ORIGINAL ARTICLE

WILEY

Prognostic and predictive value of a mRNA signature in peripheral T-cell lymphomas: A mRNA expression analysis

Jiannan Tu[1] | Zhixing Kuang[2] 🆔 | Xiaoliang Xie[3] | Shizhen Wu[1] | Ting Wu[1] |
Shengchi Chen[1] 🆔

[1]Department of Oncology, Nanping First Hospital Affiliated to Fujian Medical University, Nanping, China

[2]Department of Radiation Oncology, Nanping First Hospital Affiliated to Fujian Medical University, Nanping, China

[3]Department of Orthopedics, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

Correspondence
Zhixing Kuang, Department of Radiation Oncology, Nanping First Hospital Affiliated to Fujian Medical University, 353000, Nanping, China.
Email: zhixingkuang@gmail.com

Shengchi Chen, Department of Oncology, Nanping First Hospital Affiliated to Fujian Medical University, Nanping, 353000, China.
Email: npcscchen@gmail.com

Funding information
Starup Fund for scientific, fujian Medical University, Grant/Award Number: 2018QH1173

Abstract

Current international prognostic index is widely questioned on the risk stratification of peripheral T-cell lymphoma and does not accurately predict the outcome for patients. We postulated that multiple mRNAs could combine into a model to improve risk stratification and helping clinicians make treatment decisions. In this study, the gene expression profiles were downloaded from the Gene Expression Omnibus (GEO) database. Weighted gene co-expression network analysis (WGCNA) was used to screening genes in selected module which most closely related to PTCLs, and then built a mRNA signature using a LASSO Cox regression model and validated the prognostic accuracy of it. Finally, a nomogram was constructed and the performance was assessed. A total of 799 WGCNA-selected mRNAs in black module were identified, and a mRNA signature which based on DOCK2, GSTM1, H2AFY, KCNAB2, LAPTM5 and SYK for PTCLs was developed. Significantly statistical difference can be seen in overall survival of PTCLs between low-risk group and high-risk group (training set:hazard ratio [HR] 4.3, 95% CI 2.4-7.4, $P < .0001$; internal testing set:hazard ratio [HR] 2.4, 95% CI 1.2-4.8, $P < .01$; external testing set:hazard ratio [HR] 2.3, 95% CI 1.10-4.7, $P = .02$). Furthermore, multivariate regression demonstrated that the signature was an independently prognostic factor. Moreover, the nomogram which combined the mRNA signature and multiple clinical factors suggesting that predicted survival probability agreed well with the actual survival probability. The signature is a reliable prognostic tool for patients with PTCLs, and it has the potential for clinicians to implement personalized therapeutic regimen for patients with PTCLs.

KEYWORDS
mRNA, peripheral T-cell lymphomas, predictive, Prognostic, signature

---

Tu and Kuang authors contributed equally to this work.

1 | INTRODUCTION

Non-Hodgkin lymphomas are clonal neoplasms that arise from lymphocyte at various stages of maturation,[1] it estimated that 77 240 new cases of non-Hodgkin lymphoma are expected in the United States, and 19 940 patents will die for this disease in 2020.[2] Peripheral T-cell lymphomas (PTCLs) are a subgroup of non-Hodgkin lymphomas which also characterized as a infrequency and heterogeneous aggressive behaviour diseases that associated with very dismal prognosis, representing 10%–15% of non-Hodgkin lymphomas (NHLs) in Western countries but up to 35% in some countries of Asian.[3] Peripheral T-cell lymphomas (PTCLs) comprise more than 30 distinct histologic subtypes including anaplastic lymphoma kinase (ALK)-positive anaplastic large cell lymphoma (ALCL) and ALK-negative ALCL, extranodal natural killer (NK)/T-cell lymphoma (ENKTL), angioimmunoblastic T-cell lymphoma(AITL), and PTCL, not otherwise specified (PTCL-NOS) according to World Health Organization (WHO) classification system 2017.[4] Numerous attempts have been made to optimize the treatment approach, but no definitive standard therapy has been reached.[5] The traditionally combination regimens such as CHOP or a CHOP-like regimen which initially established for aggressive B-cell lymphomas are most widely used in PTCLs patient.[6] However, outcomes for most patients treated with CHOP are still poor, with only 33%-43% with PTCLs achieving a complete response (CR) and 5-year overall survival (OS) barely exceeds achieving 38.5%.[7] Given the poor outcomes in PTCLs, several novel drugs such as pralatrexate, Mogamulizumab, Chidamide, romidepsin, brentuximab vedotin, and Forodesine have been approved by FDA for the treatment of relapsed and refractory PTCLs recently,[8] but none of these new drugs led to improvement of survival.[9,10] Moreover, the role of stem-cell transplantation for PTCLs remains controversial in front-line settings.[11] There may be a role for prognostic biomarkers in risk classification of PTCLs patients. High-risk patients could receive more intensive treatment to avoid insufficient treatment, whereas low-risk patients should choose low-intensity treatment regime to avoid excessive drug toxicity. Therefore, it is urgent to identify robust biomarkers for predict the prognosis of PTCLs, and discriminate patients who might benefit from the therapy.

To date, the most widely used model for evaluating the prognosis of peripheral T-cell lymphoma is international prognostic index (IPI) that based on performance status, lactate dehydrogenase, extranodal involvement, stage and age, which was initially established for diffuse large B-cell lymphoma (DLBCL). However, Given the marked heterogeneity among the patients that diagnosed with PTCLs, the IPI score is far less satisfactory for distinguishing recurrence risk for PTCLs patients than for aggressive B-cell lymphoma.[12] For example, even patients which categorized in the best risk group (IPI 0) still experience an extremely unfavourable outcome, the cause of this phenomenon is attributed to that IPI score only focused on clinical characteristics, with very few genomic information reflecting the molecular mechanism underlying the PTCLs biology. On the other hand, the lack of information on risk stratification brings the merits

of limitations for clinicians to conduct individualized treatment strategies. Recently, several gene expression biomarker signatures that based on gene expression profiling (GEP) and whole-genome methylation profiling have been build and used to predict the prognosis of human cancer,[13-16] but none mRNA signatures have been utilized for PTCLs patients.

Weighted gene co-expression network analysis (WGCNA) is powerful screening approach and has been gradually valued in discovery of novel biomarkers or therapeutic targets via construct free-scale gene co-expression networks.[17] In this study, we explore the correlation between PTCLs and gene sets by WGCNA. Furthermore, the univariate proportional hazards analysis and LASSO Cox regression were carried out to identify a mRNA signature which beyond clinical parameters and significant associated with PTCLs prognosis. Finally, a prognostic nomogram was established based on the combination of signature and clinical characteristics.

2 | MATERIALS AND METHODS

2.1 | Data sources and data processing

The raw data of GSE59307, GSE58445, GSE19069, GSE90597 and GSE53798 were downloaded from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database, and all datasets except GSE90597 were built based on the GPL570 platform [HG-U133_Plus_2]. A total of 14 samples of cutaneous T-cell lymphoma (CTCL) and 8 cases of healthy control specimens were obtained in GSE59307 (**Figure 1A**), whereas the GSE58445 and GSE19069 comprises 193 and 137 samples of PTCLs, respectively. In addition to this, 66 cases of ENKTL which is a subtype of PTCLs were included in GSE90597 and GSE53798 comprise by 26 cases of Diffuse large B-cell lymphoma.According to current WHO classification, CTCL is a subtype of PTCLs; therefore, GSE59307 was chosen to construct the co-expression network. The packages of 'simpleaffy',[18] 'affy-PLM' and 'arrayQualityMetrics' were utilized to perform the process of quality assessment (QA), quality control (QC), background correction and normalization. The probe id in datasets which based on GPL570 platform was annotated by the 'hgu133Plus2' package, and probe id of GSE90597 was annotated by GPL10739 files.

2.2 | Co-expression network construction

The top 5000 variant of expression profiles in GSE59307 were used to construct a co-expression network by using the package of WGCNA, and the network topology was analysed with soft-threshold power from 1 to 30. After determining the optimal beta value for the soft threshold parameter, the relational matrix can be converted into adjacent matrix, and then, it can be transformed into topological overlap matrix (TOM). Finally, average linkage hierarchical clustering was conducted to classify the highly correlated into modules according to the measure of TOM-based dissimilarity measure.
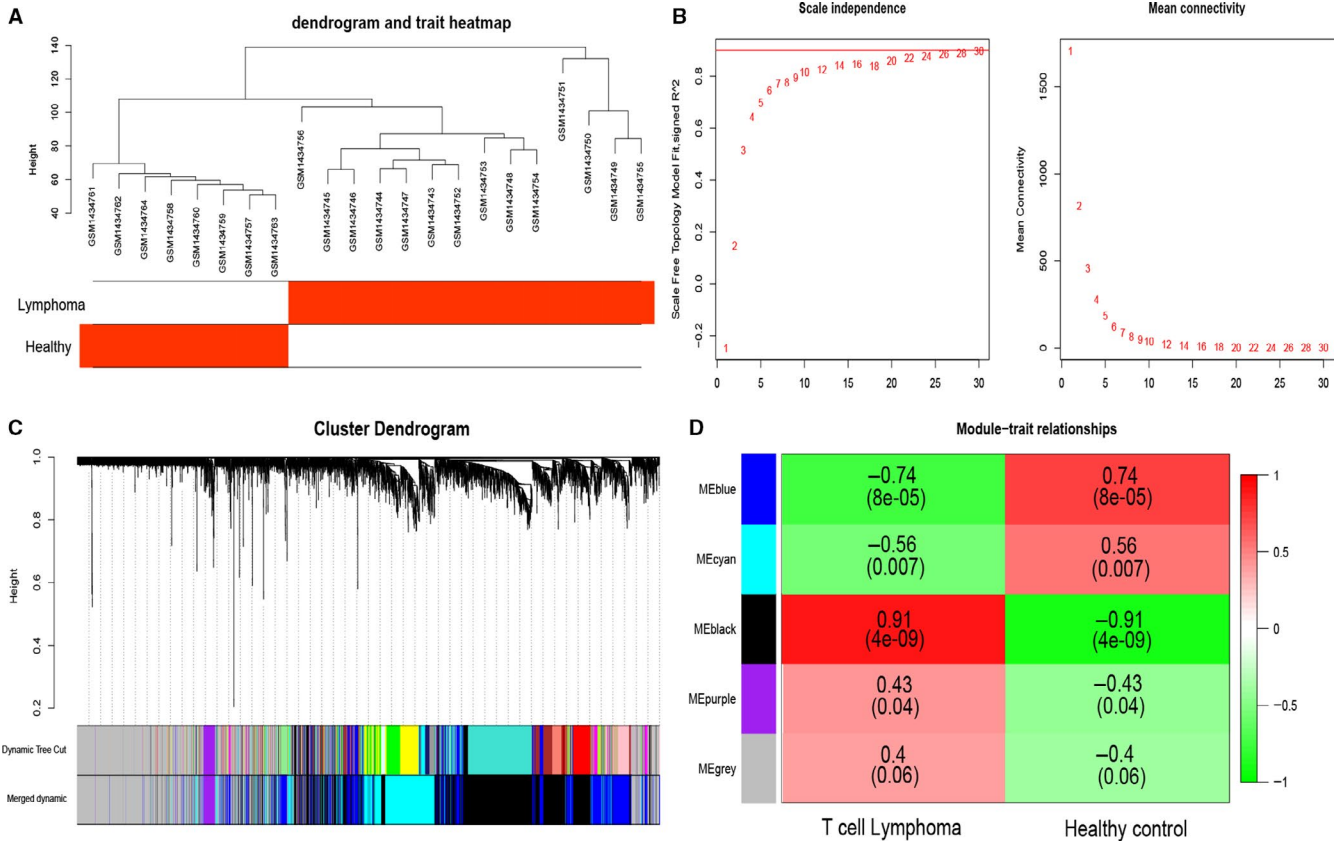
FIGURE 1 Identification of candidate genes in PTCLs. A, Clustering dendrogram of PTCLs and normal control. B, analysis of scale-free fit for soft-thresholding powers and 28 were selected as the best value. C, Dendrogram of genes clustered on a dissimilarity measure. D, Heat map of the relationships between modules and PTCLs by Pearson correlation

2.3 | Clinically significant modules visualization and identity the hub genes

To identify the modules which are significantly related to PTCLs, the module eigengene (ME) was used to characterize the expression profiles of each module and the correlation between PTCLs. The relationship of each genes with PTCLs was measured by gene significance (GS) value. Module significance (MS) represents the average GS of all genes that in the module. Finally, the module that highly related to PTCLs was chosen for further analysis. The connectivity of genes in module was quantified by the absolute value of the Pearson's correlation, and the module membership (MM) was defined as the correlation of module eigengenes (MEs) with genes. In present study, the intramodular hub genes were chosen with the criterion of GS > 0.2 and MM > 0.8 to ensure the reliability of the results.

2.4 | LASSO Cox Regression conduction and identification of a mRNA signature

Univariate Cox proportional hazards regression analysis was applied to assess the relationship between the expression of WGCNA-selected genes and the overall survival (OS) of patients with PTCLs, genes which calculated with $P < .05$ were sorted out and chosen

to screening the most valuable predictable mRNAs by performing the LASSO Cox regression analysis which depends on the R package 'glmnet'. The optimal values of the penalty parameter λ were estimated through 10-time cross-validations. The risk score of mRNA signature for each patient was calculated by the coefficient that from LASSO regression analysis and expression level of each mRNA. The risk score was constructed as follows:

$$\text{Risk score} = \sum_{a=1}^{n} \exp_a * \beta_a$$

n was the number of prognostic genes, $\exp_a$ was the expression value of gene a, and β was the regression coefficient. All PTCLs patients were separated into high- and low-risk groups according to median risk score that used as cut-off value. Kaplan-Meier estimator was carried out to assess the prognostic value of the mRNA signature. Survival prediction based on the risk score was illustrated by using the 'survivalROC' package. Wilcoxon signed-rank was applied to compare the differential expression between high-risk group and low-risk group of PTCLs. In addition, the protein expression levels of the six genes in the mRNA signature were validated by immunohistochemistry through using the Human Protein Atlas database (https://www.proteinatlas.org/) and their mutation status was explored by cbioportal database (https://www.cbiop ortal.org/).

2.5 | Immune landscape difference between high-risk and low-risk PTCLs patients

CIBERSORT is a tool based on the principle of linear support vector regression to deconvolute the expression matrix and calculate the proportions of 22 types of infiltrating immune cells in each sample.[19] And only if P value of each sample less than 0.05 will be retained for subsequent analysis. Considering the small sample size of GSE90597, we combined the two datasets of GSE19069 and GSE90597 and used the combat function of the R software package of sva to remove batch effects and calculate the distribution of immune cells.[20] Then, we analysed the immune differences between the high-risk and low-risk groups in the combined dataset and the GSE58445 dataset.

2.6 | Integrated analysis by combining the clinical factors and mRNA signature

To investigate the effect of the risk signature on the prognosis of PTCLs patients, univariate and multivariate Cox regression analyses were conducted. The risk scores of six-mRNA signature and other clinical characteristics, including gender and age, were used as covariates. Moreover, the six mRNA which screened by LASSO Cox regression also were selected as candidate mRNAs to explore the difference in survival between high and low expression groups which performed by Kaplan-Meier survival analysis. Furthermore, the analysis concerning the correlation between risk score and currently available clinical characteristics was conducted.

2.7 | Association of a mRNA signature and response to chemotherapy

The dataset of GSE53798 was originally established for predicting sensitivity to chemotherapy drugs in CHOP (cyclophosphamide, doxorubicin hydrochloride, vincristine, prednisone) regime for diffuse large B-cell lymphoma cell.[21] Considering that the degree of malignancy of diffuse large B-cell lymphoma is comparable to that of some subtype of peripheral T-cell lymphoma, and currently treatment of peripheral T-cell lymphoma is similarly to which used for diffuse large B-cell lymphoma (DLBCL). So, in this part, we investigated whether the mRNA signature could predict patients' responses to chemotherapy of vincristine.

2.8 | Nomogram development and validation

The Cox regression model was used to perform the multivariable survival analysis and build nomograms. Calibration curves were selected to assess the consistency between the actual survival and the predicted survival for the nomogram. Nomogram and calibration curves were performed with the package named rms. The C-index was utilized to measure the discrimination of the nomogram.

3 | RESULTS

3.1 | Pre-processing of the datasets

All microarray data were converted into expression matrix after processing, 31 cases in GSE58445 and 11 cases in GSE90597 which lacking survival data were excluded in this study. In addition, after excluding unqualified samples, 162 patients in GSE58445 were randomly divided into the training set ($n = 98$) and testing set ($n = 64$) according to a ratio of 6:4.

3.2 | Construction of weighted co-expression network and identification of key modules

To ensure build a scale-free network, the power of $\beta = 28$ (scale-free $R2 = 0.84$) was selected as the best soft-thresholding parameter (Figure 1B). Next, co-expression modules were produced by method of dynamic tree cutting and make sure that the number of genes in each module is not <30 (Figure 1C). Additionally, by setting the parameter of MEDissThresas as 0.25, the modules that closely associated were merged into a larger one. Ultimately, there are five modules were generated in co-expression network, and black module demonstrated the strongest positive correlation with PTCLs samples (weighted correlation = 0.91, $P = 4e -9$) (Figure 1D).

3.3 | Identification of the six-mRNA signature in training group patients

All 799 WGCNA-selected hub genes used to identify survival-related mRNA by univariable Cox survival analysis in training group dataset, 15 genes were pre-filtered based on P values $< .05$, and then, those genes were selected to preform LASSO Cox regression analysis in GSE58445 cohort (Figure S1). The risk score for predicting the outcome of patients was calculated with the following formula which based on the six mRNA: risk score = (0.2554 × DOCK2 expression) + (0.2334 × GSTM1 expression) + (0.3123 × H2AFY expression) + (0.1719 × KCNAB2 expression) + (−0.2820 × LAPTM5 expression) + (−0.1399 × SYK expression). According to the median of the risk score, all PTCLs patients were divided into high-risk (n = 49) and low-risk groups (n = 49). 5-year os was 12.2% for the high-risk group and 32.6% for the low-risk group, which were significantly different in terms of overall survival(OS) ([HR] :5.6, 95% CI 2.75-11.6, $P < .0001$).The 1-year, 2-year, 3-year, 4-year and 5-year areas under the curve were 0.793, 0.831, 0.778, 0.753 and 0.753, respectively (Figure 2B). Additionally, the mRNA signature can function as a novel indicator of the survival of PTCLs patients, which was confirmed by Kaplan-Meier curves (Figure 2A). Among these six mRNA, DOCK2, GSTM1, H2AFY and KCNAB2 significantly overexpressed in high-risk PTCLs patients compare to high-risk group and were associated with poor prognosis; LAPTM5 and SYK significantly

overexpressed in low-risk patients compare to high-risk patients and related to prolonged prognosis(Figure 2C).

3.4 | Validation of prognostic and predictive accuracy of the six-mRNA signature in the internal and external testing group

The prognostic value of six-mRNA signature was further evaluated in the internal test set and external testing set. In the internal testing cohort, the PTCLs were categorized 32 (50%) of 64 patients into the low-risk group and 32 patients (50%) into the high-risk group, and 5-year os was 9.37% for the high-risk group and 25% for the low-risk group, which were significantly different in terms of overall survival(OS) ([HR] :2.4, 95% CI 1.2-4.8, $P < .01$. Figure 2D). In the external testing cohort, the ENKTL which is a subtype of PTCLs

was categorized 27 (49.09%) of 55 patients into the low-risk group and 28 patients (50.91%) into the high-risk group, and 5-year os was 7.4% for the high-risk group and 28.57% for the low-risk group, ([HR] 2.3, 95% CI 1.10-4.7, $P = .02$. Figure 2G). We also noted similar results in the total set of GSE58445, and 5-year os was 11.1% for the high-risk group and 29.6% for the low-risk group ([HR]: 3.3, 95% CI 2.2-5.0, $P < .0001$ Figure S2). Prognostic accuracy of the six-mRNA based signature is also assessed by time-dependent ROC analysis. The 1-year, 2-year, 3-year, 4-year and 5-year areas under the curve for internal testing group were 0.655, 0.672, 0.663, 0.731 and 0.701 Figure 2E). Similarly, the 1-year, 2-year, 3-year, 4-year and 5-year areas under the curve for external testing cohort set were 0.656, 0.612, 0.656, 0.656 and 0.656, respectively (Figure 2H).

In the internal verification cohort, the expression distribution of 6 genes between the high and low two groups is consistent with the test group (Figure 2F). In the external verification cohort, DOCK2
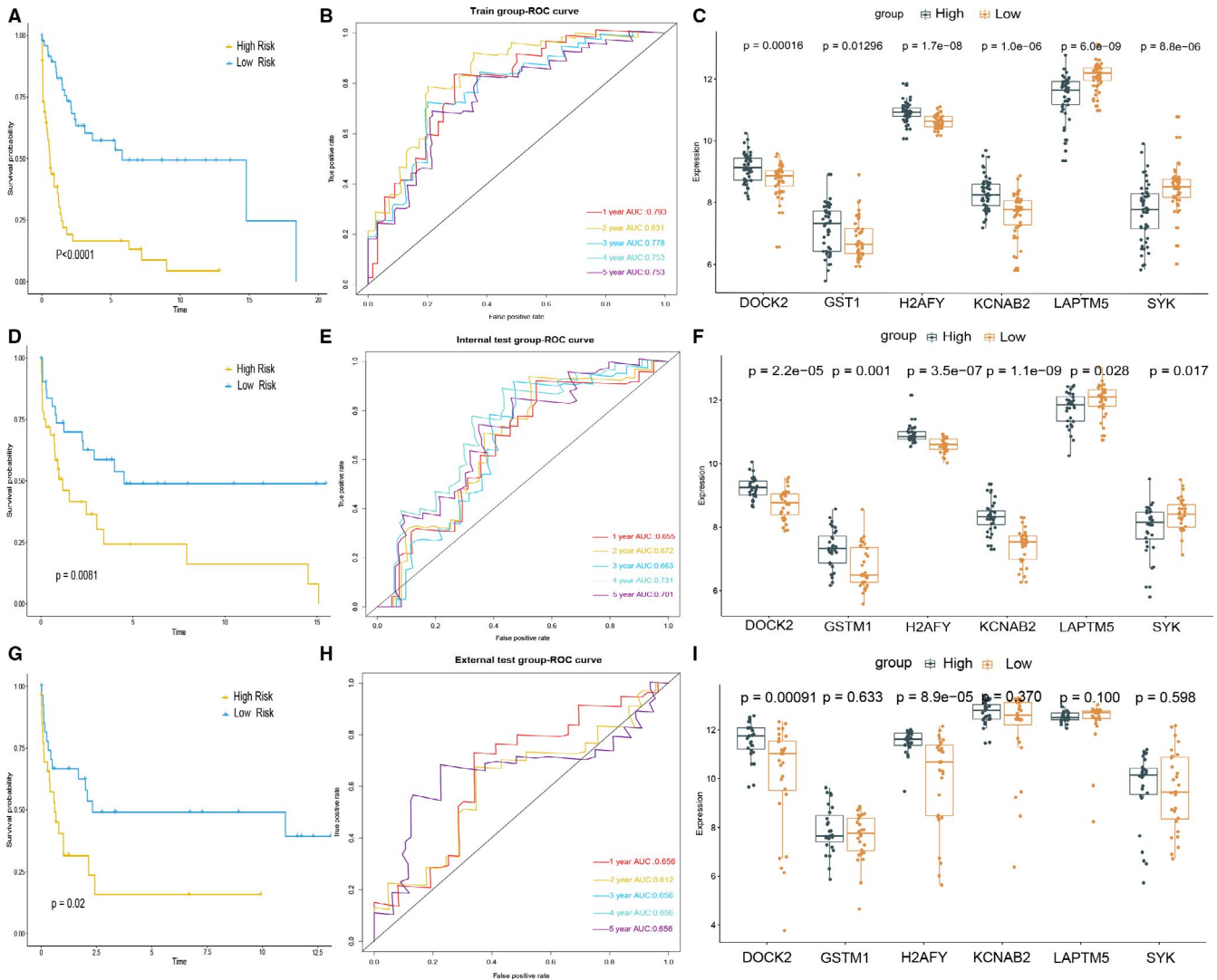

FIGURE 2 Prognostic and predictive value of the mRNA signature. A,D,G, Kaplan-Meier survival curves for training group, internal testing group and external testing cohort of PTCLs patients. B,E,H, Time-dependent ROC curves of 1, 2, 3, 4, 5 years for the six-mRNA signature in training group, testing group and all cohort. C,F,I, Box plot visualization of the expression levels of DOCK2, GSTM1, H2AFY, KCNAB2, LAPTM5 and SYK in different risk group

and H2AFY significantly overexpressed in high-risk PTCLs patients compare to high-risk group and there is no statisticaldifference in the expression level of the remaining four genes between the high and low two groups (Figure 2I). Moreover, we explored the impact of expression of these six mRNA on the prognosis of all PTCLs, and the median expression value of the selected genes was set up as cut-off value in GSE58445, and fond high expression of LAPTM5 and SYK is a protective factor for prognosis of PTCLs; however, high expression of DOCK2, GSTM1, H2AFY and KCNAB2 is a risk factor to prognosis (Figure 3). Furthermore, we analysed the correlation between the expression of six genes and survival in GSE90597, and X-tile software was used to find the best cut-off value. We found that the high expression of GSTM1, H2AFY and KCNAB2 is negatively correlated with the prognosis, and LAPTM5 and SYK are positively correlated with the prognosis, which in line with the result in dataset of GSE58445 (Figure S3).

3.5 | External validation of the protein expression levels and genetic alteration of the six mRNA

Protein expression levels in PTCLs which obtained and visualized by Human Protein Atlas database showed that DOCK, GSTM1, H2AFY and SYK represent medium to high degree positive in immunohistochemistry staining results. However, KCNAB2 shows weak positive and LAPTM5 shows negative staining (Figure 4A). All the results of protein expression levels are basically consistent with genes coefficient in our mRNA results. Among the 43 PTCLs patients enrolled in Cancer Genomics database of cBioportal, all six mRNA have no genetic alterations (Figure 4B).

3.6 | Distribution of immune cells in different risk groups

After completion of CIBERSORT immune analysis, we found that the two cohorts of PTCLs patients generally have similar immune cell distribution; additionally, naive B cells are statistically different in the high-risk and low-risk groups in both two cohorts. In the GSE58445 cohort, 81 cases PTCLs in each of the low-risk group and the high-risk group showed a significant difference in the presence of 6 immune cells types (naive B cells, memory B cells, resting natural killer cells, M1 macrophages, resting mast cells, eosinophils) (Figure 5A). In the combinational cohort which incorporated by GSE19069 and GSE90597 dataset, 99 cases PTCLs in the low-risk group and 92 PTCLs cases in the high-risk group showed a significant difference in distribution of 4 immune cells types (naive B cells, activated CD4 + T cells, activated NK cells and resting dendritic cells) (Figure 5B).

3.7 | Independent prognostic role of the mRNA signature

To confirm the value of mRNA signature in assessing PTCLs patients' prognosis, we performed univariate and multivariate Cox regression analyses in training group and testing dataset by including age, gender and mRNA signature as explanatory variables. Clinical characteristic parameters were grouped according to the International Prognostic Score (IPI) criteria: Age ≥ 60 Years. In training group, gender and mRNA signature were significantly correlated with OS by using univariate Cox regression. After multivariate adjustment using the factors above, the mRNA signature remained a powerful and
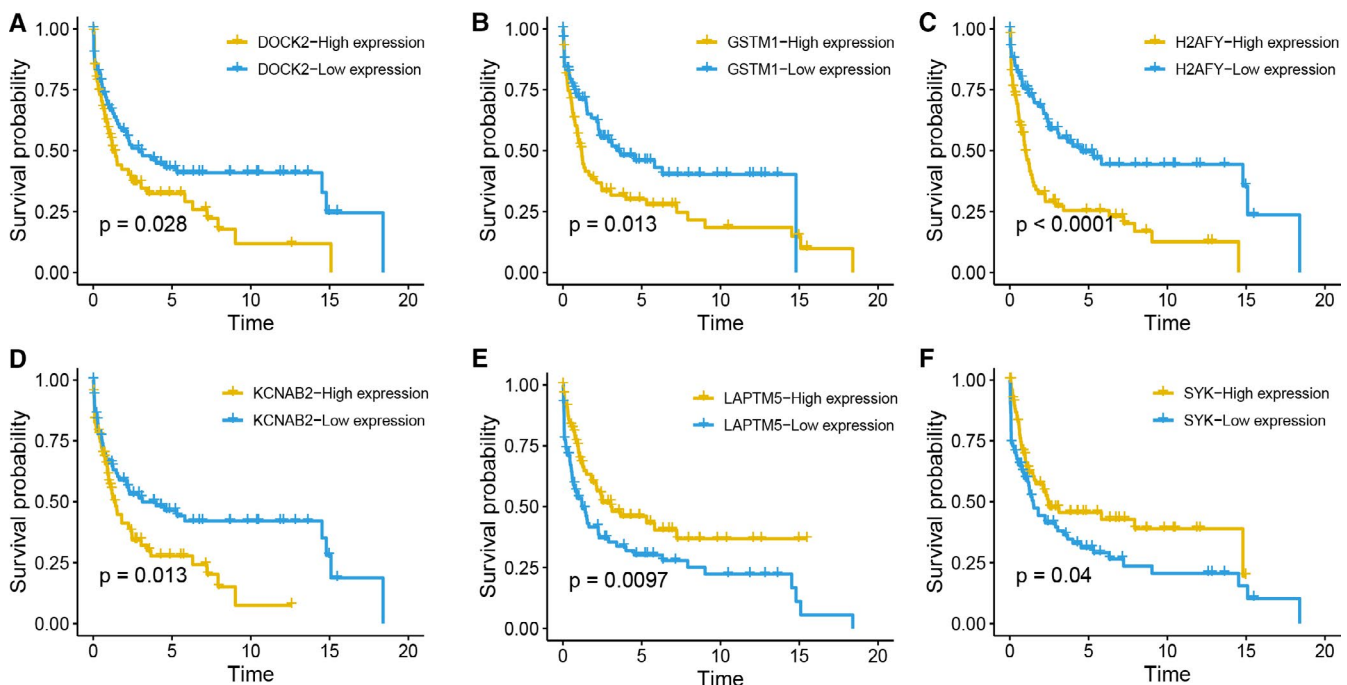

FIGURE 3 Survival analysis of DOCK2, GSTM1, H2AFY, KCNAB2, LAPTM5 and SYK in GSE58445 PTCLs cohorts (A, DOCK2; B, GSTM1; C, H2AFY; D, KCNAB2; E, LAPTM5; F, SYK)
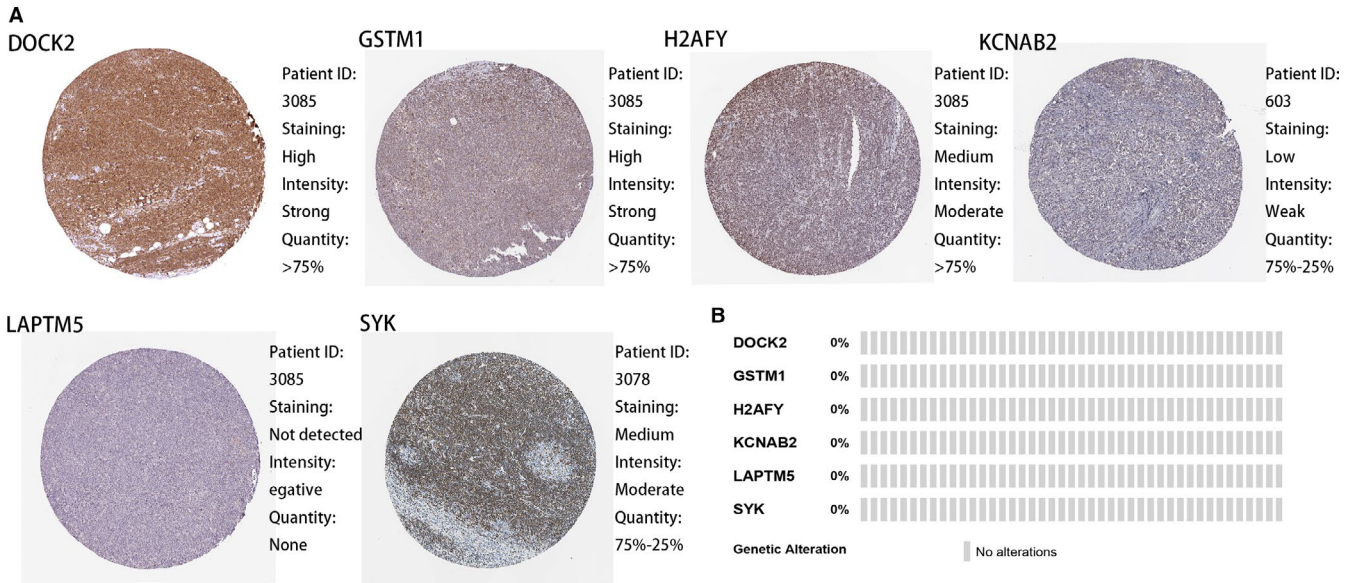
FIGURE 4 Expression and genetic alterations of the four predictive genes. A, The representative protein expression of the six mRNA in PTCLs . *Data were from the Human Protein Atlas (*http://www.proteinatlas.org*) online database. B, Genetic alterations of the six mRNA in PTCLs*
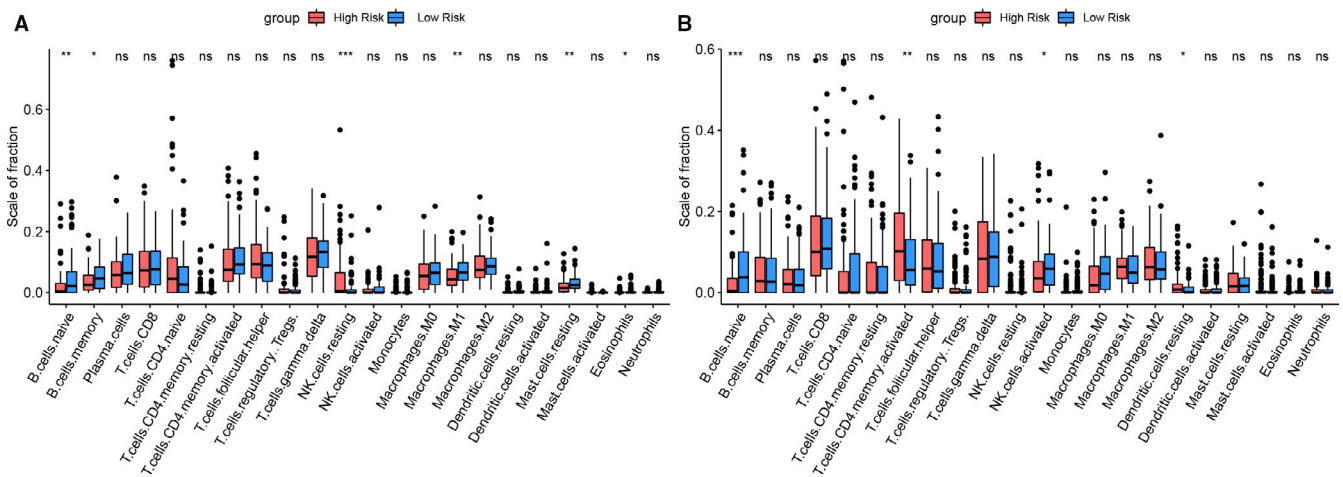

FIGURE 5 Differential distribution of immune cells between the high risk and low risk group. A, GSE58445 dataset. B, Combinational cohort which incorporated by GSE19069 and GSE90597 dataset

independent prognostic factor for PTCLs patients (mRNA signature: HR = 5.6, 95% CI = 2.75-11.6, $P < .0001$) (Table 1), the same results were also seen in testing group (internal testing cohort: HR = 2.7, 95% CI = 1.31-5.6, $P = .0007$; external testing cohort: HR = 2.64, 95% CI = 2.64, $P = .02$),suggesting that the risk prognostic signature independent impact on prognostic of PTCLs patients.

3.8 | Association of the mRNA risk score with Clinical characteristics and the role of risk stratification on response to chemotherapy

In order to figure out the impact of different PTCLs subtypes and clinical features on risk scores, we analysed the risk differences between samples with different pathological types and different

clinical characteristics according to the available information in the GEO dataset. The risk score of AITL, ALTL, ENKTL and PTCL-NOS in dataset of GSE58445 has no obvious difference (Figure 6A), as the same results of status of gender and age impact on risk score (Figure 6B,C). Compared with other types of lymphoma, adult T-cell leukaemia/lymphoma (ATLL) has the lowest risk score and T-cell leukaemia/lymphoma (T-ALL) has the highest score in dataset of GSE19069(Figure 6D), and there is no statistical difference in risk score between different ages and different genders in GSE19069(Figure 6E,F). Apart from this, we investigated whether the mRNA signature could predict patients' response to chemotherapy in GSE53798. More patients with high risk exhibited resistant to vincristine compare to low-risk patients (Figure 6G), which may partly explain why high-risk PTCLs patients have a worse prognosis than low-risk PTCLs patients.

TABLE 1 Univariate and multivariate Cox regression analyses of the mRNA signature in PTCLs patients

| Variable	Univariate analysis			Multivariate analysis		
	HR	95% CI P	P	HR	95% CI	P
Training group						
Risk (High/Low)	4.30	2.40-7.40	<.001	5.60	2.75-11.6	<.001
Gender (Male/Female)	3.00	1.50-5.90	.0016	1.90	0.88-3.90	.10
Age (≥60/<60 y)	1.20	0.61-2.10	0.65	1.70	0.88-3.30	.118
Internal testing group						
Risk (High/Low)	2.40	1.20-4.80	.01	2.70	1.31-5.60	.007
Gender (Male/Female)	0.69	0.33-1.40	.32	0.82	0.39-1.80	.62
Age (≥60/<60 y)	1.70	0.82-3.60	.15	1.86	0.89-3.9	.10
External testing group						
Risk (High/Low)	2.30	1.10-4.70	.023	2.64	1.16-62	.02
Gender (Male/Female)	1.70	0.75-3.70	.21	1.67	0.70-3.95	.24
Age (≥60/<60 y)	3.00	1.40-6.10	.003	3.15	1.41-7.02	.005
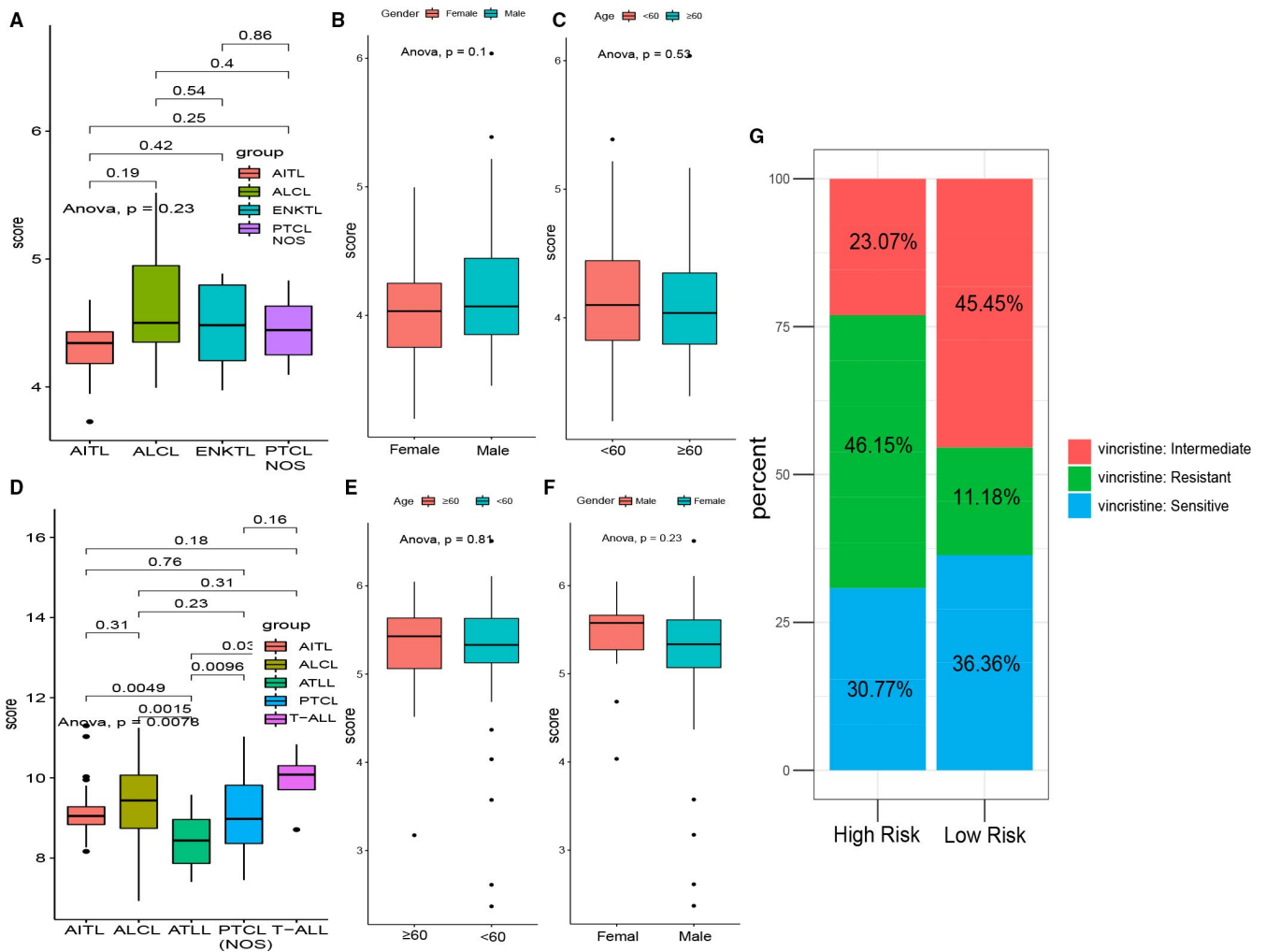
FIGURE 6 Distribution of the mRNA risk score in distinct clinical characteristics and the role of risk stratification on response to chemotherapy. A,D, Differences in risk score among different PTCLs subtypes. B,E, The risk score was group by age. C,F, The risk score was group by gender. G, Risk stratification on response to vincristine chemotherapy

3.9 | Establishment of the nomogram and assessment of predictive value of mRNA signature

In order to develop a convenient clinical tool that could facilitate clinician to predict overall survival (OS) probability of every patient, a nomogram which included a mRNA signature, age and gender was constructed to predict the 1-, 3- and 5-year OS of PTCLs patients (Figure 7A,C).The calibration curve also illustrates high consistency between predictive survival time and observation survival time for the probabilities of 3- and 5-year OS in the PTCLs cohort. In the GSE58445 dataset, the Harrell's concordance index for OS was 0.722 (Figure 7B).In the GSE90597 dataset, the Harrell's concordance index for OS was 0.684 (Figure 7D), it means that the calibration plots for the 3- and 5-year OS rate were estimated well in entire PTCLs patients.

4 | DISCUSSION

Peripheral T-cell lymphoma is an aggressively lymphoproliferative disease that seriously threatens human health, and most patients with PTCLs have a poor prognosis due to the combination of the lack of specific treatment and an aggressive clinical process.[22] However, molecular risk stratification which based on gene expression profile

(GEP) into some type of human cancer has opened an avenue for clinicians to personalized medicine and brought enthusiasm for researchers to applicate to other cancer types.[23] Until recently, PTCLs were lagged behind in terms of risk classification unfortunately. In the present research, we developed a prognostic signature that based on six genes (DOCK2, GSTM1, H2AFY, KCNAB2, LAPTM5 and SYK) for PTCLs and validated it in internal test datasets and external test datasets. Complementary value of clinical characteristics and molecular were further leveraged and showed that combination of both could accurately predict the overall survival of PTCLs.

There is an increasing application of risk signature used for predicting prognosis of cancer patients due to the carcinogenesis and development of tumours are the interaction of multiple genes.[23,24] On the other hand, the risk signature based on multigenes usually shows better performance in predicting the prognostic value than an individual gene or clinical characteristic risk classifier.[25] Therefore, in our study, we built a multi-mRNA-based signature with the LASSO Cox regression model to predict overall survival of PTCLs, and the prognostic and predictive accuracy of this signature was assessed in training group and testing patient groups. By utilizing this mRNA signature to the PTCLs patients, significantly statistical difference was depicted in the survival curve between high-risk group and low-risk group. Compared with wen yin's report of 5 genes signature predicting the survival of Glioblastoma
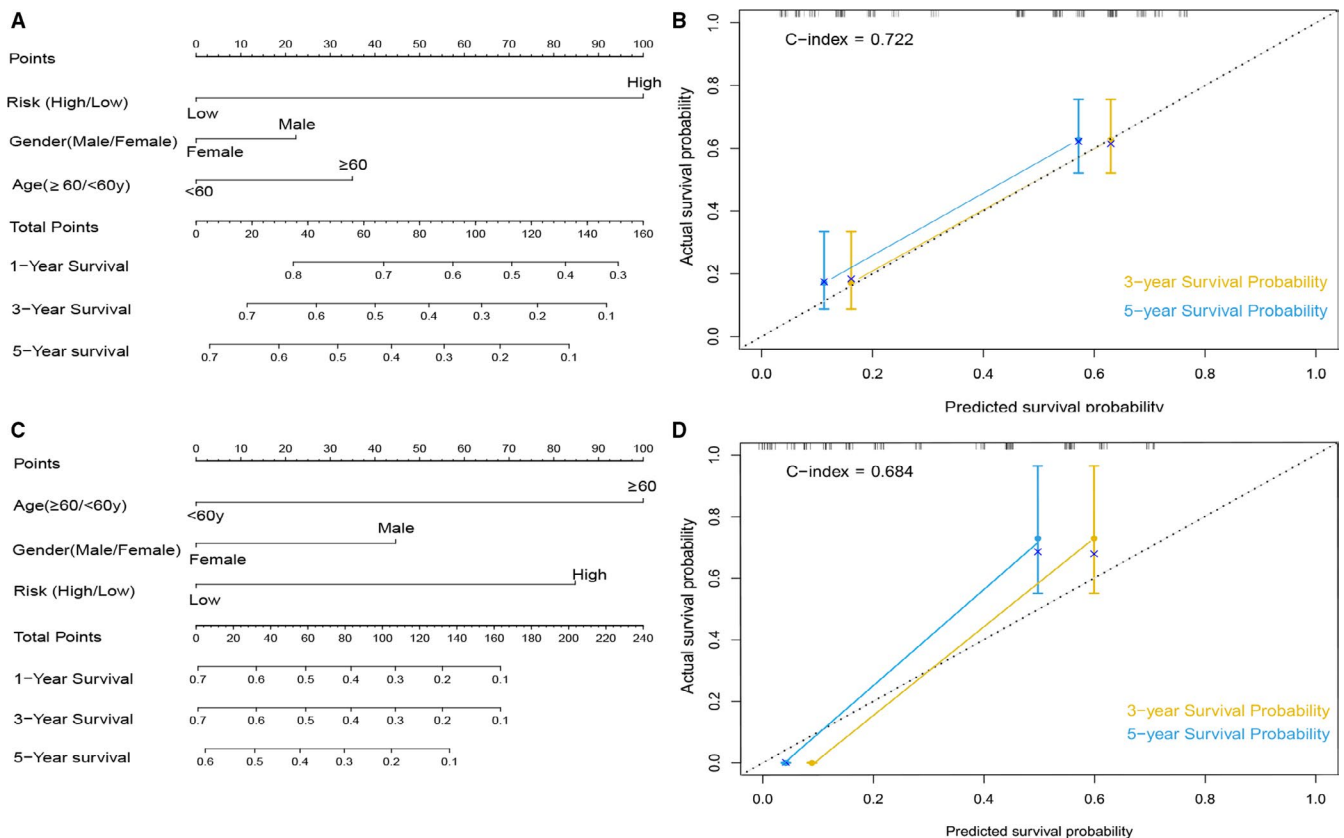

FIGURE 7 Nomogram and calibration plot for GSE58445 cohort and GSE90597 cohort. (A,C) The nomogram was constructed for predicting1, 3, 5-year survival rate of PTCLs patients. (B,D)The calibration curves for predicting patient survival at 3 and 5 years in the cohort

multiforme and jie zhu's research of 6 genes signature discriminating the high-risk and low-risk group of lung cancer,[26,27] our signature classifier show better performance. Additionally, we also demonstrated the predictive value of the mRNA signature for chemotherapy of vincristine in DLBCL and that may helpful for predicting the sensitivity of PTCLs to vincristine. Moreover, to accurately predict the outcome of each individual PTCLs patient, we combine clinical characteristics and 6-mRNA signature to construct nomograms, and we had evaluated the calibration of the nomogram according to the calibration curve. In our study, the c-index for the nomogram in two group patient was 0.722 and 0.684, respectively (3-, 5-year OS), significantly higher than previous research that used to predict the prognosis of non-small-cell lung cancer patients,[28] showing that there is distinguished consistency between predicted survival probability and actual survival proportion, and indicating our nomogram that based on the six genes signature is a promising tool for predicting the outcome of PTCLs patients and can be useful for clinicians to implement personalized treatment.

In this context, KCNAB2, H2AFY, DOCK2 and GSTM1 in our signature significant high expression in high-risk group compare to low-risk group and related to poor prognosis, and all genes except KCNAB2 in signature have been reported to be involved in cancer. Potassium voltage-gated channel subfamily A regulatory beta subunit 2 (KCNAB2) is a component of voltage-dependent potassium channels (KCh) proteins,[29] and it has been documented that KCh proteins play an important role in controlling tumour cell proliferation in the early stages of G1/S transition and even during.[30] Among all potassium channel proteins, some have been considered as promising tumour markers, such as KCNK5, KCNQ1OT1, KCNH2 and KCNN4.[31,32] KCNAB2 in our study significant highly expressed in high-risk group, indicating it may be involved in the disease progression of peripheral T-cell lymphoma. KCNAB2 in our mRNA signature almost never be reported in cancer pathogenesis or disease progression may be due to its tissue specificity.[29] H2AFY is one alternatively exon-spliced isoform of macroH2A, generally expressed in tissues that with active cell proliferation[33] and the overexpression of the H2AFY in tumour sample could further increase aggressiveness of tumour cells and gave rise to metastasis by decreased the expression level of SOD3.[34] It has been reported that H2AFY is significantly augmented in breast cancer cells and hepatocellular carcinoma (HCC) compared with normal control cells,[35] up-regulated H2AFY also related to poor prognosis of breast cancer by driving overexpression of HER-2 and might favour HCC progression through pathway of p38 MAPK.[36,37] DOCK2 as a guanine nucleotide exchange factor (GEF) belongs to the dedicators of cytokinesis (DOCK) family, which originally identified in hematopoietic cell, and now it is also studied in B-cell lymphoma and prostate cancer.[38] DOCK2 has the functions of activating small G proteins such as Rac1/2 and subsequently activates downstream pathways which involved in survival, proliferation and migration of cancer.[39] It has also been demonstrated that DOCK2 was abnormally elevated expressed in B-cell lymphoma

and the overexpressed DOCK2 correlated with the reduced prognosis of chronic lymphocytic leukaemia.[40,41] GSTM1 (glutathione S-transferase M1) is a member of the family of cytosolic GSTs, and the null genotype of GSTM1 has been proven to be associated with risk of colorectal cancer, renal cell carcinoma, oesophageal cancer, nasopharyngeal cancer and bladder cancer.[42-46] LAPTM5 (lysosomal-associated protein transmembrane 5) is a membrane protein that can inhibit the expression of T-cell receptor (TCR) and play a positive role in migration and invasion of ovarian cancer cell but play a negative regulator of T-cell or B-cell receptor downstream signalling.[47-49] SYK (spleen tyrosine kinase) is an important component involved in immune receptor signal transduction and is found to be highly expressed in most PTCLs.[50] Moreover, the inhibitor of SYK was shown to not only inhibit T-cell lymphoma cell lines proliferation but also induce apoptosis.[51] In our study, the prognosis of the SYK high-risk group is better than that of the low-risk group, which may be attributed to the absent expression of SYK in some lymphoma with worse prognosis.[52] But it cannot be ruled out that it has a protective effect in some subtype of PTCLs, because it has been reported that SYK has a protective effect in some solid tumour.[53-55]

Genomic changes have been shown as the cause of carcinogenic and progression of tumours, but in recent analyses infer that the changes in the tumour microenvironment (TME) are also closely related to cancer prognosis and have influence on the response of immunotherapy.[56] The high infiltration of B cells in tumours has been demonstrated to be associate with patients prolonged survival[57] and unique role for B cells in antitumour immunity may be responsible for this phenomenon.[58] To explore the composition of the immune microenvironment of PTCLs, the scale of value of immune cells in the high- and low-risk groups was calculated and analysed. The proportion of naïve B cells is significant higher in the low-risk group than in the high-risk group, which in line with the Javeed Iqbal's research that the signatures of B cell predicted a favourable outcome of PTCLs.[59] In addition to this, the presence of B cell in tumours could promote immunotherapy response,[60] and it suggests that low-risk group PTCLs may be more effective for immunotherapy.

Limitations of the present study should be acknowledged. Firstly, the sample size might not be adequate and may lead to selection bias. Secondly, lack of complete clinical characteristics and absent comprehensive analysis of signature and clinical features. What's more, additional genetic and experimental studies are required to elucidate the mechanism and the function of these genes that are included in signature which in the carcinogenesis and progression of PTCLs. Finally, our results in more larger samples or more external independent datasets need further validation.

5 | CONCLUSION

In conclusion, this is the first study to investigate the ability of mRNA risk signature as novel prognostic biomarkers for PTCLs.

In present research, we identified a six-mRNA based signature for predicting OS of PTCLs and the mRNA signature has showed power performance to stratify all PTCLs patients into low and high risk group. Moreover, A nomogram which integrated mRNA signature and clinical characteristics potentially offers good value for clinicians implementing personalized therapeutic regimen for patients with PTCLs.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTION

Jiannan Tu: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Methodology (supporting); Software (equal); Visualization (equal); Writing-original draft (equal). **Zhixing Kuang:** Data curation (equal); Formal analysis (supporting); Methodology (lead); Project administration (equal); Software (equal); Visualization (equal); Writing-original draft (equal). **Xiaoliang Xie:** Formal analysis (supporting); Resources (supporting); Visualization (equal); Writing-original draft (equal). **Shizhen Wu:** Investigation (supporting); Software (supporting); Visualization (supporting); Writing-original draft (equal). **Ting Wu:** Writing-original draft (equal). **Shengchi Chen:** Conceptualization (supporting); Funding acquisition (lead); Methodology (supporting); Project administration (lead); Resources (supporting); Software (supporting); Writing-review & editing (equal).

DATA AVAILABILITY STATEMENT

All data generated or analysed during this study are included in this article.

ORCID

Zhixing Kuang https://orcid.org/0000-0003-0553-379X
Shengchi Chen https://orcid.org/0000-0001-9659-4744

REFERENCES

1. Elenitoba-Johnson KSJ, Lim MS. New insights into lymphoma pathogenesis. *Annu Rev Pathol*. 2018;13:193-217.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70:7-30.
3. Vose JM. Peripheral T-cell lymphoma: novel backbone. *Blood*. 2018;131:375-376.
4. Heavican TB, Bouska A, Yu J, et al. Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma. *Blood*. 2019;133:1664-1676.
5. Armitage JO, Gascoyne RD, Lunning MA, Cavalli F. Non-Hodgkin lymphoma. *Lancet*. 2017;390:298-310.
6. Zhang Y, Xu W, Liu H, Li J. Therapeutic options in peripheral T cell lymphoma. *J Hematol Oncol*. 2016;9:37.
7. Gleeson M, Peckitt C, To YM, et al. CHOP versus GEM-P in previously untreated patients with peripheral T-cell lymphoma (CHEMO-T): a phase 2, multicentre, randomised, open-label trial. *The Lancet Haematology*. 2018;5:e190–e200.
8. Ito Y, Makita S, Tobinai K. Development of new agents for peripheral T-cell lymphoma. *Expert Opin Biol Ther*. 2019;19:197-209.
9. Laribi K, Alani M, Truong C, Baugier de Materre A. Recent advances in the treatment of peripheral T-cell lymphoma. *Oncologist*. 2018;23:1039-1053.
10. Broccoli A, Zinzani PL. Peripheral T-cell lymphoma, not otherwise specified. *Blood*. 2017;129:1103-1112.
11. Fossard G, Broussais F, Coelho I, et al. Role of up-front autologous stem-cell transplantation in peripheral T-cell lymphoma for patients in response after induction: an analysis of patients from LYSA centers. *Ann Oncol*. 2018;29:715-723.
12. Huh SJ, Oh SY, Lee S, et al. The Glasgow Prognostic Score is a significant predictor of peripheral T-cell lymphoma (PTCL) treated with CHOP-based chemotherapy and comparable with PTCL prognostic scores. *Int J Hematol*. 2019;110:438-446.
13. Dimitrakopoulos C, Vrugt B, Flury R, et al. Identification and validation of a biomarker signature in patients with resectable pancreatic cancer via genome-wide screening for functional genetic variants. *JAMA Surg*. 2019;154:e190484.
14. Criscitiello C, Bayar MA, Curigliano G, et al. A gene signature to predict high tumor-infiltrating lymphocytes after neoadjuvant chemotherapy and outcome in patients with triple-negative breast cancer. *Ann Oncol*. 2018;29:162-169.
15. Hu X, Martinez-Ledesma E, Zheng S, et al. Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. *Neuro Oncol*. 2017;19:786-795.
16. Smyth EC, Nyamundanda G, Cunningham D, et al. A seven-Gene Signature assay improves prognostic risk stratification of perioperative chemotherapy treated gastroesophageal cancer patients from the MAGIC trial. *Ann Oncol*. 2018;29:2356-2362.
17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
18. Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*. 2005;21:3683-3685.
19. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Meth*. 2015;12:453-457.
20. Leek J, Johnson W, Parker H, Jaffe A, Storey J. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England)*. 2012;28:882-883.
21. Falgreen S, Dybkær K, Young KH, et al. Predicting response to multidrug regimens in cancer patients using cell line experiments and regularised regression models. *BMC Cancer*. 2015;15:235.
22. Stephenson R, Singh A. Drug discovery and therapeutic delivery for the treatment of B and T cell tumors. *Adv Drug Deliv Rev*. 2017;114:285-300.
23. Mo Q, Nikolos F, Chen F, et al. Prognostic power of a tumor differentiation gene signature for bladder urothelial carcinomas. *J Natl Cancer Inst*. 2018;110:448-459.
24. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol*. 2017;3:1529-1537.
25. Mitra AP, Lam LL, Ghadessi M, et al. Discovery and validation of novel expression signature for postcystectomy recurrence in high-risk bladder cancer. *J Natl Cancer Inst*. 2014;106.
26. Zhu J, Wang M, Hu D. Development of an autophagy-related gene prognostic signature in lung adenocarcinoma and lung squamous cell carcinoma. *PeerJ*. 2020;8:e8288.

27. Yin W, Tang G, Zhou Q, et al. Expression profile analysis identifies a novel five-gene signature to improve prognosis prediction of glioblastoma. *Front Genet*. 2019;10:419.

28. Wu J, Zhou L, Huang L, et al. Nomogram integrating gene expression signatures with clinicopathological features to predict survival in operable NSCLC: a pooled analysis of 2164 patients. *J Exp Clin Cancer Res*. 2017;36:4.

29. Felipe A, Vicente R, Villalonga N, et al. Potassium channels: new targets in cancer therapy. *Cancer Detect Prev*. 2006;30:375-385.

30. Ko J, Ko E, Gu W, Lim I, Bang H, Zhou T. Expression profiling of ion channel genes predicts clinical outcome in breast cancer. *Mol Cancer*. 2013;12:106.

31. Anderson K, Cormier R, Scott P. Role of ion channels in gastrointestinal cancer. *World J Gastroenterol*. 2019;25:5732-5772.

32. Lastraioli E, Iorio J, Arcangeli A. Ion channel expression as promising cancer biomarker. *Biochem Biophys Acta*. 2015;1848:2685-2702.

33. Sporn J, Jung B. Differential regulation and predictive potential of MacroH2A1 isoforms in colon cancer. *Am J Pathol*. 2012;180:2516-2526.

34. Dardenne E, Pierredon S, Driouch K, et al. Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nat Struct Mol Biol*. 2012;19:1139-1146.

35. Cantariño N, Douet J, Buschbeck M. MacroH2A–an epigenetic regulator of cancer. *Cancer Lett*. 2013;336:247-252.

36. Borghesan M, Fusilli C, Rappa F, et al. DNA hypomethylation and histone variant macroH2A1 synergistically attenuate chemotherapy-induced senescence to promote hepatocellular carcinoma progression. *Can Res*. 2016;76:594-606.

37. Li X, Kuang J, Shen YI, et al. The atypical histone macroH2A1.2 interacts with HER-2 protein in cancer cells. *J Biol Chem.*. 2012;287:23171-23183.

38. Wang L, Nishihara H, Kimura T, et al. DOCK2 regulates cell proliferation through Rac and ERK activation in B cell lymphoma. *Biochem Biophys Res Commun*. 2010;395:111-115.

39. Steele AJ. Roring ahead with DOCK2. *Blood*. 2018;132:115-116.

40. Hasan MK, Yu J, Widhopf GF, et al. Wnt5a induces ROR1 to recruit DOCK2 to activate Rac1/2 in chronic lymphocytic leukemia. *Blood*. 2018;132:170-178.

41. Chen Y, Meng F, Wang B, He L, Liu Y, Liu Z. Dock2 in the development of inflammation and cancer. *Eur J Immunol*. 2018;48:915-922.

42. Albarakati N, Khayyat D, Dallol A, Al-Maghrabi J, Nedjadi T. The prognostic impact of GSTM1/GSTP1 genetic variants in bladder Cancer. *BMC Cancer*. 2019;19:991.

43. Li J, Xu W, Liu F, Huang S, He M. GSTM1 polymorphism contribute to colorectal cancer in Asian populations: a prospective meta-analysis. *Sci Rep*. 2015;5:12514.

44. Li Y, Wan W, Li T, Cao J, Xu G. GSTM1 null genotype may be associated with an increased nasopharyngeal cancer risk in South China: an updated meta-analysis and review. *Onco Targets Ther*. 2015;8:2479-2484.

45. Huang W, Shi H, Hou Q, Mo Z, Xie X. GSTM1 and GSTT1 polymorphisms contribute to renal cell carcinoma risk: evidence from an updated meta-analysis. *Sci Rep*. 2015;5:17971.

46. Lu QJ, Bo YC, Zhao Y, et al. Glutathione S-transferase M1 polymorphism and esophageal cancer risk: An updated meta-analysis based on 37 studies. *World J Gastroenterol*. 2016;22:1911-1918.

47. Nuylan M, Kawano T, Inazawa J, Inoue J. Down-regulation of LAPTM5 in human cancer cells. *Oncotarget*. 2016;7:28320-28328.

48. Ouchida R, Kurosaki T, Wang JY. A role for lysosomal-associated protein transmembrane 5 in the negative regulation of surface B cell receptor levels and B cell activation. *J Immunol*. 2010;185:294-301.

49. Gao Y, Chen Q, Yue W. LAPTM5 protein can regulate TGF-β mediated MAPK and smad signaling pathways in ovarian cancer cell. *Ann Oncol*. 2019;30:v9.

50. Liu D, Mamorska-Dyga A. Syk inhibitors in clinical development for hematological malignancies. *J Hematol Oncol*. 2017;10:145.

51. Wilcox RA, Sun DX, Novak A, Dogan A, Ansell SM, Feldman AL. Inhibition of Syk protein tyrosine kinase induces apoptosis and blocks proliferation in T-cell non-Hodgkin's lymphoma cell lines. *Leukemia*. 2010;24:229-232.

52. Bisig B, Gaulard P, de Leval L. New biomarkers in T-cell lymphomas. *Best Pract Res Clin Haematol*. 2012;25:13-28.

53. Peng C, Sun Q, Hao Y, Cong B, Zhao Y, Zhao X. Syk is low-expressed in non-small-cell lung cancer and inversely correlates with patient's survival. *Acta Biochim Biophys Sin (Shanghai)*. 2013;45:149-151.

54. Krisenko MO, Geahlen RL. Calling in SYK: SYK's dual role as a tumor promoter and tumor suppressor in cancer. *Biochim Biophys Acta*. 2015;1853:254-263.

55. Fueyo J, Alonso MM, Parker Kerrigan BC, Gomez-Manzano C. Linking inflammation and cancer: the unexpected SYK world. *Neuro Oncol*. 2018;20:582-583.

56. Zeng D, Li M, Zhou R, et al. Tumor microenvironment characterization in gastric cancer identifies prognostic and immunotherapeutically relevant gene signatures. *Cancer Immunol Res*. 2019;7:737-750.

57. Petitprez F, de Reyniès A, Keung EZ, et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature*. 2020;577:556-560.

58. Bruno TC. New predictors for immunotherapy responses sharpen our view of the tumour microenvironment. *Nature*. 2020;577:474-476.

59. Iqbal J, Wright G, Wang C, et al. Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. *Blood*. 2014;123:2915-2923.

60. Helmink BA, Reddy SM, Gao J, et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature*. 2020;577:549-555.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.