



OPEN Exploring the prognostic value of T follicular helper cell levels in chronic lymphocytic leukemia

Rui Zhang, Sha Guo & Jianhua Qu✉

Chronic lymphocytic leukemia (CLL) presents with heterogeneous clinical outcomes, suggesting varied underlying pathogenic mechanisms. This study aims to elucidate the impact of T follicular helper (Tfh) cells on CLL progression and prognosis. Gene expression profile data for CLL were collected from GSE22762 and GSE39671 datasets. Patients were divided into high and low groups using Tfh levels using the optimal cutoff value based on overall survival (OS) and time-to-first treatment (TTFT). Differential expression analysis was performed between these groups, followed by co-expression network analysis and single-sample Gene Set Enrichment Analysis (ssGSEA). Marker genes of Tfh cells were used to construct prognostic models. Additionally, 40 CLL patients were recruited and categorized based on median Tfh levels. Marker gene expression was assessed using RT-qPCR and Western Blot, and immune cell levels were determined through flow cytometry. The high group showed better prognosis compared to the low group. Among the 1121 differentially expressed genes identified, five co-expression networks were constructed, with the turquoise module showing the highest correlation with Tfh cells. Genes within this module significantly participate in cytokine-cytokine receptor interaction, PI3K-Akt signaling pathway, and natural killer cell mediated cytotoxicity. Tfh cells were significantly negatively correlated with activated B cells and positively correlated with Tregs. The Random Survival Forest (RSF) model identified 10 marker genes, and further analysis using Lasso regression and nomogram selected CLEC4A, RAE1, CD84, and PRDX1 as prognostic markers. In the high group, levels of CLEC4A and RAE1 were higher than in the low group, whereas CD84 and PRDX1 were lower. Flow cytometry revealed that the level of activated B cells in the high Tfh group was significantly lower than in the low Tfh group, while the level of Tregs is significantly higher in the high Tfh group. This study seeks to contribute to a more detailed understanding of the pathogenesis of CLL, delving into the prognostic significance of Tfh.

Keywords Chronic lymphocytic leukemia, Tfh, Time-to-first treatment, Prognostic genes

Chronic lymphocytic leukemia (CLL) is the most common B-cell malignancy in adults, characterized by the progressive accumulation of mature B-lymphocytes¹. The clinical course of CLL is highly variable, with some patients experiencing an indolent disease course while others face aggressive disease requiring immediate treatment². The interval between diagnosis and the development of symptomatic diseases that require treatment may vary greatly among patients³. Identifying CLL patients with unfavorable Time To First Treatment (TTFT) or overall survival in the early stages of the disease remains a significant challenge.

Recent studies have highlighted the complex tumor microenvironment in CLL, which involves interactions between malignant B cells and various components of the host's immune system, including T cells, macrophages, leading to hypogammaglobinemia, and pseudo-exhausted immune function⁴. Among these, T follicular helper (Tfh) cells, a subset of CD4+ T cells, have emerged as key players in B cell help and antibody production⁵. Tfh cells support B cell maturation and differentiation within germinal centers and have been implicated in autoimmunity and infection⁶. Tfh is considered the most common subgroup of T cells in lymphoid organs, Tfh expand abnormally and produce a mass of cytokines, potentially supporting complex immune effects^{7,8}. Studies have suggested that Tfh cells might play a dual role in CLL: they could potentially support malignant B cell survival and proliferation⁹, but they might also contribute to anti-tumor immune responses¹⁰. Despite the recognized importance of the immune microenvironment in CLL, the specific contributions of Tfh cells to CLL pathogenesis and their potential as prognostic markers or therapeutic targets remain inadequately understood.

Hematology Center, The First Affiliated Hospital of Xinjiang Medical University, No. 137 Liyushan South Road, Urumqi 830054, Xinjiang, China. ✉email: jhuaqu@163.com

Recently, machine learning model was developed to identify patients at risk of severe infection within 2 years after CLL diagnosis⁴. Although required large, high-quality datasets for training, machine learning model can handle large, complex datasets and identify patterns that traditional statistical methods might miss, this enables more accurate predictions of clinical outcomes.

This study seeks to elucidate the role of Tfh cells in CLL by integrating gene expression profiles from publicly available datasets and novel single-cell sequencing data. By integrating bulk and single-cell RNA sequencing data with flow cytometry analysis of immune cell populations in CLL patients, we aimed to elucidate the impact of Tfh cells on the CLL immune microenvironment, identify Tfh cell-associated gene markers that correlate with patient prognosis, and highlight potential therapeutic targets.

Materials and methods

Data collection and preprocessing

GSE22762¹¹, GSE39671¹², and GSE165087¹³ datasets were collected from Gene Expression Omnibus (GEO) database, bulk RNA-seq datasets (GSE22762 and GSE39671) were selected for comprehensive gene expression profiles and prognostic information in CLL. GSE22762 dataset including gene expression profile of peripheral blood mononuclear cells from 151 patients with CLL. GSE39671 dataset including gene expression profile of peripheral blood samples from 8 patients with CLL. Raw data were preprocessed using the DESeq2 R package, included normalization to account for differences in sequencing depth and RNA composition between samples. GSE165087 dataset including single cell sequence data of peripheral blood samples from 9 patients with CLL.

Sample grouping and survival analysis

Kaplan-Meier (K-M) survival curve analysis is utilized to assess each patient's Overall Survival (OS) and Time-To-First Treatment (TTFT). Single-sample Gene Set Enrichment Analysis (ssGSEA) is applied to evaluate the levels of T follicular helper (Tfh) cells, conducted using the GSVA R package. The maxstat R package is employed to determine the optimal cutoff value for Tfh levels with OS and TTFT as the primary endpoints, subsequently dividing patients into high Tfh and low Tfh groups. The prognostic differences between the high Tfh and low Tfh groups within the GSE22671 dataset for OS and TTFT, as well as TTFT differences within the GSE39671 dataset, are compared using K-M curves.

Differential expression and co-expression network analysis

The limma R package was used to conduct differential analysis between the high Tfh and low Tfh groups for OS and TTFT, identifying differentially expressed genes (DEGs) with a selection threshold of $P < 0.05$. Volcano plots were drawn using the ggplot2 R package. Intersection analysis was performed on the DEGs from the high Tfh and low Tfh groups for OS and TTFT in the GSE22671 dataset, as well as the DEGs from the high Tfh and low Tfh groups for TTFT in the GSE39671 dataset, to identify common DEGs.

The Weighted Gene Co-expression Network Analysis (WGCNA) was then utilized to construct a co-expression network of common DEGs. Briefly, the optimal soft threshold (β) was selected based on network topology invariance using the pickSoftThreshold function of the WGCNA R package, followed by the calculation of correlation between all gene pairs. The topological overlap measure (TOM) between every pair of genes in the network was calculated, and genes were clustered using average linkage hierarchical clustering based on TOM. Gene modules were identified by cutting the dendrogram using the dynamicTreeCut function. Pearson correlation was calculated to assess the relationship between module eigengenes (MEs) and clinical traits.

Enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for module genes were performed using ClusterProfiler R package¹⁴. $P < 0.05$ was considered statistically significant.

Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) were used to evaluate the differences in activation of signaling pathways between the high Tfh and low Tfh groups. GSEA was performed using ClusterProfiler R package. The enrichment score (ES) was calculated, which is normalized based on the size of the gene set to obtain the normalized enrichment score (NES). GSVA was performed using GSVA R package to estimate the activity of pathways in high Tfh group.

Estimation of immune cell infiltration

ssGSEA was used to estimate the infiltration abundance of 28 immune cell types in high Tfh group and low Tfh group. Correlations between genes and immune cells were calculated by Pearson correlation analysis. $P < 0.05$ was considered statistically significant.

Building machine learning models

A total of 12 machine learning algorithms and 113 algorithm combinations were integrated¹⁵, and based on the marker genes of Tfh cells¹⁶, we have constructed models that can predict the TTFT. All algorithm combinations were performed on all samples for training data and 70 samples for validation data. The area under the receiver operating characteristic (ROC) curve (AUC) was calculated for each model, the model with the highest average AUC was considered optimal.

Genes in optimal model were selected to construct risk model in GSE22762 and GSE39671 datasets. Then nomogram prognostic model was built based on the above variables. Calibration curves were used to evaluate the calibration of the nomogram.

Screening and evaluation of prognostic markers

Lasso regression was established to predict prognostic genes based on optimal model using glmnet R package. A 10-fold cross validation was performed to determine the optimal value of Lasso penalty parameter. Features were selected to construct nomogram prognostic model and risk model. Additionally, random survival forest (RSF) was established to predict survival probabilities for prognostic markers.

Single cell RNA sequencing (scRNA-seq) analysis

The top 2000 highest variance genes were selected after excluding low-quality cells with mitochondrial genes $\leq 20\%$ and gene count ≥ 8000 . NormalizeData in Seurat R was used to normalize data to make gene expression levels comparable across cells. Clustering algorithms of k-means categorize cells into clusters. The t-distributed stochastic neighbor embedding (t-SNE) method was used for visualization. Cell type annotation was performed using a combination of marker gene expression and reference databases. The FindMarkers function in Seurat was used to identify differentially expressed genes (DEGs) for each cluster. Key marker genes were matched to known cell type markers from the publicly available databases, such as CellMarker database (<http://117.50.127.228/CellMarker/>), which we utilized in our analysis.

Sample collection

Peripheral blood mononuclear cells (PBMCs) were collected from 40 CLL patients at the First Affiliated Hospital of Xinjiang Medical University. The diagnosis of CLL patients follows the latest international standards¹⁷. This study was approved by the Ethics Committee of the First Affiliated Hospital of Xinjiang Medical University (No. 20211015-47), all methods were performed in accordance with the Declaration of Helsinki. All participants were aware of the content of this study and gave written informed consent.

Patients were divided into high and low Tfh cell level groups based on the median value of Tfh cell levels determined through flow cytometry analysis.

RT-qPCR

Total RNA was extracted from PBMCs using Trizol reagent (Invitrogen, CA, USA). RNA concentration and purity were assessed using a NanoDrop spectrophotometer. cDNA was synthesized from total RNA using Primescript RT master mixture (Invitrogen). A 20 μL reaction mixture containing 1 μg of total RNA, random hexamers, and oligo(dT) primers was incubated at 42 °C for 30 min, followed by 85 °C for 5 min to inactivate the reverse transcriptase. The RT-qPCR reaction was carried out using SYBR Green PCR master mix (Invitrogen) on Applied Biosystems 7500. Each 20 μL PCR reaction contained 10 μL of SYBR Green mix, 1 μL of cDNA, 0.4 μL of each primer (10 μM), and 8.2 μL of nuclease-free water. The thermal cycling conditions were as follows: initial denaturation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Primer sequences are provided in Table S1. Gene expression levels were normalized to GAPDH using $2^{-\Delta\Delta\text{Ct}}$ method.

Western blot

Total protein was extracted from PBMCs using RIPA buffer (Thermo Fisher Scientific, CA, USA) supplemented with protease and phosphatase inhibitors. Protein concentration was determined using the BCA protein assay kit (Beyotime, Shanghai, China). Equal amounts of proteins (30 μg) were separated on SDS-PAGE gels and transferred to PVDF membranes. Membranes were blocked with 5% non-fat dry milk in TBST (Tris-buffered saline with 0.1% Tween-20) for 1 h at room temperature. Membranes were incubated with primary antibodies specific to the marker genes and a housekeeping protein (GAPDH) as a loading control overnight at 4 °C. Membranes were washed with TBST and then incubated with HRP-conjugated secondary antibodies for 1 h at room temperature. Protein bands were visualized using enhanced chemiluminescence and imaged with a ChemiDoc MP Imaging System (Bio-Rad). Band intensities were quantified using ImageJ software and normalized to GAPDH.

Flow cytometry

Flow cytometry was performed to detect the levels of activated B Cells, Tregs, and Tfh cells. PBMCs were resuspended in staining buffer (PBS with 2% fetal bovine serum) at a concentration of 1×10^6 cells/mL. Cells were then stained with the following fluorochrome-conjugated antibodies: CD19 FITC and CD23 PE for activated B cells, CD4 PC7, CD127 PE, and CD25 FITC for Tregs, and CD4 FITC and CXCR5 PE for Tfh cells. For each staining, 100 μL of the cell suspension was incubated with the antibodies at the recommended dilution for 30 min at 4 °C in the dark. Following incubation, cells were washed twice with staining buffer and resuspended in 500 μL of staining buffer for acquisition. Flow cytometry data were acquired using a BD FACSCanto II flow cytometer (BD Biosciences). Data analysis was performed using FlowJo software (Tree Star). Gating strategies were applied to identify and quantify the populations of activated B cells, Tregs, and Tfh cells based on their respective surface markers.

Statistical analysis

All statistical analyses were performed using the R 3.6.1 or GraphPad Prism 9.0 software. Data were shown as the mean \pm SD and compared using Student's *t* test. P value < 0.05 was considered statistically significant.

Results

Patient stratification based on tfh indicates prognostic significance

The flowchart of this study is shown in Fig. 1. Utilizing gene expression profiles from the GSE22762 and GSE39671 datasets, we stratified CLL patients into high and low Tfh groups based on the optimal cutoff value determined with OS and TTFT. In the GSE22671 dataset, the high Tfh group demonstrated significantly better

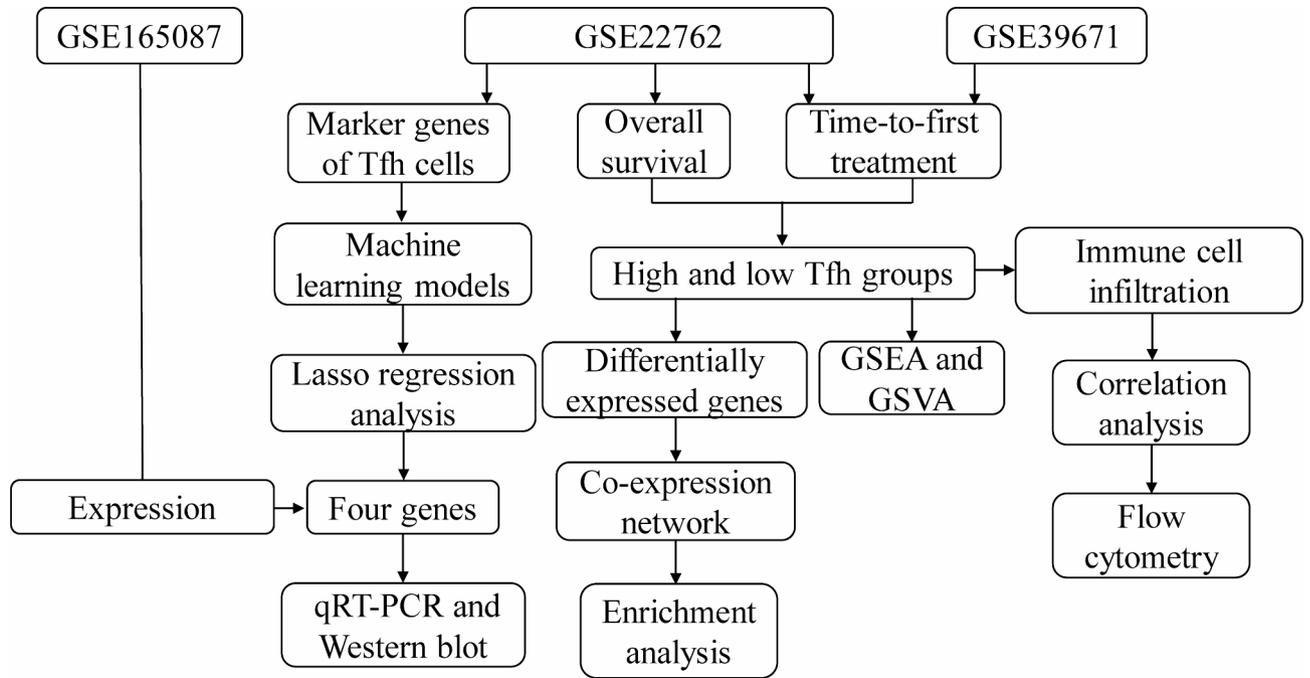


Fig. 1. The flowchart of this study.

prognosis in terms of both OS (Fig. 2A) and TTFT (Fig. 2B) compared to the low Tfh group. In the GSE39671 dataset, the high Tfh group demonstrated significantly better prognosis in terms of TTFT compared to the low Tfh group (Fig. 2C).

Through differential analysis, we found 3696 DEGs in terms of OS in GSE22671 between high Tfh group and low Tfh group (Fig. 2D), 3120 DEGs in terms of TTFT in GSE22671 between high Tfh group and low Tfh group (Fig. 2E), 7577 DEGs in terms of TTFT in GSE39671 between high Tfh group and low Tfh group (Fig. 2F). Intersection analysis showed 1121 common DEGs among three groups of DEGs (Fig. 2G).

Gene networks and biological functions

Common DEGs were further analyzed to construct five co-expression networks through WGCNA (Fig. 3A). Among which the turquoise module showed the highest correlation with Tfh cell levels and TTFT (Fig. 3B). Enrichment analysis showed that genes within turquoise module were significantly enriched in GO functions of immune system process, and leukocyte activation (Fig. 3C). The KEGG pathways including cytokine-cytokine receptor interaction, PI3K-Akt signaling pathway, and natural killer cell mediated cytotoxicity (Fig. 3D).

Additionally, GSEA revealed distinct enrichment profiles between the high and low Tfh groups. In, Immune-related pathways were enriched in patients with high Tfh cell levels in GSE22671 (Fig. 4A) and GSE39671 (Fig. 4B). Including the chemokine signaling pathway, cytokine-cytokine receptor interaction, leukocyte transendothelial migration, natural killer cell mediated cytotoxicity, and T cell receptor signaling pathway. Correlation results showed higher correlation coefficients between Tfh cell levels and immune pathways (Fig. 4C and D).

Correlation with immune cell populations

In the results of ssGSEA, we found that the expression levels of various immune cell types in two distinct patient groups characterized by high and low Tfh group in GSE22671 (Fig. 5A) and GSE39671 (Fig. 5B). The correlation showed that the activated B cells showing a significant negative correlation with Tfh levels, whereas regulatory T cells (Treg), macrophage, and natural killer (NK) T cells exhibited a positive correlation in GSE22671 (Fig. 5C) and GSE39671 (Fig. 5D).

Comparative analysis of predictive models for CLL prognosis

Further integration using machine learning algorithms identified a consensus signature among Tfh marker genes. The best-performing model was the Random Survival Forest (RSF), which achieved an AUC of 0.909 in the training data and an AUC of 0.637 in the validation data, with an average AUC of 0.773 (Fig. 6A). This model includes ten genes: RAE1, PRDX1, CD84, CSF1R, CEBPA, CLEC4A, SIGLEC7, PDCD6, TYRO3, and DPP4 (Figure S1). Ten genes in RSF were used to construct risk model in GSE22762 (Fig. 6B). Low-risk group had better survival probabilities than high-risk group (Fig. 6C). Genes in RSF were also used to construct risk model in GSE39671 (Fig. 6D), and low-risk group had better survival probabilities than high-risk group (Fig. 6E). A nomogram integrating ten gene expression levels with clinical features was developed to predict individual survival probabilities (Fig. 6F). The calibration curve showed observed survival probabilities closely aligning with predictions across different risk groups (Fig. 6G).

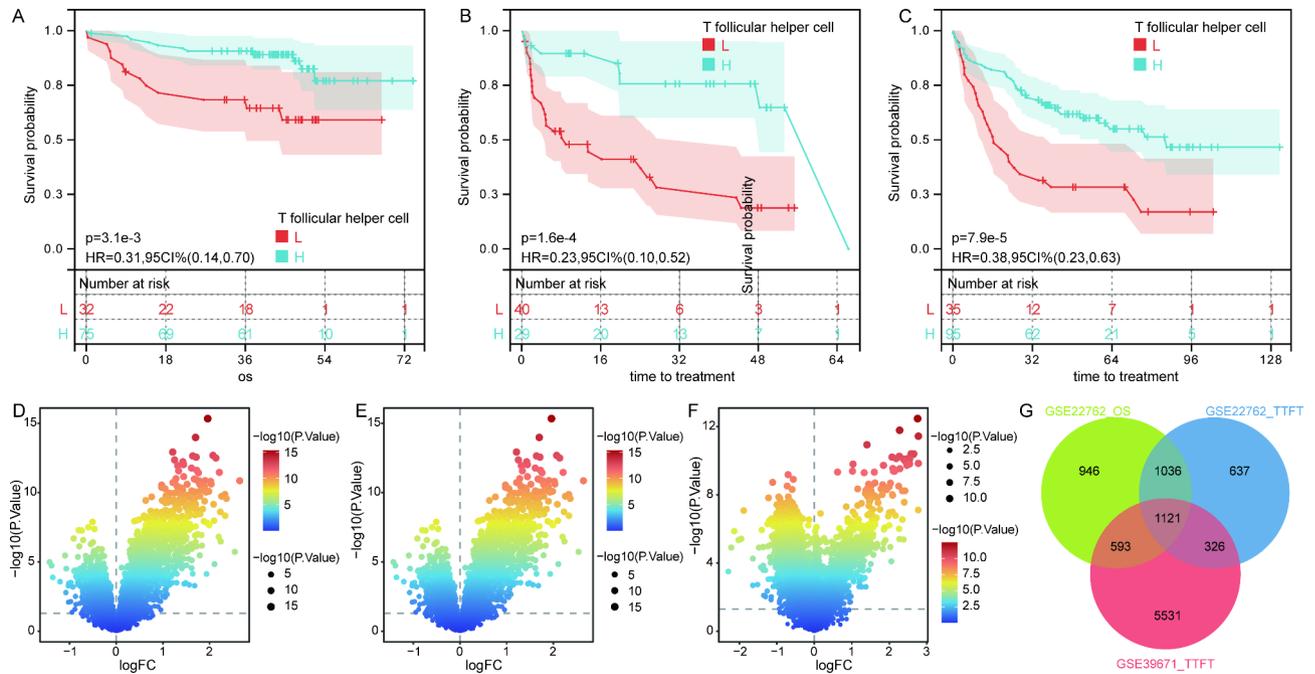


Fig. 2. Identification of high and low Tfh groups in CLL patients. (A) Kaplan-Meier plot showing OS for CLL patients with high (H) versus low (L) T follicular helper groups based on the optimal cutoff value determined with OS in GSE22671. (B) Kaplan-Meier plot showing OS for CLL patients with high (H) versus low (L) T follicular helper groups based on the optimal cutoff value determined with time to first treatment (TTFT) in GSE22671. (C) Kaplan-Meier plot showing OS for CLL patients with high (H) versus low (L) T follicular helper groups based on the optimal cutoff value determined with time to first treatment (TTFT) in GSE39671. (D) Volcano plots of differentially expressed genes (DEGs) between high and low Tfh groups with OS in GSE22671. (E) Volcano plots of differentially expressed genes (DEGs) between high and low Tfh groups with TTFT in GSE22671. (F) Volcano plots of differentially expressed genes (DEGs) between high and low Tfh groups with TTFT in GSE39671. (G) Venn diagram showing the overlap of DEGs associated with OS in GSE22762, TTFT in GSE22762, and TTFT in GSE39671.

Employing Lasso regression yielded a subset of genes with non-zero coefficients at the optimal penalty value ($\lambda = 0.03$), and identified 8 genes (SIGLEC9, CDK5R1, PRDX1, CD84, KIR2DL2, CLEC4A, NCAM1, and RAE1) significantly affecting patient prognosis (Fig. 7A and B). We then constructed a risk model based on 8 genes from the Lasso model in GSE39671, and divided CLL samples into high-risk and low-risk groups (Fig. 7C). The area under the ROC curve (AUC) indicated the risk score had high predictive accuracy across time points for 12 months (AUC = 0.96), 36 months (AUC = 0.88), and 60 months (AUC = 0.91) (Fig. 7D). The high-risk group showed markedly reduced survival probabilities, with a pronounced hazard ratio indicating increased risk (Fig. 7E). A nomogram was established to predict individual survival probabilities for 8 genes and highlighted four genes (CLEC4A, RAE1, CD84, and PRDX1) with a notable impact on OS (Fig. 7F). The calibration curve showed a close alignment along the 45-degree line indicates a well-calibrated model (Fig. 7G).

Evaluation of prognostic markers

Furthermore, the RSF results showed that CLEC4A, and RAE1 were protective factors for CLL (OR < 1), while CD84, and PRDX1 were risk factors (OR > 1) in terms of both OS in GSE22671 (Fig. 8A) and TTFT in GSE39671 (Fig. 8B). Compared to low Tfh group, CLEC4A, and RAE1 were higher expression and CD84, and PRDX1 were lower expression in high Tfh group in terms of OS in GSE22671 (Fig. 8C), TTFT in GSE22671 (Fig. 8D), and in GSE39671 (Fig. 8E).

Through analysis of scRNA-seq data from GSE165087 dataset, we identified 42 cell clusters (Fig. 9A). The annotation of clusters was performed and we obtained 6 cell types, including CD4 + T cells, CD8 + T cells, NK cells, NKT cells, B cells, and monocytes (Table S2, Figure 9B). The expression distribution of CLEC4A, RAE1, CD84, and PRDX1 across cell types (Fig. 9C and D) showed that CLEC4A were mainly expressed in monocytes, PRDX1 were mainly expressed in NK cells, NKT cells, B cells, and monocytes, CD84, and RAE1 were mainly expressed in B cells.

Validation of prognostic markers expression and immune cell levels

Experimental validation through RT-qPCR and Western blot confirmed the differential expression of the prognostic markers (CLEC4A, RAE1, CD84, and PRDX1) between the high and low Tfh groups (Fig. 10A and B). Consistent with the bioinformatics analysis, CLEC4A, and RAE1 were higher expression in high Tfh group than low Tfh group, CD84, and PRDX1 were lower expression high Tfh group. Flow cytometry results revealed

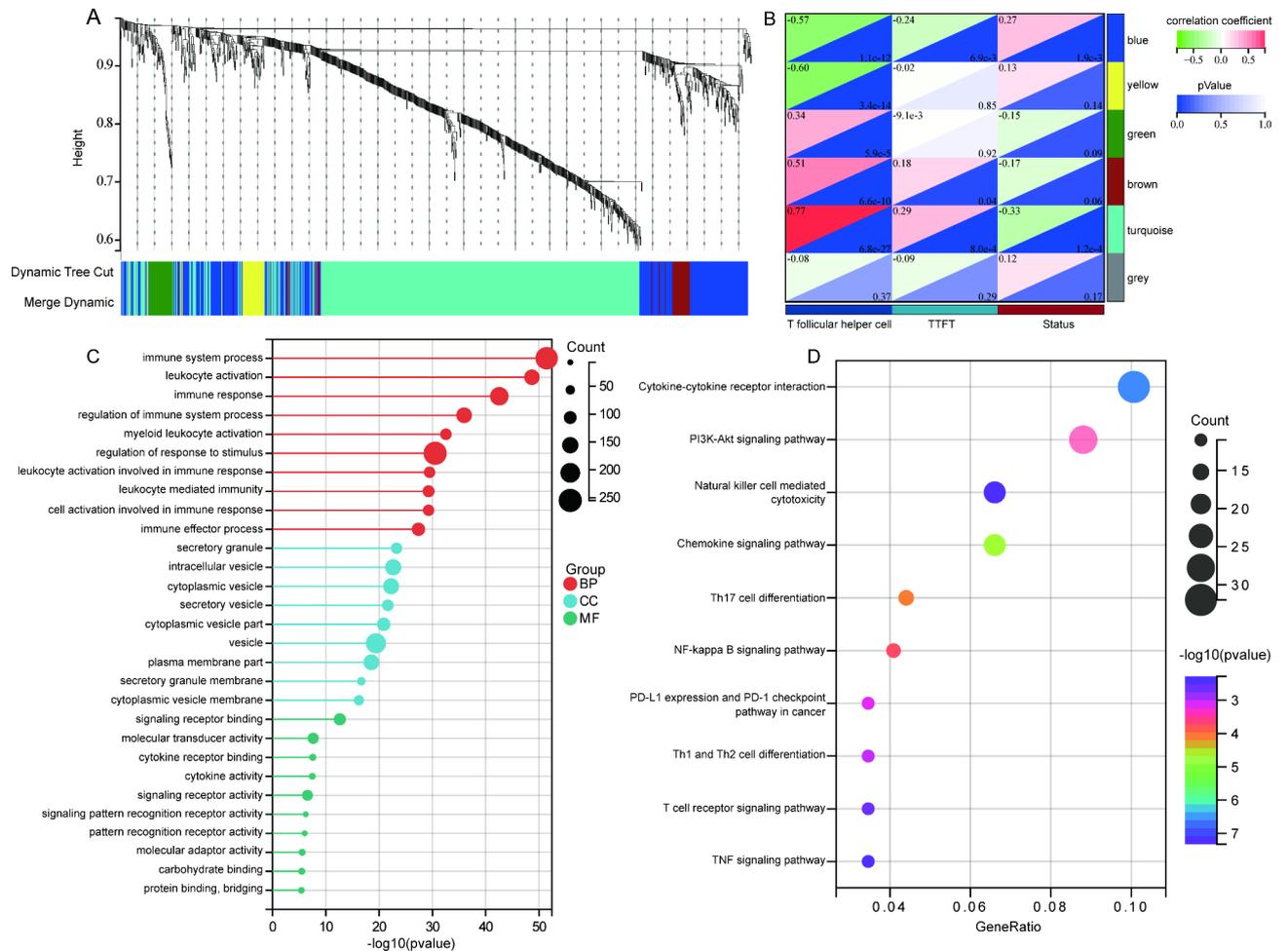


Fig. 3. Co-expression network analysis and biological function characterization. **(A)** Dendrogram obtained from hierarchical clustering using WGCNA, with branches representing genes and colors below indicating identified modules through dynamic tree cutting. **(B)** Heatmap of module-trait relationships. **(C)** GO enrichment analysis for turquoise module. The most significantly enriched biological processes (BP), cellular components (CC), and molecular functions (MF). **(D)** KEGG pathways enriched by genes within turquoise module.

a lower level of activated B cells and a higher level of Tregs in the high Tfh group compared to the low Tfh group (Table S3, Fig. 10C).

Discussion

T cell immune dysfunction is a prominent feature of CLL and a major cause of failure in immunotherapy and multidrug resistance¹⁸. This study analyzed the role of Tfh cells in CLL based on gene expression profiles and immune cell levels, several key points should be covered to interpret the results, contextualize them within the existing literature, and propose directions for future research.

Our study demonstrates a significant association between Tfh cell levels and CLL patient prognosis, with higher Tfh cell levels correlating with improved OS and TTFT. This contradicts the theory that previous studies have shown that Tfh may promote the progression of CLL⁹. However, this may be related to the characteristics of Tfh cell subpopulations⁵. This difference may reflect the complex role of Tfh cells in the tumor microenvironment, which may affect tumor development in multiple ways. IL-21, a cytokine that is highly expressed by Tfh cells, which can promote B cell apoptosis and activate the innate immune system^{19,20}. In addition, IL-21 effectively amplified NK cells from healthy volunteers and CLL patients, which have high cytotoxicity to CLL cells²¹.

The differentially expressed genes between high Tfh group and low Tfh group were used to construct five co-expression networks. Particularly, the turquoise module showed the highest correlation with Tfh cell level and TTFT. Enrichment analysis suggested further identified genes closely related to immune system processes and cell activation, supporting the importance of Tfh cells in regulating immune responses. Especially, cytokine-cytokine receptor interaction, PI3K-Akt signaling pathway, and NK cell mediated cytotoxicity may underlie the impact of Tfh cells on CLL. IL-21 and IL-4 are the main cytokines produced by Tfh, and their mediated cytotoxicity contributes to anti-tumor effects²²⁻²⁴. PI3K plays a crucial role in promoting the BCR pathway of CLL proliferation and survival, and the potential efficacy of PI3K inhibitors has been widely recognized in

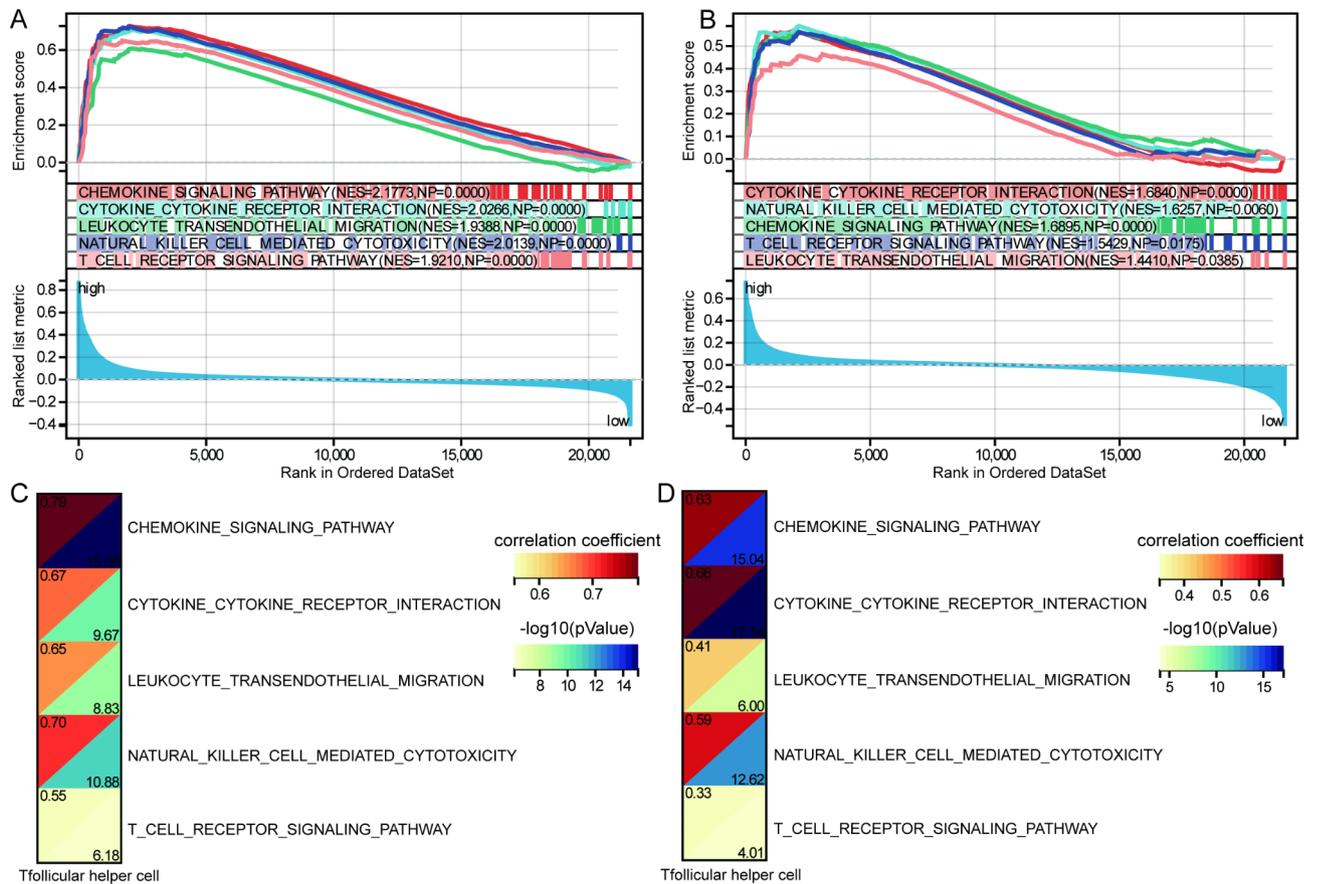


Fig. 4. Correlation of Tfh cell levels with pathway activation in CLL. **(A)** GSEA for the high and low Tfh group based on TTFT in GSE22671 showing enriched pathways at the top-ranked portion of the gene list. **(B)** GSEA for the high and low Tfh group based on TTFT in GSE39671 showing enriched pathways at the top-ranked portion of the gene list. NES normalized enrichment score, NP nominal p value. **(C)** Heatmap displaying the correlation coefficients between Tfh cell levels and pathways in GSE22671. **(D)** Heatmap displaying the correlation coefficients between Tfh cell levels and pathways in GSE39671.

CLL treatment^{25,26}. NK cells play a crucial role in defending against malignant tumors, including CLL²⁷. These pathways have been implicated in immune response modulation and could offer insights into the mechanisms through which Tfh cells influence CLL biology.

The observed correlations between Tfh cell levels and other immune cell populations, including a negative correlation with activated B cells and a positive correlation with Tregs and NK cells, further elucidate the complex immune landscape of CLL. Compared to the healthy control group, the Treg percentage of CLL subjects decreased, further confirming the immune complexity of CLL²⁸. The results of this study suggest a significant positive correlation between Tfh cells and NK cells, which are innate lymphocytes that induce direct cytotoxicity against tumor cells by releasing cytotoxic particles and death receptor signals²⁹. These interactions highlight the intricate balance between pro-tumorigenic and antitumorigenic forces within the CLL microenvironment.

By integrating machine learning algorithms, we successfully identified prognostic features associated with Tfh marker genes and constructed a risk model based on these features. The identification of four genes (CLEC4A, RAE1, CD84, and PRDX1) as significant prognostic markers underscores the potential for molecular profiling to enhance CLL prognosis and treatment strategies. CLEC4A was target of IL-4 with the highest levels of change in CLL³⁰. CLEC4A is expressed on monocytes, neutrophils, DCs, and plasma like dendritic cells, and is involved in anti-tumor immunity^{31,32}. Reducing RAE1 expression induces NK cell tolerance and weakens the anti-tumor activity of NK cells in CLL^{33,34}. CD84 bridges the gap between CLL cells and their microenvironment, blocking as a therapeutic strategy to reverse CLL induced immunosuppression³⁵. NK cells with high expression of antioxidant enzyme PRDX1 exhibit resistance to oxidative stress and improved anti-tumor function³⁶. Their varied expression levels between high and low Tfh cell groups underscore the molecular heterogeneity of CLL and the potential for targeted therapies to disrupt pathways critical to CLL progression and immune evasion.

This study leverages multiple high-quality gene expression datasets and single-cell RNA sequencing data, offering a robust and comprehensive analysis of Tfh cells in CLL. By identifying and validating prognostic markers associated with Tfh cells, this research provides potential biomarkers for predicting patient outcomes, which could be pivotal for personalized treatment strategies. The use of advanced machine learning algorithms to develop predictive models enhances the accuracy and reliability of the prognostic markers identified in this

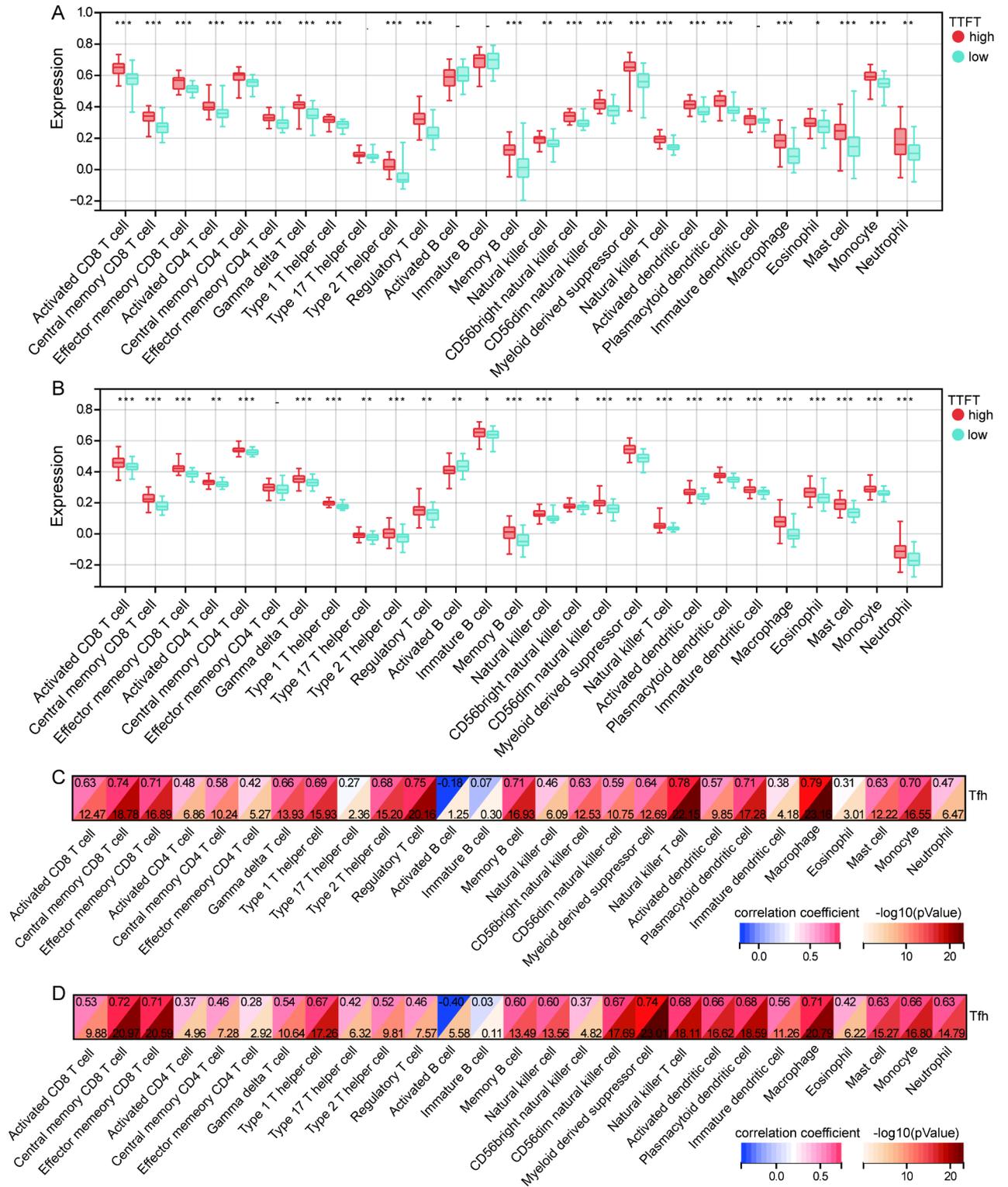


Fig. 5. Comparative immune cell type expression and correlation with Tfh levels in CLL. **(A)** Boxplots displaying the expression levels of immune cell types categorized into high and low Tfh groups based on TTFT in GSE22671. **(B)** Boxplots displaying the expression levels of immune cell types categorized into high and low Tfh groups based on TTFT in GSE39671. **(C)** Heatmap of correlation coefficients between the expression levels of different immune cell types and Tfh cell levels in GSE22671. **(D)** Heatmap of correlation coefficients between the expression levels of different immune cell types and Tfh cell levels in GSE39671.

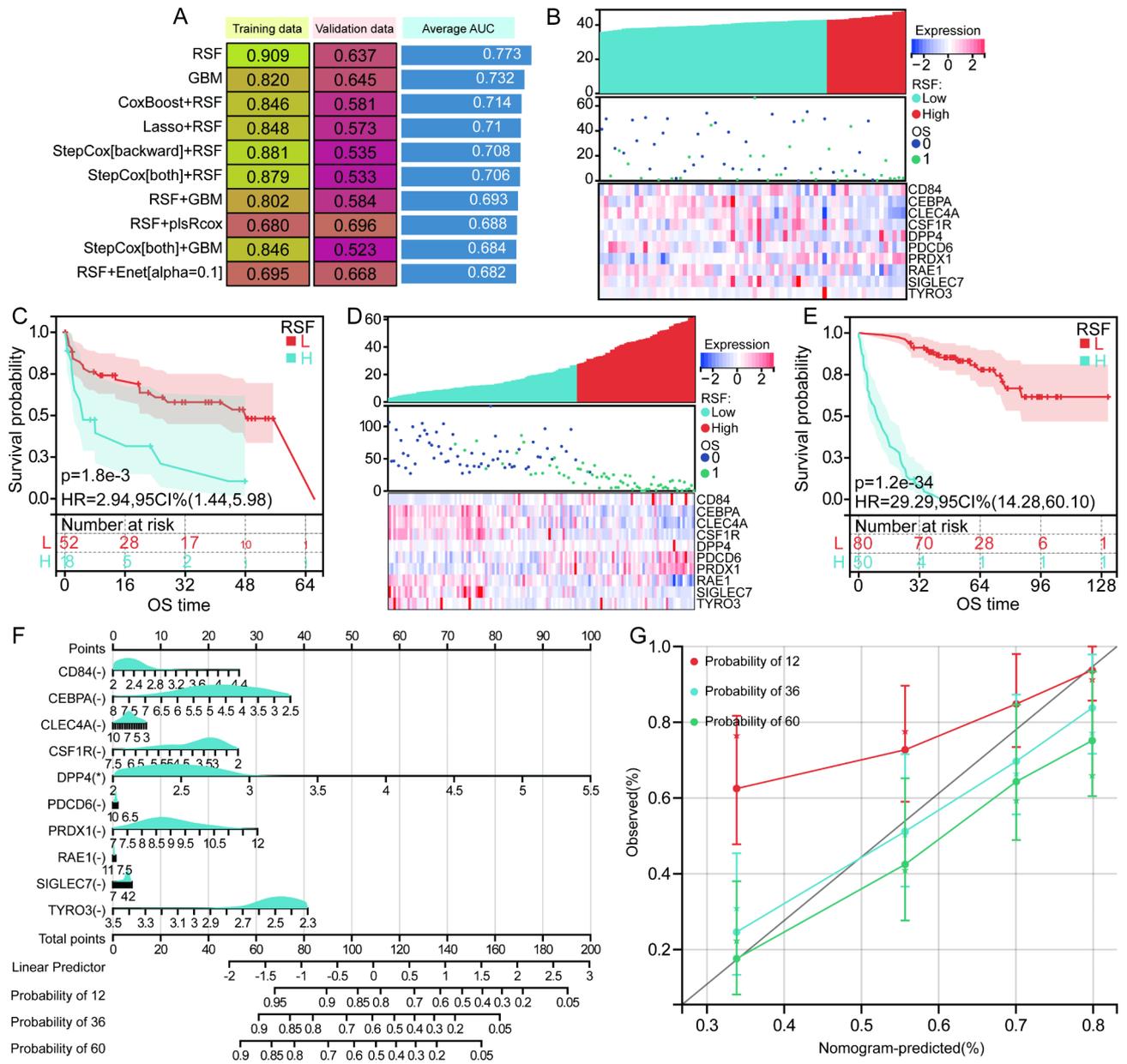


Fig. 6. Performance of prognostic models and correlation with clinical outcomes in CLL. **(A)** Bar chart comparing the performance of main machine learning models in predicting clinical outcomes for CLL. **(B)** Risk model constructed by ten genes of RSF in GSE22762. **(C)** Kaplan-Meier survival curve detailing the survival probability for the high-risk and low-risk group. HR, hazard ratios. **(D)** Risk model constructed by ten genes of RSF in GSE39671. **(E)** Kaplan-Meier survival curve detailing the survival probability for the high-risk and low-risk group. HR, hazard ratios. **(F)** A nomogram predicted survival probabilities at 12, 36, and 60 months based on ten genes of RSF model. **(G)** Calibration curves assessing the nomogram’s predictive accuracy.

study. However, it is important to acknowledge limitations and challenges. Including samples from multiple centers for experimental validation would enhance the generalizability and robustness of the findings. A more detailed analysis of the different subsets of Tfh cells and their specific roles in CLL could provide deeper insights and identify more targeted therapeutic strategies. Expanding the immunophenotyping to include additional immune cell types and markers could offer a more comprehensive view of the immune landscape in CLL and its interaction with Tfh cells. Utilizing in vivo models, such as patient-derived xenografts or genetically engineered mouse models, could provide further validation and mechanistic insights into the role of Tfh cells in CLL.

Conclusion

Our findings highlight the prognostic significance of Tfh cells in CLL and suggest mechanisms by which Tfh cells may influence CLL progression. The identified genes and pathways offer new insights into the CLL immune

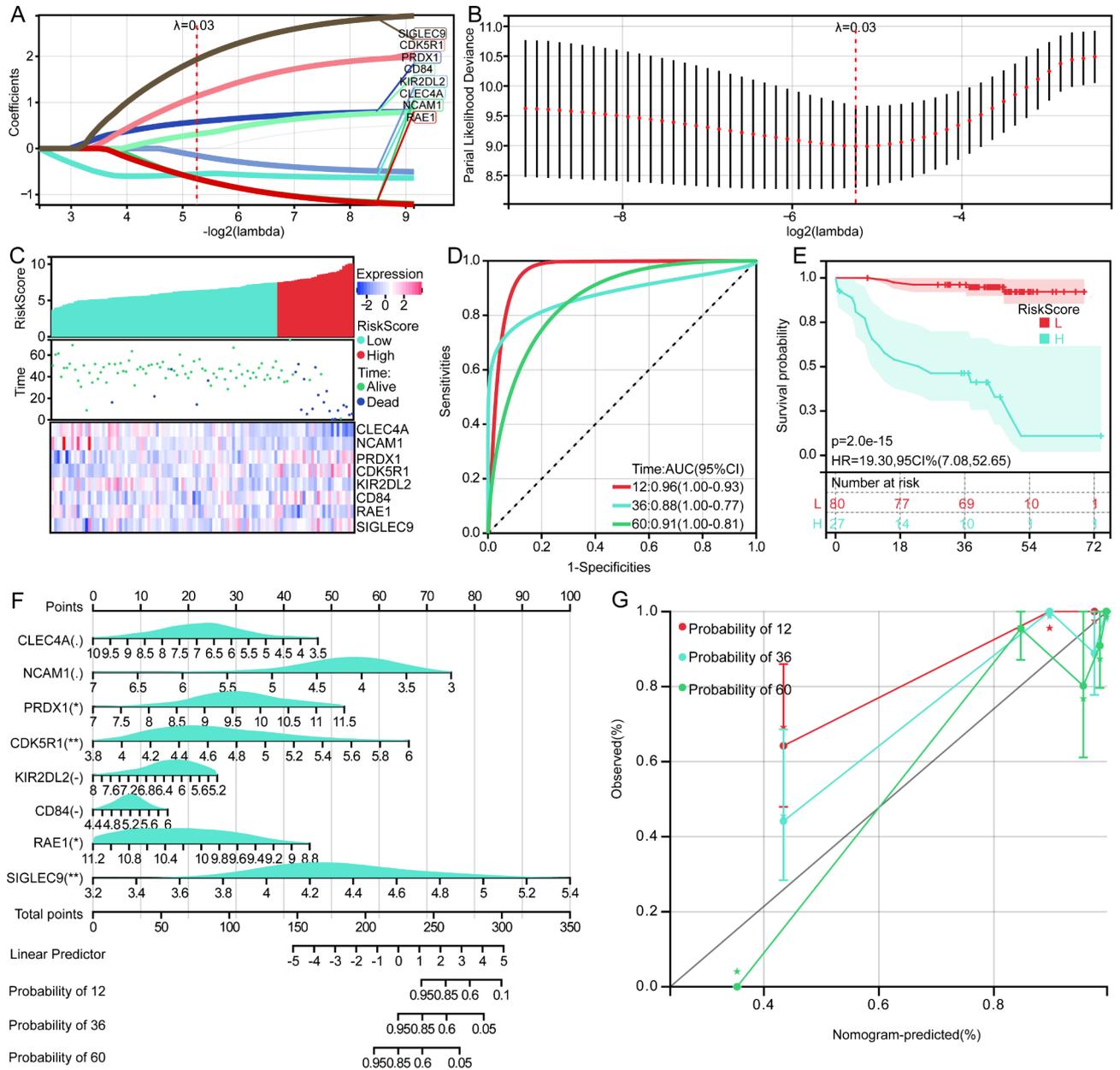


Fig. 7. Predictive modeling and survival analysis selecting prognostic markers. **(A)** Lasso coefficient profiles of prognostic markers across a range of lambda values in CLL. **(B)** Partial likelihood deviance plot for model tuning across different λ values. **(C)** Risk model constructed based on markers of Lasso model. **(D)** Receiver operating characteristic curves evaluating the model's discriminative ability. AUC, area under the curve. **(E)** Kaplan-Meier survival plots stratified by low and high-risk groups calculated from the risk model. **(F)** A prognostic nomogram integrating the expression of selected markers to predict the probability of survival at 12, 36, and 60 months. **(G)** Calibration plots demonstrating the accuracy of the nomogram's predictions compared to actual observed outcomes.

landscape and potential targets for therapeutic intervention, paving the way for improved CLL management and patient outcomes.

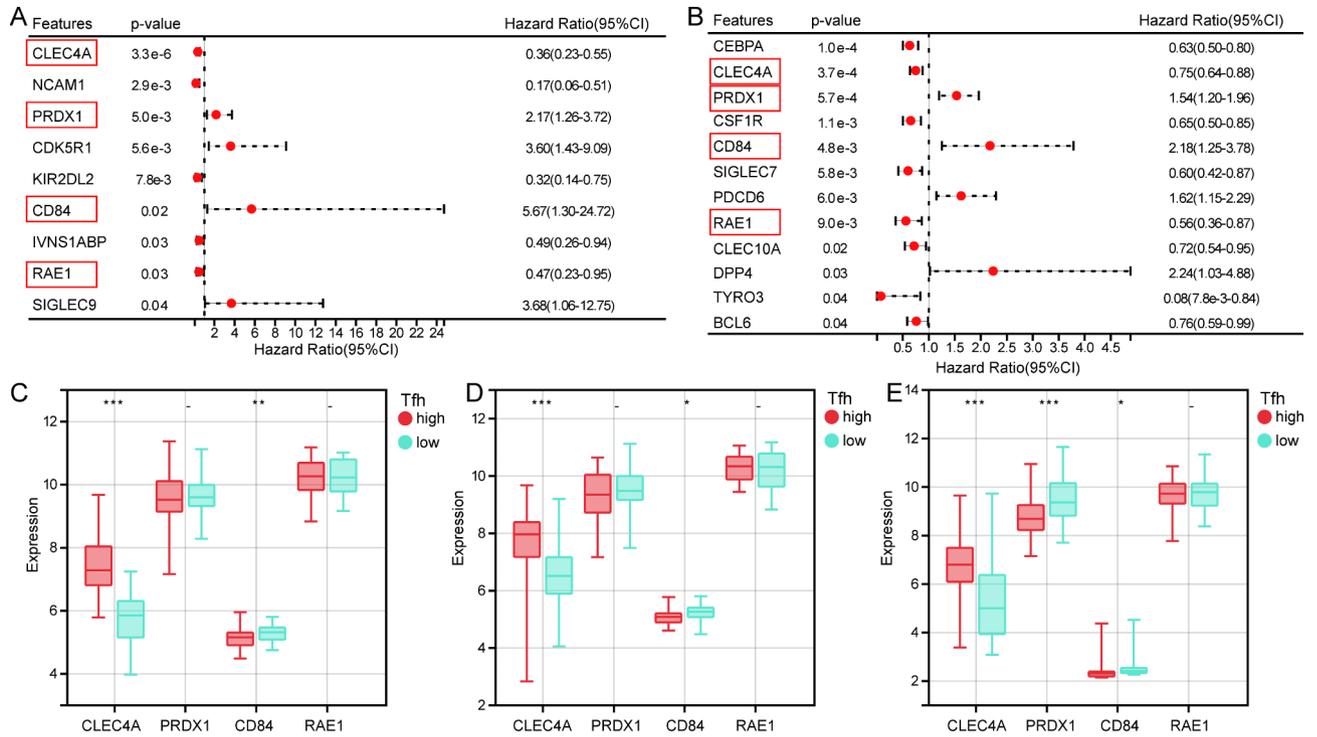


Fig. 8. Analysis of prognostic markers in CLL. **(A)** Forest plot illustrating the hazard ratios and 95% confidence intervals (CIs) for prognostic markers in predicting CLL outcomes in terms of OS in GSE22671. HR, hazard ratio. **(B)** Forest plot illustrating the hazard ratios and 95% confidence intervals (CIs) for prognostic markers in predicting CLL outcomes in terms of TTFT in GSE39671. HR, hazard ratio. Expression of prognostic markers in high Tfh and low Tfh group in terms of OS in GSE22671 **(C)**, TTFT in GSE22671 **(D)**, and in GSE39671 **(E)**. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

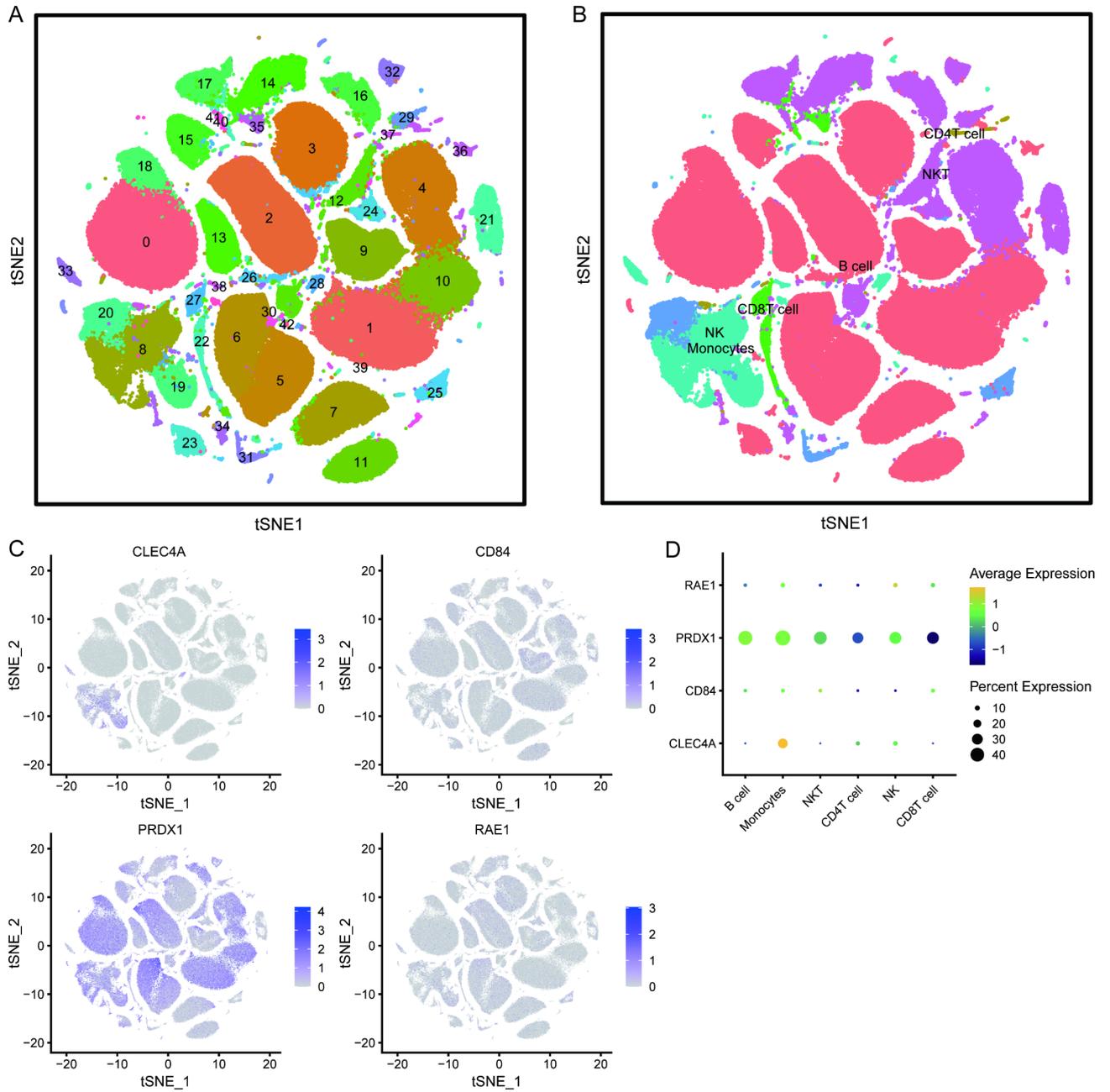


Fig. 9. Expression mapping and immune cell landscape in scRNA-seq data in GSE165087. **(A)** t-SNE showing distinct cellular clusters. **(B)** t-SNE showing identified immune cell types. **(C)** Expression density maps for CLEC4A, CD84, PRDX1, and RAE1 across the t-SNE landscape. **(D)** Expression bubble plot for CLEC4A, CD84, PRDX1, and RAE1 across cell types.

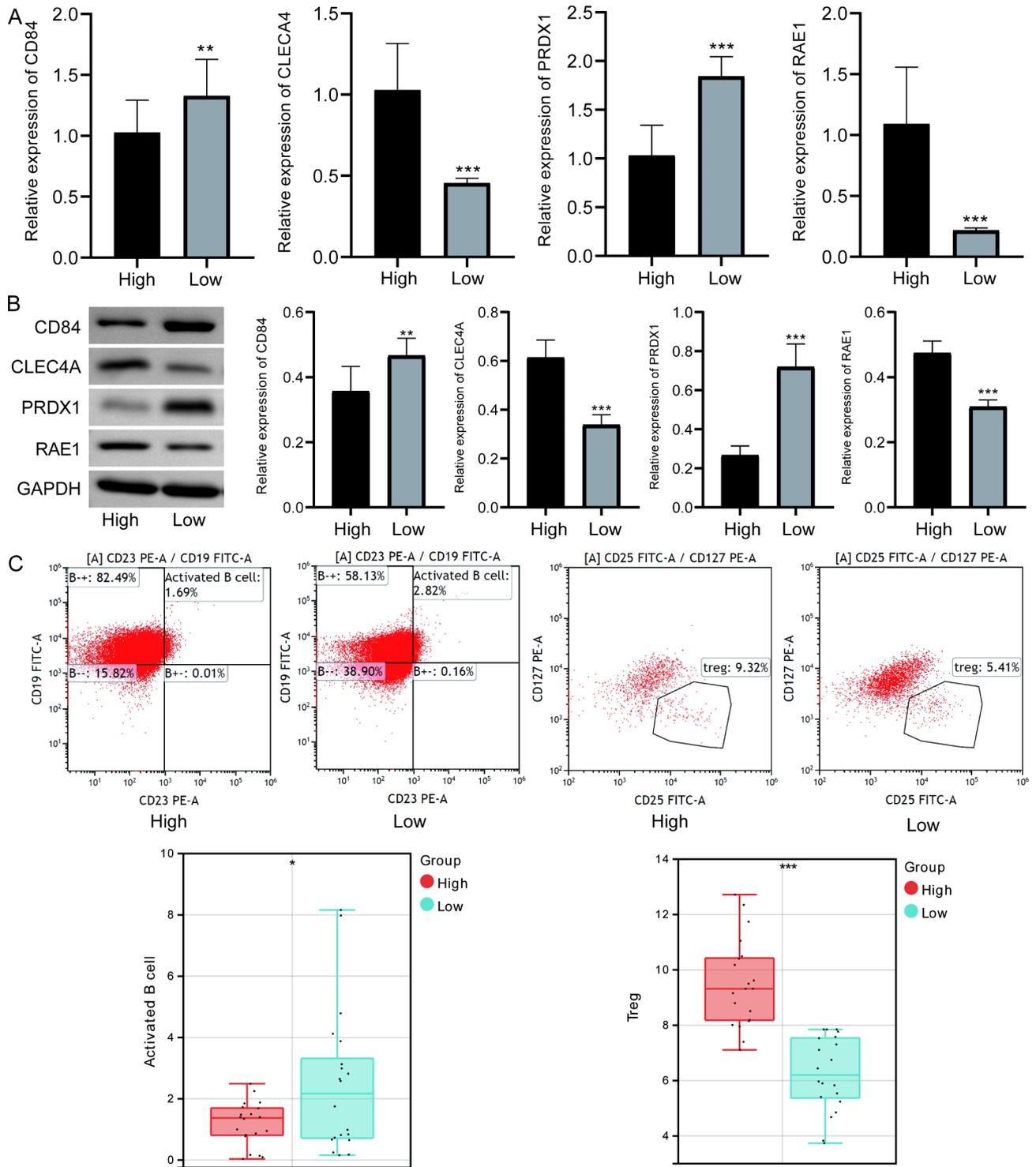


Fig. 10. The levels of prognostic markers and immune cells in 20 high Tfh patients and 20 low Tfh patients with three biological replicates. **(A)** The mRNA levels of prognostic markers detected by RT-qPCR. **(B)** The protein expression of prognostic markers detected by Western blot. Original blots are presented in Figure S2. **(C)** The levels of immune cells detected by flow cytometry. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Data availability

The datasets generated and/or analyzed during the current study are available from GEO database with GSE22762, GSE39671, and GSE165087.

Received: 12 April 2024; Accepted: 16 September 2024

Published online: 28 September 2024

References

- Kennedy, E. et al. TLR9 expression in chronic lymphocytic leukemia identifies a promigratory subpopulation and novel therapeutic target. *Blood* **137** (22), 3064–3078 (2021).
- Kipps, T. J. et al. Chronic lymphocytic leukaemia. *Nat. Rev. Dis. Primers* **3**, 16096 (2017).
- Hengeveld, P. J. et al. High-throughput proteomics identifies THEMIS2 as independent biomarker of treatment-free survival in untreated CLL. *Hemasphere* **7** (10), e951 (2023).
- Agius, R. et al. Machine learning can identify newly diagnosed patients with CLL at high risk of infection. *Nat. Commun.* **11** (1), 363 (2020).
- Wu, X. et al. Altered T follicular helper cell subsets and function in chronic lymphocytic leukemia. *Front. Oncol.* **11**, 674492 (2021).
- Patel, P. S. et al. Translational regulation of TFH cell differentiation and autoimmune pathogenesis. *Sci. Adv.* **8** (25), eabo1782 (2022).
- Nie, Y., Yun, X., Zhang, Y. & Wang, X. Targeting metabolic reprogramming in chronic lymphocytic leukemia. *Exp. Hematol. Oncol.* **11** (1), 39 (2022).
- Vaca, A. M. et al. Activation and expansion of T-follicular helper cells in chronic lymphocytic leukemia nurselike cell co-cultures. *Leukemia* **36** (5), 1324–1335 (2022).
- Le Saos-Patrinis, C. et al. Elevated levels of circulatory follicular T helper cells in chronic lymphocytic leukemia contribute to B cell expansion. *J. Leukoc. Biol.* **113** (3), 305–314 (2023).
- Koh, C. H. et al. G1TR agonism triggers antitumor immune responses through IL21-Expressing follicular helper T cells. *Cancer Immunol. Res.* **8** (5), 698–709 (2020).
- Herold, T. et al. An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia. *Leukemia* **25** (10), 1639–1645 (2011).
- Chuang, H. Y. et al. Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood* **120** (13), 2639–2649 (2012).
- Penter, L. et al. Longitudinal single-cell dynamics of chromatin accessibility and mitochondrial mutations in chronic lymphocytic leukemia Mirror Disease History. *Cancer Discov.* **11** (12), 3048–3063 (2021).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16** (5), 284–287 (2012).
- Liu, Z. et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat. Commun.* **13** (1), 816 (2022).
- Charoentong, P. et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell. Rep.* **18** (1), 248–262 (2017).
- Hallek, M. & Al-Sawaf, O. Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures. *Am. J. Hematol.* **96** (12), 1679–1705 (2021).
- Liang, T. et al. Decreased TCF1 and BCL11B expression predicts poor prognosis for patients with chronic lymphocytic leukemia. *Front. Immunol.* **13**, 985280 (2022).
- Roessner, P. M. & Seiffert, M. T-cells in chronic lymphocytic leukemia: guardians or drivers of disease? *Leukemia* **34** (8), 2012–2024 (2020).
- Gowda, A. et al. IL-21 mediates apoptosis through up-regulation of the BH3 family member BIM and enhances both direct and antibody-dependent cellular cytotoxicity in primary chronic lymphocytic leukemia cells in vitro. *Blood* **111** (9), 4723–4730 (2008).
- Yano, M. et al. Evaluation of allogeneic and autologous membrane-bound IL-21-expanded NK cells for chronic lymphocytic leukemia therapy. *Blood Adv.* **6** (20), 5641–5654 (2022).
- Ahearn, M. J. et al. Enhancement of CD154/IL4 proliferation by the T follicular helper (tfh) cytokine, IL21 and increased numbers of circulating cells resembling tfh cells in chronic lymphocytic leukaemia. *Br. J. Haematol.* **162** (3), 360–370 (2013).
- Bhatt, S. et al. Direct and immune-mediated cytotoxicity of interleukin-21 contributes to antitumor effects in mantle cell lymphoma. *Blood* **126** (13), 1555–1564 (2015).
- Qian, F. et al. Interleukin-4 treatment reduces leukemia burden in acute myeloid leukemia. *FASEB J.* **36** (5), e22328 (2022).
- Guarente, V. & Sportoletti, P. Lessons, challenges and Future Therapeutic opportunities for PI3K Inhibition in CLL. *Cancers (Basel)* **13** (6) (2021).
- Chen, S. S. et al. Duvelisib eliminates CLL B cells, impairs CLL-Supporting cells, and overcomes Ibrutinib Resistance in a Xenograft Model. *Clin. Cancer Res.* **29** (10), 1984–1995 (2023).
- Barshidi, A. et al. The role of exhausted natural killer cells in the immunopathogenesis and treatment of leukemia. *Cell. Commun. Signal.* **22** (1), 59 (2024).
- Rubino, V. et al. Adaptive and innate cytotoxic effectors in chronic lymphocytic leukaemia (CLL) subjects with stable disease. *Int. J. Mol. Sci.* **24** (11) (2023).
- Fisher, J. G. et al. XPO1 inhibition sensitises CLL cells to NK cell mediated cytotoxicity and overcomes HLA-E expression. *Leukemia* **37** (10), 2036–2049 (2023).
- Ruiz-Lafuente, N. et al. The gene expression response of chronic lymphocytic leukemia cells to IL-4 is specific, depends on ZAP-70 status and is differentially affected by an NFκB inhibitor. *PLoS ONE* **9** (10), e109533 (2014).
- Chiffolleau, E. C-Type lectin-like receptors as emerging orchestrators of sterile inflammation represent potential therapeutic targets. *Front. Immunol.* **9**, 227 (2018).
- Yu, J., Sun, H., Cao, W., Song, Y. & Jiang, Z. Research progress on dendritic cell vaccines in cancer immunotherapy. *Exp. Hematol. Oncol.* **11** (1), 3 (2022).
- Lemasson, Q., Akil, H., Feuillard, J. & Vincent-Fabert, C. Genetically Engineered Mouse models support a major role of Immune checkpoint-dependent immunosurveillance escape in B-Cell lymphomas. *Front. Immunol.* **12**, 669964 (2021).
- Wang, Z. H., Li, W., Dong, H. & Han, F. Current state of NK cell-mediated immunotherapy in chronic lymphocytic leukemia. *Front. Oncol.* **12**, 1077436 (2022).
- Lewinsky, H. et al. CD84 regulates PD-1/PD-L1 expression and function in chronic lymphocytic leukemia. *J. Clin. Invest.* **128** (12), 5465–5478 (2018).
- Klopotowska, M. et al. PRDX-1 supports the survival and antitumor activity of primary and CAR-Modified NK cells under oxidative stress. *Cancer Immunol. Res.* **10** (2), 228–244 (2022).

Author contributions

Rui Zhang contributed to design the study and write the manuscript. Sha Guo analyzed data and performed experiment. Jianhua Qu drafted and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the “Tianshan Innovation Team Plan” by the Science and Technology Department

of Xinjiang Uygur Autonomous Region (no. 2022D14008).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This study was approved by Ethics Committee of the First Affiliated Hospital of Xinjiang Medical University (No. 20211015-47).

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-73325-8>.

Correspondence and requests for materials should be addressed to J.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024