.05. Gene expression differences were visualized using heatmaps and volcano plots generated by the"heatmap"package (https://cran.r-project.org/web/packages/tidyHeatmap/vignettes/intro

duction.html). Following the identification of DEGs through intersection analysis, univariate Cox regression was employed to select those DEGs associated with prognosis, considering $p < 0.05$ as statistically significant.

2.3 Immune infiltration analysis

The evaluation of immune cell infiltration was conducted through dual algorithmic validation: the CIBERSORT platform (https://github.com/topics/cibersort) and single-sample gene set enrichment analysis (ssGSEA) within the"GSVA"package [12]. Correlation between risk scores and immunophenotypic infiltration features was established using a multi-dimensional bioinformatics workflow integrating data restructuring tools (reshape2), visualization systems (ggplot2, v3.4.4), and statistical plotting extensions (ggpubr v0.6.0, https://cran.r-project.org/web/packages/ggpubr/index.html).

2.4 Unsupervised consensus clustering identified of SUMOylation patterns

The SUMOylation regulatory gene set was curated through molecular signatures from the MSigDB database, yielding 216 genes. To delineate SUMOylation-driven molecular cluster, we implemented an ensemble-based classification framework via the"ConsensusClusterPlus package"(v1.70.0, https://cran.r-project.org/) [13], with transcriptomic inputs normalized using variance-stabilizing transformation.

The clustering architecture was configured as follows: Partitioning method: k-means centroid optimization (k = 2); Distance metric: Euclidean pairwise dissimilarity matrix; Resampling control: 80% subsampling rate with fixed seed (123,456) for reproducibility. Cluster stability was rigorously assessed through: Cumulative distribution function (CDF): Optimal k = 2 determined by delta area threshold > 0.15; Multidimensional validation: Principal component analysis (PCA) using 95% variance-explained components. SUMOylation pathway activity was quantified via Gene Set Variation Analysis (GSVA) with the"SUMOylation_Process"gene set. The analysis of DEGs was carried out in the same manner as described above.

2.5 Independent prognostic factors and nomogram construction

2.5.1 Establishment of Riskscore model

TCGA-COAD samples were randomly divided into training and testing sets in a 7:3 ratio. The GSE39582 dataset (N = 562) from GEO served as an external validation set. The training set supported model development while both testing and validation sets were used for validation. All datasets were normalized for consistency.

In the training set: We first utilized univariate Cox regression to identify differentially expressed genes with prognostic significance ($p < 0.05$) between SUMO Cluster 1 and SUMO Cluster 2. Potential genes were further refined using the Least Absolute Shrinkage and Selection Operator (LASSO) method, implemented with the R package"glmnet"(v4.1–6, https://cran.r-project.org/src/contrib/Archive/glmnet/) [14] retaining those with non-zero coefficients. A risk score model was then constructed by weighting the expression levels of these selected genes with their corresponding coefficients derived from multivariate Cox regression. The risk score was calculated using the following formula:

$$\text{Riskscore} = \sum \beta\text{gene} \times \text{Expgene}$$

βgene represents the LASSO regression coefficient of the gene, and Expgene indicates the relative expression level of the gene in the training set.

To enhance clinical applications, patients were divided into two groups, namely high-risk and low-risk, using the median risk score as the threshold. Survival analysis was performed with the R package"survival"(https://cran.r-project.org/web/packages/survival/index.html) to evaluate prognostic differences between the groups. The R package"survivalROC"(version 1.0.2, https://cran.r-project.org/src/contrib/Archive/survivalROC) was utilized to create ROC curves and compute AUC values for 1, 3, and 5 years. To confirm the model's reliability, the TCGA test set and the GEO external validation set were computed using identical regression coefficients based on the risk score formula.

2.5.2 Nomogram construction

Cox regression analyses, including univariate and multivariate, were used to assess independent prognostic factors, incorporating clinical data from TCGA tumor samples and the previously obtained RiskScore. The predictive

performance of the Riskscore was compared with various clinicopathological factors by generating receiver operating characteristic (ROC) curves.

2.6 Prediction of drug susceptibility

Pharmacodynamic responsiveness to 129 antineoplastic compounds was computationally predicted using the Genomics of Drug Sensitivity in Cancer (GDSC) resource. Drug sensitivity profiling was implemented via the "pRRophetic" package to estimate half-maximal inhibitory concentrations (IC 50) [15].

2.7 Statistical analysis

Statistical analyses were performed with R software (v4.2.1) using a significance threshold of $p < 0.05$. For continuous variables, parametric (Student's t-test) and non-parametric (Wilcoxon rank-sum test) tests were applied, whereas Pearson's chi-square test was used for categorical variable comparisons. A schematic representation of the analytical workflow is depicted in Fig. 1.

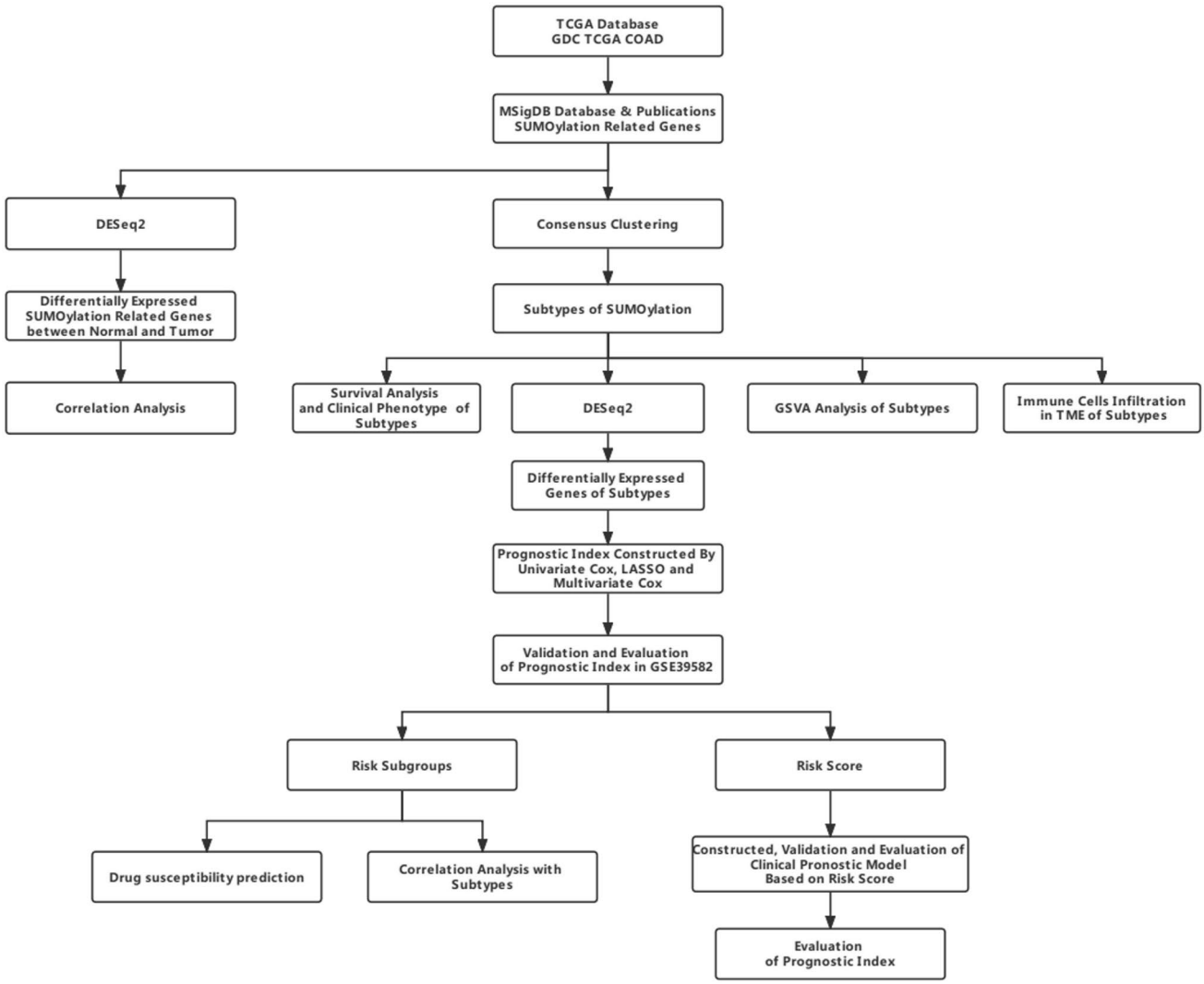

Fig. 1 The workflow of this study

3 Results

3.1 Expression analysis of SUMOylation-regulated genes in colon cancer

We utilized the TCGA-COAD dataset to analyze 216 SUMOylation-regulated genes identified from the MsigDB database (Table S1), comparing tumor tissues to normal tissues in CRC.

Differential expression analysis revealed 46 genes exhibiting significant differences in expression levels between tumor and normal tissues (FDR < 0.05, |log2 FC|> 1). Specifically, 33 genes were upregulated in tumor tissues, while 13 genes were downregulated (Fig. 2A) (Table S2). Additionally, a heatmap displayed the expression profiles of these 46 genes (Fig. 2B), and we conducted an expression correlation analysis for the differentially expressed SUMOylation-regulated genes in CRC tumor samples (Fig. 2C).

3.2 Identification of different SUMOylation patterns in colon cancer

To explore the expression patterns of SUMOylation in CRC patients, we analyzed clinical data from 427 samples with survival information in the TCGA database using unsupervised consensus clustering based on a set of 216 SUMOylation regulatory genes. We employed the cumulative distribution function (CDF) curve alongside a scree plot to assess the validity of the clustering, (Fig. 2D, Supplement Fig. 1 A, B). Our analysis determined that the optimal number of clusters was 2 (K = 2), leading to the classification of patients into two distinct groups: SUMO Cluster 1, comprising 221 cases, and SUMO Cluster 2, which included 206 cases.

We calculated the enrichment scores of the SUMOylation-regulated gene sets using the GSVA algorithm to validate the distinction between the identified clusters. A subsequent t-test comparison between SUMO Clusters 1 and 2 revealed significant differences (Fig. 2E).

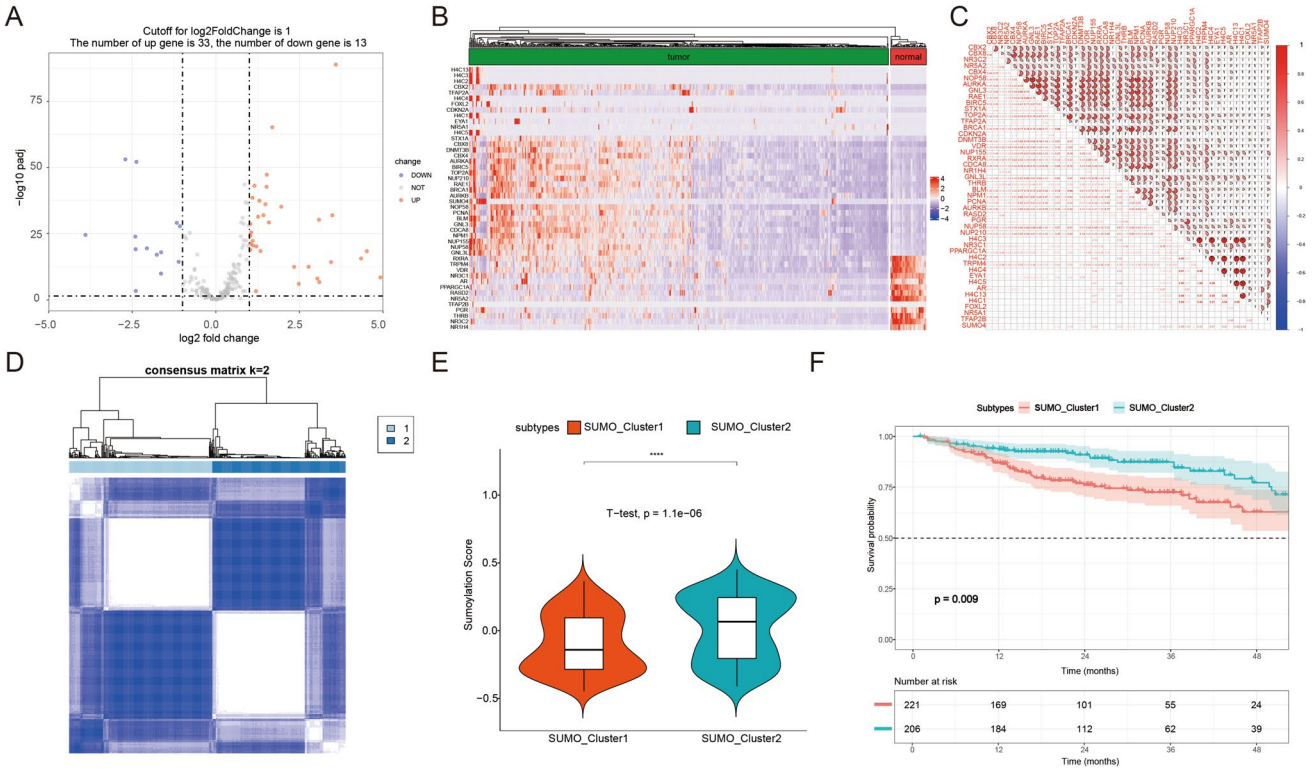

Fig. 2 Identification of SUMOylation regulatory subtypes in colon cancer. **A** Volcano plot showing differential gene expression of SUMOylation regulatory genes. **B** Heatmap of differential gene expression of SUMOylation regulatory genes. **C** Correlation between differential SUMOylation regulatory genes. **D** Unsupervised clustering results based on SUMOylation regulatory genes (Pam Pearson k = 2). **E** Comparison of enrichment scores of SUMOylation regulatory gene sets between different subtypes. **F** Kaplan–Meier curves showing prognosis correlation between molecular subtypes of SUMOylation regulatory genes

The prognostic differences between SUMO Cluster 1 and SUMO Cluster 2 were assessed utilizing Kaplan–Meier survival analysis. As illustrated in Fig. 2F, SUMO Cluster 1 exhibited a significantly poorer prognosis compared to Cluster 2. ($p = 0.009$).

3.3 Clinical correlation analysis between two different SUMOylation patterns

Moreover, we conducted a comparative analysis of the distribution of various clinical characteristics among CRC patients. Two SUMOylation patterns showed statistically significant differences in tumor stage ($p < 0.01$, Supplement Fig. 2 A), with SUMO Cluster 1 having 32.7% of patients in stage II and 37.4% in stage III, while SUMO Cluster 2 showed higher proportions of stage II patients at 45.5% and a lower representation of stage III patients at 21.8%. Further analysis of the TNM staging system indicated that only the N stage exhibited significant differences between the subtypes ($p < 0.001$, Supplement Fig. 2B), while no significant discrepancies were noted in the T and M stages (Supplement Fig. 2 C, D). Additionally, age and gender distributions were found to be comparable across both subtypes (Supplement Fig. 2E, F).

3.4 Gene expression differences between SUMOylation patterns

Differential gene expression analysis was conducted to compare SUMO Cluster 1 and Cluster 2. The analysis revealed that, relative to Cluster 1, Cluster 2 demonstrated a significant upregulation of 48 genes and a significant downregulation of 412 genes. (Supplement Fig. 3 A).

Furthermore, the Wilcoxon test was utilized to compare the expression data of immune checkpoint genes and HLA family genes across the two distinct SUMOylation patterns. The analysis identified significant differential expression in 16 immune checkpoint genes, such as TNFSF14, TNFRSF8, TNFRSF4, TNFRSF25, and TNFRSF14, as well as in seven genes from the HLA family, including HLA-G, HLA-E, HLA-DQB2, and HLA-DPB1 ($p < 0.05$) (Supplement Fig. 3B, C).

3.5 Comparison of immune microenvironment between SUMOylation patterns

To investigate the impact of SUMOylation on the CRC immune microenvironment, we used CIBERSORT to analyze immune cell infiltration. The analysis revealed that the infiltration of neutrophils, eosinophils, activated dendritic cells, resting dendritic cells, γδ T cells, activated CD4 memory T cells, and resting CD4 memory T cells was significantly higher in the SUMO-Cluster 2 compared to SUMO-Cluster 1. Conversely, the SUMO-Cluster 1 group exhibited significantly increased infiltration of M0 macrophages and Treg cells (Supplement Fig. 4 A).

To delineate immune heterogeneity between the SUMO-cluster subtypes, we implemented ssGSEA quantifying 23 leukocyte subsets via Wilcoxon rank-sum testing. SUMO-Cluster2, exhibiting significant enrichment of γδT cells, CD4 + effector memory T lymphocytes, memory B cells, immature dendritic cells, and type II helper T cells. Comparatively, SUMO-Cluster1 demonstrated a pronounced accumulation of myeloid-derived suppressor cells, CD56 dim natural killer cells, and activated B cell compartments ($p < 0.05$) (Supplement Fig. 4B).

3.6 Prognostic significance of the riskscore model

To evaluate the prognostic value of SUMOylation patterns in CRC and to develop a personalized prognostic model for individual patients, we initially conducted univariate Cox regression analysis using 70% of the samples from the TCGA COAD database as the training cohort to identify genes with significant prognostic relevance. This analysis identified 20 prognosis-related genes from the differentially expressed genes of SUMO Cluster 1 and Cluster 2 (Fig. 3A). Next, we applied LASSO regression to select the best variables, narrowing it down to 8 prognosis-associated genes (Fig. 3B, C). Finally, multivariate Cox regression analysis identified 5 key genes linked to prognosis: MC1R, LRRC4 C, SAGE1, GJB6, and HOXC5. These genes were used to construct a SUMOylation-based prognostic signature model (Riskscore model) (Fig. 3D). The risk score calculation formula was set as follows:

$$\begin{aligned}\text{Riskscore} &= \text{MC1R} \times 0.304907298 + \text{LRRC4C} \times 0.231643319 + \text{SAGE1} \times 0.225151075 \\ &+ \text{GJB6} \times 0.34961565 + \text{HOXC5} \times 0.14103721\end{aligned}$$

Patients with Riskscore above the median experienced significantly worse survival outcomes compared to their lower-scoring counterparts ($p < 0.05$) (Fig. 4A). The AUC values for predicting 1-year, 3-year, and 5-year survival rates were 0.693, 0.695, and 0.689, respectively (Fig. 4D).

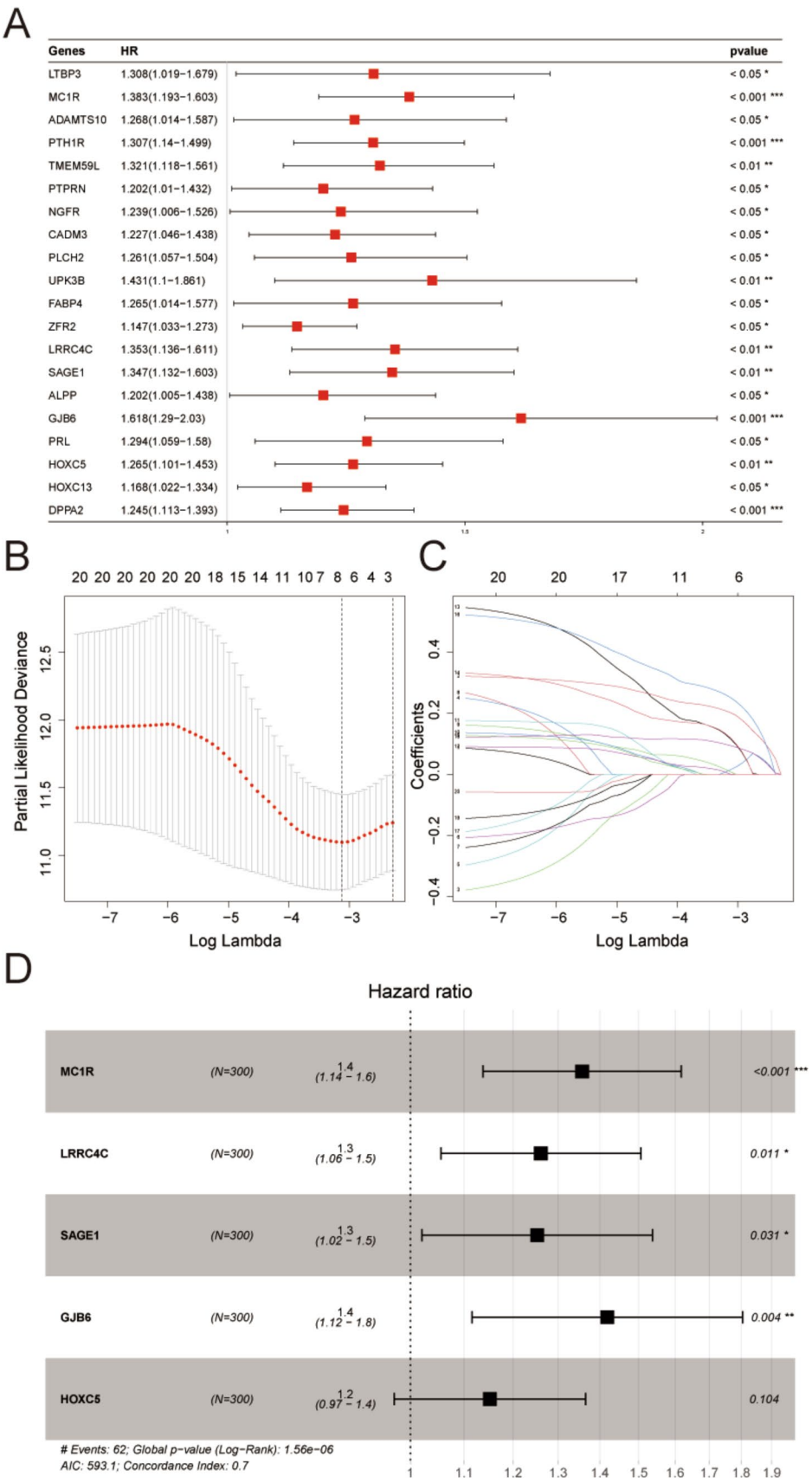

Fig. 3 Gene selection for the riskscore prognostic model. **A** Univariate Cox regression analysis of differentially expressed genes between subtypes. **B** LASSO selection of prognostic-related genes. LASSO Regression analysis. **C** LASSO Coefficient distribution map-LASSO coefficient distribution of all variables. **D** Multivariate Cox regression analysis

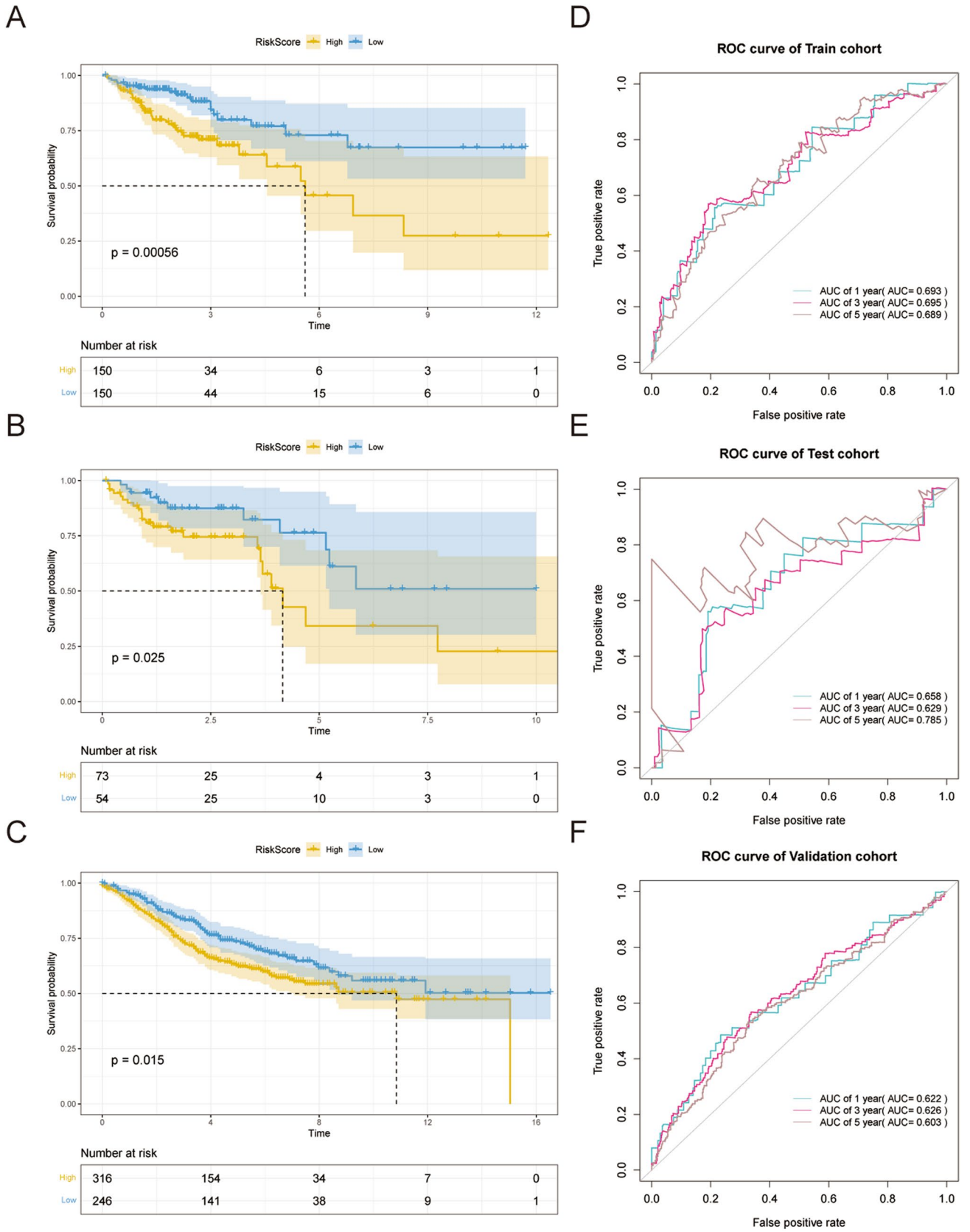

Fig. 4 Correlation analysis between riskscore and prognosis. Kaplan–Meier curves in (**A**) TCGA training set; **B** TCGA testing set; **C** GEO external validation set and ROC curve in **D** TCGA training set; **E** TCGA testing set; **F** GEO external validation set

The prognostic accuracy of the Riskscore model was validated in both a TCGA test cohort (30% of TCGA-COAD) and an independent GEO validation cohort. Results consistently demonstrated a significantly worse prognosis for the high-Riskscore group compared to the low-Riskscore group ($p < 0.05$) (Fig. 4B, C). The predictive accuracy of the Riskscore model was assessed by calculating AUC values for 1-, 3-, and 5-year survival. In the TCGA test cohort (Fig. 4E), AUCs were 0.658, 0.629, and 0.785, respectively. The GEO validation cohort (Fig. 4F) yielded AUCs of 0.622, 0.626, and 0.603.

3.7 Construction and validation of a nomogram

Univariate Cox regression analysis demonstrated that the risk score, along with age, stage, and TNM classification (AJCC_T, AJCC_N, and AJCC_M stages), were independent prognostic factors (Fig. 5A). Multivariate Cox regression analysis confirmed that the risk score, age, and stage were independent prognostic factors ($p < 0.05$) (Fig. 5B). A nomogram incorporating the risk score, age, and stage was developed, demonstrating good performance with a C-index of 0.763 (Fig. 5C).

Based on this nomogram, we calculated the clinical risk scores (Nomoscore) for TCGA COAD tumor samples. The median Nomoscore served as a threshold to categorize samples into the Nomoscore-high group (Nomoscore > median) and Nomoscore-low group (Nomoscore ≤ median). A comparison of the prognostic outcomes revealed that patients in the Nomoscore-high group had a significantly worse prognosis than those in the Nomoscore-low group ($p < 0.0001$) (Fig. 6A).

The model's performance was further assessed using time-dependent ROC curves, with AUC values of 0.774, 0.785, and 0.728 for predicting 1-year, 3-year, and 5-year survival, respectively. (Fig. 6B). The Nomoscore model was further evaluated by plotting calibration curves as well as DCA decision curves, and as shown in Fig. 6C–F, the Nomoscore model performed well in predicting prognosis than clinical model (including age, clinical stage, and TNM stage) (Supplement Fig. 5 A, B).

3.8 Prediction of chemotherapy drug sensitivity between Nomoscore groups

Using the GDSC data, we assessed the sensitivity of each sample to chemotherapeutic drugs by analyzing mRNA expression levels in CRC tumor tissues. We calculated the IC50 values for 129 chemotherapeutic agents using the pRRophetic package and classified patients into different Nomoscore groups. Our analysis indicated that the Nomoscore-low group exhibited greater sensitivity to AMG.706 and ABT.888, with significant differences in IC50 values compared to Nomoscore-high groups ($p < 0.05$) (Supplement Fig. 6 A, B).

4 Discussion

In this study, we investigated the roles of SUMOylation-regulated genes in CRC, revealing their distinct expression patterns, clinical relevance, and prognostic implications. Our findings highlight the critical involvement of SUMOylation in shaping tumor progression, immune microenvironment modulation, and therapeutic responses in CRC, and could serve as both a biomarker and a therapeutic target.

Post-translational modifications (PTMs) exert a profound influence in fine-tuning cellular functions, but their dysregulation contributes to pathological processes including carcinogenesis. Among PTMs, SUMOylation has emerged as a critical regulator of tumor biology, though its dual roles in cancer progression remain context-dependent. SENP1 is a deSUMOylating enzyme that interacts with c-Myc to stabilize its levels and activity by reducing its polyubiquitination and degradation. Studies suggest that inhibiting SUMOylation may be a potential therapeutic strategy for Myc-driven cancers, as decreased SAE1/2 activity leads to the death of Myc-dependent tumor cells [16, 17]. PIAS1 is a chromatin-bound coregulator of the androgen receptor in prostate cancer, influencing the expression of a specific subset of AR-regulated genes and affecting AR chromatin occupancy, thereby impacting VCaP cell proliferation and unveiling additional genes to androgen regulation [18]. MANF acts as a crucial link between endoplasmic reticulum stress and liver inflammation by inhibiting the NF-κB/Snail signaling pathway, thereby suppressing epithelial-mesenchymal transition and hepatocellular carcinoma progression, suggesting its potential as a therapeutic target for HCC [19]. However, there were serval studies investigating the role of SUMOylation in CRC.

In comparative transcriptomic profiling of CRC and paired normal tissues in the TCGA-COAD cohort, we identified 46 SUMOylation-associated differentially expressed genes, comprising 33 significantly upregulated genes and 13 downregulated genes. This distinct expression pattern underscores the pivotal role of SUMOylation in modulating CRC pathogenesis, potentially through the dysregulation of key cellular processes.

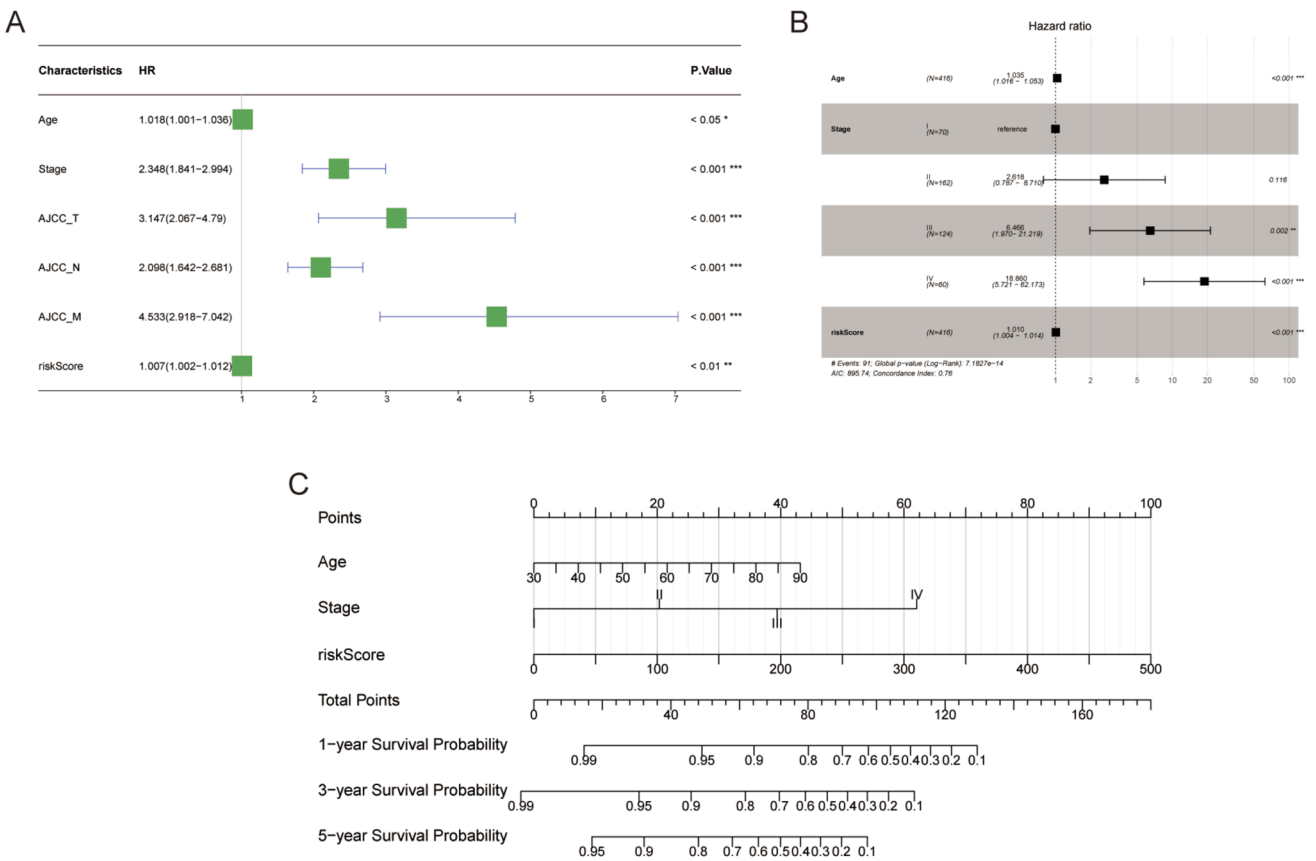

Fig. 5 Independent prognostic factor analysis. **A** Univariate Cox analysis of riskscore and clinical indicators. **B** Multivariate Cox analysis of riskscore and clinical indicators. **C** Construction of a nomogram for the clinical risk model

Then we using NMF clustering according to the expression level of these SUMOylation-related genes divide the patients from the TGCA-COAD cohort into two SUMOylation patterns. The molecular rationality of this subtyping was validated by significant differences in SUMOylation pathway enrichment scores (GSVA, $p < 0.05$) and prognostic divergence (log-rank $p = 0.009$) between clusters. Patients in Cluster 1 demonstrated a 37.4% prevalence of stage III tumors and significantly inferior survival outcomes, in contrast to Cluster 2, which was characterized by a predominance of stage II tumors (45.5%) and a more favorable prognosis. Gene expression analysis identified a set of differentially expressed genes between SUMO Clusters 1 and 2. Notably, immune checkpoint genes and HLA family genes showed significant differences in expression between subtypes. Dysregulation of immune checkpoints and alterations in the HLA system have been extensively studied in the context of cancer immunotherapy [20].

Furthermore, we investigated the differences in the immune microenvironment between the two SUMOylation-regulated gene subtypes. Cluster 1 showed increased infiltration of M0 macrophages and Tregs. M0 macrophages can polarize into either anti-tumor M1 or pro-tumor M2 types [21], suggesting their elevated presence may indicate a specific polarization pathway that requires further investigation. Although Tregs are generally associated with poor prognosis in many caners [22, 23], they correlate with better outcomes in CRC [24]. Thus, elevated M0 macrophages and Tregs in Cluster 1 may influence CRC prognosis through unique mechanisms. In contrast, Cluster 2 exhibited higher levels of activated dendritic cells, gamma delta T cells, and memory $CD4^+$ T cells, which are linked to anti-tumor immunity [25]. SUMOylation dysregulation has previously been implicated in immune modulation and the evasion of immune surveillance in cancer [26, 27]. These immune differences likely contribute to the prognostic disparities between clusters. Future research should elucidate the specific roles of these immune cells within SUMOylation patterns to fully understand their functions in CRC.

To determine the clinical relevance of the findings, a risk score prognostic model using 70% of TCGA COAD samples identified 5 key prognostic genes—MC1R, LRRC4 C, SAGE1, GJB6, and HOXC5—through regression analyses. This model effectively predicted survival across the training cohort (70% TCGA-COAD cohort), testing cohort (30% TCGA cohort), and external validation cohort (GEO cohort). Furthermore, a nomogram incorporating risk score, age, and tumor stage

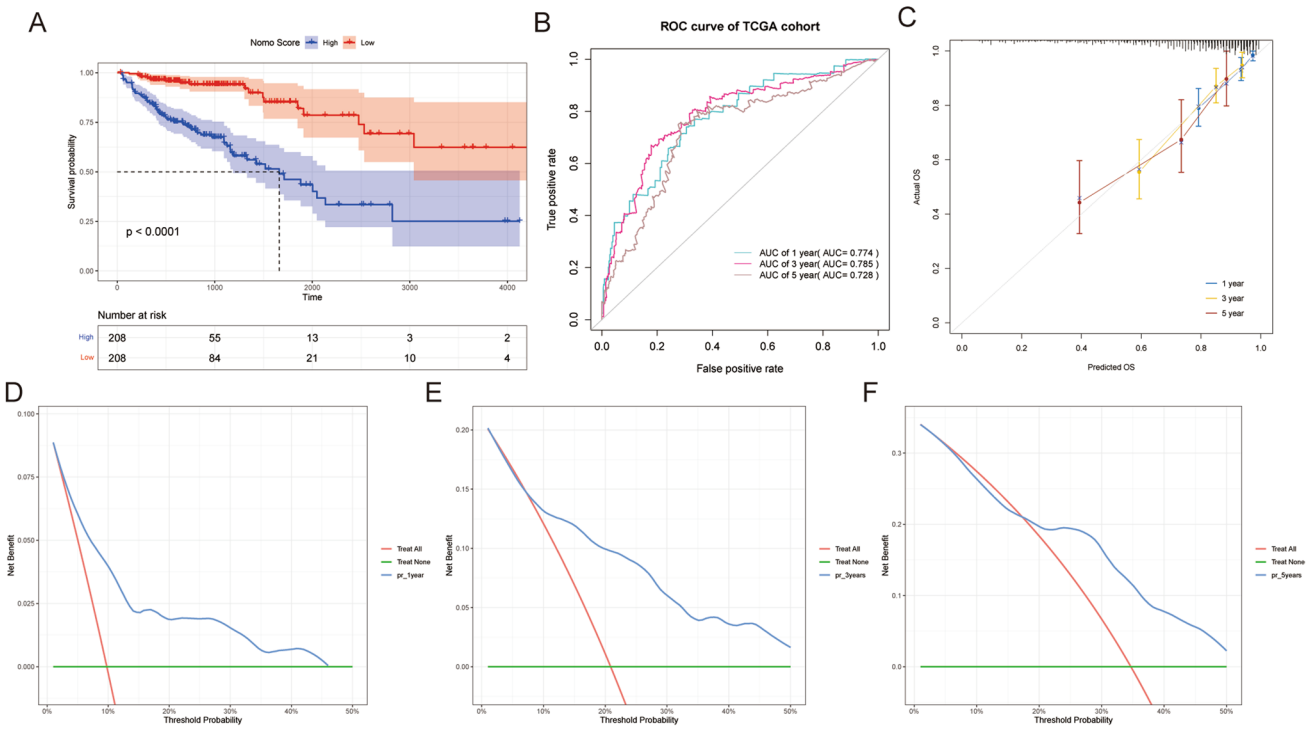

Fig. 6 Correlation analysis between Nomoscore and prognosis. **A** Kaplan–Meier curves showing prognosis correlation with Nomoscore. **B** ROC curve. **C** Calibration curve. **D**–**F** Decision curve analysis (DCA) curves for 1, 3, and 5-year survival

was developed to provide a comprehensive tool for predicting individual patient outcomes. The nomogram performed well in predicting prognosis and may aid clinical decision-making and patient counseling. Given that the liver and lungs are the most common metastatic sites of colon cancer and are associated with poor prognosis, the genes in our model may serve as early predictors of these risks, allowing for timely interventions. Future studies should validate their ability to forecast organ-specific metastases, thereby enhancing personalized treatment strategies and improving patient outcomes. Importantly, the Nomoscore group's sensitivity to AMG.706 and ABT.888 suggests that SUMOylation patterns may guide chemotherapy selection. AMG.706 (a multi-kinase inhibitor) and ABT.888 (a PARP inhibitor) target pathways potentially dysregulated in SUMOylation-high tumors, such as angiogenesis and DNA repair [28–31].

While our study provides valuable insights, several limitations should be acknowledged. First, the retrospective nature of the TCGA and GEO datasets may introduce selection bias, and the exclusion of the READ (rectum adenocarcinoma) cohort further limits the generalizability of our findings. Prospective validation in independent cohorts is necessary to confirm the clinical utility of the SUMOylation signature. Second, the use of CIBERSORT and ssGSEA for immune cell infiltration analysis has limited reliability compared to immunohistochemistry-based methods, further validation of the immune analysis results is needed. Mechanistic studies using both in vitro and in vivo models are needed to elucidate the contribution of SUMOylation to tumorigenesis and therapy resistance. Finally, the interplay between SUMOylation and other post-translational modifications in shaping the CRC microenvironment warrants further investigation.

5 Conclusion

In this study, we identified differentially expressed SUMOylation-regulated genes in colon cancer and demonstrated their association with clinical outcomes. These findings highlight the role of SUMOylation in molecular subtyping, prognosis, and the immune microenvironment of colon cancer. The risk score prognostic model and nomogram provided additional tools for assessing patient outcomes. However, further experimental and clinical studies are required to elucidate the underlying mechanisms and validate the clinical implications of SUMOylation dysregulation in colon cancer.

Acknowledgements This project was supported by Guizhou Provincial Science and Technology Department: Qian Ke He JC (2024) 472, and Science and Technology Fund of Guizhou Provincial Health Commission, gzwkj2023-011). The results here are partly based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga and GEO database: https://www.ncbi.nlm.nih.gov/geo.

Author contributions WP, ZY, and RY were involved in conceptualization, data analysis, and original draft preparation. LM, LL and SJ were involved in data analysis and methodology. ST wrote, reviewed, and edited the manuscript. All the authors have read and agreed to the published version of the manuscript.

Funding Guizhou Provincial Science and Technology Department, Qian Ke He JC ZK (2024) 472, Science and Technology Fund of Guizhou Provincial Health Commission,gzwkj2023-011.

Data availability The public data set of this study can be found in the TCGA database (https://portal.gdc.cancer.gov/) and GEO database (https://www.ncbi.nlm.nih.gov/geo).

Declarations

Competing interests The authors declare no competing interests.

References

1. Chen H, Lu B, Dai M. Colorectal cancer screening in china: status, challenges, and prospects-China, 2022. China CDC Weekly. 2022;4:322–8. https://doi.org/10.46234/ccdcw2022.077.
2. Wang R, Lian J, Wang X, Pang X, Xu B, Tang S, Shao J, Lu H. Survival rate of colorectal cancer in China: a systematic review and meta-analysis. Front Oncol. 2023;13:1033154. https://doi.org/10.3389/fonc.2023.1033154.
3. Benson AB, Venook AP, Adam M, Chang G, Chen YJ, Ciombor KK, Cohen SA, Cooper HS, Deming D, Garrido-Laguna I, et al. Colon cancer, version 32024, NCCN clinical practice guidelines in oncology. J Nat Comprehens Cancer Netw. 2024. https://doi.org/10.6004/jnccn.2024.0029.
4. Sinicrope FA, Sargent DJ. Molecular pathways: microsatellite instability in colorectal cancer: prognostic, predictive, and therapeutic implications. Clin Cancer Res. 2012;18:1506–12. https://doi.org/10.1158/1078-0432.Ccr-11-1469.
5. Lange SM, Armstrong LA, Kulathu Y. Deubiquitinases: from mechanisms to their inhibition by small molecules. Molecular cell Deubiquitinases. 2022;82:15–29. https://doi.org/10.1016/j.molcel.2021.10.027.
6. Eifler K, Vertegaal ACO. SUMOylation-mediated regulation of cell cycle progression and cancer. Trends Biochem Sci. 2015;40:779–93. https://doi.org/10.1016/j.tibs.2015.09.006.
7. Mo YY, Yu Y, Theodosiou E, Ee PL, Beck WT. A role for Ubc9 in tumorigenesis. Oncogene. 2005;24:2677–83. https://doi.org/10.1038/sj.onc.1208210.
8. Moschos SJ, Jukic DM, Athanassiou C, Bhargava R, Dacic S, Wang X, Kuan SF, Fayewicz SL, Galambos C, Acquafondata M, et al. Expression analysis of Ubc9, the single small ubiquitin-like modifier (SUMO) E2 conjugating enzyme, in normal and malignant tissues. Hum Pathol. 2010;41:1286–98. https://doi.org/10.1016/j.humpath.2010.02.007.
9. Xia QD, Sun JX, Xun Y, Xiao J, Liu CQ, Xu JZ, An Y, Xu MY, Liu Z, Wang SG, Hu J. SUMOylation pattern predicts prognosis and indicates tumor microenvironment infiltration characterization in bladder cancer. Front Immunol. 2022;13: 864156. https://doi.org/10.3389/fimmu.2022.864156.
10. Sun JX, An Y, Xiang JC, Xu JZ, Hu J, Wang SG, Xia QD. The prognosis-predictive and immunoregulatory role of sumoylation related genes: potential novel targets in prostate cancer treatment. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms241713603.
11. Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. J Vis Exp. 2021. https://doi.org/10.3791/62528.
12. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformat GSVA. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7.
13. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformat ConsensusClusterPlus. 2010;26:1572–3. https://doi.org/10.1093/bioinformatics/btq170.
14. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33:1–22.
15. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS ONE. 2014;9: e107468. https://doi.org/10.1371/journal.pone.0107468.

16. Kessler JD, Kahle KT, Sun T, Meerbrey KL, Schlabach MR, Schmitt EM, Skinner SO, Xu Q, Li MZ, Hartman ZC, et al. A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. Science. 2012;335:348–53. https://doi.org/10.1126/science.1212728.
17. Sun XX, Chen Y, Su Y, Wang X, Chauhan KM, Liang J, Daniel CJ, Sears RC, Dai MS. SUMO protease SENP1 deSUMOylates and stabilizes c-Myc. Proceed Nat Acad Sci. 2018;115:10983–8. https://doi.org/10.1073/pnas.1802932115.
18. Toropainen S, Malinen M, Kaikkonen S, Rytinki M, Jääskeläinen T, Sahu B, Jänne OA, Palvimo JJ. SUMO ligase PIAS1 functions as a target gene selective androgen receptor coregulator on prostate cancer cell chromatin. Nucleic Acids Res. 2015;43:848–61. https://doi.org/10.1093/nar/gku1375.
19. Liu J, Wu Z, Han D, Wei C, Liang Y, Jiang T, Chen L, Sha M, Cao Y, Huang F, et al. Mesencephalic astrocyte-derived neurotrophic factor inhibits liver cancer through small ubiquitin-related modifier (SUMO)ylation-related suppression of NF-κB/Snail signaling pathway and epithelial-mesenchymal transition. Hepatology. 2020;71:1262–78. https://doi.org/10.1002/hep.30917.
20. Puttick C, Jones TP, Leung MM, Galvez-Cancino F, Liu J, Varas-Godoy M, Rowan A, Pich O, Martinez-Ruiz C, Bentham R, et al. MHC hammer reveals genetic and non-genetic HLA disruption in cancer evolution. Nat Genet. 2024;56:2121–31. https://doi.org/10.1038/s41588-024-01883-8.
21. Zhang Y, Zhao Y, Li Q, Wang Y. Macrophages, as a promising strategy to targeted treatment for colorectal cancer metastasis in tumor immune microenvironment. Front Immunol. 2021;12: 685978. https://doi.org/10.3389/fimmu.2021.685978.
22. Wang YN, Wang YY, Wang J, Bai WJ, Miao NJ, Wang J. Vinblastine resets tumor-associated macrophages toward M1 phenotype and promotes antitumor immune response. J Immunother Cancer. 2023. https://doi.org/10.1136/jitc-2023-007253.
23. Kang JH, Zappasodi R. Modulating Treg stability to improve cancer immunotherapy. Trends in cancer. 2023;9:911–27. https://doi.org/10.1016/j.trecan.2023.07.015.
24. Shang B, Liu Y, Jiang SJ, Liu Y. Prognostic value of tumor-infiltrating FoxP3+ regulatory T cells in cancers: a systematic review and meta-analysis. Sci Rep Prognostic. 2015;5:15179. https://doi.org/10.1038/srep15179.
25. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic cells in cancer immunology and immunotherapy. Nat Rev. 2020;20:7–24. https://doi.org/10.1038/s41577-019-0210-z.
26. Demel UM, Böger M, Yousefian S, Grunert C, Zhang L, Hotz PW, Gottschlich A, Köse H, Isaakidis K, Vonficht D, et al. Activated SUMOylation restricts MHC class I antigen presentation to confer immune evasion in cancer. J Clin Investigat. 2022. https://doi.org/10.1172/jci152383.
27. Jiang X, Chen ZJ. The role of ubiquitylation in immune defence and pathogen evasion. Nat Rev Immunol. 2011;12:35–48. https://doi.org/10.1038/nri3111.
28. Polverino A, Coxon A, Starnes C, Diaz Z, DeMelfi T, Wang L, Bready J, Estrada J, Cattley R, Kaufman S, et al. AMG 706, an oral, multikinase inhibitor that selectively targets vascular endothelial growth factor, platelet-derived growth factor, and kit receptors, potently inhibits angiogenesis and induces regression in tumor xenografts. Can Res. 2006;66:8715–21. https://doi.org/10.1158/0008-5472.Can-05-4665.
29. Shen Q, Han Y, Wu K, He Y, Jiang X, Liu P, Xia C, Xiong Q, Liu R, Chen Q, et al. MrgprF acts as a tumor suppressor in cutaneous melanoma by restraining PI3K/Akt signaling. Signal Transduct Target Ther. 2022;7:147. https://doi.org/10.1038/s41392-022-00945-9.
30. Murai J, Huang SY, Das BB, Renaud A, Zhang Y, Doroshow JH, Ji J, Takeda S, Pommier Y. Trapping of PARP1 and PARP2 by clinical PARP inhibitors. Can Res. 2012;72:5588–99. https://doi.org/10.1158/0008-5472.Can-12-2753.
31. Somnay Y, Lubner S, Gill H, Matsumura JB, Chen H. The PARP inhibitor ABT-888 potentiates dacarbazine-induced cell death in carcinoids. Cancer Gene Ther. 2016;23:348–54. https://doi.org/10.1038/cgt.2016.39.