Transcriptomic Analysis Identified two Subtypes of Brain Tumor Characterized by Distinct Immune Infiltration and Prognosis

Xilin Shen
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Hongru Shen
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Mengyao Feng
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Dan Wu
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Yichen Yang
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Yang Li
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Meng Yang
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Wei Ji
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Wei Wang
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Qiang Zhang
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Fangfang Song
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Ben Liu
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Kexin Chen
Tianjin Medical University Cancer Institute and Hospital: Tianjin Tumor Hospital
Xiangchun Li (lixiangchun2014@foxmail.com)
Tianjin Medical University Cancer Institute and Hospital https://orcid.org/0000-0002-2230-2152

Research

Keywords: Brain tumor, molecular subtype, immune infiltration, prognosticator, transcriptome

Posted Date: June 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-630718/v1

Abstract

Background Brain tumor ranks the most devastating cancer type. The complex tumor immune microenvironment prevents brain tumor from therapeutic benefits. The purpose of this study was to stratify brain tumors based on their distinct immune infiltration signatures to facilitate better clinical decision making and prognosis prediction.

Methods We developed a deep learning model to characterize immune infiltration from transcriptome. The developed model was applied to distill expression signatures of transcriptome of brain tumor samples. We performed molecular subtyping with the extracted expression signatures to unveil brain tumor subtypes. Computational methods including gene set enrichment analysis, Kaplan-Meier survival and multivariate Cox regression analyses were employed.

Results We identified two distinctive subtypes (i.e. C1/2) of brain tumor featured by distinct immune infiltration signatures. The C1 subtype is characterized by protective immune infiltration signatures, including high infiltration of CD8+ T cells and activation of *CX3CL1*. The C2 subtype has an extensive infiltration of tumor-associated macrophages and microglia, and was enriched with immune suppressive, wound healing and angiogenic signatures. The C1 subtype had significantly better prognosis as compared with C2 (Log-rank test, HR: 2.5, 95% CI: 2.2 – 2.7; P = 8.2e-78). This difference remained statistically significant (multivariate Cox model, HR: 2.2, 95% CI: 1.7 – 2.9; P = 3.7e-10) by taking into account age, gender, recurrent/secondary status at sampling time, tumor grade, histology, radio-chemotherapy, *IDH* mutation, *MGMT* methylation and co-deletion of 1p and 19q. This finding was validated in 6 datasets. The C2 subtype of glioblastoma patients with *IDH* mutation has poor survival analogous to those without *IDH* mutation (Log-rank test, adjusted P = 0.8), while C1 has favorable prognosis as compared with glioblastoma of C2 subtype with *IDH* mutation (Log-rank test, adjusted P = 1.2e-3) or without *IDH* mutation (Log-rank test, adjusted P = 1.3e-6).

Conclusions We identified two distinctive subtypes of brain tumor with different immune infiltration signatures and prognosis. Our finding is helpful for better understanding of brain tumor and has potential clinical utilities.

Background

Brain tumors are highly aggressive and rank among the most fatal and devastating diseases (1). The standard treatments for brain tumors include chemotherapy and radiotherapy in addition to surgical removal (2). However, the efficacy is varying considerably with some patients exhibited rapid resistance while others response durably (3). In addition, a range of postoperative complications including seizure, loss of movement ability, visual impairment, impairment of speech and comprehension occur after surgery.

Better understanding of key genomic alteration in brain tumor leads to effective treatment options for patients. For example, lack of *MGMT* methylation is associated with reduced benefit from temozolomide

(4). *IDH1*/*IDH2* mutation and co-deletion of chromosome arms 1p and 19q are associated with radio-chemotherapy response and survival outcome (5). Meanwhile, high-throughput analyses of genomic and transcriptomic data have led to a refined classification system of brain tumor to promote effective clinical therapeutics. The WHO classification of central nervous system (CNS) introduced in 2016 defined tumor entities based on molecular characteristics in addition to traditional morphologic findings (3). Nevertheless, clinical heterogeneity remains an intractable issue. For instance, patients of astrocytoma without *IDH* mutation have diverse clinical outcomes (6).

Brain tumor microenvironment is immunologically distinct from other cancer types (7). Tumorigenesis can cause damage to blood-brain barrier, facilitating infiltration of immune cells from peripheral circulation into brain (7). Compromised blood-brain barrier can activate wound healing and angiogenesis, which promotes cancer progression and confers immune suppression (8, 9). Meanwhile, brain tumor is immunologically “cold” in that tumor-associated macrophages and microglia (TAMs) prevent tumor from immune response (7).

In this study, we established a model to distill expression signatures from the transcriptome of brain tumor tissues. We revealed two subtypes of brain tumor with distinct immune infiltration signatures, genomic alteration and prognosis. Our findings were validated in 11 previous datasets.

Methods

Data collection

We collected 3810 transcriptomes of brain tumor samples from 11 datasets. Each dataset has more than 100 patients accompanied vital status and period of follow-up. Histology, tumor grade, radio-chemotherapy treatment, recurrent/secondary status at sampling time, *IDH* mutation, *MGMT* methylation and co-deletion of 1p and 19q data were collected if available (**Supplementary Table 1**). Meanwhile, we collected 93,293 single-cell RNA profiles subjected to Smart-Seq2 sequencing protocol from 16 previous studies (**Supplementary Table 2**) from brain cancer, lung cancer, colorectal cancer, ovarian cancer, melanoma and head and neck squamous cell carcinoma. These single cell datasets encompass T cells, B cells, monocytes, macrophages, natural killer cells, dendritic cells, cancer cells, and other nonmalignant cells including fibroblast, epithelia cells, gliocyte and neurons. In addition, we manually curated a list of genes related to tumor microenvironment, immune cells, immune checkpoint blockade therapy response and prognosis (**Supplementary Table 3**).

Data preprocessing

For each single-cell dataset, we performed logarithmic transformation as $\log_2(\text{TPM}/10 + 1)$. We clipped gene expression values at 99% quantile values of all genes. Subsequently, we employed R package *preprocessCore* (version 1.40.0) (10) to perform quantile normalization. We applied ComBat routine implemented in R package *sva* (version 3.26.0) (11) to perform batch effect correction for the normalized expression data of bulk brain tumors.

Feature representation learning of single-cell transcriptome

We employed a neural network as feature encoder to learn the nonlinear feature representations of transcriptomes in a reduced dimensional space. A self-supervised deep learning algorithm (12) was adopted to train the feature encoder through contrastive learning. Contrastive learning allows feature encoder to learn representations in a label-free manner through contrast samples (**Supplementary Figure 1**). Specifically, two different views of the same sample form a positive pair, and two different samples form a negative pair. Contrastive learning set labels to 1 for positive pairs and 0 for negative pairs. The feature encoder was driven by contrastive loss to minimize the distance between positive pair and maximize the distance between negative pair. In our task, the feature encoder was trained to learn the same representation of the different noise-adding views of the same single-cell transcriptome and dissimilar representation of different cells. We made noise through random zero out, shuffling and add Gaussian noise (mean: 0, standard deviation: 0.1) to 20% of genes for all transcriptomes at each epoch. We logarithmically transformed the preprocessed transcriptomes of the preselected genes and scaled to a range of 0 to 1 before fed them into feature encoder. The feature encoder was an 18 layers deep residual networks (13) with a project head of 128 output neurons. The architecture of the feature encoder was provided as **Supplementary Figure 2**. We employed stochastic gradient descent algorithm as optimizer. The weight decay of the optimizer is 1e-4 and the momentum is 0.9. We set the contrastive learning queue size of 3072, the momentum of 0.999, and the softmax temperature of 0.2. The model was trained in parallel on two graphic processing units with the batch size of 256 and an initial learning rate of 0.03. We trained for 300 epochs with the learning rate decay with a cosine annealing schedule. The model was developed with *PyTorch* (v1.3.0) package.

Molecular subtyping of brain tumor

The developed feature encoder was applied to extract expression signatures from bulk sample transcriptomes. Specifically, we extracted feature representations from the developed feature encoder applied to the expression data of TCGA pan-cancer. The feature encoder transformed the expression profile of each bulk sample into 128 features, which was determined by the output neurons of feature encoder. The extracted features were hierarchically clustered through R package *ConsensusClusterPlus* (version 1.42.0) (14). The obtained clusters were further grouped into expression signatures because of the high negative correlations among these clusters (**Supplementary Figure 3**). Subsequently, we dichotomized brain tumor patients based on each expression signature and selected one that can best represent unique immune infiltration signature of brain tumor. Specifically, we used *R* package *Ckmeans.1d.dp* (version 4.2.1) (15) to perform k-means clustering. The k-means clustering cutoff value closest to the median value of signature was select as the optimal cutoff to dichotomize samples. We used R package *fgsea* (version 1.6.1) (16) to performed gene set enrichment analysis (GSEA) for a gene set related to unique immune infiltration properties of brain tumor such as microglia and reactive gliosis (**Supplementary Table 4**). We kept the signature that ranked on the top to dichotomize patients as mentioned above for downstream analysis. A flowchart illustrating this procedure was provided in **Supplementary Figure 3**.

Linear feature encoder comparison

To examine the advantage of self-supervised learning paradigm, we employed principle component analysis (PCA) as a linear feature encoder and compare the PCA features with the deep learning features. Specifically, we performed PCA on single cell transcriptomes of the 2616 filtered genes with python package *sklearn* (v0.24.1). The principle components of single- cells were then projected to brain tumor transcriptomes. Then, brain tumor patients were dichotomized through hierarchical clustering based on R package *ConsensusClusterPlus* (version 1.42.0) (14).

Association between molecular subtypes and clinical data

We analyzed the association between molecular subtypes with immune and genomic alteration signatures, which include immune cellular fractions, immunomodulatory expressions, oncogenic and immune pathways, genomic alterations, driver mutations and molecular subtypes of glioblastoma proposed by the Cancer Genome Atlas (TCGA) (17). We used CIBERSORT to estimate the proportions of 22 immune cell types based on LM22 matrix (18). We performed paired t-test for 78 genes related to immunomodulation (19) in the 11 collected datasets. In addition, we performed GSEA based on R pakage *fgsea* (version 1.6.1) (16) for cancer hallmark (20) and immune-related gene sets (19) (**Supplementary Table 4**). Continuous variables were evaluated by Wilcoxon rank sum test, while discrete variables were evaluated via Chi-square test if not specified. For GSEA, *P*-values were calculated based on 10,000 permutations. Kaplan-Meier survival analysis and multivariate Cox hazards model were utilized to analyze the association of subtypes and prognosis, which were carried out with R *survival* package (2.40-3). Survival curve and forest plot were visualized with R packages *survminer* (0.4.6) and *forestmodel* (0.5.0), respectively. A *P*-value < 0.05 or adjusted *P*-value < 0.05 was considered to be statistically significant.

Results

Patients and analytic pipeline

RNA profiles and clinical data of 3810 brain tumor patients were collected from 11 public studies. The baseline characteristics are shown in **Supplementary Table 5**. Glioma and medulloblastoma account for 83% (3146) and 16% (624), respectively. Among 2951 patients with cancer type information, the primary, recurrent, secondary and post-treatment tumor account for 71% (2105), 11% (332), 1% (40) and 16% (474), respectively. In the glioma cohort (3146), patients with tumor grade II, III, IV respectively accounted for 24% (743), 28% (872) and 39% (1226); 10% (305) of patients did not have tumor grade information. Meanwhile, there were 2960 glioma patients have pathological information. This glioma cohort consisted of diverse pathological subtypes such as astrocytoma (27%), oligodendroglioma (21%), oligoastrocytoma (8%) and glioblastoma (44%). Among 2289 of glioma patients with *IDH* mutation examined, 55% of them (1254) have *IDH* mutation. Among 1098 of these 1254 patients with co-deletion of 1p/19q tested, 42% (459) carried co-deletion of 1p and 19q. Among 1731 patients tested for *MGMT* methylation, 59% (1028) of them were positive for hypermethylation of *MGMT* promoter. Among 1226 patients with radio-

chemotherapy treatment information, the proportion of patients treated with chemotherapy, radiotherapy and a combination of both were 15% (184), 23% (281) and 62% (761), respectively.

A flowchart depicting the whole procedures of this study was shown in **Figure 1**. We collected 93,293 single-cell RNA profiles from 16 published datasets and manually curated 2,616 genes that were associated with tumor microenvironment, immune cells, immune checkpoint blockade therapy response and prognosis. We developed a self-supervised deep learning model on single-cell RNA profiles of these 2616 genes to decipher gene expression signatures from transcriptomes. Subsequently, we applied this developed feature encoder to extract expression signature from transcriptome of bulk brain tumor samples (See **Methods** and **Supplementary Figure 3**). We then examined the association of expression signature with immune signatures, genomic alteration and prognosis.

Differences of immune infiltration signatures in C1 versus C2 subtype

The results obtained from CIBERSORT (18) showed that 18 of 22 types of immune cells were significantly different between C1/2 subtypes (**Figure 2A**). All types of TAMs (i.e. M0, M1, M2), CD4+ follicular helper T cells and neutrophils had higher infiltration rate in C2 as compared with C1 subtype (**Figure 2A**). Contrastively, C1 had higher infiltration of CD8+ T cells, plasma cells and dendritic cells than C2 subtype (**Figure 2A**). The infiltration of the other cell types was provided in **Supplementary Table 6**.

We observed that 24 immunomodulatory genes were differentially expressed in C1 versus C2 subtype (**Figure 2B**). Specifically, *C10orf54*, *CX3CL1* and *EDNRB* were highly expressed in C1 versus C2 subtype (**Figure 2B**). *CD276*, *CCL5*, *CXCL10*, *HMGB1* and the other 17 immunomodulatory genes were significantly upregulated in C2 versus C1 subtype (**Figure 2B**). The detailed expression of all immunomodulatory genes were provided in **Supplementary Table 7**. Enrichment analysis of 50 cancer hallmarks and 132 immune signaling modules showed that CSF-1, MYC, TGF-β, JAK/STAT3, IFN-α and the other 28 signaling pathways were enriched in C2 versus C1 subtype (**Supplementary Table 4**).

C1/2 subtypes were significantly associated with genomic alterations

In the TCGA low-grade glioma, non-silent mutation burden, intratumor heterogeneity, aneuploidy and the other 6 types of genomic variation were significantly higher in C2 versus C1 subtype (**Figure 2C** and **Supplementary Table 8**). In TCGA glioblastoma cohort, there was no difference among the aforementioned variations except for segments of copy number variation (**Supplementary Figure 4**).

We also examined the association of C1/2 subtypes and driver gene mutations of brain tumors that linked to prognosis and therapeutic resistance (**Supplementary Table 6**). Our finding showed that 4 driver events were significantly higher in C1 versus C2 subtype, including *IDH* mutation, hypermethylation of *MGMT* promoter, high CpG island methylation phenotype (G-CIMP) and co-deletion of 1p and 19q (**Figure 2D**). Four driver events were significantly higher in C2 versus C1 subtype such as *EGFR* amplification, deletion of *CDKN2A*/*CDKN2B* and *PTEN*, gain of chromosome 7 and/or loss of chromosome 10 (**Figure 2D**).

In addition, we found that C1/2 subtypes were linked to TCGA molecular subtypes, namely classical, neural, proneural and mesenchymal subtypes (17) (**Figure 2E**). Neural (168(37%) versus 104(11%); *P*-value = 6.2e-29) and proneural subtypes (186(41%) versus 243(26%); *P*-value = 6.2e-29) were significantly enriched in C1 versus C2 subtype. C2 has higher proportions of classical (307(33%) versus 56(12%); *P*-value = 1.4e-16) and mesenchymal subtypes (265(29%) versus 44(10%); *P*-value = 2.3e-15) as compared with C1 subtype.

C1/2 subtypes were significantly associated with clinical characteristics

Clinical characteristics of brain tumor patients were provided in **Supplementary Table 9**. C2 subtype had lower Karnofsky scores (Median: 80 vs. 90, Wilcoxon rank sum test, *P*-value = 3.4e-6) and higher tumor microvascular infiltration rate versus C1 subtype (61/76, 80% vs. 31/65, 48%; OR: 4.2, 95% CI: 2.0 – 8.7; Chi-squared test, *P*-value = 1.8e-4). Among patients with recurrence, C1 subtype has marginally significant lower distant recurrence rate (4/23, 17% vs. 19/48, 40%; OR: 0.3, 95% CI: 0.1 – 1.1) and higher local recurrence rate (19/23, 83% vs. 29/48, 60%; OR: 3.1, 95% CI: 0.9 – 10.6) as compared with C2 subtype (Chi-squared test, *P*-value = 0.1). There were no significant differences in family history of cancer, prediagnostic symptoms and tumor location between C1/2 subtypes (Chi-squared test, all *P*-values > 0.5).

Kaplan-Meier survival analysis showed that C1 subtype has better survival than C2 subtype (**Figure 3A**; Log-rank test, *P*-value = 8.2e-78) in the combined cohort of 3810 patients. This result was also observed in each individual of the 11 datasets (**Figure 3A**; Log-rank test, all *P*-values < 0.05). Moreover, the difference remained significant in the combined cohort after controlling for confounding factors such as age, gender, tumor, histology, radio-chemotherapy, recurrent/secondary status, *IDH* mutation status, *MGMT* methylation status and co-deletion of 1p and 19q (**Figure 3B** and **Supplementary Figure 6**; Multivariate Cox model, HR: 2.2, 95% CI: 1.7 – 2.9; *P* = 3.7e-10). The independent association of C1/2 subtypes with prognosis from multivariate model remained significant in 6 individual datasets and exhibited the same trend in the other 4 datasets (**Figure 3B** and **Supplementary Figure 6**). In TCGA glioma cohort, surgery type was taking into consideration additively. In the medulloblastoma cohort (i.e. GSE85217), clinically relevant confounding factors such as age, gender and molecular subtypes were included. In addition, we observed that the association between prognosis and expression signatures derived from deep learning is more generalizable as compared with PCA (**Supplementary Table 10**).

We also examined the association between C1/2 subtypes and prognosis of glioma patients with respect to histology, genomic alteration and grade. The glioma patients were divided into 9 subgroups such as astrocytoma, oligodendroglioma, glioma with or without *IDH* mutation, glioma with *IDH* mutation with or without co-deletion of 1p and 19q, tumor grade II, III and IV (**Figure 4A**). The C2 subtype has significantly poor survival outcome than C1 in all subgroups (**Figure 4A**; Log-rank test, *P*-values < 0.05). In addition, the difference remained significant in 8 out of these 9 subgroups and marginal significant in grade IV glioma after taking into account age, gender, histology, *IDH* mutation status, *MGMT* methylation status and co-deletion of 1p and 19q (**Figure 4B** and **Supplementary Figure 6**). The dataset was taken as strata variable

in multivariate Cox model. The C2 subtype of glioblastoma with *IDH* mutation has poor survival outcome analogous to glioblastoma without *IDH* mutation (**Figure 5A;** Log-rank test, adjusted *P*-value = 0.8). While the C1 subtype of glioblastoma with *IDH* mutation has favorable survival outcome versus C2 subtype (Log-rank test, adjusted *P*-value = 1.2e-3) or glioblastoma without *IDH* mutation (Log-rank test, adjusted *P*-value = 1.3e-6). The result remained significant after ruling out confounding impact of age, gender and co-deletion of 1p and 19q (**Figure 5B**).

Kaplan-Meier survival analysis showed that C2 subtype had worse progression-free survival as compared with C1 subtype in TCGA glioma cohort (**Supplementary Figure 5,** Log-rank test, adjusted *P*-value = 6.1e-4). The difference remained significant in radio-chemotherapy patients (**Supplementary Figure 5,** Log-rank test, adjusted *P*-value = 5.3e-3) and show the same trend in radiotherapy along patients (**Supplementary Figure 5,** Log-rank test, adjusted *P*-value = 0.4). Progression-free survival was not analyzed for the chemotherapy group due to the limited sample size (**Supplementary Table 11**).

Discussion

Despite substantial efforts to characterize molecular signatures of brain tumor, there remain a gap between immune heterogeneity and clinical behaviors (8, 21, 22). This study represents an attempt to identified new subtypes of brain tumor based on distinct immune infiltration signature, to spanning the previous classification systems that are mainly defined on histology and genome (3). The importance of identifying the C1/2 subtypes lies in their markedly different survival outcomes due to their distinct immune infiltration. Our findings will facilitate the elucidation of distinct immune infiltration in the development and prognosis prediction of brain tumors.

The feature encoder derived from self-supervised learning is akin to PCA in that both of them can extract representation features in a label-free manner. However, the deep neural network is able to capture the non-linear feature in contrastive to linear feature reduction of PCA. The non-linear feature modeling capability of deep neural network may be better in distilling immune infiltration signatures and provide new insights as compared with PCA. This was demonstrated by the identification of C1/2 subtypes that can better serve as an independent prognosticator compared with subtypes obtained from PCA. The C1/2 subtypes can distinguish glioma patients with different prognosis stratified by histology, tumor grade and genomic alteration. It is generally accepted that glioblastoma with *IDH* mutation has better prognosis than those without (23). The C2 subtype of glioblastoma with *IDH* mutation has poor survival outcomes analogous to those without *IDH* mutation, whereas the C1 subtype of glioblastoma with *IDH* mutation has significantly better survival outcomes (Fig. 5). Thus, the C1/2 subtypes proposed in our study may improve the current glioblastoma classification system based on *IDH* mutation status to more accurately reflect prognostic discrepancy among glioblastoma patients. Besides, the C1/2 subtypes can also reflect differences in microvascular infiltration, distant metastasis and radio-chemotherapy response of patients.

The intrinsic distinctiveness in immunity may explain the different prognosis of C1/2 subtypes. The C1 subtype enriched for a constellation of protective markers for prognosis such as high infiltration of CD8 +

T cells, plasma cells, dendritic cells and activation of *CX3CL1*. CD8 + T cells are main force in maintaining anti-tumor immune responses (24). CX3CL1 can inhibit the migration of tumor cells (25). Protective genomic alteration events including *IDH* mutations and CpG island methylation were also occur frequently in C1 subtype. The *IDH* mutation causes aberrant methylation of DNA and histone (26) to force the appearance of CpG island methylation phenotype in glioma, both of which are favorable prognosticators in brain tumor (27).

The C2 subtype was characterized by enrichment of immune infiltration signatures. A striking characteristic of C2 subtype is the extensive infiltration of TAMs. TAM functions in immunosuppression to promote the development of a “cold” microenvironment for brain tumor (28). TAMs recruitment signatures such as upregulated CSF-1 response circuits and highly expressed macrophage chemokine genes including *CCL5* and *VEGF* (28) were also enriched in C2 subtype. Besides, abundant signatures related to immune suppression, wound healing and angiogenesis were detected in C2 subtype. For example, C2 subtype was poorly infiltrated with CD8 + T cells and enriched for immune suppression genes such as *IL-10*, *TGF-β*, *HAVCR2* and *ENTPD1* (28, 29). Wound healing programs including core serum response and JAK/STAT3 circuit were overrepresented in C2. Angiogenic signatures including high expressions of *VEGFA*, *CD276* and amplification of *EGFR* (30–32) were abundant in C2 subtype. Aberrant tumor angiogenesis contribute to immunosuppression and tumorigenesis through subvert effector CD8 + T cells and promote regular CD4 + T cells infiltration (33). CD276 is a putative target for CAR T-cell therapy of pediatric glioma (31). Apart from upregulation of E2F, MYC and G2M circuits, C2 subtype was enriched for deletion of *CDKN2A*/*CDKN2B* and *PTEN*. C2 subtype was highly infiltrated by neutrophil. Neutrophil has been reported to be associated with acquired resistance to radio-chemotherapy in brain tumor (34).

Our analysis has several limitations. First, the limited availability of clinical information restricts the association analysis between C1/2 subtype with therapy response. The association between therapy outcome and progression-free survival can only be explored in TCGA glioma cohort. We cannot examine connection between C1/2 subtypes and chemotherapy due to the limited sample size ($N = 9$). A differential trend in progression-free survival of radiation-alone patients ($N = 21$) was observed between C1/2 subtypes, and further study should include more patients to demonstrate this difference. Second, the immune infiltration differences between C1/2 subtypes and their relationship with prognosis are still preliminary. The detail mechanisms still unclear and require further biomedical experiments.

Conclusions

We revealed two molecular subtypes (i.e. C1/2) of brain tumor featured by distinct immune infiltration signatures and prognosis. Our finding is helpful for better understanding of brain tumor and has potential clinical utilities.

Abbreviations

CNS: Central nervous system; TAMs: Tumor-associated macrophages and microglia; GSEA: Gene set enrichment analysis; PCA: Principle component analysis; TCGA: The Cancer Genome Atlas.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

We presented an R package *Brammer* (https://github.com/xilinshen/brammer) that can identify C1/2 subtypes based on expression matrix. The datasets generated and/or analysed during the current study are available in the GlioVis (35), the Chinese Glioma Genome Atlas (36), the Gene Expression Omnibus (37) (accession number GSE16011, GSE13041, GSE108474, GSE85217, GSE70630, GSE89567, GSE102130, GSE103322, GSE108989, GSE98638, GSE146771, GSE99254, GSE115978, GSE72056, GSE94820, GSE120575, GSE89232, GSE70580, GSE146026, GSE136732), the European Bioinformatics Institution (38) (accession number E-MTAB-3892), the International Cancer Genome Consortium (39) and the Genomic Data Commons (40) data portal.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China [31801117]; the Program for Changjiang Scholars and Innovative Research Team in University in China [IRT_14R40]; the Tianjin Science and Technology Committee Foundation [17JCYBJC25300]; and the Chinese National Key Research and Development Project [2018YFC1315600]; Tianjin Municipal Health Commission Foundation [RC20027].

Authors' contributions

LXC and CKX designed and supervised the study. LXC and SXL wrote the manuscript. LXC, CKX and SXL revised the manuscript. LXC and SXL analyzed the data. SXL, SHR, FMY, WD, YYC, LY and YM collected data. LXC, CKX, SXL, JW, WW, ZQ, SFF and LB interpreted the results. All authors reviewed and approved the submission of this manuscript.

Acknowledgements

We want to thank all the researchers for their generosity to make their data publicly available.

Authors' information

[1]Tianjin Cancer Institute, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China; [2]Department of Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China; [3]Department of Maxillofacial and Otorhinolaryngology Oncology, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2020. CA: a cancer journal for clinicians. 2020;70(1):7–30.
2. Nabors LB, Portnow J, Ahluwalia M, Baehring J, Brem H, Brem S, et al. Central Nervous System Cancers, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. D – 101162515. 2020;18(11):1537–70.
3. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016;131(6):803–20.
4. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. D – 0255562. 2005;352(10):997–1003.
5. Cancer Genome Atlas Research N. Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. D – 0255562. 2015;372(26):2481–98.
6. Brat DJ, Aldape K, Colman H, Holland EC, Louis DN, Jenkins RB, 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(5):805–10.
7. Sampson JH, Gunn MD, Fecci PE, Ashley DM. Brain immunology and immunotherapy in brain tumours. Nature reviews Cancer. 2020;20(1):12–25.
8. Quail DF, Joyce JA. The Microenvironmental Landscape of Brain Tumors. D – 101130617. 2017;31(3):326–41.
9. Jain RK, di Tomaso E, Duda DG, Loeffler JS, Sorensen AG, Batchelor TT. Angiogenesis in brain tumours. Nat Rev Neurosci. 2007;8(8):610–22.
10. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19(2):185–93.

11. Parker HS, Leek JT, Favorov AV, Considine M, Xia X, Chavan S, et al. Preserving biological heterogeneity with a permuted surrogate variable analysis for genomics batch correction. Bioinformatics. 2014;30(19):2757–63.
12. Chen X, Fan H, Girshick R, He K. Improved Baselines with Momentum Contrastive Learning. arXiv. 2020.
13. He K, Zhang X, Ren S, Sun J, editors. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016.
14. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26(12):1572–3.
15. Wang H, Song M. Ckmeans.1d.dp: Optimal k-means Clustering in One Dimension by Dynamic Programming. R J. 2011;3(2):29–33.
16. Korotkevich G, Sukhov V, Sergushichev A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv. 2019.
17. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. D – 101130617. 2010;17(1):98–110.
18. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature methods. 2015;12(5):453–7.
19. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity. 2018;48(4):812–30. e14.
20. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–40.
21. White K, Connor K, Clerkin J, Murphy BM, Salvucci M, O'Farrell AC, et al. New hints towards a precision medicine strategy for IDH wild-type glioblastoma. Ann Oncol. 2020;31(12):1679–92.
22. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. D – 101130617. 2017;32(1):42–56. e6.
23. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. D – 0404511. 2008;321(5897):1807–12.
24. Broekman ML, Maas SLN, Abels ER, Mempel TR, Krichevsky AM, Breakefield XO. Multidimensional communication in the microenvirons of glioblastoma. Nat Rev Neurol. 2018;14(8):482–95.
25. Sciume G, Soriani A, Piccoli M, Frati L, Santoni A, Bernardini G. CX3CR1/CX3CL1 axis negatively controls glioma cell invasion and is modulated by transforming growth factor-beta1. Neuro Oncol. 2010;12(7):701–10.
26. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and molecular epidemiology of adult diffuse glioma. Nat Rev Neurol. 2019;15(7):405–17.

27. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. D – 101130617. 2010;17(5):510–22.
28. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. Nat Rev Clin Oncol. 2017;14(7):399–416.
29. Hambardzumyan D, Gutmann DH, Kettenmann H. The role of microglia and macrophages in glioma maintenance and progression. Nat Neurosci. 2016;19(1):20–7.
30. Claesson-Welsh L, Welsh M. VEGFA and tumour angiogenesis. J Intern Med. 2013;273(2):114–27.
31. Majzner RG, Theruvath JL, Nellan A, Heitzeneder S, Cui Y, Mount CW, et al. CAR T Cells Targeting B7-H3, a Pan-Cancer Antigen, Demonstrate Potent Preclinical Activity Against Pediatric Solid Tumors and Brain Tumors. Clin Cancer Res. 2019;25(8):2560–74.
32. Bonomi PD, Gandara D, Hirsch FR, Kerr KM, Obasaju C, Paz-Ares L, et al. Predictive biomarkers for response to EGFR-directed monoclonal antibodies for advanced squamous cell lung cancer. Ann Oncol. 2018;29(8):1701–9.
33. Schaaf MB, Garg AD, Agostinis P. Defining the role of the tumor vasculature in antitumor immunity and immunotherapy. Cell Death Dis. 2018;9(2):115.
34. Fossati G, Ricevuti G, Edwards SW, Walker C, Dalton A, Rossi ML. Neutrophil infiltration into human gliomas. Acta Neuropathol. 1999;98(4):349–54.
35. Bowman RL, Wang Q, Carro A, Verhaak RG, Squatrito M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. Neuro Oncol. 2017;19(1):139–41.
36. The Chinese Glioma Genome Atlas. [Available from: http://www.cgga.org.cn/.
37. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):D991-5.
38. The European Bioinformatics Institute. [Available from: https://www.ebi.ac.uk/.
39. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The International Cancer Genome Consortium Data Portal Nat Biotechnol. 2019;37(4):367–9.
40. Genomic Data Commons. [Available from: https://portal.gdc.cancer.gov/.
41. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550–8.
42. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. Nat Med. 2018;24(10):1545–9.

Figures

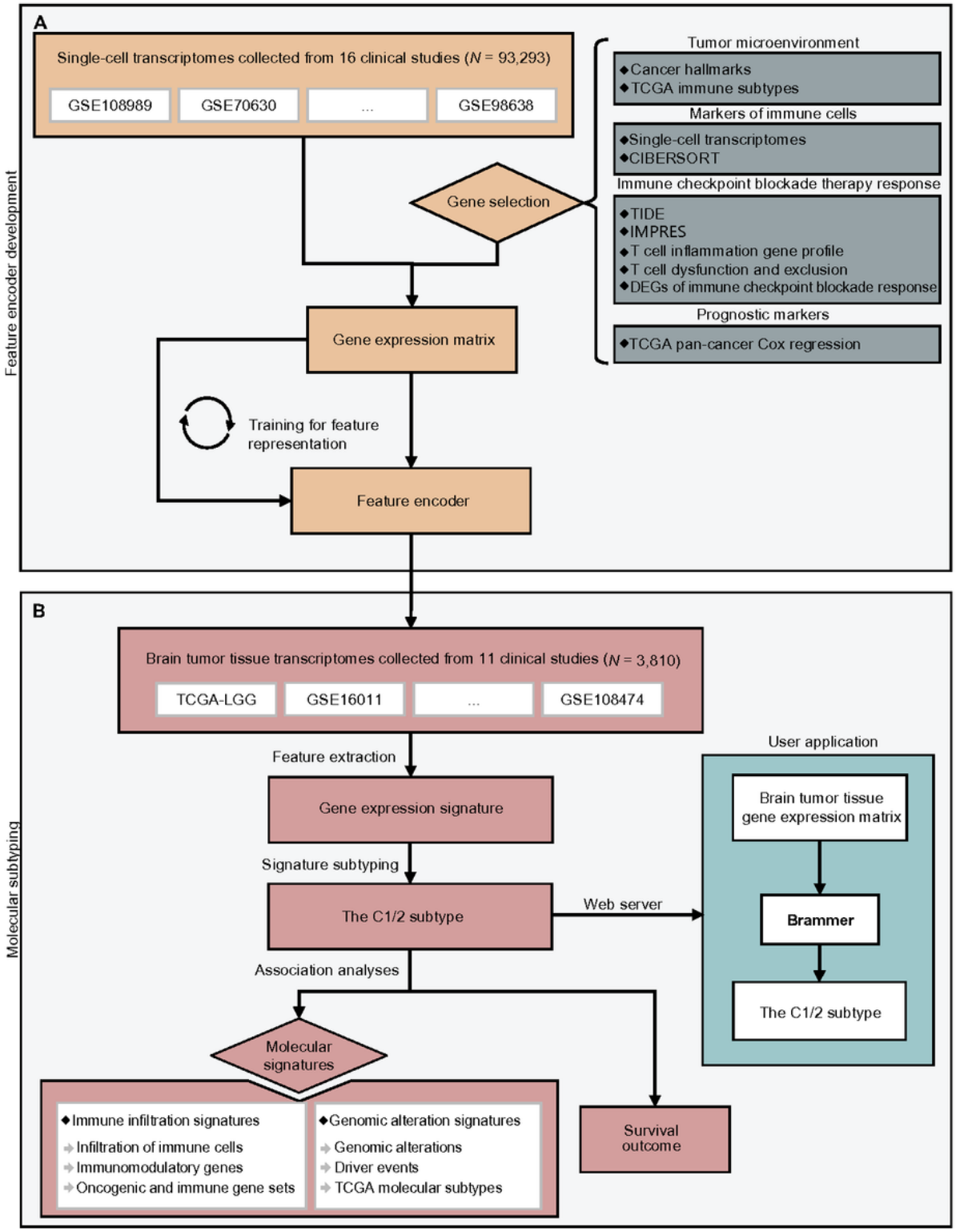

Figure 1

A flowchart depicting the whole procedures conducted in this study. The upper panel (A) describes steps involved in the development of a deep learning model to learn feature representation from single-cell transcriptomes. The lower panel (B) depicts molecular subtyping of brain tumors and downstream analysis tasks. CIBERSORT, TIDE, IMPRES were referenced from (18, 41, 42). DEG, differentially expressed gene; TCGA, the Cancer Genome Atlas.

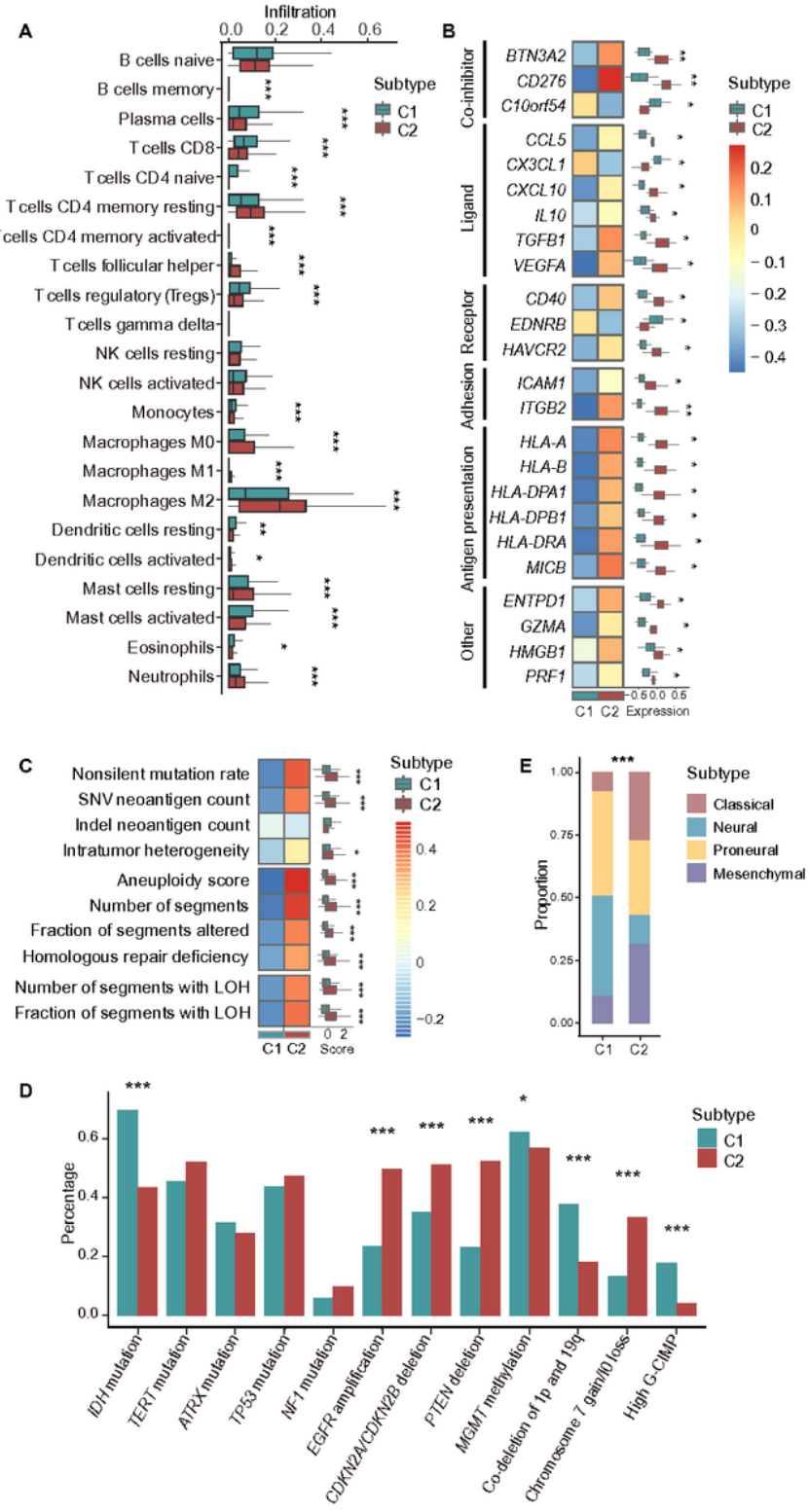

Figure 2

Association between C1/2 subtypes with genomic and transcriptomic signatures. (A) The proportion of infiltrated immune cell types in C1 versus C2 subtype. (B) The median expression levels of immunomodulatory genes across 11 brain tumor datasets in C1 versus C2 subtype. (C) Genomic alteration signatures in C1 versus C2 subtype in TCGA low-grade glioma cohort. (D) Alteration prevalence of driver events in C1 versus C2 subtype. (E) The proportion of TCGA molecular subtypes in C1 versus C2

subtype. P values were subjected to multiple hypothesis correction. *P < 0.05, **P < 0.01, ***P < 0.001. LOH, loss of heterozygosity; G-CIMP, CpG island methylation phenotype.

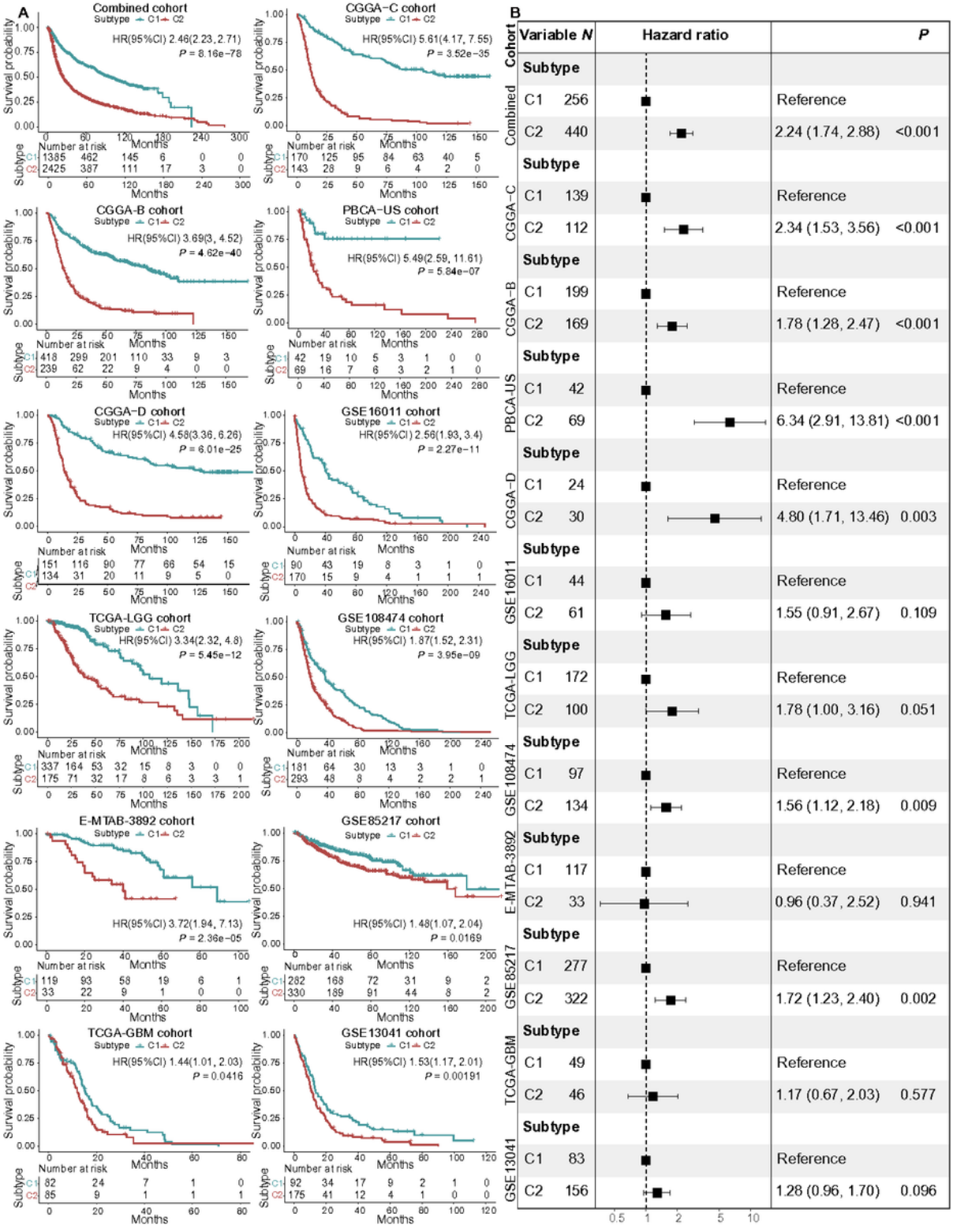

Figure 3

Prognostic significance of C1/2 subtypes in the combined and 11 individual brain tumor cohorts. (A) Kaplan-Meier survival analysis of C1 versus C2 subtype in the combined and each individual brain tumor cohorts. (B) Combined forest plot depicting multivariate Cox regression analysis of C1/2 subtypes by

taking into account age, gender, tumor grade, histology, radio-chemotherapy status, IDH mutation, MGMT methylation and co-deletion of 1p and 19q. HR, hazard ratio; CI, confidence interval.

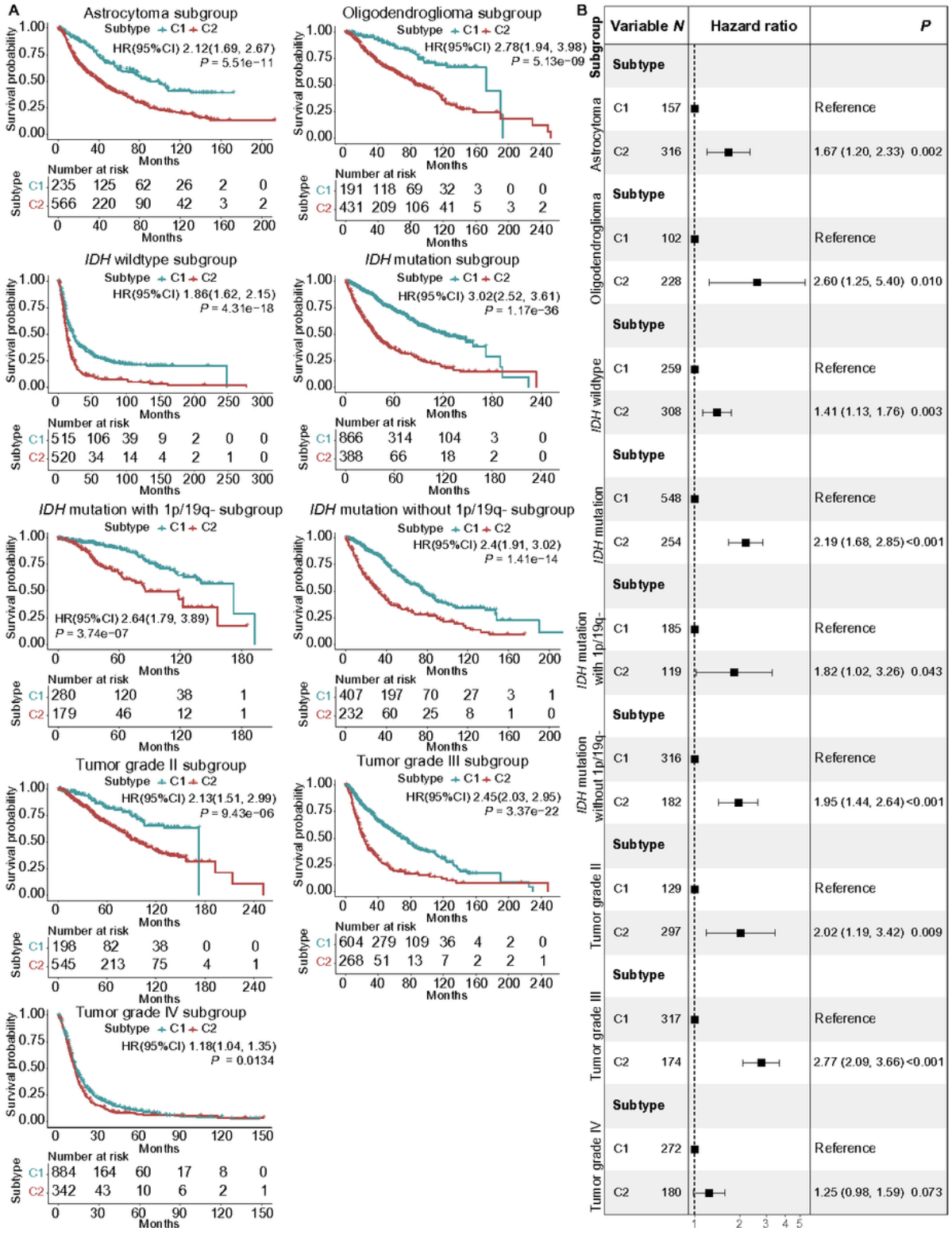

Figure 4

Prognostic significance of C1/2 subtypes stratified by different clinical variables. (A) Kaplan-Meier survival analysis of C1 versus C2 subtype. (B) Combined forest plot portraying multivariate Cox regression analysis of C1/2 subtypes after controlling age, gender, histology, IDH mutation, MGMT

methylation and co-deletion of 1p and 19q. 1p/19q-, co-deletion of 1p and 19q; HR, hazard ratio; CI, confidence interval.

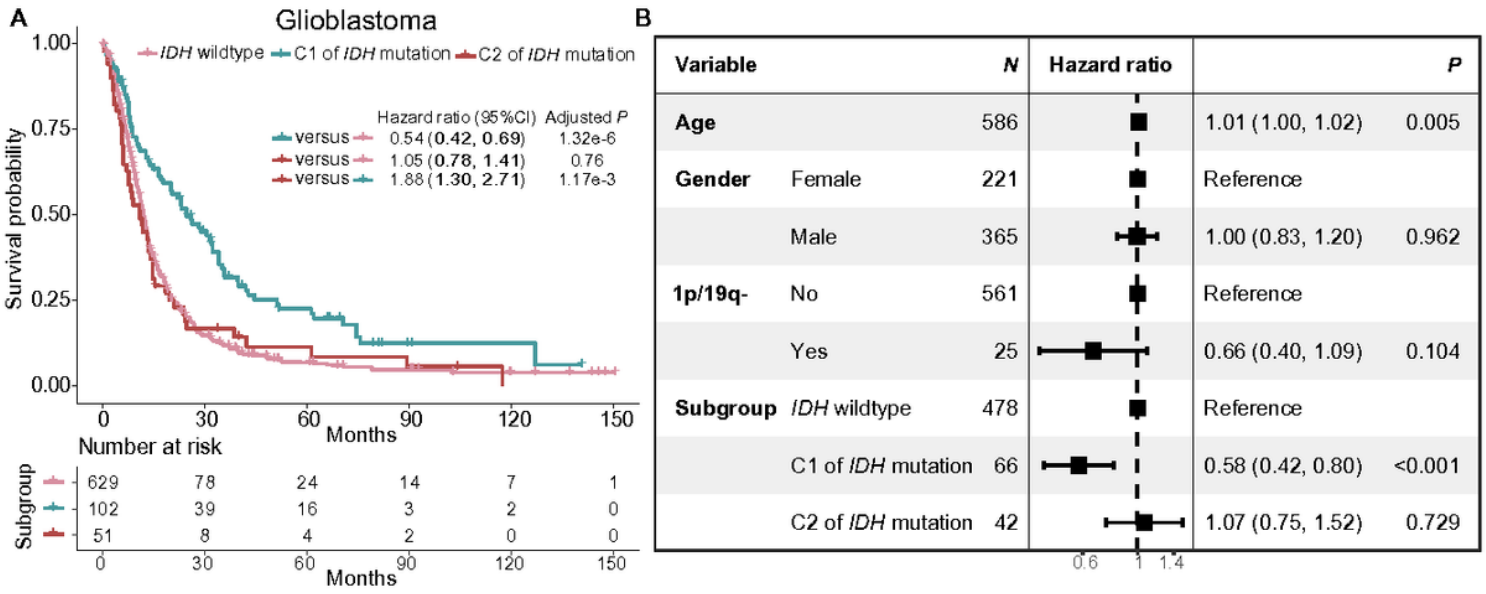

Figure 5

Prognostic significance of IDH mutation plus C1/2 subtype in glioblastoma patients. (A) Kaplan-Meier survival analysis of glioblastoma in patients without IDH mutation, C1 subtype with IDH mutation and C2 subtype with IDH mutation. (B) Multivariate Cox regression analysis of C1/2 subtypes by ruling out confounding impacts such as age, gender and co-deletion of 1p and 19q. 1p/19q-, co-deletion of 1p and 19q; CI, confidence interval.

Supplementary Files

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

- SupplementaryFigure1.pdf
- SupplementaryFigure2.png
- SupplementaryFigure3.pdf
- SupplementaryFigure4.pdf
- SupplementaryFigure5.pdf
- SupplementaryFigure6.pdf
- SupplementaryTable1.xls
- SupplementaryTable2.xls
- SupplementaryTable3.xls
- SupplementaryTable4.xls
- SupplementaryTable5.xls

- SupplementaryTable6.xls
- SupplementaryTable7.xls
- SupplementaryTable8.xls
- SupplementaryTable9.xls
- SupplementaryTable10.xls
- SupplementaryTable11.xls