

Bioinformatics Identification of SPAG5 as a Potential Prognostic Biomarker in Diffuse Large B-Cell Lymphoma

Xinyu Yan, Lanxiang Liu, Junnan Li, Hongbin Zhang, Li Wang, Lin Liu

Department of Hematology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, People's Republic of China

Correspondence: Lin Liu, Department of Hematology, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuzhong District, Chongqing, 400016, People's Republic of China, Tel +86 13608357776, Email liulin@cqmu.edu.cn

Background: Diffuse large B-cell lymphoma (DLBCL) is the most prevalent form of non-Hodgkin's lymphoma globally. *SPAG5*, a mitotic spindle protein, plays a significant role in DLBCL, where its abnormal expression is often associated with tumor growth, chemotherapy resistance, local recurrence, and poor prognosis.

Methods: A comprehensive analysis of *SPAG5* expression across various cancer types was conducted using Timer 2.0 and Sanger Box 3.0. Subsequently, the expression levels of *SPAG5* in DLBCL were investigated in comparison to normal samples. Receiver operating characteristic (ROC) curve was then generated to evaluate the diagnostic performance of *SPAG5* for DLBCL. Furthermore, the functional role of *SPAG5* was characterized, and its impact on the immune microenvironment of DLBCL patients was analyzed. Its potential in predicting immune checkpoint status and responses to immunotherapy was also evaluated.

Results: *SPAG5* expression demonstrated significant heterogeneity across various cancer types, with a marked upregulation in DLBCL. The diagnostic efficacy of *SPAG5* was moderate, yielding an area under curve (AUC) of 0.75. *SPAG5* exerted a multifaceted influence on DLBCL progression by regulating critical cellular processes, including cell cycle dynamics, chromosomal segregation, and DNA homeostasis. Notably, patients with elevated *SPAG5* expression had poorer survival outcomes than those with low expression. Analysis of the tumor immune microenvironment revealed a distinct pattern: high *SPAG5* expression correlated with increased infiltration of resting natural killer (NK) cells, while being associated with reduced presence of regulatory T cells (Tregs) and follicular helper T cells (Tfh).

Conclusion: Our bioinformatics study elucidated the expression profile, diagnostic potential, and prognostic significance of *SPAG5* in DLBCL, emphasizing the complex interplay between *SPAG5* expression and the tumor immune landscape. Our findings suggested *SPAG5* could be a candidate prognostic marker and potential therapeutic target for DLBCL.

Keywords: *SPAG5*, diffuse large B cell lymphoma, diagnostic markers, immune infiltration, therapeutic target

Introduction

Diffuse large B cell lymphoma (DLBCL) is the most common type of non-Hodgkin lymphoma (NHL), accounting for approximately 31%-34% of all NHL cases.¹ Originating from mature B cells, DLBCL is an aggressive and heterogeneous malignancy predominantly affecting individuals aged 50 to 70 years.² Meanwhile, relapse and drug resistance remain major challenges in the treatment of DLBCL. Current treatment modalities for DLBCL include chemotherapy, targeted therapy, and hematopoietic stem cell transplantation.³ Whereas, approximately 30%-40% of patients relapse within two years post-treatment⁴ —a statistic that underscores the limitations of existing therapeutic strategies. Consequently, there is an imperative necessity to identify dependable, non-invasive biomarkers for the early differentiation of DLBCL prognosis or the assessment of recurrence risk and ongoing disease monitoring, which could facilitate prompt therapeutic intervention, consequently enhancing long-term patient outcomes.

Sperm-associated antigen 5 (*SPAG5*) is a protein associated with mitotic spindles, playing a critical role in their function and dynamic regulation.⁵ High expression levels of *SPAG5* have been reported in various cancers, where it

serves as a biomarker for cancer development.⁶ And meanwhile, emerging evidence links SPAG5 expression to the tumor immune microenvironment (TIME) and immunotherapy responses across malignancies. For instance, in lung adenocarcinoma (LUAD), SPAG5 expression correlates with immune cell infiltration and the levels of immune checkpoint molecules such as PD-L1, CTLA4, GZMB, LAG3, and TIGIT.⁷ Mechanistically, *SPAG5* has been shown to promote the progression of gastric cancer and confer chemotherapy resistance through activation of the PI3K/AKT pathway.⁸ Additionally, targeting *SPAG5* has demonstrated potential therapeutic benefits in treating multiple myeloma.⁹ These collective findings strongly suggest that *SPAG5* plays a pivotal role in tumorigenesis.

Given its involvement in other malignancies, *SPAG5* expression may be closely related to DLBCL progression, and could serve as a potential target for auxiliary diagnosis or prognostic prediction. This bioinformatics study aims to investigate the diagnostic and prognostic value of *SPAG5* in DLBCL, as well as its impact on the immune landscape. Through enrichment analysis, we explore the pathways by which *SPAG5* affects DLBCL progression, thereby further establishing its foundation as a therapeutic target. [Supplementary Figure 1](#) depicted a flowchart detailing the cohort selection procedure and integrated design.

Materials and Methods

Data Preprocessing and Normalization

To guarantee the reliability and reproducibility of the analysis results, rigorous preprocessing was carried out on the raw data downloaded from public databases. The officially provided normalized expression data (transcripts per million, TPM) for 48 patients diagnosed with diffuse large DLBCL were gained from The Cancer Genome Atlas (TCGA) database (<https://tcga-data.nci.nih.gov/tcga/>). Clinical parameters included gender, age, weight, overall survival (OS), progression-free survival (PFS), and tumor stage. Samples with missing survival information or a survival time of 0 were excluded, leaving 47 samples for analysis. For comparative analysis, transcriptome data and clinical information of 107 lymphocyte samples and 337 whole blood samples were obtained from the Genotype-Tissue Expression (GTEx) database (<https://www.gtexportal.org>), amounting to 444 normal samples in total. Additionally, to validate our findings, the GSE56315 and GSE181063 datasets were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), which contained transcriptome microarray data from 88 and 1303 patients diagnosed with DLBCL, respectively, probes were converted to gene symbols based on probe annotations. Probes corresponding to multiple genes were eliminated, and for multiple probes corresponding to the same gene symbol, the average value was calculated. Expression values were normalized using the “normalizeBetweenArrays” function from the “limma” R package (v 3.56.2) to ensure similar distributions across arrays.

Expression Level of SPAG5 in Pan-Cancer

The differential expression (Diffexp) module of TIMER 2.0 (<http://timer.cistrome.org/>) and SangerBox 3.0 (<http://vip.sangerbox.com/login.html>) were utilized to analyze and visualize the expression levels of *SPAG5* across various cancer types. The expression data were displayed using both box plots and violin plots to provide comprehensive insights into the distribution and variability of *SPAG5* expression.

Expression Level of SPAG5 in DLBCL

We conducted an analysis to evaluate the expression profile of *SPAG5* between DLBCL and normal tissues by leveraging multiple datasets. Specifically, we utilized TCGA cohort combined with the GTEx cohort, as well as additional data from the GSE10846 dataset, and box plots were generated through ggplot2 package (v 3.3.6).¹⁰ For further analysis, the association between *SPAG5* expression and various clinicopathological parameters was examined in DLBCL patients from TCGA cohort.

Diagnostic Value Analysis of SPAG5

The receiver operating characteristic (ROC) curve was constructed through pROC package (v 1.18.0)¹¹ to evaluate the sensitivity and specificity of *SPAG5* expression levels in distinguishing between DLBCL patients and normal controls

within TCGA cohort. The area under curve (AUC) was calculated as a quantitative measure of this diagnostic performance. An AUC value greater than 0.7 was considered indicative of favorable diagnostic accuracy, suggesting that *SPAG5* could serve as a potential diagnostic biomarker for DLBCL.

Functional Characterization of SPAG5

The STRING database (<https://www.string-db.org/>) was utilized to extract data related to the functional network of *SPAG5*, applying a confidence score threshold greater than 0.4 to ensure reliable interactions. The protein-protein interaction (PPI) network was then visualized using Cytoscape software (v 3.8.2).¹² Following that, the correlation between *SPAG5* expression levels and other mRNA expression levels was determined within TCGA cohort. We then selected the top 200 genes exhibiting the most significant positive correlation with *SPAG5* for further Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis making use of clusterProfiler package (v 4.7.1.001),¹³ with thresholds set at adj. $P < 0.05$ and $q < 0.05$. Additionally, to explore the signaling pathways associated with *SPAG5* that may influence DLBCL development, gene set enrichment analysis (GSEA) was performed on TCGA cohort. Based on the median *SPAG5* expression level, samples were divided into high and low expression groups for GSEA, considering $P < 0.05$, FDR $q < 0.9$, and $|\text{NES}| \geq 1$ as significantly enriched criteria. The enrichplot package (v 1.18.3)¹⁴ was used to visualize the aforementioned results.

Survival and Prognosis Analysis for SPAG5

To evaluate the prognostic value of *SPAG5* expression, survival analysis were conducted using OS and PFS data obtained from TCGA database, applying the survival (v 3.7-0)¹⁵ and survminer (v 0.4.9)¹⁶ packages. Patients were stratified into high- and low-*SPAG5* expression groups using the optimal cutoff value determined by the “surv_cutpoint” function in the survminer package (with minprop = 0.3 to define group proportions).

For PFS analysis: samples with a PFS time of less than 30 days were excluded (three samples removed), resulting in 44 eligible patients; median PFS was calculated using the survival package: the median PFS was 17.49 years in the low-*SPAG5* expression group and 8.92 years in the high-*SPAG5* expression group. Additionally, the reverse Kaplan-Meier (K-M) method (implemented via the survival package in R) was used to estimate median follow-up time: the median follow-up time was 5.84 years in the low-expression group and 2.59 years in the high-expression group. For OS analysis: samples with an OS time of less than 30 days were excluded (three samples removed), resulting in 45 eligible patients; the median OS was 17.6 years in the low-*SPAG5* expression group, while the median OS was not reached in the high-*SPAG5* expression group. The median follow-up time was 5.84 years in the low-expression group and 2.59 years in the high-expression group. Given the relatively small sample size in TCGA, we further incorporated a large-cohort dataset (GSE181063) for additional OS-based survival analysis. This dataset included 1303 samples. For this cohort, the median OS was 11.20 years in the low-*SPAG5* expression group and 5.49 years in the high-*SPAG5* expression group. The reverse K-M method revealed a median follow-up time of 7.93 years in the low-expression group and 8.74 years in the high-expression group. K-M survival curves were generated to visualize survival differences between the two groups. A Log rank test was used to assess between-group differences ($P < 0.05$).

For comprehensive genomic analysis and visualization of *SPAG5* mutations, cBioPortal (<https://www.cbioportal.org>) was utilized. The relationship between *SPAG5* mutations and OS was specifically assessed in TCGA-DBLC, with a reference z-score threshold of ± 1.8 applied to identify significant alterations.

Meantime, to evaluate the independent impact of clinicopathological characteristics and *SPAG5* expression on patients' OS, we performed univariate and multivariate Cox proportional hazards regression analyses. Specifically, to address potential confounding factors, we first identified OS-associated variables via univariate Cox regression (with a screening criterion of hazard ratios (HR) $\neq 1$, $P < 0.1$ to enhance the practicality of the model and avoid missing potentially meaningful variables). Subsequently, along with *SPAG5* expression levels, these variables were incorporated into a multivariate Cox regression model to adjust for confounding effects (HR $\neq 1$, $P < 0.2$). For the TCGA dataset, four clinical features (Age, Stage, Gender, Weight) were sequentially included in the analysis, and six cases lacked Stage information and were excluded, leaving 41 samples for analysis. For the GSE181063 dataset, three clinical features (Age, Stage, Gender) were incorporated simultaneously. Samples without Stage information were excluded, resulting in 1083

eligible samples for analysis. Results were reported as HR with their 95% confidence intervals (CI). By integrating these analytical approaches, we aimed to comprehensively determine the prognostic significance of *SPAG5* expression in DLBCL patients.

Immune Cell Infiltration Analysis

The single-sample GSEA (ssGSEA) algorithm among GSVA package (v 1.42.0)¹⁷ was employed to analyze immune infiltration in DLBCL samples of TCGA cohort. This approach allowed us to quantify the relative abundance of various immune cell types within tumor samples. For the purpose of investigating the relationship between *SPAG5* expression levels and immune cell infiltration, we performed Spearman correlation analysis was operated. Subsequently, we utilized CIBERSORT algorithm (v 1.03)¹⁸ to evaluate the proportion of 22 immune cell types¹⁹ in both high- and low-*SPAG5* expression groups. The Pearson correlation coefficients was also executed by corrplot package (v 0.92).¹⁴ This enabled us to visualize and analyze correlations between different immune cell populations.

Immune Checkpoint Analysis and Immunotherapy Prediction

The correlation between *SPAG5* expression and eight key immune checkpoint genes (ICGs)—*SIGLEC15*, *TIGIT*, *CD274*, *HAVCR2*, *PDCD1*, *CTLA4*, *LAG3*, and *PDCD1LG2*—was investigated adopting Spearman correlation analysis. According to the TCGA cohort, tumor immune dysfunction and exclusion (TIDE) scores were subsequently calculated for DLBCL patients to predict potential responses to immune checkpoint blockade (ICB) therapy. At last, The protein expression profile of *SPAG5* was obtained through HPA database (<https://www.proteinatlas.org>).

Statistical Analysis

All analysis was carried out in R software (v 4.2.1). For comparisons of continuous variables—including gene expression levels, immune cell infiltration ratios, and other characteristic values—intergroup differences were evaluated using the Wilcoxon rank-sum test, with a significance threshold set at $P < 0.05$. K-M survival curves were constructed to visualize survival differences between groups, and intergroup survival differences were assessed via the Log rank test ($P < 0.05$ for significance). The median follow-up time was estimated using the reverse K-M method (implemented via the survival R package (v 3.7–0)). To identify factors associated with OS and determine independent prognostic factors, univariate and multivariate Cox proportional hazards regression models were employed. The proportional hazards (PH) assumption of the Cox model was validated using the Schoenfeld residuals test; a P -value > 0.05 indicated that the PH assumption was satisfied.

Result

Broad Spectrum Expression of SPAG5 Across Multiple Tumor Types

At the outset, *SPAG5* expression across multiple tumor types as analyzed using TIMER 2.0 (Figure 1A). Concurrently, data from Sangerbox3.0 database corroborated these findings, showing significant up-regulation of *SPAG5* in several cancers, including breast cancer (BRCA), bladder cancer (BLCA), uterine corpus endometrial carcinoma (UCEC), lung squamous cell carcinoma (LUSC), and cervical squamous cell carcinoma (CESC) ($P < 0.05$) (Figure 1B). Conversely, *SPAG5* was significantly down-regulated in kidney chromophobe (KICH) and kidney renal papillary cell carcinoma (KIRP) ($P < 0.05$). These results indicated that *SPAG5* exhibited a broad spectrum of expression differences across various tumors, highlighting its potential role in cancer biology.

Analysis of SPAG5 Expression in DLBCL and Its Diagnostic Value

Using the TCGA cohort combined with the GTEx cohort, the expression levels of *SPAG5* in tumor and normal samples were analyzed, clearly revealing significant upregulation of *SPAG5* in DLBCL ($P < 0.001$). To avoid heterogeneity, validation was performed using GSE56315 dataset, which confirmed the upregulated expression trend of *SPAG5* in DLBCL ($P > 0.05$) (Figure 2A and B). Further analysis involved evaluating the diagnostic value of *SPAG5* in DLBCL using a ROC curve, which yielded an AUC of 0.75, indicating a moderate diagnostic potential

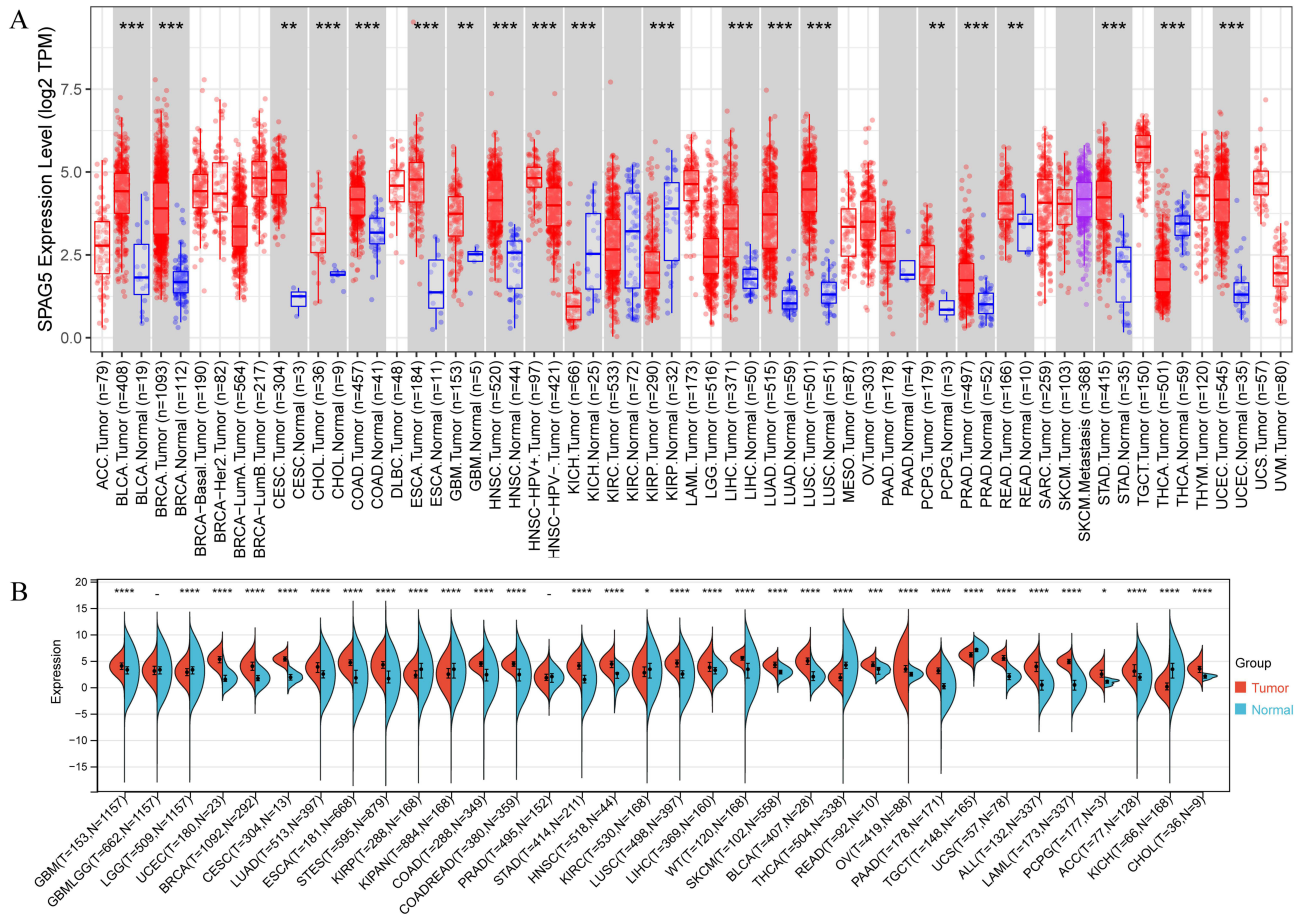


Figure 1 Expression level of SPAG5 in pan-cancer. **(A)** Box plots showing the expression levels of SPAG5 in various tumor tissues (red) and corresponding normal tissues (blue) across multiple cancer types. **(B)** Violin plots illustrating the expression levels of SPAG5 in tumor tissues (red) and normal tissues (blue) for different cancer types. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

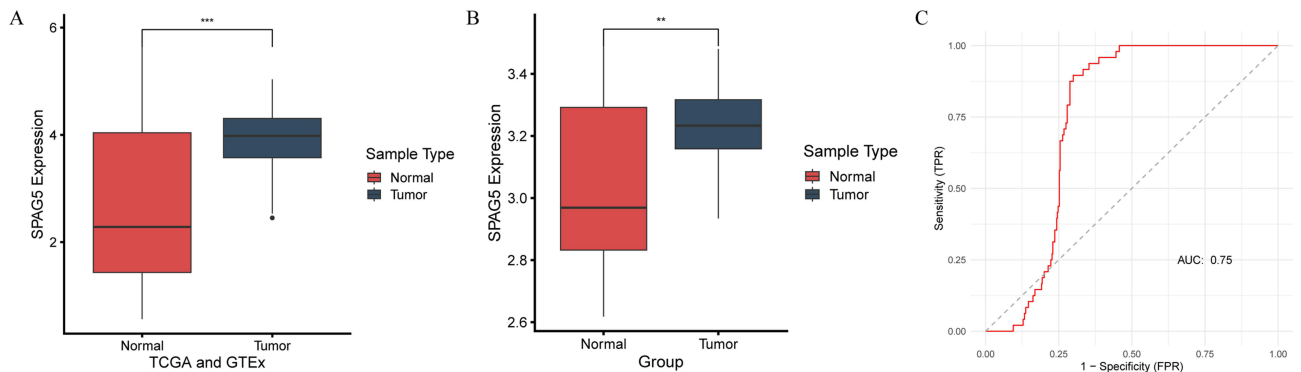


Figure 2 Expression level of SPAG5 in DLBCL. **(A)** Expression of SPAG5 in TCGA (n = 47, Tumor) and GTEx (n = 444, Normal) databases; **(B)** Expression of SPAG5 in GSE56315 (n = 88); **(C)** ROC curve of SPAG5. ** $P < 0.01$, *** $P < 0.001$.

(Figure 2C). We also examined the association between *SPAG5* expression levels and various clinicopathological parameters of DLBCL patients within the TCGA database. The analysis revealed no significant correlations between *SPAG5* expression and any of these clinicopathological factors ($P > 0.05$) (Supplementary Figure 2).

Revealing Pathways Impacted by SPAG5 in the Progression of DLBCL

After establishing the significance of *SPAG5* expression in DLBCL, we utilized STRING database to identify genes that interact with *SPAG5*. In total, 10 genes were found: *KIF11*, *CDK1*, *ASPM*, *CDCA8*, *TOP2A*, *KNSTRN*, *BUB1B*, *DLGAP5*, *BUB1*, and *PLK1* which exhibited strong interactions with *SPAG5* (Figure 3A). Subsequently, the top 200 genes were selected that showed the most significant positive correlation with *SPAG5* and performed GO and KEGG analyses on them. The GO analysis revealed significant enrichment in pathways related to chromosome segregation including mitotic nuclear division, DNA replication, mitotic sister chromatid segregation, and regulation of cell cycle phase transition (Figure 3B). Similarly, KEGG pathway analysis highlighted enrichment in several critical cellular processes such as cell cycle control mechanisms and DNA damage response pathways including the Fanconi anemia pathway and mismatch repair system (Figure 3C). Furthermore, we stratified DLBCL samples from the TCGA cohort into high- and low-expression groups based on median *SPAG5* expression levels to perform GSEA. The visualization displayed the top 6 pathways which included Chagas disease-related processes; cytochrome P450-associated metabolic functions; DNA adducts; gluconeogenesis; complement and coagulation cascades; and metabolism xenobiology mediated by cytochrome (Figure 3D). These findings suggest that *SPAG5* may play a crucial role in regulating various biological processes related to cell division and genome stability.

Studying the Prognostic Value of SPAG5 in DLBCL

The prognostic significance of *SPAG5* expression levels on OS and PFS in DLBCL patients was assessed. The results of TCGA cohort indicated that patients with high *SPAG5* expression had poorer survival outcomes ($P_{PFS} = 0.038$; $P_{OS} = 0.24$) (Figure 4A). Consistently, the OS-based survival analysis using the GSE181063 dataset also demonstrated the same result, *SPAG5* high expression was significantly associated with poorer OS ($P_{OS} < 0.0001$) (Figure 4B). This further validated the reproducibility of our findings regarding the prognostic role of *SPAG5* in DLBCL across independent cohorts.

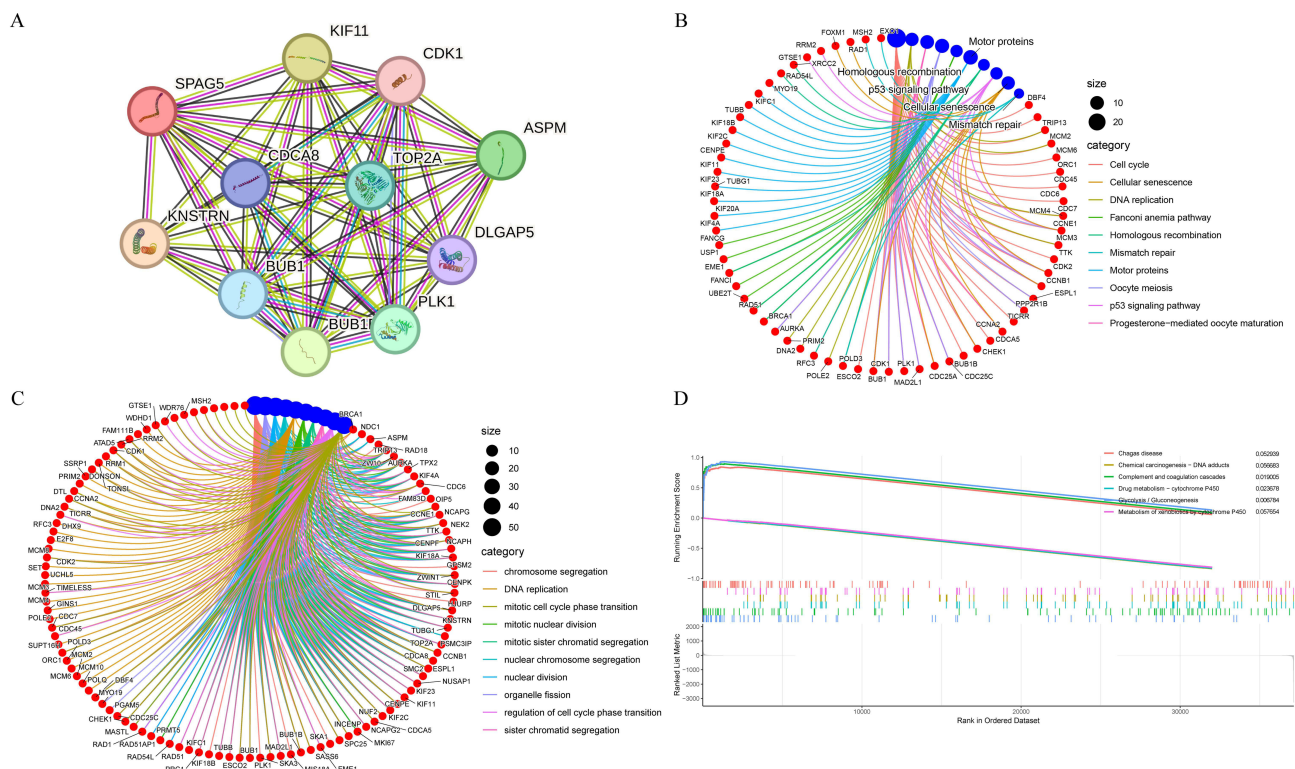


Figure 3 PPI network of *SPAG5* and enrichment analysis. **(A)** PPI network: different colors represent distinct gene names, different colors of lines typically distinguish types of interactions, and a greater number of lines indicates a highly connected network; **(B and C)** GO and KEGG: lines connect genes (red dots) to biological functions/categories (blue dots). **(D)** GSEA: the vertical bars mark the positions where genes from each set appear in the ordered list.

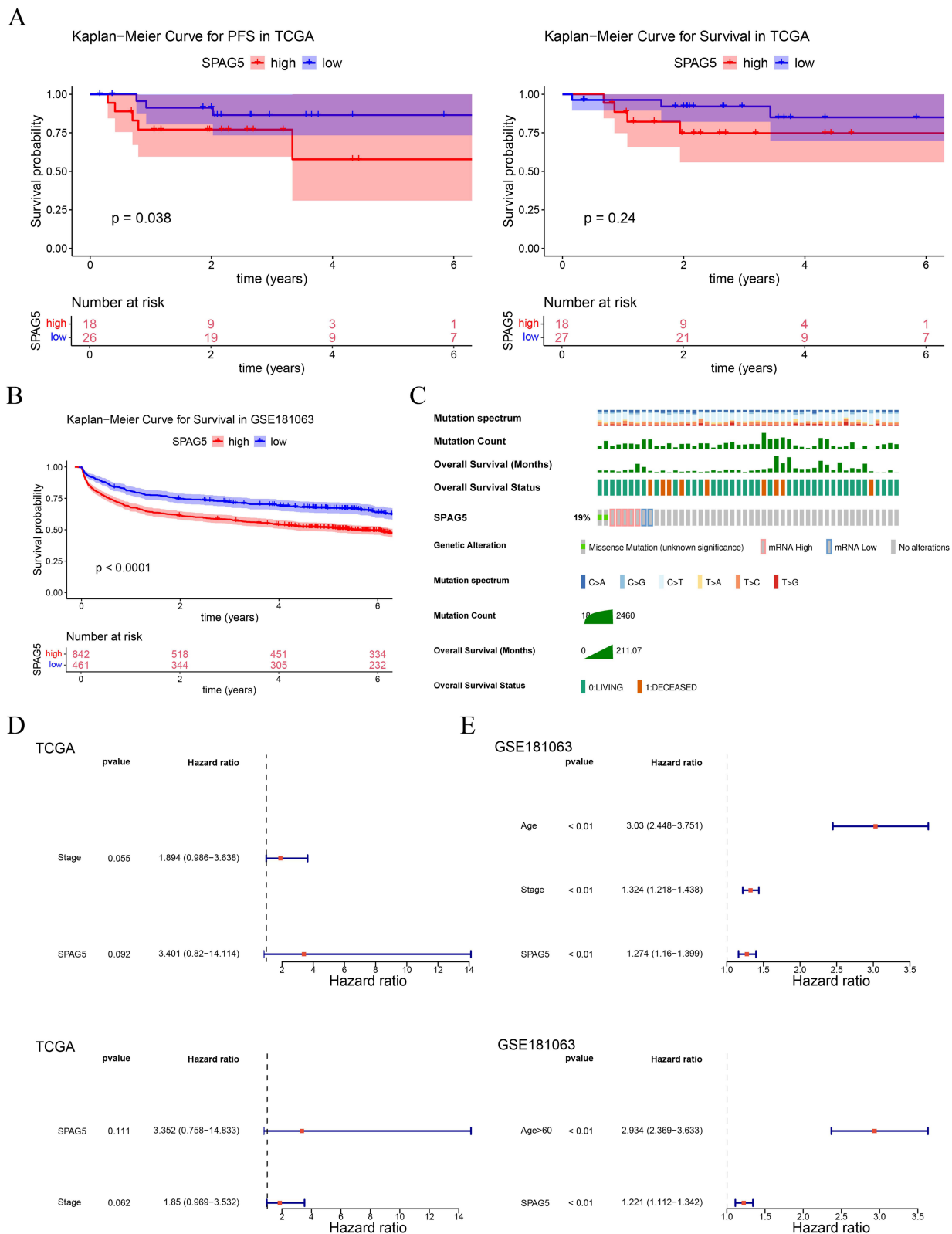


Figure 4 Survival analysis of SPAG5 expression level (**A**) OS (n = 45) and PFS (n = 41) analyses in TCGA; (**B**) OS analysis in GSE181063 (n = 1303). The shaded areas typically represent the confidence interval (CI) for the survival probability at each time point. (**C**) Gene mutation of SPAG5 in TCGA-DLBCL; (**D** and **E**) Univariate and multivariate Cox regression analyses related to OS in TCGA (n = 41) and GSE181063 (n = 1083).

Table 1 The Results of the Proportional Hazards (PH) Assumption Test for Univariate Analysis

Datasets	id	P value
TCGA	Stage	0.310
	SPAG5	0.244
GSE181063	Age	0.852
	Stage	<0.0001
	SPAG5	0.112

To further investigate the impact of *SPAG5*, we analyzed its mutation status and its association with OS using data from TCGA cohort. The gene mutations were identified in 37 patients. From the figure, about 19% of samples exhibit *SPAG5* alterations, including missense mutations and changes in mRNA expression levels (high or low) (Figure 4C).

Moreover, univariate and multivariate Cox regression analysis was employed to evaluate the influence of clinico-pathological features and *SPAG5* expression on mortality risk. The results revealed that only *SPAG5* expression [TCGA: HR = 3.352 (0.758–14.833), $P = 0.111$; GSE181063: HR = 1.221 (1.112–1.342), $P < 0.01$] served as independent prognostic factor for DLBCL (Figure 4D and E). The results of the PH assumption test for univariate analysis are presented in Table 1.

Impact of *SPAG5* Expression on Immune Cell Infiltration

In an effort to investigate the immune landscape associated with *SPAG5* expression in DLBCL, ssGSEA was utilized to assess immune cell infiltration within tumor samples from TCGA cohort. The relationship between *SPAG5* expression levels and specific immune cell populations was analyzed using Spearman correlation coefficient. Results of these analyses were visually represented through bubble charts and scatter plots (Figures 5A and B, Supplementary Figure 3). Notably, *SPAG5* expression exhibited a strong positive correlation with monocytes and M1 macrophages ($\text{cor} > 0$). Subsequently, we employed CIBERSORT to compare immune infiltration profiles between high- and low-expression groups of *SPAG5* (Figure 5C). The analysis revealed that there were significant differences in three types of immune cells between the two groups ($P < 0.05$). Specifically, regulatory T cells (Tregs) and follicular helper T cells (Tfh) exhibited significantly higher infiltration abundance in the *SPAG5* low-expression group ($P < 0.05$), whereas resting natural killer (NK) cells demonstrated notably higher infiltration abundance in the *SPAG5* high-expression group ($P < 0.01$). In addition, we calculated Pearson correlation coefficients to quantify the relationships among different immune cell populations based on their relative abundances across samples. A heatmap was generated to visually depict these correlations, illustrating a strong positive correlation between resting NK cells and activated memory CD4 T cells ($\text{cor} > 0$) (Figure 5D). These findings indicate that *SPAG5* may play a pivotal role in modulating immune responses within the tumor microenvironment (TME) of DLBCL, potentially influencing both innate and adaptive immunity components.

Association of *SPAG5* Expression with ICGs

Immune checkpoints are critical regulators of the immune system, essential for maintaining autoimmune tolerance and modulating the intensity and extent of immune responses in peripheral tissues. In this study, we focused on eight ICGs. Our analysis revealed that *SPAG5* expression positively correlated with *CD274* ($\text{cor} > 0$). Conversely, it showed negative correlations with several other ICGs including *CTLA4*, *LAG3*, *PDCD1*, *SIGLEC15*, *TIGIT*, *HAVCR2*, and *PDCD1LG2* ($\text{cor} < 0$) (Figure 6).

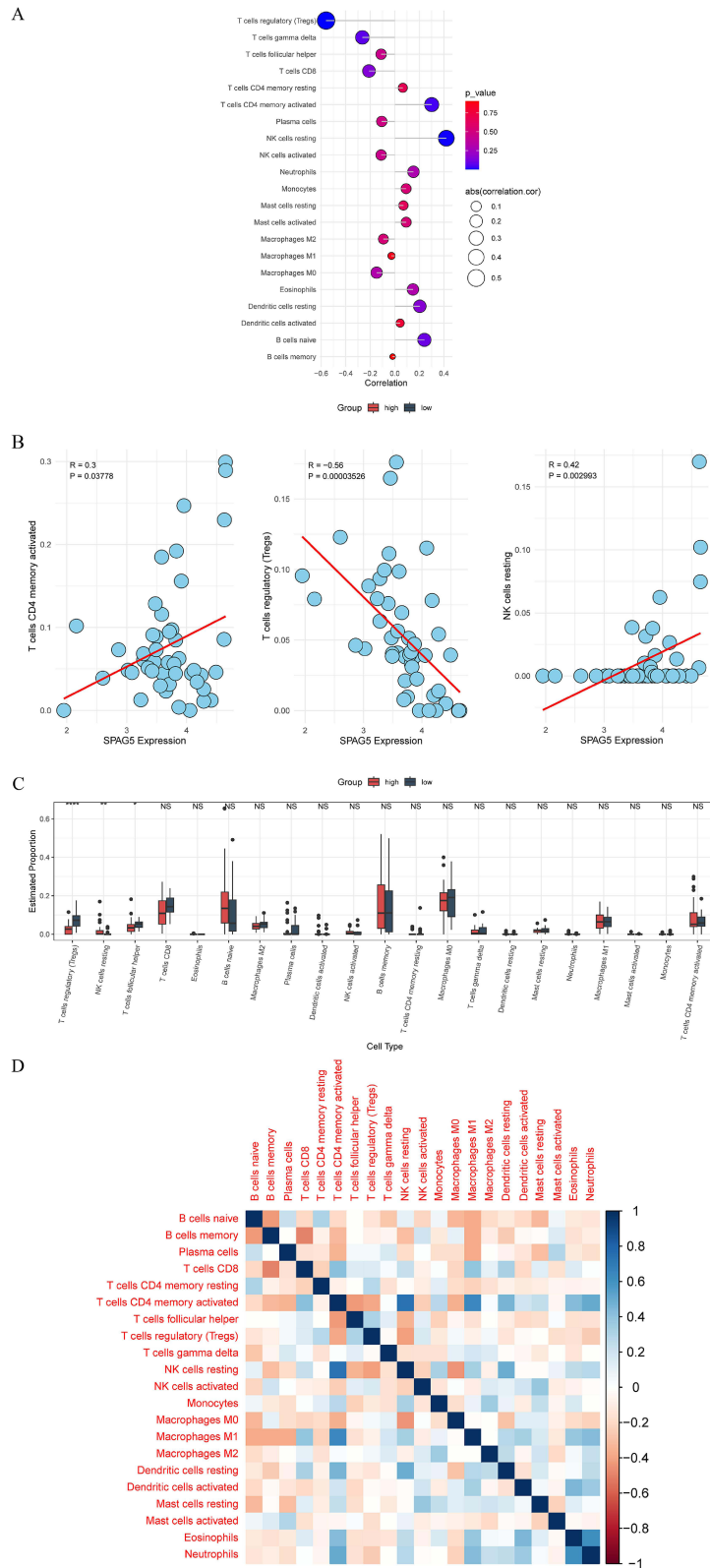


Figure 5 Immune cell infiltration analysis. **(A)** Correlation bubble diagram between SPAG5 and immune cells; **(B)** Scatter diagram of correlation between SPAG5 and immune cells (the red line is the regression line derived using linear regression analysis); **(C)** Difference of high- and low expression groups of SPAG5 in immune cell infiltration; **(D)** Immune cell correlation heatmap. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0001$, NS: $P > 0.05$.

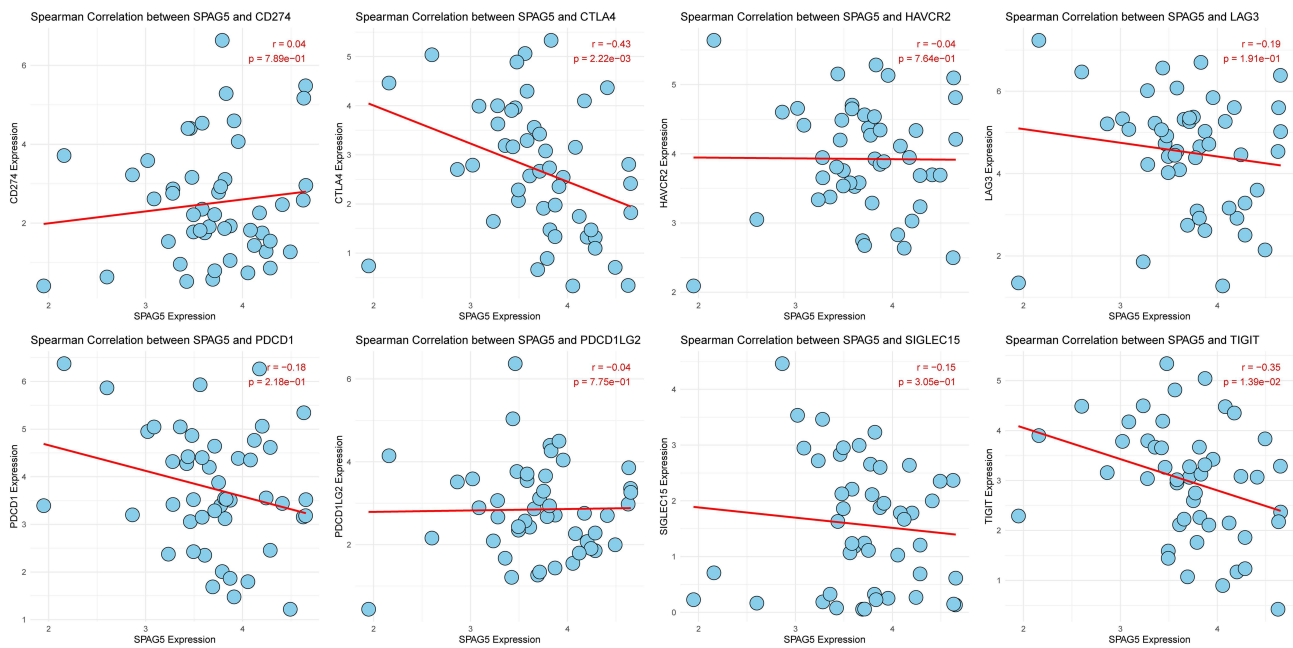


Figure 6 Correlation between SPAG5 expression level and ICGs. Red lines represent the linear regression fit, and each plot includes the Spearman correlation coefficient (r) and statistical significance (P value) to indicate the strength and significance of the correlation.

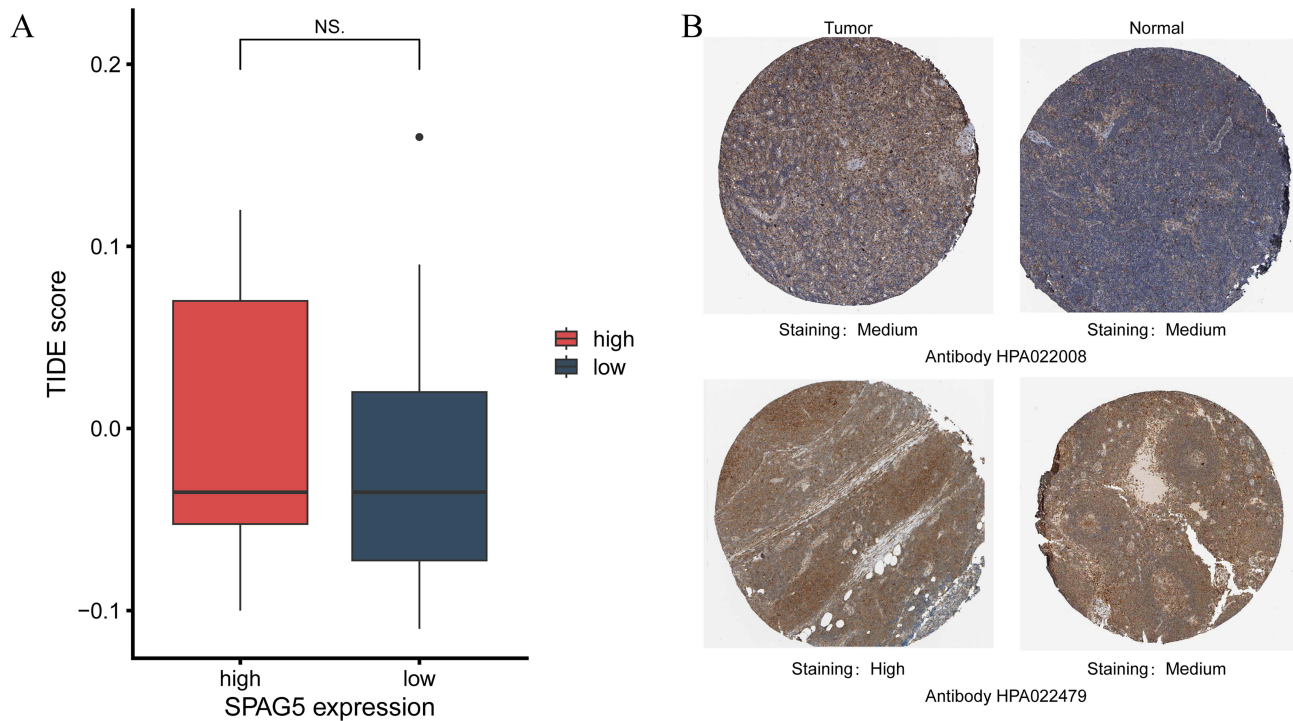


Figure 7 ICB treatment response and protein atlas of SPAG5. **(A)** TIDE scores of the low and high SPAG5 groups of DLBCL patients from the TCGA database, NS: $P > 0.05$; **(B)** SPAG5 protein expression was detected by immunohistochemistry from the HPA database.

ICB Treatment Response and Protein Atlas of SPAG5

TIDE score of DLBCL patients revealed no statistically significant differences in TIDE scores between the high- and low-SPAG5 expression groups ($P > 0.05$) (Figure 7A). However, a trend was observed where a higher proportion of patients in the high-SPAG5 expression group responded to ICB therapy compared to those in the low-expression group.

To further validate *SPAG5* protein expression levels in DLBCL, we consulted HPA. Our investigation confirmed that *SPAG5* was expressed at moderate or high levels in DLBCL tumor tissues. In contrast, minimal staining of *SPAG5* was detected in normal tissues (Figure 7B). These results suggest that while *SPAG5* expression may not significantly impact TIDE scores directly, its higher expression levels are associated with an increased likelihood of response to ICB therapy in DLBCL.

Discussion

DLBCL, recognized as an aggressive form of lymphoma, exhibits significant heterogeneity in cell morphology and molecular genetics.²⁰ This diversity often results in varied patient responses to chemotherapy, with many experiencing adverse effects. Currently, DLBCL is primarily diagnosed through imaging studies, bone marrow examinations, and biopsy histopathology. Furthermore, the molecular heterogeneity of DLBCL is closely associated with genetic subgroups, including MCD, BN2, N1, and EZB, each defined by distinct driver mutations and clinical outcomes.²¹ Without timely improvements in clinical strategies, the incidence of DLBCL is projected to rise.²² *SPAG5*, a gene that encodes a mitotic spindle protein, has been implicated in several critical tumor-related processes including tumor growth, chemotherapy resistance, and local recurrence.²³ Given these associations, along with the presence of chromosomal abnormalities observed in most solid tumors including DLBCL, *SPAG5* could serve as a potential detection marker for this disease and may be implicated in particular genetic subtypes of DLBCL.²⁴ Our bioinformatics work revealed that *SPAG5* is significantly elevated in DLBCL. Its diagnostic performance for DLBCL showed an AUC of 0.75, indicating modest potential as a complementary diagnostic indicator; survival analyses revealed that DLBCL patients with high *SPAG5* expression had poorer prognosis than those with low expression, suggesting a potential association between *SPAG5* expression and patient outcomes, further distinguishing high-risk patients within established groups to guide surveillance frequency; functional enrichment of the top 10 *SPAG5*-interacting genes indicated involvement in DNA replication pathways; and differential immune infiltration (altered abundances of Tregs and resting NK cells) and correlations with CD247 and multiple ICGs were observed between high and low *SPAG5* expression groups, implying *SPAG5* may be linked to TME immunity in DLBCL.

The characteristic changes of tumor cells are chromosomal abnormalities, the molecular mechanism of which is related to the abnormal expression of spindle and centrosome related genes.²⁵ Studies have demonstrated that chromosomal aberrations in patients with DLBCL generally correlate with adverse prognostic outcomes.²⁶ As a core spindle assembly protein, *SPAG5* influences chromosome dynamics through its interaction with the spindle structure, potentially causing mitotic delays and can be modulated by certain kinases to affect spindle positioning and mitotic processes. Consequently, abnormal *SPAG5* expression or its dysfunctional interactions could modulate cancer progression by impacting chromosomal behavior within tumor cells. Consistent with findings in various cancers such as prostate, bladder, breast, bladder urothelial carcinoma, and cervical cancer where *SPAG5* is up-regulated, our study also observed an overexpression of *SPAG5* in DLBCL compared to normal samples.^{27–31} This aligns with reports suggesting genomic instability as a prevalent characteristic of malignancies; factors like DNA replication stress and chromosome missegregation contribute to oncogenesis and are hallmarks of many cancers.^{32,33}

Mechanically, functional enrichment analysis revealed a significant enrichment of these genes in processes related to chromosome separation: DNA replication, mitotic nuclear division, nuclear chromosome separation and organelle fission. Bakhom et al found that the increased rate of chromosome missegregation in DLBCL was associated with poor prognosis of patients.³⁴ *SPAG5*'s aberrant expression may induce chromosome mis-segregation in DLBCL, promoting chromosomal instability and thereby facilitating disease progression. This suggests that pathways regulated by *SPAG5* involved in chromosome separation might be critically implicated in the poor prognosis observed in DLBCL patients. Additionally, abnormal metabolic activities of tumor cells support their survival under stress conditions like avoiding apoptosis and promoting angiogenesis and metastasis.³⁵ Our GSEA also revealed cytochrome P450, DNA adducts, gluconeogenesis and metabolic heterobiological pathways induced by cytochrome, further suggesting that these metabolic pathways may potentially play a crucial role in the pathogenesis of DLBCL.

Immune infiltrate in the TME have been discovered as pivotal contributors to tumorigenesis, aberrant expression and dysfunction of immune cells populations and immune checkpoint molecules have been identified as the pivotal factors

contributing to the pathogenesis of numerous malignancies.^{36,37} In-depth understanding of immune infiltrates and immune checkpoints is crucial to reveal the underlying mechanism of DLBCL. Therefore, we further analyzed immune cell infiltration patterns in DLBCL samples stratified by *SPAG5* expression levels. The distinct immune infiltration profiles observed between the two groups support *SPAG5*'s potential as a prognostic indicator—for instance, we identified notable differences in the infiltration of Tregs and resting NK cells, both of which are well-documented to correlate with DLBCL clinical outcomes.^{38,39} Meanwhile, immune checkpoint analysis also revealed a positive correlation between the expression of *SPAG5* and *CD247*, while demonstrating negative correlations with *CTLA4*, *LAG3*, *PDCD1*, *SIGLEC15*, *TIGIT* and *PDCD1LG2*, which play a role as rheostat in immune response regulation.^{40,41} This expression pattern suggests that aberrant *SPAG5* may modulate immune checkpoint signaling, which could theoretically influence DLBCL immune escape, and further studies are needed to clarify whether and how these gene-level correlations might impact anti-tumor immune responses in vivo.

A major limitation of this study is that its conclusions are primarily derived from bioinformatics analyses and in silico simulations, remaining theoretical in nature and lacking direct experimental validation. To substantiate *SPAG5*'s role as a diagnostic biomarker, further experimental validation is required. Specifically, in vivo and in vitro studies using DLBCL cell lines and tissue samples are necessary to verify these bioinformatic predictions and its role in tumor growth. Second, the sample size of the analyzed cohorts is limited, which may affect the robustness and generalizability of the findings. Third, the absence of the prognostic correlation between *SPAG5* and DLBCL outcomes, immune checkpoint regulation, and direct analysis linking *SPAG5* expression to conventional DLBCL prognostic factors necessitates validation in larger, prospective clinical cohorts with comprehensive clinical annotations. Given these limitations, *SPAG5* cannot be considered a standalone diagnostic marker for DLBCL—its value would likely be as a complementary tool alongside existing clinical and pathological indicators.

Conclusions

The diagnostic performance for DLBCL exhibited an AUC of 0.75, indicating limited potential as a supplementary diagnostic marker; survival analyses demonstrated that DLBCL patients with elevated *SPAG5* expression had a worse prognosis compared to those with low expression, implying a possible correlation between *SPAG5* expression and patient outcomes. Overall, our bioinformatics analyses provide initial insights into the potential associations of *SPAG5* expression with DLBCL prognosis, and TME immunity. It also exhibited genetics alterations in DLBCL samples from TCGA-DLBCL, this encourages subsequent research to examine the strong association of *SPAG5* expression with various genetic subtypes, thereby elucidating the role of *SPAG5* in subtype-specific processes of DLBCL. However, further experimental and clinical studies are essential to validate these preliminary findings before *SPAG5* can be translated into clinical applications for DLBCL.

Data Sharing Statement

All data were obtained from publicly available, de-identified repositories: TCGA (<https://portal.gdc.cancer.gov>) – dbGaP project phs000178; GTEx (<https://gtexportal.org>) – dbGaP project phs000424; GEO dataset GSE56315 and GSE181063 (<https://www.ncbi.nlm.nih.gov/geo/>).

Ethics Approval and Informed Consent

Given that all data utilized in this study are fully de-identified, publicly accessible, and have already undergone ethical review and informed consent procedures by the respective database institutions, this study was exempt from additional review by our institutional review board (IRB). All analyses were conducted in compliance with the principles of the Declaration of Helsinki and relevant national and institutional regulations on human research ethics.

Author Contributions

All authors made a significant contribution to the work reported in the conception, study design, execution, acquisition of data, analysis and interpretation, have drafted or written, or substantially revised or critically reviewed the article, have agreed on the journal to which the article will be submitted; reviewed and agreed on all versions of the article before

submission, during revision, the final version accepted for publication, and any significant changes introduced at the proofing stage and agree to take responsibility and be accountable for the contents of the article.

Funding

No funding was received.

Disclosure

The authors report no conflicts of interest in this work.

References

- Susanibar-Adaniya S, Barta SK. 2021 update on diffuse large B cell lymphoma: a review of current data and potential applications on risk stratification and management. *Am J Hematol*. 2021;96(5):617–629. doi:10.1002/ajh.26151
- Sehn LH, Salles G. Diffuse large B-cell lymphoma. *New Engl J Med*. 2021;384(9):842–858. doi:10.1056/NEJMra2027612
- Poletto S, Novo M, Paruzzo L, Frascione PMM, Vitolo U. Treatment strategies for patients with diffuse large B-cell lymphoma. *Cancer Treat Rev*. 2022;110:102443. doi:10.1016/j.ctrv.2022.102443
- Zhang J, Gu Y, Chen B. Drug-resistance mechanism and new targeted drugs and treatments of relapse and refractory DLBCL. *Cancer Manag Res*. 2023;15:245–255. doi:10.2147/CMAR.S400013
- He J, Green AR, Li Y, Chan SYT, Liu DX. SPAG5: an emerging oncogene. *Trends Cancer*. 2020;6(7):543–547. doi:10.1016/j.trecan.2020.03.006
- Gao X, Bu H, Gao X, Wang Y, Wang L, Zhang Z. Pan-cancer analysis: SPAG5 is an immunological and prognostic biomarker for multiple cancers. *FASEB J*. 2023;37(10):e23159. doi:10.1096/fj.202300626R
- Xiao G, Xu X, Chen Z, Zeng J, Xie J. SPAG5 expression predicts poor prognosis and is associated with adverse immune infiltration in lung adenocarcinomas. *Clin Med Insights Oncol*. 2023;17:11795549231199915. doi:10.1177/11795549231199915
- An J, Yang L, Pan Y, et al. SPAG5 activates PI3K/AKT pathway and promotes the tumor progression and chemo-resistance in gastric cancer. *DNA Cell Biol*. 2022;41(10):893–902. doi:10.1089/dna.2021.0531
- Zeng X, Xu W, Tong J, et al. SPAG5 as a novel biomarker and potential therapeutic target via regulating AKT pathway in multiple myeloma. *Leukemia Lymphoma*. 2022;63(11):2565–2572. doi:10.1080/10428194.2022.2086247
- Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics*. 2022;38(15):3844–3846. doi:10.1093/bioinformatics/btac409
- Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinf*. 2011;12(1):77. doi:10.1186/1471-2105-12-77
- Liu P, Xu H, Shi Y, Deng L, Chen X. Potential molecular mechanisms of plantain in the treatment of gout and hyperuricemia based on network pharmacology. *Evid Based Complement Alternat Med*. 2020;2020(1):3023127. doi:10.1155/2020/3023127
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic*. 2012;16(5):284–287. doi:10.1089/omi.2011.0118
- Wang L, Wang D, Yang L, et al. Cuproptosis related genes associated with Jab1 shapes tumor microenvironment and pharmacological profile in nasopharyngeal carcinoma. *Front Immunol*. 2022;13:989286. doi:10.3389/fimmu.2022.989286
- Lei J, Qu T, Cha L, et al. Clinicopathological characteristics of pheochromocytoma/paranglioma and screening of prognostic markers. *J Surg Oncol*. 2023;128(4):510–518. doi:10.1002/jso.27358
- Ramsay IS, Ma S, Fisher M, et al. Model selection and prediction of outcomes in recent onset schizophrenia patients who undergo cognitive training. *Schizophr Res Cogn*. 2018;11:1–5. doi:10.1016/j.scog.2017.10.001
- Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. 2013;14(1):7. doi:10.1186/1471-2105-14-7
- Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243–259.
- Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. doi:10.1038/nmeth.3337
- He MY, Kridel R. Treatment resistance in diffuse large B-cell lymphoma. *Leukemia*. 2021;35(8):2151–2165. doi:10.1038/s41375-021-01285-3
- Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378(15):1396–1407. doi:10.1056/NEJMoa1801445
- Tilly H, Gomes da Silva M, Vitolo U, et al. Diffuse large B-cell lymphoma (DLBCL): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2015;26 Suppl 5:v116–25. doi:10.1093/annonc/mdv304
- Li Q, Wang Y, He J. MiR-133a-3p attenuates resistance of non-small cell lung cancer cells to gefitinib by targeting SPAG5. *J Clin Lab Analysis*. 2021;35(7):e23853. doi:10.1002/jcla.23853
- Alonso S, Dow LE. Engineering chromosome rearrangements in cancer. *Dis Models Mech*. 2021;14(9). doi:10.1242/dmm.049078
- Morgan C, Nayak A, Hosoya N, Smith GR, Lambing C. Meiotic chromosome organization and its role in recombination and cancer. *Curr Topics Develop Biol*. 2023;151:91–126.
- Nakamura Y, Itoh Y, Kakegawa E, Uchida Y, Ichimura T, Sasaki A. Biallelic *BCL6* rearrangements by dual t(3;14)(q27;q32) and t(3;22)(q27;q11) translocations in diffuse large B-cell lymphoma. *J Clin Experimental Hematopathol*. 2022;62(4):268–272. doi:10.3960/jslrt.22031
- Zhang H, Li S, Yang X, Qiao B, Zhang Z, Xu Y. miR-539 inhibits prostate cancer progression by directly targeting SPAG5. *J Experimental Clin Cancer Res*. 2016;35(1):60. doi:10.1186/s13046-016-0337-8
- Canu V, Donzelli S, Sacconi A, et al. Aberrant transcriptional and post-transcriptional regulation of SPAG5, a YAP-TAZ-TEAD downstream effector, fuels breast cancer cell proliferation. *Cell Death Differ*. 2021;28(5):1493–1511. doi:10.1038/s41418-020-00677-9

29. Liu JY, Zeng QH, Cao PG, et al. SPAG5 promotes proliferation and suppresses apoptosis in bladder urothelial carcinoma by upregulating Wnt3 via activating the AKT/mTOR pathway and predicts poorer survival. *Oncogene*. 2018;37(29):3937–3952. doi:10.1038/s41388-018-0223-2
30. Li Y, Diao Y, Wang Z, Wang S, Peng J, Kong B. The splicing factor SF3B4 drives proliferation and invasion in cervical cancer by regulating SPAG5. *Cell Death Discovery*. 2022;8(1):326. doi:10.1038/s41420-022-01120-3
31. Liu J, Zhang Y, Zeng H, et al. Fe-doped chrysotile nanotubes containing siRNAs to silence SPAG5 to treat bladder cancer. *J Nanobiotechnol*. 2021;19(1):189. doi:10.1186/s12951-021-00935-z
32. Boukaba A, Liu J, Ward C, et al. Ectopic expression of meiotic cohesin generates chromosome instability in cancer cell line. *Proc Natl Acad Sci USA*. 2022;119(40):e2204071119. doi:10.1073/pnas.2204071119
33. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet*. 2020;21(1):44–62. doi:10.1038/s41576-019-0171-x
34. Bakhomou SF, Danilova OV, Kaur P, Levy NB, Compton DA. Chromosomal instability substantiates poor prognosis in patients with diffuse large B-cell lymphoma. *Clin Cancer Res*. 2011;17(24):7704–7711. doi:10.1158/1078-0432.CCR-11-2049
35. Park JH, Pyun WY, Park HW. Cancer metabolism: phenotype, signaling and therapeutic targets. *Cells*. 2020;9(10):2308. doi:10.3390/cells9102308
36. Wu W, Wang X, Le W, et al. Immune microenvironment infiltration landscape and immune-related subtypes in prostate cancer. *Front Immunol*. 2022;13:1001297. doi:10.3389/fimmu.2022.1001297
37. Bader JE, Voss K, Rathmell JC. Targeting metabolism to improve the tumor microenvironment for cancer immunotherapy. *Molecular Cell*. 2020;78(6):1019–1033. doi:10.1016/j.molcel.2020.05.034
38. Peng F, Qin Y, Mu S, Li J, Ai L, Hu Y. Prognostic role of regulatory T cells in lymphoma: a systematic review and meta-analysis. *J Cancer Res Clin Oncol*. 2020;146(12):3123–3135. doi:10.1007/s00432-020-03398-1
39. Yu M, Zhang Q, Wan L, et al. IL-1R8 expression in DLBCL regulates NK cell recruitment and influences patient prognosis. *Funct Integr Genomics*. 2023;23(4):328. doi:10.1007/s10142-023-01254-2
40. Miyawaki K, Sugio T. Lymphoma Microenvironment in DLBCL and PTCL-NOS: the key to uncovering heterogeneity and the potential for stratification. *J Clin Experimental Hematopathol*. 2022;62(3):127–135. doi:10.3960/jslrt.22027
41. Tang T, Wang J, Zhang L, et al. IQGAP2 acts as an independent prognostic factor and is related to immunosuppression in DLBCL. *BMC Cancer*. 2021;21(1):603. doi:10.1186/s12885-021-08086-y

Journal of Blood Medicine

Publish your work in this journal

The Journal of Blood Medicine is an international, peer-reviewed, open access, online journal publishing laboratory, experimental and clinical aspects of all aspect pertaining to blood based medicine including but not limited to: Transfusion Medicine; Blood collection, Donor issues, Transmittable diseases, and Blood banking logistics; Immunohematology; Artificial and alternative blood based therapeutics; Hematology; Biotechnology/nanotechnology of blood related medicine; Legal aspects of blood medicine; Historical perspectives. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/journal-of-blood-medicine-journal>

Dovepress
Taylor & Francis Group