The Journal of Pathology, ORIGINAL PAPER

Molecular subtyping reveals immune alterations in *IDH* wild-type lower-grade diffuse glioma

Fan Wu[1,2,3]†, Guan-Zhang Li[1,2]†, Han-Jie Liu[1,2]†, Zheng Zhao[1,2,3], Rui-Chao Chai[1,2,3], Yu-Qing Liu[1,2,3], Hao-Yu Jiang[1,2], You Zhai[1,2], Yue-Mei Feng[1,2], Ren-Peng Li[1,2], Wei Zhang[1,2,3,*]

[1]Department of Molecular Neuropathology, Beijing Neurosurgical Institute, Capital Medical University, Beijing, PR China

[2]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, PR China

[3]Chinese Glioma Genome Atlas Network (CGGA) and Asian Glioma Genome Atlas Network (AGGA), Beijing, PR China

†These authors contributed equally to this work.

*__Correspondence to:__ Wei Zhang, M.D.; Nan Si Huan Xi Lu 119, Fengtai District, Beijing 100070, PR China; Tel: +86 01059975624; Fax: +86 010167021832; E-mail: zhangwei_vincent@126.com.

Conflict of interest statements: The authors have no conflicts of interest to declare.

Word count: 3270.

Running title: Molecular subtyping of *IDH* wild-type diffuse LGG

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1002/path.5468

Abstract

Isocitrate dehydrogenase (*IDH*) wild-type diffuse lower-grade glioma (LGG) is usually associated with poor outcome, but there have been disputes over its clinical outcome and classification. We presented here a robust gene expression-based molecular classification of *IDH* wild-type diffuse LGG into two subtypes with distinct biological and clinical features. A discovery cohort of 49 *IDH* wild-type diffuse LGGs from Chinese Glioma Genome Atlas (CGGA) was subjected to clustering and function analysis. 73 tumors from The Cancer Genome Atlas (TCGA) were used to validate our findings. Consensus clustering of transcriptional data uncovered concordant classification of two robust and prognostically significant subtypes of *IDH* wild-type LGG. Sub1, associated with poorer outcomes, was characterized by significantly higher immune, cytolytic scores, M2 macrophages and up-regulation of immune exhaustion markers; while Sub2 who had elevated lymphocytes and plasma cells showed relatively favorable survival. Somatic alteration analysis revealed Sub1 showed more frequently deleted regions, such as locus of *CDKN2A*/*CDKN2B*, *DMRTA1*, *C9orf53* and *MTAP*. Furthermore, we developed and validated a five-gene signature for better application of this acquired stratification. Our data demonstrate the biological and prognostic heterogeneity within *IDH* wild-type diffuse LGGs and deepen our molecular understanding of this tumor entity.

Keywords: *IDH* wild-type diffuse LGG; molecular classification; immune signature; prognosis

Introduction

Diffuse World Health Organization (WHO) grade II and III gliomas, namely diffuse lower-grade gliomas (LGG), present as heterogeneous disease. They are infiltrative tumors that arise from glial or precursor cells, and show a more indolent course compared with glioblastoma (GBM, grade IV)[1]. Despite standard treatment, including neurosurgical resection, radio- and chemo-therapy, tumor recurrence and malignant progression are inevitable because of their highly invasive nature [2, 3]. The survival ranges widely (1 to 15 years) and varies considerably when stratified by tumor type.

In the new classification by the WHO, diffuse LGGs are classified into three diagnostic and prognostic subgroups based on *IDH* mutation and 1p/19q codeletion status [4]. *IDH* wild-type tumors with mostly higher grade (III) hold the worst prognosis, followed by tumors with *IDH* mutation and no 1p/19q co-deletion, and by tumors with *IDH* mutation and 1p/19q co-deletion (oligodendrogliomas). A recent study reported that the majority of *IDH* wild-type LGGs were underdiagnosed GBMs [5]. The outcome of *IDH* wild-type LGG has been shown to be indistinguishable from that of *IDH* wild-type GBM and worse than that of *IDH* mutant GBM[6]. Moreover, it is proved that not all *IDH* wild-type LGGs have a poor survival [7, 8]. This evidence inspires many groups to further stratify this tumor prognostically. Aibaidula *et al* found that *IDH* wild-type LGG was prognostically heterogeneous, and *MYB, EGFR, TERT* promoter and *H3F3A* should be examined to delineate distinct prognostic groups [9]. The cIMPACT-NOW group addressed the molecular definition of "molecular GBM", a cohort of *IDH* wild-type LGG characterized by either EGFR amplification, combined gains of chromosome 7 and loss of chromosome 10, or *TERT* promoter mutations [10]. A meta-analysis also assessed the prognostic values of several genetic markers, and found that *TERT* promoter, *H3F3A* mutation and

EGFR amplification have negative impact on survival of *IDH* wild-type LGG [11]. However, the transcriptional and biological heterogeneities within *IDH* wild-type diffuse LGGs need further dissection.

Increasing evidence has revealed that molecularly distinct subtypes of glioma differs in tumor microenvironment [12, 13]. Wang *et al* found tumor evolution of glioma-intrinsic gene expression subtypes associated with immunological changes in the microenvironment [14]. *NF1*-silenced high grade glioma exhibited increased tumor-associated macrophage infiltration [15, 16]. Bockmayr *et al* identified four distinct microenvironmental signatures of immune cell infiltration (vascular, monocytic/stromal, monocytic/T cell- and APC/NK/T cell-dominated immune clusters) in high grade glioma [17]. Zeiner *et al* reported that GBMs contain mixed M1/M2-like polarized tumor-associated microglia/macrophages (TAM) and the levels of different TAM subpopulations in tumor core are positively associated with overall survival in *IDH* wild-type GBMs [18]. A better understanding of tumor immune microenvironment is critical for improving the efficacy of current immunotherapies. Robert *et al* found that GBMs with mesenchymal gene expression signatures might be more responsive to immune-based therapies [19]. CAR T cells targeting *EGFRvIII* and *IL12R* was proved to be safe in preclinical studies of GBMs [20, 21]. Immunotherapies have received extensive attention in gliomas.

Here, we sought to determine whether molecular profiling of transcriptomic data could provide valuable stratification for *IDH* wild-type diffuse LGG. We performed molecular subtyping on 122 diffuse lower-grade gliomas of *IDH* wild-type from Chinese Glioma Genome Atlas (CGGA) and The Cancer Genome Atlas (TCGA). We revealed that *IDH* wild-type diffuse LGG could be classified into two robust subtypes with distinct prognostic and immune features. Our findings suggest that a subset of these gliomas may benefit from additional immunotherapy accompanied with the current treatments.

Materials and methods

Patients and datasets

A total of 122 *IDH* wild-type diffuse lower-grade gliomas (49 cases from CGGA microarray data and 73 cases from TCGA RNA-sequencing data) were collected in this study (supplementary material, Table S1). An informed consent was obtained from all patients. This study was carried out in accordance with the Declaration of Helsinki and approved by the ethics committee of Tiantan Hospital.

The microarray data and corresponding clinical information containing age, gender, histology, *IDH*, 1p/19q, methylguanine methyltransferase (*MGMT*) promoter status and survival information were downloaded from CGGA database as discovery cohort (http://www.cgga.org.cn) [22]. The RNA-sequencing data, somatic mutation and copy number alterations (CNAs) data and clinical information including *TERT* promoter and *ATRX* mutation status additionally were obtained from the TCGA database for validation cohort (http://cancergenome.nih.gov/) [23]. The clinical and molecular features are listed in supplementary material, Table S2.

Identification of gene expression-based subtypes

Unsupervised clustering was performed using the R package "ConsensusClusterPlus" for class discovery based on the comparison of gene expression profile[24, 25]. 80% item resampling, 50 resamplings and a maximum evaluated K of 10 were selected for clustering. The cumulative distribution function (CDF) and consensus heatmap were used to assess the optimal K.

Gene signature selection

Significance analysis of microarray (SAM) was applied to identify differentially expressed genes between Sub1 and Sub2. Univariate Cox regression analysis was performed to determine the genes with prognostic significance. Subsequently, the Cox proportional hazards model, which was suitable for high-dimensional regression analysis, was adopted for construction of the optimal gene set with R package "glmnet" [26, 27]. The linear combination of gene expression weighted by regression coefficients (Coeffs) was used to calculate the risk scores of patients.

Bioinformatic analysis

Gene ontology (GO) analysis were applied for functional annotation of differential genes between groups [28]. Gene set enrichment analysis (GSEA) was performed to determine gene sets of statistical difference [29]. Principal components analysis (PCA) was used to detect expression difference between groups with R package "princomp" [30]. Receiver operating characteristic (ROC) curve analysis was carried out for overall survival (OS) prediction with R package "pROC". ESTIMATE algorithm was performed to calculate the fraction of stromal and immune cells with R package "estimate", and tumor purity of each patient was inferred based on the formula described by Yoshihara [31]. The infiltrating immune cells were explored using CIBERSORT algorithm [32]. The cytolytic activity was calculated through an average expression of six genes, five granzymes and perforin-1 (PRF1), to assess the cytotoxic immune cell activity [33]. Tumor inflammation score were computed based on the formulas described previously [34, 35]. GISTIC2.0 analysis was adopted to assess CNAs between groups. Locus with GISTIC value more than 1 or less than -1 was defined as an amplification or deletion, respectively [36].

Immunohistochemical staining

Paraffin-embedded tissues of 12 cases from CGGA cohort were collected. Anti-CD163 [16646-1-AP, 1:300, Proteintech (Wuhan, Hubei, PR China)] was used to detect M2 macrophages. Anti-GZMA [ab10870, 1:300, Abcam (Burlingame, CA, USA)], anti-PD-1 (18106-1-AP, 1:300, Proteintech) and anti-HAVCR2 (60355-1-lg, 1:300, Proteintech) were used to evaluate the cytolytic activity and immune exhaustion, respectively. Briefly, the sections were incubated with antibody overnight at 4 °C, then incubated with secondary antibody at room temperature for 1 h. After washing, the sections were stained with DAB (MXB, Fuzhou, Fujian, PR China) for 2 min and counterstained with hematoxylin (Solarbio, Tongzhou, Beijing, PR China). Stained cells were counted three times for each photograph independently by two investigators. Five different fields of each section were examined for quantitative evaluation.

Statistical analysis

All statistical analyses were conducted using R software, GraphPad Prism 6.0 (GraphPad Inc., San Diego, CA, USA) and SPSS 16.0 (IBM, Chicago, IL, USA). $P < 0.05$ was considered statistically significant. Kaplan–Meier analyses with log-rank tests were performed to assess survival difference between groups. A Chi-square test was carried out to determine the difference of clinical and molecular parameter between groups. Univariate and multivariate Cox regression analyses were conducted to determine factors with independent prognostic value.

Results

***IDH* wild-type diffuse LGGs split into two expression subgroups**

To further stratify the patients of *IDH* wild-type lower-grade diffuse glioma, we obtained microarray data of 49 samples from the CGGA database and took an unbiased approach using consensus clustering of gene expression

profile. Hierarchical clustering of 2501 genes with highly variable expression (median absolute deviation (MAD) > 1) identified two main subtypes (Figure 1A), assessing by CDF curves and consensus matrices (supplementary material, Figure S1). Principal component analysis (PCA) found robust differences in the expression portraits between these two subtypes (Figure 1B). The group membership of each of the two subtypes was associated with distinct molecular and survival characteristics (Figure 1C). Patients in Sub1 had significantly poorer clinical outcome in comparation with ones in Sub2. In contrast, Sub2, with longer overall survival, contained more tumors of grade II (supplementary material, Figure S2A). Meanwhile, univariate Cox analysis showed that the acquired new stratification of *IDH* wild-type LGG had a significant prognostic value ($P = 0.022$, Table 1). Additionally, we also assessed the effect of three-classification on patients' prognosis, yet survival and PCA analysis revealed no obvious difference (supplementary material, Figure S3). Survival analysis based on progress free survival (PFS) also confirmed the above findings.

An independent set of 73 *IDH* wild-type diffuse LGG expression profiles was collected from TCGA database to assess subtype reproducibility. Applying a similar gene ordering from the training set (2365 genes available) in the validation set clearly recapitulated the subgroups identified in the CGGA data set (Figure 1D). PCA also confirmed the gene expression profile difference between these acquired subtypes (Figure 1E). Consistently, survival analysis revealed that Sub1 mainly consisted of grade III tumors with shorter overall survival (Figure 1F and supplementary material, S2A). Referring to EM/PM classification [37], Sub2 had more PM and EM-low/PM-low tumors which implied better outcome (supplementary material, Figure S2B). Moreover, univariate Cox analysis also demonstrated this classification was of prognostic significance in the TCGA cohort ($P = 0.039$, Table 1).

The subtypes differ in biological functions

Next, we analyzed the functional context of these two subtypes of *IDH* wild-type diffuse LGG. Gene ontology (GO) analysis based on the differentially expressed genes between groups, which were identified by SAM (Figure 2A, false discovery rate < 0.05), showed that the highly expressed genes in Sub1 were mainly enriched in immune, inflammatory response, antigen processing and presentation and interferon gamma mediated signaling pathway. Instead, the upregulated genes in Sub2 were annotated to chemical synaptic transmission, nervous system development, neurotransmitter and glutamate section (supplementary material, Figure S4A). Moreover, GSEA further found that immune and inflammatory response were significantly enriched in patients of Sub1 (Figure 2C and supplementary material, Figure S4B). Then, these functional analyses were repeated on the TCGA cohort. GO and GSEA also reproduced the enriched functions of these two subtypes (Figure 3A–C, and supplementary material, Figure S4C,D).

To further decipher the immune heterogeneity between these two subtypes, we resorted to immune-related tools published recently. We first examined the distribution of stromal and immune content of each group by computing Estimate algorithm [31]. Sub1 had significantly higher stromal and immune scores compared to Sub2 (Figures 2D and 3D). When comparing tumor purity, we observed a reduction in Sub1 (Figure 2D), suggesting that tumors in Sub1 contained higher number of immune cells in both CGGA and TCGA cohorts. Since natural anti-tumor immunity requires a cytolytic immune response, we next examined the immune cell-mediated cytolytic activity by quantifying the average expression of granzymes and *PRF1*. Consistent with the immune scores, the cytolytic scores of tumors in Sub1 were significantly higher than those of Sub2 (Figures 2E and 3E). Further, we computed the tumor inflammation signature algorithm and found that Sub1 presented

higher inflammation scores when compared to Sub2 (Figures 2F and 3F). we also looked into the expression of major histocompatibility class 1 (MHC1) complex which is required to present endogenous cellular antigens to circulating T cells [38]. The expression of *HLA-A*, *HLA-B* and *HLA-C* was significantly increased in cases of Sub1 in two cohorts (Figures 2G and 3G).

We next explored the composition of infiltrating immune cells between groups by CIBERSORT method [32]. Cases in Sub1 showed the higher enrichment of M2 macrophages, while Sub2 tumors were enriched in lymphocytes, naïve B cells and plasma cells (Figure 2H). In cohort of TCGA, we observed similar results (Figure 3H). Additionally, we detected the expression of several exhaustion marker genes, since T cell and natural killer cell exhaustion has been identified as an important mechanism by which cancer cells escape host immunity [39, 40]. We observed that most of these exhaustion marker genes (*LAG3*, *CTLA4*, *PD1*, *PD-L1* and *HAVCR2*) were expressed highly in tumors of Sub1 in both CGGA and TCGA cohorts (supplementary material, Figure S5A,B), indicating an elevated level of immune exhaustion in these tumors. We further selected some classical markers for IHC validation (*CD163* for M2 macrophages, *GZMA* for immune cytolytic activity, *PD1* and *HAVCR2* for immune exhaustion). Twelve *IDH* wild-type tumor samples from CGGA cohort were used (n = 6 per subtype). As shown in supplementary material, Figure S5C,D, these four genes showed higher protein expression level in sub1, which was consistent with results of transcriptional sequencing data. Collectively, these results identified an extensive immune heterogeneity in *IDH* wild-type diffuse LGGs, as illustrated by several immune signatures and molecular features.

An inflammatory signature is associated with outcome in *IDH* wild-type diffuse LGG

Infiltrating immune cells derived from gene expression profiles have been shown to be prognostic in many

human cancers [41, 42]. We therefore determined prognostic values of differential immune cell signatures between groups in *IDH* wild-type LGG. Out of these immune signatures, the cytolytic and inflammatory signatures were significantly associated with outcome of patients in CGGA cohort (Figure 4A), as well as PFS analyses (supplementary material, Figure S6). Figure 4B revealed that high inflammatory scores implied poor outcome in the TCGA cohort. Additionally, the inflammatory and cytolytic signatures were correlated with M1, M2 macrophage, CD8+ T cell, Neutrophils and memory activated CD4+ T cell (Figure 4C). Likely, a correlation between inflammatory scores, cytolytic scores and immune inhibitory genes (*PD-1* and *CTLA4*) was observed (Figure 4D). These findings suggested that inflammatory microenvironment and infiltrating cytolytic immune cells might impact patient outcome in *IDH* wild-type diffuse LGG.

Somatic variations in two subgroups

Several genomic alterations have negative impact on survival of *IDH* wild-type diffuse LGG [9, 11]. We also analyzed the somatic mutations and CNVs data from TCGA database to explore the difference of genomic alterations between these two subtypes. First, we compared the frequency of mutations and found no significant enrichment of mutations between two groups (Figure 5A). In contrast, GISTIC2.0 analysis revealed distinct copy number alterations between Sub1 and Sub2 tumors. Sub1 showed more frequently deleted regions, such as locus of *CDKN2A*/*CDKN2B*, *DMRTA1*, *C9orf53* and *MTAP* (Figure 5B).

Identification of a classification-related prognostic signature for *IDH* wild-type diffuse LGG

Considering the convenient application of the new classification for prognostic prediction, we proposed to develop an easier gene set based on the differential genes between Sub1 and Sub2 of TCGA cohort. SAM analysis identified 1453 differentially expressed genes (false discovery rate < 0.05), wherein 326 genes were

significantly correlated with patients' survival in univariate Cox regression analysis ($P < 0.05$, Figure 6A). Then, we performed a Cox proportional hazards model for selecting gene set with best prognostic value (Figure 6B). A five-gene signature was obtained and risk score of each case was computed with the gene expression level and regression coefficient (Figure 6C).

Subsequently, tumors were assigned into low or high-risk group based on the median value of risk scores. Kaplan–Meier analysis revealed that patients in high-risk group had significantly poorer outcome than those in low-risk group ($P < 0.001$, Figure 6D). We also calculated the risk scores of samples in CGGA cohort with the same coefficients to validate this signature (Figure 6F). Consistently, the survival curve showed that high-risk patients had shorter overall survival than low-risk ones (Figure 6G). We also detected the association between this signature and pathological features. Patients were arrayed based on the risk scores. As shown in Figure 6E and H, the risk scores distributed differently in patients stratified by grade, subtype and EM/PM classification. We further evaluated the prognostic independence of this signature by performing univariate and multivariate Cox regression analyses. As expected, the acquired signature was significantly correlated with patients' overall survival, independent of other factors, whereas only in univariate analysis of CGGA cohort (Table 1). Afterwards, we assessed the predictive accuracy by computing area under the curve (AUC) of risk score, age and group. The AUC of risk score (80.5% in TCGA cohort, 72.2% in CGGA cohort) was much higher than that of other factors (Figure 6I). These results demonstrated the superior performance of this signature for prognosis prediction.

Application and functional annotation of the acquired signature

The signature was further applied in diffuse gliomas to determine the compatibility. 550 samples from TCGA database and 299 samples from CGGA database (supplementary material, Table S2) were included in survival analyses. High risk scores conferred shorter OS in these two sets (supplementary material, Figure S7A and E). When stratified by grade and molecular subtypes (*IDH* and 1p/19q), high scores tended to be associated with poor outcome in cases of LGG, whereas in GBM and other molecular groups found no significant differences (supplementary material, Figure S7B–H). The relatively shorter follow-up period of CGGA cohort might affect the consistency of survival analyses.

We next compared gene expression between low and high-risk patients to determine the functional differences. Based on the differential genes identified by SAM, GO analysis found that immune, inflammatory response and interferon gamma mediated signaling pathway were significantly enriched in high-risk tumors, while the low-risk ones showed enrichment of chemical synaptic transmission, neurotransmitter and glutamate section (supplementary material, Figure S8A,B). Meanwhile, GSEA also confirmed this finding (supplementary material, Figure S8C,D). Correlation analyses found that this signature was associated with inflammatory, cytolytic signatures and immune inhibitory genes (*PD-1* and *CTLA4*) (supplementary material, Figure S8E–L). These results indicated that the signature derived from the new classification could represent the similar biological differences in *IDH* wild-type diffuse LGG.

Discussion

Numerous genetic alterations have been found in glioma, such as *IDH*, *TERT* promoter mutation and 1p/19q codeletion [43, 44]. The prognostic role of *IDH* mutation is widely accepted and *IDH* wild-type gliomas show

distinct biological and clinical characteristics compared with *IDH* mutant gliomas [45, 46]. The new WHO classification has segregated GBMs into *IDH* wild-type and mutant groups, which have significant difference in prognosis [4]. However, *IDH* wild-type diffuse LGG has be described as provisional entity, and its definition and typing are still controversial.

Here, we showed that transcriptome profiling defined two subtypes with distinct clinical and biological features in *IDH* wild-type diffuse LGG. The reproducibility of this classification was demonstrated in an independent validation set, and consistent phenotype was observed. Aggressive gliomas are characterized by higher stromal, immune scores and cytolytic activity, but lower purity [33, 47, 48]. We found that cases in Sub1 who had a poorer outcome showed similar immune phenotype in functional annotation. In addition, we observed M2 macrophages in tumors of Sub1, which have protumor activity in glioma [49, 50]. While Sub2 tumors displayed higher levels of naïve B cells and plasma cells, and these immune cells are favorable for patients' survival in glioma [41]. The presence of tumor infiltrating B cells has a paradoxical effect in GBMs. Candolfi *et al* showed that B cells can act as APC for T cells and play a critical role in T cell-mediated antitumor immunity [51]. In contrast, Lee-Chang *et al* recently found that GBM-associated B cells showed an immunosuppressive function toward activated CD8+ T cells [52]. Although naïve B cells exhibited elevated level in Sub2 which had better outcome, the function of B cells in *IDH* wild-type diffuse LGG needed to be further inquired. In addition, GO and GSEA analyses found that Sub2 showed higher expression genes which were involved in chemical synaptic transmission, neurotransmitter and glutamate section. A recent study reported by Venkatesh *et al* found that synaptic and electrical integration in neural circuits promotes glioma progression [53], which implied the underlying mechanism of progression in Sub2 tumors.

When we used these immune-related scores to identify predictors of survival, we found that higher cytolytic activity was associated with adverse survival in this tumor entity (Figure 4A), which was contrary to most studies of cancers. These indicated that cytolytic activity might not be in function in this tumor. On the other hand, despite high immune and cytolytic activity, the higher expression of immune exhaustion genes in Sub1 implied an immune-suppressive status, which might lead to the poorer outcome.

To summarize, our data indicated that transcriptomic profiling stratified *IDH* wild-type lower-grade diffuse glioma into two distinct subgroups with different clinical and biological phenotypes. We believe that this classification has implications for differential therapeutic strategies and will lead to targeted treatment for patients with this tumor.

Acknowledgements

The authors conducting this work represent the Chinese Glioma Cooperative Group (CGCG). We acknowledge funding from the National Natural Science Foundation of China (NSFC)/Research Grants Council (RGC) Joint Research Scheme (81761168038), National Key Research and Development Plan (2016YFC0902500) and National Natural Science Foundation of China (81672479, 81903078).

Author contributions statement

WZ conceived and designed the study. FW, G-ZL and H-JL performed most of analysis with assistance from R-CC, Y-QL, ZZ, H-YJ, YZ, Y-MF and R-PL.

References

1. Lapointe S, A Perry and NA Butowski. Primary brain tumours in adults. *Lancet* 2018; **392**: 432-446.
2. Cancer Genome Atlas Research Network, DJ Brat, RG Verhaak, *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* 2015; **372**: 2481-98.
3. Jiang T, Y Mao, W Ma, *et al.* CGCG clinical practice guidelines for the management of adult diffuse gliomas. *Cancer Lett* 2016; **375**: 263-273.
4. Louis DN, A Perry, G Reifenberger, *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 2016; **131**: 803-20.
5. Reuss DE, A Kratz, F Sahm, *et al.* Adult IDH wild type astrocytomas biologically and clinically resolve into other tumor entities. *Acta Neuropathol* 2015; **130**: 407-17.
6. Tabouret E, AT Nguyen, C Dehais, *et al.* Prognostic impact of the 2016 WHO classification of diffuse gliomas in the French POLA cohort. *Acta Neuropathol* 2016; **132**: 625-34.
7. Chan AK, Y Yao, Z Zhang, *et al.* TERT promoter mutations contribute to subset prognostication of lower-grade gliomas. *Mod Pathol* 2015; **28**: 177-86.
8. Chan AK, Y Yao, Z Zhang, *et al.* Combination genetic signature stratifies lower-grade gliomas better than histological grade. *Oncotarget* 2015; **6**: 20885-901.
9. Aibaidula A, AK Chan, Z Shi, *et al.* Adult IDH wild-type lower-grade gliomas should be further stratified. *Neuro Oncol* 2017; **19**: 1327-1337.
10. Brat DJ, K Aldape, H Colman, *et al.* cIMPACT-NOW update 3: recommended diagnostic criteria

for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV". *Acta Neuropathol* 2018; **136**: 805-810.

11. Vuong HG, TTK Tran, HTT Ngo*, et al.* Prognostic significance of genetic biomarkers in isocitrate dehydrogenase-wild-type lower-grade glioma: the need to further stratify this tumor entity - a meta-analysis. *Eur J Neurol* 2019; **26**: 379-387.
12. Chen Z and D Hambardzumyan. Immune Microenvironment in Glioblastoma Subtypes. *Front Immunol* 2018; **9**: 1004.
13. Zhang B, R Shen, S Cheng*, et al.* Immune microenvironments differ in immune characteristics and outcome of glioblastoma multiforme. *Cancer Med* 2019; **8**: 2897-2907.
14. Wang Q, B Hu, X Hu*, et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* 2017; **32**: 42-56 e6.
15. Herting CJ, Z Chen, KL Pitter*, et al.* Genetic driver mutations define the expression signature and microenvironmental composition of high-grade gliomas. *Glia* 2017; **65**: 1914-1926.
16. Doucette T, G Rao, A Rao*, et al.* Immune heterogeneity of glioblastoma subtypes: extrapolation from the cancer genome atlas. *Cancer Immunol Res* 2013; **1**: 112-22.
17. Bockmayr M, F Klauschen, CL Maire*, et al.* Immunological profiling of mutational and transcriptional subgroups in pediatric and adult high-grade gliomas. *Cancer Immunol Res* 2019;
18. Zeiner PS, C Preusse, A Golebiewska*, et al.* Distribution and prognostic impact of microglia/macrophage subpopulations in gliomas. *Brain Pathol* 2019; **29**: 513-529.

19. Prins RM, H Soto, V Konkankit*, et al.* Gene expression profile correlates with T-cell infiltration and relative survival in glioblastoma patients vaccinated with dendritic cell immunotherapy. *Clin Cancer Res* 2011; **17**: 1603-15.

20. Johnson LA, J Scholler, T Ohkuri*, et al.* Rational development and characterization of humanized anti-EGFR variant III chimeric antigen receptor T cells for glioblastoma. *Sci Transl Med* 2015; **7**: 275ra22.

21. Brown CE, B Badie, ME Barish*, et al.* Bioactivity and Safety of IL13Ralpha2-Redirected Chimeric Antigen Receptor CD8+ T Cells in Patients with Recurrent Glioblastoma. *Clin Cancer Res* 2015; **21**: 4062-72.

22. Yan W, W Zhang, G You*, et al.* Molecular classification of gliomas based on whole genome gene expression: a systematic report of 225 samples from the Chinese Glioma Cooperative Group. *Neuro Oncol* 2012; **14**: 1432-40.

23. Ceccarelli M, FP Barthel, TM Malta*, et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* 2016; **164**: 550-63.

24. Wilkerson MD and DN Hayes. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; **26**: 1572-3.

25. Wu F, RC Chai, Z Wang*, et al.* Molecular classification of IDH-mutant glioblastomas based on gene expression profiles. *Carcinogenesis* 2019;

26. Hughey JJ and AJ Butte. Robust meta-analysis of gene expression using the elastic net. *Nucleic Acids Res* 2015; **43**: e79.

27. Liu YQ, RC Chai, YZ Wang, *et al.* Amino acid metabolism-related gene expression-based risk signature can better predict overall survival for glioma. *Cancer Sci* 2019; **110**: 321-333.

28. Huang da W, BT Sherman and RA Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44-57.

29. Subramanian A, P Tamayo, VK Mootha, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**: 15545-50.

30. Zhou Z, R Huang, R Chai, *et al.* Identification of an energy metabolism-related signature associated with clinical prognosis in diffuse glioma. *Aging (Albany NY)* 2018; **10**: 3185-3209.

31. Yoshihara K, M Shahmoradgoli, E Martinez, *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013; **4**: 2612.

32. Newman AM, CL Liu, MR Green, *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015; **12**: 453-7.

33. Rooney MS, SA Shukla, CJ Wu, *et al.* Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015; **160**: 48-61.

34. Ayers M, J Lunceford, M Nebozhyn, *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017; **127**: 2930-2940.

35. Pinto JA, J Araujo, NK Cardenas, *et al.* A prognostic signature based on three-genes expression in triple-negative breast tumours with residual disease. *NPJ Genom Med* 2016; **1**: 15015.

36. Mermel CH, SE Schumacher, B Hill, *et al.* GISTIC2.0 facilitates sensitive and confident localization

of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12**: R41.

37. Sun Y, W Zhang, D Chen, *et al.* A glioma classification scheme based on coexpression modules of EGFR and PDGFRA. *Proc Natl Acad Sci U S A* 2014; **111**: 3538-43.

38. Reeves E and E James. Antigen processing and immune regulation in the response to tumours. *Immunology* 2017; **150**: 16-24.

39. Zarour HM. Reversing T-cell Dysfunction and Exhaustion in Cancer. *Clin Cancer Res* 2016; **22**: 1856-64.

40. Gonzalez-Gugel E, M Saxena and N Bhardwaj. Modulation of innate immunity in the tumor microenvironment. *Cancer Immunol Immunother* 2016; **65**: 1261-8.

41. Gentles AJ, AM Newman, CL Liu, *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015; **21**: 938-945.

42. Wei JS, IB Kuznetsov, S Zhang, *et al.* Clinically Relevant Cytotoxic Immune Cell Signatures and Clonal Expansion of T-Cell Receptors in High-Risk MYCN-Not-Amplified Human Neuroblastoma. *Clin Cancer Res* 2018; **24**: 5673-5684.

43. Brennan CW, RG Verhaak, A McKenna, *et al.* The somatic genomic landscape of glioblastoma. *Cell* 2013; **155**: 462-77.

44. Hu H, Q Mu, Z Bao, *et al.* Mutational Landscape of Secondary Glioblastoma Guides MET-Targeted Trial in Brain Tumor. *Cell* 2018; **175**: 1665-1678 e18.

45. Yan H, DW Parsons, G Jin, *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* 2009; **360**:

765-73.

46. Parsons DW, S Jones, X Zhang*, et al.* An integrated genomic analysis of human glioblastoma multiforme*. Science* 2008; **321**: 1807-12.

47. Zhang C, W Cheng, X Ren*, et al.* Tumor Purity as an Underlying Key Factor in Glioma*. Clin Cancer Res* 2017; **23**: 6279-6291.

48. Aran D, M Sirota and AJ Butte. Systematic pan-cancer analysis of tumour purity*. Nat Commun* 2015; **6**: 8971.

49. Morisse MC, S Jouannet, M Dominguez-Villar*, et al.* Interactions between tumor-associated macrophages and tumor cells in glioblastoma: unraveling promising targeted therapies*. Expert Rev Neurother* 2018; **18**: 729-737.

50. Ellert-Miklaszewska A, P Wisniewski, M Kijewska*, et al.* Tumour-processed osteopontin and lactadherin drive the protumorigenic reprogramming of microglia and glioma progression*. Oncogene* 2016; **35**: 6366-6377.

51. Candolfi M, JF Curtin, K Yagiz*, et al.* B cells are critical to T-cell-mediated antitumor immunity induced by a combined immune-stimulatory/conditionally cytotoxic therapy for glioblastoma*. Neoplasia* 2011; **13**: 947-60.

52. Lee-Chang C, A Rashidi, J Miska*, et al.* Myeloid-Derived Suppressive Cells Promote B cell-Mediated Immunosuppression via Transfer of PD-L1 in Glioblastoma*. Cancer Immunol Res* 2019; **7**: 1928-1943.

53. Venkatesh HS, W Morishita, AC Geraghty*, et al.* Electrical and synaptic integration of glioma into

neural circuits. *Nature* 2019; **573**: 539-545.

Figure legends

Figure 1. Unbiased clustering identified two subtypes with distinct outcome in *IDH* wild-type diffuse LGG. (A) Heat map of two subtypes defined by 2501 genes with highly variate expression. MAD > 1. (B) PCA analysis of two subtypes in CGGA cohort. (C) Kaplan–Meier analysis of two clusters based on overall survival (OS) and progression-free survival (PFS). (D) Gene order of training set was applied in TCGA cohort as validation set. (E) PCA analysis of two subtypes in TCGA cohort. (F) Survival difference between two subtypes in validation set.

Figure 2. Biological function analysis of two subtypes in the CGGA cohort. (A) Heatmap showing the differentially expressed genes (DEGs) between Sub1 and Sub2. FDR < 0.05. (B) Gene Ontology analysis based on the DEGs. (C) GSEA analysis in CGGA cohort. (D–F) Comparison of stromal, immune, purity, cytolytic and inflammation scores presented in boxplots. (G) The expression of human leucocyte antigen (*HLA*) genes. *, $P < 0.05$; **, < 0.01; ***, < 0.001. (H) Relative abundance fractions (%) of immune cell population in each tumor using CIBERSORT tool, lymphocytes = B cells + T cells CD4 + T cells CD8 + T cells follicular helper + Tregs + T cells gamma delta + NK cells +Plasma cells. (*t*-test, Benjamini–Hochberg method). *, FDR < 0.05; **, < 0.01.

Figure 3. Validation set displaying consensus immunological differences between two clusters. (A) Heatmap showing the DEGs between two clusters. FDR < 0.05. (B and C) Gene Ontology and GSEA analysis in TCGA cohort. (D-F) Comparison of stromal, immune, purity, cytolytic and inflammation scores presented in boxplots.

(G) The expression of human leucocyte antigen (*HLA*) genes. *, $P < 0.05$; **, < 0.01; *** (H) Relative abundance fractions (%) of immune cell population in each tumor subgroup using CIBERSORT tool. Elevated M2 macrophages occurred in Sub1 gliomas (*t*-test, Benjamini–Hochberg method). *, FDR < 0.05; **, < 0.01.

Figure 4. Immune cell signatures were associated with outcome for *IDH* wild-type diffuse LGG. (A) Kaplan–Meier analyses of *IDH* wild-type LGG patients stratified by cytolytic and inflammatory scores in CGGA cohort. (B) Kaplan–Meier analyses of *IDH* wild-type LGG patients stratified by cytolytic and inflammatory scores in the TCGA cohort. (C) Heat maps of Spearman correlation coefficients between cytolytic, inflammatory signatures and 22 immune cell signatures in *IDH* wild-type LGG patients of both cohorts. (D) Heat maps of Spearman correlation coefficients between cytolytic, inflammatory signatures and immune inhibitory molecules in both cohorts.

Figure 5. Comparison of genomic alterations between Sub1 and Sub2 in the TCGA cohort. (A) Differential somatic mutation analysis between two subgroups. (B) Distinct CNA profile between Sub1 and Sub2.

Figure 6. Identification of a prognostic signature by Cox proportional hazards model. (A) Venn diagram shows prognosis-related genes which are differentially expressed between two groups. (B) Cross-validation for tuning parameter selection. (C) Heat map of 5 signature genes. (D) Survival analyses of the signature in the TCGA cohort. (E) Distribution of risk scores in cases stratified by subtype, grade and EM/PM classification in TCGA set. (F) Heat map shows the expression of 5 signature genes in CGGA cohort. (G) Survival analyses of the

signature in the CGGA cohort. (H) Distribution of risk scores in cases stratified by subtype and grade in CGGA set. (I) ROC curve analysis of age, subtype and risk score in the TCGA cohort. AUC, area under the curve.

Table 1. Univariate and multivariate Cox regression analysis of clinical pathologic features.

	CGGA cohort						TCGA cohort					
	Univariate analysis			Multivariate analysis			Univariate analysis			Multivariate analysis		
Characteristics	**HR**	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**	**HR**	**95% CI**	***P* value**
Age	1.05	1.008-1.093	**0.018**	1.038	1.001-1.077	**0.043**	1.055	1.024-1.086	**<0.001**	1.051	1.009-1.094	**0.017**
Gender	1.162	0.516-2.617	0.718				1.272	0.621-2.603	0.51			
***MGMT* promoter**	2.634	0.928-7.476	0.069				1.635	0.79-3.385	0.186			
Grade	0.133	0.052-0.345	**<0.001**	0.165	0.062-0.439	**<0.001**	0.258	0.061-1.09	0.065			
Subtype	0.339	0.134-0.855	**0.022**	0.805	0.531-1.2214	0.308	0.47	0.23-0.962	**0.039**	0.846	0.489-1.461	0.548
***TERT* promoter**							1.482	0.568-3.871	0.421			
ATRX							1.028	0.35-3.018	0.96			
Risk score	4.35	1.125-16.819	**0.033**	1.732	0.373-8.053	0.484	16.411	5.111-52.69	**<0.001**	7.396	2.077-26.329	**<0.001**

Gender: male, female; *MGMT* promoter: methylated, unmethylated; *TERT* promoter: mutant, wild-type; Grade: II, III; *ATRX*: mutant, wild-type; Subtype: Sub1, Sub2. Values in bold, statistically significant.