J Neuropathol Exp Neurol
Vol. 81, No. 9, September 2022, pp. 707–716
doi: 10.1093/jnen/nlac059

OXFORD

ORIGINAL ARTICLE

Isocitrate Dehydrogenase Mutations Are Associated with Different Expression and DNA Methylation Patterns of *OLIG2* in Adult Gliomas

Huan Mo, MD, MS, Shino Magaki, MD, PhD, Jeremy K. Deisch, MD, and Ravi Raghavan, MBBS, MD, MRCPath

Abstract

Isocitrate dehydrogenase (IDH) mutant gliomas are associated with a better prognosis in comparison to adult IDH wild-type glioma and glioma-CpG island methylator phenotypes. Although *OLIG2* is mainly expressed in oligodendrocytes in normal adult brain, it is expressed in both astrocytomas and oligodendrogliomas. Utilizing the clinical, DNA methylation, and RNA-sequencing data from the Cancer Genome Atlas (TCGA) for lower-grade glioma and glioblastoma cohorts, we explored the association between IDH mutation status and *OLIG2* expression on transcription, DNA methylation, and gene target levels. Compared to IDH wild-type gliomas, IDH mutant gliomas showed consistently higher expression of *OLIG2* transcripts. *OLIG2* overexpression is a good surrogate marker for IDH mutation with an AUC of 0.90. At the DNA methylation level, IDH-mutant gliomas showed hyper- and hypomethylation foci upstream of the *OLIG2* transcription start site. Underexpressed OLIG2 target genes in IDH mutant glioma were enriched in cell cycle-related pathways. Thus, the differential expression of *OLIG2* between IDH mutant and wild-type gliomas reflects involvement in multiple pathways in tumorigenesis.

Key Words: Dehydrogenase, Glioma, Isocitrate, Oligodendrocyte, Transcription Factor 2.

INTRODUCTION

Isocitrate dehydrogenase (IDH) enzymes, including *IDH1* and *IDH2*, are involved in several major metabolic processes, such as the Krebs cycle, glutamine metabolism, lipogenesis, and redox regulation (1–4). Normally, IDH catalyzes the oxidative decarboxylation of isocitrate to α-ketoglutarate (α-KG). Studies have shown that a subset of glioblastomas (GBMs) and the majority of lower-grade gliomas (LGGs) show recurrent missense mutations at specific arginine residues in the active site of these enzymes, for *IDH1* most commonly *p.R132H* and for the homologous enzyme *IDH2* most commonly *p.R172H* (5). The mutant forms of IDH exhibit neomorphic activity, converting α-KG into D-2-hydroxyglutarate (2-HG), an oncometabolite that is produced at high levels and alters histone methylation patterns resulting in CpG island hypermethylation (6–8).

Assessment of the mutational status of *IDH1* and *IDH2* is essential in the classification and grading of gliomas (9). Histologically, LGGs that are wild-type for IDH are molecularly and clinically similar to IDH-wild-type GBM (10, 11). The converse also applies, as previous studies have shown that IDH mutation was associated with a cluster of GBM that has a glioma-CpG island methylator phenotype (G-CIMP) and predicted better survival (8).

The gene product of *oligodendrocyte transcription factor 2* (*OLIG2*) has been a widely used immunohistochemical marker by pathologists (12–14), and plays a pivotal role in glioma development, with positive feedback loops involving epigenetic regulation and receptor tyrosine kinases (15, 16). The expression of *OLIG2* is an early marker for oligodendroglial lineage and is seen in all grades of diffuse gliomas. An induced deletion of *OLIG2* in glioma models has been shown to decelerate tumor growth with a shift from a proneural towards an astrocyte-associated gene expression pattern (16). The relationships between *OLIG2* expression, *IDH1*/*IDH2* mutation status, and the presence of 1p/19q codeletion, a defining alteration for oligodendroglioma, have not been elucidated.

The Cancer Genome Atlas (TCGA) pilot project identified genetic changes of primary DNA sequence and copy number, DNA methylation, gene expression, and clinical information for a set of lower-grade (LGG cohort, defined as former WHO Grade II and III astrocytomas and oligodendrogliomas) (10) and high-grade (GBM cohort, defined as former WHO Grade IV) gliomas (17). With the breadth of multiomic

From the National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA (HM); Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA (SM); Department of Pathology, Loma Linda University Medical Center and School of Medicine, Loma Linda, California, USA (JKD, RR).
Send correspondence to: Ravi Raghavan, MBBS, MD, MRCPath, Department of Pathology, Loma Linda University Medical Center, 11234 Anderson St, Loma Linda, CA 92354, USA; E-mail: rraghavan@llu.edu.
Funding: None declared.
The authors have no duality or conflicts of interest to declare.
Supplementary Data can be found at academic.oup.com/jnen.

information, these adult glioma cases can be reclassified with 2021 WHO Classification of Central Nervous System (WHO CNS5) (18). In particular, we identified cases that are newly classified as *CDKN2A*/*CDKN2B* homozygous deletion in IDH-mutant astrocytomas, which were classified as LGG in TCGA but are now considered WHO Grade 4 (19, 20).

In this study, we explored differences in *OLIG2* expression (in the context of genomic transcription and methylation landscape) between IDH mutant and wild-type gliomas using a combined dataset of LGG and GBM cohorts from the TCGA database. We also explored the changes in expression of OLIG2 target genes and involved pathways.

MATERIALS AND METHODS

The overall study design is summarized in Figure 1. In this report, Roman numeral grades (i.e. II, III, and IV) were used for legacy histological grades provided by the data source that used old WHO grading systems, while Arabic numeral grades (i.e. 2, 3, and 4) were used for modified grades according to the WHO 2021 CNS5 classification (18).

GBMLGG Dataset

Previously harmonized clinical (including histologic diagnoses and survival data), mutational, DNA methylation (by Illumina HumanMethylation450 BeadChip), and gene expression (by RNA-sequencing) data from GBM and LGG cohorts (GBMLGG) (21) were obtained from Xena Platform hosted by University of California, Santa Cruz (http://xena.ucsc.edu) (22). The IDH mutant group (MT) was defined as cases with either *IDH1 p.R132* or *IDH2 p.R172* mutations. The IDH-wild-type group (WT) was defined as cases without any somatic *IDH1* or *IDH2* mutations. Three cases were excluded due to *IDH2* somatic alterations in sites other than *p.R172*.

Detection of Chromosomal Codeletion of 1p/19q and *CDKN2A*/*CDKN2B* Deletion

To molecularly identify oligodendrogliomas, we utilized DNA copy number analysis (CNA) data (with germline variations removed) as log2 (tumor/normal) scores (17, 23) to computationally detect chromosomal deletions of the short arm of chromosome 1 (1p) and the long arm of chromosome 19 (19q). For convenience, CNA segments are mapped and averaged (with weights of segment lengths) to chromosome bands. A deletion of chromosome band is defined as ≤ -0.5 in the CNA score. Although a loss of the chromosome is defined as the loss of at least half of their bands (totally 33 bands in 1p and 12 in 19q), the final cases with 1p/19q codeletion have lost at least 29 bands in 1p and 9 in 19q.

CDKN2A and *CDKN2B* are 2 closely adjacent genes located in chromosome 9p21.3. We utilized DNA CNA GISTIC2-Threshold data (23) provided by UCSC Xena repository, which produced gene-level copy number estimates that were further thresholded to −2, −1, 0, 1, and 2. Only cases with the same copy number between *CDKN2A* and *CDKN2B* and nongain copy numbers are included (only 8 out of 431 cases in the IDH-mutant group or 24 out of 809 total cases in the combined GBMLGG dataset were excluded). All cases with 1p/19q codeletions were removed from this analysis due to lack of clinical significance from previous studies (19, 20).

Differences in Genomic DNA-Methylation and Expression Between IDH Mutant and Wild-Type Groups

The DNA methylation level for an individual CpG site (with an Illumina cg identifier) was presented as the percentage of alleles methylated (ranging from 0 to 1), which was termed beta-value. Because the beta-values commonly had bimodal distribution in the 2 ends (either near-100% unmethylated or near-100% methylated for a site), they were converted to M-values (24) for normalization with a formula M-value = log2 ([beta + psc]/[1−beta + psc]), where pseudocount (psc, 0.0001) was introduced to limit the skewing from outliers. For each CpG site, we compared the M-values between the IDH mutated and wild-type groups with 2-tailed Student *t*-tests for independent groups. In the discussion of individual CpG sites (such as in the *OLIG2* gene), we reported the difference of median beta-values (percentages of methylation) between the 2 groups to facilitate understanding.

Expression levels of each gene were presented as RSEM scores (25), a normalized and base 2 logarithm-transformed score from RNA-sequencing counts. For each gene, the *RSEM scores* were compared between the IDH mutant and wild-type groups with 2-tailed Student *t*-tests for independent groups.

To study the correlation of the *OLIG2* expression with the transcriptome and DNA methylome (in M-value), we used the Empirical Bayes statistic models (26, 27), adjusted with IDH mutation status.

Statistical significance was defined by *Bonferroni* corrected cutoffs (28) with the numbers of tests ($\alpha = 0.05$/numbers of tests). All the Student *t*-tests were performed as equal variances not assumed to avoid false positivity.

Independent Gene Expression Profile Data for Corroboration

As a corroboration, we used gene expression profiles (GEPs) from an independent glioma cohort (GSE4290) (29) from Gene Expression Omnibus (GEO), which was procured with Affymetrix Human Genome U133 Plus 2.0 microarray platform. Expression levels are log 2 transformed and annotated with GPL570 file from GEO (accessed on September 27, 2021, last update on December 14, 2020). *OLIG2* expression level for each specimen was defined as the average of the log 2 readings of probes "213824_at" and "213825_at." Due to the unavailability of IDH data, the grouping was defined by reported histology and clustering shown by Uniform Manifold Approximation and Projection (UMAP) dimension reduction.

OLIG2-Protein Binding Sites

To determine OLIG2 target genes, we used OLIG2 chromatin immunoprecipitation sequencing (ChIP-Seq) data previously reported by Jolma et al (European Nucleotide Archive [ENA] accession numbers ERP001824 and ERP001826) (30). The alignment and peak-calling were performed as described in

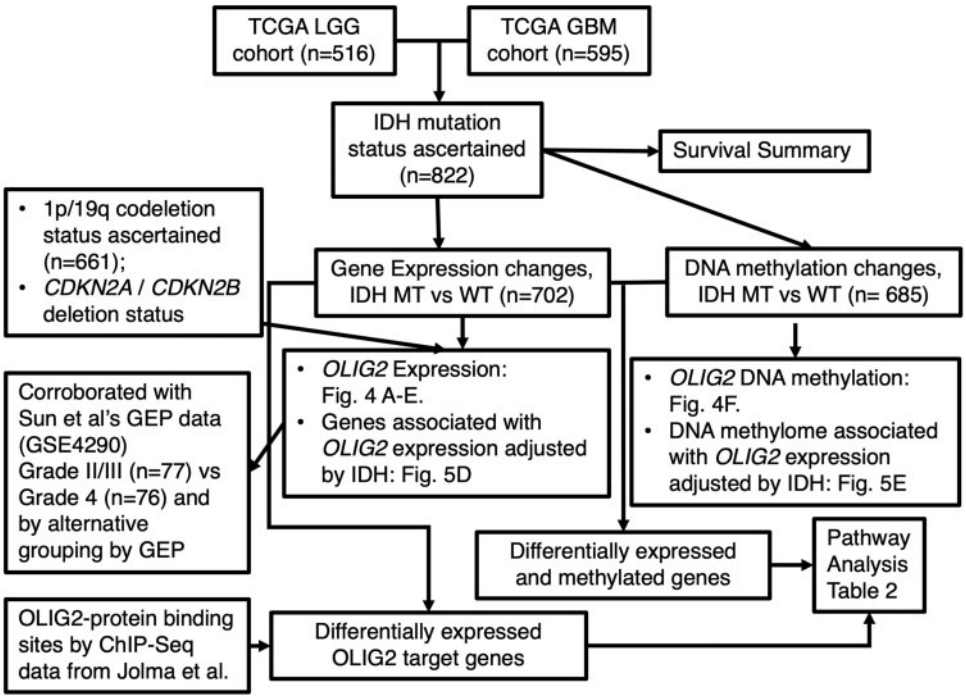

FIGURE 1. Overall flowchart of the study design. TCGA, The Cancer Genome Atlas; LGG, lower-grade glioma cohort; GBM, glioblastoma cohort; GEP, gene expression profile; MT, mutant; WT, wild-type; ChIP-Seq, chromatin immunoprecipitation sequencing.

the original paper (116 688 peaks), with GRCh38/hg38 as the reference genome. After comparison of output and peak-calling quality, we selected the ChIP-Seq dataset performed against the full OLIG2 protein (ERR193878, ERR193879, ERR193880, and ERR193881), rather than the DNA-binding domain. The genes corresponding to the peaks were defined by neighboring genes within a distance of 5000 bp from peaks as in previous mouse studies described by Satoh et al (31). A total of 13 694 genes were assigned as OLIG2 target genes (compared with ~6000 genes in mouse reported by Satoh et al). The CpG sites corresponding to the ChIP-Seq peaks were defined as being within the peaks and a total of 25 893 CpG sites were identified.

Pathway Analysis

We used Illumina Beadchip annotation file (GPL16304) to associate the genes and the neighboring CpG sites. Significant differentially expressed genes (between IDH mutant and wild-type groups) were separated into 3 lists: increased expression in IDH mutant group with associated differentially methylated CpG site(s), decreased expression in IDH mutant group with associated differentially methylated CpG site(s), and increased or decreased expression without associated differentially methylated CpG sites. Pathway enrichment analysis was performed on https://reactome.org/AnalysisService/, with all default settings.

Statistics

Data transformation and Student's *t*-tests are performed on *KNIME* v4.4 platform (www.knime.com). The Empirical Bayes statistic models were performed with the "limma" (27) package on R v3.6 (www.r-project.org). Survival analysis and diagram plotting were performed with Kaplan-Meier survival curves with the "survminer" (32) and "survival" (33) packages on R. All the UMAP models for GEP were constructed with the "umap" package (34, 35) on R with "n_neighbors = 15, random_state = 123" setting, although the plotting was performed on Python 3 for esthetic considerations.

RESULTS

Glioma Cohorts and Datasets

There were 516 patients in the LGG cohort and 595 patients in the GBM cohort. There were 822 patients whose IDH statuses were ascertained through whole exome sequencing data. There were 661 patients whose 1p/19q statuses were ascertained through CNA. There were 702 patients with gene expression data and 685 patients with DNA methylation data, although the total patient numbers for individual genes or CpG sites may vary due to the nature of high-throughput platforms (Table 1).

The IDH mutation pattern comprised 375 cases with *p.R132H*, 17 cases with *p.R132C*, 12 cases with *p.R132G*, and 9 cases with *p.R132S IDH1* mutations. In addition, there were 20 cases with *IDH2 p.R172* mutations (12 of which [or 60%] had 1p/19q codeletion).

Similar to previous observations, patients with LGG and IDH wild-type status (median OS = 758 days) had significantly worse overall survival than those with LGG and IDH mutant status (median OS = 2988 days), although better than

TABLE 1. Clinical Characteristics of the TCGA Sample Sets

	IDH Mutated			IDH Wild-Type
	Overall	1p/19q Codeleted	1p/19q Noncodeleted	
Total	433 413 *IDH1* mutated 20 *IDH2* mutated	139	283	389 (228 ascertained 1p/19q noncodeleted)
Cohorts				
LGG	417	139	277	93
GBM	16	0	6	296
Median age (years)	39.2	43.8	38.2	60.5

LGG, low-grade glioma; GBM, glioblastoma; IDH, isocitrate dehydrogenase, including *IDH1* and *IDH2*.

patients with IDH wild-type GBM (median OS = 406 days) (Fig. 2A). The histologic features of IDH mutant and IDH wild-type tumors were similar within each cohort (LGG and GBM) from reviewing the Digital Pathology Images in TCGA (Fig. 2E–H). The overall mRNA sequencing data showed distinct separation between IDH mutant and wild-type gliomas by UMAP dimension reduction (Fig. 3A, B).

OLIG2 Expression

OLIG2 expression was significantly higher in IDH mutant tumors compared to wild-type tumors overall (1.93 [1.73–2.12] fold, $p = 2.08 \times 10^{-67}$), as well as both LGG (1.54 [1.22–1.87] fold, $p = 2.64 \times 10^{-16}$) and GBM cohorts (2.00 [1.24–2.77] fold, p = 0.0002) (Fig. 4A).

Within the IDH mutant tumors, *OLIG2* expression was significantly but only slightly higher in 1p/19q codeleted tumors (oligodendrogliomas) compared to those without the codeletion (astrocytomas) (0.20 [0.05–0.35] fold, p = 0.011). After excluding oligodendrogliomas (with 1p/19q codeletion), *OLIG2* expression was significantly higher in IDH mutant gliomas compared to IDH wild-type tumors (1.87 [1.62–2.12] folds, $p = 2.76 \times 10^{-39}$) (Fig. 4B).

Within IDH mutant, 1p/19q noncodeleted astrocytoma cases in both LGG and GBM cohorts, 179 cases were *CDKN2A/CDKN2B* copy number neutral (denoted as GISTIC2-Threshold score 0), 69 cases were −1 (possibly single allelic deletion or subclonal homozygotic deletion), and 26 cases were −2 (likely homozygotic deletion). For each GISTIC2-Threshold score deletion (0, −1, and −2), OLIG2 expression was decreased by 0.29 (0.12–0.46) folds (p = 0.00086). The scale of this change is much smaller than the comparison between IDH-mutant and wild-type groups but similar to that between IDH mutant astrocytomas and oligodendrogliomas (Fig. 4C).

A receiver operating characteristic (ROC) curve for using *OLIG2* expression level to predict IDH mutation status yielded an area under the curve (AUC) of 0.899 (Fig. 4D). Within the IDH mutant tumors, an ROC curve for using *OLIG2* expression level to predict 1p/19q codeletion yielded an AUC area of 0.546 (Fig. 4E).

Differential *OLIG2* Expression Corroboration by Independent Dataset

To corroborate the finding of differential expression of *OLIG2* between IDH mutant and wild-type gliomas, we used GEP dataset of glioma samples (GSE4290) (29). This dataset did not provide IDH mutation status, so we used histologic grades as surrogate for IDH status (i.e. Grade II and III to imply IDH mutant, while Grade IV to imply IDH wild-type). As a sanity check (36), we performed a global UMAP clustering with all the genes and observed that most of the Grade IV gliomas were clustered together (66/76, 87%) and most of the Grade II and III gliomas were clustered together (54/76, 71%) (Fig. 3C, compared with Fig. 3A, B from TCGA cohorts), suggesting grouping by histologic grades is imperfect but informative. In this dataset, *OLIG2* expression is 0.83 (0.44–1.22)-fold higher in Grade II or III gliomas compared to Grade IV gliomas ($p = 4.29 \times 10^{-5}$).

Alternatively, we also attempted to reclassify the cases by GEP clusters suggested by UMAP (Fig. 3D). The cluster with enriched Grade IV gliomas was inferred as "IDH-wild-type-like" cases (n = 87); the cluster with enriched Grade II, III, and oligodendrogliomas was inferred as "IDH-mutant-like" cases (n = 49); the cluster admixed with nontumor brain tissue (from brain tissue resected for epilepsy) was inferred as "normal-like" cases (n = 16), which were removed from the analysis due to a concern about tumor purity. Compared with IDH-wild-type-like cases, IDH-mutant-like cases have 1.41 (1.06–1.76) folds higher expression of *OLIG2* ($p = 6.2 \times 10^{-13}$).

DNA Methylation of *OLIG2*

Methylation analysis revealed that, in comparison with IDH wild-type tumors, IDH mutant tumors had a significantly hypermethylated region (−6124 to −5397 nucleotides upstream from *OLIG2* transcription start site [TSS], up to 0.47 in difference of median beta-value [e.g. cg06412358]) and a significantly hypomethylated region (−1647 to −1018 nucleotides upstream from *OLIG2* TSS, up to −0.48 in difference [e.g. cg07601542]). This finding was similar in both LGG and GBM cohorts (Fig. 4F).

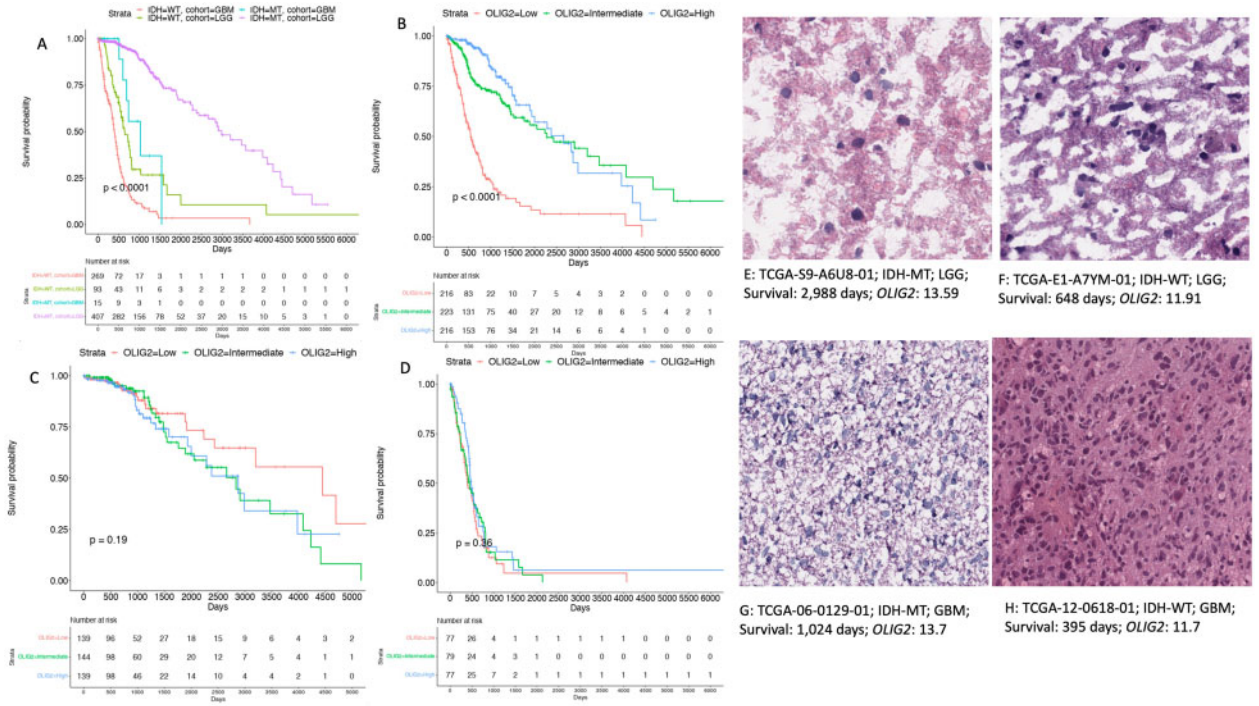

FIGURE 2. Overall clinicopathological characteristics associated with *OLIG2* expression. **(A–D)** Kaplan-Meier curves of TCGA glioblastoma (GBM) and lower-grade glioma (LGG) cohorts for overall survival, grouped by histology and IDH mutation status (WT, wild-type; MT, mutant) **(A)**, including all individuals in the merged cohort (GBMLGG) grouped by *OLIG2* expression levels **(B)**, including IDH wild-type cases **(C)**, and including Only IDH mutant cases **(D)**. **(B–D)** *OLIG2* Group 1 (red) is designated as the lowest third of *OLIG2* expression within the individuals included in each graph, Group 2 (green) is designated as middle third, and Group 3 (blue) is designated as highest third. **(E–H)** Histology images of each median-survival individual in the 4 cohorts with their *OLIG2* expression levels (in log2 scale). The images were captured from TCGA Digital Pathology Images.

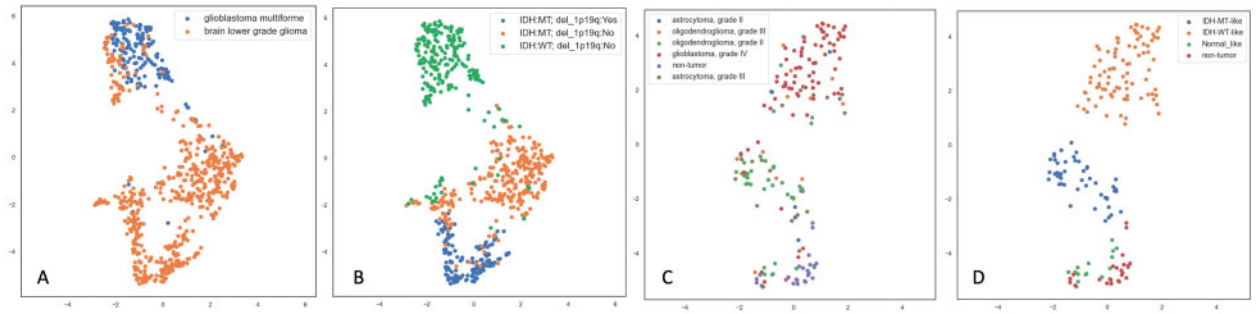

FIGURE 3. Uniform Manifold Approximation and Projection (UMAP) dimension reduction and clustering plots for gene expression profiles (GEP) for TCGA merged glioblastoma and lower-grade glioma cohort (GBMLGG), labeled by histological grade **(A)** and labeled by IDH mutant (MT) or wild-type (WT) status and chromosome 1p and 19q codeletion status **(B)**. Both plots **(A, B)** were produced based on the same UMAP output and showed good separation between IDH wild-type and mutant gliomas with GEP data. **(C)** UMAP for independent corroboration cohort (GSE4290; Sun et al [29]) of gliomas and nonneoplastic brain tissues, labeled by reported histology diagnoses. **(D)** An alternative grouping based on panel **(C)**; only IDH-wild-type-like and IDH-mutant-like gliomas were compared for *OLIG2* expression.

Outcome Analysis for *OLIG2* Expression

For each analysis, cases were divided into 3 groups based on *OLIG2* expression levels: Group 1 was the lowest third in *OLIG2* expression levels while Group 3 was the highest third. In the overall GBMLGG cohort, the median overall survivals are 548 (485–675) days for Group 1, 2282 (1762–4695) days for Group 2, and 2660 (1933–4229) days for group 3. Group 3 shows significantly longer overall survival than Group 1 with a hazard ratio of 0.418 (0.351–0.499, $p < 0.0001$) (Fig. 2B). However, there were no overall survival differences between groups within either IDH mutant (Group 3 vs Group 1 HR = 1.236 [0.9182–1.663], $p = 0.2$) or wild-type (0.8635 [0.7144–1.044], $p = 0.1$) cohorts (Fig. 2C, D).

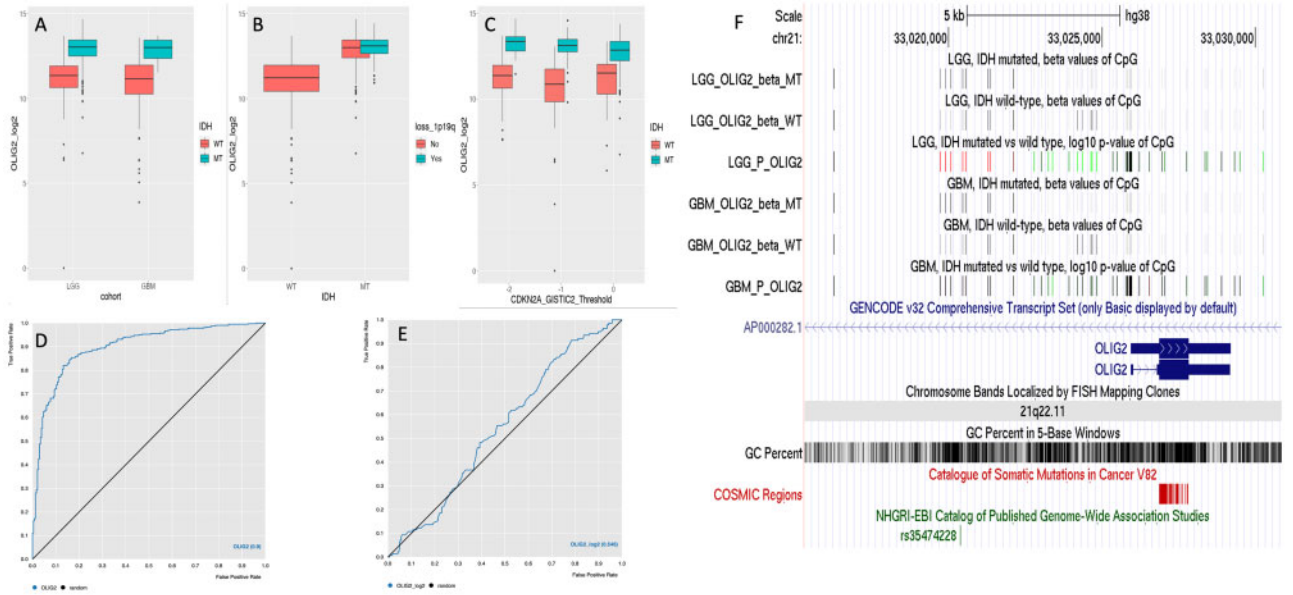

FIGURE 4. Comparison of *OLIG2* gene between IDH mutant (MT) and wild-type (WT) groups. **(A)** *OLIG2* was significantly overexpressed in the IDH mutant gliomas in both lower-grade glioma (LGG) and glioblastoma (GBM) cohorts. **(B)** *OLIG2* was significantly overexpressed in the IDH mutant gliomas, regardless of 1p/19q status, although *OLIG2* was expressed slightly (but significantly) higher in IDH mutant gliomas with 1p/19q codeletion (oligodendrogliomas) than in IDH mutant gliomas without 1p/19q codeletion (astrocytomas). **(C)** *OLIG2* expression was slightly affected by *CDKN2A*/*CDKN2B* copy number deletion in IDH-mutant astrocytomas without 1p/19q codeletion, but the effect size is much smaller than that of IDH status. **(D)** Receiver operating characteristic (ROC) curve for using *OLIG2* transcription level to predict IDH mutation status using the merged GBMLGG cohort (area under the curve [AUC]=0.9). **(E)** ROC curve for using *OLIG2* transcription level to predict 1p/19q codeletion status within IDH mutant gliomas (AUC = 0.546). **(F)** The genomic geographic relationship between differentially methylated CpG sites in *OLIG2*, generated by University of California Santa Cruz Genome Browser (http://genome.ucsc.edu/). LGG/GBM_OLIG2_beta_MT/WT: median *beta-values* (percentage of methylation) of each CpG site within the *OLIG2* gene, from 0% (white) to 100% (black). LGG/GBM_P_OLIG2: p values of each CpG site. Sites that were significantly hypermethylated in IDH mutant groups are in red, while sites that were significantly hypomethylated are in green and nonsignificant sites are in black.

Global DNA Methylation and Gene Expression Differences Between IDH Groups

In our gene expression (transcriptome) study, 20 242 genes were successfully compared between the IDH mutant and wild-type tumors, with the *Bonferroni* correction cutoff for the p value of 2.47×10^{-6}. There were 11 605 genes significantly differentially expressed between the 2 groups, with decreased expression of 6398 (31.6%) and increased expression of 5207 (25.7%) genes in the IDH mutant group (Fig. 5A).

In our genome-wide methylation (methylome) study with combined cohort GBMLGG, 393 153 CpG sites were successfully compared between IDH mutant and wild-type groups. Therefore, the *Bonferroni* correction cutoff for the p value of the methylome was 1.27×10^{-7}. There were 210 193 sites (53.5%) significantly differentially methylated CpG sites between the 2 groups, with 192 407 (48.9%) hypermethylated and 17 786 (4.5%) hypomethylated sites in the IDH mutant group (Fig. 5B). As a visual comparison, we also performed a similar comparison between IDH mutant and wild-type acute myeloid leukemia with the TCGA LAML cohort (37) (Fig. 5C).

Among the interrogated genes, 19 926 genes had at least 1 CpG site interrogated in the methylome study. Among the significantly differentially expressed genes, 11 044 genes also had at least 1 differentially methylated CpG site, while 476 genes did not. Among the nonsignificantly differentially expressed genes, 7951 genes had significantly differentially methylated CpG sites, while 455 genes did not. The p value for chi-square test was 2.3×10^{-5}, suggesting the differential methylation pattern was significantly associated with the differential gene expression pattern.

Genes and CpG Sites Associated With *OLIG2* Expression

Since IDH mutation status has been shown to be highly associated with *OLIG2* expression, as well as with the transcriptome and methylome, to assess correlation of gene expression with *OLIG2* (instead of IDH), we utilized a linear model between the expressions of *OLIG2*, and genes of interest adjusted by IDH status. Out of 20 529 genes tested, 5504 genes (26.8%) were significantly associated with *OLIG2* expression. Although there were similar numbers of positively associated genes (n = 2869, 14.0%) as the negatively associated genes (n = 2635, 12.8%) that passed the *Bonferroni* correction line, the top genes with top p values and absolute values of coefficients were exclusively positively associated with *OLIG2* expression (Fig. 5D and Supplementary Data S1). Out of 8956 genes with neighbor-

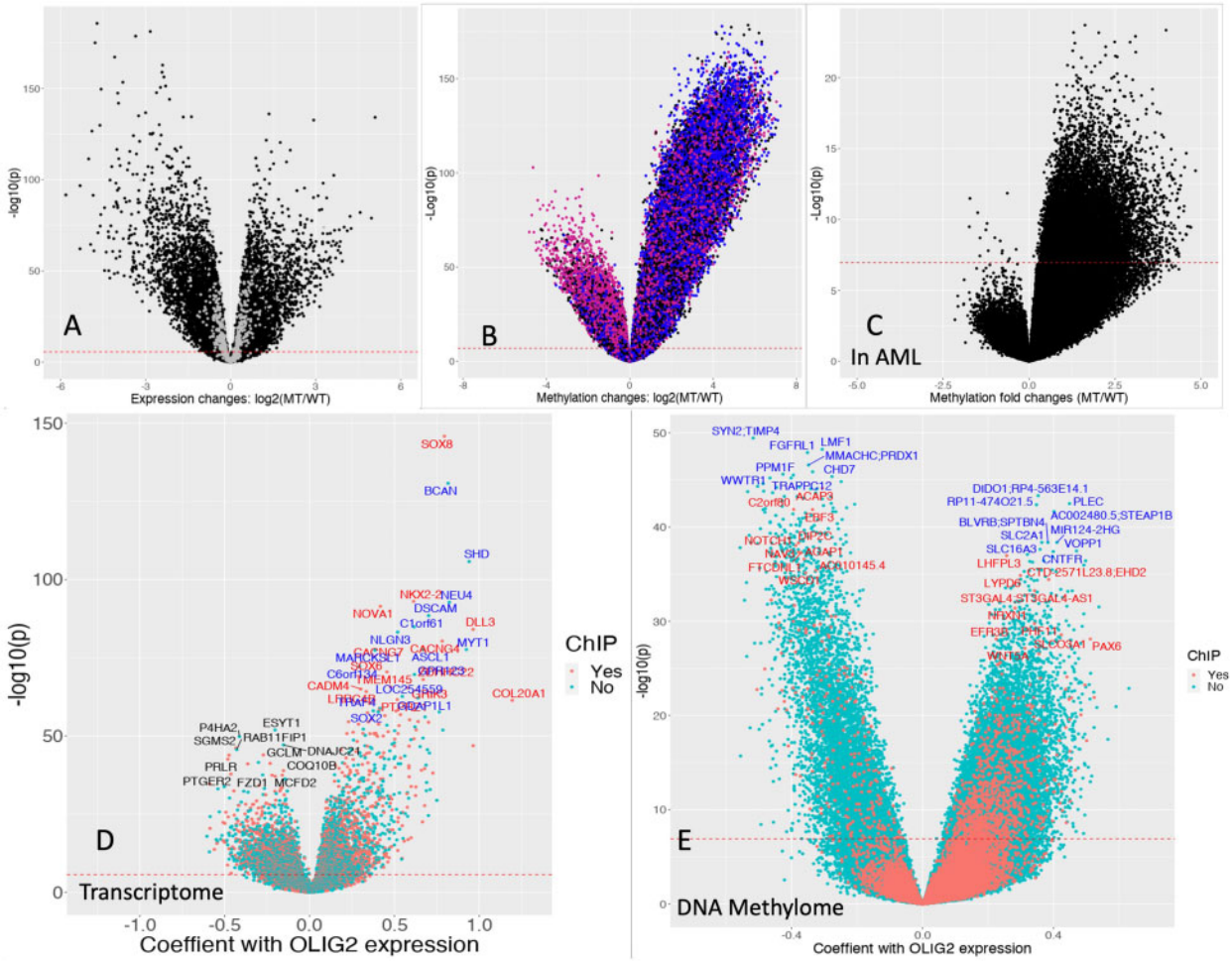

FIGURE 5. Volcano plots for differential gene expression **(A)** and DNA methylation **(B)** for merged GBMLGG cohort when IDH mutant (MT) glioma and wild-type (WT) glioma were compared. **(A)** Each dot represents one gene, and the gray ones represent genes without associated differential DNA methylation. **(B)** Each dot represents a CpG site. The violet-red dots represent overexpressed gene-associated CpG sites in the IDH mutant group, while the blue dots represent the underexpressed gene-associated CpG sites. The black dots indicate genes without assigned significantly differentially methylated CpG sites between the 2 groups. **(C)** Volcano plot comparing genomic DNA methylation changes between IDH mutant and wild-type groups in the Cancer Genome Atlas (TCGA) acute myeloid leukemia (AML) cohort (LAML). **(B, C)** The dots located on the right side of 0 on the *x* axis indicate hypermethylation in the IDH mutant group. **(D, E)** Linear association studies of gene expressions **(D)** and DNA methylations of CpG sites **(E)** with *OLIG2* expression adjusted by IDH status; ChIP: OLIG2-protein binding sites from chromatin immunoprecipitation sequencing data. **(D)** Red dots and labels represent genes that have neighboring OLIG2 protein binding sites; green dots and blue labels represent genes that do not have neighboring OLIG2 binding sites; additional black labels represent top genes that were negatively associated with *OLIG2* expression. *OLIG1* expression was highly correlated with *OLIG2* expression and off the plotted areas [coefficient = 0.091, log10(p) = −293, no OLIG2 binding sites]. **(E)** Red dots and labels represent CpG sites (and neighboring genes) that are located within OLIG2-protein binding sites; blue dots and labels represent CpG sites that are not located within binding sites. In all panels, the red dashed lines indicate *Bonferroni* corrected cutoff for statistical significance.

ing OLIG2-protein binding peaks, 2573 genes (28.7%) were significantly associated with *OLIG2* expression after a correction with the overall *Bonferroni* cutoff, with 1391 genes (15.5%) positively and 1182 genes (13.2%) negatively associated.

Similarly, we utilized a linear model to highlight the expression of *OLIG2* and DNA methylome adjusted by IDH status (Fig. 5E and Supplementary Data S2). Out of 396 065 tested CpG sites, 30 180 sites (7.6%) were significantly associated with *OLIG2* expression after *Bonferroni* correction, with 22 881 sites (5.8%) positively and 7299 sites (1.8%) negatively associated. Out of 23 002 CpG sites that locate within the OLIG2 ChIP peaks, 1144 sites (5.0%) were positively and 301 sites (1.3%) negatively associated with *OLIG2* expression after a correction with the overall *Bonferroni* cutoff (1445 sites [6.3%] in total).

Pathway Analysis

The top enriched pathways for genes (n = 4195) with increased expression in the IDH mutant group with differentially methylated CpG sites (likely driving-genes) in IDH mutant tumors, were "protein-protein interactions at synapses" ($p = 1.6 \times 10^{-6}$), "neurexins and neuroligins" ($p = 1.8 \times 10^{-5}$), and "neuronal system" ($p = 1.3 \times 10^{-3}$). On the other hand, genes (n = 5333) with decreased expression in the IDH mutant group with differentially methylated CpG sites were enriched in "interferon gamma signaling" ($p = 2.4 \times 10^{-14}$), "interferon alpha/beta signaling" ($p = 8.3 \times 10^{-12}$), and "interferon signaling" ($p = 1.7 \times 10^{-11}$) pathways.

Differentially expressed genes without differentially methylated CpG sites (n = 381) were enriched in "processing of capped intron-containing pre-mRNA" ($p = 1.3 \times 10^{-5}$), "mRNA splicing—major pathway" ($p = 2.1 \times 10^{-5}$), and "mRNA splicing" ($p = 4.3 \times 10^{-5}$). These genes were likely related to IDH-independent mechanisms in IDH wild-type tumors.

To evaluate the expressions of the genes potentially regulated by OLIG2, we also performed a gene set enrichment analysis on OLIG2 target genes that were significantly underexpressed in the IDH mutant tumors (n = 3030). Pathways for "G1/S Transition" ($p = 8.7 \times 10^{-6}$) and "DNA Replication" ($p = 1.1 \times 10^{-5}$) were enriched, while the enrichment of genes involving in "interferon gamma signaling," seen in the underexpressed genes associated with differential methylation, was absent. There were 2489 OLIG2 target genes overexpressed in the IDH mutant tumors, and pathways "neuronal system" ($p = 2.9 \times 10^{-7}$), "protein-protein interactions at synapses" ($p = 1.4 \times 10^{-6}$), and "neurexins and neuroligins" ($p = 8.8 \times 10^{-6}$) were enriched, similar to the overexpressed genes associated with differential methylation (Table 2).

DISCUSSION

In this study, we identified the differential expression of *OLIG2* on transcription levels between IDH mutant and wild-type gliomas, regardless of histologic grade. We corroborated this finding with both RNA-sequencing data from TCGA cohorts (LGG, GBM, and merged) and microarray-based gene expression profiling data from Sun et al (29). In survival analysis, higher mRNA levels of *OLIG2* were associated with better overall survival. However, this association was completely dependent on IDH mutation status. Also, mRNA levels of *OLIG2* are a good predictor for IDH mutation status with an AUC of ~0.90. Therefore, a well validated quantitative assay of *OLIG2*, at least on the transcription level, may serve as a good surrogate biomarker to predict IDH status and overall survival, although the assay modalities and validation are beyond the scope of this report.

Nevertheless, compared with 375 cases (86.6% of all IDH-mutant cases) with *IDH1 p.R132H* mutation (with available IHC assays), there were 38 cases (8.8%) with other amino acid substitutes at the *p.R132* locus and another 20 cases (4.6%) with *IDH2* p.R172 mutations (12 of which [or 60%] had 1p/19q codeletion). These statistics provide quantified estimates of the need for additional molecular tests for IHC negative cases.

In nonneoplastic adult brains, *OLIG2* is mostly expressed in oligodendrocytes but not on astrocytes. However, *OLIG2* is expressed in most gliomas including oligodendrogliomas and astrocytomas. Indeed, within the IDH mutant gliomas, *OLIG2* mRNA levels were minimally higher in oligodendrogliomas (defined by both IDH mutation and 1p/19q codeletion), compared with astrocytomas. These findings exclude the possibility that the association between *OLIG2* expression and IDH status were confounded by admixed oligodendrogliomas in the IDH mutant glioma cohort. It should be also noted that the *OLIG2* mRNA level was a poor predictor for 1p/19q codeletion status with an AUC of ~0.55 within IDH mutant gliomas. Similarly, the differential expression of *OLIG2* was not confounded by *CDKN2A*/*CDKN2B* deletion.

To explore possible associations between IDH mutation and increased *OLIG2* expression, we compared the DNA methylation patterns at the *OLIG2* locus in IDH mutant and wild-type gliomas, given that IDH mutation is associated with genomic DNA hypermethylation (or a CpG-island methylator phenotype [G-CIMP] through clustering of methylome patterns [8, 38]). We identified a hypermethylation focus located ~6000 nucleotides upstream from the *OLIG2* TSS and a hypomethylation foci located ~1500 nucleotides upstream. The differential methylation patterns in the *OLIG2* locus suggest possible roles of the global G-CIMP phenotype and epigenetic mechanisms in the regulation of *OLIG2* expression.

Next, we explored the differential GEPs for the OLIG2 target genes, which were defined with ChIP-Seq data (30). Similar to ChIP-Seq studies performed in mice (31), transcription factor OLIG2 binding sites are widely distributed in the genome (~230 000 peaks) and neighboring ~14 000 genes (~9000 genes with RNA-seq data) in human. The regulatory roles of the OLIG2 transcription factor in these target genes are complex with coregulator interactions and possible post-translational modifications of the OLIG2 protein (39), but OLIG2 generally acts as a transcriptional repressor of direct target genes (40). We identified ~3000 underexpressed and ~2500 overexpressed OLIG2-target genes in IDH mutant gliomas compared to IDH wild-type gliomas. Interestingly, the pathway enrichment profiles are different for underexpressed genes in IDH mutant gliomas between OLIG2-target genes and differential methylation-associated genes. The underexpressed OLIG2-target genes in IDH mutant gliomas were enriched in cell cycle pathways, while absent of enrichment of inflammation-related genes (such as interferon gamma signaling pathways). These findings implicate a possible antimitotic and proneurogenic role of the OLIG2 transcription factor, potentially with the effects of global differential methylation.

We also performed transcriptome and DNA methylome association studies with *OLIG2* expression adjusted by IDH status. In the volcano plot for the transcriptome, we observed an asymmetric pattern with a long tail of genes highly positively associated with OLIG2 expression. Many of these highly positively associated genes involve in neurological functions (such as *OLIG1*, *SOX8*, *BCAN*, *NKX2-2*, *NOVA1*, and *NEU4*). Similarly, we observed that many top *OLIG2*-associated CpG methylation sites neighboring neurological (such as *ACAP3*, *SYN2*, and *NAV3*), metabolic (such as *LMF1*), and proliferation (such as *FGFRL1* and *NOTCH1*)

TABLE 2. Pathway Analysis (Top 3 Hits for Each Category)

Pathway ID*	Pathway Name	p Value	FDR
	Underexpressed OLIG2 target genes†		
R-HSA-69206	G1/S transition	8.68E−6	0.012
R-HSA-69306	DNA replication	1.11E−5	0.012
R-HSA-453279	Mitotic G1 phase and G1/S transition	1.96E−5	0.014
	Overexpressed OLIG2 target genes		
R-HSA-112316	Neuronal system	2.91E−7	0.001
R-HSA-6794362	Protein-protein interactions at synapses	1.39E−6	0.001
R-HSA-6794361	Neurexins and neuroligins	8.78E−6	0.006
	Underexpressed genes associated with differential methylation		
R-HSA-877300	Interferon gamma signaling	4.43E−14	1.01E−10
R-HSA-913531	Interferon signaling	2.27E−11	2.58E−8
R-HSA-909733	Interferon alpha/beta signaling	1.02E−10	7.74E−8
	. . .		
R-HSA-453279	Mitotic G1 phase and G1/S transition	7.36E−5	0.015
	Over-expressed genes associated with differential methylation		
R-HSA-6794362	Protein-protein interactions at synapses	2.24E−6	0.005
R-HSA-112316	Neuronal system	1.44E−5	0.013
R-HSA-6794361	Neurexins and neuroligins	1.84E−5	0.013
	Differential expressed gene without association with differential methylation		
R-HSA-72203	Processing of capped intron-containing pre-mRNA	5.86E−6	0.004
R-HSA-72163	mRNA splicing—major pathway	8.17E−6	0.004
R-HSA-72172	mRNA splicing	1.75E−5	0.005

*Based on https://reactome.org/.
†Under-/overexpression is based on IDH mutant to wild-type comparison.
FDR, false discovery rate.

genes. These findings suggest that future function and comprehensive network studies will lead to a more discriminative local regulatory structure.

OLIG2 expression has been hypothesized to be involved in tumorigenesis of gliomas and is associated in PDGF and EGF receptor signaling, especially in those with proneural phenotype (16). However, in this study, we found that high *OLIG2* expression is associated with IDH mutation and present in the subgroup of gliomas with better outcome. The OLIG2 target gene included decreased expressions of cell cycle-related genes and increased expression of neuronal genes in the IDH mutated gliomas. Therefore, this study suggests more complex roles for *OLIG2* in glioma tumorigenesis.

Although expression of *OLIG2* in the in adult brain is generally restricted to the oligodendrocyte lineage, *OLIG2* expression has also been reported in putative astrocyte precursors cells (41) and proliferating reactive astrocyte precursor cells (42). *OLIG2* expression is downregulated upon terminal differentiation of astrocytes. The high expression levels of *OLIG2* in IDH mutant glioma may represent an arrest in a precursor stage (termed proneural phenotype), while the decreased expression of *OLIG2* in high-grade IDH wild-type glioma/GBM may represent loss of a lineage marker. However, further studies are needed to test this hypothesis.

REFERENCES

1. Koh HJ, Lee SM, Son BG, et al. Cytosolic NADP+-dependent isocitrate dehydrogenase plays a key role in lipid metabolism. J Biol Chem 2004; 279:39968–74
2. Hurley JH, Dean AM, Koshland DE, et al.; Stroud RM. Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: Implications from the structures of magnesium-isocitrate and NADP+ complexes. Biochemistry 1991;30:8671–8.
3. Xu X, Zhao J, Xu Z, et al. Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity. J Biol Chem 2004;279:33946–57
4. Waitkus MS, Diplas BH, Yan H. Isocitrate dehydrogenase mutations in gliomas. Neuro Oncol 2016;18:16–26
5. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. N Engl J Med 2009;360:765–73
6. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 2009;462:739–44
7. Christensen BC, Smith AA, Zheng S, et al. DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma. J Natl Cancer Inst 2011;103:143–53
8. Noushmehr H, Weisenberger DJ, Diefes K, et al.; Cancer Genome Atlas Research Network. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell 2010;17:510–22
9. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: A summary. Acta Neuropathol 2016;131:803–20
10. Brat DJ, Verhaak RGW, Aldape KD, et al.; Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. New Engl J Med 2015;372:2481–98
11. Brat DJ, Aldape K, Colman H, 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;136:805–10
12. Yokoo H, Nobusawa S, Takebayashi H, et al. Anti-human Olig2 antibody as a useful immunohistochemical marker of normal oligodendrocytes and gliomas. Am J Pathol 2004;164:1717–25
13. Popova SN, Bergqvist M, Dimberg A, et al. Subtyping of gliomas of various WHO grades by the application of immunohistochemistry. Histopathology 2014;64:365–79

14. Suzuki A, Nobusawa S, Natsume A, et al. Olig2 labeling index is correlated with histological and molecular classifications in low-grade diffuse gliomas. J Neurooncol 2014;120:283–91
15. Tsigelny IF, Kouznetsova VL, Lian N, Kesari S. Molecular mechanisms of OLIG2 transcription factor in brain cancer. Oncotarget 2016;7: 53074–101
16. Lu F, Chen Y, Zhao C, et al. Olig2-dependent reciprocal shift in PDGF and EGF receptor signaling regulates tumor phenotype and mitotic growth in malignant glioma. Cancer Cell 2016;29:669–83
17. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–8
18. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: A summary. Neuro Oncol 2021; 23:1231–51
19. Shirahata M, Ono T, Stichel D, et al. Novel, improved grading system(s) for IDH-mutant astrocytic gliomas. Acta Neuropathol 2018;136:153–66
20. Reis GF, Pekmezci M, Hansen HM, et al. CDKN2A loss is associated with shortened overall survival in lower-grade (World Health Organization Grades II–III) astrocytomas. J Neuropathol Exp Neurol 2015;74: 442–52
21. Ceccarelli M, Barthel FP, Malta TM, et al.; TCGA Research Network. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. Cell 2016;164:550–63
22. Goldman MJ, Craft B, Hastie M, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol 2020;38:675–8
23. Mermel CH, Schumacher SE, Hill B, 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
24. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 2010;11:587
25. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011; 12:323
26. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47
27. *R: Empirical Bayes Statistics for Differential Expression [Internet]*. Available at: http://web.mit.edu/~r/current/arch/i386_linux26/lib/R/library/limma/html/ebayes.html. Accessed May 25, 2022
28. Noble WS. How does multiple testing correction work? Nat Biotechnol 2009;27:1135–7
29. Sun L, Hui AM, Su Q, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. Cancer Cell 2006;9:287–300
30. Jolma A, Yan J, Whitington T, et al. DNA-binding specificities of human transcription factors. Cell 2013;152:327–39
31. Satoh JI, Asahina N, Kitano S, et al. A comprehensive profile of ChIP-Seq-based Olig2 target genes in motor neuron progenitor cells suggests the possible involvement of Olig2 in the pathogenesis of amyotrophic lateral sclerosis. J Cent Nerv Syst Dis 2015;7:1.
32. Kassambara A, Kosinski M, Biecek P, et al. Survminer: Drawing Survival Curves using "ggplot2" [Internet]. 2021. Available at: https://CRAN.R-project.org/package=survminer. Accessed May 25, 2022
33. Therneau TM, (until 2009) TL original S >R port and R maintainer, Elizabeth A, Cynthia C. *survival: Survival Analysis [Internet]*. 2022. Available at: https://CRAN.R-project.org/package=survival. Accessed May 25, 2022
34. Konopka T. umap: Uniform Manifold Approximation and Projection [Internet]. 2022. Available at: https://CRAN.R-project.org/package=umap. Accessed May 26, 2022
35. McInnes L, Healy J, Melville J, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Internet]*. arXiv. 2020. Report No.: arXiv:1802.03426. Available at: http://arxiv.org/abs/1802.03426. Accessed May 26, 2022
36. Tran PMH, Tran LKH, Nechtman J, et al. Comparative analysis of transcriptomic profile, histology, and IDH mutation for classification of gliomas. Sci Rep 2020;10:20651
37. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059–74
38. Malta TM, de Souza CF, Sabedot TS, et al. Glioma CpG island methylator phenotype (G-CIMP): Biological and clinical implications. Neuro Oncol 2018;20:608–20
39. Meijer DH, Kane MF, Mehta S, et al. Separated at birth? The functional and molecular divergence of OLIG1 and OLIG2. Nat Rev Neurosci 2012;13:819–31
40. Novitch BG, Chen AI, Jessell TM. Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. Neuron 2001;31:773–89
41. Marshall CAG, Novitch BG, Goldman JE. Olig2 directs astrocyte and oligodendrocyte formation in postnatal subventricular zone cells. J Neurosci 2005 Aug 10;25:7289–98
42. Chen Y, Miles DK, Hoang T, et al. The basic helix-loop-helix transcription factor olig2 is critical for reactive astrocyte proliferation after cortical injury. J Neurosci 2008;28:10983–9