

Research Article

# Identification of Potential Biomarkers for Low-Grade Gliomas by Analyzing the Genomics Profiles and CpG Island Methylator Phenotype of Patients in the TCGA Database

Bendahou Mohammed Amine<sup>1\*</sup>, Ibrahimi Azeddine<sup>1</sup>, Boutarbouch Mahjouba<sup>2</sup>

<sup>1</sup>Medical Biotechnology Laboratory (MedBiotech), BioInova Research Center, Medical and Pharmacy School, Mohammed V University, Rabat, Morocco

<sup>2</sup>Department of Neurosurgery, Hospital of Specialties, CHU Ibn Sina, Rabat, Medical and Pharmacy School, Mohammed V University, Rabat, Morocco

**\*Corresponding Author:** Bendahou Mohammed Amine, Medical Biotechnology Laboratory (MedBiotech), BioInova Research Center, Medical and Pharmacy School, Mohammed V University, Rabat, Morocco, Tel: +212 6 15 16 50 22; E-mail: [amine.bendahou@um5s.net.ma](mailto:amine.bendahou@um5s.net.ma)

**Received:** 25 July 2020; **Accepted:** 05 August 2020; **Published:** 08 August 2020

**Citation:** Bendahou Mohammed Amine, Ibrahimi Azeddine, Boutarbouch Mahjouba. Identification of Potential Biomarkers for Low-Grade Gliomas by Analyzing the Genomics Profiles and CpG Island Methylator Phenotype of Patients in the TCGA Database. International Journal of Applied Biology and Pharmaceutical Technology 11 (2020): 188-199.

## Abstract

Low-grade gliomas are the most common primary brain tumors which affects the central nervous system (CNS), they causes considerable morbidity and represents a therapeutic challenge due to the heterogeneity of their clinical behavior. The objective of this study is to reclassify patients in TCGA databases according to the new 2016 WHO classification by identifying their genomics profiles, glioma CpG island methylator phenotype and studying their survival according to the status of the IDH1 gene. Data were obtained for low-grade gliomas patients from The Cancer Genome Atlas (TCGA), a total of 516 samples were analyzed by TCGA Biolinks R Bioconductor. Patients with

astrocytomas had shorter overall survival than those with oligodendrogliomas, they are characterized by mutations in ATRX and TP53 and hypermethylation at the MYT1L and ELAVL4 genes, whereas patients with oligodendrogliomas have mutations in CIC and FUBP1 and hypermethylation at PLCG1 and STAB1 genes. The identification of these biomarkers will improve the diagnosis of low grade gliomas by completing the anatomical pathology tests.

**Keywords:** Gliomas; Astrocytomas; Oligodendrogliomas; Genomics; Epigenetics; Bioinformatics analysis; TCGA Database; IDH; CpG; Methylation

## Introduction

Low-grade gliomas are primary brain tumors of the central nervous system (CNS), they are more common in adults and are characterized by a poor diagnosis. The differentiation between astrocytomas and oligodendrogliomas remains a major challenge in neuro-oncology because they have a similar morphological appearance and distinct biological and therapeutic responses [1].

The histological characteristics are insufficient to discriminate systematically between the two subtypes, that's why the 2016 update of the World Health Organization (WHO) CNS tumor classification represents a change in the diagnosis of brain tumors by integrating molecular and phenotypic characteristics into the classification of tumors and thus reducing the defined sub-groups. Among the genomic and epigenetic alterations associated with low-grade gliomas, the mutation of isocitrate dehydrogenase (IDH) and the glioma CpG island methylator phenotype (G-CIMP) respectively [2].

In this study, we analyzed the genomic profiles and CpG island methylator phenotype of patients downloaded from the Cancer Genome Atlas (TCGA) database. The identification of these biomarkers in our study could provide a new means of accurate diagnosis of low-grade gliomas.

## Materials and Methods

### Data Collection

The use data were downloaded from the TCGA database (<https://gdc-portal.nci.nih.gov>), including 516 LGG (low-grade gliomas) samples. Genomic data (Illumina HiSeq 2000) and methylation data (Illumina Infinium Human Methylation 450 BeadChip) was obtained by the TCGA biolinks R package (version 2.12.6) [5]. After normalization, 501 samples were retained for genomic data and 496 samples for

methylation data. All clinical information about the patients was used for each type of tumor.

### Sample Processing

The ggplot2 package (version 3.2.1) [3] is used to describe the distribution of mutated genes within samples. Based on the new WHO 2016 classification, we reclassified all samples into astrocytomas and oligodendrogliomas using the maftools package (version 3.9) [4], by generating an oncolplot with all patients with their clinical information and the percentages of the most implicated mutated genes. We used the plot maf summary function to display the number of variants of each sample as a stacked bar chart and the types of variants as a boxplot. The titv function classifies the SNPs into transitions and transversions and returns the boxplot summary data indicating the general distribution of six different conversions. Stacked barplot showing fraction of conversions in each sample, and to identify genes that are mutually exclusive or co-occurring, the somatic interactions function that performs the fisher's exact pair test is used to detect such a significant genes pair. This function also uses cometExactTest to identify sets of potentially altered genes.

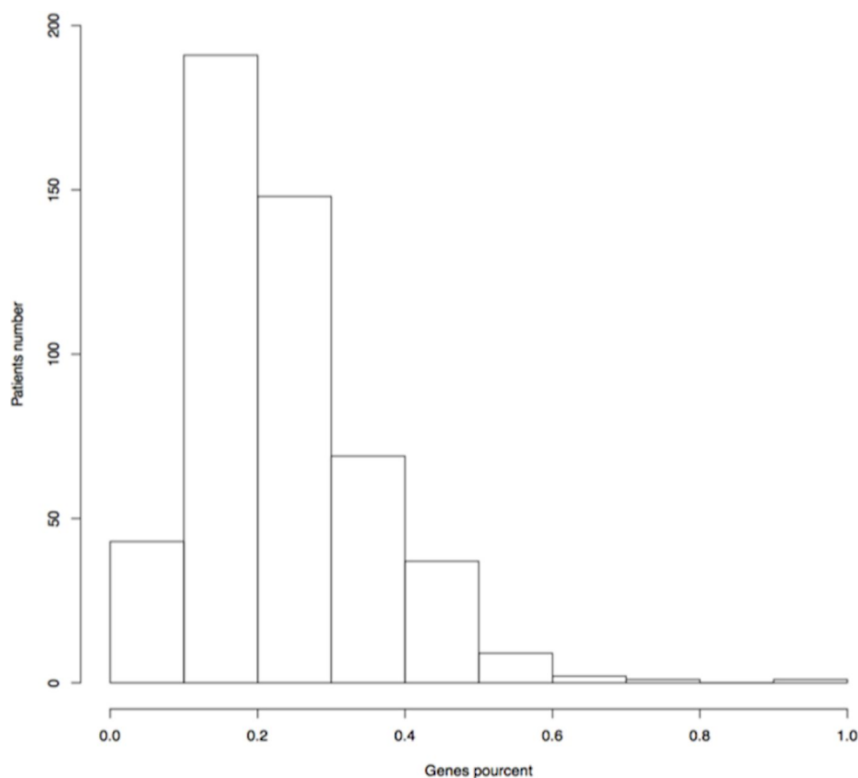
Finally, to estimate the survival function according to clinical data of the patients, we used TCGA analyze Survival KM package [5]. For the methylation data, we used TCGA Workflow Bioconductor (version 1.8.1) [5], we started by checking the average methylation of the DNA of the different groups of samples (astrocytomas and oligodendrogliomas) through the TCGA Visualize\_MeanMethylation function, then the differentially methylated regions were visualized using heat maps by the ComplexHeatmap software (version 2.0.0) [6].

## Results

### Distribution of Genes involved in Low-grade Glioma

In order to show the distribution of genes within our data from the TCGA database, we used the ggplot2 package. The asymmetric forms of kurtosis indicates

that the genes vary from patient to patient and we see that some genes are more present than others (Figure 1).

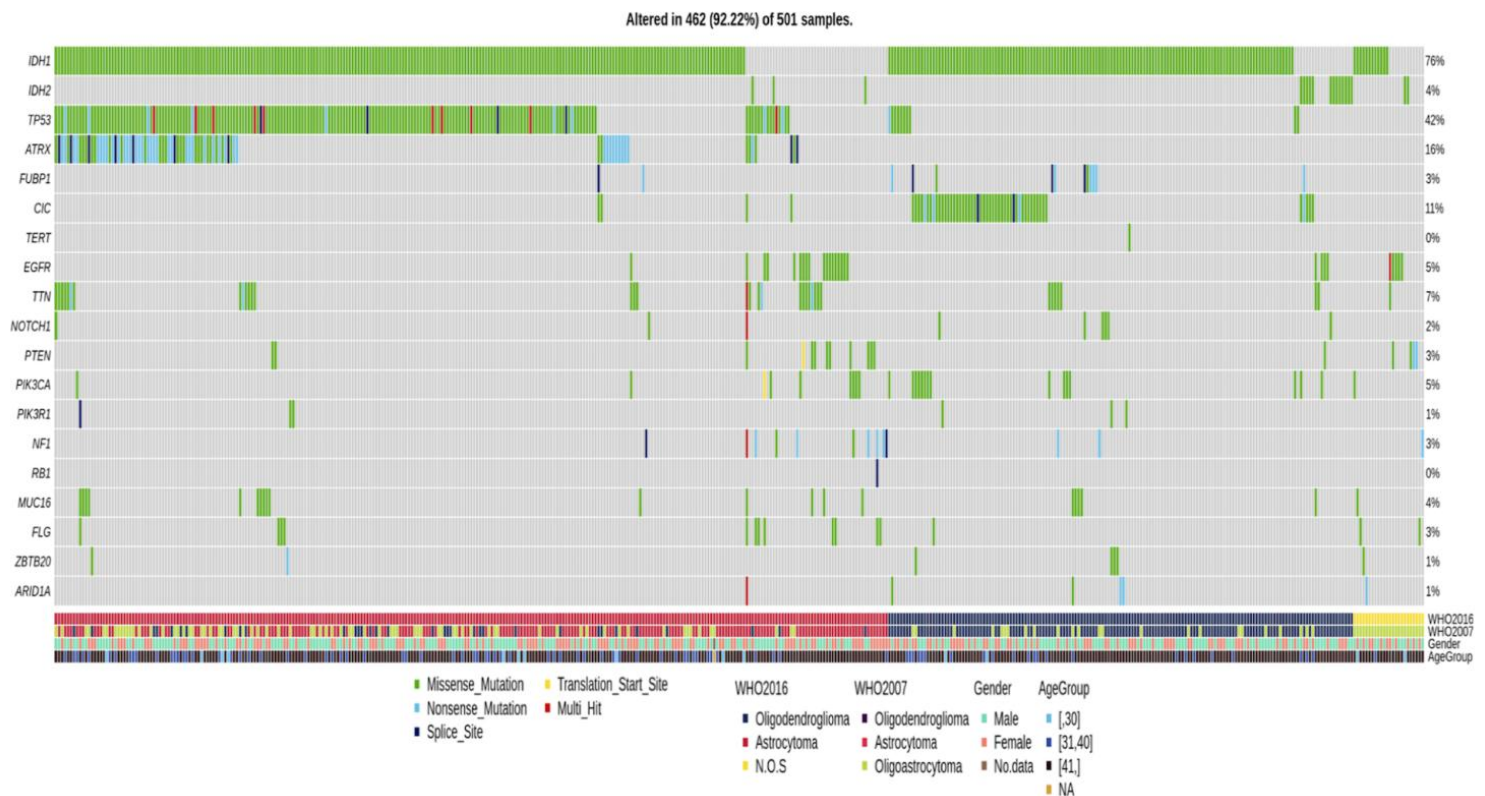


**Figure 1:** Distribution of genes involved in low-grade gliomas

### Mutational Landscape of Somatic Alterations in Low-grade Glioma

To identify significant differences in gene mutations between the two molecular subtypes astrocytomas and oligodendrogliomas, we used the maftools package (Figure 2). For the patients with astrocytomas the most mutated genes found is IDH1 78%, TP53 73%, ATRX 63%, TTN 12%, MUC16 6%, NF1 6% and PTEN 6% (Figure 3A), in the other side, patients with oligodendrogliomas most mutated genes are IDH1 81%, CIC 56%, FUBP1 23%, NOTCH1 12%, PIK3CA 12%, IDH2 9% and PIK3R1 7% (Figure 4A). In a cohort of 501 samples divided into three categories

according to the WHO 2007 classification 186 astrocytomas, 182 oligodendrogliomas and 96 oligoastrocytomas. These patients were reclassified into two categories according to the new 2016 WHO classification, 290 astrocytomas and 174 oligodendrogliomas, and there remain 37 NOS (not otherwise specified) patients whose sequencing must be repeated to have a precise diagnosis. The majority of patients are over 41 years old and the number of men is 157 with astrocytomas and 98 with oligodendrogliomas, for women we have 132 astrocytomas and 76 oligodendrogliomas (Figure 2).

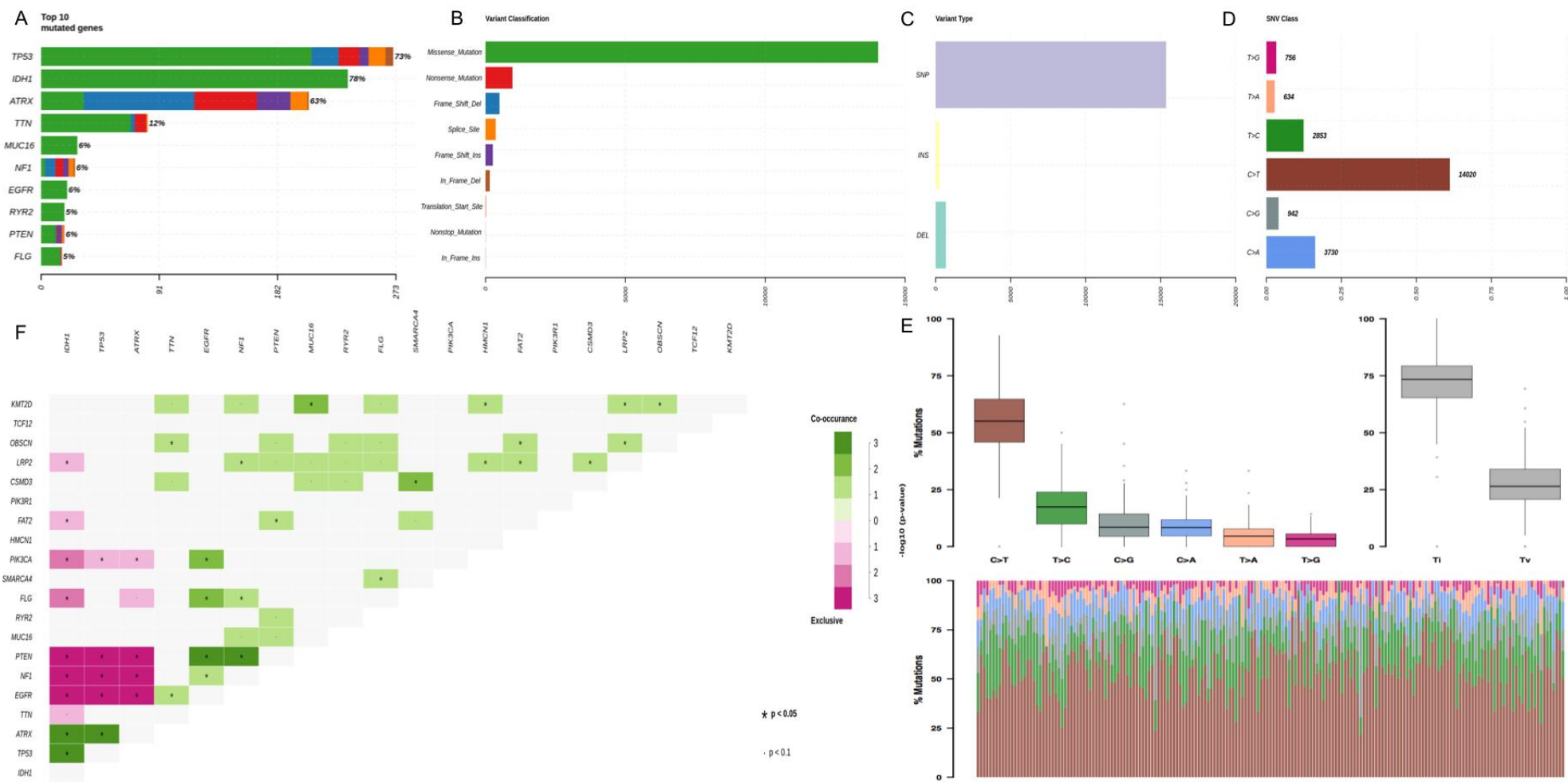


**Figure 2:** Mutational alteration of 501 Samples that have low-grade gliomas

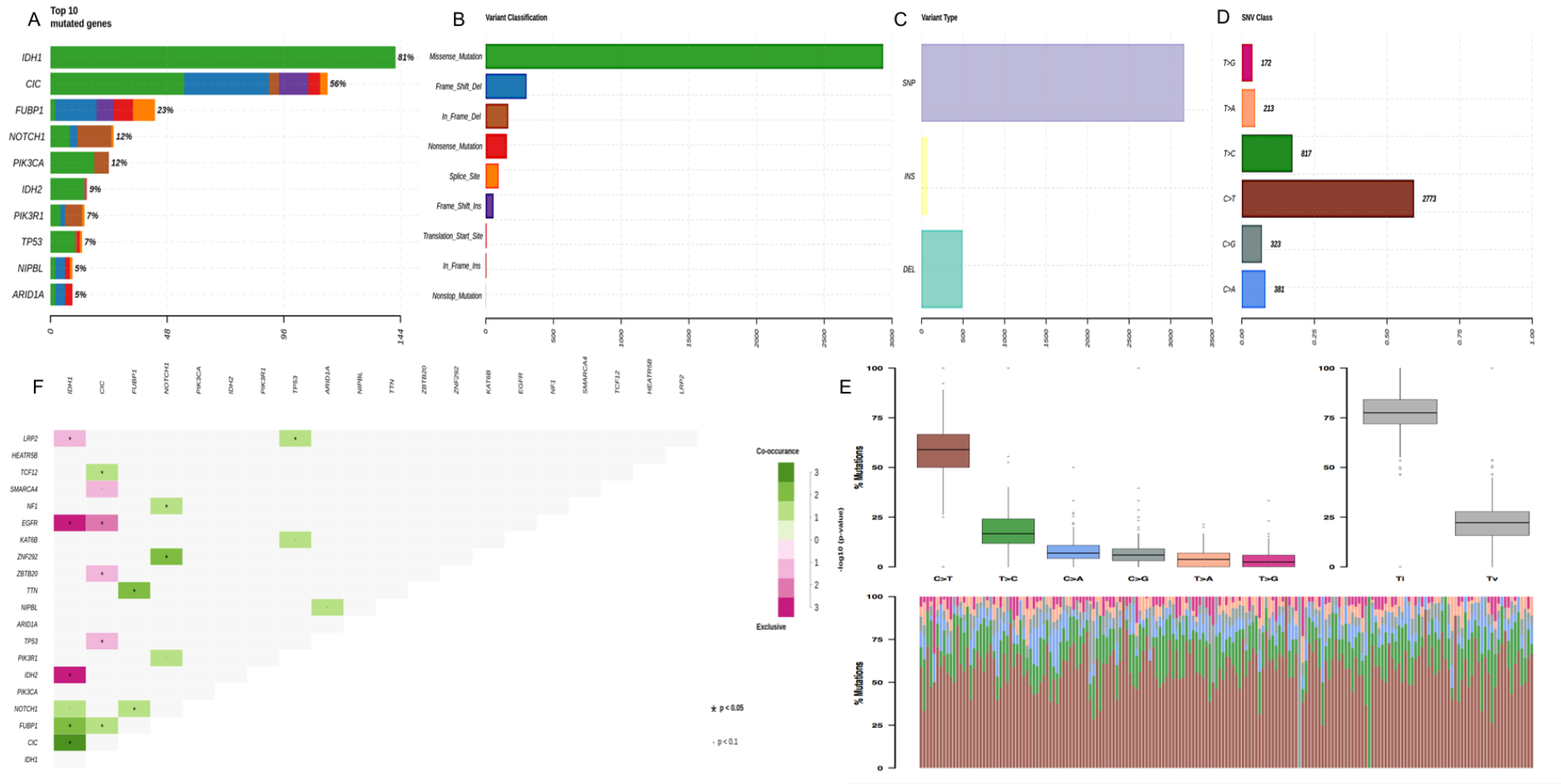
### Variant Classification

In order to establish a classification of the variants, we used the maf summary function of the Maftools package, the two subtypes are characterized by a very high rate of missense mutations represents 13834 mutations for astrocytomas and 2906 mutations for oligodendrogliomas, then we find the nonsense mutation that represents 949 mutations for astrocytomas and 154 mutations for oligodendrogliomas followed by the frame shift del which represents 484 mutations, 301 mutations and the frame shift ins represents 258 mutations, 57 mutations for astrocytomas and oligodendrogliomas respectively. Splice site represents 360 mutations for astrocytomas and 95 mutations for

oligodendrogliomas (Figure 3B and Figure 4B). For the type of variants, the most present is the SNPs followed by the Del and finally the Ins presenting 22631, 792, 321, 4635, 555 and 80 mutations for astrocytomas and oligodendrogliomas respectively (Figure 3C, Figure 4C). The SNV Class (Figure 3D, Figure 4D) and Transition Transversion plot (Figure 3E, Figure 4E) shows that the most common mutation type in both subtypes is transition mutation 74% for astrocytomas (C>T, T>C) and 76% for oligodendrogliomas (C>T, T>C) compared to transverse mutations that present 26% and 24% for astrocytomas (C>A, C>G, T>A and T>G) and oligodendrogliomas (C>A, C>G, T>A and T>G) respectively.



**Figure 3:** The genes involved in astrocytomas (A) barplot shows top 10 mutated who are ordered by their mutation frequency. (B) Variant Classification. (C) Variant Type. (D) SNV Class. (E) Transition and transversion plot displaying distribution of SNVs into six transition and transversion events. Stacked bar plot ( bottom ) shows distribution of mutation spectra for every sample. (F) Mutually exclusive and co-occurring gene pairs displayed as a triangular matrix.



**Figure 4:** The genes involved in oligodendrogliomas (A) barplot shows top 10 mutated who are ordered by their mutation frequency. (B) Variant Classification. (C) Variant Type. (D) SNV Class. (E) Transition and transversion plot displaying distribution of SNVs into six transition and transversion events. Stacked bar plot ( bottom ) shows distribution of mutation spectra for every sample. (F) Mutually exclusive and co-occurring gene pairs displayed as a triangular matrix.



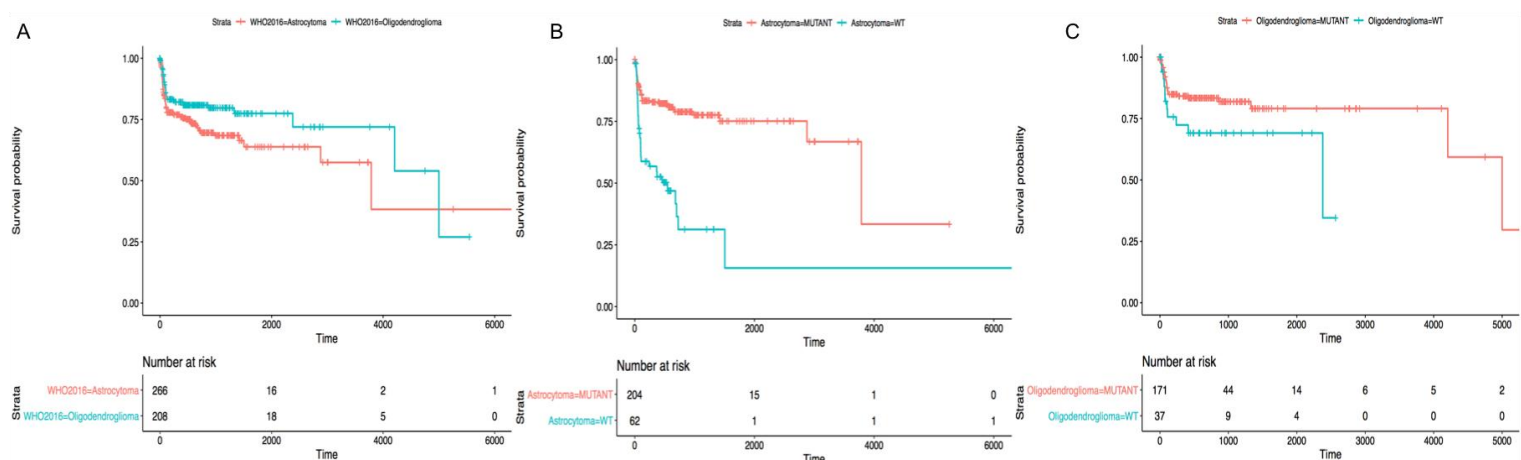
### Genes Interaction

The generation of the matrix (Figure 3F and Figure 4F) which shows the interaction between genes based on the cumulative binomial distribution function, takes into account both the length of the genes and the total number of mutations in the genes identified as potentially significant for the seed to cluster them together. A score is calculated for each cluster, the green indicates a tendency to co-occurrence, the rose has the tendency to exclusivity [7]. In the case of patients with astrocytomas, the genes that have a significant co-occurrence of  $P < 0.05$  are: IDH1-TP53; IDH1-ATRX; ATRX-TP53; PTEN-EGFR; PTEN-NF1 then the exclusive genes that presenting a significance of  $P < 0.05$  are : IDH1-PTEN; IDH1-NF1; TP53-EGFR; ATRX-PTEN; TP53-NF1.

On the other hand, oligodendrogliomas are characterized by other genes, we find that the genes which have a significant co-occurrence of  $P < 0.05$  are: IDH1-CIC; IDH1-FUBP1; CIC-FUBP; NOTCH1-ZNF292, in the other direction the exclusive genes that presenting a significance of  $P < 0.05$  are: IDH1-IDH2; IDH1-EGFR; CIC-EGFR; CIC-TP53.

### Patients Survival Analysis

To estimate patient survival, we used Survival KM package, our analysis of the clinical data showed that people with astrocytomas had shorter overall survival than those with oligodendrogliomas (Figure 5A). Specifically, patients who have an IDH1 mutant have a longer survival than those who have an IDH1 wild type and this is true for astrocytomas and oligodendrogliomas (Figure 5B, C).



**Figure 5:** Kaplan–Meier estimates of overall survival among patients with astrocytomas and oligodendrogliomas that are classified according to IDH mutation and IDH wild-type status.

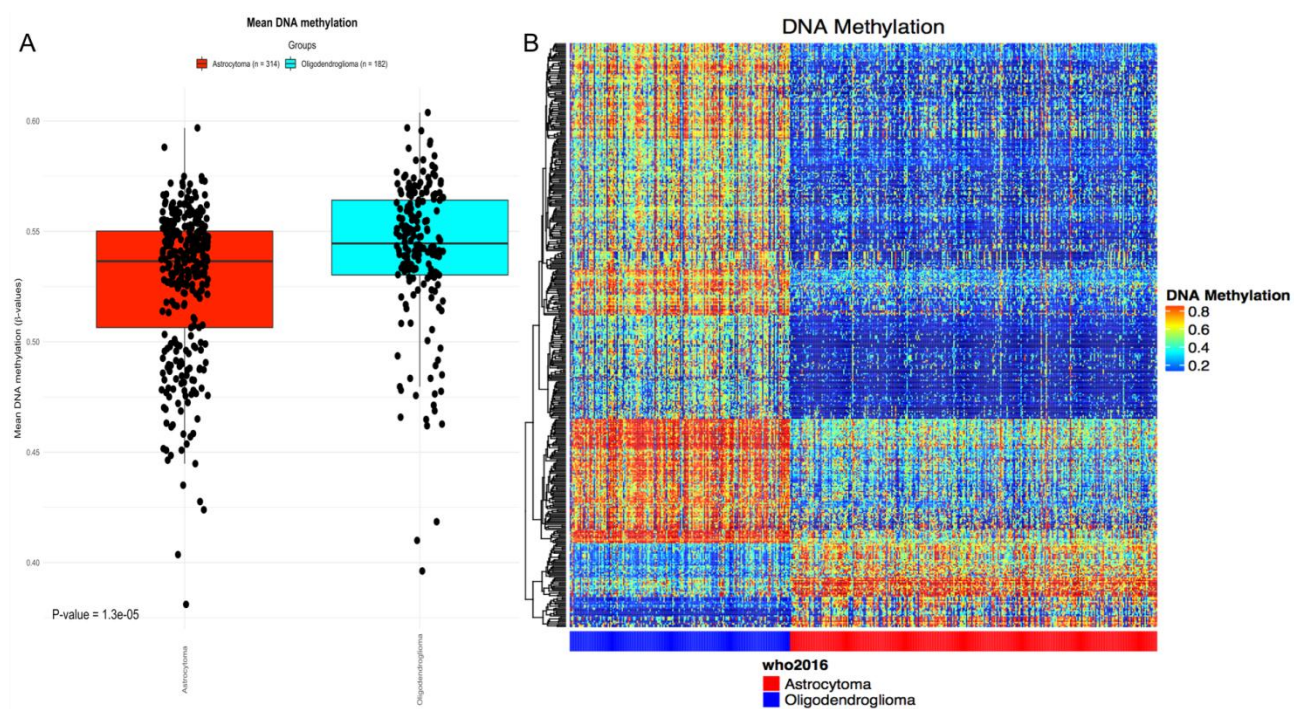
### Glioma CpG Island Methylator Phenotype Visualization

To detect the average methylation of patients, we used the TCGA Workflow Bioconductor, (Figure 6A) shows an average plot of DNA methylation for each sample of the astrocytomas group (314 samples) and an average DNA methylation for each sample of the

oligodendrogliomas group (182 samples). The genome-wide view of the data highlights a difference between the two subtypes found that oligodendrogliomas presents a hypermethylation relative to astrocytomas. To visualize differentially methylated regions, we used Complex Heatmap software, (Figure 6B) shows a heatmap that visualizes the methylation level of the DNA in

all samples. In total we find 361340 CpG locus, for astrocytomas the hypomethylated genes are REST in position cg26694713, FSCN1 in positions cg23080179, cg06123544 and cg04618002, IMPDH1 in position cg15740366, CYHR1 position cg14831162 and ATL3 in positions cg27042081 and cg15814736. On the other hand, patients with oligodendrogliomas present no hypomethylated genes identical to those of patients with astrocytomas, we find three genes coding for proteins FGGY in position cg17393296, RTN4 in position cg14640066 and PDE7B in positions cg08378567 and cg27306443, there are also genes that encode for lincRNAs like AC073283.4 in position cg25653638, LINCO1551 in position cg03505995 and

cg06779110, in the end the gene FOXG1-AS1 in position antisense-type cg00946992 for interacting with mRNA to inhibit the synthesis of the corresponding protein. In the case of hypermethylation, patients with an astrocytomas have three proteins genes encoding MYT1L in position cg20435238, ELAVL4 in position cg26888153 and cg26888153 and PSD3 in position cg22276811 and RP11-1102P16.1 a gene code for a LincRNA in the position cg 08758568. Oligodendrogliomas patients have no hypermethylated genes compared to patients with astrocytomas, PLCG1 gene in position cg03222834 and STAB1 in position cg04654429 and a gene that encodes a U91319.1 LincRNA in the position cg04654429.



**Figure 6:** (A) Plot of DNA methylation of the Astrocytoma and Oligodendroglioma. (B) Heatmap of DNA methylation.

## Discussion

In this report, our analysis of genomic low-grade glioma data from the TCGA database were based on the 2016 WHO classification. All patients who had a 2007 WHO classification were reclassified into two types of tumors (astrocytomas and

oligodendrogliomas), contrariwise the oligoastrocytomas profile was eliminated. Patients who do not present classification biomarkers such as TP53, ATRX, CIC and FUBP1 have been named NOS (not otherwise specified) whose diagnosis is not complete which insists to resequenced to better characterize their genomics



profiles. For IDH1, it is a common gene between the two types of tumors. In the case of a mutation, it catalyzes the conversion of alpha-ketoglutarate to beta-hydroxyglutarate (2-HG) [8,9]. Since the structure of 2-HG is similar to that of  $\alpha$ -KG, 2-HG inhibits a variety of  $\alpha$ -KG-dependent dioxygenases [10,11]. Among them, Ten-Eleven Translocation-2 (TET2) induces global demethylation of DNA by catalyzing the conversion of 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC). Forced mutant IDH1 caused increased 5mC concentrations, instead of decreased 5hmC [12]. IDH mutation also promotes methylation of DNA by TET2 inhibition, resulting in a glioma CpG island methylator phenotype (G-CIMP).

Figuerola ME et al [13] and Turcan S et al [14] demonstrated that the G-CIMP phenotype in glioblastoma is formed after the introduction of an IDH1 mutation in normal human astrocytes, indicating that the mutant IDH induced suppression of TET2, followed by G-CIMP, in cancer cells. Consistent with patients with mutant IDH gliomas, patients with G-CIMP gliomas are younger at diagnosis and survive longer than those without G-CIMP [13,14]. Intriguingly, about 10% of G-CIMP tumors were relapsed as G-CIMP low tumors with poor clinical outcome [15].

For astrocytomas, we find TP53 and ATRX. The wild type TP53 gene has an essential role in several cellular processes, including tumor suppression, apoptosis, DNA repair and autophagy suggesting that the mutation at the TP53 level is defined by a loss of P53 protein function [16].

Inactivation of ATRX in gliomas may be due to mutations, deletions, gene fusions or an amalgam of these causes [17]. For oligodendrogliomas there is the presence of genes CIC and FUBP1, CIC is a

transcriptional repressor that counteracts activation of genes downstream of Receptor Tyrosine Kinase (RTK)-Ras-Extracellular signal-regulated kinase (ERK) signaling [18], and FUBP1 gene has a putative tumor suppressor role in oligodendrogliomagenesis [19]. The prevalence of IDH mutations in lower-grade gliomas prompts the targeting of the mutant enzymes themselves or their downstream metabolic and epigenomic consequences, such as G-CIMP especially since patients with astrocytomas exhibit hypermethylation at the level of genes MYT1L a transcription factor that plays a key role in neuronal differentiation by specifically repressing the expression of non-neuronal genes, Wong KK et al [20], have shown that this gene directly suppress the expression of YAP1, a protein that promotes the proliferation and growth of glioblastomas, when hypermethylated it loses its functioning [20]. The hypermethylated ELAVL4 gene is also found in astrocytomas, Huang S et al [21] have indicated that during neuronal differentiation, ELAVL2, ELAVL3 and ELAVL4 induced an alternative polyadenylation of HuR (human antigen R) and consequently suppressed its translation, leading to a non-proliferative state. In the glioblastomas scenario, there is a hypothesis that downregulation of ELAVL2, ELAVL3, and ELAVL4 may be preferred to further enhance the translation of HuR, leading to a less differentiated and highly proliferative state of cancer cells [21]. On the other hand, patients with oligodendrogliomas are characterized by the presence of PLCG1 which allows the production of diacylglycerol and inositol 1,4,5-trisphosphate (IP3). Su Y et al [22] revealed that this reaction uses calcium as cofactor and plays an important role in intracellular transduction of tyrosine kinase activators, and also STAB1 which encodes a protein that binds nuclear matrix and scaffold DNAs through a single nuclear

architecture [22]. Cancer Genome Atlas Research Network [23] have shown that the protein recruits chromatin remodeling factors to regulate chromatin structure and gene expression, hypermethylation of these genes leads to the reduction of this activity [23].

Our integrative gene analysis with G-CIMP status showed that both subtypes of low-grade gliomas have some common point such as IDH1 mutation or EGFR-IDH1 exclusivity, but also points of divergence as some number of genes mutated as ATRX for astrocytomas or CIC for oligodendrogliomas, also at the level of hypermethylated genes that encode proteins such as the MYT1L gene for astrocytomas and the PLCG1 gene for oligodendrogliomas. The integration of clinical data in the analysis for the study of the patient's survival has shown that patients who mutated IDH1 have a longer survival compared to patients with the wild status of the IDH1. This study has some limitations, the data analysis was based on a bioinformatics algorithm using public TCGA data and no further experimental validation was conducted, in contrast many studies have shown that histopathological classification of low-grade gliomas is subject to strong interobserver variation, this provides insights into identifying additional biomarkers that will complement histopathological and neuroimaging diagnostics to develop a treatment strategy for each patient.

## Conclusion

The prevalence of the mutated IDH gene in low-grade gliomas insists on targeting either the mutant enzymes themselves or their downstream metabolic and epigenomic consequences, such as G-CIMP. The ATRX, CIC, FUBP1, TP53, NOTCH1 and EGFR mutations are also involved in gliomagenesis, but their specificity and prevalence

in low-grade gliomas with an IDH mutation argue for further characterization of the associated signaling networks to improve the diagnosis. These biomarkers lend themselves to a classification system based on biology and practice, which must improve the concordance between observers. This type of system also seems likely to reduce the diagnosis of oligoastrocytoma and the confusion that could result and develop new therapeutic strategies in an early stage before the malignant transformation.

## Acknowledgements

This work was carried out under National funding from the Moroccan Ministry of Higher education & Scientific research (PPR program) to AI. This work was also supported by a grant from the NIH for H3Africa BioNet to AI and Institut Research of the Foundation Lalla Salma, and scholarship of excellence from National Center for Scientific and Technical Research in Morocco.

## Author Disclosure Statement

The authors declare they have no competing financial interests.

## References

1. Gittleman H, Sloan AE, Barnholtz-Sloan JS. An independently validated survival nomogram for lower grade glioma. *Neuro-Oncology* (2019).
2. Paul Y, Mondal B, Patil V, et al. DNA methylation signatures for 2016 WHO classification subtypes of diffuse gliomas. *Clinical Epigenetics* 1 (2017): 1-18.
3. H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer New York (2009).
4. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic

- variants in cancer. *Genome Research* 28 (2018): 1747-1756.
5. Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Research* 44 (2016): e71.
6. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32 (2016): 2847-2849.
7. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* 29 (2013): 2238-2244.
8. Gussyatiner O, Hegi ME. Glioma epigenetics: from subclassification to novel treatment options. *In Seminars in Cancer Biology* 51 (2018): pp. 50-58.
9. Bendahou MA, Arrouchi H, Lakhilili W, et al. Computational Analysis of IDH1, IDH2, and TP53 Mutations in Low-Grade Gliomas Including Oligodendrogliomas and Astrocytomas. *Cancer Informatics* 19 (2020): 1176935120915839.
10. Xu W, Yang H, Liu Y, et al. Article Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of  $\alpha$ -Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* (2011): 17-30.
11. Nonoguchi N, Ohta T, Eun J. TERT promoter mutations in primary and secondary glioblastomas. *Acta Neuropathol* (2013): 931-937.
12. Gliomagenesis E, Bardella C, Al-dalahmah O, et al. Expression of Idh1 R132H in the Murine Subventricular Zone Stem Cell Niche Recapitulates Features of Article Expression of Idh1 R132H in the Murine Subventricular Zone Stem Cell Niche Recapitulates Features of Early Gliomagenesis. *Cancer Cell* (2016): 578-594.
13. Figueroa ME, Abdel-wahab O, Lu C, et al. Article Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* (2010): 553-567.
14. Turcan S, Rohle D, Goenka A, et al. IDH1 mutation is sufficient to establish the glioma hypermethylation phenotype. *Nature* (2012): 479-483.
15. Salama SR, Barnholtz-sloan JS, Souza CF De, et al. Resource A Distinct DNA Methylation Shift in a Subset of Glioma CpG Island Methylator Phenotypes during Tumor Recurrence Resource A Distinct DNA Methylation Shift in a Subset of Glioma CpG Island Methylator Phenotypes during Tumor Recurrence. *Cell Reports* (2018): 637-651.
16. Aubrey BJ, Strasser A, Kelly GL. Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harb Perspect Med* (2016): 1-16.
17. Bunda S, Heir P, Metcalf J, et al. CIC protein instability contributes to tumorigenesis in glioblastoma. *Nat Commun* (2019): 661.
18. Carpentier C, Idhah A, Dehais C, et al. SNP Array Analysis Reveals Novel Genomic Abnormalities Including Copy Neutral Loss of Heterozygosity in Anaplastic Oligodendrogliomas. *PLoS One* (2012): e45950.
19. Wang B, Li D, Yao Y, et al. The crucial role of DNA-dependent protein kinase and myelin transcription factor 1-like protein

- in the miR-141 tumor suppressor network. Cell Cycle (2019): 1-17.
20. Wong KK, Rostomily R, Wong STC. Prognostic Gene Discovery in Glioblastoma Patients using Deep Learning. Cancers (Basel) (2019): 53.
21. Huang S, Zhang L, Sung Y, et al. Recurrent CIC Gene Abnormalities in Angiosarcomas: A Molecular Study of 120 Cases With Concurrent Investigation of PLCG1, KDR, MYC, and FLT4 Gene Alterations. Am J Surg Pathol (2016): 645-655.
22. Su Y, Xiong J, Bing Z, et al. Identification of novel human glioblastoma-specific transcripts by serial analysis of gene expression data mining. Cancer Biomark (2013): 367-375.
23. Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. New England Journal of Medicine 372 (2015): 2481-2498.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC-BY\) license 4.0](https://creativecommons.org/licenses/by/4.0/)