

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

Jiannan Tu

Department of Oncology, Nanping First Hospital Affiliated to Fujian Medical University

Zhixing Kuang

Department of Radiation Oncology, Nanping First Hospital Affiliated to Fujian Medical University

Xiaoliang Xie

Department of Orthopedics, Shanghai municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine

Shizhen Wu

Department of Oncology, Nanping First Hospital Affiliated to Fujian Medical University

Ting Wu

Department of Oncology, Nanping First Hospital Affiliated to Fujian Medical University

Shengchi Chen (✉ npscchen@gmail.com)

Department of Oncology, Nanping First Hospital Affiliated to Fujian Medical University

Research

Keywords: peripheral T-cell lymphomas, Prognostic, predictive, signature

Posted Date: March 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-16089/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Current international prognostic index is widely questioned on the risk stratification of peripheral T-cell lymphoma and do not accurately predict the outcome for patients. We postulated that multiple mRNAs could combined into a single model to improve risk stratification and to guide Clinicians implementing personalized therapeutic regimen for these patients. **Methods:** The gene expression profiles with clinical characteristics were selected and 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. Then build a gene classifier using a Lasso Cox Regression model and validated the prognostic accuracy of this mRNA signature in an internal validation cohort. Finally, a prognostic nomogram was constructed and performance was assessed by calibration plot and the concordance index (C-index). **Results:** 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, $p<0.0001$; internal testing set :hazard ratio [HR] 2.4, 95% CI 1.2–4.8, $p<0.01$).Multivariate regression demonstrated that the signature was an independently prognostic factor contrast to age and gender. Furthermore, receiver operating characteristic analysis indicated that this signature exhibited excellent diagnostic efficiency for overall survival. Moreover, the nomogram which combined the six-genes risk signature and multiple clinical factors suggesting that predicted survival probability agreed well with the actual survival probability. **Conclusions:** 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 stage PTCLs.

Background

Non-hodgkin lymphomas are clonal neoplasms that arise from lymphocyte at various stages of maturation[1], it estimated that 77240 new cases of non-Hodgkin lymphoma are expected in the United States, and 19940 patents will die for this disease in 2020[2]. Peripheral T cell lymphomas (PTCLs) as a subgroup of non-Hodgkin lymphomas which also characterized as a infrequency and heterogeneous aggressive behavior 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 (PTCL) comprises more than 30 distinct histologic subtypes including anaplastic lymphoma kinase (ALK)-positive anaplastic large cell lymphoma (ALCL) and ALK-negative ALCL, angioimmunoblastic T cell lymphoma(AITL), PTCL, not otherwise specified (PTCL-NOS) according 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, while 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 unfavorable 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 gene 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 was carried out to identify a gene signature which beyond clinical parameters and significant associated with PTCLs prognosis. Finally, a prognostic nomogram was established based the combination of signature and clinical characteristics.

Materials And Methods

Data sources and data processing

The raw data of GSE59307 and GSE58445 which based on the GPL570 platform [HG-U133_Plus_2] were downloaded from the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) database, 14 samples of Cutaneous T cell Lymphoma (CTCL) and 9 cases of healthy controls pecimens were obtained in GSE59307. While, the GSE58445 comprises 193 samples of peripheral T-cell lymphomas (PTCLs). According to current WHO classification, Cutaneous T cell Lymphoma (CTCL) is a subtype of PTCLs, therefore, GSE159307 were chosen to construct the co-expression Network. The packages of

“simpleaffy”[18], “affyPLM”, “arrayQualityMetrics”[19] and “hgu133Plus2” were utilized to perform the process of quality assessment (QA), quality control (QC), background correction, normalization and probe annotation.

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 analyzed 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.

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 were measured by Gene Significance (GS) value. Module significance (MS) represent 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.

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 $P < 0.05$ were sorted out and chosen to screening the most valuable predictable mRNAs by performing the LASSO Cox regression analysis which depend on the R package “glmnet”. The optimal values of the penalty parameter λ were estimated through 10-times 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:

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

n was the number of prognostic genes, exp_a 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 were illustrated by using the “survivalROC” package. In addition, Wilcoxon Signed-Rank was applied to compare the differential expression between High risk group and Low risk group of PTCLs

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 four genes which screened by lasso cox regression also were selected as candidate mRNAs to performed by Kaplan-Meier survival analysis, and the median expression value of the selected genes were set up as cut off value.

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.

Results

Pre-processing of the data sets

All microarray data were converted into expression matrix after processing and 31 cases of PTCLs which lacking survival data included in GSE58445 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.

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 $R^2 = 0.84$) was selected as the best soft-thresholding parameter(Fig. 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 less than 30(Fig. 1C). Additionally, by setting the parameter of MEDissThresas as 0.25, the modules that closely associated were merge into a larger one. Ultimately, there are 5 modules were generated in co-expression network, and black module demonstrated the strongest positive correlation with PTCLs samples (weighted correlation = 0.91, $P = 4e^{-9}$) (Fig. 1D).

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 prefiltered based on P values < 0.05, and then those genes were selected to preform Lasso Cox Regression analysis in GSE58445 cohort(Fig. 2).. The risk score for predicting the outcome of patients was calculated with the following formula which based on the six mRNA: risk score = $(0.2554 \times \text{DOCK2 expression}) + (0.2334 \times \text{GSTM1 expression}) + (0.3123 \times \text{H2AFY expression}) + (0.1719 \times \text{KCNAB2 expression}) + (-0.2820 \times \text{LAPTM5 expression}) + (-0.1399 \times \text{SYK expression})$. according to the median of the risk score, all PTCLs patients were divided into high-risk (n = 49) and low-risk group(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, $p < 0.0001$). 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 (Fig. 3B). Additionally, the mRNA signature can function as a novel indicator of the survival of PTCLs patients, which was confirmed by Kaplan-Meier curves (Fig. 3A). Among these six mRNA, R DOCK2, GSTM1, H2AFY, 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(Fig. 3C).

Validation of prognostic and predictive accuracy of the six-mRNA signature in the testing group and the total set group

The prognostic value of six-mRNA signature was further evaluated in the internal test set and the total set. In the testing cohort, the PTCLs categorised 32 (50%) of 64 patients into the low risk group and 32 patients (50%) into the high-risk group, 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, $p < 0.01$. Figure 3D). We also noted similar results in the total set, all the PTCLs classified 81 (50%) of 192 patients into the low risk group and 81 patients (50%) into the high-risk group, 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, $p < 0.0001$ Fig. 3G). We also assessed the prognostic accuracy of the six-mRNA based signature with time-dependent ROC analysis. The 1-year, 2-year, 3-year 4-year and 5-year areas under the curve for testing group were 0.655, 0.672, 0.663, 0.731 and 0.701 Fig. 3E), Similarly, the 1-year, 2-year, 3-year 4-year and 5-year areas under the curve for total set were 0.738, 0.768, 0.736, 0.742 and 0.728, respectively (Fig. 3H). moreover, we explored the impact of expression of these six mRNA on the prognosis of all PTCLs, and found high expression of LAPTM5 and SYK is a protective factor for prognosis of PTCLs, however, high expression of DOCK2, GSTM1, H2AFY, KCNAB2 is a risk factor to prognosis(Fig. 4).

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 independent prognostic factor for PTCLs patients (mRNA signature: HR = 5.6, 95% CI = 2.75–11.6, $P < 0.0001$) (Table 1), the same results was also be seen in testing group (mRNA signature: HR = 2.7, 95% CI = 1.31–5.6, $P = .0007$), suggesting that the risk prognostic signature independent impact on prognostic of 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	HR	95%CI	P
Training group						
Risk (High/Low)	4.3	2.40–7.40	< 0.001	5.60	2.75–11.6	< 0.001
Gender(Male/Female)	3	1.50–5.90	0.0016	1.90	0.88–3.90	0.10
Age (≥ 60 / <60 y)	1.2	0.61–2.10	0.65	1.70	0.88–3.30	0.118
Testing group						
Risk (High/Low)	2.4	1.20–4.80	0.01	2.70	1.31–5.60	0.007
Gender(Male/Female)	0.69	0.33–1.40	0.32	0.82	0.39–1.80	0.62
Age (≥ 60 / <60 y)	1.7	0.82–3.60	0.15	1.86	0.89–3.9	0.10

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, gender was constructed to predict the 1, 3 and 5 year OS of PTCLs patients (Fig. 5A,C).The calibration curve also illustrate 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 training set, the Harrell's concordance index for OS were 0.765 (Fig. 5B) ; in the validation set ,the Harrell's concordance index for OS were 0.688(Fig. 5D) ,it means that the calibration plots for the 3-, 5-year OS rate were estimated well in entire PTCLs patients.

Discussion

Peripheral T-cell lymphoma is an aggressively lymphoproliferative disease that seriously threatens human health, most patients with PTCLs have a poor prognosis due to the combination of the lack of specific treatment and an aggressive clinical process[20]. 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[21].until recently, PTCLs was 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 data sets. 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 tumors are the interaction of multiple genes[21, 22]. 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[23]. 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 train group and test in internal 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 multiforme and jie zhu's research of 6 genes signature discriminating the high risk and low risk group of lung cancer[24, 25], our signature classifier show better performance. Additionally, our six-mRNA signature is an independent prognostic factor which has been corroborated by the univariate and multivariate cox analyses. 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 train group and test group patient was 0.765, and 0.688 respectively (3-, 5-year OS),significantly higher than previous research that used to predict the prognosis of non-small cell lung cancer patients[26], 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, all genes except H2AFY and KCNAB2 in signature has been reported to be involved in cancer. DOCK2, GSTM1, H2AFY and KCNAB2 significant high expression in high risk group compare to low risk group and related to poor prognosis.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's also studied in B cell lymphoma and prostate cancer[27]. 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[28].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 leukemia[29, 30]. GSTM1 (glutathione S-transferase M1) is a member of the family of cytosolic GSTs and the null genotype of GSTM1 been proven to be associated with risk of colorectal cancer, renal cell carcinoma, esophageal cancer, nasopharyngeal cancer and bladder cancer[31–35].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 signaling[36–38].SYK(spleen tyrosine kinase) is an important component involved in immune receptor signal transduction and is found to be highly expressed in most PTCLs[39]. Moreover, the inhibitor of SYK was shown to not only inhibit T-cell lymphoma cell lines proliferation but also induce apoptosis[40]. 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[41].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 tumor[42–44].

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. Thirdly, no stratified analysis in all subtype of PTCLs was performed due to the PTCLs case classification data were not available. what's more, additional genetic and experimental studies are required to elucidate the mechanism and the function of these genes that be 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

Conclusion

In conclusion, this is the first study to investigate the ability of mRNA risk signatures as novel prognostic biomarkers for PTCLs. In present research, we identified a six mRNA-based signature for predicting OS of PTCLs and the risk signature has showed power performance to stratify all PTCLs patients into low and high risk. Moreover, A nomogram which integrated risk signature and clinical characteristics potentially offers good value for clinicians implementing personalized therapeutic regimen for patients with PTCLs.

Abbreviations

PTCLs:peripheral T-cell lymphoma; GEO:Gene Expression Omnibus; WGCNA:weighted gene co-expression network analysis (WGCNA); DOCK2:dedicators of cytokinesis 2; GSTM1 glutathione S-transferase M1; LAPTM5:lysosomal-associated protein transmembrane 5; LAPTM5:lysosomal-associated protein transmembrane 5 ; IPI:International Prognostic Score ;ROC:Receiver operating characteristic; OS:Overall survival.

Declarations

ACKNOWLEDGMENTS

This work was funded by Fujian Medical University qihang Fund Project (N0.2018QH1173) and Authors would thank to author for sharing survival information of dataset GSE59307 and GSE58445.

AUTHORS' CONTRIBUTIONS

TJ and KZ: development of methodology, analysis and interpretation of data, XX ,WS and WT writing-original draft. CS: conception and design, and further drafts. All authors reviewed the manuscript and approved the final version to be published.

FUNDING

This work was funded by Fujian Medical University qihang Fund Project (N0.2018QH1173)

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERSETS

The authors declare that they have no competing interests.

AUTHOR DETAILS

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

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

³Department of Orthopedics, Shanghai municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

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**(1):7-30.
3. Vose JM: **Peripheral T-cell lymphoma: novel backbone**. *Blood* 2018, **131**(4):375-376.

4. Heavican TB, Bouska A, Yu J, Lone W, Amador C, Gong Q, Zhang W, Li Y, Dave BJ, Nairismagi ML *et al*: **Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma.** *Blood* 2019, **133**(15):1664-1676.
5. Armitage JO, Gascoyne RD, Lunning MA, Cavalli F: **Non-Hodgkin lymphoma.** *The Lancet* 2017, **390**(10091):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, Edwards L, Oates J, Wotherspoon A, Attygalle AD, Zerizer I, Sharma B, Chua S *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**(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**(3):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**(9):1039-1053.
10. Broccoli A, Zinzani PL: **Peripheral T-cell lymphoma, not otherwise specified.** *Blood* 2017, **129**(9):1103-1112.
11. Fossard G, Broussais F, Coelho I, Bailly S, Nicolas-Virelizier E, Toussaint E, Lancesseur C, Le Bras F, Willems E, Tchernonog E *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**(3):715-723.
12. Huh SJ, Oh SY, Lee S, Lee JH, Kim SH, Lee GW, Kim SJ, Kim WS, Lee HS, Jo JC *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**(4):438-446.
13. Dimitrakopoulos C, Vrugt B, Flury R, Schraml P, Knippschild U, Wild P, Hoerstrup S, Henne-Bruns D, Wuerl P, Graf 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**(6):e190484.
14. Criscitiello C, Bayar MA, Curigliano G, Symmans FW, Desmedt C, Bonnefoi H, Sinn B, Pruneri G, Vicier C, Pierga JY *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**(1):162-169.
15. Hu X, Martinez-Ledesma E, Zheng S, Kim H, Barthel F, Jiang T, Hess KR, Verhaak RGW: **Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma.** *Neuro Oncol* 2017, **19**(6):786-795.
16. Smyth EC, Nyamundanda G, Cunningham D, Fontana E, Ragulan C, Tan IB, Lin SJ, Wotherspoon A, Nankivell M, Fassan M *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**(12):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**(18):3683-3685.
 19. Kauffmann A, Gentleman R, Huber W: **arrayQualityMetrics—a bioconductor package for quality assessment of microarray data.** *Bioinformatics* 2009, **25**(3):415-416.
 20. 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.
 21. Mo Q, Nikolos F, Chen F, Tramel Z, Lee YC, Hayashi K, Xiao J, Shen J, Chan KS: **Prognostic Power of a Tumor Differentiation Gene Signature for Bladder Urothelial Carcinomas.** *J Natl Cancer Inst* 2018, **110**(5):448-459.
 22. 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**(11):1529-1537.
 23. Mitra AP, Lam LL, Ghadessi M, Erho N, Vergara IA, Alshalalfa M, Buerki C, Haddad Z, Sierocinski T, Triche TJ *et al*: **Discovery and validation of novel expression signature for postcystectomy recurrence in high-risk bladder cancer.** *J Natl Cancer Inst* 2014, **106**(11).
 24. 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.
 25. Yin W, Tang G, Zhou Q, Cao Y, Li H, Fu X, Wu Z, Jiang X: **Expression Profile Analysis Identifies a Novel Five-Gene Signature to Improve Prognosis Prediction of Glioblastoma.** *Front Genet* 2019, **10**:419.
 26. Wu J, Zhou L, Huang L, Gu J, Li S, Liu B, Feng J, Zhou Y: **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**(1):4.
 27. Wang L, Nishihara H, Kimura T, Kato Y, Tanino M, Nishio M, Obara M, Endo T, Koike T, Tanaka S: **DOCK2 regulates cell proliferation through Rac and ERK activation in B cell lymphoma.** *Biochem Biophys Res Commun* 2010, **395**(1):111-115.
 28. Steele AJ: **Roring ahead with DOCK2.** *Blood* 2018, **132**(2):115-116.
 29. Hasan MK, Yu J, Widhopf GF, 2nd, Rassenti LZ, Chen L, Shen Z, Briggs SP, Neuberg DS, Kipps TJ: **Wnt5a induces ROR1 to recruit DOCK2 to activate Rac1/2 in chronic lymphocytic leukemia.** *Blood* 2018, **132**(2):170-178.
 30. 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**(6):915-922.
 31. 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**(1):991.

32. 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.
33. 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.
34. 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.
35. Lu QJ, Bo YC, Zhao Y, Zhao EJ, Sapa WB, Yao MJ, Duan DD, Zhu YW, Lu WQ, Yuan L: **Glutathione S-transferase M1 polymorphism and esophageal cancer risk: An updated meta-analysis based on 37 studies.** *World J Gastroenterol* 2016, **22**(5):1911-1918.
36. Nuylan M, Kawano T, Inazawa J, Inoue J: **Down-regulation of LPTM5 in human cancer cells.** *Oncotarget* 2016, **7**(19):28320-28328.
37. 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**(1):294-301.
38. Gao Y, Chen Q, Yue W: **LPTM5 protein can regulate TGF- β mediated MAPK and smad signaling pathways in ovarian cancer cell.** *Annals of Oncology* 2019, **30**:v9.
39. Liu D, Mamorska-Dyga A: **Syk inhibitors in clinical development for hematological malignancies.** *J Hematol Oncol* 2017, **10**(1):145.
40. 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**(1):229-232.
41. Bisig B, Gaulard P, de Leval L: **New biomarkers in T-cell lymphomas.** *Best Pract Res Clin Haematol* 2012, **25**(1):13-28.
42. 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**(2):149-151.
43. 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**(1):254-263.
44. Fueyo J, Alonso MM, Parker Kerrigan BC, Gomez-Manzano C: **Linking inflammation and cancer: the unexpected SYK world.** *Neuro Oncol* 2018, **20**(5):582-583.

Figures

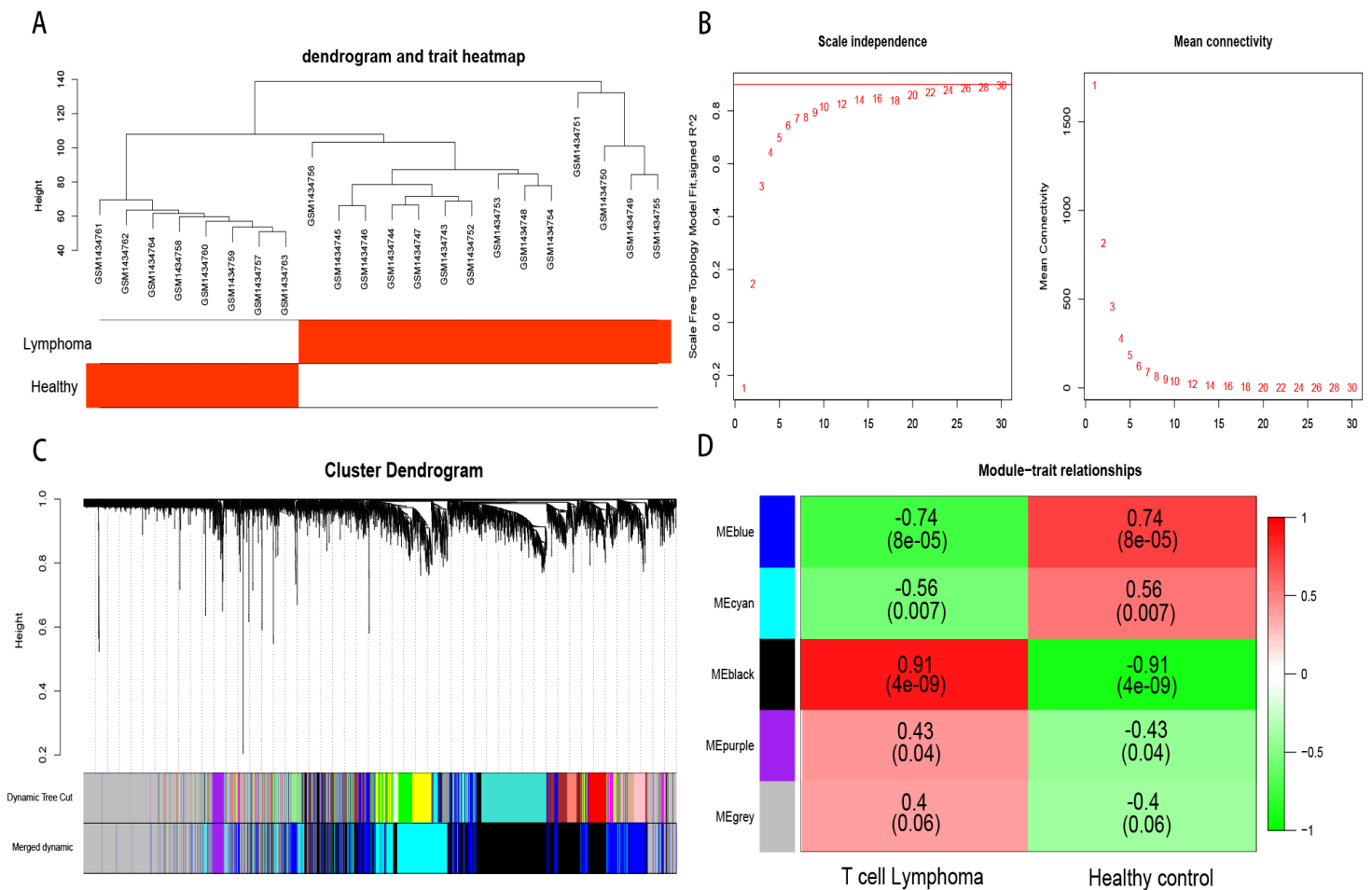


Figure 1

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 was selected as the best value; (C) Dendrogram of genes clustered on a dissimilarity measure; (D) Heatmap of the relationships between modules and PTCLs by Pearson correlation.

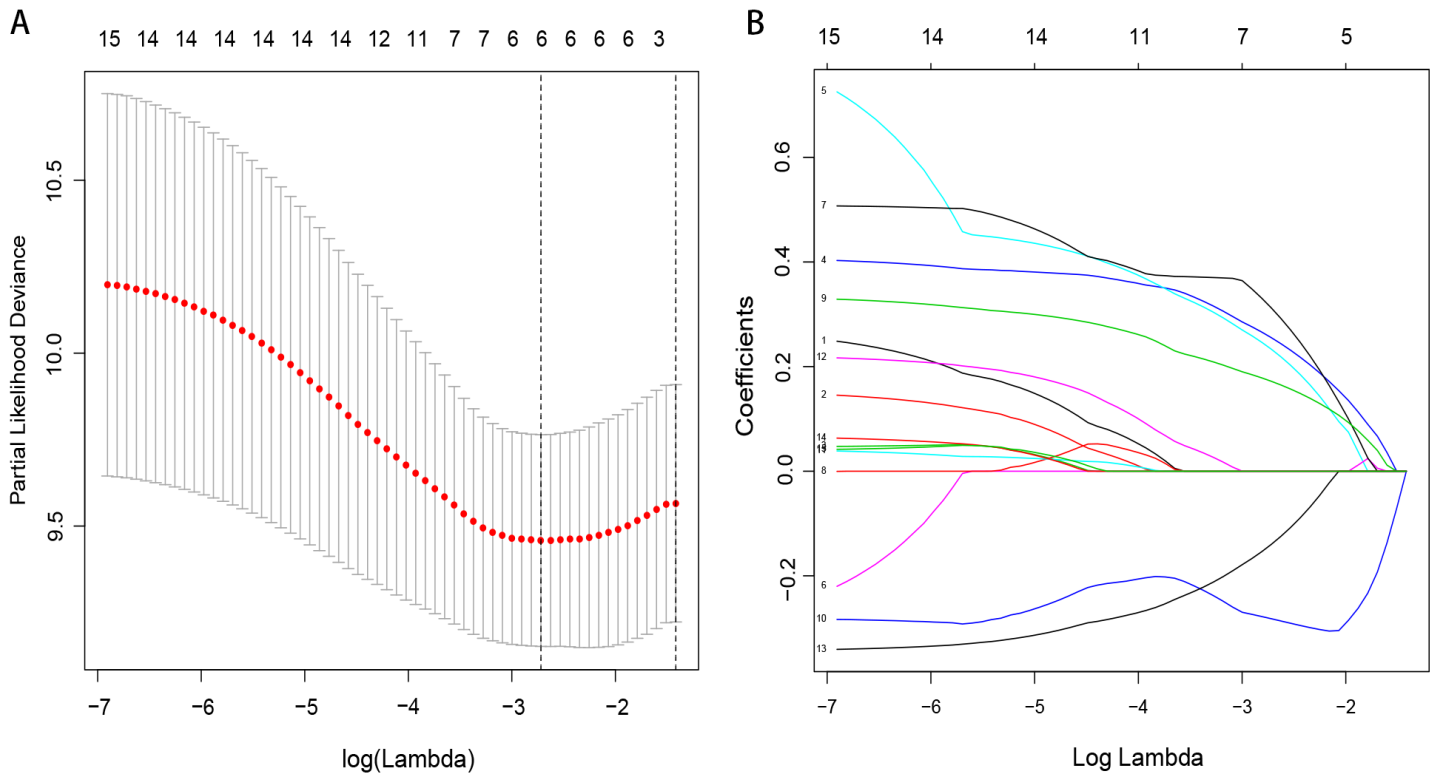


Figure 2

Constructing the six mRNA signature by Lasso cox regression model. (A) Lasso coefficient of the 15 uncox selected genes ;(B) Ten-fold cross-validation for tuning parameter selection in the Lasso module

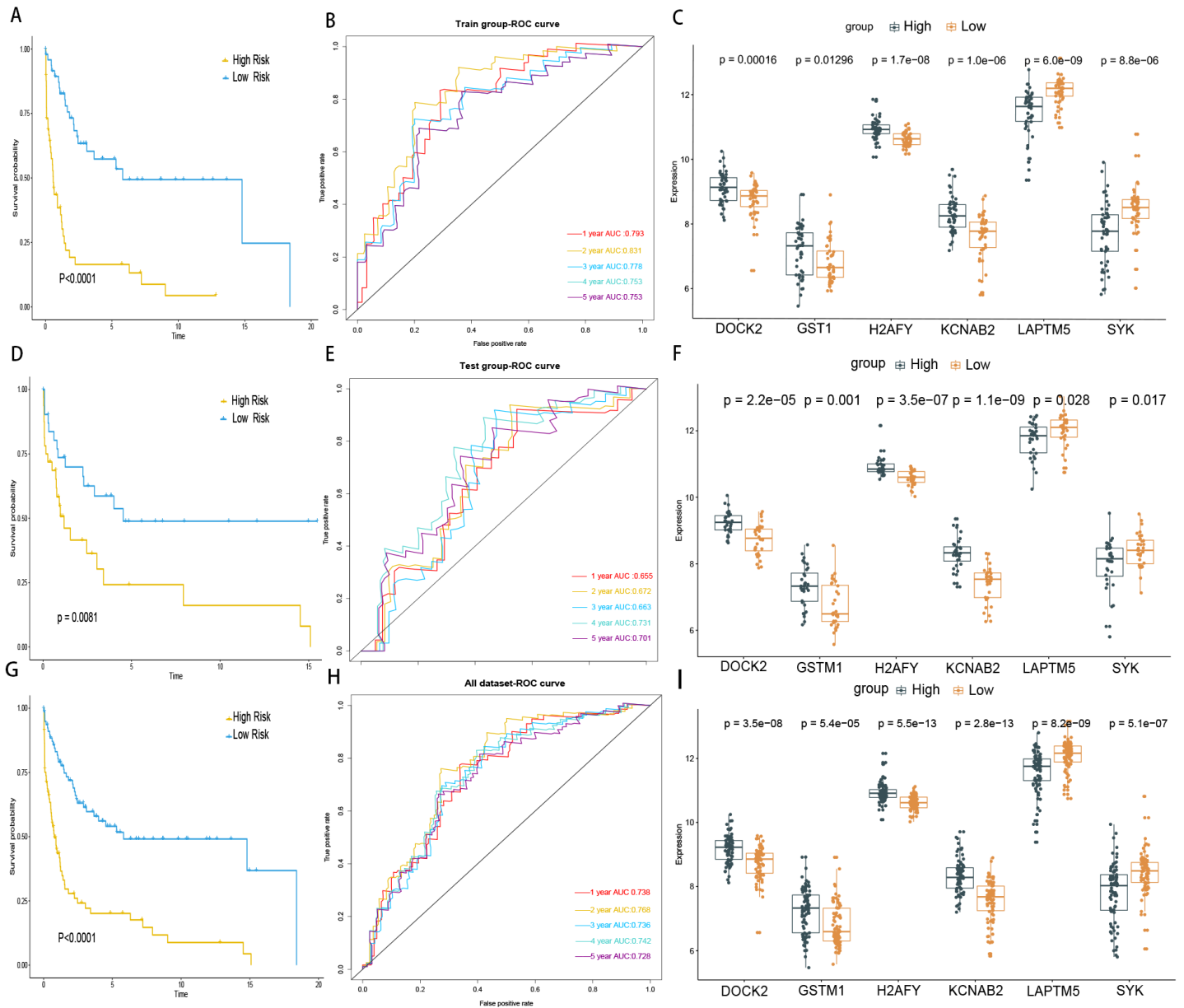


Figure 3

Prognostic and predictive value of the mRNA signature. (A,D,G) Kaplan-Meier survival curves for training group, testing group and all 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.

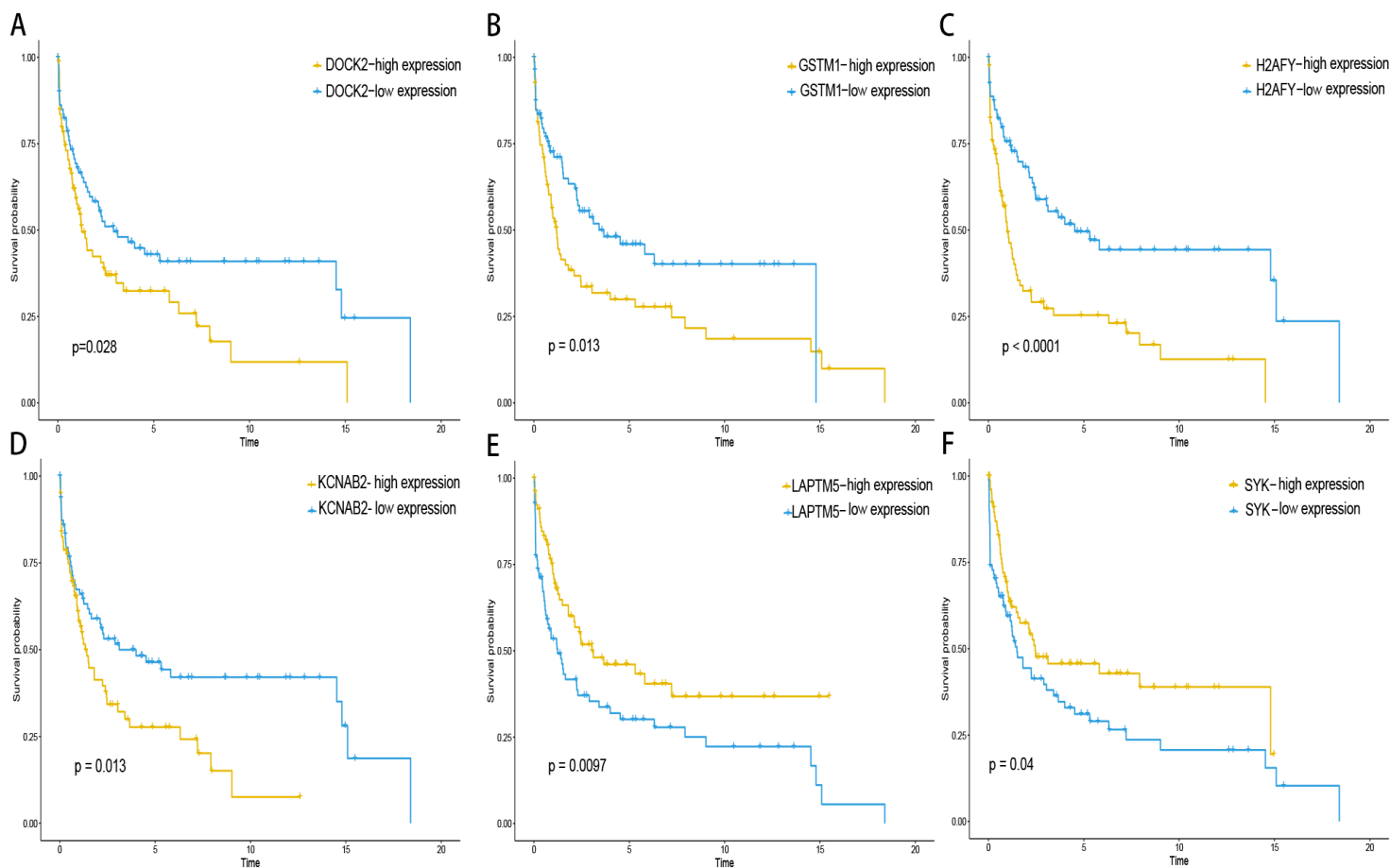


Figure 4

Survival analysis of DOCK2, GSTM1, H2AFY, KCNAB2, LPTM5 and SYK in all PTCLs cohorts(A: DOCK2; B:GSTM1, C:H2AFY; D:KCNAB2; E: LPTM5, F: SYK)

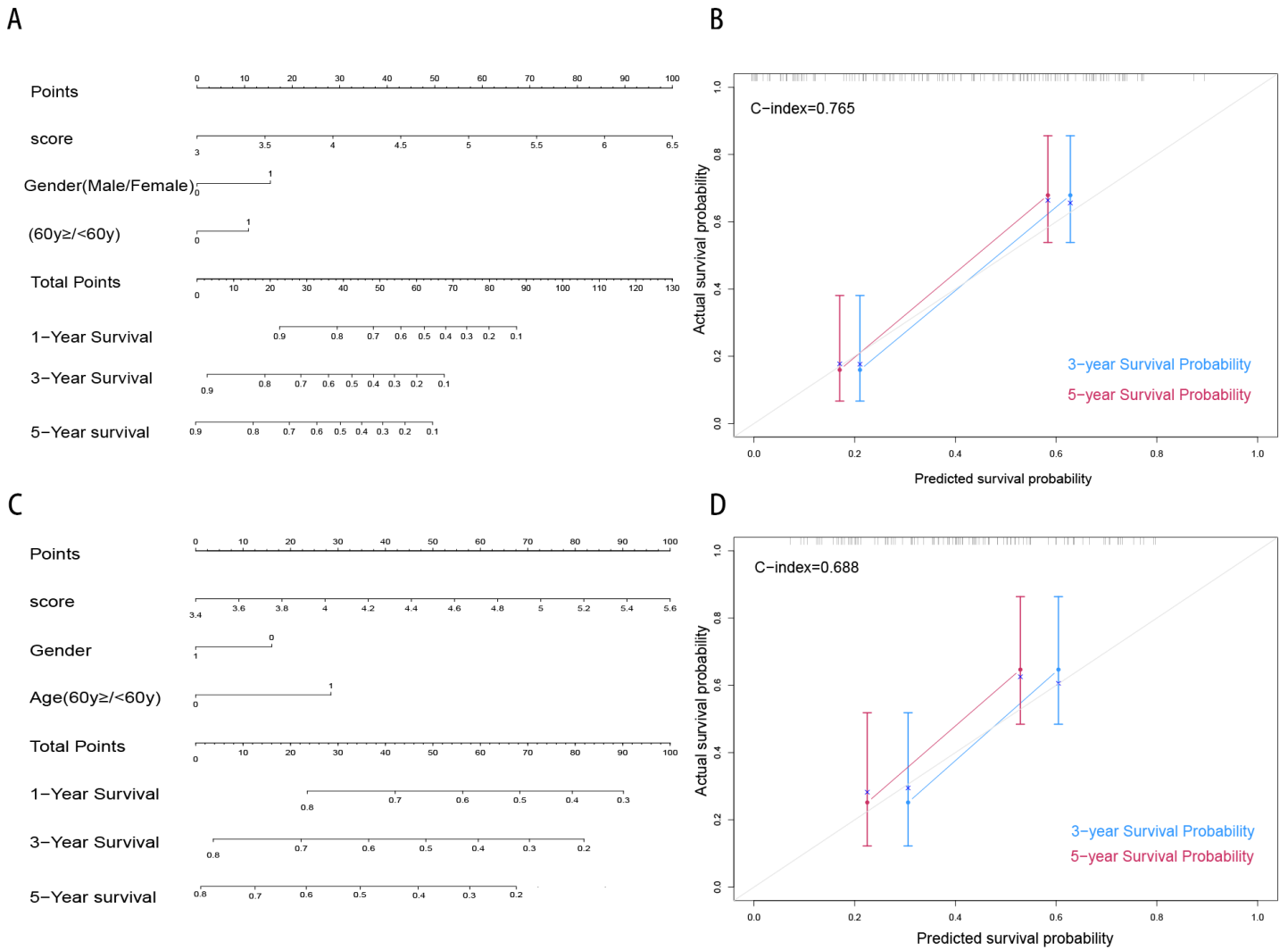


Figure 5

Nomogram and calibration plot for training group and testing group. (A,C) The nomogram was constructed for predicting 1, 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.