

Prognostic and therapeutic value of *LSM5* gene in human brain cancer Glioma: An omics database exploration approach

Md. Asad Ullah^a, Sayka Alam^{a,1}, Maisha Farzana^{b,1}, Abu Tayab Moin^c,
Chowdhury Nusaiba Binte Sayed Prapty^b, Umme Salma Zohora^a,
Mohammad Shahedur Rahman^{a,d,*}

^a Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Dhaka, Bangladesh

^b Biotechnology Program, Department of Mathematics and Natural Sciences, School of Data and Sciences, BRAC University, Dhaka, Bangladesh

^c Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

^d Wazed Miah Science Research Center, Jahangirnagar University, Dhaka, Bangladesh

ARTICLE INFO

Keywords:
Biomarker
Glioma
LSM5
Diagnosis
Treatment

ABSTRACT

This study explored the diagnostic and therapeutic values of the *LSM5* gene in two commonest types of gliomas, i.e., lower-grade glioma (LGG) and glioblastoma multiforme (GBM) by exploiting a bioinformatics-based database mining pipeline. *LSM5* gene was reported to be upregulated at both mRNA and protein levels in both LGG and GBM tissues compared to adjacent normal tissues. Moreover, overexpression of *LSM5* was later observed to be linked with glioma patients' demographic status and advancing cancer stages. Additionally, a differential and distinct methylation pattern of the *LSM5* promoter and coding sequence in glioma tissues was reported. The gene of our interest was also discovered to undergo multiple amplification events (frequency of alteration: 1.1%) in both LGG and GBM patients. Furthermore, *LSM5* overexpression was reported to be correlated with the poor overall and relapse-free survival of glioma patients. Subsequently, the immune phenotype analysis revealed that the expression of *LSM5* correlates with the abundance levels of various immune cells i.e., B cell, T cell, dendritic cell, and immunomodulators, i.e., IL10RB, CD48, CD274 in glioma microenvironment. Finally, functional enrichment analysis of the neighbor genes of *LSM5* showed that most of the genes are involved in different pivotal cellular processes like RNA splicing, DNA mismatch repair, and DNA replication. Overall, this study suggests that *LSM5* and its downstream products are potential candidates for glioma diagnosis and treatment.

1. Introduction

Gliomas, one of the most pervasive primary intracranial tumors, can be developed from both the glial cells and their precursors as well as the neural precursors or stem cells [1,2]. These tumors are classified based on their histology, location, differential patterns, and anaplastic features ranging from World Health Organization (WHO) grade I to IV [3]. About 50% of brain tumors in children develop from the region below the tentorium cerebelli, but in adults, a great majority of brain tumors originate from the cerebral hemispheres [4]. Malignant gliomas are caused by abnormalities in chromosomes and in growth-factor signaling pathways. On the basis of underlying genetic distinctions, glioblastomas (GBM; grade IV gliomas) can be categorized into primary and secondary subtypes [5]. The amplification of the epidermal growth factor receptor

(EGFR) is restricted to primary glioblastoma, whereas *TP53* deletion has been found to be the molecular characteristic of secondary glioblastoma [6]. On the other hand, individuals with pediatric neurofibromatosis type 1 (NF-1) are more likely to develop lower-grade gliomas (LGG) which involve mainly grade I tumors [7]. Among different types of gliomas, GBM and LGG are the most commonest and lethal forms of gliomas that are responsible for a significant health burden worldwide.

Cellular components such as ion channels, protease, neurotransmitters, and extracellular matrix (ECM) are needed for glioma cell penetration, whereas, the motility is associated with Ca²⁺ signaling, voltage-gated Ca²⁺ channels, inositol 1,4,5- triphosphate receptors (IP3Rs), and transient receptor potential channels [8]. ECM is one of the components of the tumor microenvironment (TME). TME plays a vital role in the formation of malignant tumors because, in a healthy state it protects

* Corresponding author. Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Dhaka, Bangladesh.
E-mail address: rahmanms@juniv.edu (M.S. Rahman).

¹ These authors contributed equally.

against tumorigenesis, and when not, will lead to the formation of tumors [9].

Due to the diffusivity of brain tumors, the effects are more devastating than strokes or other types of brain damage. The severity of the disease varies depending on the type, grade, location, or duration of the disease. It is also associated with other health conditions like depression, anxiety, and insomnia. Overall, gliomas may affect the cognitive functioning of the affected patients [10]. The prognosis in cases of high-grade glioma is still poor, despite the progress in conventional therapies. The mobility of tumor cells is a major factor in the ineffectiveness of treatments such as the removal of tumors via surgical means. Radiation therapy is also affected by the tumor's hypoxic nature. Furthermore, poor prognosis is associated with a high recurrence rate of glioma that halts the way to mitigating the disease. All of the aforementioned reasons call for novel therapeutic approaches and the establishment of new prognostic markers [11,12].

The Smith-like (Lsm) proteins alongside Sm proteins, encoded by the respective homologs of U6 small nuclear RNA and mRNA degradation-associated genes, belong to a large family of biologically important proteins involved in RNA processing inside the human body [13,14]. The *LSM* family consists of 13 different members: *LSM1-LSM12*, *LSM14A*, and *LSM14B*. While this family of proteins is involved in controlling a variety of cellular functions, they are primarily essential for the normal processing and stability of ribosomal RNAs [15]. The ring

LSM1-7 ring resides in the cytoplasm, and the *LSM2-8* ring can be seen in the nucleus. It is evident that *LSM1* and *LSM8* are responsible for specificity, as *LSM2-7* are present in both cytoplasm as well as in the nucleus. A study showed that a reduction in the level of *LSM* causes a reduction of U6snRNA. So, the *LSM2-8* complex actually functions in pre-mRNA splicing and helps to stabilize the U6 small nuclear RNA by creating the U6snRNA core [15,16]. But the defect in *LSM* genes can lead to the formation of several types of cancer. Overexpression of *LSM1* has been seen in the advancement of malignant phenotypes of mesothelioma, lung cancer, and prostate cancer [17]. Furthermore, *LSM1* overexpression has been demonstrated in 87% of human pancreatic tumor samples [18]. Similarly, elevated expression of *LSM12* in the exosomes of cancer tissues coupled with lower levels of CD4 + T cells in the tumor microenvironment suggests that high levels of *LSM12* might promote the transfer of colorectal cancer [19]. Besides, the upregulation of *LSM12* has also been found to be associated with oral squamous cell carcinoma progression [20]. In addition, *LSM4* knockdown has been found to prevent colony formation, migration, and glycolysis metabolism in ovarian cancer cells which highlights the therapeutic potential of targeting *LSM* protein knockdown [21]. Thus, it may be suggested that altered expression of *LSM5* could be seen in tumors such as glioma which provides a novel prognostic and therapeutic role of the *LSM5* gene in glioma. However, the roles of *LSM* family of genes in glioma development and progression remain unstudied to date, though aberrant RNA



Fig. 1. Strategic flowchart of the overall pipeline followed in the database mining approach in this study for the *LSM5* gene against glioma databases.

processing has been proven to be associated with gliomagenesis [22].

In this study, we studied the differential expression pattern of the *LSM5* gene in glioma patients with specific conditions alongside methylation patterns of promoter and coding sequence in LGG and GBM patients (Fig. 1). Additionally, we investigated the somatic mutation and copy number alteration frequencies in LGG and GBM tissues. We have also analyzed the correlations between *LSM5* expression and survival outcome of glioma patients alongside *LSM5* expression and the abundance levels of immune cells and immunomodulators in the glioma microenvironment. Next, we identified genes that are co-expressed with *LSM5* in glioma tissues along and carried out the functional enrichment analysis on the top positively co-expressed genes. Finally, we investigated the expression pattern and validated the effect of *LSM5* expression on the overall survival of glioma patients from independent datasets. The scientific findings of this study will help assess the diagnostic and therapeutic values of the *LSM5* gene and its downstream products in LGG and GBM and recommend *LSM5* for clinical implications.

2. Materials and methods

2.1. Examining the differential expression pattern of *LSM5* in Glioma tissues at the mRNA and protein levels

In this step, the differential expression pattern of the *LSM5* gene in LGG and GBM tissues at the mRNA level was examined from two different servers i.e., GEPIA 2 (<http://gepia2.cancer-pku.cn/>) and OncoDB server (<http://oncodb.org/>) [1,2]. GEPIA 2 offers a comparative RNA expression data analysis between cancerous and normal tissues on the data collected from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) cohorts [23]. The GEPIA2 platform preprocesses RNA-seq data collected from the GTEx and TCGA cohorts by subjecting the raw data to principal component analysis in order to eliminate the batch variations and outliers, which are then normalized to Transcript Per Million (TPM) units to allow the users to carry out differential gene expression analysis using normal and cancerous samples. Similar to the GEPIA2 server, OncoDB also retrieves the RNA-seq data of normal and cancer tissues from both TCGA and GTEx studies. The raw reads are then aligned to the human genome with human RefSeq transcriptome and STAR protocols, while unaligned reads are eliminated [24]. After that, the raw read counts are normalized and transformed into TPM units to facilitate the comparative expression value representation in the form of boxplots. Moreover, the OncoDB server also utilizes statistical methods to calculate the logarithmic fold change (log2FC) between the test and control variables. After that, we evaluated the expression levels of *LSM5* in 39 different glioma cell lines using the Expression Atlas server (<https://www.ebi.ac.uk/gxa/home>) [25]. Later on, we explored the Tissue and Pathology modules of the Human Protein Atlas server (<https://www.proteinatlas.org/>) to examine the protein level expression of the *LSM5* gene in glioma and adjacent and normal brain tissues by investigating the immunohistochemistry images obtained using the anti-*LSM5* protein antibody [26]. We also assessed the protein level expression of the *LSM5* gene by exploring the Clinical Proteomic Tumor Analysis Consortium (CPTAC) database integrated into the UALCAN server. The UALCAN server collects the proteomic expression data from the CPTAC database and then normalizes the log2 transformed spectral count ratio values to facilitate differential protein expression analysis between cancer and normal samples by performing a student's t-test.

2.2. 2.2. analyzing the *LSM5* overexpression pattern across the demographic and clinical conditions of Glioma patients

We examined the expression pattern of the *LSM5* gene across different demographic and clinical conditions of glioma patients i.e., age, sex, Isocitrate Dehydrogenase (IDH) mutation status, glioma grades, and histological subtypes using the Chinese Glioma Genome

Atlas (CGGA) server (<http://www.cgga.org.cn/>) [27]. The associations were evaluated utilizing the RNA-seq data ($n = 693$) obtained from glioma patients. This online platform includes processed RNA-seq data collected from >700 patients with glioma that were then mapped and counted with RSEM and STAR methods, respectively. Expression values of genes that show an expression level of >0 in at least half of the samples are then normalized into fragments per kilobase million transcriptome fragments (FPKM). The CGGA server then offers the association analysis between different gene/mRNA level features with the cancer clinical manifestation of brain cancer patients by performing an ANOVA. The result was considered significant based on the p-value cutoff of <0.05.

2.3. Determining the *LSM5* promoter and coding sequence methylation pattern in LGG and GBM patients

The promoter and gene body methylation pattern of the *LSM5* gene was first inspected by the OncoDB server. This analysis platform showed the significant methylated probes across the *LSM5* promoter and coding sequence in LGG ($n = 516$) and GBM tissues ($n = 148$) obtained from the Illumina methylation 450k array studies of TCGA cohorts. After that, we inspected the UALCAN database (<http://ualcan.path.uab.edu/>) to extract methylation differences in *LSM5* promoter per different age groups of LGG patients utilizing the TCGA LGG cohort [28]. Additionally, we also examined the methylation difference in *LSM5* promoters based on the sample types of GBM tissues (i.e., normal and cancerous samples) using the TCGA GBM cohort. The methylation difference in the *LSM5* gene coding promoter was determined by performing a student's t-test between the control and test samples in the UALCAN server. Later on, the results were analyzed based on the p-value of the experiment. Lastly, we explored the Gene Set Cancer Analysis (GSCA) server (<http://bioinfo.life.hust.edu.cn/GSCA/>) to investigate the correlation between *LSM5* gene methylation and its mRNA level expression and different survival rates of LGG and GBM patients [29].

2.4. Examining the somatic mutation and copy number alteration frequency in *LSM5* gene in LGG and GBM tissues

At first, we utilized the cBioPortal server (<https://www.cbioportal.org/>) to investigate the somatic mutation and copy number alteration (CNA) frequency in *LSM5* gene by investigating different LGG and GBM cohorts involving whole genome sequence from glioma patients [30]. A total of 8 studies submitted by MSK, TCGA, CPTAC, and others spanning genome sequence data from more than 3970 LGG and GBM patients were selected for the analysis. The analysis parameter values were kept at default during this experiment. Later, the OncoPrint and Cancer Types Summary modules were inspected to examine the frequency of somatic mutations and CNAs in the *LSM5* gene in LGG and GBM tissues. Lastly, we explored the mutation module of the GSCA online-based tool to examine the types of CNAs present in the *LSM5* gene in LGG and GBM patients. We also examined the association between *LSM5* mutations and their mRNA level expression and glioma patients' survival rate from the GSCA server.

2.5. Correlation analysis between *LSM5* expression and Glioma patients' survival conditions

In this step, we first used the PreCog online-based web tool (<https://p-recog.stanford.edu/>) to assess the association between the *LSM5* expression and LGG and GBM patients' overall survival (OS) [31]. Later, the GEPIA2 server was utilized to evaluate the correlation between *LSM5* gene expression and relapse-free survival (RFS) of glioma patients. Both the servers were first mounted to use TCGA LGG and GBM cohorts for analysis. The data were then divided into two groups, high-expression and low-expression, with both lower threshold values of 50% of the mean. Finally, we assessed the hypothesis by performing a

log-rank *t*-test between two groups. Lastly, we also examined the impact of *LSM5* expression on the OS of primary and recurrent glioma patients from the CGGA server.

2.6. Determining the association between *LSM5* expression and the abundance of immune cells and modulators in Glioma patients

We first utilized the TIMER 2.0 server (<http://timer.cistrome.org/>) to examine the association between *LSM5* expression and infiltration levels of different immune cells, i.e., B cell, T cell, dendritic cell in the glioma microenvironment [32]. TIMER 2.0 utilizes immune deconvolution studies on TCGA data to estimate the infiltration levels of different immune cells in cancer tissues and adjacent normal tissues with different integrated algorithms. We used the immune module of the TIMER 2.0 server to assess the association between *LSM5* gene expression and different immune cells i.e., B cell, T cell, and dendritic cell (DC) infiltration levels in LGG and GBM tissues. We later investigated the TISIDB server (<http://cis.hku.hk/TISIDB/>) to examine the association between *LSM5* expression and infiltration levels of different immunomodulators, i.e., stimulators and inhibitors in glioma patients [33]. The TISIDB server determines such association from the samples collected from cancerous and adjacent normal tissues during bulk RNA sequencing. The results were then obtained in the form of a scatter plot and interpreted based on the spearman correlation coefficient and *p*-value of the association.

2.7. Identification of the top Co-expressed genes of *LSM5* in Glioma tissues and their functional enrichment analysis

In this step, we first retrieved all the co-expressed genes of *LSM5* from TCGA LGG and GBM cohorts (Firehose Legacy) found in the cBioPortal server. Later, we subjected the top 100 positively co-expressed genes of *LSM5* from each glioma type to functional enrichment analysis i.e., KEGG (Kyoto Encyclopedia of Genes and Genomes), biological processes, molecular function, and cellular components with the help of the Clusterprofiler package in R studio [34]. Thereafter, we retrieved and visualized the top 10 enriched terms from each enrichment analysis with the help of the ggplot2 package [35]. Later, we also carried out a gene set enrichment analysis (GSEA) on all positively and negatively co-expressed genes of *LSM5* in LGG and GBM tissues using the GSEA tool to understand the biological roles of the gene cluster better [36].

2.8. Validation of *LSM5* gene expression and its impact on Glioma patients' survival from independent dataset

In this step, we investigated the expression pattern of the *LSM5* gene from two independent microarray datasets i.e., GSE35493 and GSE50161 from the National Center for Biotechnology Information-Gene Expression Omnibus (NCBI-GEO) database [37–39]. GSE35493 contains the total mRNA expression data from the pediatric brain cancer samples and adjacent normal brain tissue samples (*n* = 71). We only selected the glioma (*n* = 12) and normal brain samples (*n* = 9) for our analysis. Moreover, GSE50161 also incorporates the total mRNA expression profile of glioma (*n* = 34) and adjacent normal brain tissue samples (*n* = 13). The expression data were normalized (RMA) and log2 transformed in R studio using the Bioconductor package [40]. The log2FC between the test and control variables was calculated using the LIMMA package [41]. We also evaluated the impact of *LSM5* gene expression on the OS of glioma patients utilizing the clinical data (*n* = 74) of glioma patients from the GSE4412 dataset [42]. To validate the association, the clinical data were retrieved from the PrognScan server and a long-rank *t*-test was performed between the higher and lower *LSM5*-expressing glioma patients using the survival and survminer packages in R studio [43–45].

3. Results

3.1. Expression profile of *LSM5* in Glioma and adjacent normal tissues at the mRNA and protein level

We evaluated the mRNA level expression of *LSM5* in glioma and adjacent normal brain tissues with the help of two different servers, i.e., GEPIA 2 and OncoDB. The analysis report from GEPIA 2 suggested that the *LSM5* gene is highly expressed in LGG tissues (*n* = 518) than in adjacent normal brain tissues (*n* = 207) (*p* < 0.01) (Fig. 2a). In terms of GBM tissues, the cancerous samples (*n* = 163) again showed a higher level of expression when compared to the normal tissues (*n* = 207) (*p* < 0.01). The results obtained from the OncoDB server also indicated that LGG tissues express more *LSM5* transcripts than the normal brain tissues as observed by a log2FC of 0.8 in cancerous tissues (*p* = 2.7e-69) (Fig. 2b). Moreover, *LSM5* was also found to be highly expressed in GBM tissues (log2FC = 1.52, *p* = 7.3e-59) than in normal tissues from the OncoDB server. Later, we explored the *LSM5* expression pattern in 39 different glioma cell lines and we observed that the expression of our gene of interest ranged from 40 to 100 TPM in all glioma cell lines reflecting the noticeable higher expression of the *LSM5* gene (Fig. 2c). Overall, the GB-1 cell line showed the highest *LSM5* expression level followed by GAMG, LN-229, KNS-42, KNS-81, and so on. After that, we inspected the protein level expression of *LSM5* in glioma tissues in comparison with the corresponding normal tissues by analyzing the IHC images of cancerous and normal brain samples reflecting protein level expression of *LSM5* against the administered IHC antibody (HPA019054) from the HPA server (Fig. 3a). The normal tissues (i.e., cerebral cortex, cerebellum) showed no staining, whereas LGG and GBM tissues showed higher staining against the anti-*LSM5* antibody administered during the IHC experiment, suggesting that *LSM5* is highly expressed at the protein levels in glioma tissues compared to the normal tissues. Lastly, we examined the protein level expression of the *LSM5* gene in glioma and normal brain tissues by exploring the CPTAC database. Similar to our previous analysis report, we also found that *LSM5* proteins are significantly and highly expressed in glioma tissues than in normal brain tissues (*p* = 2.60e-02) (Fig. 3b).

3.2. Expression pattern of *LSM5* across various clinical and demographic features of Glioma patients

In this step, we discovered the expression pattern of *LSM5* across different clinical and demographic conditions of glioma patients. We found that *LSM5* expression increases gradually and significantly with glioma grade (from WHO II to WHO IV) (*p* = 4.3e-10) (Fig. 4). Moreover, given the prevalence of isocitrate dehydrogenase (IDH) mutations in glioma patients, we also evaluated the expression pattern of *LSM5* in relation to IDH mutation status. Reportedly, IDH mutant glioma patients showed lower *LSM5* expression levels than wild-type glioma patients (*p* = 3.4e-05). Additionally, *LSM5* mRNA expression was also found to be correlated to the sex of glioma patients. For instance, though the association was not found to be statistically significant, a lower *p*-value of 0.077 indicated that male patients express more *LSM5* mRNAs when compared to female glioma patients (Fig. 4). Moreover, *LSM5* expression was also reported to significantly increase with the advancing age of glioma patients as our analysis suggested that >42 years age groups express more *LSM5* transcripts than those observed for <42 years age groups (*p* = 0.03) (Fig. 4). Lastly, we evaluated the expression pattern of *LSM5* mRNA across different histological subtypes of glioma patients. Surprisingly, there was a significant and gradual increase in *LSM5* mRNA expression with increasing histological subtypes of glioma patients. As such, oligoastrocytoma patients were reported to express the least level of *LSM5* followed by an upward trend of *LSM5* expression along with a few events of the downward trend and finally reaching the maximum level of expression in terms of GBM and rGBM (*p* = 9.3e-08) (Fig. 4).

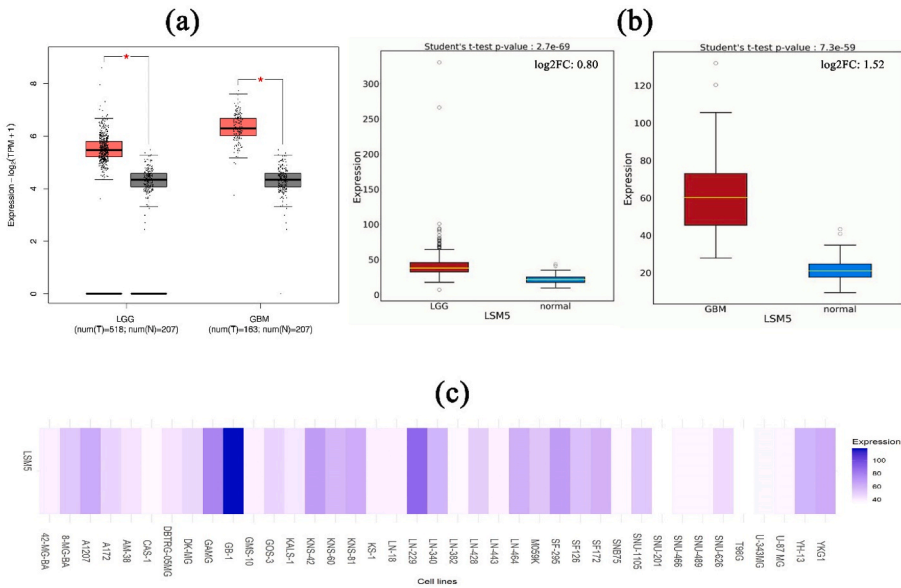


Fig. 2. (a) The expression pattern of *LSM5* mRNAs represented as a boxplot in glioma and adjacent normal tissues (GEPIA 2). (Red box: glioma samples; black box: normal samples). Expression values are presented as size factor adjusted TPM units i.e., log₂(TPM+1) units. (b) The mRNA level expression pattern of *LSM5* is represented as in glioma and adjacent normal tissues (OncoDB). Expression values are presented as the TPM units. (c) The heatmap representation of mRNA level expression pattern of *LSM5* gene (in TPM unit) in different glioma cell lines. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

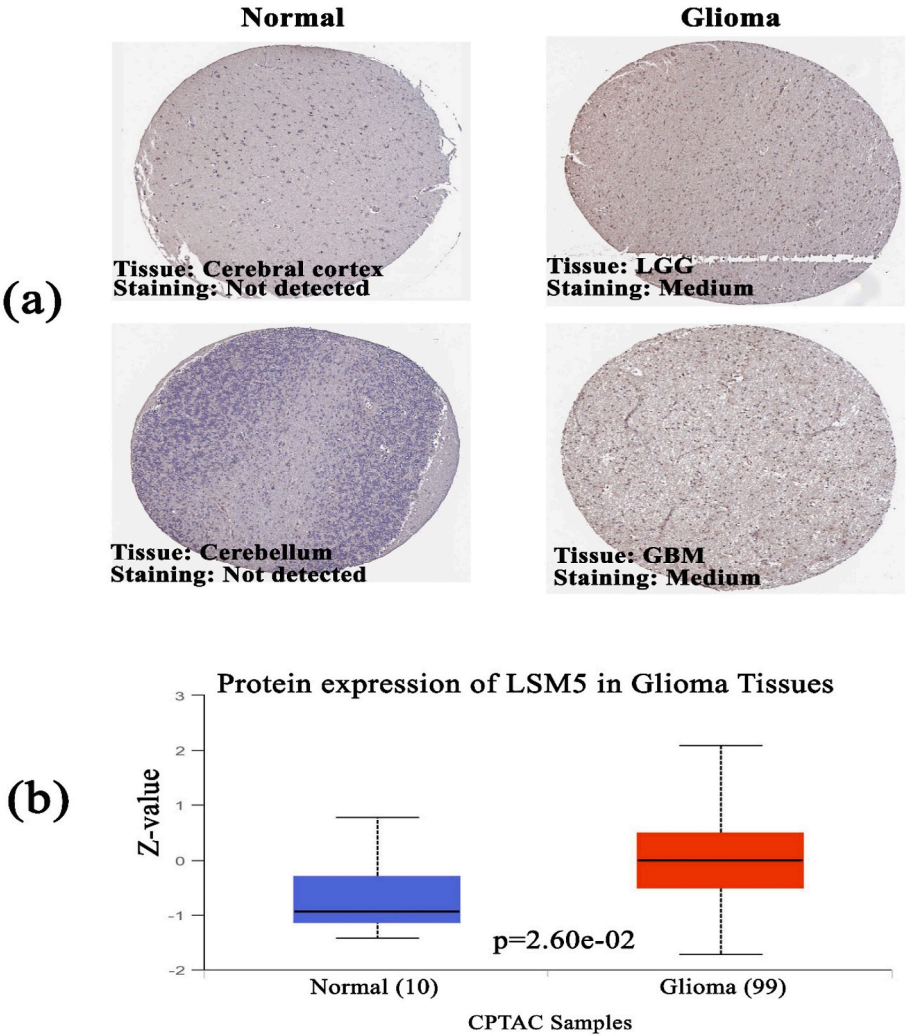


Fig. 3. (a) The IHC images (visualized at 200 μ m) of glioma and adjacent normal tissues representing the protein level expression of *LSM5* against the administered antibody (HPA019054) of IHC staining. (b) The protein level expression pattern of *LSM5* is represented as a boxplot in glioma and adjacent normal tissues from the CPTAC database. Z-values refer to the standard deviations from the median across different clinical samples (i.e., normal and cancer samples).

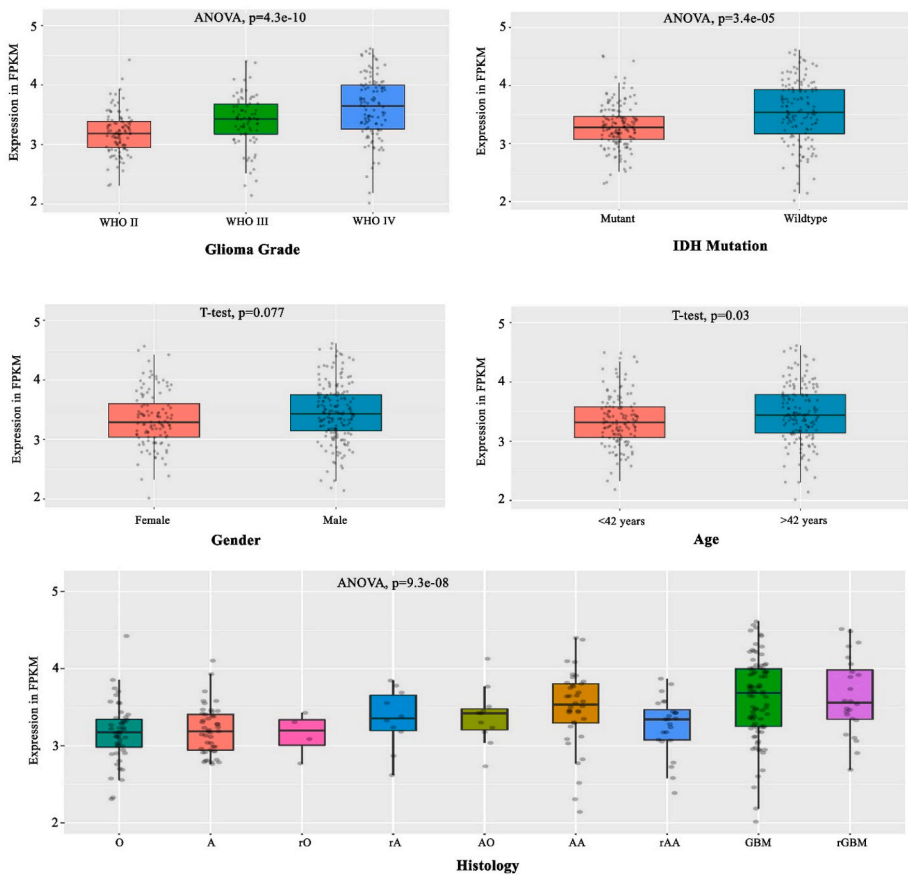


Fig. 4. The mRNA level expression pattern of *LSM5* across glioma grades, IDH mutation status, glioma patient's age, sex, and histological subtypes. *LSM5* expression was found to correlate with the selected demographic and clinical conditions of glioma patients. IDH: Isocitrate dehydrogenase, A: Astrocytoma, O: Oligoastrocytoma, rA: Recurrent Astrocytoma, AO: Anaplastic Oligoastrocytoma, rO: Recurrent Oligoastrocytoma, AA: Anaplastic Astrocytoma, rGBM: Recurrent Glioblastoma, rAA: Recurrent Anaplastic Astrocytoma, GBM: Glioblastoma.

3.3. Methylation pattern of *LSM5* gene in Glioma and adjacent normal brain tissues

The methylation pattern of the *LSM5* gene in glioma and

corresponding normal tissues was analyzed using the OncoDB and UALCAN servers. The analysis report from the OncoDB server suggested that the *LSM5* promoter and gene body contain mostly less methylated landscape in LGG tissues as observed by a higher number of

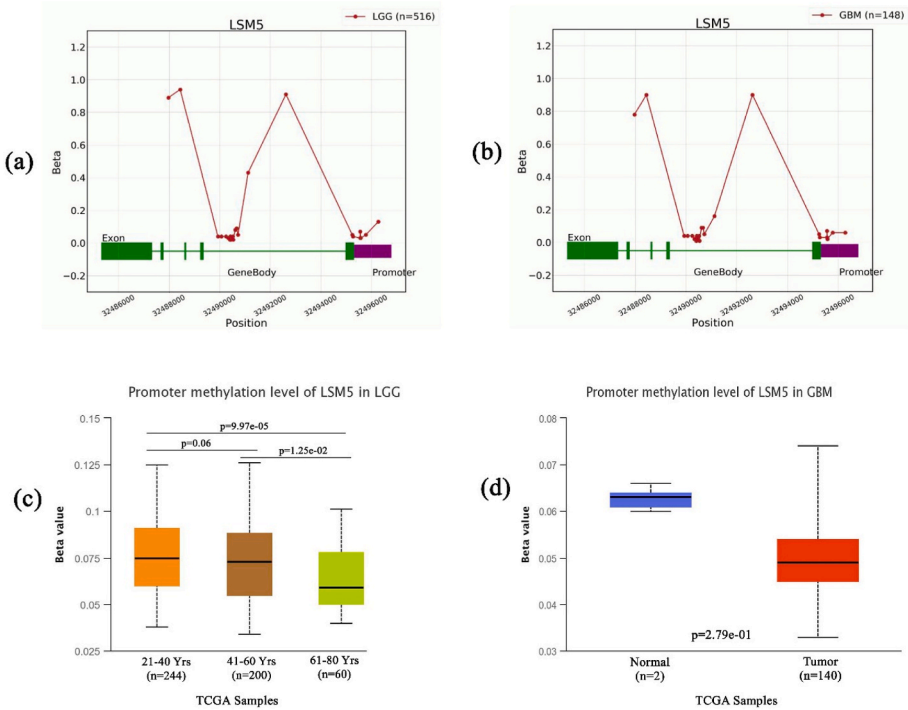


Fig. 5. The promoter and gene body methylation pattern of *LSM5* in (a) LGG and GBM tissues. (Beta value threshold: 0.7 to 0.5: hypermethylation, 0.3–0.25: hypomethylation). Significant methylated probes between normal and glioma tissues were highlighted with red color dot symbols. (c) *LSM5* methylation pattern in LGG patients based on patients' age. (d) Comparative *LSM5* methylation pattern between a normal brain and GBM tissues. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

hypomethylated probes with lower beta values (Fig. 5a). In the case of GBM tissues, similar to the LGG tissues, both the promoters and gene body of *LSM5* were also found to contain multiple hypomethylated regions across its coding sequence (Fig. 5b). However, a slight pattern variation across the orientation of CpG islands in the *LSM5* coding gene was observed between the LGG and GBM tissues. Additionally, a smaller number of methylation probes, specifically toward the downstream region of the 5' end, was found to have slightly hypermethylated regions in both LGG and GBM tissues (Fig. 5). Next, we explored the methylation difference of the *LSM5* gene in accordance with the advancing age of LGG patients from the UALCAN server. Interestingly, the 61–81-year-old group of LGG patients showed a significantly lower level of *LSM5* methylation when compared to the methylation level in the 21–40 years group ($p = 9.97e-05$) (Fig. 5c). Additionally, the most advanced LGG patients (61–81 years) also showed a lower level of *LSM5* methylation in comparison to their immediate predecessors (41–60 years). Thereafter, we examined the promoter methylation status of *LSM5* in GBM tissues based on sample types i.e., normal brain tissue and GBM tissues. Though no significant association was observed, a lower p-value of the analysis indicated that *LSM5* is hypomethylated in GBM tissues compared to the normal brain tissues ($p = 2.79e-01$) (Fig. 5e). Lastly, we inspected the impact of *LSM5* methylation on its mRNA level expression and survival rates in LGG and GBM patients. Unsurprisingly, lower methylation of our gene of interest was positively correlated to the mRNA expression of *LSM5* in LGG and GBM tissues (Supplementary Fig. S1a). Additionally, the lower methylation level of *LSM5* was negatively associated with the overall (OS), disease-specific (DSS), and progression-free (PFS) survival of LGG and GBM patients (Supplementary Fig. S1b).

3.4. Frequency of somatic mutations and copy number alterations in LGG and GBM tissues

The frequency of *LSM5* somatic mutation and CNAs was determined

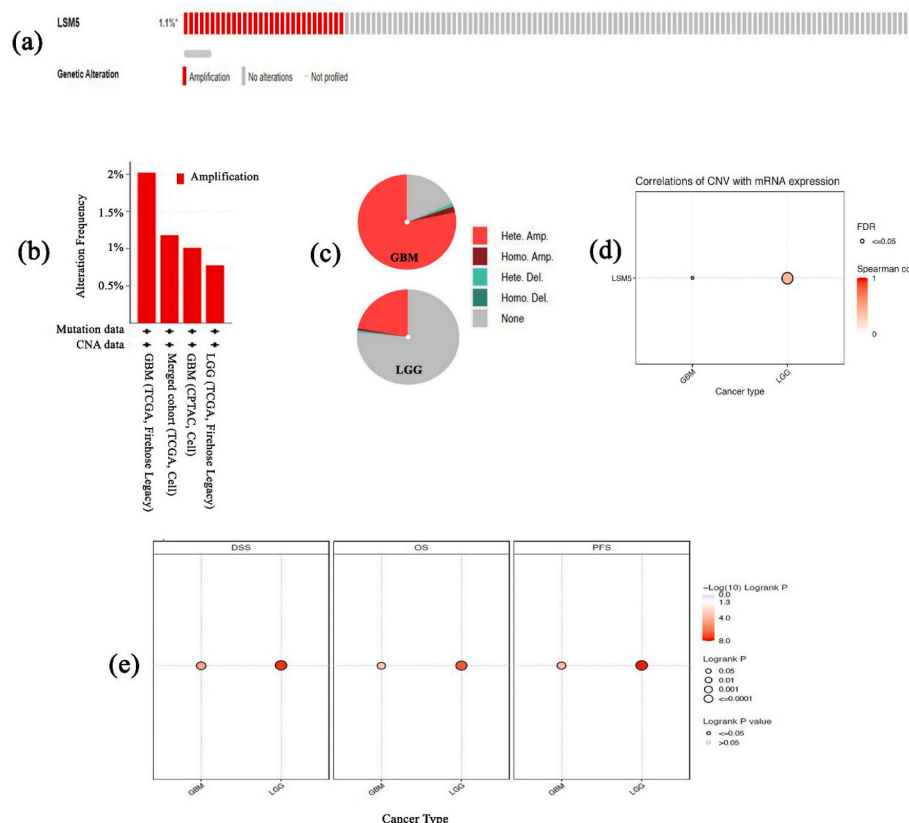


Fig. 6. (a) The distribution of the CNAs present in *LSM5* genes in LGG and GBM patients. The colored column represents the *LSM5* altered patients over the total columns. (b) Bar diagram representation of the frequency of *LSM5* alterations (y-axis) in relation to different LGG and GBM studies (x-axis). (c) Types of different CNAs present in the *LSM5* gene in LGG and GBM tissues. (d) Bubble plot representation of the correlation between *LSM5* alteration and its mRNA expression in LGG and GBM tissues. (e) Bubble plot representation of the correlation between *LSM5* alteration and different survival rates of LGG and GBM patients. PFS: progression-free survival; OS: overall survival; DSS: disease-specific survival.

using the cBioPortal server. The analysis report suggested that the *LSM5* gene underwent amplification events in most of the selected studies with an alteration frequency of 1.1% (Fig. 6a). However, no sign of point mutations, i.e., nonsense, or missense mutations was observed across the *LSM5* coding gene in LGG and GBM tissues. The analysis result on the distribution of *LSM5* CNAs in LGG and GBM tissues reflected that GBM tissue contained most of the amplification events though other cohorts, i.e., LGG and GBM also showed that the *LSM5* underwent amplification events (Fig. 6b). The analysis of the frequency of CNAs of the *LSM5* gene in LGG and GBM tissues from the GSCA server also indicated that GBM tissues contain the greatest number of amplification events compared to LGG tissues (Fig. 6c). Interestingly, a small portion of the genomic studies involving LGG and GBM tissues showed the presence of a heterozygous deletion. Later, we investigated the impact of *LSM5* mutations on the mRNA level expression of *LSM5* in LGG and GBM tissues. *LSM5* CNAs were found to be positively associated with their mRNA level expression in glioma tissues (Fig. 6d). Lastly, the impact of *LSM5* CNAs on different survival rates of LGG and GBM patients was investigated from the GSCA server too. The presence of CNAs in *LSM5* coding genes was found to be negatively associated with the OS, DSS, and PFS of both LGG and GBM patients (Fig. 6e).

3.5. Association between *LSM5* expression and Glioma patients' survival conditions

The survival analysis of cancer patients in relation to specific gene expression is paramount in the discovery of biomarkers. In this step, we utilized two different servers i.e., PreCog and GEPIA 2 to investigate the association between *LSM5* expression and OS and RFS of glioma patients. We found that *LSM5* overexpression was significantly and negatively correlated with the OS of LGG patients (HR:1.87, $p < 0.001$) (Fig. 7a). Moreover, a higher HR of 1.38 signified that *LSM5* overexpression might be a detrimental cause of GBM patients' poor

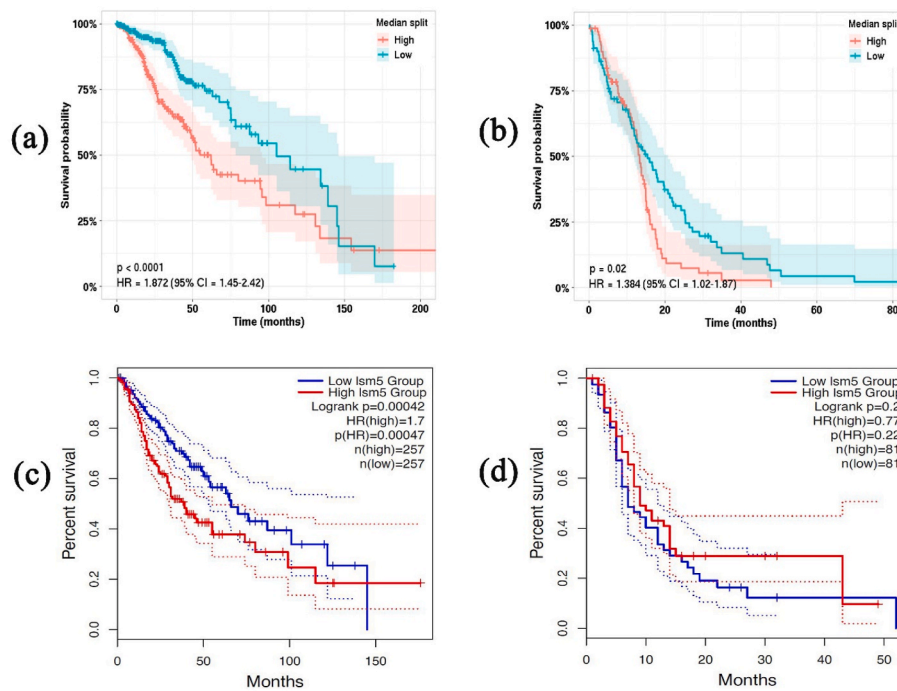


Fig. 7. Kaplan-Meier plot representation of the correlation between *LSM5* mRNA expression and the OS of (a) LGG and (b) GBM patients. Association between *LSM5* expression and the RFS of (c) LGG and (d) GBM patients. The vertical tick mark on the plot represents an event (death). *LSM5* expression was found to be associated with both the OS and RFS of glioma patients.

prognosis ($p = 0.02$) (Fig. 7b). In the case of RFS, *LSM5* overexpression was again found to be responsible for the unfavorable survival in LGG patients (HR:1.70, $p < 0.001$) (Fig. 7c). On the contrary, *LSM5* overexpression was found to be associated with favorable RFS of GBM patients (HR:0.77) (Fig. 7c). However, though the association was not found to be significant ($p = 0.22$), considering the sample number of TCGA GBM ($n > 500$) cohorts, it was observed that *LSM5* may predict the favorable RFS in a considerable number of GBM patients. Lastly, the predictive potential of *LSM5* gene expression in primary and recurrent gliomas was evaluated from the CGGA server. *LSM5* expression was found to be negatively associated with the OS of primary glioma and,

hence, its overexpression was discovered to be able to predict the unfavorable OS of the primary glioma state ($p < 0.001$) (Supplementary Fig. S2). Additionally, *LSM5* overexpression was also found to be negatively associated with the OS of recurrent glioma patients ($p < 0.01$) (Supplementary Fig. S3).

3.6. Association between *LSM5* expression and immune infiltration levels in Glioma microenvironment

The correlation between *LSM5* expression and the abundance of different immune cells in the glioma microenvironment was established

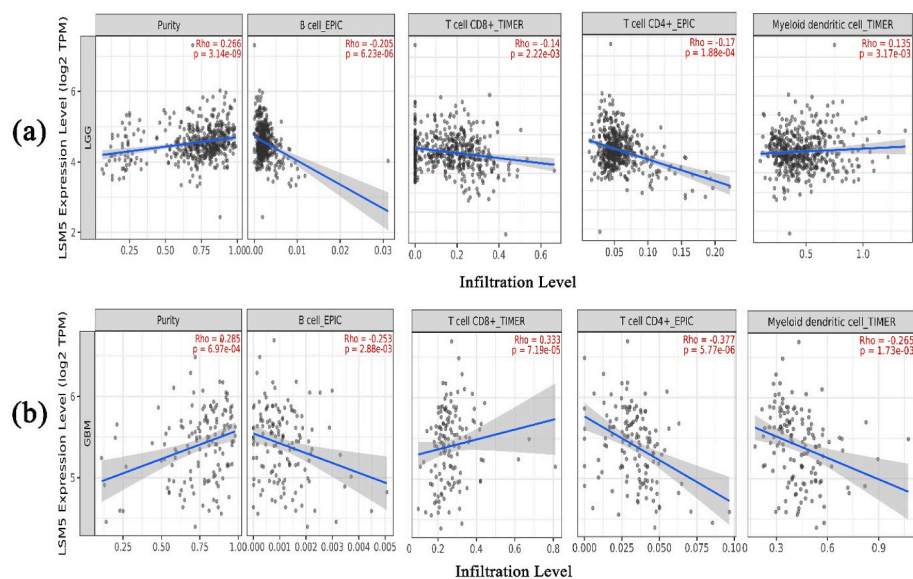


Fig. 8. Scatter plots representing the correlation between the abundance level of different immune cell population and *LSM5* mRNA expression in (a) LGG and (b) GBM patients. A disproportionate level of association between *LSM5* mRNA expression and the abundance of different immune cells was observed.

from the TIMER 2.0 server. *LSM5* expression was found to be negatively associated with the B cell infiltration level in LGG patients (Cor: 0.205, $p = 6.23 \times 10^{-6}$) (Fig. 8a). Additionally, *LSM5* also showed a negative association with the $CD8^+$ (Cor: 0.14, $p = 2.22 \times 10^{-3}$) and $CD4^+$ T (Cor: 0.17, $p = 1.88 \times 10^{-4}$) cell infiltration levels in LGG microenvironment. On the contrary, dendritic cell (DC) infiltration was observed to be positively associated with *LSM5* expression levels in LGG patients (Cor: 0.135, $p = 3.17 \times 10^{-3}$) (Fig. 8a). Similar to LGG patients, *LSM5* expression was observed to be negatively associated with the infiltration levels of B cells in GBM patients (Cor: 0.253, $p = 2.88 \times 10^{-3}$) (Fig. 8b). However, in contrast to LGG patients, *LSM5* expression was discovered to be positively associated with the $CD8^+$ T cell abundance level in the GBM microenvironment (Cor: 0.14, $p = 2.22 \times 10^{-3}$). Again, $CD4^+$ T cell abundance level showed a similar correlation pattern i.e., negative association in GBM patients as that were observed in LGG patients (Cor: 0.377, $p = 5.77 \times 10^{-6}$) (Fig. 8b). On the contrary, *LSM5* expression showed again a different pattern of association with DC in GBM patients compared to LGG patients that is its expression was discovered to have a negative correlation with the DC cell abundance level in GBM patients (Cor: 0.265, $p = 1.73 \times 10^{-3}$) (Fig. 8b). Overall, *LSM5* showed a stronger correlation with most of the immune cells in GBM patients compared to the LGG microenvironment (Cor: $\sim |2.5-3.8|$). Later, we investigated the association between *LSM5* expression and the infiltration levels of different immunomodulators such as immune inhibitors and immunostimulators from the TISIDB server. We observed that *LSM5* expression is positively associated with the infiltration levels of IL10RB and CD48 in both LGG and GBM patients (Cor: >0.19 , $p < 0.01$). On the contrary, CD274 showed a negative association with the expression levels of *LSM5*

in LGG and GBM patients (Cor: <-0.15 , $p < 0.05$) (Supplementary Fig. S3).

3.7. Functional relevance of the Co-expressed genes of *LSM5* in LGG and GBM tissues

The top 100 positive co-expressed genes of *LSM5* from LGG and GBM tissues were utilized in the KEGG pathway and functional enrichment analysis. The KEGG pathway analysis revealed that most of the genes were involved in the function of the spliceosome, DNA replication, nucleotide excision repair, mismatch repair, proteasome activity, homologous recombination, and so on (Fig. 9a). The biological processes analysis on the top co-expressed genes revealed that the genes were predominantly involved in RNA splicing. Moreover, a significant number of genes were also found to be involved in maintaining some other crucial biological processes i.e., DNA replication, DNA strand elongation, RNA localization, maintaining the function of the spliceosome, ribonucleoprotein biogenesis, and so forth (Fig. 9b). The major molecular functions of the top co-expressed genes of *LSM5* in glioma tissues were observed to be the structural constituents of the ribosomes. Additionally, a considerable number of genes were found to be involved in single-stranded DNA binding, ribonucleoprotein complex binding, catalytic activity on RNA, ATP-dependent activity on DNA, damaged DNA binding, telomeric DNA binding, DNA polymerase activity (Fig. 9c). The cellular component analysis on the top neighbor genes of *LSM5* in glioma tissues revealed that most of the gene products were predominantly involved in the spliceosome complex (Fig. 9d). Lastly, all the co-expressed genes of *LSM5* were utilized in the GSEA analysis. The

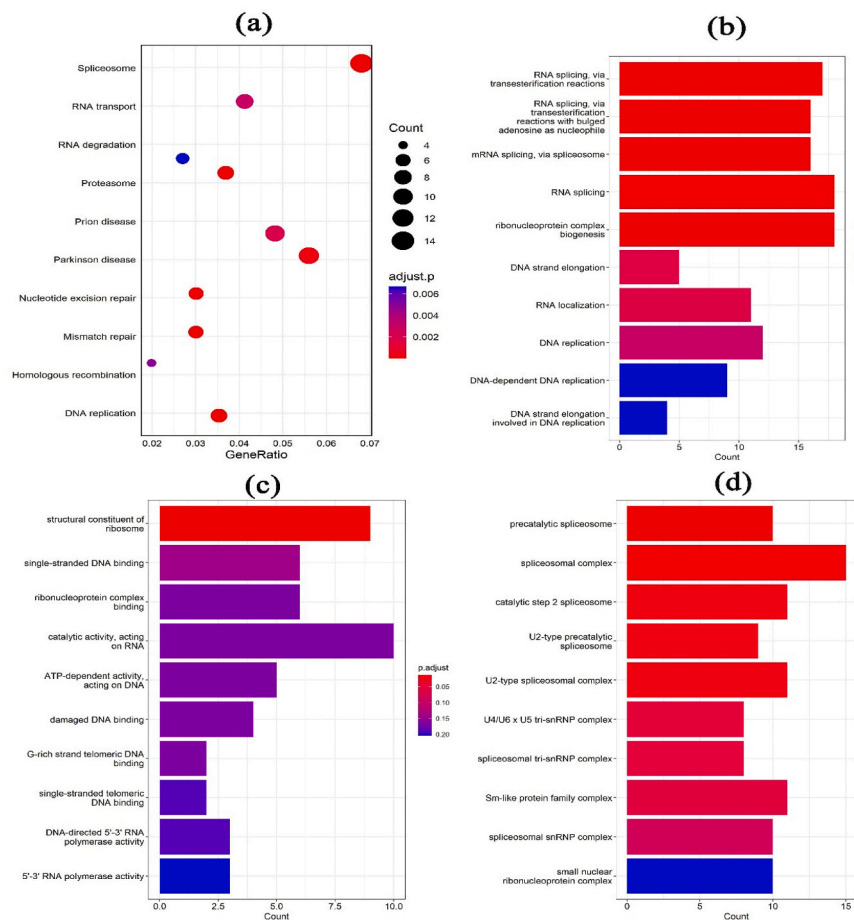


Fig. 9. (a) The bubble plot representation of the top 10 enriched terms from KEGG analysis of the top neighbor genes of *LSM5* in LGG and GBM tissues. The functional enrichment analysis report on the neighbor genes of *LSM5* in LGG and GBM tissues represented in the form of bar plots: (b) biological processes, (c) molecular function, and (d) cellular component analysis.

results indicated that most of the top positively co-expressed genes of *LSM5* in both tissues were majorly involved in crucial biological processes like G to M phase transition of the cell cycle, P53 pathway, unfolded protein response, and so forth (Supplementary Fig. S4). Additionally, the positively co-expressed genes of *LSM5* were also found to receive higher enrichment scores for functions like epithelial to mesenchymal transition, E2F targets, and DNA repair.

3.8. Expression pattern and survival impact of *LSM5* gene on Glioma patients in independent datasets

In this step, we evaluated the expression pattern of the *LSM5* gene in two different publicly available microarray datasets i.e., GSE35493 and GSE50161. The *LSM5* gene was observed to be highly expressed in glioma tissues ($n = 12$) compared to the adjacent normal brain tissues ($n = 9$) from GSE35493 dataset ($\log_2FC: 2.35$, $p = 1.44e-08$) (Fig. 10). Similarly, our gene of interest was also discovered to be highly expressed in glioma samples ($n = 34$) with a $\log_2FC$ of 1.79 than in adjacent normal samples ($n = 13$) from the GSE50161 dataset ($p = 6.45e-09$) (Fig. 10). Later, we also examined the impact of *LSM5* gene expression on the OS of glioma patients from the GSE4412 dataset ($n = 74$). *LSM5* overexpression was found to be significantly and negatively associated with the OS of glioma patients, suggesting its potential to be a prognostic marker for glioma stratification ($p = 0.044$) (Fig. 10).

4. Discussion

This study employed a comprehensive bioinformatics-based omics database exploration strategy to unravel the multi-factorial diagnostic and therapeutic values of the *LSM5* gene in human brain cancer glioma. Differential gene expression analysis can facilitate the optimum understanding of the functional roles of specific genes in healthy cells and oncogenic development [46–48]. Initially, the DEG analysis revealed

that the gene of our interest is significantly upregulated in LGG and GBM tissues compared to the adjacent normal tissues, suggesting its possible underlying mechanism in the oncogenic development of normal brain tissues and subsequent metastasis. Additionally, a disproportionate level of *LSM5* upregulation was observed across different age groups, sex statuses, and histological subtypes of glioma patients. Moreover, the IHC image inspection on the protein level expression of *LSM5* in glioma tissues suggested that the *LSM5* gene is also highly expressed in glioma tissues at the protein level. Therefore, the differential level of *LSM5* transcriptional and translational product expression may be adopted to capture the glioma condition, pathologic stages of glioma, site of metastasis, and metastatic potential of glioma cells. Additionally, *LSM5*-based diagnostic measures may also help determine the response to any *LSM5*-targeted therapeutic approaches for glioma patients.

Aberrant gene methylation pattern is one of the major epigenetic etiological factors for cancer development. Abnormal methylation patterns of different genes can subject healthy cells to tumorigenic transformation by altering the expression levels of the concerning gene [49]. More specifically, promoter hypomethylation is associated with the upregulation of a particular gene in cancer tissues [50]. In this experiment, we recorded that the *LSM5* promoter is hypomethylated in glioma tissues compared to normal tissues. Later, the hypomethylation pattern of the *LSM5* promoter was found to be correlated with its mRNA level overexpression in glioma tissues. Therefore, our analysis suggests that the hypomethylated *LSM5* promoter might be responsible for its overexpression in glioma tissues. Furthermore, the aberrant methylation patterns of different genes can provide a greater sensitive measure for glioma diagnosis and, unlike other pathologic testing options, abnormal DNA methylation is observed at an earlier stage of cancer development [51]. Additionally, tracking the gene methylation pattern from the bodily fluid of patients i.e., serum, can provide an accurate way of glioma stratification in a non-invasive manner [52]. In this study, the *LSM5* gene was found to have a distinct DNA methylation pattern in LGG and

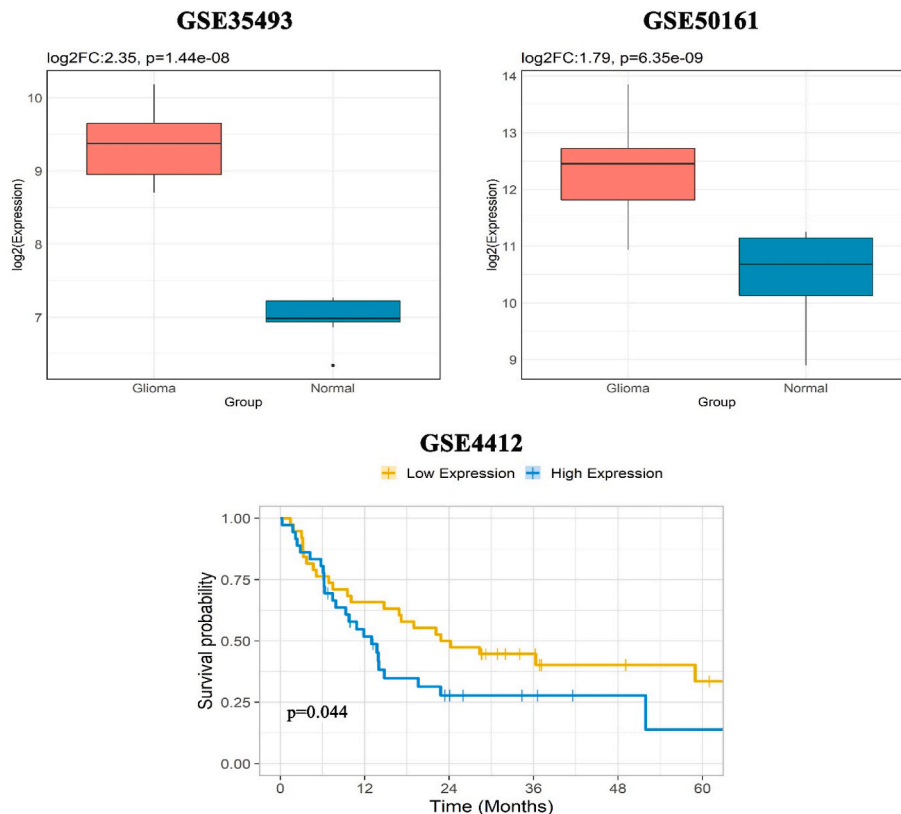


Fig. 10. The expression pattern and survival impact of the *LSM5* gene in glioma patients from independent microarray datasets. The *LSM5* gene was found to be significantly overexpressed ($\log_2FC > 1.75$) in glioma tissues and its overexpression was correlated with the unfavorable survival of glioma patients.

GBM tissues. Moreover, a significant variation in the methylation pattern of the *LSM5* coding gene was recorded among advancing age groups of LGG patients, justifying the potential of *LSM5* methylation status to be an epigenetic diagnostic measure for early-stage and non-invasive diagnosis of glioma patients. What's more, aberrant DNA methylation patterns of different genes can provide an effective therapeutic target for cancer treatment with strategies like regulating DNA methyltransferase activity [53]. In this study, we observed that the lower methylation pattern of the *LSM5* gene is negatively associated with the OS, DSS, and PFS of glioma patients, suggesting its methylation pattern to be a potential candidate for epigenetic-based clinical decisions in glioma patients.

Different somatic mutations in specific genes can drive the oncogenic processes in healthy cells and, unsurprisingly, different CNAs contribute to such deteriorating effects to a greater extent compared to other point mutations [54–56]. Such genetic alterations greatly affect cellular activity by disrupting the normal expression levels of the gene in question and also other genes (i.e., oncogenes, tumor-suppressor genes) within its interaction networks, and thus, propel the oncogenic transformation of healthy cells and subsequent metastasis, ultimately influencing the patients' survival [57]. In this experiment, we observed that *LSM5* gene underwent multiple amplification events in different glioma cohorts. Later, the CNAs present in the *LSM5* gene was found to be positively associated with its mRNA level expression in glioma tissues. We also found a negative correlation between *LSM5* CNAs and the OS, DSS, and PFS of LGG and GBM patients, which signifies that the *LSM5* alterations might further affect the prognosis of glioma patients. From the diagnostic perspective, in vitro studies with clinical specimens suggest that CNAs present in different genes collected from sources like cerebrospinal fluid can predict the condition of glioma patients [58,59]. As a result, the observed CNAs in the *LSM5* gene may further improve the precision of glioma diagnosis, and therefore, the amplification events in *LSM5* genes could be further investigated with studies involving more glioma samples from different demographic regions.

Later, we discovered that *LSM5* overexpression is negatively associated with the OS of LGG and GBM patients and RFS of LGG patients, suggesting that *LSM5* might have an underlying mechanism in the tumorigenic transformation of brain cells and glioma exacerbation. This further justifies the therapeutic potential of the *LSM5* gene in LGG and GBM patients. On the contrary, *LSM5* overexpression was favorable for the RFS of GBM patients which indicates that therapeutic intervention may contribute to improving the GBM prognosis by altering the dosage or activity of *LSM5* though such an assumption requires further laboratory investigation. Moreover, the impact of *LSM5* expression on the survival condition along with its overexpression pattern irrespective of glioma patients' demographic and clinical condition suggests that the downstream products of the *LSM5* gene might help track glioma patients throughout the clinical course, identify high-risk groups of patients, and recommend necessary follow-up measures.

The immune-cell infiltration analysis revealed that *LSM5* expression is disproportionately and strongly correlated to the abundance of different immune cells i.e., B cell, T cell, and DC in LGG and GBM patients. Given the functional roles of immune cells in maintaining immunity against cancer cells after getting produced in higher titer in the cancer microenvironment, these cell populations along with other immune mediators remain promising targets for glioma diagnosis and treatment [60–62]. For instance, high infiltration levels of CD8⁺ T cells in glioma patients predict favorable survival [63]. Therefore, the correlation between the abundance level of different immune cells and *LSM5* in glioma patients may guide the propagation of a dual diagnosis and a combinatorial therapeutic approach along with *LSM5*-targeted ones. Furthermore, we also reported that *LSM5* expression is correlated with the abundance level of different immunomodulators such as CD48, IL10RB, and CD274 in LGG and GBM patients. Previously, IL10RB overproduction was found to be associated with the unfavorable survival of glioma patients [64]. Moreover, a recent study has

systematically grounded that upregulation of CD48 in glioma patients predicts poor survival [65]. Hence, the association between *LSM5* and different immunomodulator expressions in glioma tissues could further assist the *LSM5*-based diagnostic and therapeutic discovery for glioma patients.

After that, the functional enrichment analysis on the top co-expressed genes of *LSM5* in LGG and GBM patients revealed that most of the genes were predominantly involved in RNA processing. In recent years, RNA splicing genes have been reported to act as prognostic indicators for glioma patients [66,67]. Other biological functions of the co-expressed genes of *LSM5* included damaged DNA binding, DNA replication, DNA repair and others which can give rise to the tumorigenic transformation of healthy cells upon deregulation [68,69]. Lastly, the expression pattern and impact of the *LSM5* gene on glioma patients' survival rates were validated using independent glioma microarray datasets i.e., GSE35493, GSE50161, GSE4412. The results obtained from the validation step also supported our initial findings such as *LSM5* is highly expressed and negatively impacts the OS of glioma patients.

In summary, the results obtained from the expression, methylation, and mutation analysis on the *LSM5* gene in glioma patients indicate that the selected gene in this study might have an underlying mechanism in glioma development and progression. Moreover, the overexpression pattern and impact on the survival rates of the *LSM5* gene in glioma patients suggest that the *LSM5* gene and its transcriptional and translational products could be potential targets for glioma diagnosis and treatment. Furthermore, the findings from the immunophenotype and neighbor gene analysis of *LSM5* should further help develop *LSM5*-based diagnostic and therapeutic measures for glioma and improve the precision of glioma diagnosis.

5. Conclusion

This study demonstrated that the *LSM5* gene is differentially expressed in glioma tissues. Moreover, this study also deciphered that the *LSM5* gene is aberrantly methylated in glioma tissues, predisposed to multiple alteration events, and its expression is associated with the prognostic condition of glioma patients. Multiple trajectories of the analytical outcome of this study suggest that *LSM5* and its downstream products could be a potential target for *LSM5*-based glioma diagnosis and treatment. The immune abundance analysis revealed that the infiltration of different immune cells and modulators are associated with *LSM5* expression that could help predict the glioma conditions. The functional enrichment analysis reported that the neighbor genes are involved in crucial biological processes which could drive glioma development and exacerbation upon deregulation. Therefore, the immunophenotype and neighbor genes of *LSM5* in glioma tissues should also be explored when extending further laboratory research targeting *LSM5* for glioma diagnostic measures and therapeutic intervention discovery.

Ethics approval and consent to participate

Not Applicable.

Funding

No specific grant was received for this study.

Author's contribution statement

Conceptualization and experimental design: MU. Experiment analysis and interpretation: MU, SA and AT. Data curation and visualization: MU, MF, AT and CP. Writing, review and editing: MU, SA, MF, AT, CP, US and MR. Investigation and resources: US, MR. Supervision: MR. All authors have read and agreed to the publication of this manuscript.

Consent for publication

Not Applicable.

Availability of data and material

All the data are provided within the manuscript and the supplementary material. The R codes utilized in this study were taken and adjusted from publicly available CRAN projects (<https://cran.r-project.org/>) and Bioconductor resources (<https://www.bioconductor.org/>).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Authors are thankful to the members of Bio-resources Technology and Industrial Biotechnology Laboratory, Jahangirnagar University, Dhaka, Bangladesh and Swift Integrity Computational Lab, Dhaka, Bangladesh for their supports during the preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2022.101114>.

References

- Nutt CL, Matthews RT, Hockfield S. Glial tumor invasion: a role for the upregulation and cleavage of BEHAB/brevican. *Neuroscientist* 2001 Apr;7(2):113–22.
- Chen R, Smith-Cohn M, Cohen AL, Colman H. Glioma subclassifications and their clinical significance. *Neurotherapeutics* 2017 Apr;14(2):284–97.
- Lev MH, Ozsunar Y, Henson JW, Rasheed AA, Barest GD, Harsh GR, Fitzek MM, Chiocea EA, Rabinov JD, Csavoy AN, Rosen BR. Glioma tumor grading and outcome prediction using dynamic spin-echo MR susceptibility mapping compared with conventional contrast-enhanced MR: confounding effect of elevated rCBV of oligodendrogliomas. *Am J Neuroradiol* 2004 Feb 1;25(2):214–21.
- Pollack IF. Brain tumors in children. *N Engl J Med* 1994 Dec 1;331(22):1500–7.
- Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med* 2008 Jul 31;359(5):492–507.
- Maher EA, Brennan C, Wen PY, Durso L, Ligon KL, Richardson A, Khatri D, Feng B, Sinha R, Louis DN, Quackenbush J. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res* 2006 Dec 1;66(23):11502–13.
- Hernández Driever P, von Hornstein S, Pietsch T, Kortmann R, Warmuth-Metz M, Emser A, Gnekow AK. Natural history and management of low-grade glioma in NF-1 children. *J Neuro Oncol* 2010 Nov;100(2):199–207.
- So JS, Kim H, Han KS. Mechanisms of invasion in glioblastoma: extracellular matrix, Ca²⁺ signaling, and glutamate. *Front Cell Neurosci* 2021;190.
- Wang M, Zhao J, Zhang L, Wei F, Lian Y, Wu Y, Gong Z, Zhang S, Zhou J, Cao K, Li X. Role of tumor microenvironment in tumorigenesis. *J Cancer* 2017;8(5):761.
- Lee J, Hall RC. The impact of gliomas on cognition and capacity. *J Am Acad Psychiatry Law* 2019 Aug 1;47(3):350–9.
- Fischer S, Benz J, Späth B, Maier L, Straub J, Granzow M, et al. The archaeal lsm protein binds to small RNAs. *J Biol Chem* 2010;285(45):34429–38.
- Wilusz C, Wilusz J. Eukaryotic lsm proteins: lessons from bacteria. *Nat Struct Mol Biol* 2005;12(12):1031–6.
- MacDonald TJ, Aguilera D, Kramm CM. Treatment of high-grade glioma in children and adolescents. *Neuro Oncol* 2011 Oct 1;13(10):1049–58.
- Rajesh Y, Pal I, Banik P, Chakraborty S, Borkar SA, Dey G, Mukherjee A, Mandal M. Insights into molecular therapy of glioma: current challenges and next generation blueprint. *Acta Pharmacol Sin* 2017 May;38(5):591–613.
- Perea-Resca C, Hernández-Verdeja T, López-Cobollo R, Castellano MD, Salinas J. LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal Arabidopsis development. *Plant Cell* 2012 Dec;24(12):4930–47.
- Mayes AE, Verdine L, Legrain P, Beggs JD. Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *EMBO J* 1999 Aug 2;18(15):4321–31.
- Watson P, Miller S, Fraig M, Cole D, Watson D, Boylan A. CaSm (LSM-1) overexpression in lung cancer and mesothelioma is required for transformed phenotypes. *Am J Respir Cell Mol Biol* 2008;38(6):671–8.
- Little E, Camp E, Wang C, Watson P, Watson D, Cole D. The CaSm (LSM1) oncogene promotes transformation, chemoresistance and metastasis of pancreatic cancer cells. *Oncogenesis* 2016;5(1). e182–e182.
- Wang J, Wu S, Zhang J, Chen J. Potential prognosis and diagnostic value of AKT3, LSM12, MEF2C, and RAB30 in exosomes in colorectal cancer on spark framework. *J Healthcare Eng* 2021;2021:1–14.
- Dong Y, Xue L, Zhang Y, Liu C, Zhang Y, Jiang N, et al. Identification of RNA-splicing factor lsm12 as a novel tumor-associated gene and a potent biomarker in Oral Squamous Cell Carcinoma (OSCC). *J Exp Clin Cancer Res* 2022;41(1).
- Hou W, Zhang Y. Circ 0025033 promotes the progression of ovarian cancer by activating the expression of LSM4 via targeting miR-184. *Pathol Res Pract* 2021;217:153275.
- Ding K, Ji J, Zhang X, Huang B, Chen A, Zhang D, Li X, Wang X, Wang J. RNA splicing factor USP39 promotes glioma progression by inducing TAZ mRNA maturation. *Oncogene* 2019 Sep;38(37):6414–28.
- Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019 Jul 2;47(W1):W556–60.
- Tang G, Cho M, Wang X. OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. *Nucleic Acids Res* 2022 Jan 7;50(D1):D1334–9.
- Papathodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, Burke M, Fillgrave A, Fuentes AM, George N, Huerta L. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* 2018 Jan 4;46(D1):D246–51.
- Pontén F, Jirstrom K, Uhlen M. The human protein atlas—a tool for pathology. *J Pathol: J Pathol Soc G B Irel* 2008 Dec;216(4):387–93.
- Zhao Z, Zhang KN, Wang Q, Li G, Zeng F, Zhang Y, Wu F, Chai R, Wang Z, Zhang C, Zhang W. Chinese Glioma Genome Atlas (CGGA): a comprehensive resource with functional genomic data from Chinese glioma patients. *Dev Reprod Biol* 2021 Feb 1;19(1):1–2.
- Chandrashekar DS, Bashel B, Balasubramanya SA, Creighton CJ, Ponce-Rodriguez I, Chakravarthy BV, Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* 2017 Aug 1;19(8):649–58.
- Liu CJ, Hu FF, Xia MX, Han L, Zhang Q, Guo AY. GSCALite: a web server for gene set cancer analysis. *Bioinformatics* 2018 Nov 1;34(21):3771–2.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013 Apr 2;6(269):pl1.
- Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, Diehn M. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015 Aug;21(8):938–45.
- Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* 2017 Nov 1;77(21):e108–10.
- Ru B, Wong CN, Tong Y, Zhong JY, Zhong SS, Wu WC, Chu KC, Wong CY, Lau CY, Chen I, Chan NW. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics* 2019 Oct 15;35(20):4200–2.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS A J Integr Biol* 2012 May 1;16(5):284–7.
- Wickham H, Chang W, Wickham MH. Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Versions 2016;2(1):1–89.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005 Oct 25;102(43):15545–50.
- Griesinger AM, Birks DK, Donson AM, Amani V, Hoffman LM, Waziri A, Wang M, Handler MH, Foreman NK. Characterization of distinct immunophenotypes across pediatric brain tumor types. *J Immunol* 2013 Nov 1;191(9):4880–8.
- Birks DK, Donson AM, Patel PR, Sufit A, Algar EM, Dunham C, Kleinschmidt-DeMasters BK, Handler MH, Vibhakkar R, Foreman NK. Pediatric rhabdoid tumors of kidney and brain show many differences in gene expression but share dysregulation of cell cycle and epigenetic effector genes. *Pediatr Blood Cancer* 2013 Jul;60(7):1095–102.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2012 Nov 26;41(D1):D991–5.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004 Sep;5(10):1–6.
- Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015 Apr 20;43(7). e47.
- Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liao LM, Mischel PS, Nelson SF. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* 2004 Sep 15;64(18):6503–10.
- Kassambara A, Kosinski M, Biecek P, Fabian S. Package 'survminer'. Drawing Survival Curves using 'ggplot2' (R package version 03.1). 2017 Mar 21.
- Therneau TM, Lumley T. Package 'survival'. R Top Doc 2015 Jul 2;128(10):28–33.
- Mizuno H, Kitada K, Nakai K, Sarai A. Prognoscan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genom* 2009 Dec;2(1):1–1.
- Bashyam MD. Understanding cancer metastasis: an urgent need for using differential gene expression analysis. *Cancer* 2002 Mar 15;94(6):1821–9.

- [47] Ullah MA, Sarkar B, Akter F. Prediction of biomarker signatures and therapeutic agents from blood sample against Pancreatic Ductal Adenocarcinoma (PDAC): a network-based study. *Inform Med Unlocked* 2020 Jan 1;19:100346.
- [48] Ullah MA, Islam NN, Sarkar B, Araf Y, Zohora US, Rahman MS. Bioinformatics analysis on the prognostic significance of SOX9 gene and its transcriptional product in colorectal cancer. *Hum Genet* 2022 May 10:201043.
- [49] Kulis M, Esteller M. DNA methylation and cancer. *Adv Genet* 2010 Jan 1;70:27–56.
- [50] Baylin SB. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* 2005 Dec;2(1):S4–11.
- [51] Dejaegher J, Solie L, Hunin Z, Sciort R, Capper D, Siewert C, Van Cauter S, Wilms G, van Loon J, Ectors N, Fieueus S. DNA methylation based glioblastoma subclassification is related to tumoral T-cell infiltration and patient survival. *Neuro Oncol* 2021 Feb 25;23(2):240–50.
- [52] Sabedot TS, Malta TM, Snyder J, Nelson K, Wells M, DeCarvalho AC, Mukherjee A, Chitale DA, Mosella MS, Sokolov A, Asmaro KP. A serum-based DNA methylation assay provides accurate detection of glioma. *Neuro Oncol* 2021 Sep;23(9):1494–508.
- [53] Esteller M. Relevance of DNA methylation in the management of cancer. *Lancet Oncol* 2003 Jun 1;4(6):351–8.
- [54] Tsukamoto Y, Uchida T, Karnan S, Noguchi T, Nguyen LT, Tanigawa M, Takeuchi I, Matsuura K, Hijiya N, Nakada C, Kishida T. Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer. *J Pathol: J Pathol Soc G B Irel* 2008 Dec;216(4):471–82.
- [55] Jorissen RN, Lipton L, Gibbs P, Chapman M, Desai J, Jones IT, Yeatman TJ, East P, Tomlinson IP, Verspaget HW, Aaltonen LA. DNA copy-number alterations underlie gene expression differences between microsatellite stable and unstable colorectal cancers. *Clin Cancer Res* 2008 Dec 15;14(24):8061–9.
- [56] Yoshimoto T, Matsuura K, Karnan S, Tagawa H, Nakada C, Tanigawa M, Tsukamoto Y, Uchida T, Kashima K, Akizuki S, Takeuchi I. High-resolution analysis of DNA copy number alterations and gene expression in renal clear cell carcinoma. *J Pathol: J Pathol Soc G B Irel* 2007 Dec;213(4):392–401.
- [57] Huang N, Shah PK, Li C. Lessons from a decade of integrating cancer copy number alterations with gene expression profiles. *Briefings Bioinf* 2012 May 1;13(3):305–16.
- [58] Yang RR, Shi ZF, Zhang ZY, Chan AK, Aibaidula A, Wang WW, Kwan JS, Poon WS, Chen H, Li WC, Chung NY. IDH mutant lower grade (WHO Grades II/III) astrocytomas can be stratified for risk by CDKN2A, CDK4 and PDGFRA copy number alterations. *Brain Pathol* 2020 May;30(3):541–53.
- [59] Mouliere F, Mair R, Chandrananda D, Marass F, Smith CG, Su J, Morris J, Watts C, Brindle KM, Rosenfeld N. Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients. *EMBO Mol Med* 2018 Dec;10(12):e9323.
- [60] Hussain SF, Yang D, Suki D, Aldape K, Grimm E, Heimberger AB. The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses. *Neuro Oncol* 2006 Jul 1;8(3):261–79.
- [61] Platten M, Bunse L, Wick W, Bunse T. Concepts in glioma immunotherapy. *Cancer Immunol Immunother* 2016 Oct;65(10):1269–75.
- [62] Fecci PE, Mitchell DA, Whitesides JF, Xie W, Friedman AH, Archer GE, Herndon JE, Bigner DD, Dranoff G, Sampson JH. Increased regulatory T-cell fraction amidst a diminished CD4 compartment explains cellular immune defects in patients with malignant glioma. *Cancer Res* 2006 Mar 15;66(6):3294–302.
- [63] Alexiou GA, Vartholomatos G, Karamoutsios A, Batistatou A, Kyritsis AP, Voulgaris S. Circulating progenitor cells: a comparison of patients with glioblastoma or meningioma. *Acta Neurol Belg* 2013 Mar;113(1):7–11.
- [64] Fan Y, Peng Y, Chen F, Li W. Elevated expression of IL-10RB acts as an independent unfavorable prognostic marker and predicts the polarization of tumor-associated macrophages in glioma.
- [65] Zou C, Zhu C, Guan G, Guo Q, Liu T, Shen S, Yan Z, Xu X, Lin Z, Chen L, Wu A. CD48 is a key molecule of immunomodulation affecting prognosis in glioma. *OncoTargets Ther* 2019;12:4181.
- [66] Chai RC, Li YM, Zhang KN, Chang YZ, Liu YQ, Zhao Z, Wang ZL, Chang YH, Li GZ, Wang KY, Wu F. RNA processing genes characterize RNA splicing and further stratify lower-grade glioma. *JCI insight* 2019 Sep 9;4(17).
- [67] Ding K, Ji J, Zhang X, Huang B, Chen A, Zhang D, Li X, Wang X, Wang J. RNA splicing factor USP39 promotes glioma progression by inducing TAZ mRNA maturation. *Oncogene* 2019 Sep;38(37):6414–28.
- [68] Macheret M, Halazonetis TD. DNA replication stress as a hallmark of cancer. *Annu Rev Pathol* 2015 Jan 1;10(1):425–48.
- [69] Preston BD, Albertson TM, Herr AJ. DNA replication fidelity and cancer. In: *InSeminars in cancer biology*, vol. 20. Academic Press; 2010 Oct 1. p. 281–93. No. 5.