A Hypoxia-related LncRNA Signature Predicts Survival of Primary Lower-grade glioma

Yanjia Hu
Sichuan University

Jing Zhang
Sichuan University

Jing Chen (chenjing6811@126.com)
Sichuan University https://orcid.org/0000-0002-3297-9183

Primary research

Keywords: hypoxia, long non-coding RNA, lower-grade glioma, prognosis, predictive model

Posted Date: March 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-335140/v1

Abstract

Background

Hypoxia-related long non-coding RNAs (lncRNAs) have been proven to play a role in multiple cancers and can serve as prognostic markers. Lower-grade gliomas (LGGs) are characterized by large heterogeneity.

Methods

This study aimed to construct a hypoxia-related lncRNA signature for predicting the prognosis of LGG patients. Transcriptome and clinical data of LGG patients were obtained from The Cancer Genome Atlas (TCGA) and the Chinese Glioma Genome Atlas (CGGA). LGG cohort in TCGA was chosen as training set and LGG cohorts in CGGA served as validation sets. A prognostic signature consisting of fourteen hypoxia-related lncRNAs was constructed using univariate and LASSO Cox regression. A risk score formula involving the fourteen lncRNAs was developed to calculate the risk score and patients were classified into high- and low-risk groups based on cutoff. Kaplan-Meier survival analysis was used to compare the survival between two groups. Cox regression analysis was used to determine whether risk score was an independent prognostic factor. A nomogram was then constructed based on independent prognostic factors and assessed by C-index and calibration plot. Gene set enrichment analysis and immune cell infiltration analysis were performed to uncover further mechanisms of this lncRNA signature.

Results

LGG patients with high risk had poorer prognosis than those with low risk in both training and validation sets. Recipient operating characteristic curves showed good performance of the prognostic signature. Univariate and multivariate Cox regression confirmed that the established lncRNA signature was an independent prognostic factor. C-index and calibration plots showed good predictive performance of nomogram. Gene set enrichment analysis showed that genes in the high-risk group were enriched in apoptosis, cell adhesion, pathways in cancer, hypoxia etc. Immune cells were higher in high-risk group.

Conclusion

The present study showed the value of the 14-lncRNA signature in predicting survival of LGGs and these 14 lncRNAs could be further investigated to reveal more mechanisms involved in gliomas.

Background

Lower grade gliomas account a high proportion in human central nervous system malignant tumors. Due to great intrinsic heterogeneity in biological behavior and clinical parameters, patients demonstrate different response to the standard-of-care and the prognosis differs dramatically.[1] Traditionally, WHO

grade can be used to prognosticate survival of patients. In 2016 brain tumor classification, the introduction of molecular alterations greatly improved the subclassification of LGG and the prediction of patient prognosis.[2] For example, oligodendroglioma is now defined by both isocitrate dehydrogenase (IDH) mutation and chromosome 1p and 19q (1p19q) codeletion, harboring a favorable prognosis. For other LGG, those with IDH mutation have better prognosis than those with wildtype IDH.[3] However, the established molecular characteristics are insufficient to stratify all LGG patients. For instance, Aibaidula et al.[4] found that IDH-wildtype LGG patients can be further stratified by alterations in EGFR, H3F3A, or TERT promoter. Yang et al.[5] found that IDH mutant LGG could be stratified for risk by CDKN2A, CDK4 and PDGFRA copy number alterations. Therefore, the development of novel prognostic signature in molecular level is helpful for better prognosticating the prognosis of LGG patients.

Long non-coding RNAs (lncRNAs) are non-protein-coding RNAs with at least 200 nucleotides.[6] There is increasing evidence revealing the oncogenic and tumor suppressive roles of lncRNAs in cancer development, progression, and metastasis.[7] Moreover, aberrant lncRNA expression profiles in clinical glioma specimens have been found to correlate with malignancy degree and histological differentiation, which have clinical implications in diagnosis of subclassification and prediction of prognosis.[8] Hypoxia plays a role in the malignant transformation of cells and subsequent tumor growth.[9] In multiple cancers, hypoxia is associated with metastasis,[10] tumor recurrence,[11] resistance to therapy,[12, 13] and tumor invasion. [14, 15] Previous studies have found that hypoxia-induced molecules could enhance the proliferation, renewal, malignant transition of glioma stem cells.[16–18] Thus, hypoxia-related molecules have the potential to assist diagnosis of glioma and improve risk stratification of patients. So far, many studies reported the construction of lncRNA signature or hypoxia-related gene signature to predicting prognosis of multiple cancers.[19–21] However, prognostic markers based on hypoxia-related lncRNA have not been reported in LGG.

In this study, we constructed a hypoxia-related lncRNA signature consisting of 14 lncRNAs to improve the ability to predict overall survival (OS) of LGG patients. We used external dataset to verify the prediction efficiency of the established lncRNA signature. Besides, a nomogram was constructed using risk score and other independent clinical characteristics. Finally, gene set enrichment analysis (GSEA) and immune cell infiltration analysis were performed to further uncover potential biological function and pathways. The 14-lncRNA signature demonstrated good performance in predicting prognosis of LGG patients. Moreover, the 14 lncRNAs and the results of GSEA are worthy of further investigation.

Materials And Methods

Acquisition and Processing of Data

The transcriptome and clinical data of LGG patients were downloaded from TCGA and CGGA. To construct and validate a hypoxia-related lncRNA signature, TCGA LGG cohort was chosen as training set and two LGG cohorts in CGGA served as validation sets. Hypoxia-related genes were acquired from MSigDB (http://www.gsea-msigdb.org/gsea/msigdb). Considering that LGGs in TCGA were primary, we

removed recurrent patients in CGGA cohorts. Moreover, patients with overall survival less than 30 days were also removed.

Identification of Immune-related LncRNAs

LncRNAs and hypoxia-related genes (HRGs) were identified respectively. Subsequently, hypoxia-related lncRNAs were identified using a Pearson correlation analysis between HRGs and lncRNA expression levels (Pearson correlation coefficient > 0.4, p<0.001). Meanwhile, hypoxia-related lncRNAs were also identified in validation sets from the CGGA.

Construction and Validation of a Hypoxia-related lncRNA Signature

Hypoxia-related lncRNAs in three cohorts (training set in TCGA and two validation sets in CGGA) were intersected to identify the overlapped hypoxia-related lncRNAs. To select the hypoxia-related lncRNAs for the construction of prognostic signature, a univariate Cox analysis were performed using R packages “survival” and “survminer”. In the univariate Cox analysis, hypoxia-related lncRNAs with a significant prognostic value were screened (p<0.01). Then, a LAASO Cox regression analysis was performed using R package “glmnt” to construct the prognostic model. The risk score based on the established prognostic model was calculated using the following formula: risk score = (coefficient $_{lncRNA1}$ × lncRNA1 expression) + (coefficient $_{lncRNA2}$× lncRNA2 expression) + …+ (coefficient $_{lncRNAn}$ × lncRNAn expression). The cutoff to classify the patients into high- and log-risk groups was the median of risk score in training set. Then, Kaplan-Meier survival curves were plotted using R package “survival” and overall survival of two groups was compared using log-rank test.

In order to assess the efficiency of the established hypoxia-related lncRNA signature, recipient operating characteristic (ROC) curves predicting 1-, 3-, 5-, and 7-year survival were plotted. Also, prediction efficiency of the lncRNA signature was compared with other clinical parameters by plotting ROC curves. Meanwhile, area under the curve (AUC) was calculated. Furthermore, multivariate Cox regression analysis was performed to determine whether risk score could be an independent prognostic factor. Risk score, age, gender, tumor grade, status of IDH, status of 1p19q (codeletion or not), chemotherapy, and radiotherapy were analyzed in multivariate Cox regression analysis. We also compared clinical parameters of patients between high- and low-risk groups.

Risk scores were calculated for LGG patients in validations sets (LGG cohorts from CGGA) per the same formula. Meanwhile, patients were classified into high- and low-risk groups based on the cutoff. Kaplan-Meier survival curves were generated and survival was compared using log-rank test. Subsequently, multivariate Cox regression analysis, ROC analysis, and comparison of clinical parameters between two groups were also performed to assess the efficiency of the established hypoxia-related lncRNA signature.

The Construction and Evaluation of Nomogram

A nomogram was constructed according to the independent prognostic factors in training set. The concordance index (C-index) and calibration plot were used to assess the calibration and the discrimination of the nomogram. In the nomogram, survival probability of 1-, 3-, 5-, and 7-year was displayed. Calibration plots assessing 1-, 3-, 5-, and 7-year survival prediction were generated. Moreover, LGG cohorts in CGGA were also utilized to calculate C-index and generate calibration plot to perform external verification.

Gene Set Enrichment Analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) enrichment analysis was performed on the differentially expressed genes of the high-risk and low-risk groups. KEGG gene collection file (C2.all.v6.2.symbols.gmt) and the GO functional gene collection file (C5.all.v6.2.symbols.gmt) were downloaded from GSEA software (version 4.0.0, downloaded from http://software.broadinstitute.org/gsea/index.jsp). Hallmark annotation was also performed on the differentially expressed genes between two groups (h.all.v7.2.symbols.gmt). A nominal p-value of < 0.05 and an FDR of < 0.05 was considered significant.

Analysis of Immune Cell Infiltration

For TCGA LGG cohort, the file used for the estimation of immune cell infiltration was downloaded from TIMER database (http://timer.comp-genomics.org/). Infiltration data of B cells, CD 4^+ T cells, $CD8^+$ T cells, macrophages, dendritic cells, and neutrophils were used to determine the correlation between risk score and immune infiltration (Pearson correlation analysis). In addition, immune cell infiltration was compared between high- and low-risk groups.

Statistical Analysis

The statistical analysis was performed using R software (version 3.6.3 and version 4.0.4). The correlation was determined by Pearson correlation analysis. Chi-square test, Fisher's Exact test, and t test were used to compare clinical variables between groups. Survival analysis was performed using Kaplan-Meier method and survival was compared by the log-rank test. All statistical tests were two-sided and a p value of <0.05 was considered statistically significant except special declaration.

Results

Cohort information and patient characteristic

In the present study, LGG cohort from TCGA was chosen as the training set. In total, 476 patients with survival > 30 days were included to screen prognosis-associated immune-related lncRNAs and construct prognostic model. However, 365 patients with compete clinical characteristics were used to perform multivariate Cox regression analysis to identify the independent variables. These clinical characteristics included age, gender, grade, chemotherapy, radiotherapy, status of IDH, and status of chromosome 1p

and 19q (1p19q). As for validation sets, there were two LGG cohorts in CGGA which were termed cohort 1 and cohort 2 here. There were 125 patients eligible for survival analysis and multiple Cox regression analysis in cohort 1. In cohort 2, 201 patients with complete clinical characteristics were included to perform the forementioned analyses. The cohort information and patient characteristics of three cohorts are detailed in Table 1.

Construction of the Hypoxia-related LncRNA Prognostic Signature

In total, 980 hypoxia-related lncRNAs were identified in training set. Meanwhile, 1046 and 1063 lncRNAs were identified in cohort 1 and cohort 2, respectively. Thus, 241 hypoxia-related lncRNAs were intersected in three cohorts (Figure 1A). Subsequently, 120 prognosis-specific hypoxia-related lncRNAs were screened using univariate Cox regression analysis. To establish a hypoxia-related lncRNA prognostic signature, LASSO method was utilized on the 120 prognosis-related lncRNAs. Ultimately, 14 lncRNAs were selected to construct the prognostic signature (Figure 1B). The formula calculating risk score of patients was as follow: risk score = MYLK-AS1 × (-0.0194) + FAM181A-AS1 × 0.1775 + H19 × 0.0185 + DGCR9 × (-0.0192) + LINC00641× (-0.0072) + CRNDE × 0.1336 + FAM66C × (-0.01850) + SLC25A21-AS1 × (-0.0148) + SNHG18 × 0.1931 + NR2F1-AS1 × (-0.0118) + WAC-AS1 × (-0.1600) + PAXIP1-AS2 × 0.1318 + WDFY3-AS2 × (-0.0319) + TMEM254-AS1 × (-0.1876). Among the included lncRNAs, 9 (MYLK-AS1, DGCR9, LINC00641, FAM66C, SLC25A21-AS1, NR2F1-AS1, WAC-AS1, WDFY3-AS2, and TMEM254-AS1) were prognostic protective factors, while the other 5 (FAM181A-AS1, H19, CRNDE, SNHG18, and PAXIP1-AS2) were prognostic risk factors. Based on the formula, risk score was calculated for every patient in training set and patients were divided into high- and low-risk groups according to the median of risk scores (-1.244). Next, Kaplan-Meier method was used to plot survival curve and survival of two groups was compared using log-rank test (Figure 2A). The result showed that patients in low-risk group had better prognosis (median overall survival, 11.59 vs. 4.06 years, $p<0.001$). Besides, risk score of patients, correlation of risk score with survival status, and heat map of 14 lncRNAs were plotted (Figure 2B-D). In order to assess the prognostic accuracy of this model, ROC analysis was performed and AUC was calculated. As shown in Figure 2E, AUC of ROC predicting 1-, 3-, 5-, and 7-year survival were 0.904, 0.845, 0.763, and 0.774, respectively. Moreover, ROC assessing the prognostic accuracy of clinical characteristics showed that risk score had the highest AUC (Figure 2F). Univariate and multivariate Cox regression analyses were performed to determine whether the established lncRNA signature was an independent prognostic factor (Figure 3). Univariate Cox regression analysis showed that tumor grade, age, status of IDH, status of codeletion in 1p19q, chemotherapy, radiotherapy, and risk score were eligible for multivariate Cox analysis. Subsequently, multivariate Cox regression analysis showed grade, age, and risk score (hazard ratio: 4.253, 95% confidence interval: 2.714-6.664) were independent prognostic factors. Next, we compared risk scores between high- and low-risk groups to better elucidate the correlation between clinical characteristics and risk score. As shown in Figure 4, all clinical characteristics except gender were found to harbor different risk scores. Specifically speaking, patients with older age (> 50), higher grade (WHO □), chemotherapy, radiotherapy, mutant IDH, and codeletion in 1p19q had higher risk scores. Heatmap describing the distribution of clinical characteristics in two risk groups is shown in Figure 4H.

Validation of the Hypoxia-Related LncRNA Signature in TCGA Cohorts

To verify the prognostic accuracy of the established lncRNA signature in external dataset, LGG cohorts in the CGGA were acquired. Patients were divided into high- and low-risk groups according to the same formula and cutoff. Kaplan-Meier survival analysis (Figure 5A, B) showed that low-risk group had better prognosis in both cohorts (median survival, cohort 1: 8.52 vs. 2.80 years, p<0.001; cohort 2: 8.17 vs 3.30 years, p<0,001). Risk score of patients, correlation of risk score with survival status, and heat map of 14 lncRNAs were also plotted (Figure 5C-H). ROC analysis showed that the lncRNA signature had good prognostic performance in both cohorts (Supplementary Figure S1A, B). In cohort 1, AUC of ROC predicting 1-, 3-, 5-, and 7-year survival were 0.864, 0.845, 0.817, and 0.798, respectively. In cohort 2, AUC of ROC predicting 1-, 3-, 5-, and 7-year survival were 0.797, 0.809, 0.722, and 0.706, respectively. Moreover, ROC analysis incorporating risk score and clinical characteristics showed that risk score harbored the best prognostic accuracy in both cohorts (Supplementary Figure S1C, D). Next, univariate and multivariate Cox regression analyses showed that the lncRNA signature was an independent prognostic factor in cohort 1 and 2 (Supplementary Figure S2). In cohort 1, higher age, grade LGG, chemotherapy, wildtype IDH, and non-codeletion in 1p19q were associated with higher risk score than their respective counterparts (Supplementary Figure S3). However, in cohort 2, we only found wildtype IDH, grade III LGG, and non-codeletion in 1p19q were associated with higher risk score (Supplementary Figure S4).

Construction and evaluation of a nomogram

We used independent prognostic factors in training set to establish a nomogram predicting 1-, 3-, 5-, and 7-year survival of patients. Every factor (patient age, tumor grade, and risk score) was used to obtain the corresponding score summary and the total score was calculated to prognosticate the survival probability of every patient (Figure 6A). Higher score denoted poorer prognosis. The C-index of the established nomogram was 0.854 (95% confidence interval: 0.811-0.897). Moreover, calibration curves predicting 1-, 3-, 5-, and 7-year survival were used to indicate the consistency between the actual observed survival probability and that predicted by the nomogram. As shown in Figure 6B-E, the calibration curves indicated a good fit for the nomogram.

We next assessed the prognostic accuracy of the established nomogram in cohort 1 and 2. In cohort 1, the C-index was 0.765 (95% confidence interval: 0.703-0.827). Moreover, calibration curves also indicated a good fit for the nomogram in predicting 1-, 3-, 5-, and 7-year survival (Supplementary Figure S5). In cohort 2, the C-index was 0.787 (95% confidence interval: 0.730-0.844). However, the calibration curves showed that the consistence between the actual survival probability and that predicted by the nomogram was not as good as the consistence observed in TCGA LGG cohort and cohort 1 (Supplementary Figure S6).

Gene Set Enrichment Analysis

With regard to KEGG pathways, GSEA showed that genes in high-risk group were enriched in apoptosis, cell adhesion molecules CAMs, chemokine signaling pathway, ECM receptor interaction, focal adhesion,

JAK STAT signaling pathway, pathways in cancer, toll like receptor signaling pathway, VEGF signaling pathway, etc (Figure 7). No significant KEGG pathways were found in genes in low-risk group. The GO analysis included the biological process (BP), molecular function (MF), and cellular component (CC) categories. The BP category of genes in high-risk group included leukocyte proliferation, lymphocyte activation involved in immune response, negative regulation of interleukin production, etc. No significant BP category was enriched in genes in low-risk group. The MF category of genes in high-risk group included carboxypeptidase activity, growth factor binding, cytokine binding, etc. With regard to genes in low-risk group, enriched MF was ubiquitin-like protein-specific protease activity. The CC category of genes in high-risk group included vacuolar lumen, vesicle lumen, endoplasmic reticulum, etc. The genes in low-risk group were enriched in vesicle tethering complex regarding CC category. The respective top 10 KEGG pathways and GO terms in high-risk group are displayed in Supplementary Figure S7A-D. In order to investigate tumor-associated signaling pathway, we also performed GSEA with regard to hallmarks. The results showed that tumor-associated signaling pathways enriched in high-risk group included IL2-STAT signaling, PI3K-AKT MTOR signaling, epithelial mesenchymal transition, IL6-JAK-STAT3 signaling, p53 pathway, TNFα signaling via NFκB, etc (Figure 8). However, no significant hallmark was enriched in low-risk group. The top 10 hallmark terms are showed in Supplementary S7E.

Immune Cell Infiltration

In GSEA, we found that BP terms enriched in high-risk group were mainly involved in immune cells. Thus, we performed analysis of immune cell infiltration between high- and low-risk groups. The correlation coefficients of risk scores between B cell, CD4+ T cell, CD8+ T cell, dendritic cell, neutrophil and macrophage were 0.336, 0.385, 0.229, 0.624, 0.452 and 0.487, respectively (Figure 9). In addition, abundances of B cell, CD4+ T cell, CD8+, dendritic cell, neutrophil and macrophage were higher in high-risk group than those in low-risk group (Supplementary Figure S8).

Discussion

Lower-grade gliomas are characterized by large intra-tumoral and inter-tumoral heterogeneity. Although the establishment of standard treatment, the prognosis of patients differed dramatically, ranging from 1 to 15 years. Moreover, current biomarkers and molecular alterations are insufficient to perform risk stratification for LGG patients. In this study, we developed a prognostic signature consisting of 14 hypoxia-related lncRNAs to predict OS for LGG patients. ROC analysis and univariate/multivariate Cox regression were used to evaluate the predictive accuracy of the established lncRNA signature. The results showed that patients in high-risk group had more dismal prognosis than those in low-risk group. Moreover, risk score calculated by our prognostic model was an independent prognostic factor. To comprehensively predict OS for LGG patients, we constructed a nomogram combining risk score and other clinical parameters. C-index and calibration plots showed good consistence between actual survival and predicted survival probability. Next, we performed GSEA to investigate KEGG, GO and Hallmark analyses on gene in high- and low-risk groups. The analyses identified several tumor-related biological processes or signaling pathways and immune-related GO terms.

LncRNAs regulate several biological processes including chromatin modification, gene transcription, posttranscriptional mRNA processing and nuclear-cytoplasmic trafficking.[22] In pathological conditions, a number of lncRNAs have been found to be aberrantly expressed in multiple cancers and play a part in tumor proliferation, apoptosis, metastasis, and resistance to treatment.[23–25] Tumor hypoxia is a common feature of many cancers and many mechanisms driven by hypoxia-inducible factor-1 (HIF-1) regulate multiple aspects of tumorigenesis and progression.[26–28] There are increasing evidences suggesting the involvement of hypoxia-related lncRNAs in tumor/cancer-associated biological processes and behaviors. Choudhery et al. found that NEAT1 was upregulated in breast cancer and could improve cancer cell survival in hypoxic environment.[29] Ferdin et al. found that HINCUT1 supported proliferation of cancer cell under hypoxia in colorectal carcinoma and breast cancer.[30] In gliomas, Li et al. found Linc01060 was upregulated in glioma and was significantly correlated with tumor grade and poor clinical prognosis.[31] Linc01060 activates prooncogenic signaling pathways in glioma cells to promote tumor progression. Mineo et al. found that HIF1A-AS2 was upregulated in mesenchymal glioblastoma stem-like cell and facilitated the maintenance of mesenchymal glioblastoma stem-like cells in hypoxic niches.[32] Besides, NF-kappa B interacting long noncoding RNA was showed to increase the expression level of HIF-1α and the activity of the hypoxia pathways in glioma.[33] According to these findings, we could conclude that hypoxia-related lncRNAs play a role in occurrence and progression of multiple cancers, including glioma. Thus, a prognostic signature consisting of hypoxia-related lncRNA may improve the prediction of prognosis in LGG patients and identify lncRNAs which can be targeted.

There were 14 hypoxia-related lncRNAs in our prognostic signature. Among the 14 lncRNAs, 5 were risk factors, including FAM181A-AS1, H19, CRNDE, SNHG18, and PAXIP1-AS2. The other 9 lncRNAs were protective factors, including MYLK-AS1, DGCR9, LINC00641, FAM66C, SLC25A21-AS1, NR2F1-AS1, WAC-AS1, WDFY3-AS2, and TMEM254-AS1. The study by Jiang et al. showed that FAM181A-AS1 promoted gliomagenesis by enhancing ZRANB2 expression by sponging of miR129-5p.[34] The role of H19 in tumorigenesis and/or progression has been elucidated in multiple cancers, including endometrial carcinoma,[35] ovarian cancer,[36] breast cancer,[37] acute myeloid leukemia,[38] etc. In glioma, H19 was showed to be involved in temozolomide resistance and methylguanine DNA methyltransferase expression.[39] Besides, Jia et al. showed that H19 could improve glioma angiogenesis.[40] CRNDE was showed to promote colorectal cancer cell proliferation and chemoresistance and attenuate chemoresistance in gastric cancer.[41, 42] Furthermore, this lncRNA was also found to promote malignant progression of glioma.[43] SNHG18 was showed to promote radioresistance of glioma and drive the growth and metastasis of non-small cell lung cancer.[44, 45] As for PAXIP1-AS2, no studies investigate its role in tumor or disease. In prognostic protective lncRNAs, MYLK-AS1, LINC00641, NR2F1-AS1, and WDFY3-AS2 were found to regulate multiple tumor-related biological processes in many cancers. For instance, NR2F1-AS1 demonstrated pro-tumoral functions in hepatocellular carcinoma,[46] breast cancer,[47] papillary thyroid carcinoma,[48] esophageal squamous cell carcinoma,[49] neuroblastoma,[50] and non-small cell lung cancer.[51] In addition, the investigation to DGCR9, FAM66C, SLC25A21-AS1, WAC-AS1, and TMEM254-AS1 was insufficient.[52, 53] It is noteworthy that all the 9 lncRNAs have not been reported to regulate tumorigenesis or progression of gliomas. In addition,

although the 9 lncRNAs were protective roles in our prognostic signature, they could be pro-tumoral in other cancers. Consequently, further investigation is warranted to elucidate the relationship between these hypoxia-related lncRNAs and gliomas.

In the present study, LGG cohort from TCGA was selected as the training set, while LGG cohorts from CGGA served as validation sets. Survival analysis showed that high-risk group had poorer prognosis than low-risk group in both training and validation sets. With regard to multivariate Cox regression analysis, results showed that risk score was an independent prognostic factor in all cohorts. However, one should note that hazard ratios of risk score were notably different between training and validation sets. In addition, calibrations plots of nomogram also showed obvious difference, especially between training set and cohort 2. The main reason was the heterogeneity between three cohorts.

To identify potential tumor-associated biological process and signaling pathway, we performed GSEA. We found that some oncological hallmarks, including PI3K-AKT MTOR signaling, epithelial mesenchymal transition, JAK-STAT3 signaling, p53 pathway, TNFα signaling via NFκB, were enriched in high-risk group. These oncological hallmarks have been reported to be involved in glioma, indicating potential research value of our lncRNA signature. Moreover, we also found that that BP terms enriched in high-risk group were mainly associated with immune. Therefore, immune cell infiltration analysis was performed and the result showed that the high-risk group had a higher infiltration proportion of B cell, macrophage, T cell, dendritic cell, and neutrophil. Moreover, Pearson correlation analysis showed positive correlation between risk score and immune cells. These evidences underscored the immunotherapy for LGG patients with high risk of hypoxia-related lncRNA signature.

There are some limitations existing in our study. First, the racial composition between training set and validation sets was different, causing distinction in gene expression. Second, differences in clinical characteristic also compromised the accuracy of prediction in external validation. Third, the results in our study were based on retrospective data in LGG. Thus, conclusions in our study need to be further validated in prospective research.

Conclusion

In this study, we constructed a prognostic signature consisting of 14 hypoxia-related lncRNAs to improve the prediction of prognosis for LGG patients. Relevant analyses indicated that the 14 lncRNAs could be potential biomarkers and targets in the diagnosis and treatment of LGG. Further in vitro and in vivo experiments are warranted to explore the underlying mechanisms behind hypoxia-related lncRNA and survival outcome in LGG.

Abbreviations

lncRNA: long non-coding RNA; LGG, lower-grade glioma; TCGA: The Cancer Genome Atlas; CGGA: Chinese Glioma Genome Atlas; IDH: isocitrate dehydrogenase; 1p19q, chromosome 1p and 19q; OS: overall

survival; GSEA: gene set enrichment analysis; HRG: hypoxia-related gene; ROC: recipient operation characteristic; AUC: are under the curve; C-index: concordance index; KEGG: The Kyoto Encyclopedia of Genes and Genomes; GO: gene ontology; BP: biological process; CC: cellular component; MF: molecular function; VEGF: vascular endothelial growth factor.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Hu Y contributed to conception, data analysis, and writing of this study. Zhang J contributed to data analysis, revising and writing. Chen J supervised the conduction of this study. All authors read and approved the final manuscript.

Funding

This study was supported by the Support Project Funding of Science and Technology Department of Sichuan Province under Grant Number 0040205302256.

Availability of data and materials

Some publicly available datasets were analyzed in this study. The transcriptome and clinical data of LGG patients were obtained from The Cancer Genome Atlas (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) and Chinese Glioma Genome Atlas (http://www.cgga.org.cn/). Hypoxia-related genes were obtained from MSigDB (https://www.gsea-msigdb.org/gsea/index.jsp). The file used for the estimation of immune cell infiltration was downloaded from TIMER database (http://timer.comp-genomics.org/). Other data generated in analysis process were also available from the corresponding author upon reasonable request.

Ethics approval and consent to participates

Not applicable.

Consent for publication

All authors have read the manuscript and agreed the publication of this study.

Competing interests

The authors declare that there was no conflict of interests.

References

1. Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, Rheinbay E, Miller CR, Vitucci M, Morozova O *et al*: **Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas**. *The New England journal of medicine* 2015, **372**(26):2481-2498.
2. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW: **The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary**. *Acta neuropathologica* 2016, **131**(6):803-820.
3. Tom MC, Cahill DP, Buckner JC, Dietrich J, Parsons MW, Yu JS: **Management for Different Glioma Subtypes: Are All Low-Grade Gliomas Created Equal?***American Society of Clinical Oncology educational book American Society of Clinical Oncology Annual Meeting* 2019, **39**:133-145.
4. Aibaidula A, Chan AK, Shi Z, Li Y, Zhang R, Yang R, Li KK, Chung NY, Yao Y, Zhou L *et al*: **Adult IDH wild-type lower-grade gliomas should be further stratified**. *Neuro-oncology* 2017, **19**(10):1327-1337.
5. Yang RR, Shi ZF, Zhang ZY, Chan AK, Aibaidula A, Wang WW, Kwan JSH, Poon WS, Chen H, Li WC *et al*: **IDH mutant lower grade (WHO Grades II/III) astrocytomas can be stratified for risk by CDKN2A, CDK4 and PDGFRA copy number alterations**. *Brain pathology (Zurich, Switzerland)* 2020, **30**(3):541-553.
6. Ponting CP, Oliver PL, Reik W: **Evolution and functions of long noncoding RNAs**. *Cell* 2009, **136**(4):629-641.
7. Zhou M, Zhang Z, Zhao H, Bao S, Cheng L, Sun J: **An Immune-Related Six-lncRNA Signature to Improve Prognosis Prediction of Glioblastoma Multiforme**. *Molecular neurobiology* 2018, **55**(5):3684-3697.
8. Peng Z, Liu C, Wu M: **New insights into long noncoding RNAs and their roles in glioma**. *Molecular cancer* 2018, **17**(1):61.
9. Brat DJ, Mapstone TB: **Malignant glioma physiology: cellular response to hypoxia and its role in tumor progression**. *Annals of internal medicine* 2003, **138**(8):659-668.
10. Gilkes DM, Semenza GL, Wirtz D: **Hypoxia and the extracellular matrix: drivers of tumour metastasis**. *Nature reviews Cancer* 2014, **14**(6):430-439.
11. Liu M, Zhong J, Zeng Z, Huang K, Ye Z, Deng S, Chen H, Xu F, Li Q, Zhao G: **Hypoxia-induced feedback of HIF-1α and lncRNA-CF129 contributes to pancreatic cancer progression through stabilization of p53 protein**. *Theranostics* 2019, **9**(16):4795-4810.
12. Brahimi-Horn MC, Chiche J, Pouysségur J: **Hypoxia and cancer**. *Journal of molecular medicine (Berlin, Germany)* 2007, **85**(12):1301-1307.
13. Walsh JC, Lebedev A, Aten E, Madsen K, Marciano L, Kolb HC: **The clinical importance of assessing tumor hypoxia: relationship of tumor hypoxia to prognosis and therapeutic opportunities**. *Antioxidants & redox signaling* 2014, **21**(10):1516-1554.
14. Man J, Yu X, Huang H, Zhou W, Xiang C, Huang H, Miele L, Liu Z, Bebek G, Bao S *et al*: **Hypoxic Induction of Vasorin Regulates Notch1 Turnover to Maintain Glioma Stem-like Cells**. *Cell stem cell*

2018, **22**(1):104-118.e106.

15. Kaur B, Khwaja FW, Severson EA, Matheny SL, Brat DJ, Van Meir EG: **Hypoxia and the hypoxia-inducible-factor pathway in glioma growth and angiogenesis**. *Neuro-oncology* 2005, **7**(2):134-153.
16. Almiron Bonnin DA, Havrda MC, Lee MC, Liu H, Zhang Z, Nguyen LN, Harrington LX, Hassanpour S, Cheng C, Israel MA: **Secretion-mediated STAT3 activation promotes self-renewal of glioma stem-like cells during hypoxia**. *Oncogene* 2018, **37**(8):1107-1118.
17. Soeda A, Park M, Lee D, Mintz A, Androutsellis-Theotokis A, McKay RD, Engh J, Iwama T, Kunisada T, Kassam AB *et al*: **Hypoxia promotes expansion of the CD133-positive glioma stem cells through activation of HIF-1alpha**. *Oncogene* 2009, **28**(45):3949-3959.
18. Shi QY, Zhang SJ, Liu L, Chen QS, Yu LN, Zhang FJ, Yan M: **Sevoflurane promotes the expansion of glioma stem cells through activation of hypoxia-inducible factors in vitro**. *British journal of anaesthesia* 2015, **114**(5):825-830.
19. Mao Y, Fu Z, Zhang Y, Dong L, Zhang Y, Zhang Q, Li X, Liu J: **A seven-lncRNA signature predicts overall survival in esophageal squamous cell carcinoma**. *Scientific reports* 2018, **8**(1):8823.
20. Li J, Wang W, Xia P, Wan L, Zhang L, Yu L, Wang L, Chen X, Xiao Y, Xu C: **Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer**. *International journal of cancer* 2018, **143**(9):2150-2160.
21. Sun J, Zhao T, Zhao D, Qi X, Bao X, Shi R, Su C: **Development and validation of a hypoxia-related gene signature to predict overall survival in early-stage lung adenocarcinoma patients**. *Therapeutic advances in medical oncology* 2020, **12**:1758835920937904.
22. Chang YN, Zhang K, Hu ZM, Qi HX, Shi ZM, Han XH, Han YW, Hong W: **Hypoxia-regulated lncRNAs in cancer**. *Gene* 2016, **575**(1):1-8.
23. Hu Y, Lin J, Fang H, Fang J, Li C, Chen W, Liu S, Ondrejka S, Gong Z, Reu F *et al*: **Targeting the MALAT1/PARP1/LIG3 complex induces DNA damage and apoptosis in multiple myeloma**. *Leukemia* 2018, **32**(10):2250-2262.
24. Loewen G, Jayawickramarajah J, Zhuo Y, Shan B: **Functions of lncRNA HOTAIR in lung cancer**. *J Hematol Oncol* 2014, **7**:90.
25. Ren J, Ding L, Zhang D, Shi G, Xu Q, Shen S, Wang Y, Wang T, Hou Y: **Carcinoma-associated fibroblasts promote the stemness and chemoresistance of colorectal cancer by transferring exosomal lncRNA H19**. *Theranostics* 2018, **8**(14):3932-3948.
26. Semenza GL: **Targeting HIF-1 for cancer therapy**. *Nature reviews Cancer* 2003, **3**(10):721-732.
27. Zhao S, Lin Y, Xu W, Jiang W, Zha Z, Wang P, Yu W, Li Z, Gong L, Peng Y *et al*: **Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha**. *Science (New York, NY)* 2009, **324**(5924):261-265.
28. Kimbro KS, Simons JW: **Hypoxia-inducible factor-1 in human breast and prostate cancer**. *Endocrine-related cancer* 2006, **13**(3):739-749.

29. Choudhry H, Albukhari A, Morotti M, Haider S, Moralli D, Smythies J, Schödel J, Green CM, Camps C, Buffa F *et al*: **Tumor hypoxia induces nuclear paraspeckle formation through HIF-2α dependent transcriptional activation of NEAT1 leading to cancer cell survival**. *Oncogene* 2015, **34**(34):4482-4490.
30. Ferdin J, Nishida N, Wu X, Nicoloso MS, Shah MY, Devlin C, Ling H, Shimizu M, Kumar K, Cortez MA *et al*: **HINCUTs in cancer: hypoxia-induced noncoding ultraconserved transcripts**. *Cell death and differentiation* 2013, **20**(12):1675-1687.
31. Li J, Liao T, Liu H, Yuan H, Ouyang T, Wang J, Chai S, Li J, Chen J, Li X *et al*: **Hypoxic Glioma Stem Cell-Derived Exosomes Containing Linc01060 Promote Progression of Glioma by Regulating the MZF1/c-Myc/HIF1α Axis**. *Cancer research* 2021, **81**(1):114-128.
32. Mineo M, Ricklefs F, Rooj AK, Lyons SM, Ivanov P, Ansari KI, Nakano I, Chiocca EA, Godlewski J, Bronisz A: **The Long Non-coding RNA HIF1A-AS2 Facilitates the Maintenance of Mesenchymal Glioblastoma Stem-like Cells in Hypoxic Niches**. *Cell reports* 2016, **15**(11):2500-2509.
33. Chen Z, Li S, Shen L, Wei X, Zhu H, Wang X, Yang M, Zheng X: **NF-kappa B interacting long noncoding RNA enhances the Warburg effect and angiogenesis and is associated with decreased survival of patients with gliomas**. *Cell death & disease* 2020, **11**(5):323.
34. Jiang X, Chen D: **LncRNA FAM181A-AS1 promotes gliomagenesis by sponging miR-129-5p and upregulating ZRANB2**. *Aging* 2020, **12**(20):20069-20084.
35. Zhang L, Wang DL, Yu P: **LncRNA H19 regulates the expression of its target gene HOXA10 in endometrial carcinoma through competing with miR-612**. *European review for medical and pharmacological sciences* 2018, **22**(15):4820-4827.
36. Medrzycki M, Zhang Y, Zhang W, Cao K, Pan C, Lailler N, McDonald JF, Bouhassira EE, Fan Y: **Histone h1.3 suppresses h19 noncoding RNA expression and cell growth of ovarian cancer cells**. *Cancer research* 2014, **74**(22):6463-6473.
37. Peng F, Li TT, Wang KL, Xiao GQ, Wang JH, Zhao HD, Kang ZJ, Fan WJ, Zhu LL, Li M *et al*: **H19/let-7/LIN28 reciprocal negative regulatory circuit promotes breast cancer stem cell maintenance**. *Cell death & disease* 2017, **8**(1):e2569.
38. Zhang TJ, Zhou JD, Zhang W, Lin J, Ma JC, Wen XM, Yuan Q, Li XX, Xu ZJ, Qian J: **H19 overexpression promotes leukemogenesis and predicts unfavorable prognosis in acute myeloid leukemia**. *Clinical epigenetics* 2018, **10**:47.
39. Nie E, Jin X, Miao F, Yu T, Zhi T, Shi Z, Wang Y, Zhang J, Xie M, You Y: **TGF-β1 modulates temozolomide resistance in glioblastoma via altered microRNA processing and elevated MGMT**. *Neuro-oncology* 2020.
40. Jia P, Cai H, Liu X, Chen J, Ma J, Wang P, Liu Y, Zheng J, Xue Y: **Long non-coding RNA H19 regulates glioma angiogenesis and the biological behavior of glioma-associated endothelial cells by inhibiting microRNA-29a**. *Cancer letters* 2016, **381**(2):359-369.
41. Han P, Li JW, Zhang BM, Lv JC, Li YM, Gu XY, Yu ZW, Jia YH, Bai XF, Li L *et al*: **The lncRNA CRNDE promotes colorectal cancer cell proliferation and chemoresistance via miR-181a-5p-mediated**

regulation of Wnt/β-catenin signaling. *Molecular cancer* 2017, **16**(1):9.

42. Zhang F, Wang H, Yu J, Yao X, Yang S, Li W, Xu L, Zhao L: **LncRNA CRNDE attenuates chemoresistance in gastric cancer via SRSF6-regulated alternative splicing of PICALM**. *Molecular cancer* 2021, **20**(1):6.
43. Zheng J, Liu X, Wang P, Xue Y, Ma J, Qu C, Liu Y: **CRNDE Promotes Malignant Progression of Glioma by Attenuating miR-384/PIWIL4/STAT3 Axis**. *Molecular therapy : the journal of the American Society of Gene Therapy* 2016, **24**(7):1199-1215.
44. Zheng R, Yao Q, Ren C, Liu Y, Yang H, Xie G, Du S, Yang K, Yuan Y: **Upregulation of Long Noncoding RNA Small Nucleolar RNA Host Gene 18 Promotes Radioresistance of Glioma by Repressing Semaphorin 5A**. *International journal of radiation oncology, biology, physics* 2016, **96**(4):877-887.
45. Fan H, Yuan J, Li Y, Jia Y, Li J, Wang X, Li X: **MKL1-induced lncRNA SNHG18 drives the growth and metastasis of non-small cell lung cancer via the miR-211-5p/BRD4 axis**. *Cell death & disease* 2021, **12**(1):128.
46. Huang H, Chen J, Ding CM, Jin X, Jia ZM, Peng J: **LncRNA NR2F1-AS1 regulates hepatocellular carcinoma oxaliplatin resistance by targeting ABCC1 via miR-363**. *Journal of cellular and molecular medicine* 2018, **22**(6):3238-3245.
47. Zhang Q, Li T, Wang Z, Kuang X, Shao N, Lin Y: **lncRNA NR2F1-AS1 promotes breast cancer angiogenesis through activating IGF-1/IGF-1R/ERK pathway**. *Journal of cellular and molecular medicine* 2020, **24**(14):8236-8247.
48. Yang C, Liu Z, Chang X, Xu W, Gong J, Chai F, Cui D: **NR2F1-AS1 regulated miR-423-5p/SOX12 to promote proliferation and invasion of papillary thyroid carcinoma**. *Journal of cellular biochemistry* 2020, **121**(2):2009-2018.
49. Zhang Y, Zheng A, Xu R, Zhou F, Hao A, Yang H, Yang P: **NR2F1-induced NR2F1-AS1 promotes esophageal squamous cell carcinoma progression via activating Hedgehog signaling pathway**. *Biochemical and biophysical research communications* 2019, **519**(3):497-504.
50. Liu L, Zhao H, He HH, Huang J, Xu YY, Li XL, Wu Y, Wang J, Zhu YB: **Long non-coding RNA NR2F1-AS1 promoted neuroblastoma progression through miR-493-5p/TRIM2 axis**. *European review for medical and pharmacological sciences* 2020, **24**(24):12748-12756.
51. Zhang C, Wu S, Song R, Liu C: **Long noncoding RNA NR2F1-AS1 promotes the malignancy of non-small cell lung cancer via sponging microRNA-493-5p and thereby increasing ITGB1 expression**. *Aging* 2020, **12**.
52. Ni C, Yang P, Guo J, Ye M: **Role of DiGeorge syndrome critical region gene 9, a long noncoding RNA, in gastric cancer**. *OncoTargets and therapy* 2018, **11**:2259-2267.
53. Xie Y, Gu J, Qin Z, Ren Z, Wang Y, Shi H, Chen B: **Long non-coding RNA FAM66C is associated with clinical progression and promotes cell proliferation by inhibiting proteasome pathway in prostate cancer**. *Cell biochemistry and function* 2020, **38**(8):1006-1016.

Tables

Table 1 Summary of risk scores and clinical pathological characteristics of training and validation cohorts

Variables	Training cohort (n=365)	Validation 1 (n=125)	Validation 2 (n=201)	p
Age (y)				
<=50	256 (70.1%)	103 (82.4%)	173 (86.1%)	<0.001
>50	109 (29.9%)	22 (17.6%)	28 (13.9%)	
Gender				
Male	198 (54.2%)	76 (60.8%)	123 (58.3%)	0.374
Female	167 (45.8%)	49 (39.2%)	88 (41.7%)	
Grade				
	172 (47.1%)	82 (65.6%)	103 (51.2%)	0.002
	193 (52.9%)	43 (34.4%)	98 (48.8%)	
Radiotherapy				
Yes	230 (63.0%)	117 (93.6%)	156 (77.6%)	<0.001
No	135 (37.0%)	8 (6.4%)	45 (22.4%)	
Chemotherapy				
Yes	232 (63.6%)	59 (47.2%)	126 (62.7%)	0.004
No	133 (36.4%)	66 (52.8%)	75 (37.3%)	
IDH status				
Wildtype	60 (16.4%)	34 (27.2%)	51 (25.4%)	0.007
Mutation	305 (83.6%)	91 (72.8%)	150 (74.6%)	
1p19q status				
Non-codeletion	240 (65.8%)	78 (62.4%)	139 (69.2%)	0.445
Codeletion	125 (34.2%)	47 (37.6%)	62 (30.8%)	
Survival	2.08 0.10-17.60	6.68 0.27-11.41	3.75 0.30-11.16	<0.001
Risk score				
Low risk	183 (50.1%)	79 (63.2%)	130 (64.7%)	0.001
High risk	182 (49.9%)	46 (36.8%)	71 (35.3%)	

Supplement

Supplementary Figures are not available in this version of the manuscript.

Figures

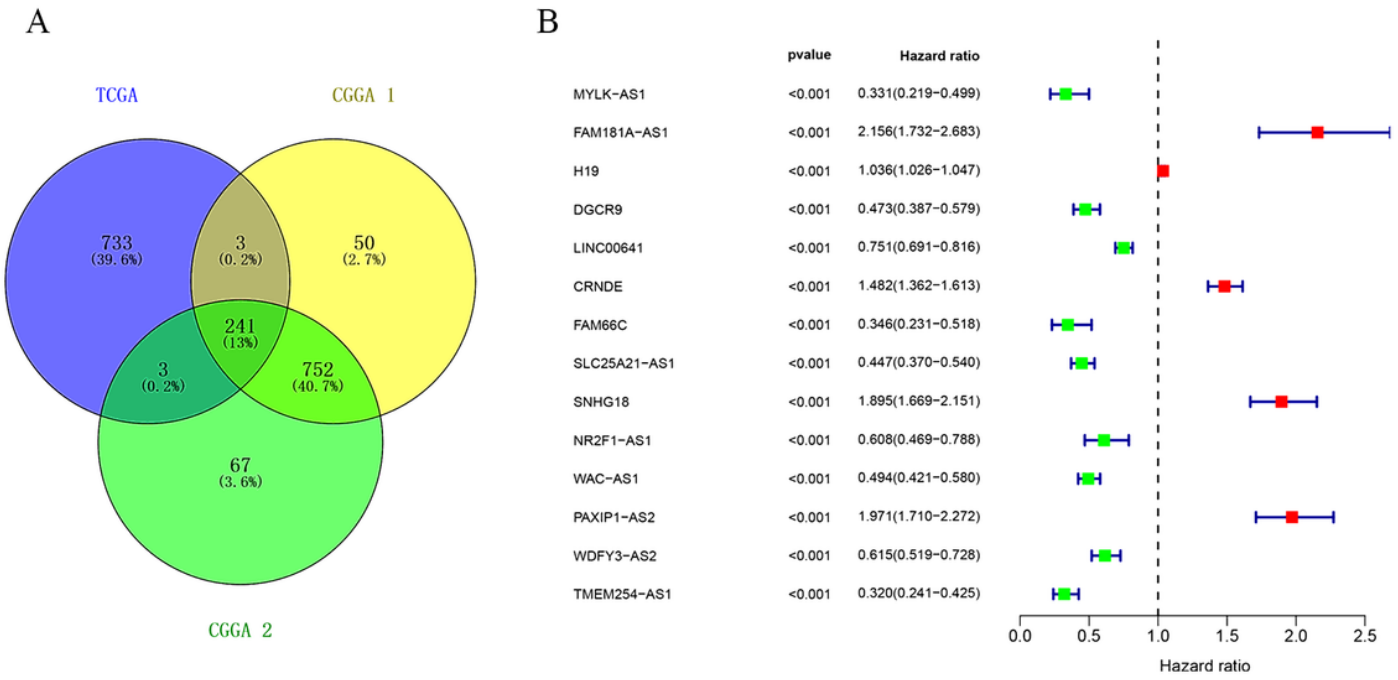

Figure 1

Construction of the hypoxia-related lncRNA signature. (A) Venn diagram showing the screening of intersected hypoxia-related lncRNA among three cohorts. (B) Forest plot of hazard ratios of 14 lncRNAs involved in the prognostic model

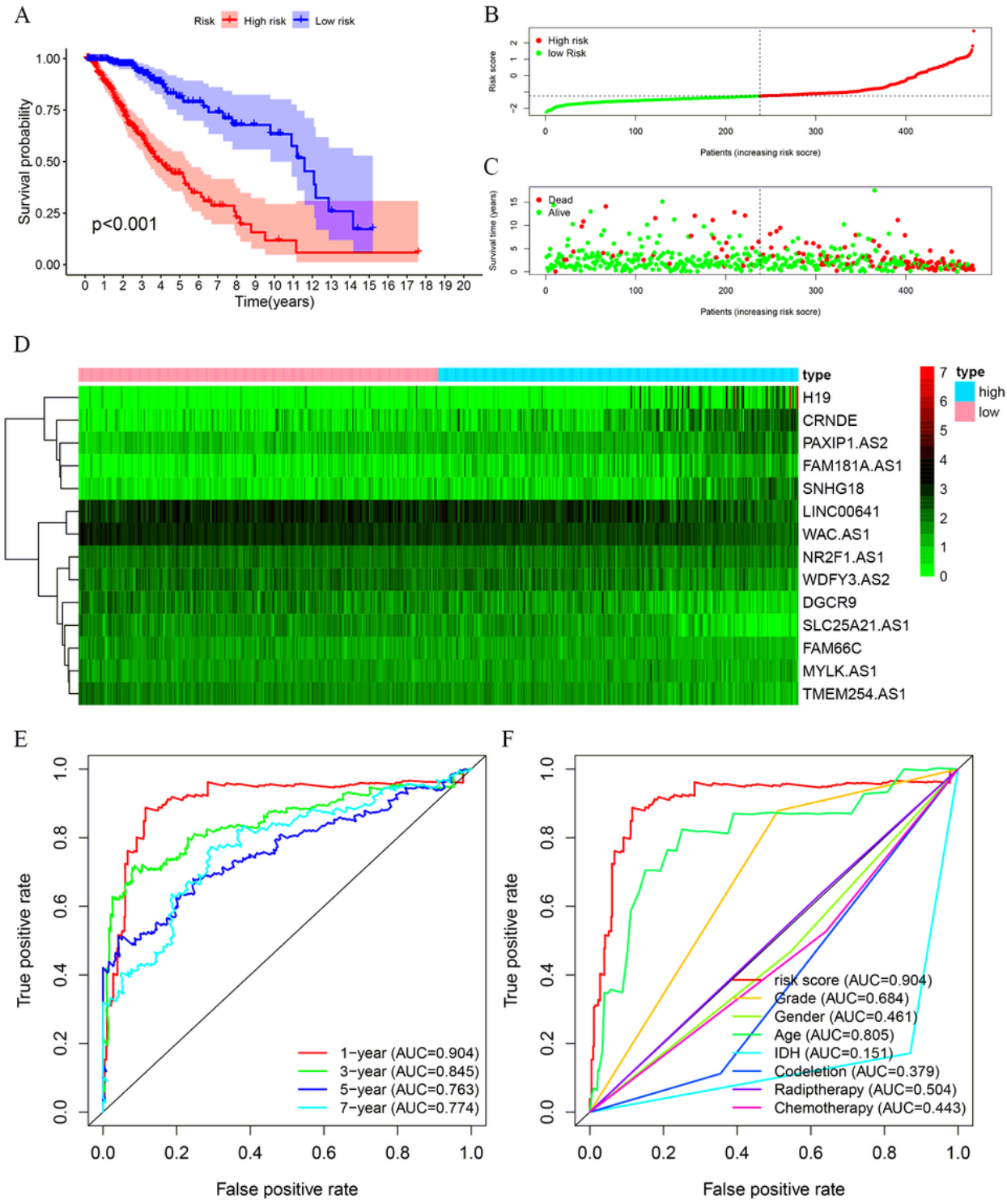

Figure 2

Validation of the established lncRNA signature in training set. (A) Kaplan-Meier survival curve showed that low-risk group had better prognosis than high-risk group. (B) The distribution of the risk score for each patient in training set. (C) Overview of the survival time for each patient in the training set. (D) Heatmap of expression profiles for 14 lncRNAs between low-risk group and high-risk group in the training set. Warm colors represented high expression, while cold colors represented low expression. (E) ROC

analysis of 1-, 3-, 5-, and 7-year survival based on risk score. (F) Comparison of ROC curves between risk score and clinical characteristics. AUC of risk score was higher than that of clinical characteristics

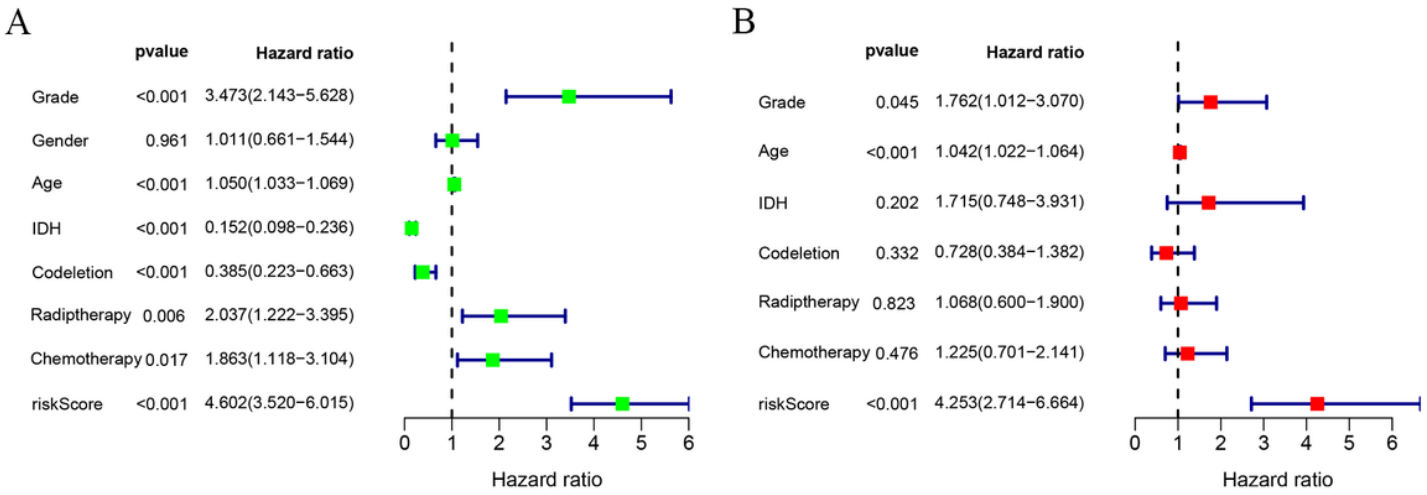

Figure 3

Cox regression analysis of training set. (A) Univariate Cox regression analysis of training set. (B) Multivariate Cox regression analysis of training set. Variables included in analysis included age, gender, tumor grade, IDH status, 1p19q status, chemotherapy, radiotherapy, and risk score

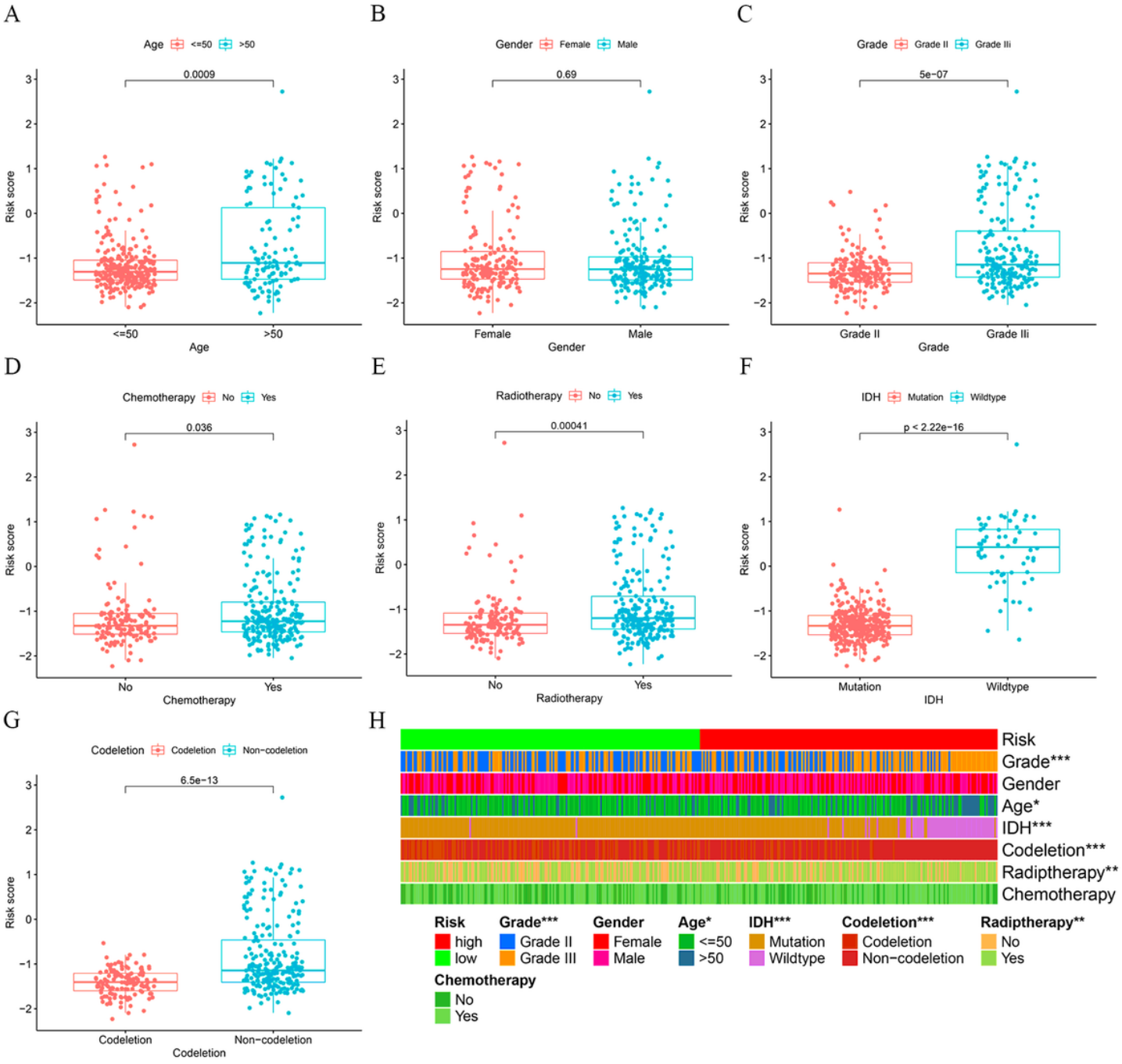

Figure 4

Comparison of clinical characteristics between high- and low-risk groups. Clinical characteristics including age (A), gender (B), grade (C), chemotherapy (D), radiotherapy (E), IDH status (F), and 1p19q status (F) were compared between two groups. (H) Heatmap showed detailed distribution of clinical characteristics between two groups

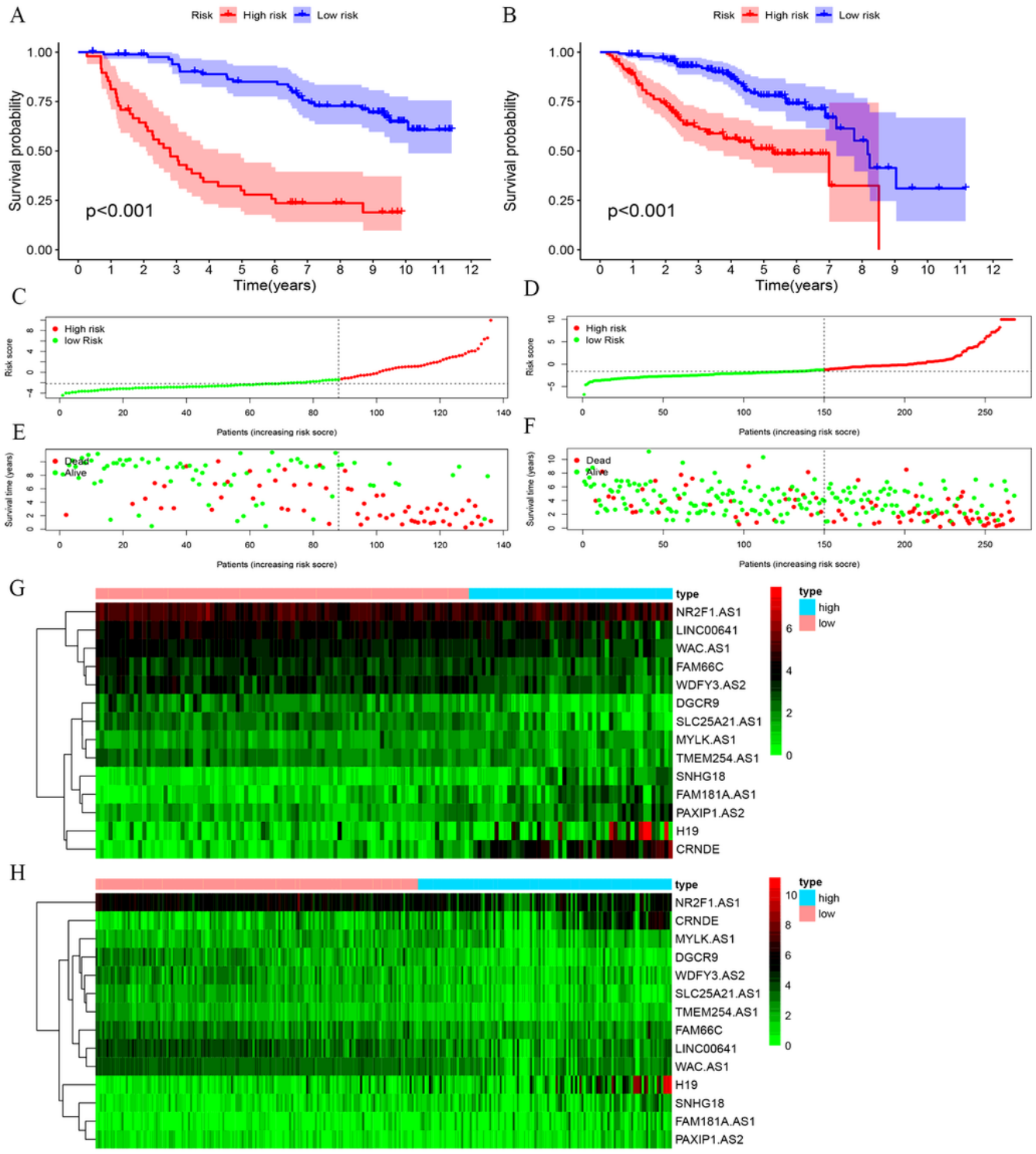

Figure 5

Validation of the established lncRNA signature in CGGA cohorts. (A) Kaplan-Meier survival curve of cohort 1showed that low-risk group had better prognosis. (B) Kaplan-Meier survival curve of cohort 2 also showed that low-risk group had better prognosis. (C-D) The distribution of the risk score for each patient in validation cohort 1 and 2. (E-F) Overview of the survival time for each patient in validation cohort 1 and 2. (G-H) Heatmaps of clinical characteristics between high- and low-risk groups in cohort 1 and 2

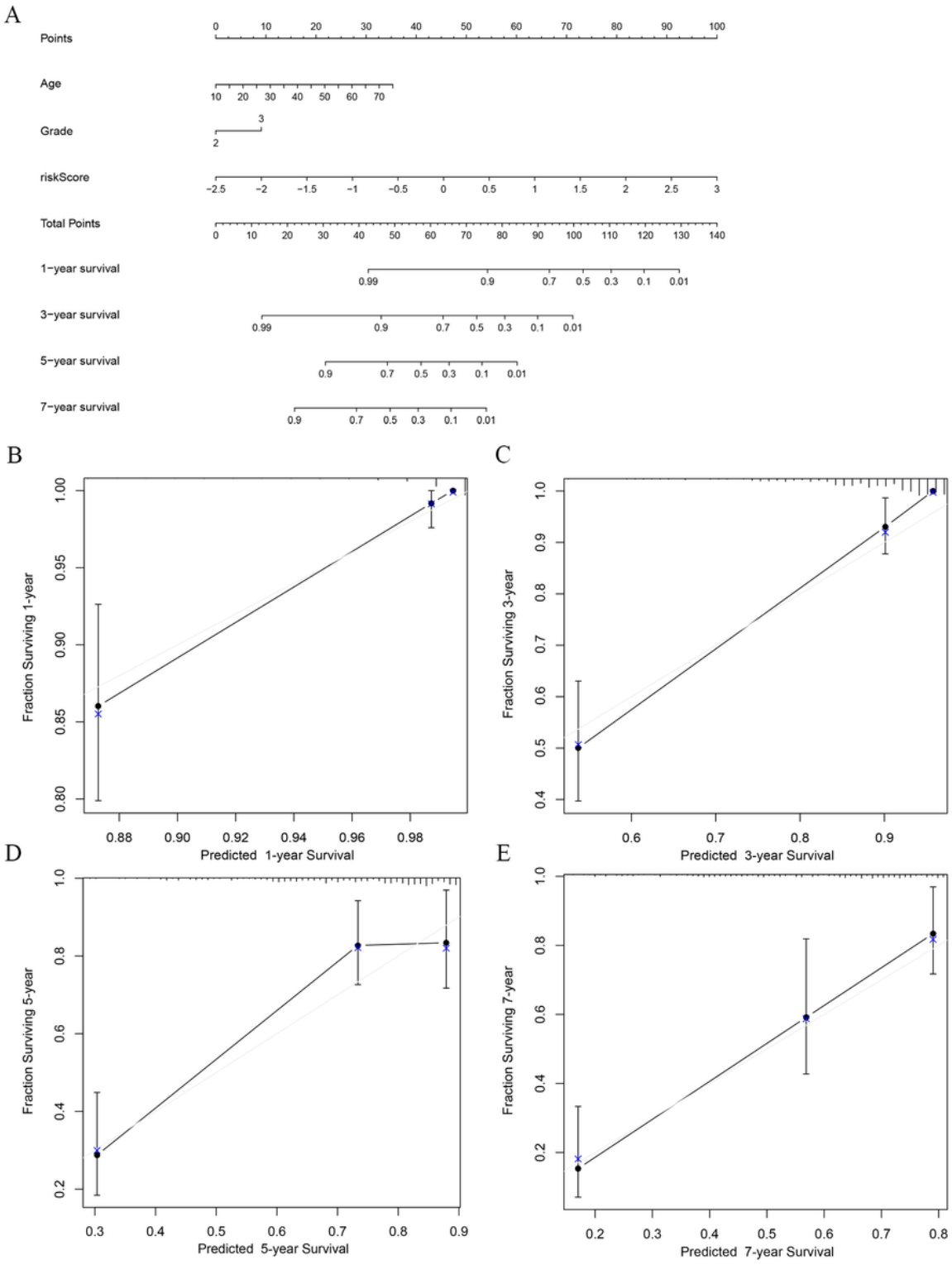

Figure 6

Construction and validation of the nomogram based on independent prognostic factors in training set. (A) The nomogram predicting overall survival of patients in the training set. (B-E) The calibration curves for 1-, 3-, 5-, and 7-year OS of the nomogram in the training set

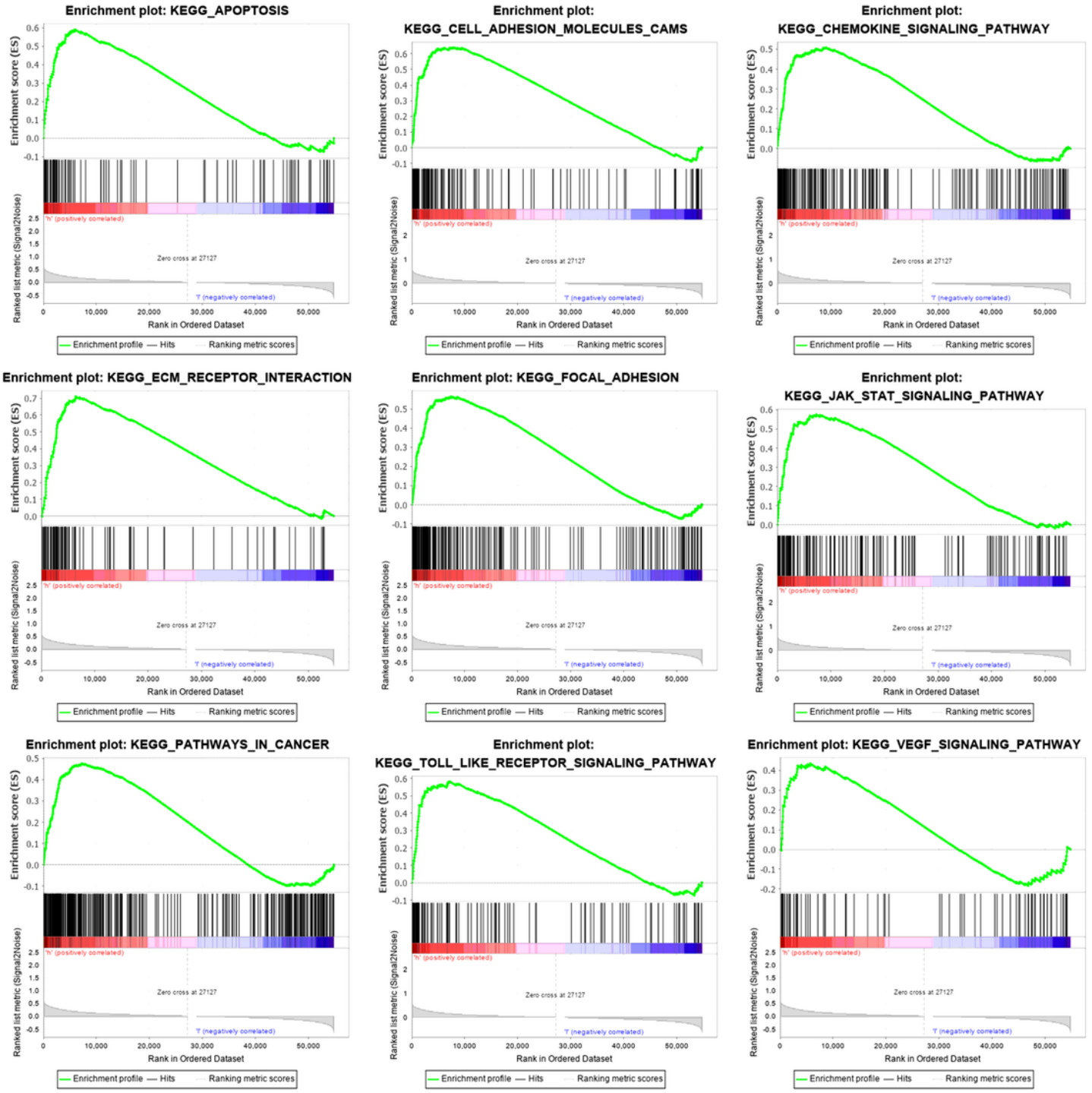

Figure 7

Representative KEGG pathways identified using GSEA. KEGG analysis showed that genes in high-risk group were involved in apoptosis, cell adhesion, chemokine signaling pathway, ECM receptor interaction, focal adhesion, JAK-STAT signaling pathway, pathways in cancer, TOLL-like receptor signaling pathway, and VEGF signaling pathway

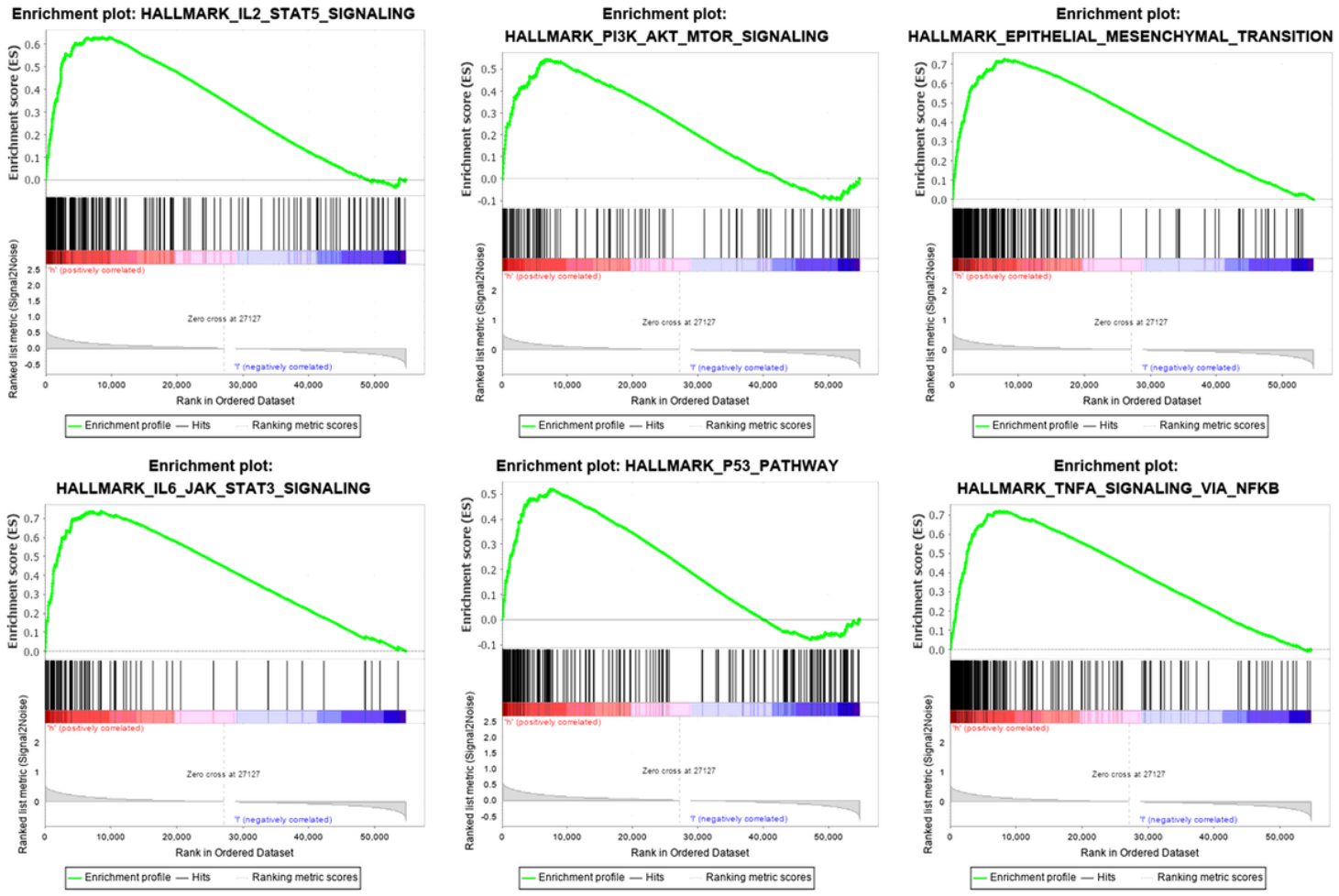

Figure 8

Representative oncological hallmarks identified using GSEA

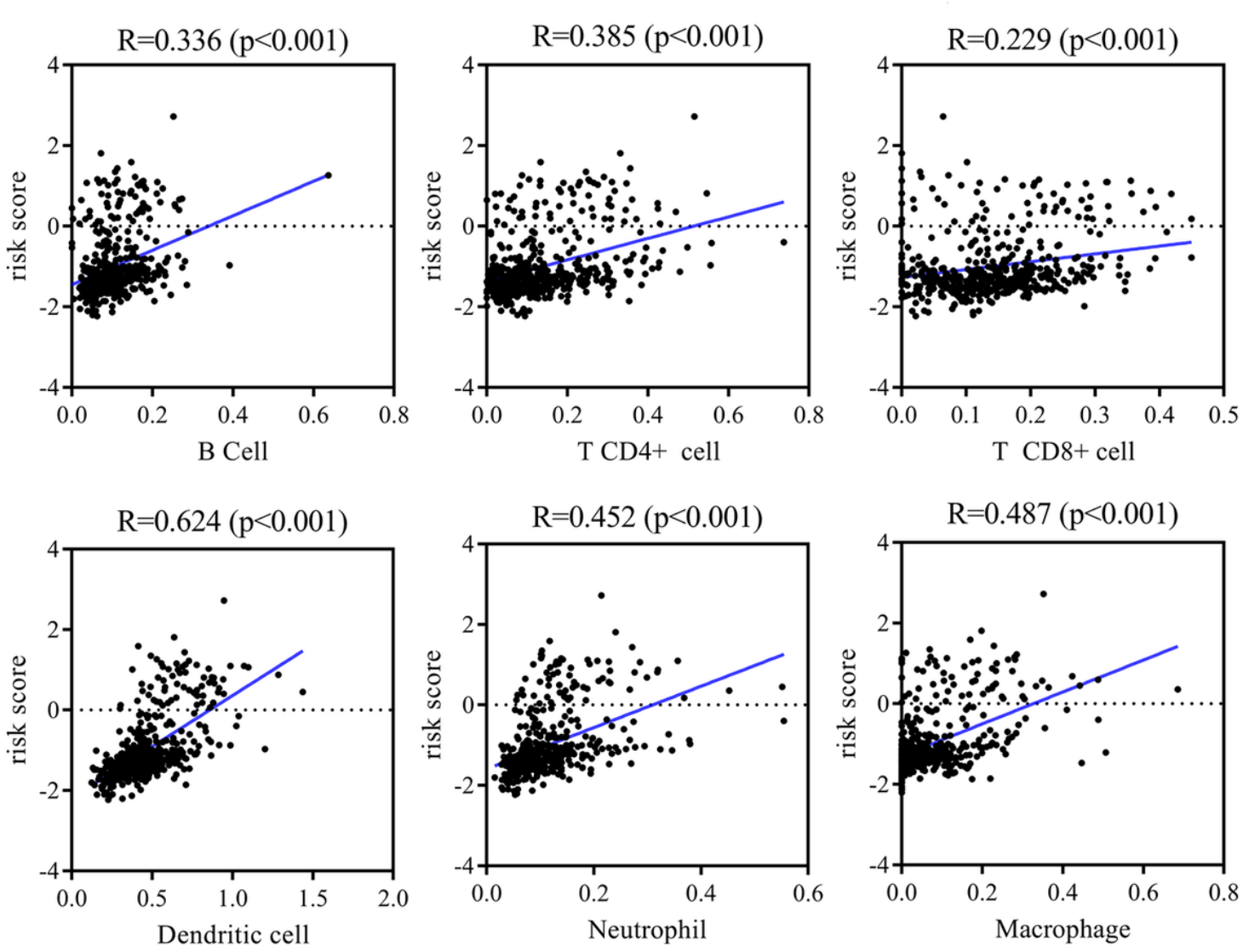

Figure 9

Pearson Correlation analyses of risk score and immune cells, including B cell, CD4+ T cell, CD8+ T cell, dendritic cell, neutrophil, and macrophage. Correlation coefficients and p values are displayed