Malignant tumor purity reveals the correlation between *CD3E* and low grade glioma microenvironment

WangRui Liu
Shanghai East Hospital https://orcid.org/0000-0002-0459-4446

Chuanyu Li
Department of Neurosurgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Guangxi, 533000, China

Wenhao Xu
Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, China

Hao Lian
Shanghai East Hospital

Yuanyuan Qu
Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, China

Kui Chen
Department of Neurosurgery, Shanghai East Hospital, Tongji University School of Medicine, 150 Jimo Road, Shanghai 200120, China

Huadong Huang
Affiliated Hospital of Youjiang Medical University for Nationalities

Haineng Huang
Department of Neurosurgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Guangxi, 533000, China

Chunlong Zhong (drchunlongzhong@126.com)
Department of Neurosurgery, Shanghai East Hospital, Tongji University School of Medicine, 150 Jimo Road, Shanghai 200120, China https://orcid.org/0000-0002-0605-7273

Research

Keywords: tumor microenvironment, tumor purity, CD3E, low grade glioma, prognosis, immune infiltrations

Posted Date: October 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-93487/v1

Abstract

Background: Tumor microenvironment (TME) contributes to the initiation and progression of low grade glioma (LGG); however, we are still unclear about the specifics of LGG's TME.

Methods: In this article, we selected 161 LGG patients from the Cancer Genome Atlas (TCGA) as data, and calculated the percentage of tumor infiltrating immune cells (TICs) in LGG and the tumor purity of LGG through ESTIMATE and CIBERSORT calculation methods. Immune-related genes were screened out through Cox regression and protein-protein interaction (PPI) network. The data in Gene Expression Omnibus (GEO) was selected to screen out clinically relevant genes. After combining the two, *CD3E* is selected as the predictor. Finally, we conducted verification at the Affiliated Hospital of YouJiang Medical University for Nationalities (AHYMUN) center.

Results: We found that the higher the expression of *CD3E*, the lower the purity of LGG tumors and the worse the prognosis of patients. Gene Set Enrichment Analysis (GSEA) showed that genes in the high-expressing *CD3E* group are mainly involved in immune-related activities. This suggests that *CD3E* may be responsible for regulating LGG's TME and tumor purity.

Conclusion: In short, the tumor purity of LGG has a considerable impact on clinical, genomic and biological status. The expression level of *CD3E* may help doctors evaluate the prognosis of LGG patients and develop personalized immunotherapy plans for patients. Evaluating the ratio of different tumor purity and the new role of *CD3E* may provide additional insights into the complex role of the LGG microenvironment and clinical treatment.

Background

Due to not comprehensively understanding of lipocytes by gene regulation and carcinogenesis, treatment and prognosis of gliomas are relatively limited [1, 2]. In clinical practice, gliomas are generally divided into four grades and low grade glioma (LGG) is grade I and II [3]. A large number of clinical studies have found that the survival rate of LGG patients is not high, and many patients have a sharp decline in survival time due to tumor deterioration in the later stage [4]. Nevertheless, high recurrence and malignancy rate of LGG still bring great pain to patients [5, 6]. Investigations on approaches to maintain the quality of life of LGG patients while prolonging the overall survival (OS) has become a common concern for clinicians and researchers [7].

The latest research finds that the tumor microenvironment (TME) facilitates the development of tumors [8]. The interaction between cancer cells, stromal cells and immune cells recruited from a distance promotes the invasion and metastasis of a variety of cancers, including proliferation, anti-apoptosis, and evasion of immune surveillance, thereby significantly affecting the treatment and prognosis of cancer patients [9, 10]. TME is mainly composed of resident stromal cells and recruited immune cells [11]. Stromal cells and immune cells affect tumor blood vessel growth and tumor proliferation, respectively. Meanwhile, tumor-infiltrating immune cells (TICs) in TME can be used to determine the prognosis of

patients [12], and the related immune genes have an impact on the survival of cancer patients. For example, immune genes affect brain tumors [13, 14]. This correlation has led to improvements in immune-based treatment methods to create immune checkpoint inhibitors and prognostic biomarkers for tumor patients [15–17]. These studies suggest that the various immune responses of LGG's TME may change the purity of the tumor, thereby affecting the invasion and metastasis of LGG. The study found that there is a deep connection between LGG and TME. The higher the stroma and immune score of LGG, the lower the purity of the tumor and the more aggressive. Low glioma purity shows a strong immunophenotype and suggests a poor prognosis [18]. Thus, clinicians and basic researches are required to identify tumor purity that accurately reflect the LGG heterogeneity and complex role of microenvironment, which may also help to explore novel biomarkers of LGG.

We selected 161 LGG patients from the Cancer Genome Atlas (TCGA) as data, and calculated the percentage of tumor infiltrating immune cells (TICs) in LGG and the tumor purity of LGG through ESTIMATE and CIBERSORT calculation methods, as well as the ratio of immune and matrix components, and selected the inter-sample screening in the Gene Expression Omnibus (GEO). LGG genes associated with prognosis were identified and the predictive biomarker *CD3E* was found. The T cell antigen receptor epsilon subunit (*CD3E*) gene is located at 11q23.3, composed of 9 exons, and is associated with autosomal recessive hereditary early-onset immunodeficiency 18 phenotype, which is a severe combined immunodeficiency variant [19]. Moreover, *CD3E* is overexpressed in certain solid tumors and is associated with immunity [20, 21]. We started by the differentially expressed genes (DEGs) produced by comparing immunological and matrix components in LGG samples, and revealed that *CD3E* was a potential indicator of TME status changes in LGG.

Methods

Data collection

We downloaded 161 LGG patient's RNA-Seq data and clinical data from the TCGA SpliceSeq dataset (bioinformatics.mdanderson.org)[22].

In the discovery step, we only select the data set that includes the LGG tissue and normal brain tissue, the titles and abstracts of these data sets were screened, and all information of the data sets of interest were further evaluated. Finally, we select three data sets, GSE107850 on GPL14951, GSE26576 on GPL6801 and GPL570, GSE20395 on GPL9183, were selected for analysis. All data sets are downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo) [23].

Score calculation

We use the estimate R software [24] (version 4.0.0) to estimate the proportion of TME immune cells and stromal cells in each LGG sample, we set ImmuneScore, StromalScore and ESTIMATEScore according to the proportion of the corresponding cells in TME.

Subsistence analysis

This study included 161 patients from TCGA database;459 patients from GEO database and 100 patients from AHYMUN database. Survival analysis by R, $p < 0.05$ was considered significant.

We performed Cox univariate analysis on the clinical data of patients in the Affiliated Hospital of YouJiang Medical University for Nationalities (AHYMUN) Center to evaluate all events that may affect the OS and disease-free survival (DFS) of LGG patients, including age, gender, epilepsy history, Karnofsky score, tumor envelope infiltration, *CD3E* expression, etc.

Screening for prognosis-related differential expressed genes (DEGs)

Using "LIMMA" [25] in R software, the data were standardized and miRNA differential expression analysis. Put the relevant code into R, and analyze the DEGs in the meningioma samples and normal brain tissue samples through the limma software package. *P* value < 0.05 and | fold change (FC) | > 1 was set as the threshold for identifying Clinical-related DEGs.

Screening for Immune-related DEGs

According to the median of ImmuneScore and StromalScore we calculated, the 161 LGG samples in the TCGA database were marked as high or low. Use package limma to conduct differential analysis of gene expression, and generate Immune-related DEGs by comparing high and low score samples. Immune-related DEGs (high/low score group) and false discovery rate < 0.05 with a fold change greater than 1 after log2 conversion were considered significant. We calculated the TIC value in all LGG data by the CIBERSORT method, and the samples with $P < 0.05$ can be further analyzed.

Bioinformatics Analysis

The protein-protein interaction (PPI) network is constructed from the STRING database. All gene interaction networks were drawn by Cytoscape (version 3.8.0.) [26]. We performed gene ontology (GO) enrichment analysis of DEGs through R software, and determined the biological processes (BPs), cell components (CCs) and molecular functions (MFs) of each gene. We also performed Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis to show enrichment for related genes.

Gene Set Enrichment Analysis (GSEA) We use GSEA software (vision 4.0.3) to analyze the entire transcriptome of all tumor samples [27], and only genomes with $p < 0.05$ are considered important.

Immunohistochemistry

Immunohistochemistry streptavidin peroxidase method was used to detect the expression of *CD3E* in LGG and nearby normal tissues. The LGG samples were scored according to the degree of cell staining: 0, cytoplasmic yellow particles; 1. Light brown particles; 2. Obvious brown particles; 3. A large number of dark brown particles. The LGG samples were also scored according to the percentage of positive cells, 0 points: 0%, 1: points: <10%, 2: points: 11%-50%, 3: points: 51–80%, 4: points:> 80%. Calculate the final IHC Score by multiplying the two scores [28].

Results

As shown in Fig. 1, our research is divided into three stages. To estimate the proportion of TICs in LGG samples and tumor purity, transcriptome RNA-seq data from 516 patients were downloaded from TCGA; then ESTIMATE and CIBERSORT algorithms were performed. DEGs shared by ImmuneScore and StromalScore were used to construct a PPI network. Significant hub genes in the PPI network were evaluated using univariate Cox regression cross-analysis. Meanwhile, we selected a qualified data set from the GEO database and conducted a difference analysis to obtain clinical-related DEGs; then the association between all the DEGs and the survival of LGG patients were evaluated and screened. Next, *CD3E* was identified and validated as the most relevant gene after combination of the two datasets of DEGs. Further studies focused on impact of *CD3E* on survival, GSEA and correlation with TICs. Functional annotations of neighbor genes and clinical validation of *CD3E* was elaborated. Finally, we put the research conclusions in our own AHYMUN center for clinical cohort study.

TME-related scores are related to survival of LGG patients

In order to confirm whether the proportion of cells in TME and tumor purity will affect the survival time of LGG patients, we calculated ImmuneScore, StrromalScore and ESTIMATEScore, and drew a Kaplan-Meier survival curve. The higher the Score, the higher the proportion of the corresponding component in TME. The sum of ImmuneScore and StromalScore is ESTIMATEScore, which also reflects tumor purity from the side. In Fig. 2, TME scores are related to overall survival. ImmuneScore ($P = 0.003$), StromalScore ($P < 0.001$) and ESTIMATEScore ($P = 0.006$) were positively correlated with OS. These results show that we can infer the prognosis of LGG patients based on the proportion of immune cells in TME and formulate personalized treatment plans.

TME-related scores are related to the Clinical features of LGG Patients

We combined the corresponding clinical information of TCGA's LGG patients with the above calculated scores to determine whether the LGG's TME and tumor purity are related to the patient's clinical characteristics. ImmuneScore positively correlated to high grade of LGG (Fig. 3C, $P < 0.001$); StromalScore was positively correlated to high grade of LGG (Fig. 3F, $P < 0.001$), and ESTIMATEScore accompanied with high grade of LGG (Fig. 3I, $P < 0.001$). These results indicate that tumor purity and the ratio of immune/stromal cells in TME are related to the deterioration of LGG. The higher the ratio of immune/stromal cells in TME, the lower the purity of the tumor, and the worse the prognosis of LGG patients.

The Enrichment Analyses of Immune-Related DEGs

In order to determine the exact changes in the genetic profiles of immune and matrix components in TME, we compared high- and low-scoring samples based on the median. We got 297 DEGs through ImmuneScore, 201 genes were upregulated, and 96 genes were downregulated (Fig. 4A, C, D). We also got 518 DEGs from StromalScore, which contained 461 upregulated genes and 57 downregulated genes (Fig. 4B-4D). We found through Venn diagram that 199 upregulated genes with high score and 19 downregulated genes with low score were both in ImmuneScore and StromalScore. These 218 immune-

related DEGs may play a decisive role in LGG's TME. We found through GO enrichment analysis and KEGG analysis that the biological functions of these genes are mainly related to immunity. (Fig. 4E-F).

Identify key Immune-Related genes

In order to further study the underlying mechanism of the above genes and find the key genes, We drew the PPI network diagram through String. The interaction between the genes is shown in Fig. 5A. We selected the top 30 genes ranked by the number of nodes and plotted them into a bar graph (Fig. 5B). We performed univariate COX regression analysis on the survival of Immune-Related DEGs and LGG patients to determine which genes are at high risk for LGG patients and which are low risk. (Fig. 5C). Finally, we combined the main nodes in the PPI and the top 75 genes ranked by the p value to analyze them, we have obtained 30 intersecting genes. (Fig. 5D).

Filter clinical-related DEGs and Lock the Target Gene

We use the R language package to screen all the genes that affect survival in three GSE sets. We screened 114 clinical-related DEGs ($P < 0.001$) that were significantly related to survival from 13299 related genes, and compared them with the previous immune-related DEGs to obtain 7 genes: *CD3E*, TLR2, CCR5, CXCL9, CXCL10, FCGR2A, and ITGAL (Fig. 5E). We mapped the PPI network for these 7 genes (Fig. 5F). 78.89% terms were in co-expression (lavender line), 7.65% terms were shared protein domains (yellow line), 7.11% terms were in co-localization (deep blue line), and 7.11% terms were predicted (khaki line). We also performed GO and KEGG pathway analyses on these 7 genes, finding that the genes were related to immune diseases and inflammatory response (Fig. 5G). Based on the hazard ratio (HR) value of each gene and the survival-related *p* value, we targeted *CD3E* for further study.

Identification of Clinical-Related DEGs

According to the median of *CD3E* expression in the sample, we divided the data set into a high and a low expression groups and screened using "log fold change = 1, and $P < 0.05$". A total of 114 related differential genes were obtained. The 15 genes with the most significant up-regulation and the 11 genes with the most significant down-regulation were selected for further analysis (Table 1), which were visualized by volcano map (Fig. 6A) and heat map (Fig. 6B).

Table 1
The 15 genes with the most significant up-regulation and the 11 genes with the most significant down-regulation were selected for further analysis

id	logFC	AveExpr	t	P.Value	adj.P.Val	B
CD3E	0.503296	8.649803	14.40339	1.42E-32	1.79E-28	61.13156
WDR3	0.540363	10.02952	7.664638	8.06E-13	5.09E-09	18.49936
CRABP2	0.574073	9.887544	6.228971	2.80E-09	2.95E-06	10.83195
MED25	0.573891	11.43372	5.924861	1.38E-08	8.78E-06	9.336745
SMAD6	0.550032	10.46	5.684687	4.69E-08	2.19E-05	8.192201
E2F2	0.693043	10.64766	5.252812	3.88E-07	8.76E-05	6.219642
GINS2	0.533677	11.25321	5.242128	4.08E-07	9.05E-05	6.172292
IGSF5	0.516926	10.51717	5.105327	7.78E-07	0.000138	5.57241
KCNIP2	-0.53684	11.71034	-4.94083	1.66E-06	0.000225	4.867026
LILRB4	0.506478	10.80824	4.702932	4.82E-06	0.000444	3.878589
DCT	0.502274	9.540816	4.647528	6.15E-06	0.000529	3.653898
ODF3L2	0.576656	10.64615	4.512396	1.10E-05	0.000801	3.114778
TIMP4	-0.51087	13.07861	-4.48168	1.26E-05	0.000864	2.994023
CCDC102A	0.509778	9.87002	4.438659	1.51E-05	0.000969	2.825997
OGDHL	-0.63799	10.63297	-4.37422	1.98E-05	0.001135	2.5768
SLC15A3	0.503854	11.32423	4.240455	3.43E-05	0.001631	2.069005
UPK1A	0.538824	9.495527	4.145965	5.03E-05	0.002087	1.718163
REM1	0.608585	10.15157	3.694196	0.000286	0.006442	0.133289
SYN2	-0.54007	11.02732	-3.67165	0.000311	0.006821	0.05829
RIT2	-0.5668	9.842119	-3.65283	0.000333	0.007176	-0.00398
P2RY1	-0.51652	11.04178	-3.59096	0.000416	0.008445	-0.20681
CAMK4	-0.53104	11.17374	-3.53893	0.000501	0.00946	-0.37504
KCNC2	-0.55738	9.807489	-3.53712	0.000505	0.00949	-0.38087
SNCB	-0.51772	11.36075	-3.43972	0.000711	0.011889	-0.68982
CALY	-0.6215	11.73797	-3.33767	0.001011	0.015188	-1.00539
ALDH1A3	0.520521	10.04314	3.110935	0.002143	0.025217	-1.67609

id	logFC	AveExpr	t	P.Value	adj.P.Val	B
ZFR2	-0.54908	11.97667	-2.95741	0.003483	0.034902	-2.10604

Correlation Analyses of Clinical-Related DEGs

As illustrated in Fig. 6C, gene-gene interaction between Clinical-Related DEGs and related genes was performed. 95.20% terms were in co-expression (lavender line), and 4.80% terms were in co-localization (deep blue line). In Fig. 6D-6F, We conducted a biological function enrichment analysis of DEGs. The results showed that enrichments of biological processes were positive regulation of voltage-gated potassium channel activity, positive regulation of potassium ion transmembrane transporter activity and regulation of pry-miRNA transcription by RNA polymerase II (Fig. 6D); enrichments of cellular components were ion glutamatergic synapse, apical plasma membrane and apical part of cell (Fig. 6E); enrichments of molecular functions were oxidoreductase activity, calmodulin binding and copper ion binding (Fig. 6F). Enrichments in KEGG pathway were glioma, tyrosine metabolism and citrate cycle (Fig. 6G).

We correlated the 20 most significantly up-regulated genes and the 20 most significantly down-regulated genes with *CD3E*. Red for positive correlation, and green represents a negative correlation. The deeper the color, the greater the relevance. *CD3E* is positively correlated with LILRB4, UPK1A, and REM1, negatively correlated with RIT2, OGDHL, and KCNC2 (Fig. 6H).

CD3E Expression is Negatively Related to the Survival of LGG Patients

CD3E is an epsilon subunit of T cell antigen receptor. According to the median of *CD3E* expression, all LGG samples were divided into *CD3E* high, median and low expression groups. Survival analysis showed that in TCGA ($P = 0.0011$; Fig. 7A) and GSE ($P < 0.001$; Fig. 7B), the survival rate of LGG patients with high *CD3E* expression was lower than that of *CD3E* low expression. Similarly, in GEPIA, the OS of the *CD3E* high expression was lower than that of the low expression ($P < 0.001$; Fig. 7C)

CD3E is a Potential Indicator of TME Modulation

Considering that *CD3E* expression is negatively correlated with the survival rate of LGG patients, we performed GSEA analysis on the high expression group. We found that the genes in the *CD3E* high expression group mainly participated in immune-related activities, such as B cell receptor signaling pathway, chemokine signaling pathway and T cell receiver signaling pathway (Fig. 7D). Furthermore, *CD3E* was positively related to glioma and immune cell response. These results suggest that *CD3E* may be a potential indicator of TME status for LGG.

Correlation of *CD3E* With the Proportion of TICs

We used the CIBERSORT algorithm to analyze the proportion of TICs of 22 immune cells in LGG to further study the correlation between *CD3E* and the immune microenvironment of LGG. (Fig. 8). We found that the expression of *CD3E* is related to the TIC of 10 LGG (Fig. 9).Seven kinds of TICs were positively

correlated with *CD3E* expression, including macrophages M0, macrophages M1, mast cells resting, NK cells resting, T cells CD4 memory activated, T cells CD8 and T cells regulatory; three kinds were negatively correlated with *CD3E* expression, including eosinophils, monocytes and NK cells activated. These results prove that *CD3E* is related to the immune activity of TME, thereby affecting the tumor purity of LGG.

Clinicopathological features related to *CD3E* expression

To verify *CD3E* expression in LGG, we performed immunohistochemistry (IHC) (Fig. 10A-10B). The scatter plot of the IHC scores revealed that *CD3E* expression increased in LGG tissues in the AHYMUN cohort ($P < 0.01$). In Table 2, we found that higher *CD3E* expression is with patients' age ($P = 0.027$), grade ($P < 0.001$), microvascular invasion ($P = 0.009$), history of epilepsy ($P < 0.001$) and Karnofsky score ($P = 0.002$). This seems to indicate that the higher the expression of *CD3E* in patients, the worse the prognosis.

Table 2
Clinicopathological characteristics in relation to CD3E expression level in AHYMUM cohort.

Characteristics	AHYMUM cohort (N = 100)	CD3E expression		χ^2	*P*
		Low IHC score (N = 50)	High IHC score (N = 50)		
N (%)					
Age				4.889	**0.027**
60 years	55(0.55)	33(0.60)	22(0.40)		
≥ 60 years	45(0.45)	17(0.38)	28(0.72)		
Gender				0.271	0.603
Male	82(0.82)	40(0.49)	42(0.51)		
Female	18(0.18)	10(0.56)	8(0.44)		
Grade				14.924	**< 0.001**
G1	69(0.69)	39(0.57)	30(0.43)		
G2	31(0.31)	11(0.35)	20(0.65)		
Seizure history				12.148	**< 0.001**
yes	61(0.61)	39(0.64)	22(0.36)		
no	39(0.39)	11(0.28)	28(0.72)		
Microvascular invasion				6.828	**0.009**
Absent	55(0.55)	34(0.62)	21(0.38)		
Present	45(0.45)	16(0.36)	29(0.64)		
Capsular invasion				1.961	0.161
Absent	51(0.51)	29(0.57)	22(0.43)		
Present	49(0.49)	21(0.43)	28(0.57)		
Karnofsky score				9.180	**0.002**
≥ 80	61(0.61)	36(0.59)	21(0.41)		
80	39(0.39)	14(0.36)	29(0.64)		

Cox regression analysis

We use univariate Cox regression analysis to show the relationship between *CD3E* and AHYMUN patients, we found that *CD3E* is not very relevant to age and gender. (Fig. 10C). In the multivariate model, we also found that patients in the high expression group had worse OS (HR = 3.22; *P* = 0.001). Moreover, in the AHNTU cohort, the microvascular invasion (HR = 1.52; *P* = 0.024), the presence of capsular infiltration (HR = 1.63; *P* = 0.016), and the Karnofsky scores (ref < 80) (HR = 1.46; *P* = 0.023) were associated with low OS (Table 3).

Table 3
Multivariate Cox regression analysis of DFS and OS in AHYMUM cohorts (DFS: disease-free survival; OS: overall survival)

Covariates	**OS**			**DFS**		
	HR	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**
Grade (ref. G1)	1.97	2.25–3.68	**0.043**	2.31	1.94–4.02	**0.037**
Microvascular invasion (ref. Absent)	1.52	1.61–2.54	**0.024**	1.98	1.73–3.64	**0.031**
Capsular invasion (ref. Absent)	1.63	2.17–3.21	**0.016**	1.54	2.31–3.16	**0.017**
Karnofsky score (ref. ≥80)	1.46	2.31–3.27	**0.023**	1.56	1.66–2.64	**0.044**
CD3E expression (ref. low)	3.32	2.48–9.91	**0.001**	4.33	2.64–12.21	**< 0.001**

We found that the patient's gender and epilepsy history were not related to DFS (Fig. 10D). Multivariate We found through Cox analysis that the high expression of the *CD3E* gene caused a significant decrease in OS (HR = 4.33; *P* < 0.001) (Table 3). Including grade, capsular infiltration, microvascular invasion and Karnofsky scores are related to OS (*P* < 0.05). In Fig. 10E-F, the higher the *CD3E* expression level, the lower the OS and DFS of LGG patients.

Discussion

In our study, we first screened the immune genes related to TME in LGG patients from TCGA. Next, we screened out genes related to the prognosis of LGG patients from GEO. After combining the above genes, we determined that CD3E is the target gene. Then we conducted a series of bioinformatics analysis and verified the research results from our own center. We found that CD3E may be an indicator gene of the TME status of LGG patients, and by affecting the TME of LGG, thereby changing the tumor purity and affecting the prognosis of patients.

The combination of the cancer cell genotype and the expression program related to the cell phenotype and the influence of the TME determines the tumor's adaptability, evolution, and resistance to treatment [29]. In recent years, studies on TCGA and GSE have mapped the genetic picture and overall expression status of numerous tumors, identified driver mutations and defined tumor subtypes based on specific transcription profiles [30, 31].

LGG is a common brain tumor, and the prognosis of patients is often poor [32]. However, whether it is surgery, radiation therapy or chemotherapy (usually using temozolomide), can't improve the prognosis and survival of patients [7, 33, 34]. The reasons for the lack of progress include the growth of invasive tumors in basic organs, which limits the utility of local therapies, and the protection of tumor cells by the blood-brain barrier, their inherent resistance to induced cell death, and the lack of dependence on a single, can targeted carcinogenic pathways [35]. Besides, when pursuing immune-based glioblastoma treatment methods, the unique immune environment of the central nervous system needs to be considered [36–38]. Therefore, we need to study novel LGG immunotherapy candidates. Here, we start from the transcription analysis of LGG in TCGA and find that the decreased expression of *CD3E* is closely related to poor prognosis of patients. Therefore, *CD3E* is a potential prognostic indicator and treatment target in LGG patients.

CD3E protein is encoded polypeptide CD3-ε, which together with the CD3-γ, -δ and -ZETA and T-cell receptor α / β and γ / δ T cell receptor heterodimer -CD3 complex. The complex plays an important role in coupling antigen recognition to several intracellular signal transduction pathways, so defects in *CD3E* can lead to immunodeficiency [39]. *CD3E* also participates in proper T cell development. TCR-CD3 complex assembly is initiated by forming two heterodimers CD3D/*CD3E* and CD3G/*CD3E*. It also participates in the internalization of TCR-CD3 complexes and cell surface down-regulation by endocytic sequences present in the cytoplasmic region of *CD3E* [40–42]. The relationship between the abundance of tumor infiltrating lymphocytes and the expression, copy number, methylation or mutation of *CD3E* in LGG was shown in **Supplement** Fig. 1.

In LGG patients, the higher the expression of *CD3E*, the worse the patient's survival. It might be attributed to immune cells with *CD3E* high expression promoting anti-tumor immunity except T cells regulatory. Similarly, *CD3E* acts as a T cell receptor, its high expression in many cancers indicates better clinical results (longer survival), with the exception of LGG alone [43]. This may be related to the cause of LGG and the immune environment of the brain, or it may be due to the interconnection between isocitrate dehydrogenase and TME [36, 44–46]. Therefore, *CD3E* may play a dual role in tumors, either promoting survival or inducing apoptosis. In addition, in the TME of glioma, the proliferation of malignant cells is enhances, the pool of undifferentiated glioma cells increases, and macrophage expression exceeds microglial expression [43]. We used GSEA and found that the *CD3E* high expression enriched immune-related signaling pathways, such as B/T cell receptor signaling pathway and chemokine signaling pathway. These results indicate that *CD3E* may be involved in the transition of TME from immune-based to metabolic-based. More and more studies prove that *CD3E* is related to tumor treatment [43, 47–49]. Our research also found that the balance between tumor pathways, sugar metabolism and lactic acid

formation can affect the immune status of LGG. Therefore, we suspect that in the development of LGG, the up-regulation of *CD3E* promotes the decline of tumor purity, and at the same time the transition of TME from immunotype to metabolite type further promotes the deterioration of LGG.

In general, we use the ESTIMATE algorithm to determine the TME-related genes in LGG by analyzing LGG samples in TCGA. Through the analysis of LGG samples in GEO, the prognostic related genes in LGG were determined. The above studies confirmed that *CD3E* is not only a potential prognostic factor for LGG patients, but also a driving factor for TME to transform from an immune state to a metabolic state.

Conclusion

In conclusion, the purity of LGG has a considerable impact on clinical, genomic and biological status. The expression of *CD3E* may help doctors analyze the purity of tumors in LGG patients, thereby predicting the tumor development and prognosis of patients, and formulating personalized treatment plans for patients. Assessing proportions of different tumor purity and novel role of *CD3E* may offer an extra insight for complex role of LGG microenvironment and clinical treatments.

Abbreviations

AHYMUN: Affiliated Hospital of YouJiang Medical University for Nationalities; GEO: Gene Expression Omnibus; GO: Gene Ontology; GSEA: Gene Set Enrichment Analysis; HR: hazard ratio; IHC: immunohistochemistry; KEGG: Kyoto Encyclopedia of Genes and Genomes; LGG: low grade glioma; OS: overall survival; PPI: Protein-protein interaction; TCGA: the Cancer Genome Atlas; TIC: tumor-infiltrating immune cells; TME: tumor microenvironment.

Declarations

Acknowledgements

None

Author Contributions

WL, CZ, Haineng Huang and Huadong Huang conceived the idea, performed analysis and drafted the manuscript; WL, CL and WX performed data acquisition and figure preparations, HL, YQ and KC interpreted the results and help to revise the manuscript. The authors read and approved the final manuscript.

Funding

CZ received funding from the Outstanding Leaders Training Program of Pudong Health Bureau of Shanghai (PWR12018-07), the Top-level Clinical Discipline Project of Shanghai Pudong (PWYgf2018-05),

the Key Discipline Construction Project of Pudong Health Bureau of Shanghai (PWZxk2017-23). CL received funding from the Guangxi Medicine and Health Self-financing Research Project (20201558).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request or online database.

Ethical Approval and Consent to participate

All of the study designs and test procedures were performed in accordance with the Helsinki Declaration II. The Ethics approval and participation consent of this study was approved and agreed by the ethics committee of Affiliated Hospital of Youjiang Medical University for Nationalities (Baise, Guangzhou, China).

Consent for publication

Not applicable.

Conflict of Interest

The authors declare no competing interests.

Author details

1 Department of Neurosurgery, Shanghai East Hospital, Tongji University School of Medicine, 150 Jimo Road, Shanghai 200120, China. 2 Department of Neurosurgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Guangxi, 533000, China. 3 Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, China. 4Department of Oncology, Shanghai Medical University, Fudan University, Shanghai 20032, China.

References

1. Ceccarelli, M., et al., *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell, 2016. **164**(3): p. 550-63.
2. Alexander, B.M. and T.F. Cloughesy, *Adult Glioblastoma.* J Clin Oncol, 2017. **35**(21): p. 2402-2409.
3. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary.* Acta Neuropathol, 2016. **131**(6): p. 803-20.
4. Kesari, S., et al., *Phase II study of protracted daily temozolomide for low-grade gliomas in adults.* Clin Cancer Res, 2009. **15**(1): p. 330-7.
5. McCormack, B.M., et al., *Treatment and survival of low-grade astrocytoma in adults--1977-1988.* Neurosurgery, 1992. **31**(4): p. 636-42; discussion 642.

6. Turkoglu, E., et al., *Clinical outcome of surgically treated low-grade gliomas: a retrospective analysis of a single institute.* Clin Neurol Neurosurg, 2013. **115**(12): p. 2508-13.
7. Rathore, S., et al., *Glioma Grading via Analysis of Digital Pathology Images Using Machine Learning.* Cancers (Basel), 2020. **12**(3).
8. Wang, Q., P. Li, and W. Wu, *A systematic analysis of immune genes and overall survival in cancer patients.* BMC Cancer, 2019. **19**(1): p. 1225.
9. Wood, S.L., et al., *The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets.* Cancer Treat Rev, 2014. **40**(4): p. 558-66.
10. Quail, D.F. and J.A. Joyce, *Microenvironmental regulation of tumor progression and metastasis.* Nat Med, 2013. **19**(11): p. 1423-37.
11. Bussard, K.M., et al., *Tumor-associated stromal cells as key contributors to the tumor microenvironment.* Breast Cancer Res, 2016. **18**(1): p. 84.
12. Gajewski, T.F., H. Schreiber, and Y.X. Fu, *Innate and adaptive immune cells in the tumor microenvironment.* Nat Immunol, 2013. **14**(10): p. 1014-22.
13. Quail, D.F. and J.A. Joyce, *The Microenvironmental Landscape of Brain Tumors.* Cancer Cell, 2017. **31**(3): p. 326-341.
14. Roesch, S., et al., *When Immune Cells Turn Bad-Tumor-Associated Microglia/Macrophages in Glioma.* Int J Mol Sci, 2018. **19**(2).
15. Xu, W.H., et al., *Screening and Identification of Potential Prognostic Biomarkers in Adrenocortical Carcinoma.* Front Genet, 2019. **10**: p. 821.
16. Rizvi, N.A., et al., *Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial.* Lancet Oncol, 2015. **16**(3): p. 257-65.
17. Carbone, D.P., et al., *First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer.* N Engl J Med, 2017. **376**(25): p. 2415-2426.
18. Zhang, C., et al., *Tumor Purity as an Underlying Key Factor in Glioma.* Clin Cancer Res, 2017. **23**(20): p. 6279-6291.
19. Sottini, A., et al., *Simultaneous quantification of recent thymic T-cell and bone marrow B-cell emigrants in patients with primary immunodeficiency undergone to stem cell transplantation.* Clin Immunol, 2010. **136**(2): p. 217-27.
20. Bacac, M., et al., *A Novel Carcinoembryonic Antigen T-Cell Bispecific Antibody (CEA TCB) for the Treatment of Solid Tumors.* Clin Cancer Res, 2016. **22**(13): p. 3286-97.
21. Gaffney, S.G., et al., *The landscape of novel and complementary targets for immunotherapy: an analysis of gene expression in the tumor microenvironment.* Oncotarget, 2019. **10**(44): p. 4532-4545.
22. Ryan, M., et al., *TCGASpliceSeq a compendium of alternative mRNA splicing in cancer.* Nucleic Acids Res, 2016. **44**(D1): p. D1018-22.

23. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets–update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.
24. Chan, B.K.C., *Data Analysis Using R Programming.* Adv Exp Med Biol, 2018. **1082**: p. 47-122.
25. Smyth, G.K., J. Michaud, and H.S. Scott, *Use of within-array replicate spots for assessing differential expression in microarray experiments.* Bioinformatics, 2005. **21**(9): p. 2067-75.
26. Wu, G., X. Feng, and L. Stein, *A human functional protein interaction network and its application to cancer data analysis.* Genome Biol, 2010. **11**(5): p. R53.
27. Subramanian, A., et al., *GSEA-P: a desktop application for Gene Set Enrichment Analysis.* Bioinformatics, 2007. **23**(23): p. 3251-3.
28. Varghese, F., et al., *IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples.* PLoS One, 2014. **9**(5): p. e96801.
29. Kreso, A. and J.E. Dick, *Evolution of the cancer stem cell model.* Cell Stem Cell, 2014. **14**(3): p. 275-91.
30. Brat, D.J., et al., *Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas.* N Engl J Med, 2015. **372**(26): p. 2481-98.
31. Lawrence, M.S., et al., *Discovery and saturation analysis of cancer genes across 21 tumour types.* Nature, 2014. **505**(7484): p. 495-501.
32. Capelle, L., et al., *Spontaneous and therapeutic prognostic factors in adult hemispheric World Health Organization Grade II gliomas: a series of 1097 cases: clinical article.* J Neurosurg, 2013. **118**(6): p. 1157-68.
33. Verburg, N. and P.C. de Witt Hamer, *State-of-the-art imaging for glioma surgery.* Neurosurg Rev, 2020.
34. Chang, J., et al., *The effect of operations in patients with recurrent diffuse low-grade glioma: A qualitative systematic review.* Clin Neurol Neurosurg, 2020. **196**: p. 105973.
35. Belykh, E., et al., *Blood-Brain Barrier, Blood-Brain Tumor Barrier, and Fluorescence-Guided Neurosurgical Oncology: Delivering Optical Labels to Brain Tumors.* Front Oncol, 2020. **10**: p. 739.
36. Lim, M., et al., *Current state of immunotherapy for glioblastoma.* Nat Rev Clin Oncol, 2018. **15**(7): p. 422-442.
37. McGranahan, T., et al., *Current State of Immunotherapy for Treatment of Glioblastoma.* Curr Treat Options Oncol, 2019. **20**(3): p. 24.
38. Huang, B., et al., *Advances in Immunotherapy for Glioblastoma Multiforme.* J Immunol Res, 2017. **2017**: p. 3597613.
39. Fischer, A., G. de Saint Basile, and F. Le Deist, *CD3 deficiencies.* Curr Opin Allergy Clin Immunol, 2005. **5**(6): p. 491-5.
40. Barber, E.K., et al., *The CD4 and CD8 antigens are coupled to a protein-tyrosine kinase (p56lck) that phosphorylates the CD3 complex.* Proc Natl Acad Sci U S A, 1989. **86**(9): p. 3277-81.

41. Borroto, A., et al., *The CD3 epsilon subunit of the TCR contains endocytosis signals.* J Immunol, 1999. **163**(1): p. 25-31.
42. Martin-Blanco, N., et al., *CD3ε recruits Numb to promote TCR degradation.* Int Immunol, 2016. **28**(3): p. 127-37.
43. Bacolod, M.D., et al., *Pathways- and epigenetic-based assessment of relative immune infiltration in various types of solid tumors.* Adv Cancer Res, 2019. **142**: p. 107-143.
44. Venteicher, A.S., et al., *Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq.* Science, 2017. **355**(6332).
45. Zhang, Y., et al., *An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex.* J Neurosci, 2014. **34**(36): p. 11929-47.
46. Matcovitch-Natan, O., et al., *Microglia development follows a stepwise program to regulate brain homeostasis.* Science, 2016. **353**(6301): p. aad8670.
47. Chen, Y., et al., *Calnexin Impairs the Antitumor Immunity of CD4(+) and CD8(+) T Cells.* Cancer Immunol Res, 2019. **7**(1): p. 123-135.
48. Bacolod, M.D., et al., *Immune infiltration, glioma stratification, and therapeutic implications.* Transl Cancer Res, 2016. **5**(Suppl 4): p. S652-s656.
49. Berger, M.F. and E.R. Mardis, *The emerging clinical relevance of genomics in cancer medicine.* Nat Rev Clin Oncol, 2018. **15**(6): p. 353-365.

Figures

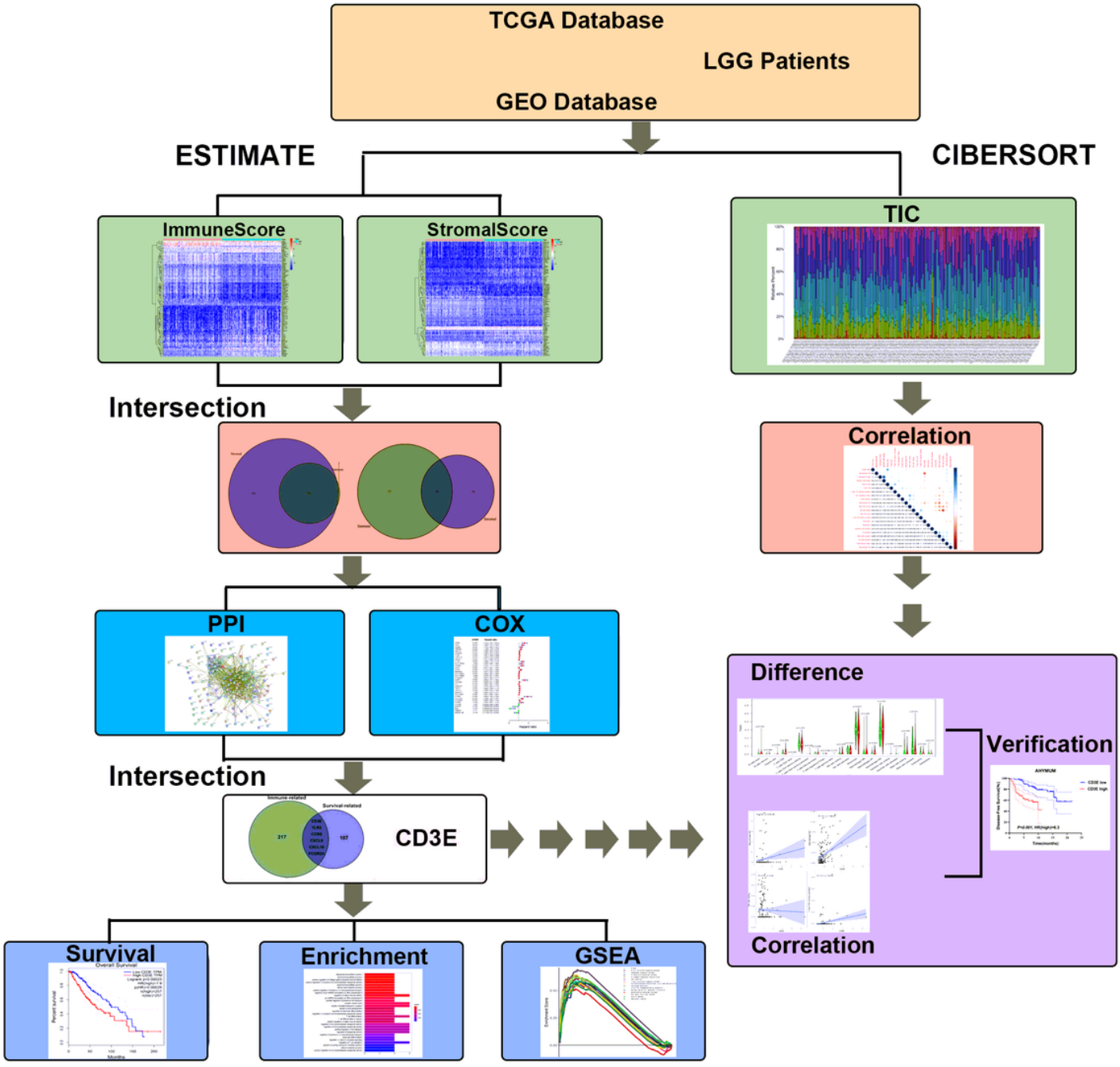

Figure 1

Flowchart of this study. TCGA, the Cancer Genome Atlas; GEO, Gene Expression Omnibus; LGG, low grade glioma;

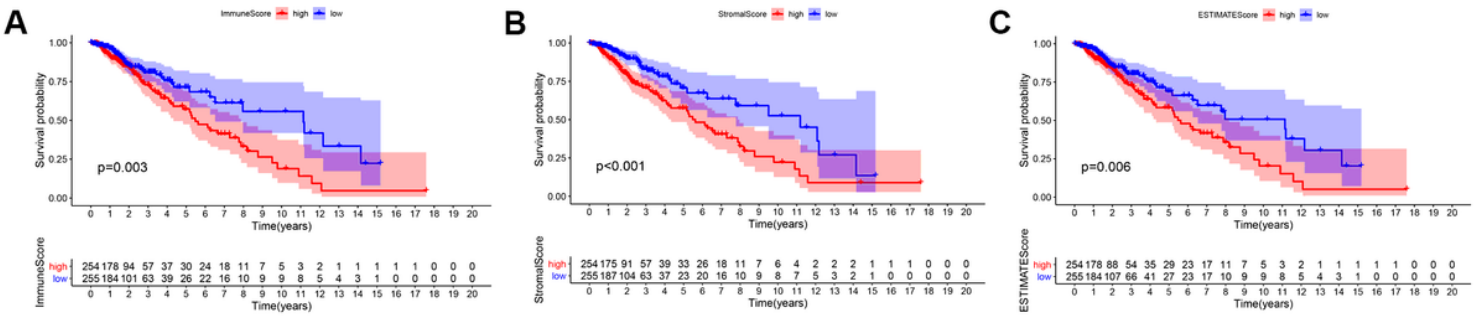

Figure 2

Correlation of scores with the survival of patients with LGG. (A) Kaplan–Meier survival analysis for Lgg patients grouped into high or low score in ImmuneScore determined by the comparison with the median, P = 0.003. (B) Kaplan–Meier survival curve for StromalScore, p 0.001 . (C) Survival analysis with Kaplan–Meier method for LGG patients grouped by ESTIMATEScore, P = 0.006.

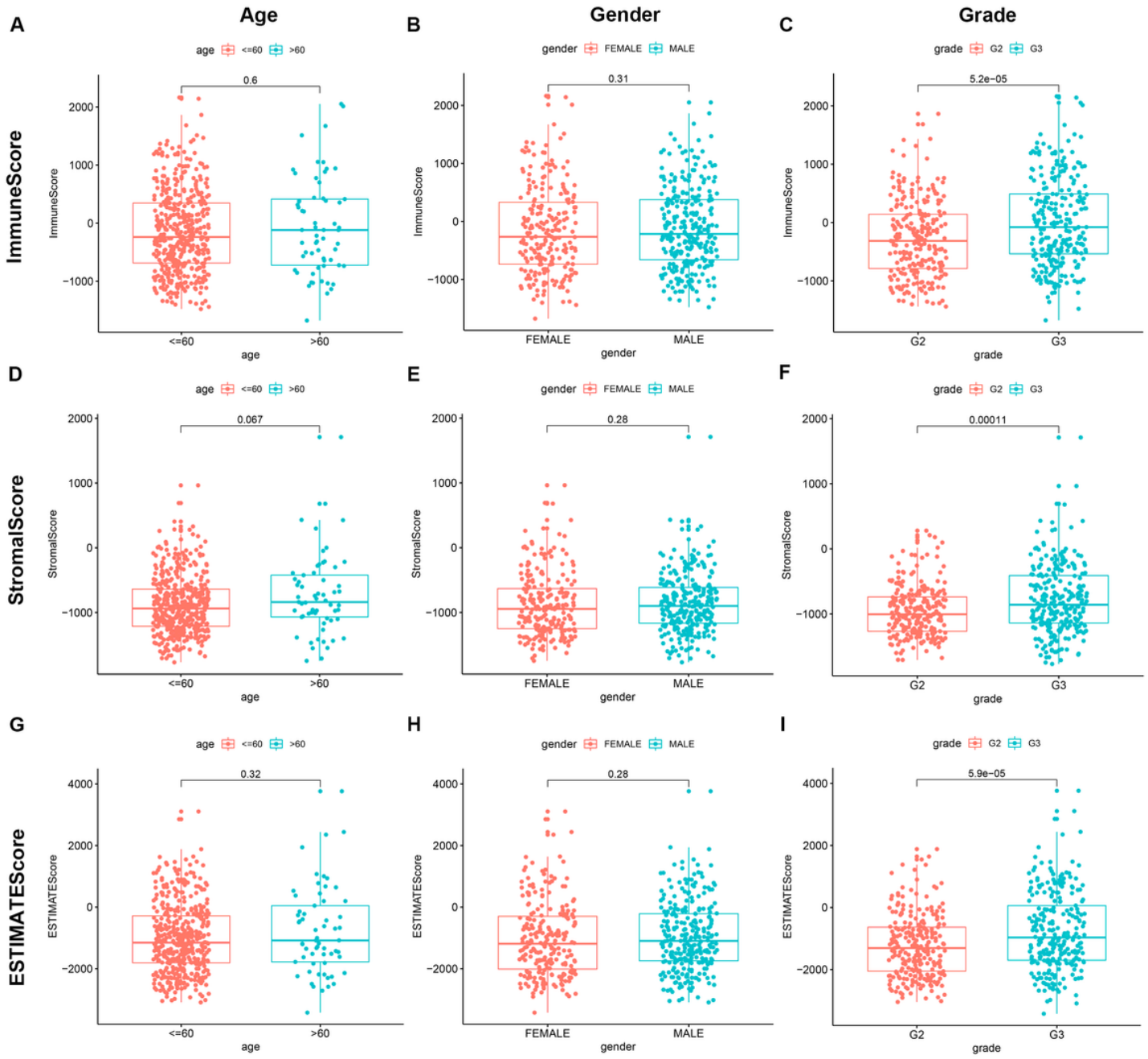

Figure 3

Correlation of ImmuneScore and StromalScore with clinicopathological staging characteristics. (A, D, G) Distribution of ImmuneScore, StromalScore, and ESTIMATEScore in age. P = 0.6, 0.067, and 0.32. (B, E, H) Distribution of three kinds of scores in gender. P = 0.31, 0.28, 0.28. (C, F, I) Distribution of scores in grade. p 0.001

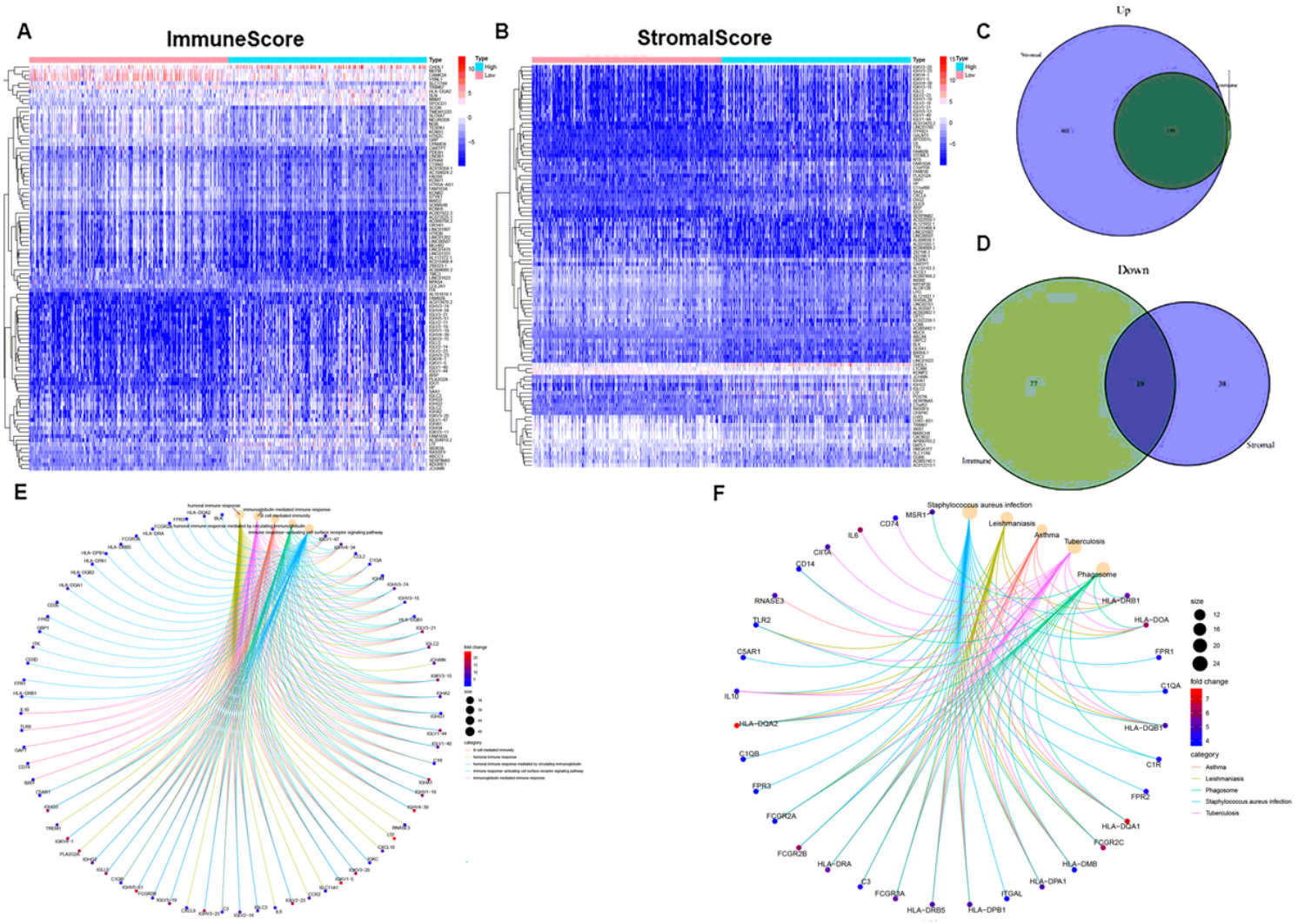

Figure 4

Heatmaps, Venn plots, and enrichment analysis of GO and KEGG for DEGs. (A) Heatmap for DEGs generated by comparison of the high score group vs. the low score group in ImmuneScore. Row name of heatmap is the gene name, and column name is the ID of samples which not shown in plot. Differentially expressed genes were determined by Wilcoxon rank sum test with q = 0.05 and fold-change >1 after log2 transformation as the significance threshold. (B) Heatmap for DEGs in StromalScore, similar with (A). (C,D) Venn plots showing common up-regulated and down-regulated DEGs shared by ImmuneScore and StromalScore (E,F) GO and KEGG enrichment analysis for DEGs

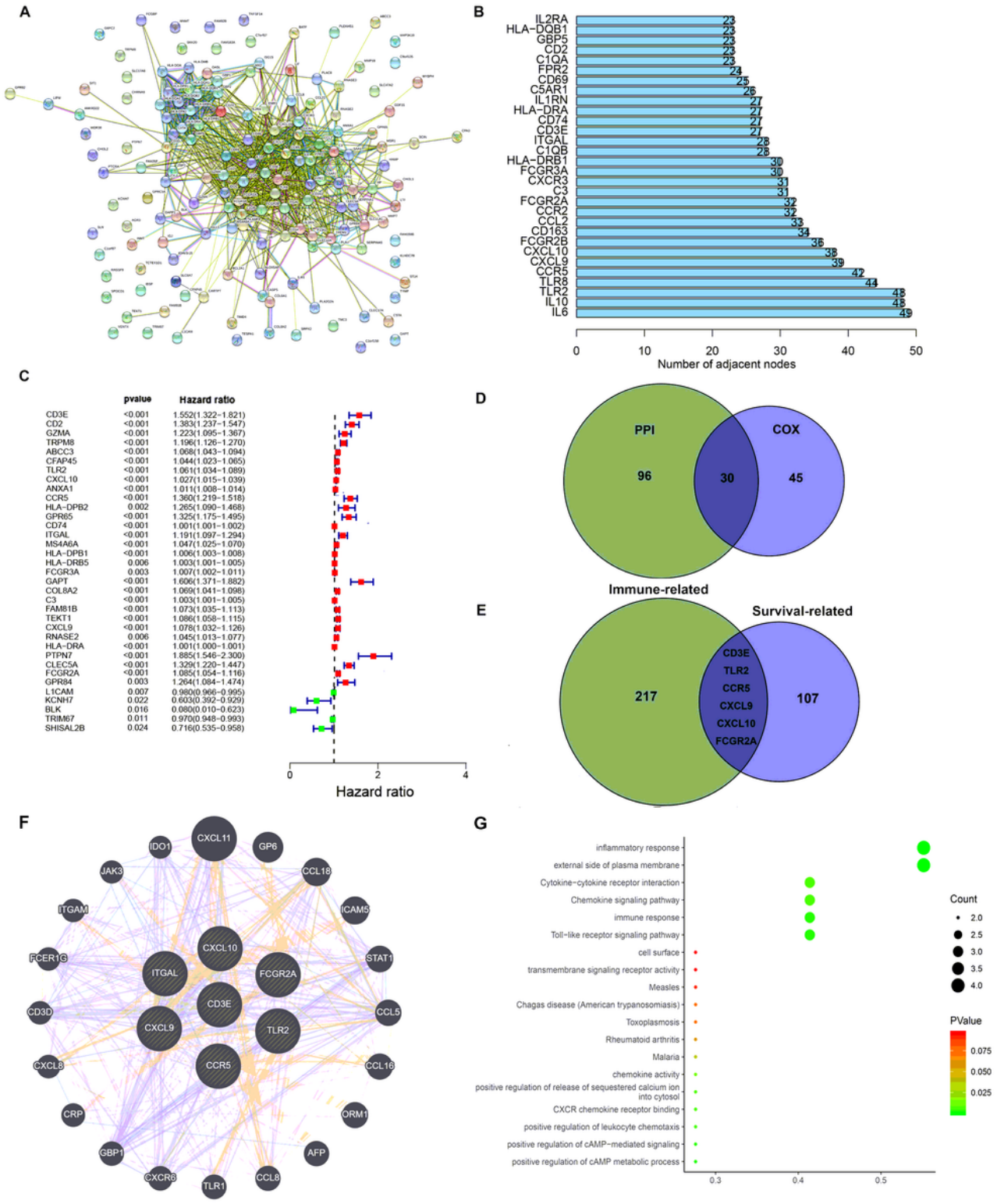

Figure 5

Protein–protein interaction network and univariate COX. (A) Interaction network constructed with String. (B) The top 30 genes ordered by the number of nodes. (C) Univariate COX regression analysis with DEGs, listing the top significant factors with P < 0.001. (D) Venn plot showing the common factors shared by nodes in PPI and top significant factors in univariate COX. (E) Venn plot showing the common factors

shared by nodes in Immune-related DEGs and Clinical-related DEGs. (F)Interaction network constructed with 7 genes. (G)GO and KEGG pathway analyses on 7 genes.

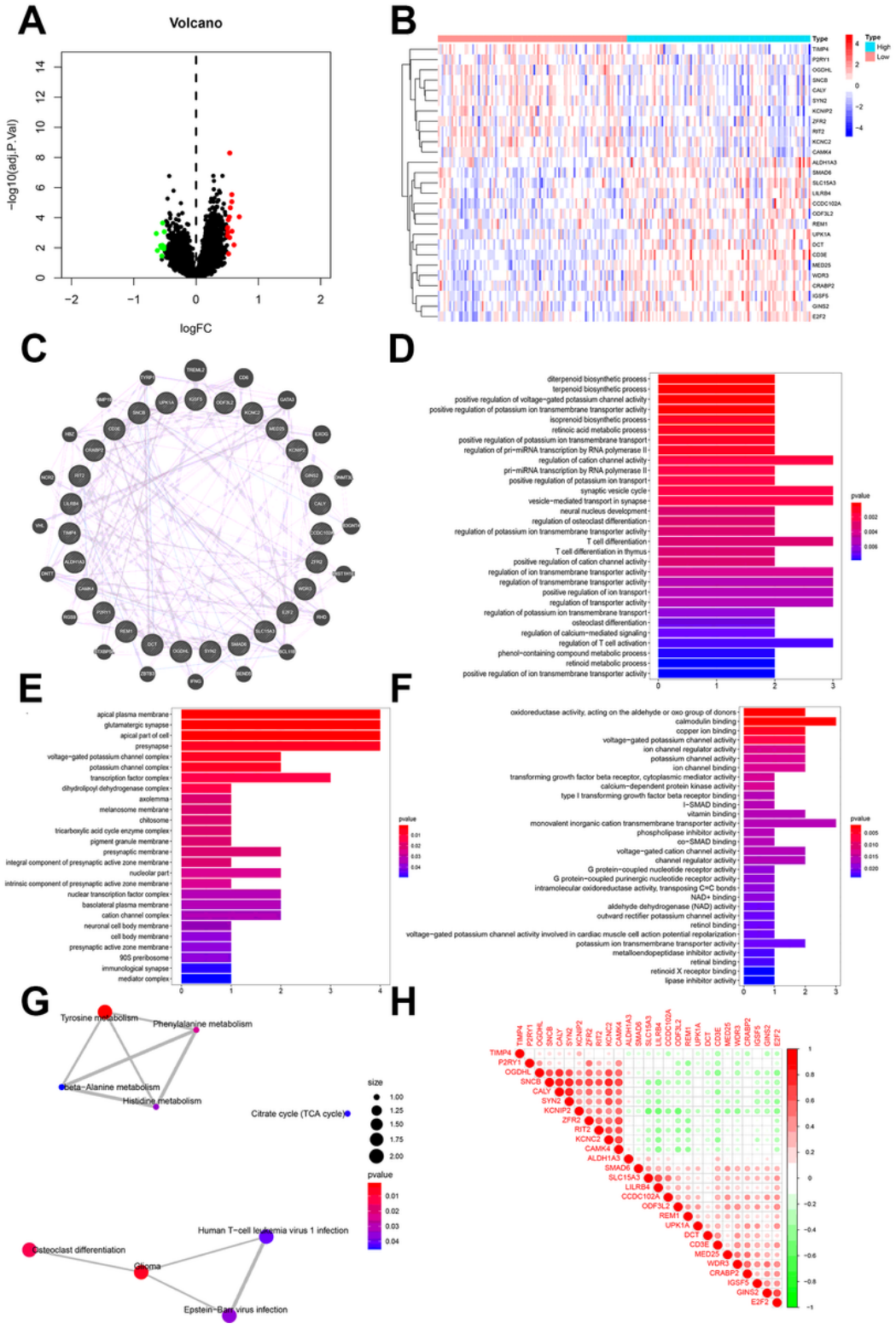

Figure 6

Correlation Analyses of Clinical-Related DEGs (A) volcano map of Clinical-Related DEGs. (B) heat map of Clinical-Related DEGs. (C)PPI of Clinical-Related DEGs co-expression (lavender line), co-localization (deep blue line). (D)enrichments of biological processes of DEGs. (E)enrichments of cellular components of

DEGs. (F)enrichments of molecular functions of DEGs. (G)Enrichments in KEGG pathway of DEGs. (H)The 20 most significantly up-regulated genes and the 20 most significantly down-regulated genes with CD3E. Red for positive correlation, and green represents a negative correlation. The deeper the color, the greater the relevance.

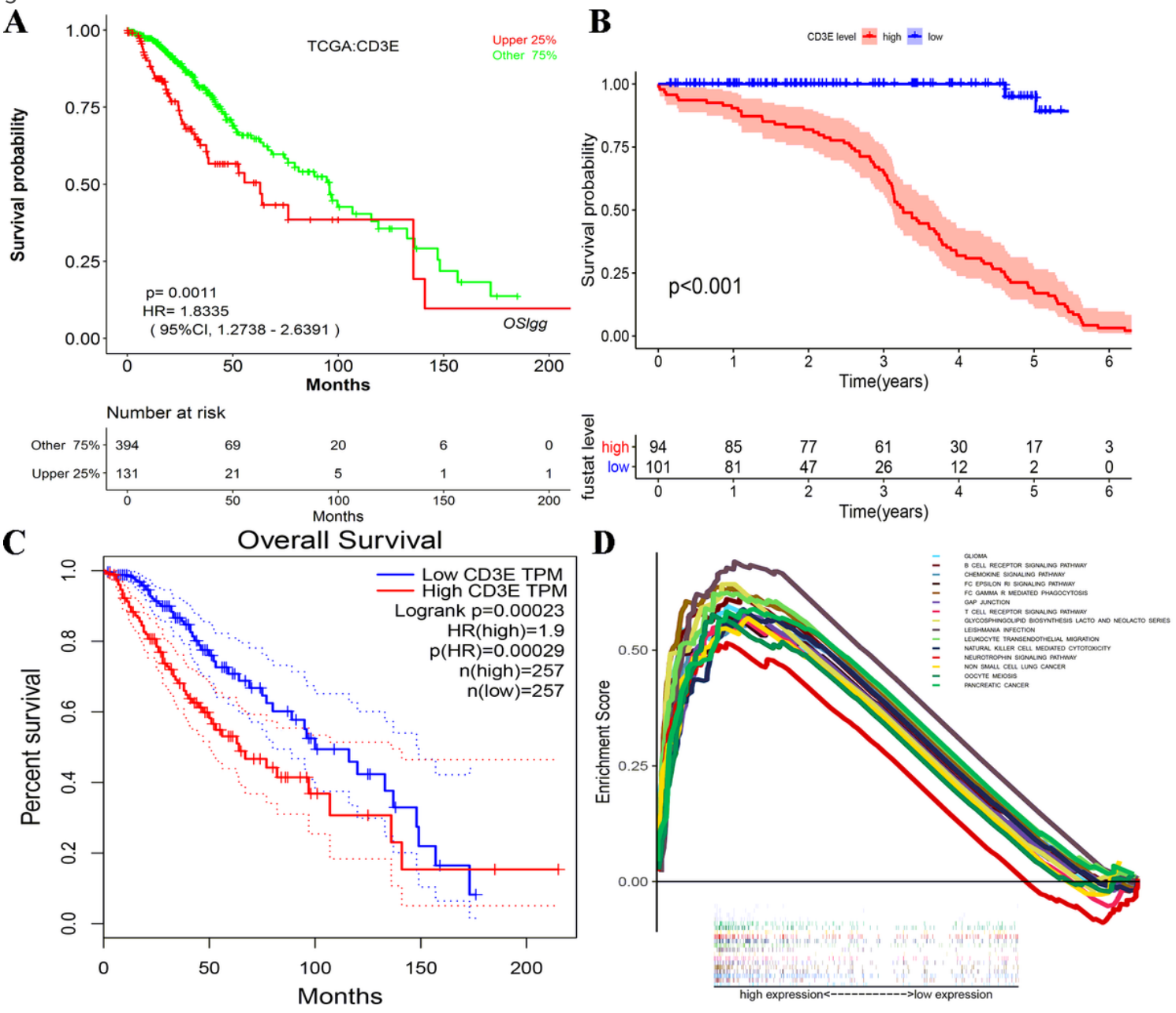

Figure 7

Relationship between CD3E expression and survival of LGG patients (A) Relationship between CD3E expression and survival of LGG patients in the TCGA database (P = 0.0011). (B) Relationship between CD3E expression and survival of LGG patients in the GSE database (P <0.001). (C) The relationship between CD3E expression and survival of LGG patients in the GEPIA database (P <0.001). (D)GSEA for samples with high CD3E expression.

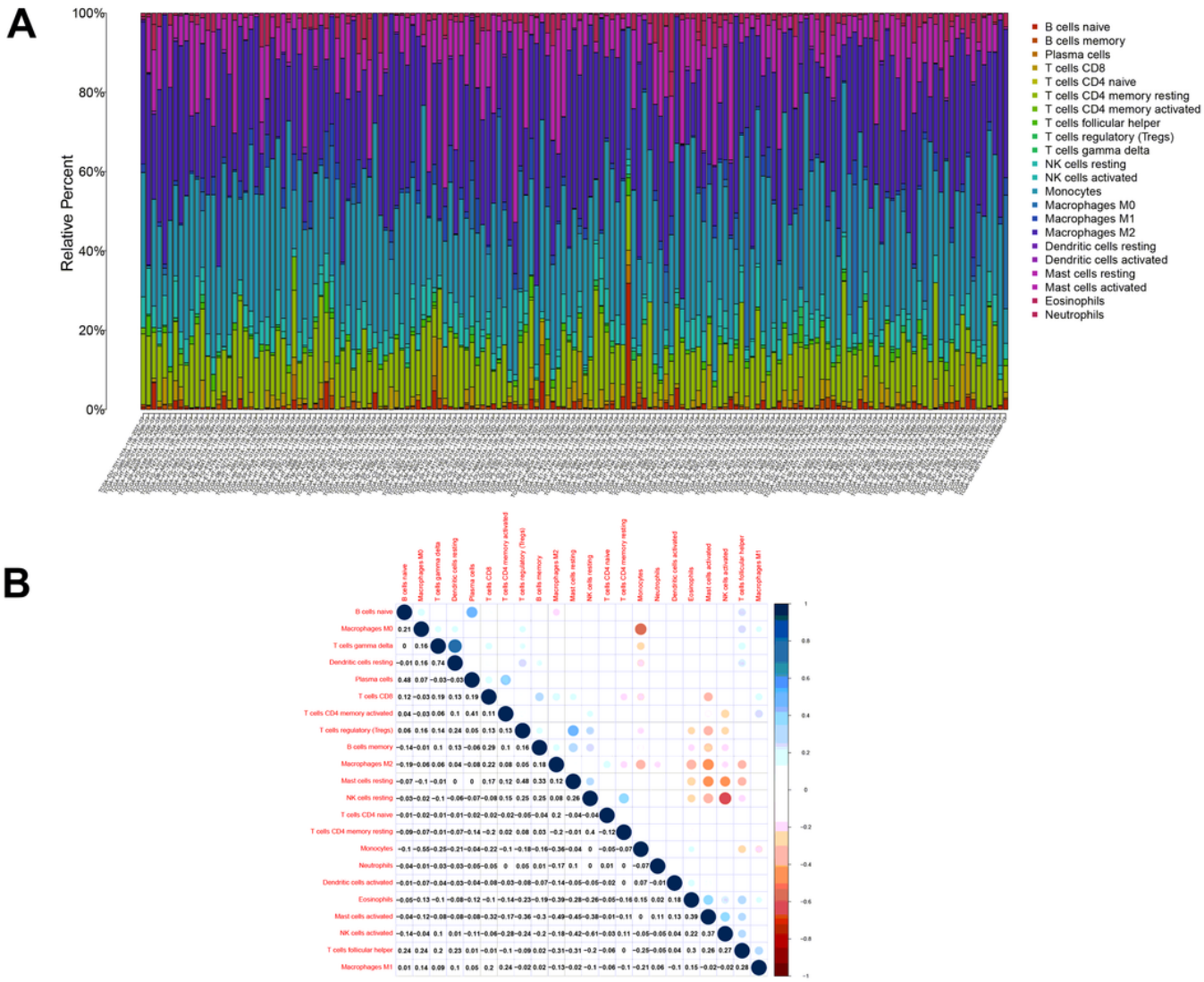

Figure 8

TIC profile in tumor samples and correlation analysis. (A) Barplot showing the proportion of 22 kinds of TICs in LGG tumor samples. Column names of plot were sample ID. (B) Heatmap showing the correlation between 22 kinds of TICs and numeric in each tiny box indicating the p value of correlation between two kinds of cells. The shade of each tiny color box represented corresponding correlation value between two cells, and Pearson coefficient was used for significance test.

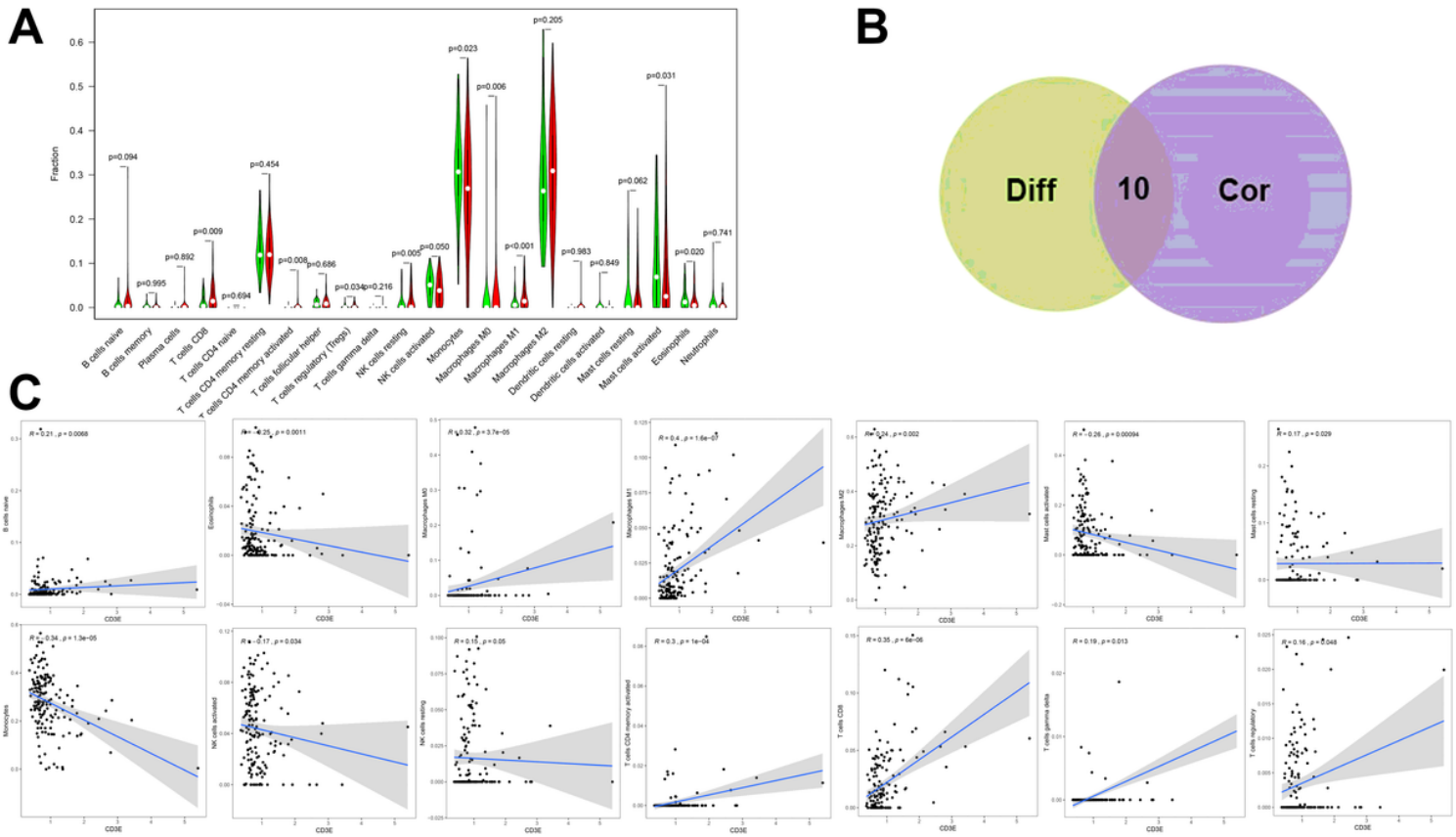

Figure 9

Correlation of TICs proportion with CD3E expression. (A) Violin plot showed the ratio differentiation of 22 kinds of immune cells between LGG tumor samples with low or high CD3E expression relative to the median of CD3E expression level, and Wilcoxon rank sum was used for the significance test. (B) Venn plot displayed ten kinds of TICs correlated with CD3E expression codetermined by difference and correlation tests displayed in violin and scatter plots, respectively. (C) Scatter plot showed the correlation of 14 kinds of TICs proportion with the CD3E expression ($P < 0.05$). The red line in each plot was fitted linear model indicating the proportion tropism of the immune cell along with CD3E expression, and Pearson coefficient was used for the correlation test.

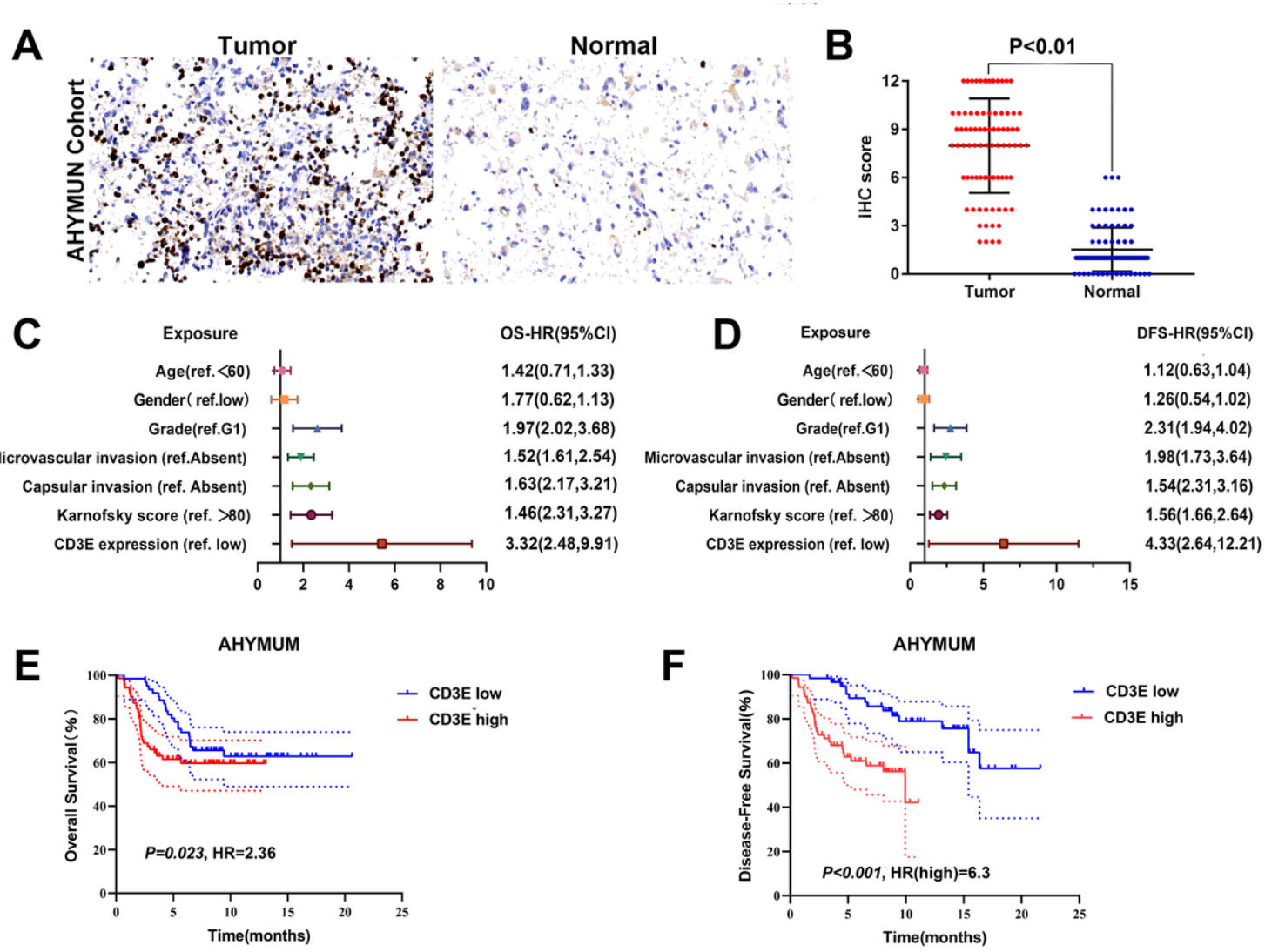

Figure 10

Further verification of CD3E gene and LGG prognosis. (A) IHC on collected LGG tissue. (B) The scatter plot of the IHC scores (P < 0.01). (C-D) Forest plots were used to visualize the univariate Cox regression analysis of OS and DFS in the AHYMUM cohorts. (E-F) Survival curves showed that LGG patients with elevated CD3E expression levels in the AHYMUN cohort showed poorer OS (P = 0.023) and poorer DFS (P < 0.001).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementfigure1.tif