

Integrating machine learning and multi-omics to identify key SUMOylation molecular signature in sarcoma

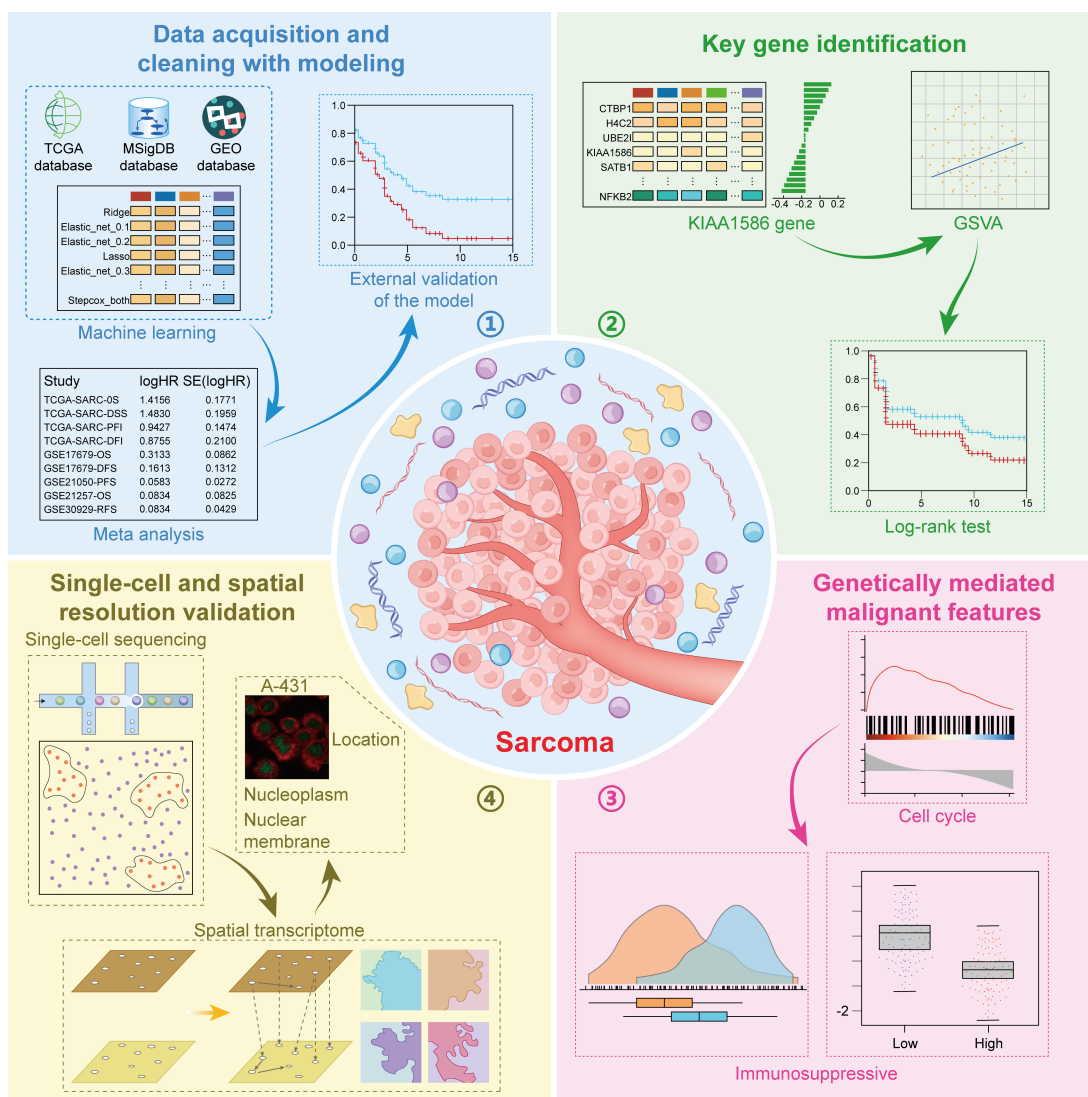
Authors

Jinhao Cheng, Zhitao He, Juehua Jing, Yuyao Liu, Haoxue Zhang

Correspondence

jjhhu@sina.com (J. Jing), ahmulyy@163.com (Y. Liu), haoxue0127@foxmail.com (H. Zhang)

Graphical Abstract



<https://doi.org/10.71321/lcfx.00001>

© 2024 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Integrating machine learning and multi-omics to identify key SUMOylation molecular signature in sarcoma

Jinhao Cheng¹, Zhitao He¹, Juehua Jing^{1*}, Yuyao Liu^{2*}, Haoxue Zhang^{3,4,5*}

Received: 2024-10-01 | Accepted: 2024-12-21 | Published online: 2024-12-30



Background: Sarcoma (SARC) is a rare and heterogeneous cancer originating from mesenchymal tissue. Due to its complex molecular mechanisms and limited treatment options, patients often have poor prognoses. Protein SUMOylation is an important post-translational modification process that plays a key role in regulating cellular functions and is closely related to the onset and progression of various cancers. However, the specific mechanisms by which SUMOylation affects SARC progression are not fully understood.

Methods: In this study, comprehensive bioinformatics approaches were utilized to analyze multiple datasets of SARC samples. By screening and identifying SUMOylation-related genes, we further explored the expression patterns of these genes in SARC and their association with prognosis and then constructed a consensus prognostic model. In particular, we focused on the KIAA1586 gene, which has attracted increasing attention in cancer biology, and conducted an in-depth study of its role in SARC.

Results: The study revealed that 19 SUMOylation-related genes were significantly correlated with the prognosis of SARC. Subsequently, the consensus prognostic model constructed by ridge regression could accurately predict the survival of patients in multiple data sets. Afterward, we identified KIAA1586 as the key gene, and its expression level was closely related to the prognosis of patients. GSEA enrichment analysis demonstrated that KIAA1586 might affect the progression of SARC by regulating the cell cycle and immune-related pathways, providing new insights into the molecular mechanism of SARC.

Conclusion: We have constructed a SUMOylation signature model that can accurately predict the prognosis of SARC patients, and identified KIAA1586 as a key SUMOylation gene that plays a crucial role in the onset and development of tumors by participating in cell cycle regulation and immune suppression.

Keywords: SUMOylation; Sarcoma; KIAA1586; prognosis; bioinformatics

Introduction

As a type of malignant tumor originating from mesenchymal tissue, sarcoma is highly heterogeneous and complex [1]. Epidemiologically, sarcomas are relatively rare. Current treatments mainly include surgery, (neo)adjuvant chemotherapy, and/or radiotherapy. However, the mortality rate cannot be ignored [2]. Due to the diverse pathological types, varied clinical manifestations, and significant differences in the responsiveness to conventional treatments of sarcomas, the prognosis of patients is often poor, with a survival rate of approximately 12 to 18 months [3]. Although certain progress has been made in the diagnosis and treatment of sarcomas in recent years, the overall survival rate still needs to be improved [4-6]. An in-depth understanding of the molecular mechanisms of sarcomas, and the search for new therapeutic targets and prognostic markers, are of great significance for improving the prognosis of sarcoma patients.

Post-translational protein modification (PTM) is an important regulatory mechanism that occurs in cellular proteins during or after translation. These modifications can alter the conformation, stability, hydrophobicity, and charge state of proteins, thus influencing their functions in various biological

processes [7]. To date, more than 450 types of PTMs have been identified, including ubiquitination, acetylation, phosphorylation, and SUMOylation [8]. SUMOylation is a post-translational modification process that depends on the conjugation of the small ubiquitin-like modifier (SUMO) to the target protein. As a member of the ubiquitin-like protein family, SUMO weighs approximately 11 kDa and can conjugate to the target protein in the form of a single monomer, multiple monomers, or different types of polymers [9]. SUMOylation plays a crucial role in cellular life activities and is involved in regulating multiple biological processes such as transcription, DNA repair, cell cycle progression, and signal transduction [9]. Abnormal SUMOylation processes are closely related to the occurrence and development of various diseases, including cancer. In cancer, abnormal SUMOylation may lead to the activation of oncogenes, the inactivation of tumor suppressor genes, and cell cycle disorders, thus promoting the progression of cancer [10].

Although the role of protein SUMOylation in cancer has been extensively studied [11-15], the specific mechanism of its action in sarcoma remains unclear. As a highly heterogeneous type of cancer, the occurrence and development of sarcoma

¹ Department of Orthopedics, The Second Affiliated Hospital of Anhui Medical University, Hefei 230601, China.

² Bioinformatics R&D Department, Hefei GuangRe Biotechnology Co., Ltd, Hefei, Anhui Province, China.

³ Department of Dermatovenerology, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui Province, China.

⁴ Key Laboratory of Dermatology, Ministry of Education, Hefei, Anhui Province, China.

⁵ Inflammation and Immune Mediated Diseases Laboratory of Anhui Province, Anhui Medical University, Hefei, Anhui Province, China

* Corresponding Author.

may involve a variety of complex molecular mechanisms, including SUMOylation. Therefore, an in-depth exploration of the role of protein SUMOylation in sarcoma is of great significance for understanding the molecular mechanisms of sarcoma and identifying new therapeutic targets, and prognostic markers. In addition, developing personalized therapeutic signatures using SUMOylation-related genes also has potential application value for achieving individualized treatment of sarcoma.

This study aims to deeply explore the role of protein SUMOylation in the progression of sarcoma, with particular attention paid to the KIAA1586 gene, which has been attracting increasing attention in cancer biology. Through bioinformatics methods, we analyzed a comprehensive data set containing sarcoma samples to identify key SUMOylation-related genes and evaluate the expression pattern of KIAA1586 in sarcoma and its relationship with the prognosis of patients. Our study found that KIAA1586 plays an important role in sarcoma, and its expression level is closely related to the prognosis of patients. This novel finding not only provides a new perspective for understanding the molecular mechanism of sarcoma but also offers new ideas and potential intervention targets for the treatment of sarcoma. We believe that the results of this study will provide strong support for the individualized treatment of sarcoma.

Methods

Data acquisition

The TCGA data we utilized underwent rigorous standardization, normalization, batch correction, and platform correction as part of The Pan-Cancer Atlas, accessible via <https://gdc.cancer.gov/about-data/publications/pancanatlas>. We meticulously extracted samples from the SARC (Sarcoma) dataset and retrieved corresponding survival data from the UCSC Xena database (<https://xena.ucsc.edu/>). To validate survival outcomes, we incorporated external datasets including overall survival (OS) data from GSE17679 (88 tumor samples), GSE21257 (53 tumor samples), E-TABM-1202 (101 tumor samples), GSE59455 (122 tumor samples), and GSE119041 (50 tumor samples); progression-free survival (PFS) data from GSE21050 (310 tumor samples), GSE71118 (312 tumor samples), GSE71119 (132 tumor samples), and GSE71120 (41 tumor samples); disease-free survival (DFS) data from GSE17679 (also included in the OS datasets); and relapse-free survival (RFS) data from GSE39055 (37 tumor samples) and GSE30929 (140 tumor samples). By harmonizing these data with the expression files from the SARC dataset, we curated a comprehensive cohort for downstream analysis. The 17 SUMOYLATION gene sets sourced from the MSigDB database (<https://www.gsea-msigdb.org/gsea/msigdb>) encompass a comprehensive range of processes related to protein SUMOylation. We have obtained a total of 227 genes that are related to SUMOylation.

Data cleaning and preprocessing

Data cleaning steps included removing missing values and non-tumor samples to ensure data integrity and accuracy. Additionally, survival times were converted from days to years to standardize the time units. All validation datasets underwent z-score normalization to transform the data into

a normal distribution with a mean of 0 and a variance of 1, eliminating dimensional differences between variables. To further optimize the data distribution characteristics, the exponential function \exp was used to convert z-scores into their exponential form.

Identifying key genes involved in SUMOylation

We conducted univariate Cox survival analysis on overall survival, disease-specific survival, and progression-free interval using the `Coxph` function, and screened for intersecting genes that showed significant correlation across all three survival outcomes through a Venn diagram.

Constructing a consensus prognostic model

We chose a linear model to model the input genes due to its simplicity, intuitiveness, and interpretability, which clearly reveals the specific contribution of each gene to prognosis. Multiple modeling algorithms were employed to construct prognostic models, including Lasso regression, Elastic Net, Ridge regression, stepwise Cox regression, and CoxBoost. Lasso regression was implemented using the `glmnet` package, with the family parameter specified as `Cox` and the alpha parameter set to 1. The `cv.glmnet` function was used to perform ten-fold cross-validation to select the optimal λ value. Model coefficients and feature names corresponding to the optimal λ value were extracted from the training results, and non-zero coefficients and their corresponding gene names were filtered out. Elastic Net and Ridge regression were also implemented using the `glmnet` package, with the alpha parameter for Elastic Net ranging from 0 to 1 (i.e., 0.1 to 0.9) and set to 0 for Ridge regression. Stepwise Cox regression involved first constructing a multivariate Cox regression model using the `Coxph` function, followed by stepwise regression analysis using the `stepAIC` function, with direction parameters including both, forward, and backward. For the CoxBoost model, we first optimized the penalty parameter using the `optimCoxBoostPenalty` function, then performed cross-validation using the `cv.CoxBoost` function to select the optimal number of steps. Finally, the CoxBoost function was used to construct the final CoxBoost model with the optimal number of steps and penalty parameter. Model coefficients and corresponding genes could be extracted using the `coef` function or obtained from the regression coefficient slot of the corresponding model. The linear combination of gene expression data and model coefficients was calculated to generate a risk score for each sample.

Evaluating the predictive performance of the prognostic model

The area under the receiver operating characteristic curve (ROC) and the area under the curve (AUC) were used as evaluation metrics, and the `timeROC` package was used to calculate AUC values at different time points to assess the performance of multiple prognostic models. Univariate Cox analysis was performed using the `Coxph` function to calculate the hazard ratio (HR) of the risk score calculated by the top-ranked algorithm across different datasets, followed by Meta-analysis and Kaplan-Meier survival analysis of the risk score.

Assessment of KIAA1586 as a key risk gene for SARC

We employed a multi-step approach for identification and validation. Firstly, we calculated the coefficients for each

gene using various algorithms and presented the results in a heatmap, where a coefficient of 0 indicates that the gene was not included in the model for a particular algorithm. Next, we conducted a univariate Cox survival analysis for KIAA1586 and performed a meta-analysis using the inverse variance method, with the log-hazard ratio (HR) as the primary measure. The HR indicates the tendency of the gene's effect, with values less than 1 suggesting a tumor-suppressive effect and values greater than 1 indicating a pro-oncogenic effect. Statistical analyses and visualizations were performed in the R environment using the "Meta" package. Additionally, we reflected the activity of given pathways by integrating characteristic gene expression, calculating combined z-scores for 14 different functional states of tumor cells based on the z-score algorithm. These scores were standardized and defined as gene set scores, followed by the calculation of Pearson correlations between KIAA1586 and each gene set score. Finally, to further explore the relationship between KIAA1586 expression levels and patient survival rates, we performed a Kaplan-Meier survival analysis using the survival package. The optimal cutoff value for distinguishing between high and low-expression cohorts was determined using the R package "survminer," ensuring that the proportion of high and low-expression groups was no less than 30% of the total sample. The log-rank test was used to assess the significance of survival differences between the two groups. DEPMap (Cancer Dependency Map) is a research project aimed at creating a detailed map of cancer cell dependencies. CERES stands for CRISPR Essentiality Screen, a method used to quantify gene essentiality in cancer cells. The CERES score is an indicator derived from this method, reflecting the impact of gene knockout on cell survival or proliferation. A negative CERES score indicates that knockout of the gene leads to growth arrest or death of cancer cells, typically implying that the gene is crucial for the survival or proliferation of cancer cells. We have identified the top 200 cancer cell lines with negative CERES scores.

Exploring the carcinogenic pathway of KIAA1586.

This study initially classified samples into high and low-expression groups based on the expression level of the KIAA1586 gene, with the top 30% of expressers designated as the high-expression group and the bottom 30% as the low-expression group. Subsequently, the limma package was utilized to perform differential analysis, calculating the log₂ fold change (log₂FC) for each gene and ranking all genes according to their log₂FC values. Following this, the clusterProfiler package was employed to conduct gene set enrichment analysis based on KEGG gene sets, GO-BP gene sets, GO-MF gene sets, GO-CC gene sets, Reactome gene sets, and WikiPathways gene sets, computing the enrichment score (ES) for each gene set and performing significance testing and multiple hypothesis testing on the ES values. To validate whether the cell cycle was enriched in the KIAA1586 high expression group (using the median value as the cutoff), we downloaded the cell cycle gene set from the MsigDB database (https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/WP_CELL_CYCLE) and conducted gene set enrichment analysis, calculating the ES and performing significance testing and multiple hypothesis testing. Additionally, the PROGENY

method from the easier package was used to compute scores for four immune-related pathways: JAK-STAT, NF-κB, TNF-α, and Trail. The specific features of these pathways were derived from studying gene expression changes during pathway perturbation experiments, and a linear regression model was employed for fitting. Finally, across multiple cohorts, various algorithms (including CIBERSORT, CIBERSORT ABS, EPIC, ESTIMATE, MCP-counter, Quantiseq, TIMER, and xCell) were applied to calculate the Spearman correlation between the KIAA1586 gene and different immune infiltrating cells.

Verifying the immunosuppressive potential of KIAA1586

To delve into the intrinsic relationship between KIAA1586 and immune-related expression signatures, we meticulously obtained a valuable dataset encompassing 68 published immune-related expression features from the UCSC Xena database. To precisely assess the correlation between KIAA1586 gene expression and these immune features, we skillfully employed the cor.test function in R to conduct detailed, pairwise correlation calculations between KIAA1586 expression levels and each immune-related feature and carefully extracted the Spearman correlation coefficients along with their corresponding p-values. To visually present the results of this correlation analysis, we cleverly utilized the hplot1 function from the fromto package to craft a vivid heatmap. Immune regulatory molecules play a pivotal role in cancer immunotherapy, and numerous agonists and antagonists of these molecules are undergoing rigorous evaluation in clinical oncology. To advance this research, there is an urgent need to understand the expression patterns and regulatory mechanisms of these molecules under different KIAA1586 states. To this end, we comprehensively investigated the expression of immune regulatory molecules, somatic copy number alterations (SCNA), and expression regulation through epigenetic pathways. To accurately assess MeTIL signatures, we adopted a principal component analysis (PCA) approach, skillfully converting individual methylation values of MeTIL markers into unitless MeTIL scores [16]. We standardized the data into Z-scores using the formula $(x-\mu)/\sigma$ and divided the samples into high and low-expression groups based on the median value of KIAA1586. Subsequently, we employed the Wilcoxon Rank Sum Tests to rigorously compare the statistical differences in MeTIL scores between the two groups. Furthermore, we adopted Spearman correlation analysis to meticulously calculate the correlation between genes and TIP scores, as well as the autocorrelation between TIP scores and utilized the linkET package for intuitive visualization [17].

Visualization and correlation analysis of gene expression in pan-cancer single-cell data

We obtained gene expression files at pan-cancer single-cell resolution from the TISCH database and constructed heatmaps using the pheatmap package to visualize the gene expression landscape at the pan-cancer single-cell level. Hierarchical clustering was performed using Euclidean distance as the metric and Ward's minimum variance method, which helped us to more clearly identify patterns and trends in the data and assess the conservation of KIAA1586 expression sources. For visualization of the gene expression data, we employed the Uniform Manifold Approximation and

Projection (UMAP) dimensionality reduction technique to display the distribution of cell types. For the visualization of single-cell transcriptome genes, we used the Nebulosa package. Due to the presence of numerous dropout events in single-cell transcriptome data, some genes have expression levels of zero or near-zero, even for marker genes. The Nebulosa package estimates weighted kernel densities, incorporates similarities between cells, and allows for the convolution of cellular features, thereby recovering lost gene signals and better presenting single-cell data. In addition, we calculated the average expression levels of KIAA1586, MKI67, CENPF, and PCNA in each single-cell dataset from the TISCH2 database, and used Spearman correlation analysis to assess the correlations between KIAA1586 and these three genes (MKI67, CENPF, PCNA).

Pan-cancer spatial transcriptome analysis revealed KIAA1586 expression landscape

Using the Sparkle database (<https://grswsci.top/>) and SpatialTME (<https://www.spatialtme.yelab.site/>), we conducted a pan-cancer spatial transcriptomics analysis [18]. The SpatialTME database utilized the Cottrazm package to deconvolute the cellular composition of the tumor microenvironment (TME) [19]. The Sparkle database integrated 10xVisium sequencing data from the SpatialTME database to construct a pan-cancer spatial transcriptomics atlas. Based on the proportion of cell types in each microregion, we named or characterized the microregions according to the dominant cell type and used a heat map to display the expression landscape of KIAA1586 across different cellular microregions. Subsequently, we selected sections from five tumors for further analysis. We presented the tissue sections, localized the dominant cell types after deconvolution, and identified the tumor boundaries. Additionally, we calculated the Spearman correlation between KIAA1586 expression and the proportion of different cell types in the microregions, as well as the difference in KIAA1586 expression among tumor, tumor boundary, and normal groups.

KIAA1586 expression and subcellular localization analysis

We systematically analyzed the immunohistochemistry staining results of KIAA1586 in various tumor tissues from the HPA database. HPA categorizes the staining intensity into four levels: High, Medium, Low, and Not detected, and we calculated the proportion of each level in detail. To more intuitively demonstrate the expression of KIAA1586 in specific tumors, we specifically selected immunohistochemistry staining sections of breast cancer and ovarian cancer for display. In addition, to further investigate the localization of KIAA1586 within cells, we conducted KIAA1586 immunofluorescence experiments using three cell lines (A-431, U-251MG, and U2OS) from the HPA database. By observing the immunofluorescence staining results, we visualized the subcellular localization of KIAA1586 within the cells.

Delve into the transcription and regulation of KIAA1586

We made use of Cistrome DB, also referred to as the Cistrome Data Browser, an interactive database that facilitates the visualization of public ChIP-seq, DNase-seq, and ATAC-seq data [20-21]. With the robust support of the extensive

Cistrome DB database, the Toolkit enables users to swiftly test their hypotheses regarding gene regulation using publicly available ChIP-seq data (for protein factors and histone marks) and chromatin accessibility data (DNase-seq and ATAC-seq). The Cistrome DB Toolkit website provides three core functionalities. It leverages the BETA algorithm [22] to compute a regulatory potential (RP) score, which gauges the probability of a factor regulating a gene. For ChIP-seq, DNase-seq, or ATAC-seq samples, BETA adopts a distance-weighted method to assess the regulatory potential of all binding sites of the factor within a specified distance from the target gene. Factors with high RP scores are likely candidates for regulating the given gene. We downloaded the potential regulatory scores for specific transcription factors from <http://dbtoolkit.cistrome.org/> and utilized the ggplot2 package for visualization. Pearson correlation analysis was applied to evaluate the correlation between transcription factors and KIAA1586. To examine the correlation between copy number variation scores and gene expression levels, we employed scatter plot analysis in conjunction with the Spearman rank correlation coefficient. A scatter plot serves as a graphical depiction that intuitively illustrates the relationship between two variables, enabling us to discern whether a linear or nonlinear relationship exists between them. Meanwhile, the Spearman rank correlation coefficient constitutes a non-parametric statistical method used to measure the monotonic relationship between two variables, irrespective of the data distribution. The Spearman rank correlation coefficient spans from -1 to 1, with values nearer to 1 or -1 indicating a more pronounced correlation, and the p-value is utilized to ascertain the significance of this correlation.

Results

Nineteen key SUMOylation Genes were identified and a robust SUMOylation signature was constructed

The Venn diagram reveals that there are 19 genes, specifically UBE2I, NDC1, SEC13, RELA, TOLLIP, UHRF2, AURKA, AURKB, BLM, CDCA8, CTBP1, DNMT3B, H4C2, MDC1, MRTFA, NFKB2, NR3C2, SATB1, and KIAA1586, that exhibit significant associations across multiple survival metrics (Figure 1A). Supplementary Figure 1 presents a forest plot displaying the results of univariate Cox analyses for SUMOylation-related genes in terms of overall survival, disease-specific survival, and progression-free survival. The heatmap presents the average AUC values at 1, 3, and 5 years for prognostic models constructed using different algorithms, with the algorithms ranked from highest to lowest average AUC values. The top-performing algorithm, Ridge Regression, exhibits the best performance across multiple datasets (Figure 1B). The meta-analysis results synthesize the findings of the Ridge Regression model across different survival outcomes in multiple datasets, enhancing statistical power and the reliability of the conclusions. The results indicate that the Ridge Regression model is a significant risk factor for various survival outcomes across multiple datasets (Figure 1C). Figure 1D-F present the Kaplan-Meier (KM) survival analysis results of the Ridge Regression model across different datasets, demonstrating that patients in the high-score or high-risk group have poorer prognoses ($p < 0.05$).

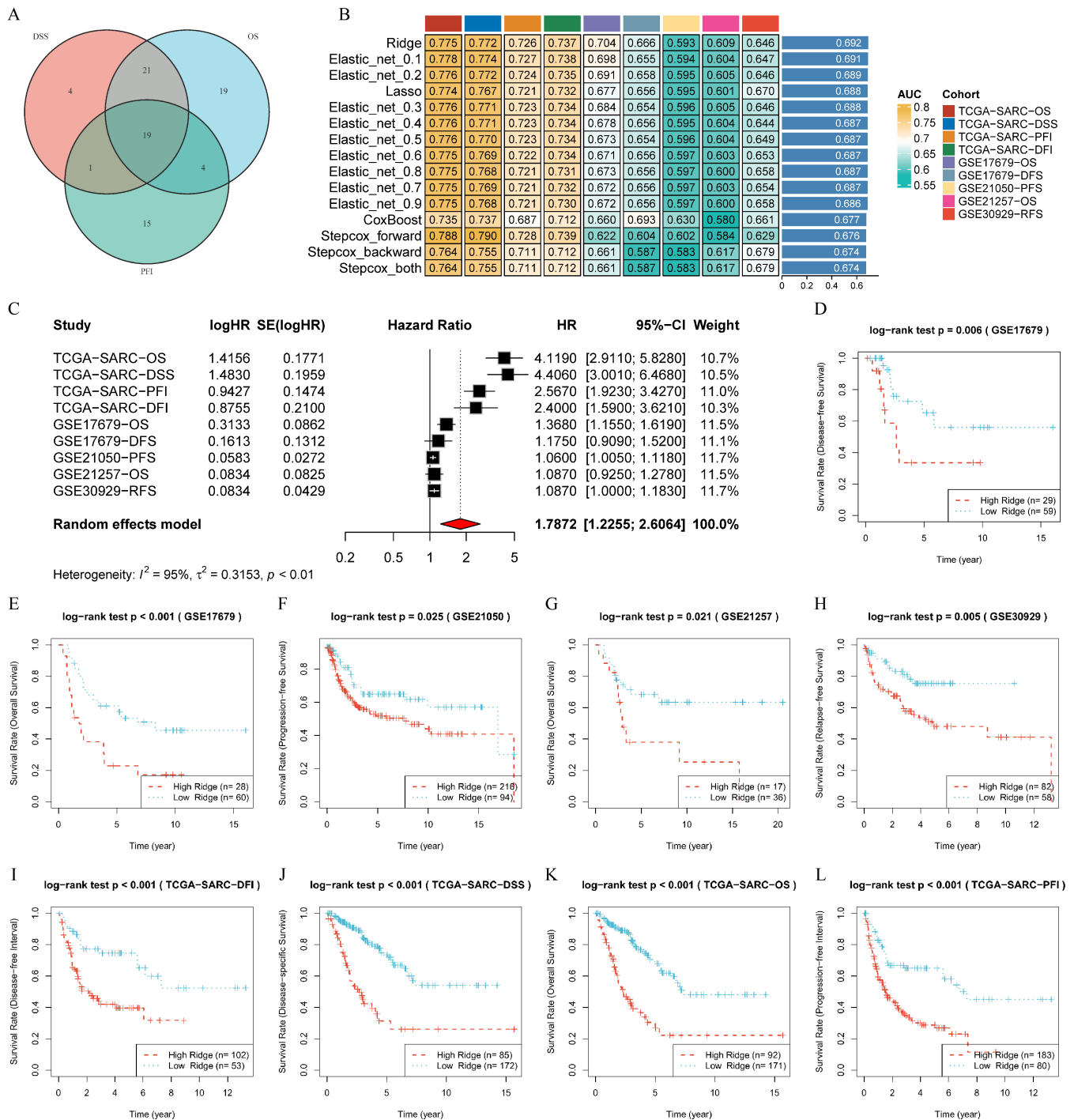


Figure 1. Identification of Key SUMOylation Genes and Construction of a Prognostic Prediction Model. (A) Presents 19 intersecting genes that exhibited significant correlation across overall survival, disease-specific survival, and progression-free interval, screened through univariate Cox survival analysis. (B) Comparison of the average AUC values at 1, 3, and 5 years for prognostic models constructed using different algorithms: This heatmap displays the average AUC (Area Under the Curve) values at 1, 3, and 5 years for prognostic models built with various algorithms. The algorithms are ranked from top to bottom based on their average AUC values, with the top-performing algorithm exhibiting the highest average AUC across multiple datasets. This ranking aids in identifying the most stable prognostic model algorithm over time. (C) Meta-analysis results of the hazard ratios for the prognostic model are presented, which helps synthesize findings from multiple independent studies, enhancing statistical power and the reliability of conclusions. (D) Kaplan-Meier survival analysis results for the final prognostic model, namely the Ridge Regression model, are shown, revealing a significant difference between the high-risk and low-risk groups.

KIAA1586 is a risk gene for SARC

Across all models, KIAA1586 exhibited remarkable consistency with positive coefficients, indicating that higher expression levels contribute to increased risk scores, suggesting that KIAA1586 is a risk gene for SARC (Figure 2A). The results of the meta-analysis revealed that the combined HR for KIAA1586 was 1.23, with a 95% confidence interval ranging from 1.11 to 1.37, confirming KIAA1586 as a risk gene for SARC, despite moderate heterogeneity observed across different survival outcomes (Figure 2B). Pearson correlation analysis uncovered associations between KIAA1586 expression levels and multiple

cancer-related gene set scores, particularly showing significant positive correlations with cell cycle, DNA damage, and repair pathways (Figure 2C). Kaplan-Meier survival analysis further substantiated the significant link between KIAA1586 expression levels and the prognosis of SARC patients, demonstrating a consistent pattern across 10 survival cohorts where higher expression was associated with poorer outcomes, reinforcing the evidence that KIAA1586 is a risk gene. Supplementary Figure 2 suggests that KIAA1586 has a negative CERES score in a large number of cell lines, implying that knockout of KIAA1586 leads to growth arrest or death of these cell lines.

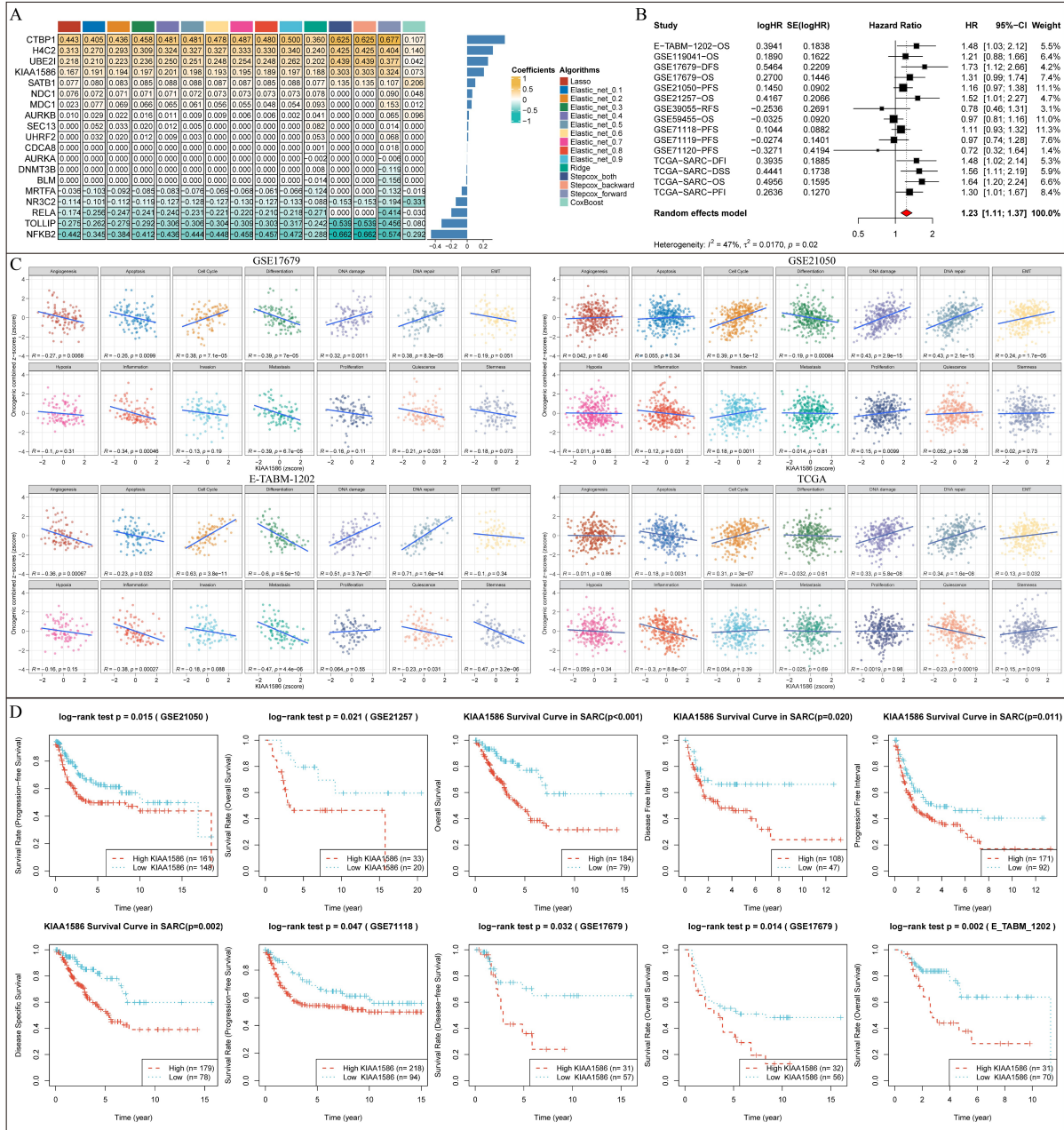


Figure 2. Identification and Validation of KIAA1586 as a Key Risk Gene for SARC. (A) The heatmap presents the coefficients of each gene across multiple algorithms, with a coefficient of 0 indicating that the gene was not involved in the model construction. (B) The results of the meta-analysis reveal the combined log-HR value from the univariate Cox survival analysis of KIAA1586, suggesting its role as a risk gene for SARC. (C) Dataset from the CancerSEA database, combined with the z-score algorithm, were used to calculate the combined z-scores reflecting 14 functional states of tumor cells, and their correlation with KIAA1586 expression. (D) The Kaplan-Meier curves illustrate the survival distribution of SARC patients grouped by KIAA1586 expression levels, with the log-rank test assessing the significance of survival differences between the two groups.

KIAA1586 is involved in tumor progression by participating in cell cycle and inhibiting immunity

Gene Set enrichment analysis revealed significant differences in biological processes and pathways between the high and low-expression groups of KIAA1586 (Figure 3A, B). Specifically, cell cycle-related pathways were significantly enriched in the high-expression group of KIAA1586, whereas immune-related pathways were more prominently enriched in the low-expression group. These findings suggest that the expression level of KIAA1586 may play a regulatory role in cellular proliferation and immune responses. To further validate these observations, we conducted an independent enrichment analysis focusing on the cell cycle gene set, which was corroborated by external datasets (Figure 3C). This analysis reaffirmed the association between high KIAA1586 expression

and increased cell cycle activity. Moreover, the scoring of four key immune signaling pathways—JAK-STAT, NF-κB, TNF-α, and Trail—indicated higher activity levels in the low-expression group of KIAA1586 (Figure 3D). This suggests that reduced expression of KIAA1586 may enhance immune system activation. Finally, Spearman correlation analysis provided additional support, demonstrating a negative correlation between KIAA1586 and components of the immune cells with tumor-killing functions (Figure 3E). This not only underscores the potential immunosuppressive capability of KIAA1586 but also highlights its importance as a potential target for cancer therapy. Collectively, our analyses offer new insights into the role of KIAA1586 in tumorigenesis and development, suggesting it could be a critical factor influencing immune reactions within the tumor microenvironment.

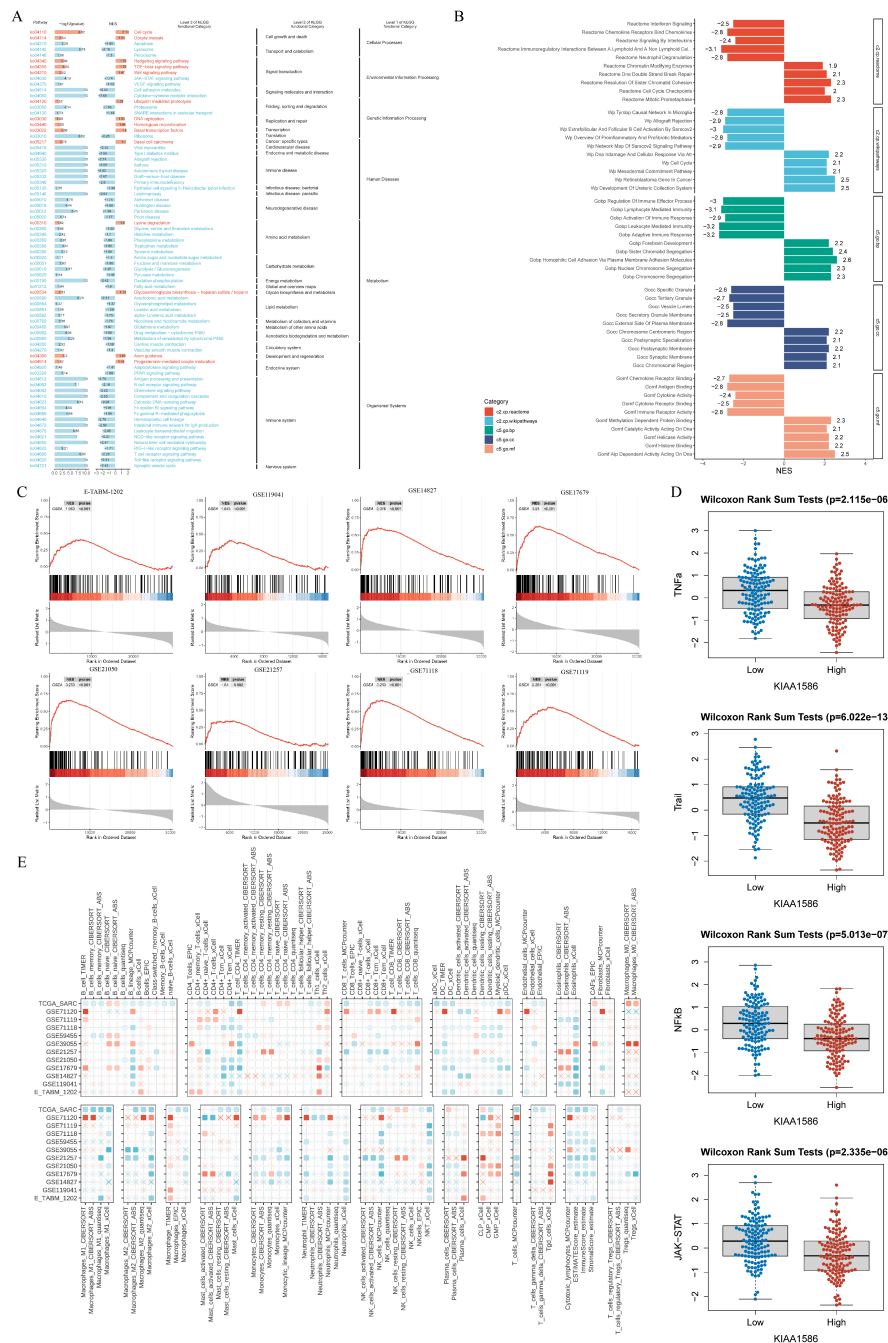


Figure 3. Impact of KIAA1586 Expression Levels on Gene Expression Patterns and Pathway Activity. (A) The pathways significantly enriched in the KIAA1586 high-expression group are represented in red, while those enriched in the low-expression group are shown in blue. A more significant p-value indicates a greater number of enriched genes, resulting in a longer bar graph. Additionally, a semantic summary of each pathway is provided. (B) The distribution of the top five significantly enriched pathways from the GO and Reactome/Wikipathways databases in the KIAA1586 high/low-expression groups is illustrated. Different colors represent distinct gene sets. Bars pointing to the left indicate significant enrichment in the low-expression group, whereas bars pointing to the right signify enrichment in the high-expression group. (C) The enrichment of gene sets at the top indicates that the core molecules within the custom gene set are primarily concentrated in the high-expression group on the left side of the KIAA1586 spectrum. This suggests that the target pathways are significantly enriched, and thus activated, in the high-expression group, while they are inhibited in the low-expression group. (D) A box plot compares the Z-Score values of four key immune pathways in the KIAA1586 high/low-expression groups, reflecting the relative activity of these pathways in different expression groups. (E) A heatmap presents the Spearman correlation between KIAA1586 and various types of immune infiltrating cells, highlighting its negative correlation with tumor-killing immune cell components. Red represents a positive correlation, blue indicates a negative correlation, and squares denote significant correlations ($p < 0.05$); otherwise, the correlation is considered insignificant.

KIAA1586 expression correlates with immune regulation and suggests potential as an immunotherapy target

Through in-depth exploration using Spearman correlation analysis, we revealed a widespread and significant negative correlation between KIAA1586 expression and various immune-related features, strongly suggesting that KIAA1586 may play a crucial role in the immune regulation process (Figure 4A). To provide a more comprehensive picture, we meticulously mapped the expression profile of immune regulatory molecules, somatic copy number alterations (SCNA), and the expression landscape regulated by epigenetic mechanisms (Figure 4B). The results indicated that the expression levels of immune molecules were significantly elevated in samples with KIAA1586 expression levels in the top 25% and between 25% and 50%. Based on previous research findings, we know that the median MeTIL

score is significantly higher in tumors enriched with functional cytotoxic T lymphocytes (CTLs), establishing MeTIL score as a valid indicator for assessing CTL function. Further analysis showed that the MeTIL score was significantly lower in the KIAA1586 high-expression group, implying a close association between high KIAA1586 expression and diminished CTL function (Figure 4C). Additionally, we found that multiple immune scores, including CYT, also exhibited a decreasing trend in the KIAA1586 high-expression group (Figure 4D-G). More notably, there was a significant correlation between KIAA1586 expression and the scores of cancer immune steps (Figure 4H). These findings not only provide strong support for our in-depth exploration of KIAA1586 as a potential target for immunotherapy but also open up new research avenues and therapeutic strategies in the field of cancer treatment.

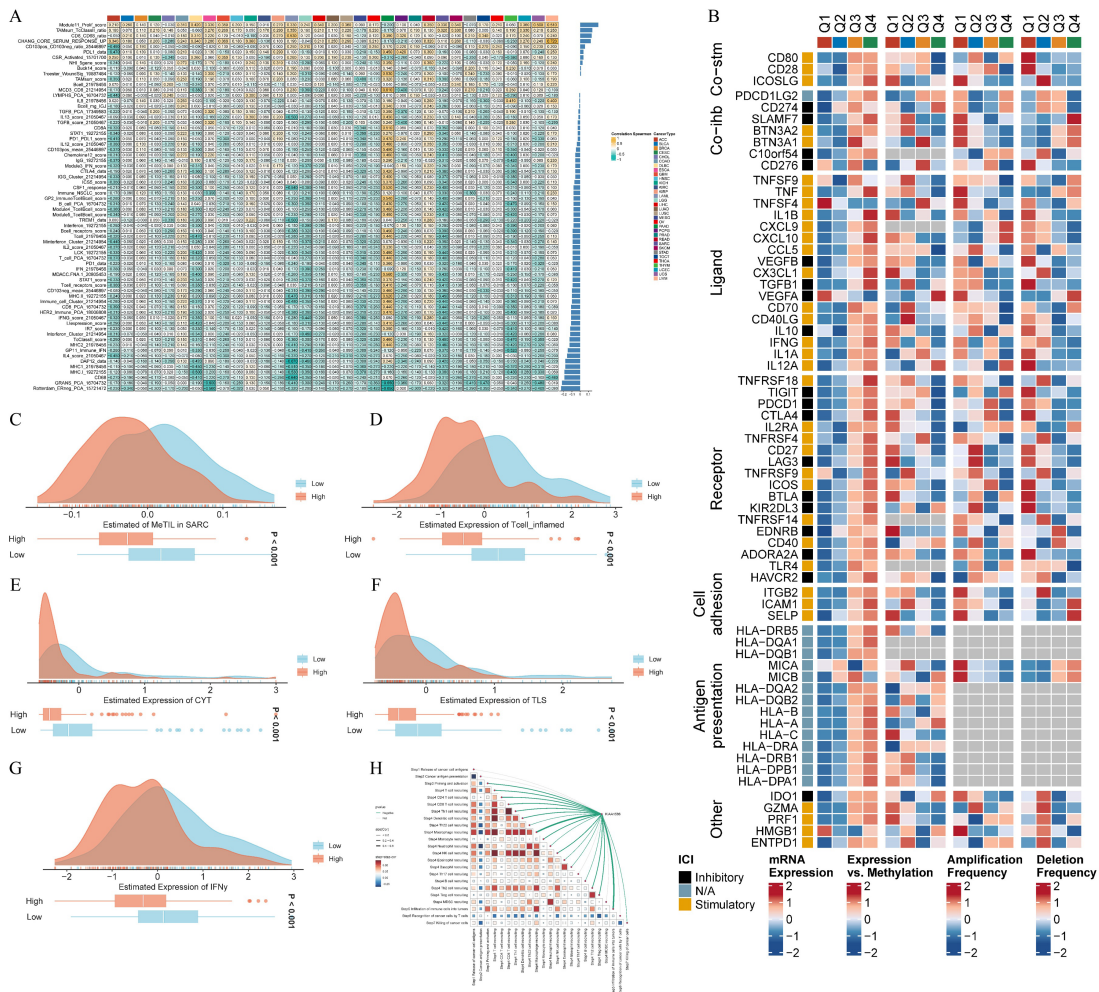


Figure 4. Validation of KIAA1586's Immunosuppressive Characteristics. (A) The heatmap depicts the Spearman correlation coefficients between specific genes and immune features. A larger absolute value of the correlation coefficient is represented by a deeper color, with shades closer to blue indicating negative correlation and otherwise indicating positive correlation. (B) Regulation of immune modulators. KIAA1586 expression is divided into quartiles: Q1, Q2, Q3, and Q4. Each component of the heatmap is arranged from left to right. mRNA expression levels are represented by the median of standardized expression values. The "Expression vs. Methylation" section shows the correlation between gene expression and DNA methylation β -values. Amplification frequency represents the difference between the proportion of samples with amplification of the immune modulator in a specific subtype and the proportion across all samples. Deletion frequency indicates the difference between the proportion of samples with deletion of the immune modulator in a specific subtype and the proportion across all samples. (C-G) Comparison of immune scores, including MeTIL, Tcell_inflamed, CYT, and TLS levels, between KIAA1586 high- and low-expression groups. The distribution of immune score levels for individual samples in the high- and low-expression groups is provided above each plot. The ends of the boxes below each plot represent the interquartile range of the values. The line within each box indicates the median value. Wilcoxon Rank Sum Tests were performed to compare the statistical differences in expression levels between the two groups.

KIAA1586 was expressed in malignant cells and co-expressed with proliferation genes

Our analysis revealed that KIAA1586 is predominantly expressed in malignant cells, demonstrating a significant advantage over immune cells (Figure 5A). To further explore the distribution of cell types and gene expression patterns, we employed the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique for visualizing the gene expression data. UMAP illustrated the distribution of cells and genes, clearly showing higher expression of KIAA1586

in clusters associated with malignant cells (Figure 5B-C). Additionally, we calculated the average expression levels of KIAA1586, MKI67, CENPF, and PCNA in each single-cell dataset from the TISCH2 database, and used Spearman correlation analysis to assess the correlation between KIAA1586 and these three proliferation-related genes. The results indicated a significant positive correlation between KIAA1586 and the three proliferation genes at the pseudo-bulk level, further confirming our previous findings that KIAA1586 is associated with proliferation (Figure 5D-F).

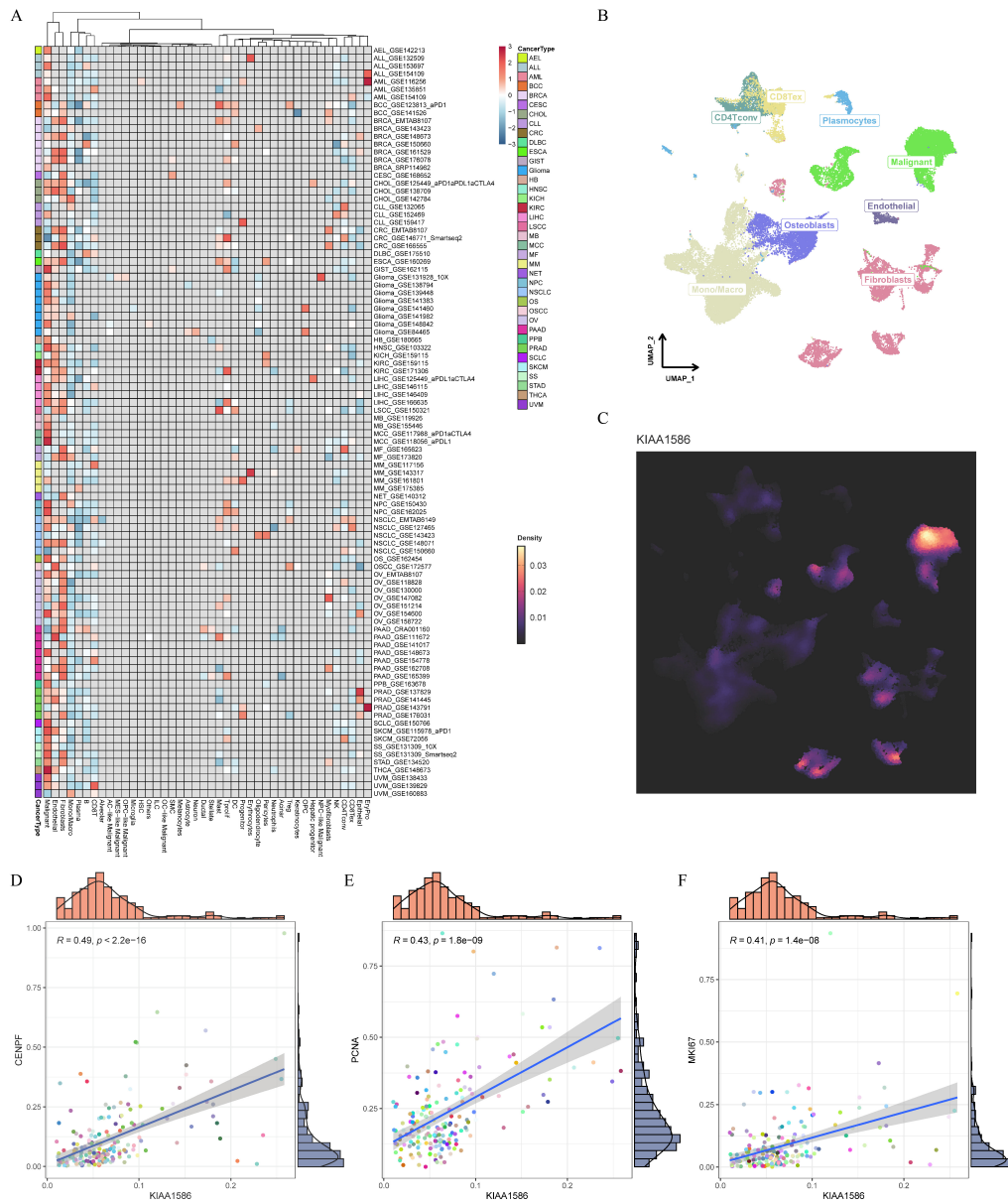


Figure 5. Expression of KIAA1586 in Malignant Cells and Its Co-expression with Proliferation Genes. (A) The heatmap visually presents the relative expression of KIAA1586 across different datasets and cell types. We observe a striking consistency in the expression pattern of KIAA1586 across different tumor types or datasets: it is predominantly highly expressed in malignant cells, demonstrating a significant advantage over immune cells. (B-C) In the osteosarcoma dataset, we employed UMAP dimensionality reduction technique to vividly illustrate the distribution of cells and genes. The results show that regions with higher expression of KIAA1586 highly overlap with malignant cell clusters, further confirming the specific expression of KIAA1586 in malignant cells. (D-F) The scatter plots clearly display the Spearman correlation between the average expression levels of KIAA1586 and the three proliferation genes, MKI67, CENPF, and PCNA, in each single-cell dataset. This result once again demonstrates the close association between KIAA1586 and proliferation genes.

KIAA1586 expression predominant in malignant regions of tumors

KIAA1586 Predominantly Expressed in Malignant Regions (Supplementary Figure 3). In spatial transcriptomics sections of breast cancer, clear cell renal carcinoma, hepatocellular carcinoma, ovarian cancer, and cutaneous melanoma, we observed a significant positive correlation between KIAA1586 expression and the proportion of malignant cells in the

microregions, while a negative correlation was noted with other components, particularly immune cell content. Furthermore, we found that KIAA1586 expression exhibited a gradient decrease from tumor regions to tumor boundaries, and finally to normal regions (Figure 6A-E). These consistent results indicate that the deregulated expression of KIAA1586 and its resulting biological effects in the tumor microenvironment can be attributed to malignant cells.

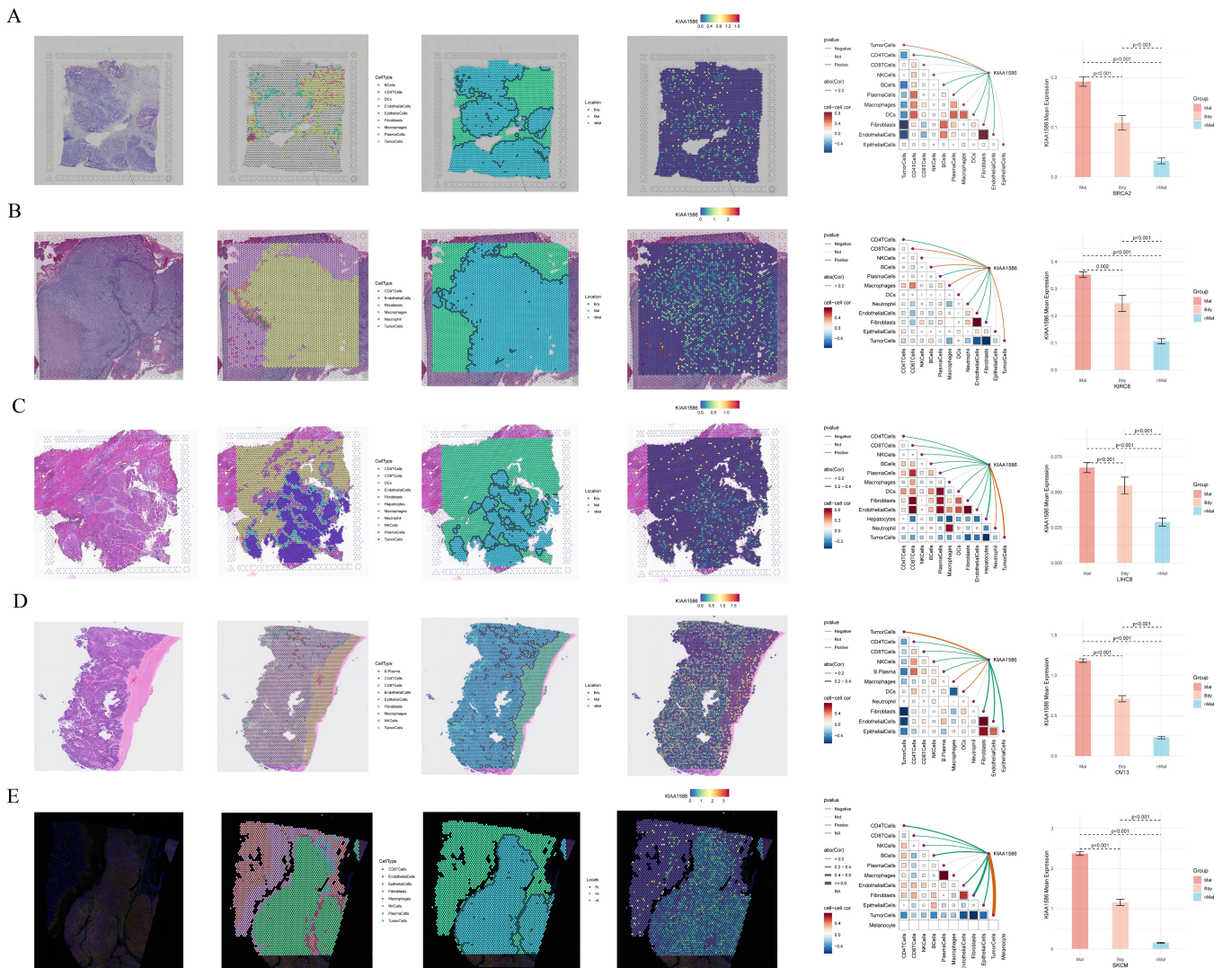


Figure 6. Predominant Expression of KIAA1586 in Malignant Regions of Tumors. (A-E) Each row of images is derived from breast cancer, clear cell renal carcinoma, hepatocellular carcinoma, ovarian cancer, and cutaneous melanoma, respectively. Each row contains six images, from left to right: 1) Tissue section serving as a blank control; 2) Each scatter point represents a microregion, named after the most abundant cell type, with different cell types represented by different colors; 3) Tumor boundary analysis, with different microregion types represented by different colors; 4) Each dot represents a spot from spatial transcriptomics sequencing, with deeper red indicating higher expression of the gene in that spot; 5) Correlation analysis, where red lines indicate positive correlation, green lines indicate negative correlation, gray lines indicate no significant correlation, and line thickness represents the absolute value of the correlation coefficient; the correlation in the triangular region is represented by the color intensity and size of the squares, with red indicating positive correlation, blue indicating negative correlation, and more significant p-values resulting in darker colors, larger absolute correlation coefficients, and larger squares; 6) Microregion differential analysis, with different groups represented by different colors, and the height of the bars representing the average expression level of each group, with differences analyzed using the Wilcoxon rank-sum test.

Expression and localization analysis of KIAA1586 in different tumors and cell lines.

As illustrated in Figure 7A, the expression levels of KIAA1586 vary across different tumor tissues. In tissues such as breast cancer, colorectal cancer, and ovarian cancer (Figure 7B-C), the expression distribution of KIAA1586 ranges from high to not detected. For instance, in breast cancer, 27% of the cases exhibit high expression, while 58% show medium expression, and the rest fall into other categories. Furthermore, in the A-431, U-251MG, and U2OS cell lines, KIAA1586 is primarily localized to the nucleoplasm and nuclear membrane (Figure 7D-F).

HDAC2 is a potential regulator of KIAA1586 expression in Sarcomas

The Cistrome DB database suggests that HDAC2 serves as a potential transcriptional regulator of KIAA1586. Additionally, there exists a correlation between the mRNA levels of these two genes, with a correlation coefficient of 0.31. In sarcomas (SARC), the Spearman's rank correlation coefficient between the copy number score of KIAA1586, as calculated by Gistic2, and its mRNA expression level is 0.45. This indicates that the expression of KIAA1586 in sarcomas is influenced by copy number variations (Supplementary Figure 4A-C).

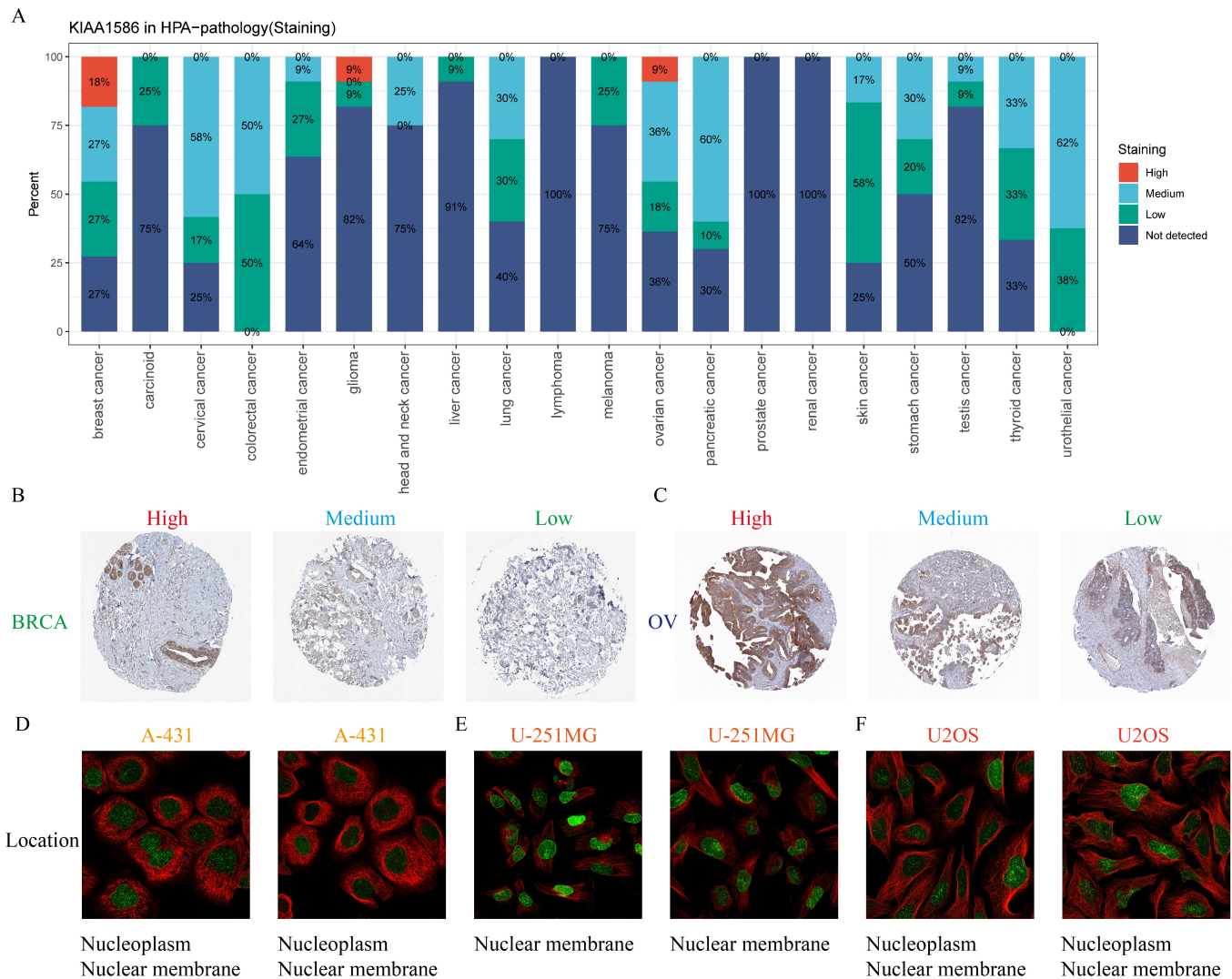


Figure 7. Comprehensive Analysis of KIAA1586 Expression and Localization in Different Tumor Tissues and Cell Lines. (A) The x-axis represents different tumor types, and the y-axis indicates the proportion of different staining intensities within a specific tumor. (B-C) Immunohistochemistry staining sections are presented, with darker brown indicating higher expression levels of KIAA1586. (D-E) Immunofluorescence staining sections are shown, with green fluorescence representing the target protein KIAA1586 and red fluorescence representing the microtubule structure.

Discussion

In this study, through comprehensive bioinformatics analysis, we deeply explored the role of protein SUMOylation in cancer progression, with a particular focus on SARC. Using the TCGA dataset and an external validation cohort, after careful data extraction and preprocessing, we identified 19 key SUMOylation-related genes. These genes showed significant correlations with multiple survival indicators, highlighting their importance in prognosis. Among these genes, KIAA1586 stood out as an important risk gene for SARC. This was confirmed by the consistent positive coefficients of KIAA1586 in various prognostic models and the results of meta-analysis.

We established a consensus prognostic model that includes 19 genes, namely AURKA, BLM, MRTFA, NR3C2, RELA, TOLLIP, NFKB2, UBE2I, NDC1, SEC13, UHRF2, AURKB, CDCA8, CTBP1, DNMT3B, H4C2, MDC1, SATB1, and KIAA1586. It is noteworthy that the risk coefficients of AURKA, BLM, MRTFA, NR3C2, RELA, TOLLIP, and NFKB2 are negative, indicating that these are protective genes. The risk coefficients of the remaining genes are positive, meaning that these are risk genes. Most of these genes have been found to be risk factors for multiple tumors or to be associated with tumor progression. UBE2I is upregulated in various cancers and is associated with cancer progression and poor prognosis. In hepatocellular carcinoma, the high expression of UBE2I is related to the enhanced migration and invasion ability of tumors. The deletion of UBE2I can significantly inhibit the migration and invasion of hepatocellular carcinoma cells [23]. AURKB plays an important role in the process of cell mitosis. Its abnormal expression is closely related to the occurrence and development of tumors. In some cancers, the high expression of AURKB is related to the proliferation of tumor cells and poor prognosis [24]. NDC1 promotes hepatocellular carcinoma tumorigenesis by targeting BCAP31 to activate PI3K/AKT signaling [25]. UHRF2 promotes hepatocellular carcinoma progression by upregulating the ErbB3/Ras/Raf signaling pathway [26], and it also promotes intestinal tumorigenesis through stabilization of TCF4 mediated Wnt/ β -catenin signaling [27]. CDCA8 facilitates tumor proliferation and predicts a poor prognosis in hepatocellular carcinoma. Additionally, KIF18B promotes the proliferation of pancreatic ductal adenocarcinoma via activating the expression of CDCA8. Moreover, a cell cycle-regulated chromosomal passenger protein with aberrant expression and nuclear accumulation is linked to poor prognosis for gastric cancer [28-30]. However, there are few studies on KIAA1586, and we have carried out key analysis and exploration.

The importance of KIAA1586 in cancer biology is further supported by its association with key cancer-related pathways, particularly the cell cycle and DNA damage repair pathways. Cancer is a group of diseases in which cells continue to divide excessively. Cancer-related mutations that disrupt cell cycle control achieve continuous cell division mainly by impairing the ability of cells to exit the cell cycle [31]. Gene Set Enrichment Analysis shows that high expression of KIAA1586 is significantly correlated with increased cell cycle activity. Through spatial transcriptomics and single-cell transcriptomics analyses, we found that KIAA1586 is mainly expressed in malignant tumor cells and has a significant positive correlation with proliferation-related genes such as MKI67, CENPF, and PCNA. This indicates

that KIAA1586 may influence cancer progression by promoting the proliferation of tumor cells. In addition, the high expression of KIAA1586 in the malignant regions of tumors further supports its central role in cancer development.

The JAK-STAT signaling pathway serves as a pivotal regulator of immune homeostasis [32]. NF- κ B transcription factors play a central role in regulating immunity and inflammation [33]. Anticancer immune surveillance and immunotherapies initiate the activation of cytotoxic cytokine signaling, encompassing the tumor necrosis factor alpha (TNF- α) and TNF-related apoptosis-inducing ligand (TRAIL) pathways [34]. Conversely, KIAA1586 is negatively correlated with immune-related pathways such as JAK-STAT, NF- κ B, TNF- α , and Trail. This suggests that KIAA1586 may affect cancer progression by suppressing the immune response. In terms of immune regulation, our study shows that the expression of KIAA1586 is widely and significantly negatively correlated with various immune-related features, indicating that KIAA1586 may play an important role in the process of immune regulation. According to previous research, in tumors enriched with functional cytotoxic T lymphocytes (CTLs), the median MeTIL score is significantly higher, indicating that the MeTIL score serves as a measure of CTL function [16]. Conversely, the MeTIL score is lower in the group with high KIAA1586 expression. This finding implies a close association between KIAA1586 and weakened CTL function, further underscoring the potential of KIAA1586 as a target for immunotherapy. Through further analysis of the expression profiles of immunomodulatory molecules, we provided more evidence for the role of KIAA1586 in immunosuppression.

Notably, our study also revealed the possibility of HDAC2 being a potential transcriptional regulator of KIAA1586. This regulatory relationship was not only verified at the mRNA level but also further supported by the correlation between copy number variations and the expression of KIAA1586. This finding provides a new perspective for understanding the abnormal expression of KIAA1586 in cancer and offers potential intervention targets for future research.

In summary, through multi-dimensional bioinformatics analysis, this study has deeply revealed the crucial role of KIAA1586 in SARC and its potential immunosuppressive mechanisms. These findings not only provide a new perspective for understanding cancer progression but also offer a scientific basis for the development of novel therapeutic strategies. In the future, research on KIAA1586 and its regulatory network may bring new breakthroughs in cancer treatment.

Conclusions

In this study, bioinformatics analysis showed that KIAA1586 was an important risk gene for SARC, and its high expression was associated with poor prognosis of patients. KIAA1586 affects sarcoma progression by promoting cell proliferation and inhibiting immune response. Spatial transcriptomics and single-cell analysis showed that KIAA1586 was mainly expressed in malignant tumor cells and positively correlated with proliferation genes. In addition, we found that HDAC2 may be a transcriptional regulator of KIAA1586. These findings provide new ideas for the treatment of sarcomas and a scientific basis for the development of new therapeutic

strategies for KIAA1586.

Acknowledgements

The authors are grateful for the valuable contribution of the TCGA and GEO project.

Author Contributions

Juehua Jing, Yuyao Liu and Haoxue Zhang designed the whole project and wrote the manuscript.

Jinhao Cheng performed the most of the analysis.

Zhitao He contributed to the investigation.

Jinhao Cheng and Yuyao Liu contributed to the methodology.

All authors read and approved the final manuscript.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

Abbreviations

Sarcoma: SARC; The Cancer Genome Atlas: TCGA; Molecular Signatures Database: MSigDB; Overall Survival: OS; Progression-Free Survival: PFS; Disease-Free Survival: DFS; Relapse-Free Survival: RFS; Receiver Operating Characteristic: ROC; Area Under the Curve: AUC; Hazard Ratio: HR; CRISPR Essentiality Screen: CERES; Cancer Dependency Map: DEPMAP; Kyoto Encyclopedia of Genes and Genomes: KEGG; Gene Ontology: GO; Principal Component Analysis: PCA; Methylation-based Immune Response Signature: MeTIL; Tumor Immunophenotype Profiling: TIP; Uniform Manifold Approximation and Projection: UMAP; Somatic Copy Number Alterations: SCNA; Cytotoxic T Lymphocytes: CTLs; Post-translational protein modification: PTM; Kaplan-Meier: KM; Tumor Microenvironment: TME

References

- [1] Chaudhary H, D'Angelo S. Role of Virus-Directed therapy in soft tissue sarcoma. *Curr Treat Options Oncol.* (2022) 23:404–14. doi: 10.1007/s11864-022-00956-2
- [2] HaDuong JH, Martin AA, Skapek SX, Mascarenhas L. Sarcomas. *Pediatr Clin North Am.* (2015) 62:179–200. doi: 10.1016/j.pcl.2014.09.012
- [3] Dajsakdipon T, Siripoon T, Ngamphaiboon N, Ativitavas T, Dejthepaporn T. Immunotherapy and biomarkers in sarcoma. *Curr Treat Options Oncol.* (2022) 23:415–38. doi: 10.1007/s11864-022-00944-6
- [4] Hall F, Villalobos V, Wilky B. Future directions in soft tissue sarcoma treatment. *Curr Probl Cancer.* (2019) 43:300–7. doi: 10.1016/j.crrproblcancer.2019.06.004
- [5] Nakata E, Fujiwara T, Kunisada T, Ito T, Takihira S, Ozaki T. Immunotherapy for sarcomas. *Jpn J Clin Oncol.* (2021) 51:523–37. doi: 10.1093/jjco/hyab005
- [6] Pollack SM, Ingham M, Spraker MB, Schwartz GK. Emerging targeted and Immune-Based therapies in sarcoma. *J Clin Oncol.* (2018) 36:125–35. doi: 10.1200/JCO.2017.75.1610
- [7] Venne A.S., Kollipara L., Zahedi R.P. The next level of complexity: Crosstalk of posttranslational modifications. *Proteomics.* 2014;14:513–524. doi: 10.1002/pmic.201300344.
- [8] Han Z.J., Feng Y.H., Gu B.H., Li Y.M., Chen H. The post-translational modification, SUMOylation, and cancer (Review) *Int. J. Oncol.* 2018;52:1081–1094. doi: 10.3892/ijo.2018.4280.
- [9] Vertegaal A.C.O. Signalling mechanisms and cellular functions of SUMO. *Nat. Rev. Mol. Cell Biol.* 2022;23:715–731. doi: 10.1038/s41580-022-00500-y.
- [10] Du L., Li Y.-J., Fakhri M., Wiatrek R.L., Duldulao M., Chen Z., Chu P., Garcia - Aguilar J., Chen Y. Role of SUMO activating enzyme in cancer stem cell maintenance and self-renewal. *Nat. Commun.* 2016;7:12326. doi: 10.1038/ncomms12326.
- [11] Seeler J.-S., Dejean A. SUMO and the robustness of cancer. *Nat. Rev. Cancer.* 2017;17:184–197. doi: 10.1038/nrc.2016.143.
- [12] Eifler K., Vertegaal A.C. SUMOylation - Mediated Regulation of Cell Cycle Progression and Cancer. *Trends Biochem. Sci.* 2015;40:779–793. doi: 10.1016/j.tibs.2015.09.006.
- [13] Moschos S.J., Jukic D.M., Athanassiou C., Bhargava R., Dacic S., Wang X., Kuan S.-F., Fayewicz S.L., 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;41:1286–1298. doi: 10.1016/j.humpath.2010.02.007.
- [14] McDoniels - Silvers A.L., Nimri C.F., Stoner G.D., Lubet R.A., You M. Differential gene expression in human lung adenocarcinomas and squamous cell carcinomas. *Clin. Cancer Res.* 2002;8:1127–1138.
- [15] Tomasi M.L., Tomasi I., Ramani K., Pascale R.M., Xu J., Giordano P., Mato J.M., Lu S.C. S-adenosyl methionine regulates ubiquitin-conjugating enzyme 9 protein expression and SUMOylation in murine liver and human cancers. *Hepatology.* 2012;56:982–993. doi: 10.1002/hep.25701.
- [16] Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, Koch A, Larsimont D, Salgado R, Van den Eynden G, Willard Gallo K, Bontempi G, Defrance M, Sotiriou C, Fuks F. DNA methylation-based immune response signature improves patient diagnosis in multiple cancers. *J Clin Invest.* 2017 Aug 1;127(8):3090-3102. doi: 10.1172/JCI91095.
- [17] Xu L, Deng C, Pang B, Zhang X, Liu W, Liao G, Yuan H, Cheng P, Li F, Long Z, Yan M, Zhao T, Xiao Y, Li X. TIP: A Web Server for Resolving Tumor Immunophenotype Profiling. *Cancer Res.* 2018 Dec 1;78(23):6575-6580. doi: 10.1158/0008-5472.CAN-18-0689.
- [18] Shi J, Wei X, Xun Z, Ding X, Liu Y, Liu L, Ye Y. The Web-Based Portal SpatialTME Integrates Histological Images with Single-Cell and Spatial Transcriptomics to Explore the Tumor Microenvironment. *Cancer Res.* 2024 Apr 15;84(8):1210-1220. doi: 10.1158/0008-5472.CAN-23-2650.
- [19] Xun, Z., Ding, X., Zhang, Y. et al. Reconstruction of the tumor spatial microenvironment along the malignant-

- boundary-nonmalignant axis. *Nat Commun* 14, 933 (2023). <https://doi.org/10.1038/s41467-023-36560-7>.
- [20] Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, Chen CH, Brown M, Zhang X, Meyer CA, Liu XS. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res*, 2018 Nov 20. Doi: 10.1093/nar/gky1094
- [21] Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, Zhu M, Wu J, Shi X, Taing L, Liu T, Brown M, Meyer CA, Liu XS. Cistrome data browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res*, 2017 Jan 4;45(D1):D658-D662. Doi: 10.1093/nar/gkw983.
- [22] Wang S, Sun H, Ma J, Zang C, Wang C, Wang J, Tang Q, Meyer CA, Zhang Y, Liu XS. Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat Protoc*. 2013 Dec;8(12):2502-15. doi: 10.1038/nprot.2013.150.
- [23] Yang H, Gao S, Chen J, Lou W. UBE2I promotes metastasis and correlates with poor prognosis in hepatocellular carcinoma. *Cancer Cell Int*. 2020;20:234. Published 2020 Jun 12. doi:10.1186/s12935-020-01311-x
- [24] Ahmed A, Shamsi A, Mohammad T, Hasan GM, Islam A, Hassan MI. Aurora B kinase: a potential drug target for cancer therapy. *J Cancer Res Clin Oncol*. 2021;147(8):2187-2198. doi:10.1007/s00432-021-03669-5
- [25] Liu YP, Guo G, Ren M, et al. NDC1 promotes hepatocellular carcinoma tumorigenesis by targeting BCAP31 to activate PI3K/AKT signaling. *J Biochem Mol Toxicol*. 2024;38(2):e23647. doi:10.1002/jbt.23647
- [26] Sun J, Wu K, Chen S, Jiang S, Chen Y, Duan C. UHRF2 promotes Hepatocellular Carcinoma Progression by Upregulating ErbB3/Ras/Raf Signaling Pathway. *Int J Med Sci*. 2021;18(14):3097-3105. doi:10.7150/ijms.60030
- [27] Li L, Duan Q, Zeng Z, et al. UHRF2 promotes intestinal tumorigenesis through stabilization of TCF4 mediated Wnt/ β -catenin signaling. *Int J Cancer*. 2020;147(8):2239-2252. doi:10.1002/ijc.33036
- [28] Cui Y, Jiang N. CDCA8 Facilitates Tumor Proliferation and Predicts a Poor Prognosis in Hepatocellular Carcinoma. *Appl Biochem Biotechnol*. 2024;196(3):1481-1492. doi:10.1007/s12010-023-04603-w
- [29] Li B, Liu B, Zhang X, Liu H, He L. KIF18B promotes the proliferation of pancreatic ductal adenocarcinoma via activating the expression of CDCA8. *J Cell Physiol*. 2020;235(5):4227-4238. doi:10.1002/jcp.29201
- [30] Chang JL, Chen TH, Wang CF, et al. Borealin/Dasra B is a cell cycle-regulated chromosomal passenger protein and its nuclear accumulation is linked to poor prognosis for human gastric cancer. *Exp Cell Res*. 2006;312(7):962-973. doi:10.1016/j.yexcr.2005.12.015
- [31] Matthews HK, Bertoli C, de Bruin RAM. Cell cycle control in cancer. *Nat Rev Mol Cell Biol*. 2022;23(1):74-88. doi:10.1038/s41580-021-00404-3
- [32] Chaimowitz NS, Smith MR, Forbes Satter LR. JAK/STAT defects and immune dysregulation, and guiding therapeutic choices. *Immunol Rev*. 2024;322(1):311 - 328. doi:10.1111/imr.13312
- [33] Capece D, Verzella D, Flati I, Arboretto P, Cornice J, Franzoso G. NF- κ B: blending metabolism, immunity, and inflammation. *Trends Immunol*. 2022;43(9):757 - 775. doi:10.1016/j.it.2022.07.004
- [34] Sukocheva OA, Neganova ME, Aleksandrova Y, et al. Signaling controversy and future therapeutical perspectives of targeting sphingolipid network in cancer immune editing and resistance to tumor necrosis factor - α immunotherapy. *Cell Commun Signal*. 2024;22(1):251. doi:10.1186/s12964-024-01626-6