Original paper

Folia Neuropathologica

The neutrophil extracellular traps-related gene signature predicts the prognosis of glioblastoma multiforme

Guanghui Sun[1,2], Wei Liu[3]

[1]Weifang Medical University, Weifang, Shandong, China, [2]Brain Intensive Care Unit, Sunshine Union Hospital, Weifang, Shandong, China, [3]Department of Neurosurgery, Affiliated Hospital of Weifang Medical University, Weifang, Shandong, China

Folia Neuropathol 2023; 61: 1-17 *DOI: https://doi.org/10.5114/fn.2023.132980*

Abstract

Introduction: *This research hoped to explore the molecular mechanism of neutrophil extracellular traps (NETs) on glioblastoma multiforme (GBM) progression, and develop a promising prognostic signature for GBM based on NETs-related genes (NETGs).*

Material and methods: *Gene expression data and clinical data of GBM tumour samples were downloaded from TCGA and CGGA databases. NETs-related molecular subtypes were explored by using ConsensusClusterPlus. The NETGs with a prognostic value were identified, and then a prognostic model was constructed using LASSO Cox regression. The predicted performance of the prognostic model was evaluated using TCGA training and CGGA validation cohorts. Moreover, independent prognostic indicators were identified by univariate and multivariate analysis to generate the nomogram model. The sensitivities for antitumor drugs and immunotherapy were predicted. Finally, hub genes in the prognostic model were validated using qPCR analysis.*

Results: *GBM patients were divided into two molecular subtypes with significant differences in tumour microenvironment (TME) score, survival, and immune infiltration. A NETGs signature was constructed based on seven genes (CPPED1, F3, G0S2, MME, MMP9, MAPK1, and MPO), which had a high value for predicting prognosis. A nomogram was constructed by two independent prognostic factors (age and risk score), which could be used to predict 1-, 2-, and 3-year survival probability of GBM. Patients in the high-risk group were more sensitive to bicalutamide, gefitinib, and dasatinib; patients in the low-risk group were associated with poor response to immunotherapy. The validation of the six genes in the prognostic model was consistent with the results of bioinformatics analysis.*

Conclusions: *The NETs-based prognostic model and nomogram proposed in this study are promising prognostic prediction tools for GBM, which may provide new ideas for the development of precise tumour targeted therapy.*

Key words: *glioblastoma multiforme, neutrophil extracellular traps, molecular subtypes, prognostic model, biomarkers, survival analysis.*

Introduction

Glioblastoma multiforme (GBM) is commonly knowns as a human aggressive cancer that takes up almost 15% of all brain tumours [30]. Statistically, the survival for patients with GBM is about 12 to 15 months, with its five-year survival rate of less than 5% [47]. Chemotherapy or radiation treatment followed by the clinical surgery are standard treatment strategies and procedures for patients with GBM [28]. Unfortunately, due to the limited effect of drugs or therapy, prevention of GBM recurrence remains a challenge. [11].

Neutrophils are a vital member of the innate immune system, and one of their defence mechanisms, neutro-

Communicating author:

Wei Liu, Department of Neurosurgery, Affiliated Hospital of Weifang Medical University, 2428 Yuhe Road, Kuiwen District, Weifang, Shandong, China, phone: 86-0536-3081100, e-mail: liuwdandy@163.com

phil extracellular traps (NETs), is a hot topic of the current research [31]. NETs are network structures consisting of DNA histones and proteins released by activated neutrophils that can trap bacteria, fungi, protozoa and viruses [27]. In recent years, a large number of animal models and tumour patient data have shown that NETs are involved in tumour progression and metastasis [1,7]. The prognostic value of NETs has also been revealed in various human diseases including glioma [50]. In tumour microenvironment (TME), tumour associated neutrophils recruited by tumour cells are closely related to the formation of NETs [19]. NETs usually play an antibacterial role, but excessive NETs can damage surrounding tissues by increasing the proinflammatory response, promote tumour angiogenesis and cancer-related thrombosis [20]. During this process, certain NETs-related genes (NETGs) have proved to play a crucial role. For example, a previous study shows that NETGs CCDC25 senses extracellular DNA to promote the tumour cell motility [44]. Deng *et al*. indicated that DDR1 overexpression contributed to the formation of NETs, which further led to the cancer cell invasion and metastasis [8]. Although neutrophil is associated with GBM process via systemic inflammation, the literature analysing the mechanism of NETs in GBM is still scarce [25]. Thus, recognition of the unique altered pattern of NETs in GBM will enhance the understanding of tumour development and provide implications for innovation in GBM treatment and prognostic strategies.

In this study, the gene expression data and clinical information data of GBM tumour samples were downloaded from TCGA database, and GBM patients were classified based on the NETGs expression profiles. The clinical characteristics, TME, immune cells, and GSEA enrichment were compared between subtypes. Next, a prognostic signature was constructed using the TCGA training set and then its predictive performance was validated in the TCGA and CGGA cohorts. The correlation between prognostic signature and immune features were further explored. Moreover, independent prognostic indicators were identified to develop a nomogram model. The sensitivities for antitumor drugs and immunotherapy between different risk groups were predicted. Finally, based on three GBM cell lines, the gene expression of certain signature genes obtained in this study were validated using qPCR analysis. The flow chart of this research is displayed in Figure 1. We hope to reveal the potential molecular mechanism of NETs on GBM progression, and investigate a promising prognostic model of GBM based on NETGs.

Material and methods

Microarray data and pre-processing

The RNA-seq data (log2(fpkm+1)) of TCGA-GBM including 168 tumour and five paracancerous tissue samples were obtained from the UCSC Xene database [13]. At the same time, the clinical data of these samples, including age, sex, Karnofsky performance score, shortest dimension, longest dimension, and disease-free status, were also downloaded. A total of 167 GBM tissue samples with prognostic information were finally enrolled for the follow-up study. In addition, the RNA-seq (fpkm) data and corresponding clinical prognosis information of GBM in mRNAseq_693 datasets were downloaded from the CGGA database [55] as the validation data. According to the clinical information of Grade, WHO IV samples and samples with OS prognosis information were selected. Finally, totally 237 GBM tissue samples were enrolled for subsequent analysis.

NETs subtypes exploration and rationality validation

A total of 69 NETGs were obtained from a previous study provided by Zhang *et al*. [54]. After matching with the gene expression matrix of TCGA-GBM, the expres-

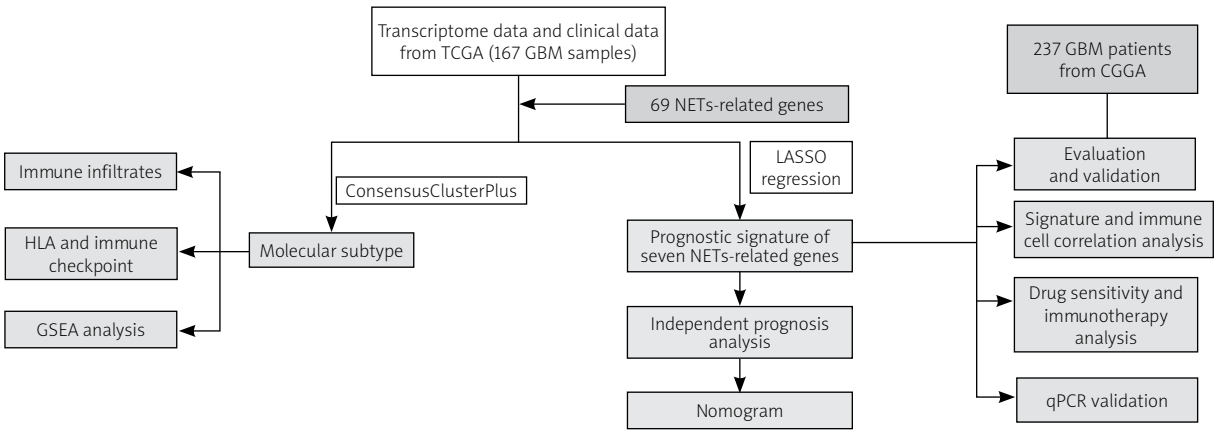

Fig. 1. The flowchart of the study design in this work.

sion values of NETGs in 167 cancer patients with prognostic information were obtained. Based on the expression of NETGs, the ConsensusClusterPlus package [53] in R software (version: 3.6.1) was applied for the NETs subtypes investigation with the following parameters: cluster algorithm = pam; correlation method = Pearson; item subsampling proportion = 0.8; feature subsampling proportion = 1. Then, utilizing ssGSEA algorithm and Gene Set Variation Analysis (GSVA) in R, the enrichment fraction of each immune cell in different samples was calculated to represent the relative abundance of infiltrating cells in samples. Meanwhile, the difference of enrichment score (*p* value) between subtypes was compared by Wilcoxon signed-rank test. The result was visualized by violin plot.

Correlation analysis between prognosis of subtypes and clinical information

To reveal the difference in prognosis between subtypes, the K-M survival investigation was performed by using survival package (version: 3.4) [38]. Then, the relationship between prognosis of NETs subtypes and clinical characteristics was further evaluated. For factor variables, chi square test was used for statistical significance test. For numerical variables including age, Karnofsky performance score, shortest dimension and longest dimension, the significance was revealed by Wilcoxon signed-rank test.

Immune cell infiltration analysis among different subtypes

To reveal the association between NETs subtypes and TME, algorithms including ESTIMATE [46], CIBERSORT [4] and MCPcounter [46] were used to evaluate the immune microenvironment state of different subtypes. Then, the enrichment fraction of each immune cell in different samples was calculated. Finally, the Wilcoxon signed-rank test was used to compare differences in immune cells among subtypes.

Analysis of immune checkpoint genes and human leukocyte antigen genes among subtypes

Based on the expression data in tumour samples, the expression of 15 immune checkpoint genes and 19 human leukocyte antigen (HLA) genes were extracted. Then, their expression difference between NETs subtypes was revealed by using *T* test.

HALLMARK gene set enrichment analysis between subtypes

Based on h.all.v7.4.symbols.gmt enrichment background in MSigDB v7.1 database [22], the enrichment scores of the HALLMARK gene set in each sample were calculated and sorted using GSEA algorithm in clusterProfiler package in R [14] with Benjamini & Hochberg (BH) adjusted $p < 0.05$.

Screening for NETGs with prognostic value

Using survival package (version: 2.41-1) [40] in R software, the univariate Cox regression analysis was performed to identify NETGs that are related to prognosis (prognostic NETGs) based on the expression value of NETGs. $P < 0.05$ was considered as the cut-off value for further analysis.

Establishment and validation of the prognostic NETs-signature model

Based on LASSO Cox regression in R (version: 3.6.1) [39], the optimal gene set was investigated from the prognostic NETGs in the TCGA-GBM training dataset. Briefly, glmnet package (version: 2.0-18) was applied to select the optimal penalty parameter λ associated with 20-fold cross-validation, and then the optimization prognostic model was constructed. Moreover, the risk score (RS) of each patient was calculated using the following formula:

$$RS = \sum \beta_{NETG} \times Exp_{NETG}$$

In this formula, β_{NETG} represents Cox regression prognostic coefficient and Exp_{NETG} represents the expression levels of NETGs in each sample.

According to the prognostic coefficient of signature genes, the RS value of each sample in the TCGA and CGGA was calculated as appropriate. All samples in TCGA and CGGA were divided into High_Risk and Low_Risk sample groups according to the median value of RS. The association between the grouping (High_Risk and Low_Risk) and survival prognostic information was evaluated by using the KM method (version: 2.41-1) in survival package of R. Further, the ROC curves of the 1-, 2-, and 3-year survival in the TCGA and CGGA cohorts were plotted to evaluate the predictive performance of the models.

Correlation analysis between RS and clinical characteristics

The RS under each clinical grouping was visualized by using the box diagram, followed by the calculation of a significant *p* value between the two groups based on Wilcoxon signed-rank test. Spearman correlation coefficient was applied to calculate the correlation between clinical features and RS. Finally, the result was visualized by using scatter plots.

Screening of independent risk factors for prognosis

The univariate and multivariate Cox regression analysis based on age, sex, Karnofsky performance score, shortest dimension, longest dimension, and RS were performed to screen the independent risk factors. $P < 0.05$ was set as the cut-off value. To provide a reference for predicting the prognosis of patients with GBM, nomogram was generated based on the independent prognostic factors. Meanwhile, the prognostic ability of nomogram was evaluated by using the correction curve and ROC analysis.

Correlation analysis between immune cell and signature genes

Based on the relative abundance of infiltrating cell obtained between subtypes, the correlation between immune cell and signature genes was explored. The result was visualized by heatmap according to the Spearman coefficient and *p* value.

Drug sensitivity analysis between two risk groups

To explore the differences in drug sensitivity between High_Risk and Low_Risk groups, the semi-inhibitory concentration (IC_{50}) values of common chemotherapy drugs were calculated by using the pRRophic algorithm according to the GDSC cell lines and TCGA-GBM gene expression profile [12]. Data of the two groups were analyzed by using the Wilcoxon signed-rank test.

Immunotherapeutic response prediction

Tumour immune dysfunction and exclusion (TIDE) analysis can be used to predict immunotherapeutic responses based on two major tumour immune escape mechanisms [17]. In this study, TIDE value of each sample in the High_Risk or Low_Risk groups were evaluated by using the TIDE tool (http://tide.dfci.harvard.edu/), and then analyzed by the Wilcoxon signed-rank test.

Correlation analysis between NETs subtype and different risk groups

The distribution proportion of NETs subtypes in High_ and Low_Risk groups was counted, and the significance was calculated by χ^2 test.

Quantitative real-time PCR (qPCR) verification analysis

To further investigate the expression of prognostic genes (*CPPED1*, *F3*, *G0S2*, *MME*, *MMP9* and *MPO*) revealed in bioinformatics analysis, a verification experiment was performed based on the GBM and normal glial cell lines. Briefly, three GBM cell lines including U251, U87 and T98G were purchased from the Cell Bank of Chinese Academy of Sciences (Shanghai, China). The complete medium used for cell culture was a mixture of Roswell Park Memorial Institute (RPMI) 1640 medium (Hyclone, USA) and foetal bovine serum (Hyclone, USA), with 10% foetal bovine serum. Cells were maintained in an incubator at 37°C with 5% CO_2. Then, total RNAs were extracted from cells in GBM cell lines and normal cells using TRIZOL reagent (Invitrogen, USA), and then reverse transcription was conducted using RevertAid™ First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) in accordance with manufacturers' instructions. The qPCR was performed on an ABI7500 Real-time PCR system (Applied Biosystems, USA). The amplification primers for detected genes are as follows: CPPED1 (forward (F): 5'-CTGGAGGATGAGAAGGCAAC-3'; reverse (R): 5'- CTGGAGGATGAGAAGGCAAC-3'); F3 (F: 5'-GCTATGGGTTCAACCTGCAT-3'; R: 5'-AAGCCCGCTTGAAGTGTTCAT-3'); G0S2 (F: 5'-GCTGTTGCTCGCAGTCCT-3'; R: 5'- GACTTTCCTCGCATCACCTC-3'); MME (F: 5'-TGTAAAGCCACCCACAAACA-3'; R: 5'- GTTGCTGCCTGTTGACTTGA-3'); MMP9 (F: 5'-AGGGAAGAAGCGTCATGAGA-3'; R: 5'-TCACTGGTTTAGGCGATTCC-3'), and MPO (F: 5'-TCACCAACTTCAACCGTGAC-3'; R: 5'-GGGAACTGGGACAGGTAGGT-3'). GAPDH was used as an internal control (F: 5'-ATCATCAGCAATGCCTCCTG-3'; R: 5'-ATGGACTGTGGTCATGAGTC-3'). The PCR program included 95°C for 5 min, 35 cycles of 95°C for 30 s and 52°C for 30 s. The relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method [23].

Statistical analysis

All statistical analyses were conducted using the R software (version: 3.6.1) with corresponding packages and GraphPad Prism 8.0 software. Univariate Cox regression and LASSO analyses were used to construct the prognosis model. Univariate and multivariate regression analyses were applied to identify independent prognostic risk factors. The gene expression differences between GBM and normal cells were calculated using the Wilcoxon signed-rank test. $P < 0.05$ was considered be of statistical significance.

Results

Investigation and validation of two NETs subtypes

Based on 69 NETGs expression profile, 167 GBM patients from TCGA were divided into two subtypes: cluster 1 and cluster 2 (Fig. 2A-C). The PCA analysis showed a significant difference of gene expression between these two NETs subtypes (Fig. 2D). Moreover,

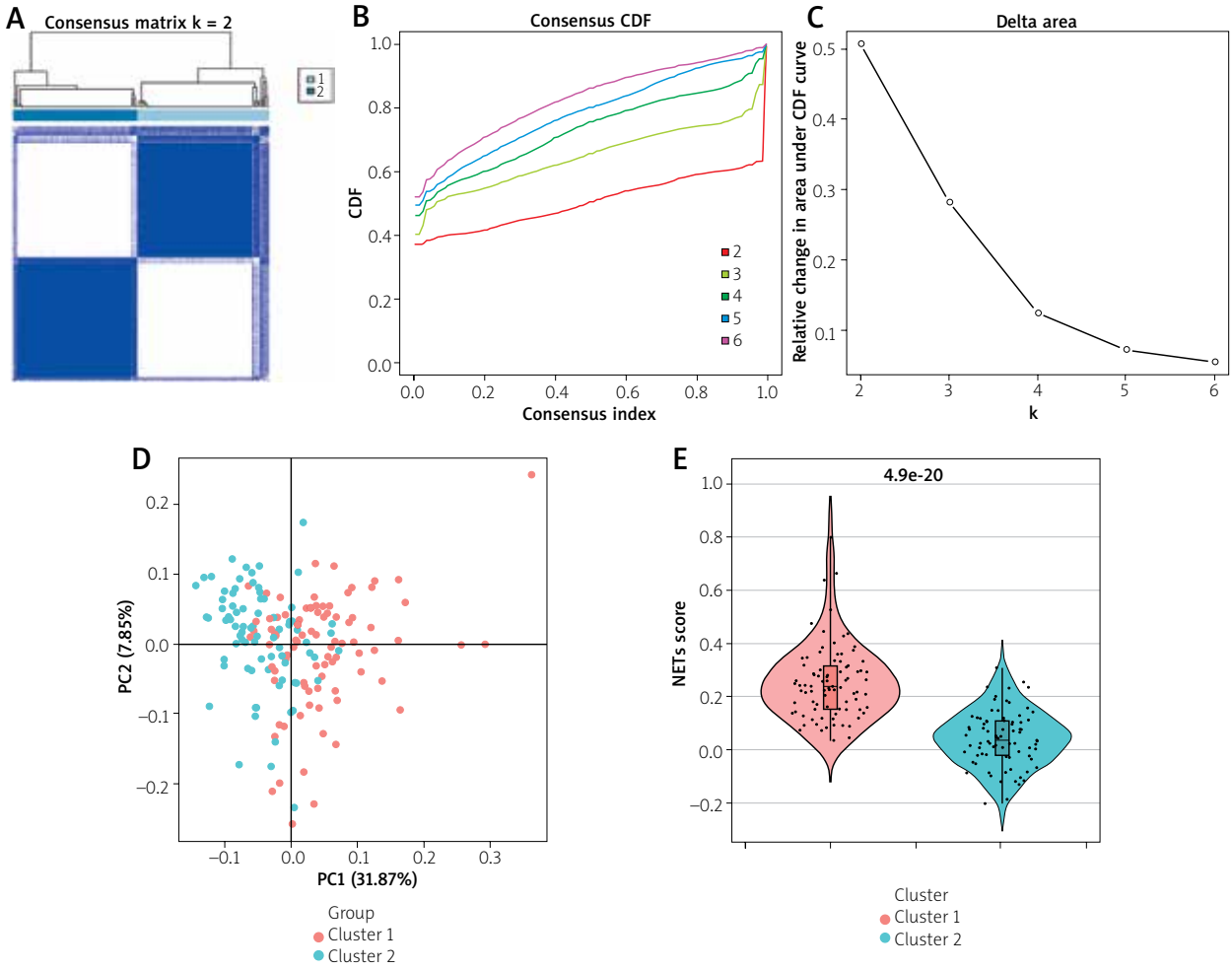

Fig. 2. Two neutrophil extracellular traps (NETs) subtypes revealed by ConsensusClusterPlus. **A**) The consensus matrix showed two clusters of NETs. **B**) The consensus of cumulative density function showed the optimal K = 2. **C**) Delta area plot showed the relative change value of CDF area under curve. **D**) The PAC showed a significant difference between two subtypes of NETs. **E**) The violin chart showed a significant difference in NETs score between two subtypes. X-axis represents different clusters, Y-axis represents the score, and the top number represents the significance – *p* value.

the Wilcoxon signed-rank test showed that the enrichment score in cluster 1 was significantly higher than that in cluster 2 ($p < 0.01$) (Fig. 2E).

Correlation between NETs subtypes and clinical characteristics

The correlation between NETs subtypes and clinical characteristics of GBM patients showed that the clinical characteristics among the subtypes was not significant (Table I). Moreover, the K-M analysis indicated that there was a significant difference ($p < 0.01$) between two subtypes in clinical prognosis (Fig. 3A). Compared with cluster 1, the survival probability of GBM patients was higher than that in cluster 2. In addition, the expression heatmap of 69 NETGs in each sample is shown in Figure 3B.

Immune cell infiltration analysis between two NETs subtypes

A total of 10 kinds of immune cells including CD4 T cell were significantly different between two subtypes based on CIBERSORT (all $p < 0.05$) (Fig. 4A). Moreover, the results of MCPCounter analysis showed that seven immune cells (such as T cells) and one stromal cell (fibroblasts) were significantly different between two subtypes (all $p < 0.05$) (Fig. 4B). Furthermore, the results of ESTIMATE analysis showed that stromal score, immune score, and ESTIMATE score in

Table I. Correlation between NETs subtypes and clinical characteristics of GBM patients

Characteristics	Cluster 1 (*n* = 86)	Cluster 2 (*n* = 81)	Total (*N* = 167)	*P* value
Age (years)				0.33
Mean ±SD	60.65 ±12.32	57.73 ±14.70	59.23 ±13.56	
Median [min, max]	60.50 [21.00, 89.00]	60.00 [21.00, 85.00]	60.00 [21.00, 89.00]	
Sex, *n* (%)				0.97
Female	31 (18.56)	28 (16.77)	59 (35.33)	
Male	55 (32.93)	53 (31.74)	108 (64.67)	
Karnofsky performance score				0.89
Mean ±SD	76.62 ±13.50	75.59 ±16.00	76.13 ±14.69	
Median [min, max]	80.00 [40.00, 100.00]	80.00 [40.00, 100.00]	80.00 [40.00, 100.00]	
Shortest dimension				0.08
Mean ±SD	0.48 ±0.25	0.41 ±0.22	0.45 ±0.24	
Median [min, max]	0.40 [0.10, 1.00]	0.30 [0.10, 0.90]	0.40 [0.10, 1.00]	
Longest dimension				0.35
Mean ±SD	1.16 ±0.34	1.12 ±0.43	1.14 ±0.39	
Median [min, max]	1.00 [0.60, 2.00]	1.00 [0.50, 2.50]	1.00 [0.50, 2.50]	
Disease-free status, *n* (%)				0.53
Disease free	8 (6.72)	11 (9.24)	19 (15.97)	
Recurred/progressed	53 (44.54)	47 (39.50)	100 (84.03)	

NETs – neutrophil extracellular traps, SD – standard deviation, GBM – glioblastoma multiforme

cluster 1 were all significantly higher than those in cluster 2 (all $p < 0.01$) (Fig. 4C).

HLA gene expression, immune checkpoint gene expression and Hallmark enrichment analysis between two NETs subtypes

The result of the HLA analysis showed that except for HLA-DOB, the other 18 genes were differentially expressed between two subtypes (Fig. 5A). For example, the expression of HLA-B in cluster 1 was significantly higher than that in cluster 2 ($p < 0.001$). Moreover, the expression of BTLA, CD274, CD47, CTLA4, HAVCR2, ICOS, PDCD1, PDCD1LG2, TNFRSF4 and TNFRSF9 in cluster 1 was significantly higher than that in cluster 2 (all $p < 0.05$) (Fig. 5B). In addition, Hallmark gene set enrichment analysis of cluster 1 vs. cluster 2 revealed that 28 up-regulated Hallmark gene sets and eight down-regulated Hallmark gene sets were enriched in cluster 1. Top 10 up-regulated and all down-regulated Hallmark gene sets are displayed in Figure 5C and D. For example, cluster 1 subtype was mainly associated with IL6/JAK/STAT3 signalling and inflammatory response pathways, while cluster 2 subtype was mainly involved in G2M checkpoint and E2F targets pathways.

Development of the prognostic model based on the NETGs

Univariate Cox regression analysis was performed on 69 NETGs to screen genes with a prognostic value. With $p < 0.05$, a total of 15 prognostic NETGs were identified (Fig. 6A). Then, LASSO regression was applied to determine the optimal λ value, and seven optimal risk genes were screened, including six up-regulated genes (*CPPED1*, *F3*, *G0S2*, *MME*, *MMP9* and *MPO*) and one down-regulated gene (MAPK1) (Fig. 6B and C). The prognostic model was constructed using these genes and RS was calculated as follows: RS = Exp (CPPED1) * (0.223340457) + Exp (F3) * (0.1517013) + Exp (G0S2) * (0.046311722) + Exp (MAPK1) * (–0.219913858) + Exp (MME) * (0.192197058) + Exp (MMP9) * (0.016505054) + Exp (MPO) * (0.414575274). Then, samples from TCGA training data and CGGA validation data were assigned into two different risk groups (High_Risk and Low_Risk) based on the median RS.

In the TCGA training cohort, patients in the High_Risk group had worse survival time than those in the Low_Risk group (Fig. 6D). A greater proportion of deaths were observed in the High_Risk group compared to the Low-Risk group (Fig. 6E). Moreover, the AUC values

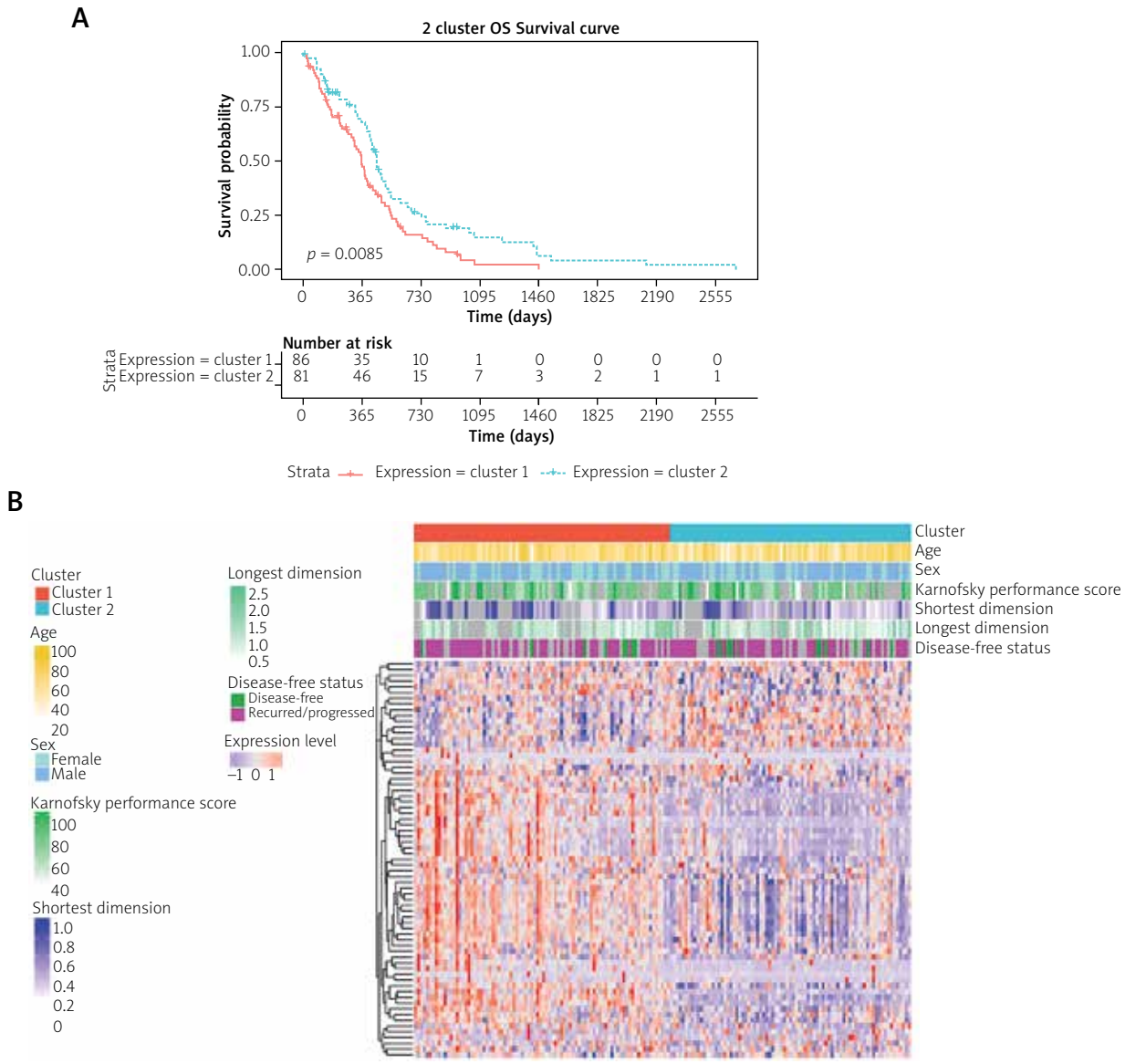

Fig. 3. Correlation between NETs subtypes and clinical characteristics. **A**) Kaplan-Meier (KM) survival curve revealed the significant difference between two subtypes of NETs. The blue line represents a sample in cluster 2, while red line represents a sample in cluster 1. **B**) Expression heatmap of 69 NETs related genes in each sample.

of ROC curves for 1-, 2-, and 3-year overall survival were 0.695, 0.75, and 0.808, respectively (Fig. 6F).

In the CGGA validation cohort, overall survival was significantly worse in the High_Risk group than the Low_Risk group (Fig. 6G). The RS and patient survival status distributions are displayed in Figure 6H. Meanwhile, the AUC values for 1-, 2-, and 3-year overall survival were 0.619, 0.63, and 0.657, respectively (Fig. 6I). Collectively, these results indicated that the prognostic model might have good predictive performance for overall survival of patients with GBM.

Correlation analysis of RS with clinical characteristics of GBM patients

The correlation analysis of RS with NETs subtypes or RS with clinical characteristics (age, sex, Karnofsky performance score, and disease-free status) were performed. Results showed that significant difference of RS among different subtypes (Supplementary Fig. 1A). However, the correlation between RS and each clinical characteristic was not significant (Supplementary Fig. 1B-E). Scatter diagram analysis of Spearman cor-

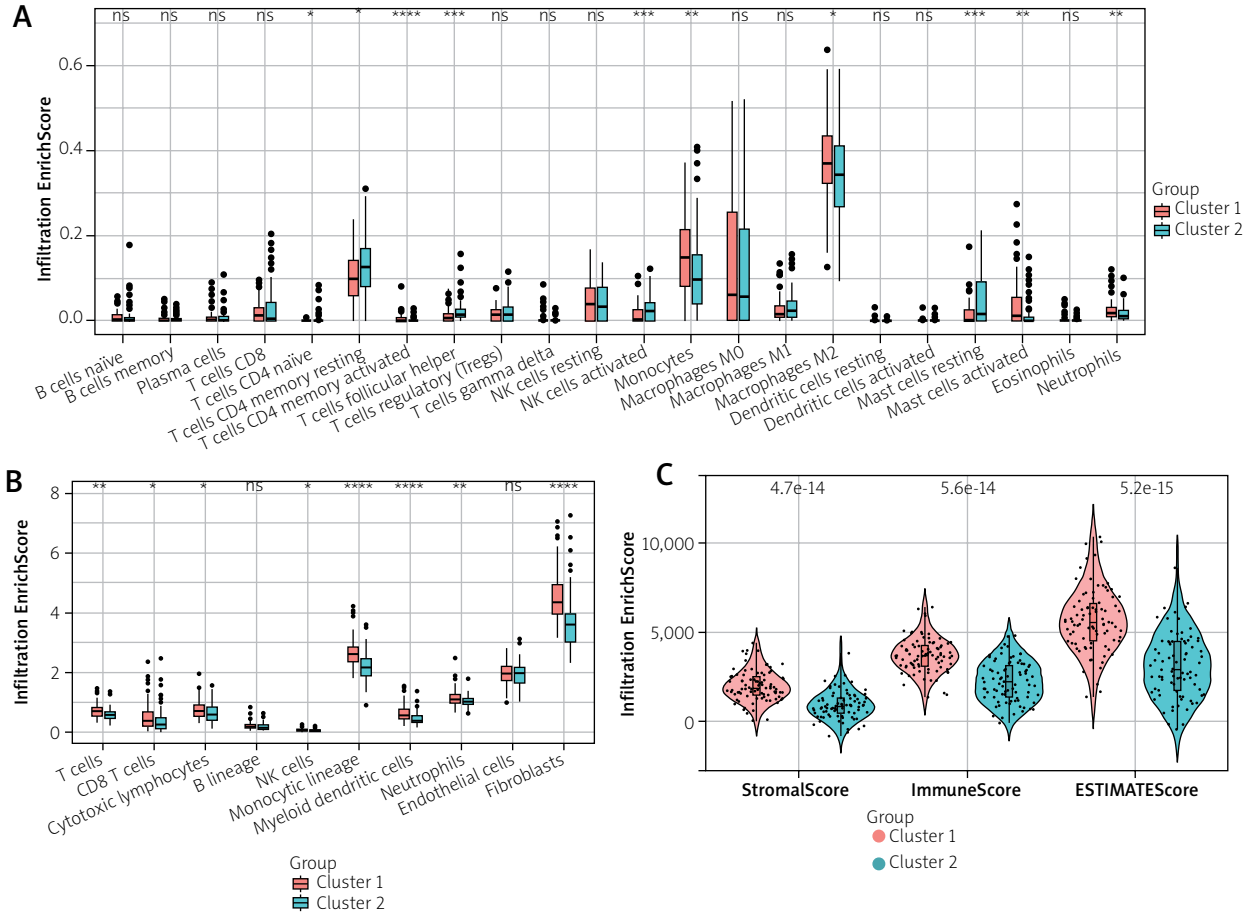

Fig. 4. Immune-associated analysis of two NETs subtypes. **A**) Box diagram revealed the result of immune cell infiltration between two NETs subtypes by using CIBERSORT analysis. X-axis represents different immune cells, while Y-axis represents EnrichScore. **B**) Box diagram revealed the result of immune cell infiltration between two NETs subtypes by using MCPCounter analysis. X-axis represents different immune cells, while Y-axis represents EnrichScore. **C**) Box diagram revealed the result of immune cell infiltration between two NETs subtypes by using ESTIMATE analysis. X-axis represents different items, while Y-axis represents score. *0.01 < p < 0.05, **0.001 < p < 0.01, ***p < 0.001, ****p < 0.0001; ns – not significant.

relation between RS, longest dimension, and shortest dimension showed that there was no significant correlation between RS and clinical characteristics (Supplementary Fig. 1F, G).

The independent analysis for the prognostic model

Age, sex, Karnofsky performance score, shortest dimension, longest dimension, and RS were used for independent analysis. With p < 0.05, the univariate and multivariate Cox regression analyses revealed that two indicators (RS and age) could be used as independent prognostic factors (Fig. 7A, B). Subsequently, a nomogram incorporating the RS and age was constructed (Fig. 7C). Each variable was scored on the point scale axis. The total score could be easily calculated by adding each individual score to estimate the GBM's survival probabilities. The result showed that the current nomogram could be used to predict 1-, 2- and 3-year survival probabilities of GBM.

Further, we evaluated the predictive performance of the nomogram. The calibration curve for 1- and 2-year overall survival indicated high consistency between nomogram predictions and actual observations (Fig. 7D). Meanwhile, the AUC of the nomogram in predicting 1-, 2-, and 3-year survival were 0.735, 0.751, and 0.809, respectively (Fig. 7E).

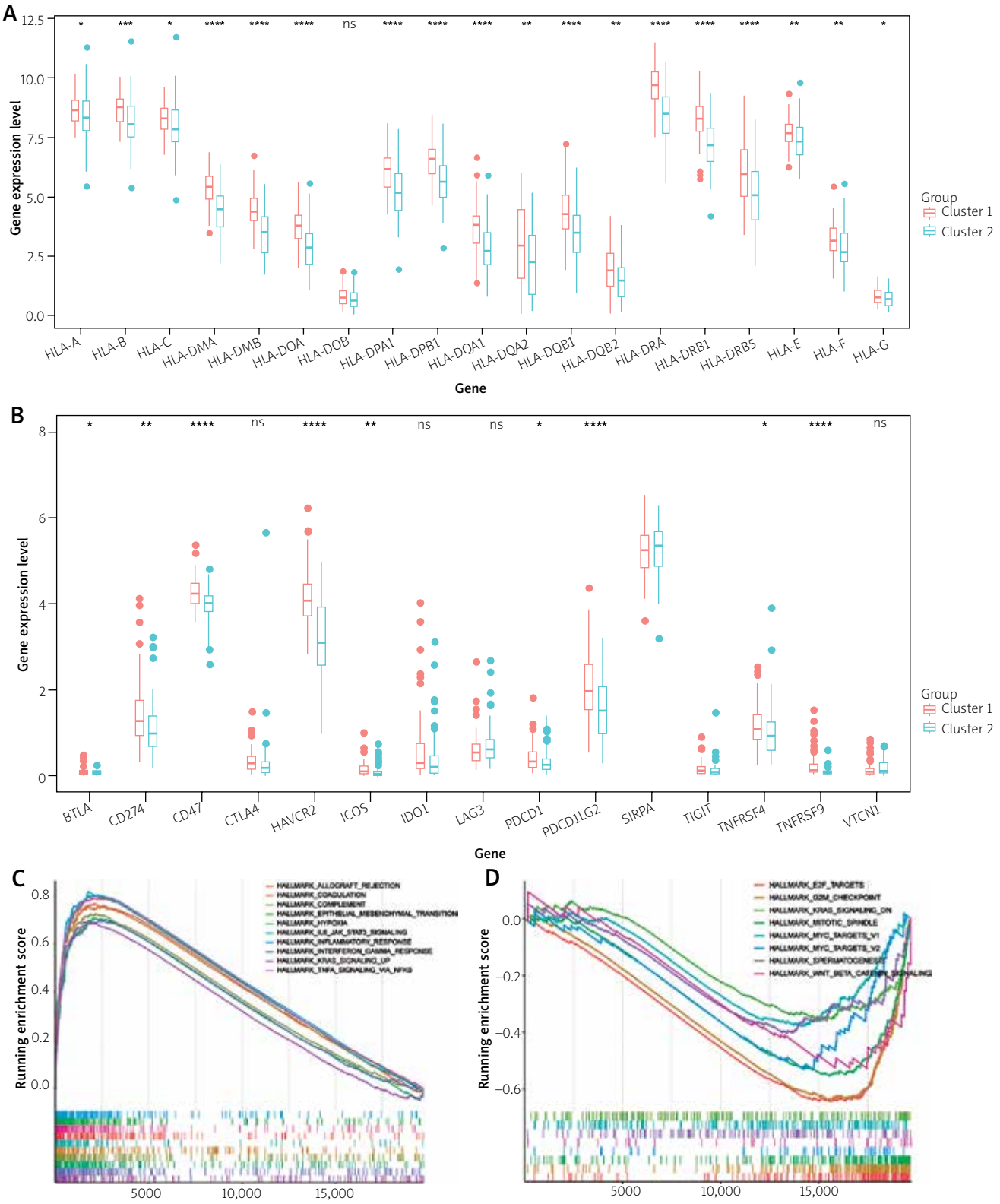

Fig. 5. HLA gene expression, immune checkpoint gene expression and Hallmark enrichment analysis between two NETs subtypes. **A**) Box diagram for HLA family genes expression between two NETs subtypes. X-axis represents the gene name of HLA family, while Y-axis represents the gene expression level. **B**) Box diagram revealed the differences in expression of immune checkpoint gene between two NETs subtypes. X-axis represents an immune checkpoint gene, while Y-axis represents a gene expression level. **C**) Top 10 up-regulated Hallmark gene set in cluster 1. **D**) Eight down-regulated Hallmark gene set in cluster 1. *0.01 < p < 0.05, **0.001 < p < 0.01, ***p < 0.001, ****p < 0.0001; ns – not significant.

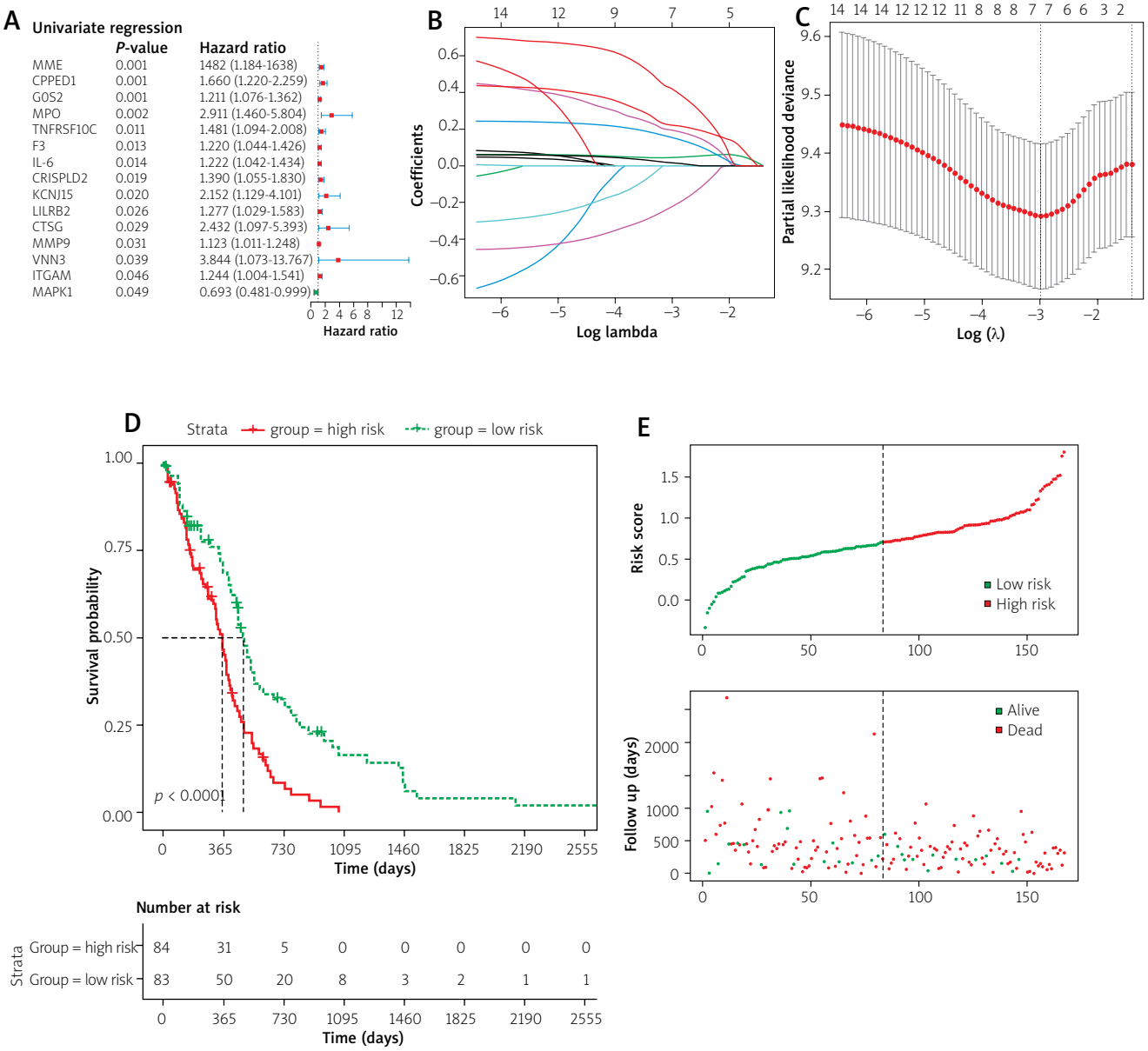

Fig. 6. Construction and validation of the prognostic model based on NETGs. **A**) Forest plot of univariate Cox regression analysis showed 15 prognostic NETGs. **B**) Plot of LASSO coefficient profiles. X-axis represents the value of log (Lambda), and Y-axis represents the coefficient of the variable. **C**) Screening of optimal parameters in the lasso regression model based on 20-fold cross validation. X-axis represents the value of log (Lambda), and Y-axis in the right represents the value of binomial deviance. **D**) KM survival curve based on the RS model in the TCGA training dataset. The red dot represents a sample in the High_Risk group, while the green dot represents a sample in the Low_Risk group. **E**) Distribution of RS and survival status of samples in the TCGA training dataset.

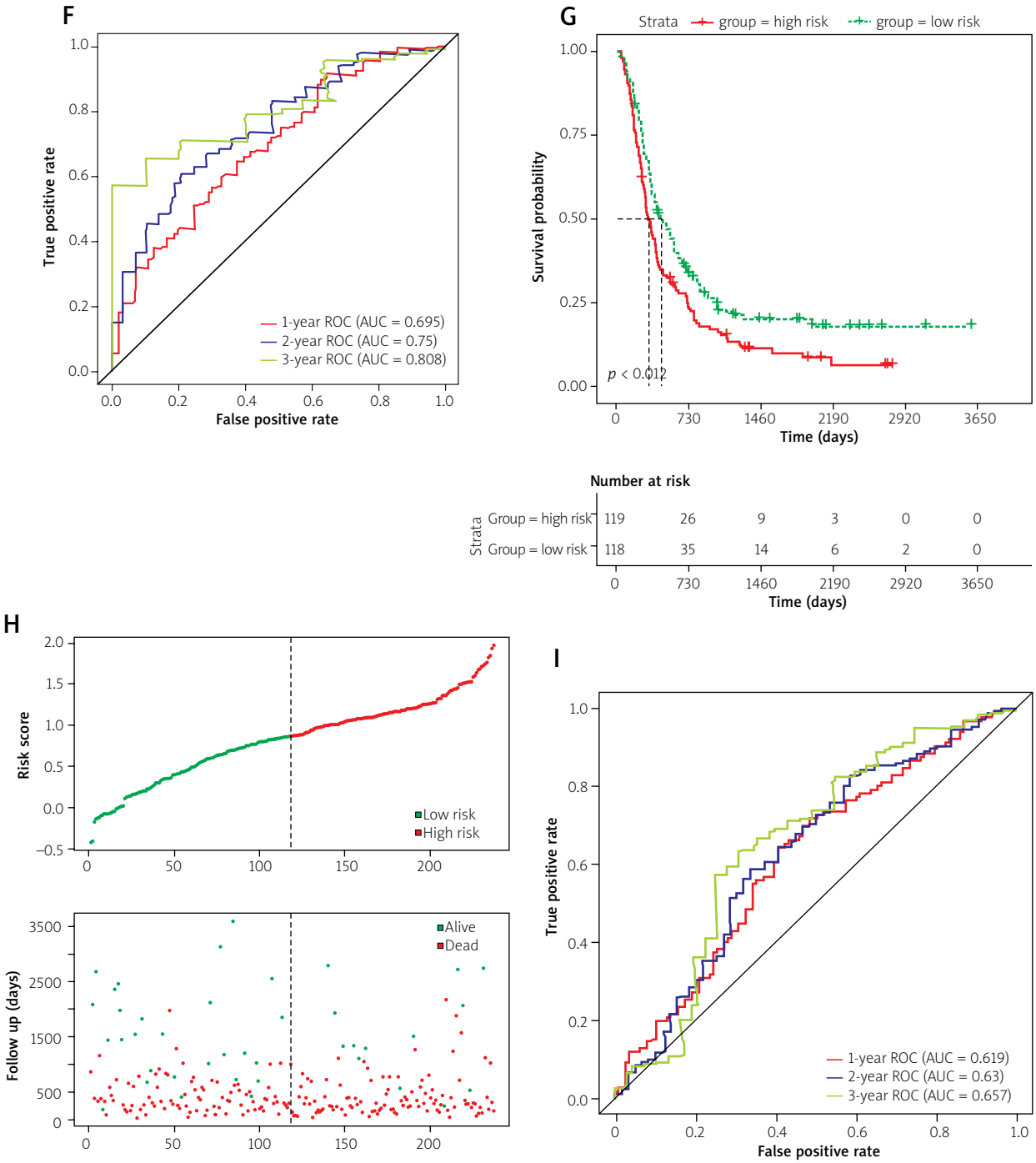

Fig. 6. Cont. **F**) ROC curve of 1-, 2-, and 5-year survival based on samples in the TCGA training dataset. **G**) KM survival curve based on the RS model in the CGGA validation dataset. **H**) distribution of RS and survival status for samples in the CGGA validation dataset. **I**) ROC curve of 1-, 2-, and 5-year survival based on samples in the CGGA validation dataset.

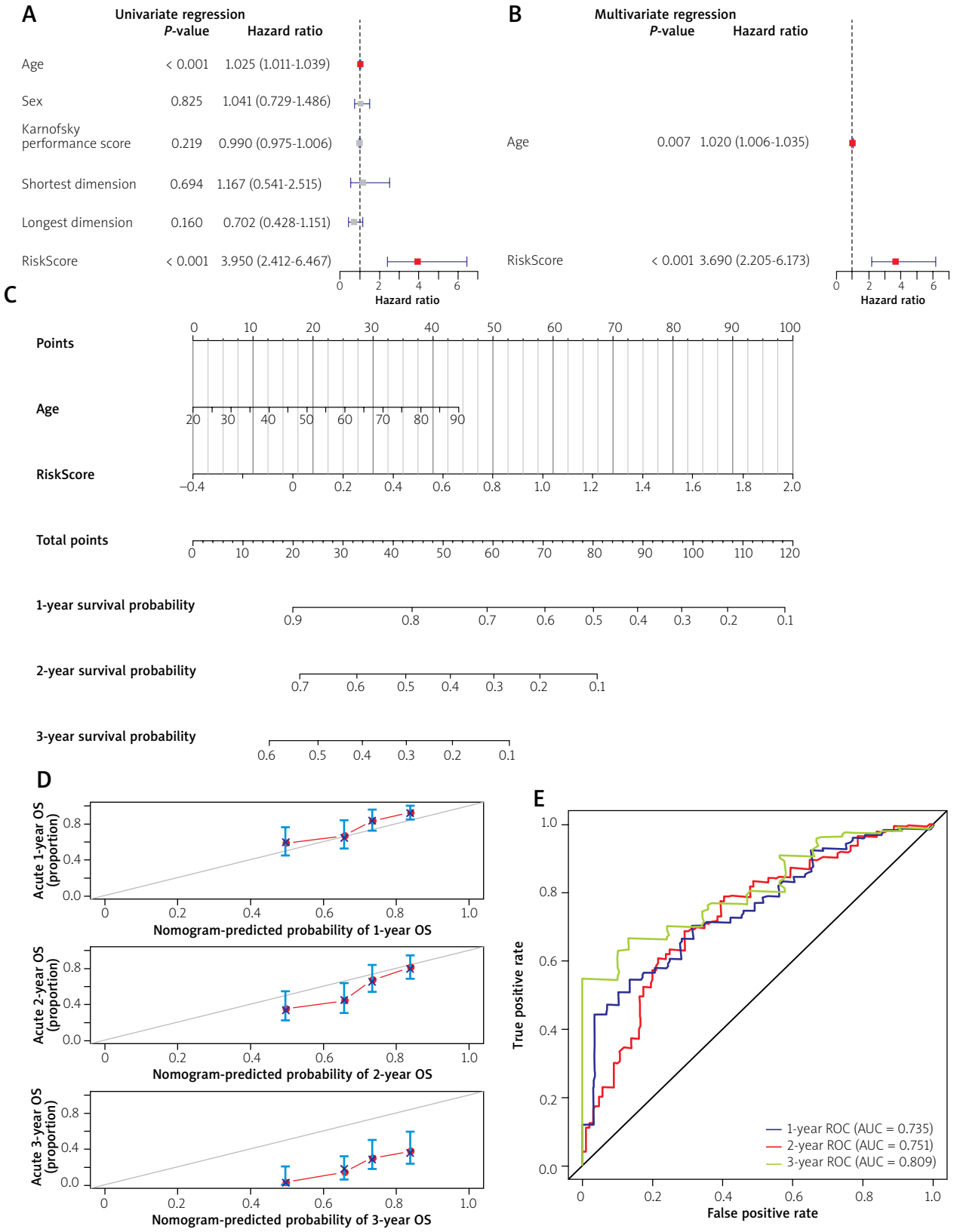

Fig. 7. Independent prognostic factor investigation and nomogram development of GBM. **A**) Univariate regression of the RS and clinical factors (age, sex, Karnofsky performance score, shortest dimension, and longest dimension). **B**) Multivariate regression revealed the independent factors for GBM. **C**) Nomogram model predicting the 1-, 2-, and 3-year OS in patients with GBM: the nomogram was used by summing all points identified on the scale for each variable; the total points projected on the bottom scales indicate the probabilities of 1-, 2-, and 3-year survival. **D**) Nomogram calibration curve analysis of 1-, 2-, and 3-year survival. **E**) The ROC curve of 1-, 2-, and 3-year survival.

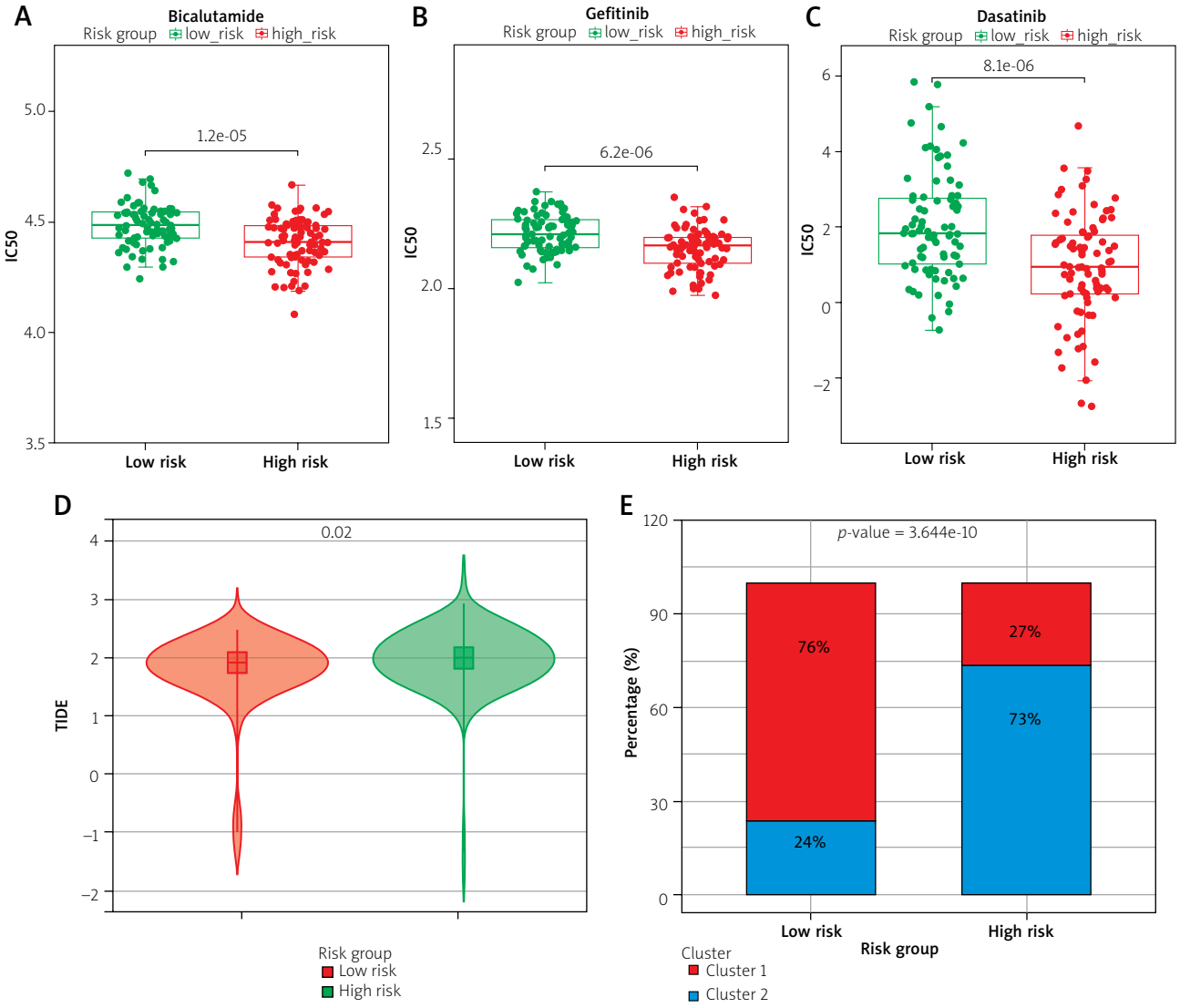

Fig. 8. Drug sensitivity, immunotherapeutic response, and NETs subtypes correlation analysis based on the high-risk group and the low-risk group determined by the prognostic model. **A**) Box diagram showed the IC50 difference of bicalutamide sensitivity between High_Risk and Low-Risk groups. The red block represents the High_Risk group, and the green block represents the Low_Risk group. **B**) Box diagram showed the IC50 difference of gefitinib sensitivity between two risk groups. **C**) Box diagram showed the IC50 difference of dasatinib sensitivity between two risk groups. **D**) Distribution of tumour immune dysfunction and exclusion (TIDE) scores between two risk groups. **E**) Bar chart showed the distribution of subtypes in risk groups. Red and blue boxes represent NETs cluster 1 and cluster 2.

Correlation analysis between immune cell and signature genes

Based on the relative abundance of different immune cells obtained between NETs subtypes, the correlations between immune cells and seven signature genes were explored. The correlation heatmap was showed in Supplementary Figure 2. For example, CPPED1 and monocytic lineage showed the strongest positive correlation, while MMP9 and monocytes showed the strongest negative correlation.

Drug sensitivity and immunotherapeutic response analysis

The difference in IC_{50} values of sensitivity to drugs between two risk groups was revealed. The result showed that the IC_{50} values of 82 drugs were signifi-

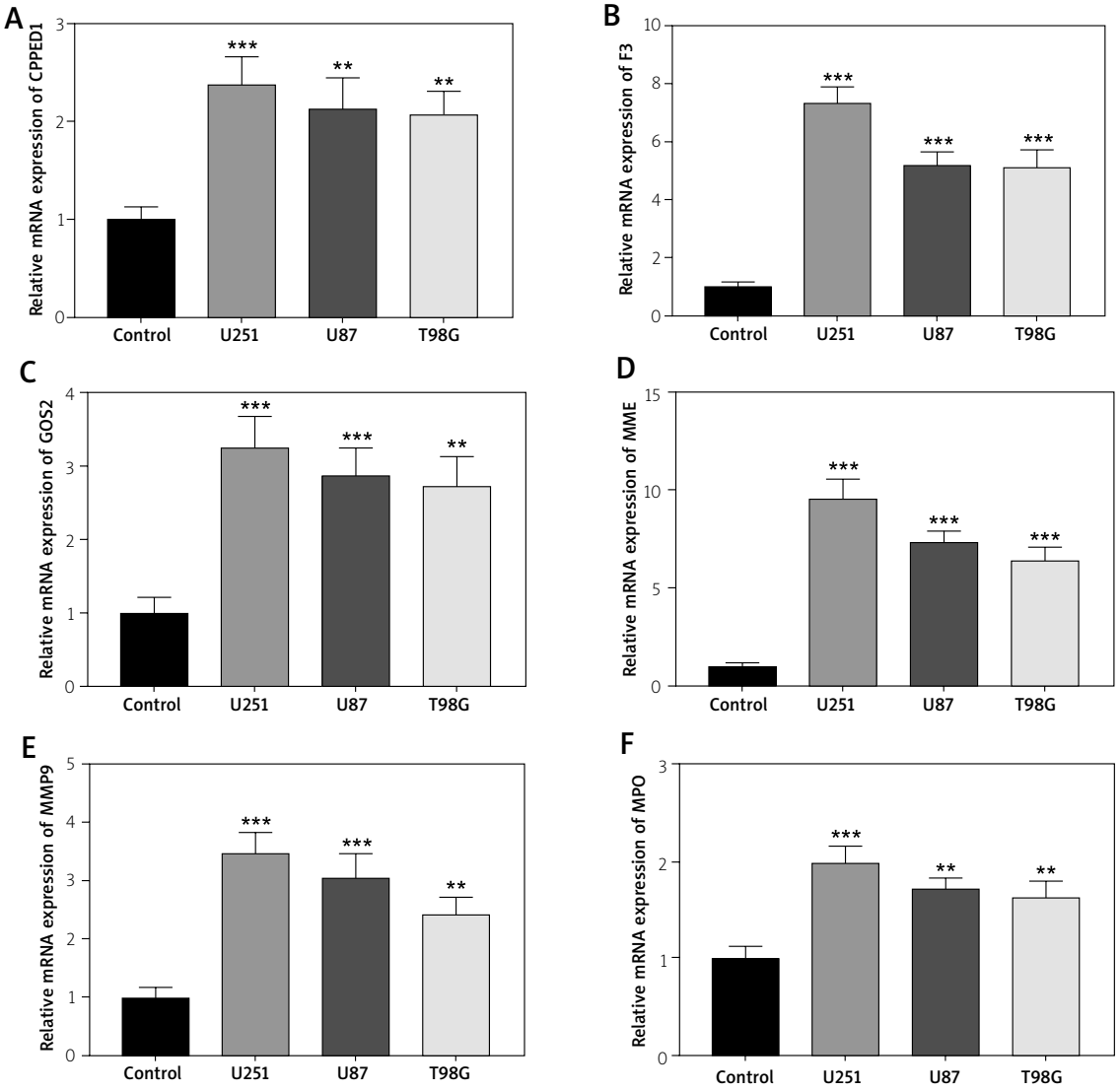

Fig. 9. A-E) Results of qPCR validation analysis for six signature genes in the prognostic model. X-axis represents different cells, while the Y-axis represents the relative mRNA expression of genes. Control, normal glial cells; U251, U87 and T98G, three different GBM cell lines. $^{**}p < 0.05$, $^{***}p < 0.01$.

cantly different between two risk groups (all $p < 0.05$, data not shown). For example, the IC_{50} values of bicalutamide, gefitinib, and dasatinib in the High_Risk group were significantly lower than those in the Low_Risk group (all $p < 0.05$), indicating that these antitumor drugs would be more efficient for GBM patients in the High_Risk group (Fig. 8A-C). Moreover, the TIDE value of each sample in two risk groups was compared. The result indicated that TIDE score of the Low_Risk group was significantly higher than that of the High_Risk group ($p < 0.05$), suggesting that the greater the potential for immune escape and the worse the efficacy of immunotherapy for patients in the Low_Risk group (Fig. 8D). Further, the correlation between different risk groups and subtype distribution was analyzed by using the χ^2 test (Fig. 8E). The results showed that GBM patients in the Low_Risk group tended to be concentrated in cluster 2, while patients in the High_Risk group are inclined to be assigned in cluster 1.

Validation of signature genes expression in cell lines by qPCR

Expression of six signature genes (including *CPPED1*, *F3*, *GOS2*, *MME*, *MMP9*, and *MPO*) in three kinds of GBM cell lines and one normal cell (control) were investigated based on qPCR analysis. The result showed

that the expression of six genes was significantly higher in three GBM cell lines (U251, U87, and T98G) than those in controls (all $p < 0.05$) (Fig. 9). The expression of 6 genes in verification analysis was in accordance with the result of our current bioinformatic study, which indicated a reliable result for this study.

Discussion

Glioblastoma multiforme is the deadliest form of brain tumour with a poor prognosis worldwide [15]. Although NETs play vital roles in the development and prognosis of various human cancers, the detail molecular mechanism of NETs in GBM remains unclear. In this study, the molecular mechanism of NETs in GBM pathogenesis was revealed by molecular subtype investigation and prognostic signature identification, which provide a basis for future studies on early diagnosis and treatment of GBM.

NETs have been proved to be successfully used for the identification and validation of molecular subtypes in human cancer such as renal cell carcinoma [34]. During this process, the inflammation and immune response of NETs is of vital importance. It is known that NETs participate in inflammation and immune response of human glioma [52]. As an immune check point gene, cell death (CD)47-mediated immune evasion is closely related with GBM progression [16]. It has been proved that CD47 is a promising antibody target for GBM since it enhances tumour cell phagocytosis [49]. Mast cells are widely recognized as critical effector cells in immunoglobulin E-associated acquired immune responses. A previous study shows that mast cells can modulate proliferation, migration and stemness of GBM cells [2]. Meanwhile, as a cell type in TME, neutrophils can promote tumour growth in glioma [26]. The important role of neutrophils and NETs in host defence, chronic inflammation, and tissue disrepair has been revealed in a previous study [3]. It has been shown that the different immune score can be used to define subtypes with different clinical and micro-environmental cell infiltration characteristics in glioma [56]. In fact, these differences in immune response and inflammation are important indicators of different subtypes in human cancer. In this study, two NETs-related subtypes in GBM were revealed based on NETGs. The immune cell including neutrophils and mast cells, as well as immune check point genes like CD47 were all significantly differentially expressed between two NETs subtypes. Meanwhile, the immune score calculated by ESTIMATE also exhibited a difference between two NETs subtypes. Importantly, the survival and clinical characteristics correlation analysis showed visible difference between patients in different NETs subtypes. Thus, we speculated that NETs classification could be used to define subtypes in GBM.

It is known to all that NETs and related genes can be used as prognostic markers in veracious kinds of human diseases like COVID-19 and terminal cancer [29,35]. Based on NETs, Quan and Huang successfully identified and validated several prognostic signatures in renal cancer including G0S2 and MMP9 [34]. In this research, we also constructed a reliable prognostic model for GBM based on NETGs. The G0/G1 switch gene 2 (G0S2) is dysregulated in the brain metastases of patients with breast cancer [5]. A previous study shows that G0S2 is a breast cancer tumour suppressor, which could be used as a novel prognostic marker [45]. Matrix metallopeptidase 9 (MMP9) is a reliable diagnostic marker in various kinds of human glioma [6]. As a neutrophil marker, MMP9 together with MPO is elevated with self-reported fever in malaria [36]. Myeloperoxidase (MPO) is commonly observed to be differentially expressed in human cancer [24]. As a marker for NETs levels, MPO-DNA complexes are closed associated with the cardiovascular risk [9]. In addition, Calcineurin Like Phosphoesterase Domain Containing 1 (CPPED1) is a novel molecule involved in glucose uptake in adipocytes. A previous integrated analysis of miRNA-mRNA regulatory network showed that CPPED1 is one of the hub genes in osteonecrosis [48]. Coagulation Factor III (F3) is one of the proteins that participate in haemostatic and inflammatory processes [32]. As a transmembrane glycoprotein, membrane metalloendopeptidase (MME) can interrupt tumour cell adhesion and lead to a beneficial outcome for patients, indicating a promising prognostic marker for esophageal squamous cell carcinoma [21]. In this study, a prognostic signature constructed by seven signature genes including *CPPED1*, *F3*, *G0S2*, *MME*, *MMP9*, *MAPK1*, and *MPO* were developed to stratify GBM patients into High_Risk and Low_Risk groups, and the time-dependent ROC analysis proved a good predictive ability of this signature. The expression of six genes in qPCR verification analysis was in accordance with the result of our current bioinformatic study. Meanwhile, the High-Risk patients were associated with poor survival and low IC_{50} value in chemical drugs like bicalutamide. It is reported that the sensitivity to chemotherapy drugs can be used to determine the clinical prognosis [33]. In a cuproptosis-based signature study for glioma, the IC_{50} values of bicalutamide was lower in the high-CuproptosisScore group than those in the low-CuproptosisScore group [41], suggesting that analysing the drug sensitivity of different risk groups will guide the development of personalized treatment. All these results in the current study indicate that NETs-related prognostic signature could predict the survival of patients with GBM. Genes including *CPPED1*, *F3*, *G0S2*, *MME*, *MMP9* and *MPO* might be novel prognostic markers for GBM prognosis.

The nomogram is widely used as a predictive device in oncology and medicine [42]. It is commonly used to explore the risk of human cancers [51]. In a previous study, the nomogram accurately classified GBM patients into two risk groups based on gene signature, indicating the value of the nomogram in predicting survival in GBM [43]. A previous study shows that a nomogram has been successfully established for predicting the OS of patients with GBM [18]. It has been proved that factors including age and RS were independent predictors of GBM participating in the nomogram construction [10,37]. In this study, the current nomogram constructed by screening two independent indicators including age and RS could be used to predict 1-, 3- and 5-year survival probability of GBM. The calibration curve of the nomogram indicated that the predicted survival rates of 1-, 2- and 3-year have superior accuracy. Thus, the novel nomogram established by two independent predictors including age and RS were valuable in predicting the survival of GBM.

In conclusion, a NETs-related prognostic signature was developed, which could independently predict GBM survival. NETs classification might be used to define subtypes with distinct clinical and microenvironment cell infiltration characteristics in GBM. *CPPED1*, *F3*, *G0S2*, *MME*, *MMP9* and *MPO* might be novel biomarkers for GBM prognosis.

Disclosure

The authors report no conflict of interest.

References

1. Albrengues J, Shields MA, Ng D, Park CG, Ambrico A, Poindexter ME, Upadhyay P, Uyeminami DL, Pommier A, Küttner V, Bružas E, Maiorino L, Bautista C, Carmona EM, Gimotty PA, Fearon DT, Chang K, Lyons SK, Pinkerton KE, Trotman LC, Goldberg MS, Yeh JT, Egeblad M. Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice. Science 2018; 361: eaao4227.
2. Attarha S, Roy A, Westermark B, Tchougounova E. Mast cells modulate proliferation, migration and stemness of glioma cells through downregulation of GSK3β expression and inhibition of STAT3 activation. Cell Signal 2017; 37: 81-92.
3. Castanheira FVS, Kubes P. Neutrophils and NETs in modulating acute and chronic inflammation. Blood 2019; 133: 2178-2185.
4. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol 2018; 1711: 243-259.
5. Cho E, Kwon YJ, Ye DJ, Baek HS, Kwon TU, Choi HK, Chun YJ. G0/G1 Switch 2 induces cell survival and metastasis through integrin-mediated signal transduction in human invasive breast cancer cells. Biomol Ther (Seoul) 2019; 27: 591-602.
6. Comincini S, Paolillo M, Barbieri G, Palumbo S, Sbalchiero E, Azzalin A, Russo MA, Schinelli S. Gene expression analysis of an EGFR indirectly related pathway identified PTEN and MMP9 as reliable diagnostic markers for human glial tumor specimens. J Biomed Biotechnol 2009; 2009: 924565.
7. Cristinziano L, Modestino L, Antonelli A, Marone G, Simon HU, Varricchi G, Galdiero MR. Neutrophil extracellular traps in cancer. Semin Cancer Biol 2022; 79: 91-104.
8. Deng J, Kang Y, Cheng CC, Li X, Dai B, Katz MH, Men T, Kim MP, Koay EA, Huang H, Brekken RA, Fleming JB. DDR1-induced neutrophil extracellular traps drive pancreatic cancer metastasis. JCI Insight 2021; 6: e146133.
9. Donkel SJ, Wolters FJ, Ikram MA, Maat M. Circulating myeloperoxidase (MPO)-DNA complexes as marker for neutrophil extracellular traps (NETs) levels and the association with cardiovascular risk factors in the general population. PLoS One 2021; 16: e0253698.
10. Ferguson SD, Hodges TR, Majd NK, Alfaro-Munoz K, Al-Holou WN, Suki D, de Groot JF, Fuller GN, Xue L, Li M, Jacobs C, Rao G, Colen RR, Xiu J, Verhaak R, Spetzler D, Khasraw M, Sawaya R, Long JP, Heimberger AB. A validated integrated clinical and molecular glioblastoma long-term survival-predictive nomogram. Neurooncol Adv 2021; 3: vdaa146.
11. Gallego O. Nonsurgical treatment of recurrent glioblastoma. Curr Oncol 2015; 22: e273.
12. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS One 2014; 9: e107468.
13. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, Zhu J, Haussler D. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol 2020; 38: 675-678.
14. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics 2013; 14: 7.
15. Huang B, Li X, Li Y, Zhang J, Zong Z, Zhang H. Current Immunotherapies for glioblastoma multiforme. Front Immunol 2020; 11: 603911.
16. Jiang N, Xie B, Xiao W, Fan M, Xu S, Duan Y, Hamsafar Y, Evans AC, Huang J, Zhou W, Lin X, Ye N, Wanggou S, Chen W, Jing D, Fragoso RC, Dugger BN, Wilson PF, Coleman MA, Xia S, Li X, Sun LQ, Monjazeb AM, Wang A, Murphy WJ, Kung HJ, Lam KS, Chen HW, Li JJ. Fatty acid oxidation fuels glioblastoma radioresistance with CD47-mediated immune evasion. Nat Commun 2022; 13: 1511.
17. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpfennig KW, Liu XS. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018; 24: 1550-1558.
18. Kudulaiti N, Zhou Z, Luo C, Zhang J, Zhu F, Wu J. A nomogram for individualized prediction of overall survival in patients with newly diagnosed glioblastoma: a real-world retrospective cohort study. BMC Surg 2021; 21: 238.
19. Kwak SB, Kim SJ, Kim J, Kang YL, Ko CW, Kim I, Park JW. Tumor regionalization after surgery: Roles of the tumor microenvironment and neutrophil extracellular traps. Exp Mol Med 2022; 54: 720-729.
20. Lee KH, Kronbichler A, Park DD, Park Y, Moon H, Kim H, Choi JH, Choi Y, Shim S, Lyu IS, Yun BH, Han Y, Lee D, Lee SY, Yoo BH, Lee KH, Kim TL, Kim H, Shim JS, Nam W, So H, Choi S, Lee S, Shin JI. Neutrophil extracellular traps (NETs) in autoimmune diseases: A comprehensive review. Autoimmun Rev 2017; 16: 1160-1173.
21. Li M, Wang L, Zhan Y, Zeng T, Zhang X, Guan XY, Li Y. Membrane Metalloendopeptidase (MME) Suppresses Metastasis of Esophageal Squamous Cell Carcinoma (ESCC) by Inhibiting FAK-RhoA Signaling Axis. Am J Pathol 2019; 189: 1462-1472.
22. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011; 27: 1739-1740.

23. Livak KJ ST. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 2001; 25: 402-408.
24. Lockhart JS, Sumagin R. Non-canonical functions of myeloperoxidase in immune regulation, tissue inflammation and cancer. Int J Mol Sci 2022; 23: 12250.
25. Lopes M, Carvalho B, Vaz R, Linhares P. Influence of neutrophil-lymphocyte ratio in prognosis of glioblastoma multiforme. J Neurooncol 2018; 136: 173-180.
26. Massara M, Persico P, Bonavita O, Mollica Poeta V, Locati M, Simonelli M, Bonecchi R. Neutrophils in gliomas. Front Immunol 2017; 8: 1349.
27. Masucci MT, Minopoli M, Del Vecchio S, Carriero MV. The emerging role of neutrophil extracellular traps (NETs) in tumor progression and metastasis. Front Immunol 2020; 11: 1749.
28. Nam JY, de Groot JF. Treatment of glioblastoma. J Oncol Pract 2017; 13: 629-638.
29. Ondracek AS, Lang IM. Neutrophil extracellular traps as prognostic markers in COVID-19: A welcome piece to the puzzle. Arterioscler Thromb Vasc Biol 2021; 41: 995-998.
30. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, Stroup NE, Kruchko C, Barnholtzsloan JS. CBTRUS Statistical Report: Primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. Neuro Oncol 2012; 14 Suppl 5: v1.
31. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. Nat Rev Immunol 2018; 18: 134-147.
32. Peña E, de la Torre R, Arderiu G, Slevin M, Badimon L. mCRP triggers angiogenesis by inducing F3 transcription and TF signalling in microvascular endothelial cells. Thromb Haemost 2017; 117: 357-370.
33. Pondé NF, Zardavas D, Piccart M. Progress in adjuvant systemic therapy for breast cancer. Nat Rev Clin Oncol 2019; 16: 27-44.
34. Quan J, Huang B. Identification and validation of the molecular subtype and prognostic signature for clear cell renal cell carcinoma based on neutrophil extracellular traps. Front Cell Dev Biol 2022; 10: 1021690.
35. Rosell A, Aguilera K, Hisada Y, Schmedes C, Mackman N, Wallén H, Lundström S, Thålin C. Prognostic value of circulating markers of neutrophil activation, neutrophil extracellular traps, coagulation and fibrinolysis in patients with terminal cancer. Sci Rep 2021; 11: 5074.
36. Sarr D, Oliveira LJ, Russ BN, Owino SO, Middii JD, Mwalimu S, Ambasa L, Almutairi F, Vulule J, Rada B. Myeloperoxidase and other markers of neutrophil activation associate with malaria and malaria/HIV coinfection in the human placenta. Front Immunol 2021; 12: 682668.
37. Shu C, Yan X, Zhang X, Wang Q, Cao S, Wang J. Tumor-induced mortality in adult primary supratentorial glioblastoma multiforme with different age subgroups. Future Oncol 2019; 15: 1105-1114.
38. Therneau TM, Lumley T. Package 'survival'. R Top Doc 2015; 128: 28-33.
39. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997; 16: 385-395.
40. Wang P, Wang Y, Hang B, Zou X, Mao JH. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget 2016; 7: 55343-55351.
41. Wang W, Lu Z, Wang M, Liu Z, Wu B, Yang C, Huan H, Gong P. The cuproptosis-related signature associated with the tumor environment and prognosis of patients with glioma. Front Immunol 2022; 13: 998236.
42. Wang Y, Shao Q, Luo S, Fu R. Development of a nomograph integrating radiomics and deep features based on MRI to predict the prognosis of high grade Gliomas. Math Biosci Eng 2021; 18: 8084-8095.
43. Wang Z, Gao L, Guo X, Feng C, Lian W, Deng K, Xing B. Development and validation of a nomogram with an autophagy-related gene signature for predicting survival in patients with glioblastoma. Aging (Albany NY) 2019; 11: 12246-12269.
44. Yang L, Liu Q, Zhang X, Liu X, Zhou B, Chen J, Huang D, Li J, Li H, Chen F, Liu J, Xing Y, Chen X, Su S, Song E. DNA of neutrophil extracellular traps promotes cancer metastasis via CCDC25. Nature 2020; 583: 133-138.
45. Yim CY, Bikorimana E, Khan E, Warzecha JM, Shin L, Rodriguez J, Dmitrovsky E, Freemantle SJ, Spinella MJ. G0S2 represses PI3K/mTOR signaling and increases sensitivity to PI3K/mTOR pathway inhibitors in breast cancer. Cell Cycle 2017; 16: 2146-2155.
46. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013; 4: 2612.
47. Young RM, Jamshidi A, Davis G, Sherman JH. Current trends in the surgical management and treatment of adult glioblastoma. Ann Transl Med 2015; 3: 121.
48. Yu L, Yao T, Jiang Z, Xu T. Integrated analysis of miRNA-mRNA regulatory networks associated with osteonecrosis of the femoral head. Evid Based Complement Alternat Med 2021; 2021: 8076598.
49. Zhang M, Hutter G, Kahn SA, Azad TD, Gholamin S, Xu CY, Liu J, Achrol AS, Richard C, Sommerkamp P, Schoen MK, McCracken MN, Majeti R, Weissman I, Mitra SS, Cheshier SH. Anti-CD47 treatment stimulates phagocytosis of glioblastoma by M1 and M2 polarized macrophages and promotes M1 polarized macrophages in vivo. PLoS One 2016; 11: e0153550.
50. Zhang S, Guo M, Liu Q, Liu J, Cui Y. Neutrophil extracellular traps induce a hypercoagulable state in glioma. Immun Inflamm Dis 2021; 9: 1383-1393.
51. Zhang W, Ji L, Wang X, Zhu S, Luo J, Zhang Y, Tong Y, Feng F, Kang Y, Bi Q. Nomogram predicts risk and prognostic factors for bone metastasis of pancreatic cancer: a population-based analysis. Front Endocrinol (Lausanne) 2021; 12: 752176.
52. Zhang W, Qin T, Yang Z, Yin L, Zhao C, Feng L, Lin S, Liu B, Cheng S, Zhang K. Telomerase-positive circulating tumor cells are associated with poor prognosis via a neutrophil-mediated inflammatory immune environment in glioma. BMC Med 2021; 19: 277.
53. Zhang X, Ren L, Yan X, Shan Y, Liu L, Zhou J, Kuang Q, Li M, Long H, Lai W. Identification of immune-related lncRNAs in periodontitis reveals regulation network of gene-lncRNA-pathway-immunocyte. International immunopharmacology 2020; 84: 106600.
54. Zhang Y, Guo L, Dai Q, Shang B, Xiao T, Di X, Zhang K, Feng L, Shou J, Wang Y. A signature for pan-cancer prognosis based on neutrophil extracellular traps. J Immunother Cancer 2022; 10: e004210.
55. Zhao Z, Zhang KN, Wang Q, Li G, Zeng F, Zhang Y, Wu F, Chai R, Wang Z, Zhang C, Zhang W, Bao Z, Jiang T. Chinese Glioma Genome Atlas (CGGA): A comprehensive resource with functional genomic data from Chinese glioma patients. Genomics Proteomics Bioinformatics 2021; 19: 1-12.
56. Zhou Q, Yan X, Liu W, Yin W, Xu H, Cheng D, Jiang X, Ren C. Three immune-associated subtypes of diffuse glioma differ in immune infiltration, immune checkpoint molecules, and prognosis. Front Oncol 2020; 10: 586019.