

Identification of aberrantly methylated differentially expressed genes in glioblastoma multiforme and their association with patient survival

MIAO ZHANG¹, XINTONG LV¹, YUANJUN JIANG², GUANG LI¹ and QIAO QIAO¹

Departments of ¹Radiation Oncology and ²Urology, The First Hospital of China Medical University, Shenyang, Liaoning 110001, P.R. China

Received February 17, 2019; Accepted June 6, 2019

DOI: 10.3892/etm.2019.7807

Abstract. Glioblastoma multiforme (GBM) is the most malignant primary tumour type of the central nervous system with limited therapeutic options and poor prognosis, and its pathogenic mechanisms have remained to be fully elucidated. Aberrant DNA methylation is involved in multiple biological processes and may contribute to the occurrence and development of GBM by affecting the expression of certain genes. However, the specific molecular mechanisms remain to be fully elucidated. The present study focused on uncovering differentially expressed genes with altered methylation status in GBM and aimed to discover novel biomarkers for the diagnosis and treatment of GBM. These genes were identified by combined analysis of multiple gene expression and methylation datasets from gene expression omnibus (GSE16011, GSE50161 and GSE 50923) to increase the reliability. In addition, The Cancer Genome Atlas (TCGA) dataset for GBM was used to test the stability of the results. Overall, 251 hypomethylated upregulated genes (Hypo-UGs) and 199 hypermethylated downregulated genes (Hyper-DGs) were identified in the present study. Functional enrichment analysis revealed that the Hypo-UGs are involved in the regulation of immune- and infection-associated signalling, while the Hyper-DGs are involved in the regulation of synaptic transmission. The three hub genes for Hyper-DGs (somatostatin, neuropeptide Y and adenylate cyclase 2) and five hub genes for Hypo-UGs [interleukin-8, matrix metalloproteinase (MMP)9, cyclin-dependent kinase 1, 2'-5'-oligoadenylate synthetase 1, C-X-C motif chemokine ligand 10 and MMP2] were identified by protein-protein interaction network analysis. Among the Hypo-UGs and Hyper-DGs, overexpression of C-type

lectin domain containing 5A, epithelial membrane protein 3, solute carrier family 43 member 3, STEAP3 metalloredutase, tumour necrosis factor α -induced protein 6 and apolipoprotein B mRNA editing enzyme catalytic subunit 3G was significantly associated with poor prognosis in the TCGA and GSE16011 datasets ($P < 0.001$). In conclusion, the present study uncovered numerous novel aberrantly methylated genes and pathways associated with GBM. Methylation-based markers, including the hub genes and prognostic genes identified, may potentially serve as markers for the diagnosis of GBM and targets for its treatment.

Introduction

Glioblastoma multiforme (GBM) is one of the most common, malignant and deadly of the primary brain tumour types, with a typical survival time from identification of ~1 year (1,2). Despite significant efforts to increase the knowledge and improve the treatment regimens for GBM, the prognosis remains unsatisfactory (3,4). Due to the multifactorial and complex pathogenesis of GBM (5), its complex pathogenesis has remained to be fully elucidated. In-depth exploration of the pathogenic mechanisms may provide novel clues for diagnosis and therapy.

Tumour epigenetics is defined as heritable or non-heritable alterations that affect gene expression and genome stability through inappropriate regulation of the local chromatin structure; the mechanisms primarily include DNA methylation, histone acetylation and the action of non-coding RNAs (6). According to certain studies, epigenetic modification may be involved in the earliest phases of tumourigenesis and tumour promotion (7,8). Aberrant DNA methylation, the most widely investigated aspect of cancer epigenetics, mainly includes the gain in methylation of tumour suppressor genes and loss of methylation of oncogenes, which has an important role in the regulation of gene expression and various biological functions (9). Multiple studies have demonstrated that certain genes with altered DNA methylation and gene expression, e.g. the O-6-methylguanine-DNA methyltransferase gene, promote the development of GBM (10,11). However, the comprehensive profiles, pathways and interaction networks associated with these aberrantly methylated genes in GBM remain largely

Correspondence to: Dr Qiao Qiao, Department of Radiation Oncology, The First Hospital of China Medical University, 155 Nanjing Street, Shenyang, Liaoning 110001, P.R. China
E-mail: braveheart8063@outlook.com

Key words: glioblastoma multiforme, gene expression, methylation, enrichment, diagnosis

elusive. In addition, previous studies performed analyses of independent databases containing data from individual investigations, and the false-positive rates make it difficult to identify reliable results. Combined analysis of multiple gene expression and methylation profiles may provide more meaningful and credible results.

Gene expression microarray and transcriptome sequencing have been used to detect thousands of genes simultaneously, and are commonly used to identify differentially expressed genes (DEGs) and gene alterations (12,13). For the gene expression microarray, the detected hybrid signal has a good linear correlation with the abundance of the target sequence. In addition, microarrays have high sensitivity for short sequences, which is suitable for identifying biomarkers for the diagnosis of diseases or associated predictions (14). In addition, the public database Gene Expression Omnibus (GEO) contains extensive genetic information and serves as a repository for microarray data retrieval and deposition. Thus, in the present study, a combined analysis of multiple datasets from the GEO and further databases was performed to increase the reliability of the results.

In the present study, a Bioinformatics analysis was performed to uncover aberrantly methylated DEGs (AMDEGs) in GBM. Based on the combined analysis of multiple microarray datasets in GEO, a number of reliable biomarkers were identified. In addition, the present study focused on revealing the functional terms and pathways through enrichment analysis and aimed to filter hub genes, which may serve as important diagnostic markers. Eventually, several genes with a stable impact on prognosis with a threshold of $P < 0.001$ were screened in order to provide a set of useful therapeutic targets for future research. Finally, the potential genes worthy of further study were validated in another independent database, The Cancer Genome Atlas (TCGA).

Materials and methods

Microarray data retrieval and processing. The keywords 'GBM' and 'gene expression' or 'GBM' and 'methylation' were used to search the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) (15,16). Finally, two gene expression profiling datasets (accession no., GSE16011 and GSE50161) and one methylation profiling dataset (accession no., GSE50923) were selected according to the following inclusion criteria: Gene expression detected using the Affymetrix Human Genome U133 Plus 2.0 Array; methylation expression detected using Illumina Methylation27 platform and sample number of tumours or controls > 5 . Of the dataset GSE16011 deposited by Gravendeel *et al* (17), only 167 samples, including 8 control brain tissues and 159 GBM tissues, were included. Of the dataset GSE50161 deposited by Griesinger *et al* (18), only 47 samples comprising 34 GBM tumours and 13 control brain tissues were incorporated. The dataset GSE50923 deposited by Lai *et al* (19) was based on the platform GPL8490 and included 54 GBM samples and 24 control brain tissues. The details for these datasets are provided in Table I. The online interactive tool GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>), an R-based web tool, was used to filter DEGs or differentially methylated genes (DMGs) by comparing GBM and normal brain tissue samples separately in each

dataset, as reported by Sang *et al* (13). As the threshold for differential expression, $|\text{fold change (FC)}| > 2$ and $P < 0.05$ were set, and $\log_2 \text{FC} > 0.1$ and $P < 0.05$ were used as the criteria for identification of DMGs. The 'MATCH' function was applied to identify overlapping DEGs between GSE16011 and GSE50161. These overlapping DEGs were then intersected with the DMGs to obtain genes affected by altered DNA methylation status. Overall, 450 AMDEGs were identified, of which 199 were hypermethylated and downregulated genes (Hyper-DGs) and 251 were hypomethylated and upregulated genes (Hypo-UGs) in GBM. A flow diagram depicting the study design is provided in Fig. 1.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. The Database for Annotation, Visualization and Integrated Discovery (DAVID; version 6.8; <https://david.ncifcrf.gov/>) (20), a web-based online comprehensive functional annotation tool, was used to perform the GO and KEGG pathway enrichment analyses on the screened Hyper-DGs and Hypo-UGs (Fig. 1; enrichment analysis module). GO functional enrichment analysis was performed in the categories biological process (BP), cellular component (CC) and molecular function (MF). All 199 Hyper-DGs and 251 Hypo-UGs genes were uploaded separately for exploration with the default settings, and only those enriched terms with > 5 gene hits and $P < 0.05$ were regarded as significant.

Generation of protein-protein interaction (PPI) networks and module analysis. To predict the functional and physical protein interactions, the PPI networks of the AMDEGs identified were predicted and constructed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; version 10.5; <http://string-db.org/>) under the default settings (PPI network construction module; Fig. 1), which integrates a variety of predicted and experimentally validated interactions of proteins (21). Nodes lacking a connection in the network were excluded. The PPI networks were then further analysed with the plug-in Molecular Complex Detection (MCODE) clustering tool in Cytoscape software version 3.4.0 (<http://www.cytoscape.org/>) (22) to screen significant modules with default settings. A MCODE score > 4 and number of nodes > 5 were the criteria used to define a significant module. Genes in the module with a degree centrality > 10 were considered hub genes. Module enrichment analysis was then performed with STRING.

Analysis of the effect of the AMDEGs on overall survival (OS). Prognostic assessments for AMDEGs as continuous variables were performed separately for the GSE16011 dataset and the TCGA GBM validation dataset using univariate Cox proportional hazard regression (survival analysis module; Fig. 1). The hazard ratio (HR) was calculated and displayed. $P < 0.001$ was considered to indicate statistical significance. Clinical and gene expression data of GSE16011 used in the present study were obtained from the official GEO website (<https://www.ncbi.nlm.nih.gov/geo/>) and the relevant study (17). Normalized mRNA array data and phenotype files of TCGA data on GBM were downloaded from the University of California Santa Cruz (UCSC) Xena browser (<https://genome-cancer.ucsc.edu>) (23).

Table I. Details of the datasets used in the present study.

Dataset	Sample	Array	Platform	Tumour	Normal	Detail
GSE16011	GBM	mRNA	GPL8542	159	8	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array
GSE50161	GBM	mRNA	GPL570	34	13	Affymetrix Human Genome U133 Plus 2.0 Array
GSE50923	GBM	Methylation	GPL8490	54	24	Illumina HumanMethylation27 BeadChip
TCGA (Oncomine)	GBM	mRNA	AffyU133a	542	10	Affymetrix HT Human Genome U133a
TCGA ^a	GBM	Methylation	Methylation27k	163	140 (GEO) ^b	Illumina Methylation27 platform

^aThe differentially methylated genes of the TCGA GBM cohort were available from the study by Lai *et al* (19). ^bThe methylation values of normal tissues were obtained from the GEO database. GBM, glioblastoma multiforme; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas.

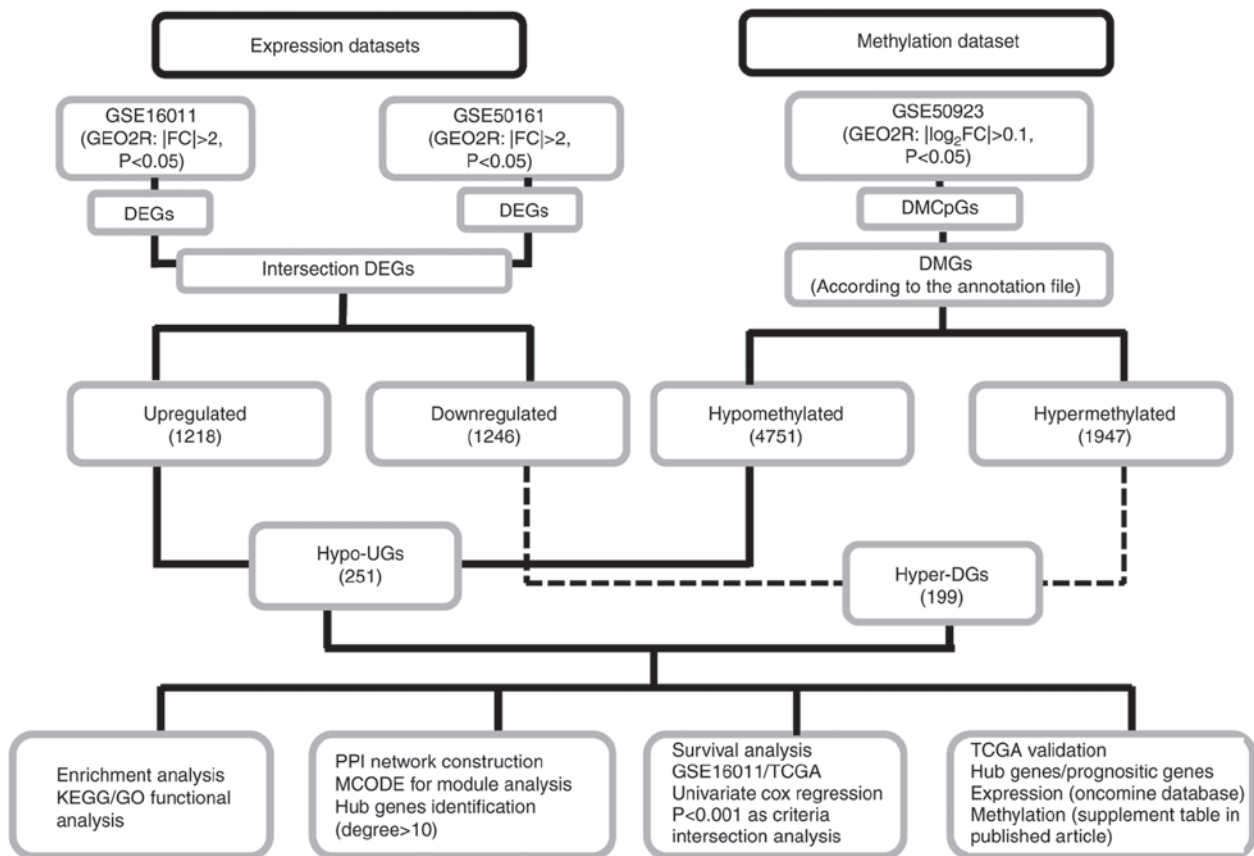


Figure 1. Flow chart depicting the design of the present study. DEGs, differentially expressed genes; DMCpGs, differentially methylated CpG sites; DMGs, differentially methylated genes; Hyper-DGs, hypermethylated and downregulated genes; Hypo-UGs, hypomethylated and upregulated genes; FC, fold change; PPI, protein-protein interaction; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TCGA, The Cancer Genome Atlas.

Samples were divided into two groups based on the median expression value of specific genes. The Kaplan-Meier analysis and the log-rank test were utilized to predict the effect of high or low levels of certain genes on survival. All analyses and plots were performed with the R software (version 3.4.4; <https://www.R-project.org/>) using the packages 'survival' and 'ggplot2'.

Verification of the expression and methylation levels of hub genes and prognostic genes. The expression and methylation levels of hub genes and prognostic genes were then validated in TCGA database to confirm the validity (TCGA validation module; Fig. 1). First, the expression levels of these hub genes and prognosis-associated genes were explored using TCGA GBM U133 array dataset in the Oncomine database

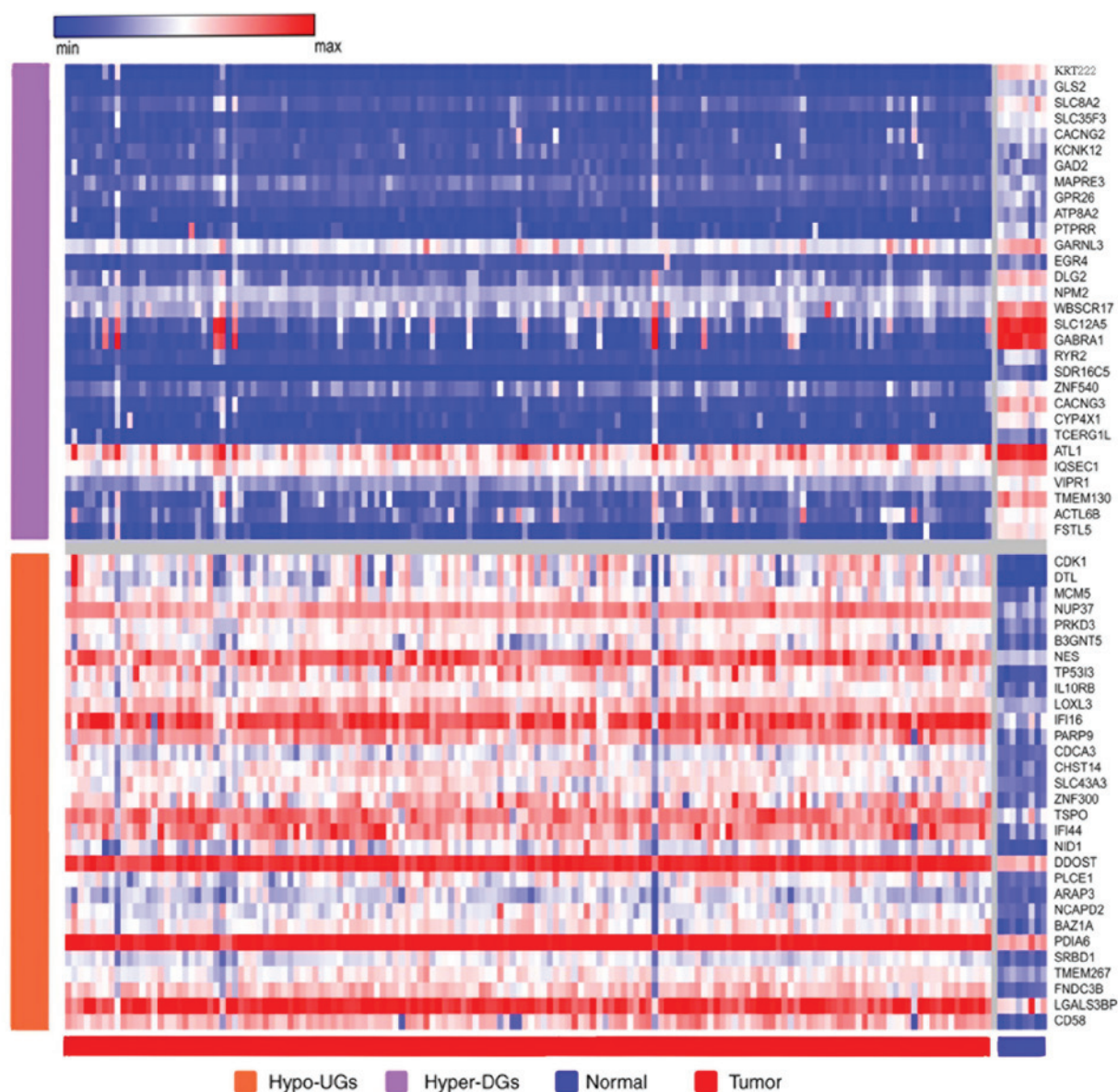


Figure 2. Heat map depicting the expression differences in the top 30 significant Hypo-UGs and Hyper-DGs between normal and tumour tissues in GSE16011. Red represents high expression of genes and blue represents low expression of genes. Each column represents each individual GBM sample. Hyper-DGs, hypermethylated and downregulated genes; Hypo-UGs, hypomethylated and upregulated genes.

(<https://www.oncomine.org>). The DMGs between normal tissue and GBM in the TCGA GBM datasets were obtained from a previous study (19). The median methylation level (β -values) of specific CpG sites in normal brain tissue and tumour tissue, median β differences between normal tissue and tumour tissue, and P-values (computed by a non-parametric Wilcoxon Rank-Sum test) were available from the supplementary materials of the published study (19). The details of the TCGA GBM gene expression dataset and methylation dataset are presented in Table I.

Results

Identification of DMGs and DEGs in GBM. The flowchart depicting the design of the present study is provided in Fig. 1. To characterize the AMDEGs in GBM, two expression profile

datasets (GSE16011 and GSE50161) and one methylation profile (GSE50923) from the GEO database were analysed. A total of 1,218 overlapping upregulated genes (1,574 in GSE16011 and 2,854 in GSE50161) and 1,246 overlapping downregulated genes (1,494 in GSE16011 and 2,518 in GSE50161) were identified. In terms of differentially methylated CpG sites from the methylation array data (GSE50923), 1,947 hypermethylated CpG sites and 4,751 hypomethylated CpG sites were identified. A total of 199 Hyper-DGs and 251 Hypo-UGs were then categorized by comparing the 2,464 DEGs with the 6,698 DECPGs. The representative heatmap depicting the expression differences of AMDEGs between normal tissue and tumour tissue in GSE16011 is provided in Fig. 2. For Hypo-UGs, the expression of GBM was higher than in normal tissues. For Hyper-DGs, the expression of GBM was lower than in normal tissues. The 450 AMDEGs identified are provided in Table SI.

Table II. GO and KEGG pathway enrichment analysis of aberrantly methylated differentially expressed genes.

Category	Term	Count (%)	P-value
GOTERM_BP_FAT	GO:0006955~immune response	82 (0.24)	2.08x10 ⁻²⁷
GOTERM_BP_FAT	GO:0006952~defense response	73 (0.21)	1.25x10 ⁻²¹
GOTERM_BP_FAT	GO:0002682~regulation of immune system process	61 (0.18)	5.18x10 ⁻¹⁶
GOTERM_BP_FAT	GO:0034097~response to cytokine	45 (0.13)	3.18x10 ⁻¹⁵
GOTERM_BP_FAT	GO:0009605~response to external stimulus	74 (0.21)	1.20x10 ⁻¹⁴
GOTERM_MF_FAT	GO:0008009~chemokine activity	8 (0.02)	4.32x10 ⁻⁶
GOTERM_MF_FAT	GO:0032403~protein complex binding	27 (0.08)	1.82x10 ⁻⁵
GOTERM_MF_FAT	GO:0042379~chemokine receptor binding	8 (0.02)	2.37x10 ⁻⁵
GOTERM_MF_FAT	GO:0042605~peptide antigen binding	6 (0.02)	3.50x10 ⁻⁵
GOTERM_MF_FAT	GO:0032395~MHC class II receptor activity	5 (0.01)	4.15x10 ⁻⁵
GOTERM_CC_FAT	GO:0005615~extracellular space	57 (0.17)	2.26x10 ⁻¹¹
GOTERM_CC_FAT	GO:0005576~extracellular region	118 (0.34)	3.30x10 ⁻¹¹
GOTERM_CC_FAT	GO:0044421~extracellular region part	104 (0.30)	9.56x10 ⁻¹¹
GOTERM_CC_FAT	GO:0031012~extracellular matrix	27 (0.08)	1.30x10 ⁻⁷
GOTERM_CC_FAT	GO:0070062~extracellular exosome	73 (0.21)	1.51x10 ⁻⁶
KEGG_PATHWAY	hsa05150:Staphylococcus aureus infection	13 (0.04)	1.50x10 ⁻¹⁰
KEGG_PATHWAY	hsa05323:Rheumatoid arthritis	13 (0.04)	5.32x10 ⁻⁸
KEGG_PATHWAY	hsa05164:Influenza A	16 (0.05)	5.42x10 ⁻⁷
KEGG_PATHWAY	hsa04145:Phagosome	14 (0.04)	3.07x10 ⁻⁶
KEGG_PATHWAY	hsa05152:Tuberculosis	15 (0.04)	3.72x10 ⁻⁶
Category	Term	Count (%)	P-value
GOTERM_BP_FAT	GO:0098916~anterograde trans-synaptic signaling	32 (0.10)	3.62x10 ⁻¹³
GOTERM_BP_FAT	GO:0099536~synaptic signaling	32 (0.10)	3.62x10 ⁻¹³
GOTERM_BP_FAT	GO:0007268~chemical synaptic transmission	32 (0.10)	3.62x10 ⁻¹³
GOTERM_BP_FAT	GO:0099537~trans-synaptic signaling	32 (0.10)	3.62x10 ⁻¹³
GOTERM_BP_FAT	GO:0007399~nervous system development	58 (0.19)	4.32x10 ⁻¹¹
GOTERM_MF_FAT	GO:0022836~gated channel activity	13 (0.04)	2.61x10 ⁻⁴
GOTERM_MF_FAT	GO:0005216~ion channel activity	14 (0.05)	7.37x10 ⁻⁴
GOTERM_MF_FAT	GO:0022838~substrate-specific channel activity	14 (0.05)	1.03x10 ⁻³
GOTERM_MF_FAT	GO:0005200~structural constituent of cytoskeleton	7 (0.02)	1.19x10 ⁻³
GOTERM_MF_FAT	GO:0015267~channel activity	14 (0.05)	1.95x10 ⁻³
GOTERM_CC_FAT	GO:0097458~neuron part	53 (0.17)	1.18x10 ⁻¹⁶
GOTERM_CC_FAT	GO:0043005~neuron projection	43 (0.14)	7.03x10 ⁻¹⁵
GOTERM_CC_FAT	GO:0045202~synapse	33 (0.11)	4.11x10 ⁻¹¹
GOTERM_CC_FAT	GO:0030424~axon	24 (0.08)	2.18x10 ⁻¹⁰
GOTERM_CC_FAT	GO:0044456~synapse part	27 (0.09)	2.87x10 ⁻⁹
KEGG_PATHWAY	hsa04727:GABAergic synapse	7 (0.02)	3.76x10 ⁻⁴
KEGG_PATHWAY	hsa04080:Neuroactive ligand-receptor interaction	11 (0.04)	1.06x10 ⁻³
KEGG_PATHWAY	hsa05032:Morphine addiction	6 (0.02)	3.58x10 ⁻³
KEGG_PATHWAY	hsa04723:Retrograde endocannabinoid signaling	6 (0.02)	5.59x10 ⁻³
KEGG_PATHWAY	hsa04726:Serotonergic synapse	6 (0.02)	8.29x10 ⁻³

If there were >5 terms enriched in the respective category, the top 5 terms were selected according to the P-value. CC, cellular component; BP, biological process; MF, molecular function; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MHC, major histocompatibility complex; hsa, *Homo sapiens*.

GO and KEGG pathway enrichment analysis. The significant KEGG pathways and GO terms in the enrichment analysis performed with DAVID are illustrated in Table II and Fig. 3.

Hyper-DGs were mainly enriched in the functional terms in the category BP of anterograde trans-synaptic signalling, synaptic signalling, chemical synaptic transmission, trans-synaptic

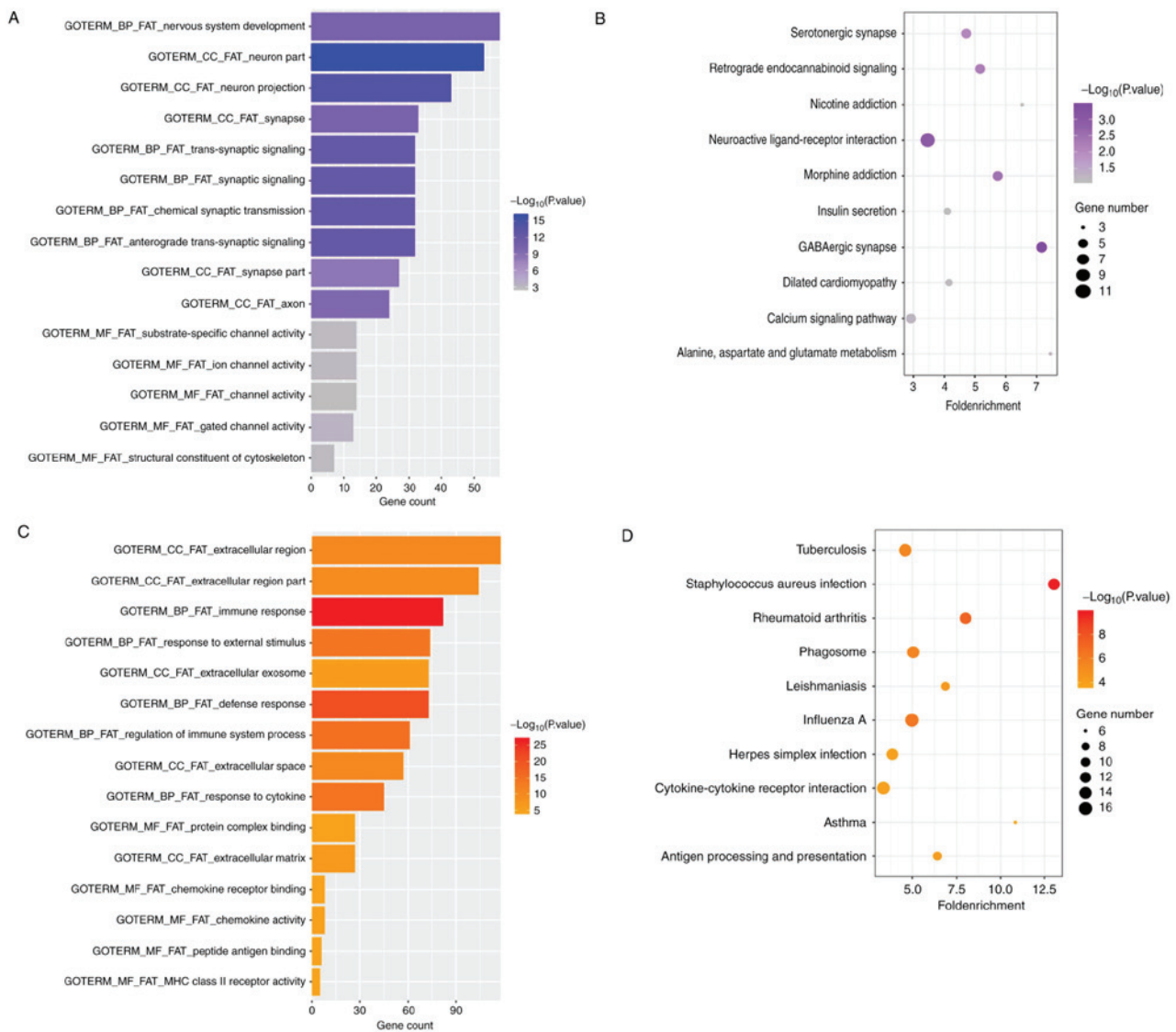


Figure 3. Bar charts and bubble charts for Gene Ontology and KEGG pathway enrichment analysis of aberrantly methylated differentially expressed genes. The top 5 gene ontology enrichment results (A) and top 10 KEGG enrichment results (B) for Hyper-DGs and the top 5 gene ontology enrichment results (C) and top 10 KEGG enrichment results (D) for Hypo-UGs. GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Hyper-DGs, hypermethylated and downregulated genes; Hypo-UGs, hypomethylated and upregulated genes; CC, cellular component; BP, biological process; MF, molecular function.

signalling and nervous system development. In the category MF, these genes exhibited enrichment in the terms gated channel activity, ion channel activity, substrate-specific channel activity, structural constituent of cytoskeleton and channel activity. Furthermore, in the category CC, enrichment was predominantly in the neuron part, neuron projection, synapse, axon and synapse part (Fig. 3A). KEGG pathway analysis of Hyper-DGs indicated that these genes were significantly enriched in GABAergic synapses, neuroactive ligand-receptor interactions, morphine addiction, retrograde endocannabinoid signalling and serotonergic synapses (Fig. 3B).

Among the Hypo-UGs, the most significantly enriched terms in the category BP included immune response, defence response, regulation of immune system process, response to cytokine and response to external stimulus. In the category MF, these genes were accumulated in the terms chemokine activity, protein complex binding, chemokine receptor binding, peptide antigen binding and major histocompatibility complex class II

receptor activity. In addition, in the category CC, the Hypo-UGs were enriched in the terms extracellular space, extracellular region, extracellular matrix and extracellular exosomes (Fig. 3C). These results indicated that Hypo-UGs may have a critical role in the tumour immune microenvironment of GBM. In addition, KEGG pathway analysis indicated enrichment in *Staphylococcus aureus* infection, rheumatoid arthritis, influenza A, phagosome and tuberculosis pathways (Fig. 3D).

PPI network construction, module analysis and selection of hub genes. PPI network construction based on identification of protein functions and interactions may help screen local protein interaction networks with specific functions and selection of hub genes with highly connected nodes or edges. For Hyper-DGs, the PPI network, which includes 94 connected nodes and 148 edges, is presented in Fig. 4A. The PPI network of Hypo-UGs, which included 87 connected nodes and 168 edges, is illustrated in Fig. 4B, The only significant

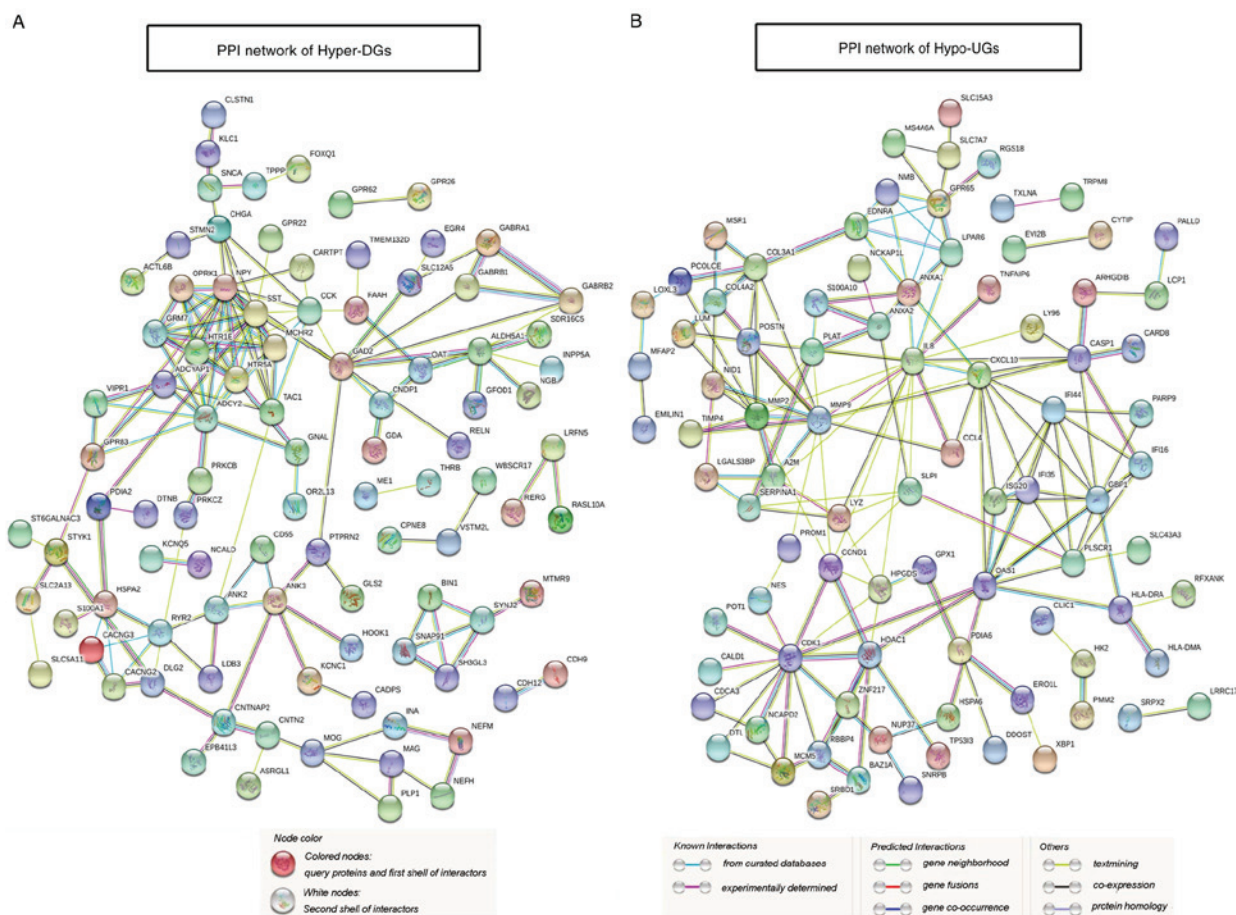


Figure 4. (A and B) PPI network of identified differentially expressed genes with altered methylation status. PPI network of (A) Hyper-DGs and (B) Hypo-UGs. Disconnected nodes were hidden in the network. PPI, protein-protein interaction; Hyper-DGs, hypermethylated and downregulated genes; Hypo-UGs, hypo-methylated and upregulated genes.

module for Hyper-DGs containing 8 nodes and 28 edges, which is linked to G-protein-coupled receptor signalling pathways and neuroactive ligand-receptor interactions, is provided in Fig. 5A. The hub genes of Hyper-DGs identified included somatostatin (SST), neuropeptide Y and adenylate cyclase 2 with 14, 14 and 12 degrees of connectivity, respectively. The most significant module for Hypo-UGs containing 12 nodes and 30 edges is presented in Fig. 5B. Significant vital modules exhibited functions in several signalling pathways, including response to virus, response to cytokine and cytokine-mediated signalling pathways. The significant hub genes of Hypo-UGs mainly included interleukin (IL)-8, matrix metalloproteinase (MMP)-9, cyclin-dependent kinase (CDK) 1, 2'-5'-oligoadenylate synthetase 1 (OAS1), C-X-C motif chemokine ligand 10 and MMP2 with 16, 14, 13, 11, 11, 11 and 10 degrees of connectivity, respectively. These screened hub genes and modules may have important roles in the occurrence and development of GBM.

Survival analysis of AMDEGs. As presented in Table III, univariate Cox regression analysis revealed that several genes were significantly associated with OS in the GBM cohort from the GSE16011 dataset. Comparison of these results with those for the GBM cohort from TCGA revealed that C-type lectin domain containing 5A (CLEC5A), epithelial

membrane protein 3 (EMP3), solute carrier family 43 member 3 (SLC43A3), STEAP3 metalloenductase, tumour necrosis factor α -induced protein 6 (TNFAIP6) and apolipoprotein B mRNA editing enzyme catalytic subunit 3G (APOBEC3G) were significant genes that stably associated with survival in the TCGA and GSE16011 cohorts of GBM patients. Kaplan-Meier plots with the log-rank test P-values and HRs of these prognosis-associated genes identified in the GBM cohorts from TCGA and GSE16011 are provided in Fig. 6. Compared with the low expression groups, high expression of STEAP3 (Fig. 6A), EMP3 (Fig. 6B), CLEC5A (Fig. 6C), SLC43A3 (Fig. 6D), APOBEC3G (Fig. 6E) and TNFAIP6 (Fig. 6F) indicated poor prognosis in the TCGA GBM cohort and the GSE16011 cohort.

Validation of hub and prognosis-associated genes identified.

To test the stability of the results, the expression and methylation levels of hub genes and prognosis-associated genes were further explored in the TCGA database for verification. As presented in Table IV, the expression levels of all of these genes concurred with the data presented in Supplementary Table I. The differences in the methylation levels of these genes, including CDK1, MMP2 and TNFAIP6, between normal and tumour tissues did not reach statistical significance in the TCGA cohort, which may be due to the different standards

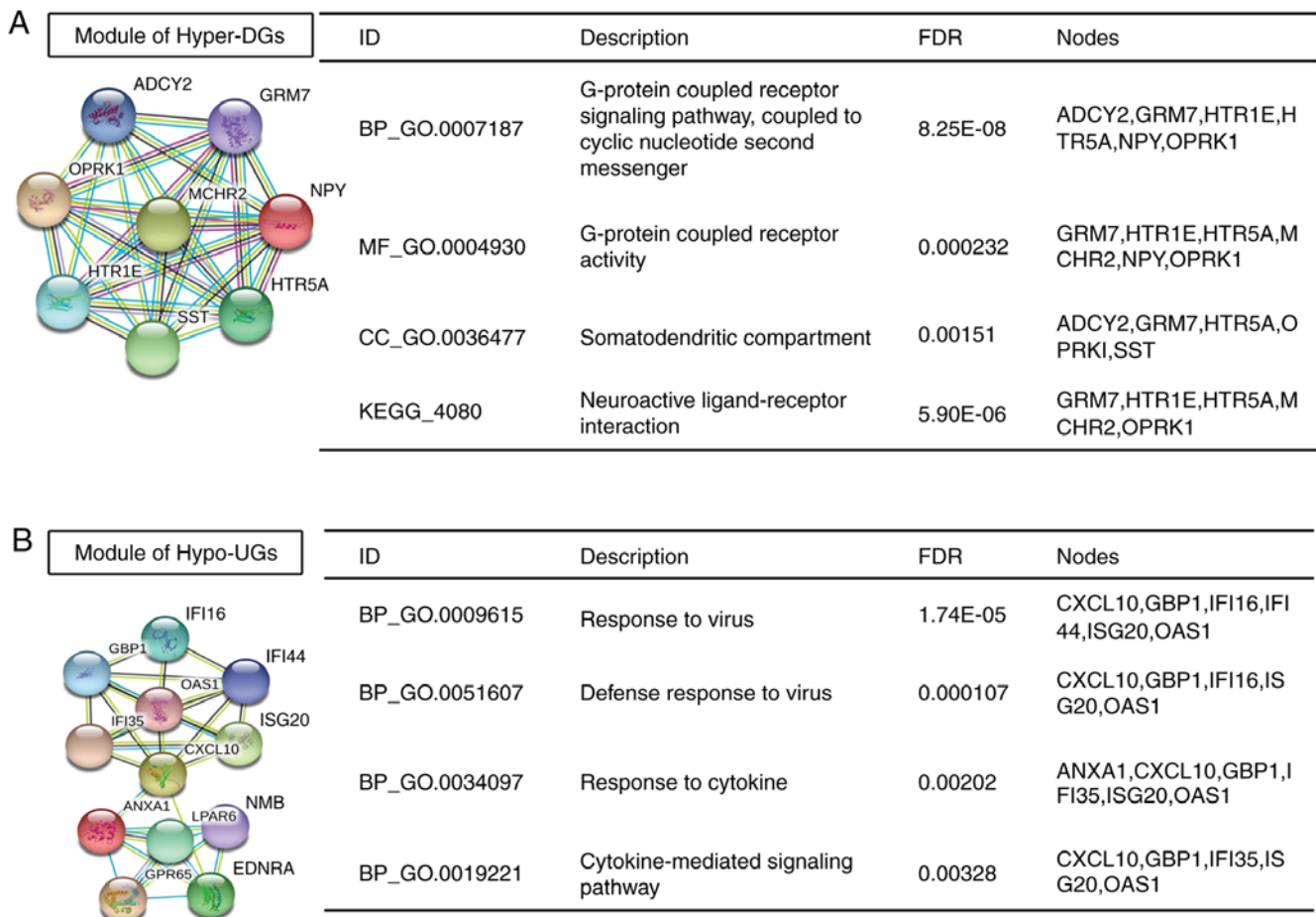


Figure 5. Significant modules in the protein-protein interaction network and enrichment results. (A and B) Significant modules and associated enriched GO terms and KEGG pathways among (A) Hyper-DGs and (B) Hypo-UGs. ADCY2, adenylate cyclase 2; GRM7, glutamate metabotropic receptor 7; OPRK1, opioid receptor kappa 1; MCHR2, melanin concentrating hormone receptor 2; NPY, neuropeptide Y; HTR1E, 5-hydroxytryptamine receptor 1E, HTR5A, 5-hydroxytryptamine receptor 5A; SST, somatostatin; IFI16, interferon gamma inducible protein 16; GBP1, guanylate binding protein 1; IFI14, interferon induced protein 44; IFI35, interferon induced protein 35; OAS1, 2'-5'-oligoadenylate synthetase 1; ISG20, interferon stimulated exonuclease gene 20; CXCL10, C-X-C motif chemokine ligand 10; ANXA1, annexin A1; LPAR6, lysophosphatidic acid receptor 6; NMB, neuromedin B; GPR65, G protein-coupled receptor 65; EDNRA, endothelin receptor type A; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Hyper-DGs, hypermethylated and down-regulated genes; Hypo-UGs, hypomethylated and upregulated genes; CC, cellular component; BP, biological process; MF, molecular function; FDR, false discovery rate.

used to identify DMGs, as mentioned in the Methods section. In general, most of the data concurred with those presented in Supplementary Table I and the results of the present study are therefore reliable and reproducible.

Discussion

DNA methylation-based central nervous system (CNS) tumour classification markedly enhances the diagnostic precision and may correctly modify the primary diagnosis in 12% of cases (24). Global DNA demethylation is one of the most important characteristics of glioma. With the recurrence and progression of glioma, whole DNA methylation undergoes a marked loss. By contrast, hypermethylation of CpG islands in promoter regions is an important mechanism of inducing transcriptional silencing of several genes, which may be important during glioma formation (25). Etcheverry *et al* (12) first performed a genome-wide integrative analysis of methylation and gene expression profiles in the same GBM cohort. They determined that 25% of genes associated with DMCPG

sites in GBM vs. control brain tissues were differentially expressed in a concordant manner. In addition, Wen *et al* (26) reported that with the increase in the degree of malignancy of GBM, the number of genes whose methylation degree was negatively correlated with mRNA expression increased as well. Wang *et al* (27) recently established a methylation-based eight-gene signature predicting the survival outcomes of GBM patients. However, the roles of aberrant methylation in the pathogenic mechanisms of GBM have remained to be fully elucidated.

In the present study, 251 Hypo-UGs and 199 Hyper-DGs that may be associated with the development of GBM were identified by conjunctively analysing gene expression and methylation profiles. Through functional analysis, it was determined that Hyper-DGs were associated with synapsis-associated signalling, which is consistent with the commonly accepted knowledge that GBM cells derived from normal brain cells lose numerous functions, including synaptic transmission performed by normal brain cells. Hypo-UGs were primarily involved in immune-associated signalling.

Table III. Significant genes according to the univariate Cox survival analysis of differentially expressed genes with altered methylation status in the GSE16011 cohort.

A, Hypomethylated and upregulated genes				
Gene	HR	95% CI	Z	P-value
CAPG	1.44	0.20-0.54	4.22	2.40x10 ⁻⁵
PTGFRN	1.51	0.21-0.61	4.03	5.68x10 ⁻⁵
STEAP3 ^a	1.32	0.14-0.42	4.02	5.92x10 ⁻⁵
PHLDA1	0.68	-(0.58-0.20)	-3.97	7.08x10 ⁻⁵
EMP3 ^a	1.22	0.10-0.30	3.84	1.22x10 ⁻⁴
CLEC5A ^a	1.25	0.11-0.34	3.83	1.26x10 ⁻⁴
AQP1	1.19	0.08-0.27	3.70	2.14x10 ⁻⁴
PLA2G2A	1.12	0.06-0.18	3.69	2.25x10 ⁻⁴
SLC43A3 ^a	1.51	0.19-0.64	3.57	3.59x10 ⁻⁴
GPX1	1.77	0.25-0.89	3.52	4.36x10 ⁻⁴
APOBEC3G ^a	1.35	0.13-0.47	3.44	5.76x10 ⁻⁴
TNFAIP6 ^a	1.22	0.08-0.32	3.37	7.61x10 ⁻⁴
OLR1	1.25	0.09-0.36	3.33	8.76x10 ⁻⁴
C1orf54	1.47	0.16-0.62	3.32	8.89x10 ⁻⁴
B, Hypermethylated and downregulated genes				
Gene	HR	95% CI	Z	P-value
GNAL	0.64	-(0.70-0.20)	-3.55	3.85x10 ⁻⁴
ABCA5	1.30	0.11-0.42	3.43	6.14x10 ⁻⁴

^aThese genes were also significant in the The Cancer Genome Atlas cohort. HR, hazard ratio; CI, confidence interval; CAPG, capping actin protein, gelsolin like; PTGFRN, prostaglandin F2 receptor inhibitor; STEAP3, STEAP3 metalloredutase; PHLDA1, pleckstrin homology like domain family A member 1; EMP3, epithelial membrane protein 3; CLEC5A, C-type lectin domain containing 5A; AQP1, aquaporin 1 (Colton blood group); PLA2G2A, phospholipase A2 group IIA; SLC43A3, solute carrier family 43 member 3; GPX1, glutathione peroxidase 1; APOBEC3G, apolipoprotein B mRNA editing enzyme catalytic subunit 3G; TNFAIP6, tumour necrosis factor α -induced protein 6; OLR1, oxidized low density lipoprotein receptor 1; GNAL, G protein subunit alpha L; ABCA5, ATP binding cassette subfamily A member 5.

Of note, pathway analysis of Hypo-UGs indicated enrichment in infection-associated pathways. Increasing evidence supports the view that human herpes virus 6, cytomegalovirus and Epstein-Barr virus infection may participate in glioma pathogenesis (28-30). A previous study has demonstrated by electron microscopy that virus-like particles were present in glioma tissues (31), emphasizing the role of viruses in the aetiology and pathogenesis of GBM. The present results provide complementary evidence supporting the theory that a virus may be a risk factor for the pathogenesis of GBM. In addition, one significant module identified in Hyper-DGs was linked to G-protein coupled receptor signalling pathways. Accumulating evidence has suggested that aberrant expression of G-protein coupled receptors may participate in BP involved in the initiation, progression and metastasis of tumours (32,33). The present results suggest that the function of the G-protein coupled receptor in the regulation of tumour promotion may be due to the aberrant methylation of associated genes. However, this warrants further confirmation by future studies.

Comprehensive analysis of the PPI network identified 6 hub genes in Hypo-UGs and 3 in Hyper-DGs. A previous study suggested that IL-8 may participate in numerous CNS

abnormalities, including gliomas, by enhancing synaptogenesis, influencing synaptic transmission and contributing to neuroinflammation (34). Upregulation of IL-8 has been observed in glioma tissues (35), which may be linked to prostaglandin E2-mediated DNA demethylation of CpG islands in IL-8 genes (24). OAS1 is a gene that encodes several essential proteins that are involved in the innate immune response to viral infection. These encoded molecules may activate latent RNase L, which results in viral and endogenous RNA degradation, and inhibition of viral replication. The present results were consistent with those of a previous study suggesting that OAS1 expression is significantly inversely correlated with its methylation level (12). Somatostatin, encoded by the SST gene, is a peptide hormone secreted by parts of the CNS that affect neurotransmission and cell proliferation (36,37). A hypermethylation-associated decrease in somatostatin expression was reported to be important for uncontrolled proliferation of colorectal cancer and gastric cancer (38-40).

The present results suggested that APOBEC3G, TNFAIP6, SLC43A3, EMP3, CLEC5A and STEAP3 were stable prognostic factors for GBM. STEAP3 is a metalloredutase that has an important role in the function of

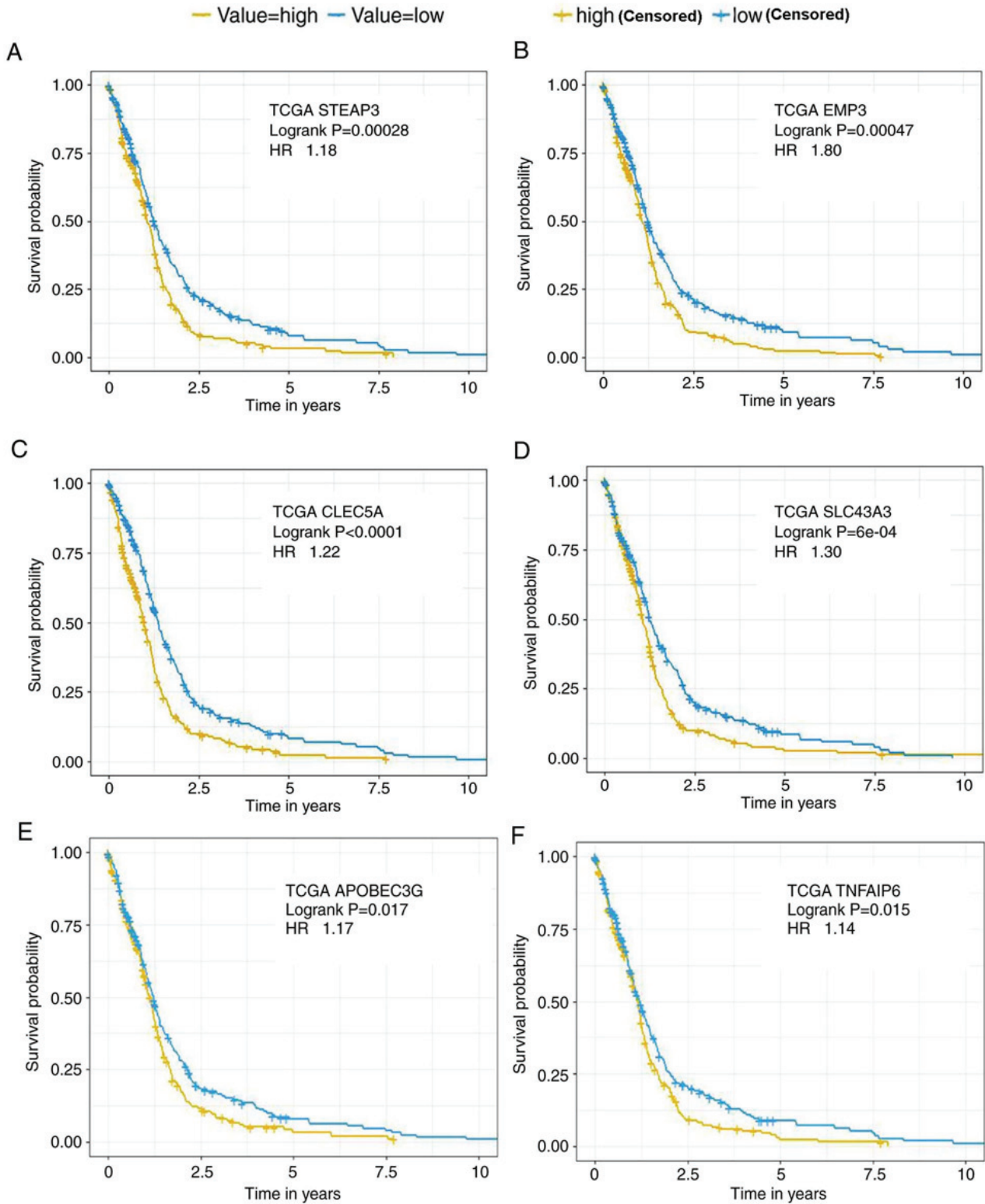


Figure 6. Kaplan-Meier survival analysis. Kaplan-Meier plots were generated to analyse the overall survival of patients with different expression levels of (A) STEAP3, (B) EMP3, (C) CLEC5A, (D) SLC43A3, (E) APOBEC3G and (F) TNFAIP6 in the TCGA datasets and log-rank P-values were compared. The median values of the expression of certain genes were used as the cut-offs to define the high and low expression groups. HR, hazard ratio; TCGA, The Cancer Genome Atlas; CLEC5A, C-type lectin domain containing 5A; EMP3, epithelial membrane protein 3; SLC43A3, solute carrier family 43 member 3; TNFAIP6, tumour necrosis factor α -induced protein 6; APOBEC3G, apolipoprotein B mRNA editing enzyme catalytic subunit 3G.

cellular iron uptake and maintenance of homeostasis. One study suggested that STEAP3 was significantly over-expressed in malignant glioma and that it was associated with poor prognosis (41). Knockdown of STEAP3 inhibited

cell proliferation and progression, indicating a tumour promotion function of the STEAP3 gene. EMP3, which has been reported to exhibit frequent promoter methylation in high-grade glioma (42), is thought to participate in cell

Table IV. Validation of the expression and methylation level of hub genes and prognosis-associated genes in The Cancer Genome Atlas.

A, Hypomethylated and upregulated genes

Gene	Expression status	Fold change	P-value	Methylation status	Delta β (tumour-normal)	P-value
IL8	Upregulated	2.603	0.049	Hypomethylated	-0.23191	1.50x10 ⁻¹⁵
MMP9	Upregulated	3.065	6.65x10 ⁻¹¹	Hypomethylated	-0.232866	7.10x10 ⁻¹³
CDK1	Upregulated	4.275	4.03x10 ⁻¹⁴	-	-	-
OAS1	Upregulated	2.20	2.94x10 ⁻¹³	Hypomethylated	-0.294876	2.70x10 ⁻¹⁵
CXCL10	Upregulated	3.716	1.62x10 ⁻⁵	Hypomethylated	-0.236382	3.70x10 ⁻¹³
MMP2	Upregulated	4.818	4.06x10 ⁻¹⁰	-	-	-
CLEC5A	Upregulated	3.149	6.08x10 ⁻⁹	Hypomethylated	-0.283954	1.10x10 ⁻¹¹
EMP3	Upregulated	6.029	3.25x10 ⁻¹²	Hypomethylated	-0.270879	3.70x10 ⁻¹⁵
SLC43A3	Upregulated	4.864	5.48x10 ⁻⁸	Hypomethylated	-0.231417	3.20x10 ⁻¹⁵
STEAP3	Upregulated	4.968	2.93x10 ⁻¹²	Hypomethylated	-0.210999	3.50x10 ⁻¹⁵
TNFAIP6	Upregulated	2.949	4.86x10 ⁻¹⁵	-	-	-
APOBEC3G	Upregulated	3.055	1.14x10 ⁻⁶	Hypomethylated	-0.269254	2.20x10 ⁻¹²

B, Hypermethylated and downregulated genes

Gene	Expression status	Fold change	P-value	Methylation status	Delta β (tumour-normal)	P-value
SST	Downregulated	-80.516	3.95x10 ⁻²¹	Hypermethylated	0.353893	3.5x10 ⁻⁵
NPY	Downregulated	-22.749	1.58x10 ⁻¹⁰	Hypermethylated	0.454559	2.60x10 ⁻⁷
ADCY2	Downregulated	-3.360	1.54x10 ⁻²²	Hypermethylated	0.213494	1.50x10 ⁻¹³

IL8, interleukin-8; MMP9, matrix metalloproteinase-9; CDK1, cyclin-dependent kinase-1; OAS1, 2'-5'-oligoadenylate synthetase 1; CXCL10, C-X-C motif chemokine ligand 10; MMP2, matrix metalloproteinase-2; CLEC5A, C-type lectin domain containing 5A; EMP3, epithelial membrane protein 3; SLC43A3, solute carrier family 43 member 3; STEAP3, STEAP3 metalloredutase; TNFAIP6, tumour necrosis factor α -induced protein 6; APOBEC3G, apolipoprotein B mRNA editing enzyme catalytic subunit 3G; SST, somatostatin; NPY, neuropeptide Y; ADCY2, adenylate cyclase 2.

proliferation and cell-cell interactions (43,44). Studies of other genes in glioma are limited and these genes require further exploration.

The GBM datasets from TCGA were then used to validate the expression and methylation levels of the hub and prognostic genes identified. The expression levels of all of these genes were consistent with the previously obtained results. However, the methylation levels of CDK1, MMP2 and TNFAIP6 in the TCGA cohort were not consistent with the previous results, which may be due to different methods used to identify the DMGs as aforementioned.

Several limitations of the present study should be considered. First, the influence of AMDEGs lacks further experimental validation. Supplementary molecular experiments are required to better confirm the results of the candidate genes and pathways identified. Second, the sample size was insufficient, although multiple datasets were analysed in combination. Furthermore, the intersection analysis between GSE16011 and GSE50161 may exclude certain important predictors that may result in limited screening ability.

However, the indicators highlighted are biologically reasonable and reliable. In addition, the quality of the chip was not strictly evaluated and the batch effect may have occurred in the interior of each dataset. However, through validation in the TCGA database, the results are relatively stable. The present study represents a significant step in the systematic assessment of the GBM microenvironment.

In conclusion, numerous AMDEGs and associated pathways in GBM were revealed in the present study through integrative analyses of gene expression and methylation profiling. These results will help identify valuable therapeutic targets and diagnostic markers for GBM and promote the understanding of the cumulative roles of epigenetic mechanisms in the aetiology and pathogenesis of GBM. In addition, a set of hub genes in a PPI network was identified, which may be used as methylation-based biomarkers for the precise diagnosis of GBM. Through survival analysis, it was determined that CLEC5A, EMP3, SLC43A3, STEAP3 and APOBEC3G may be used as potential methylation-based prognostic biomarkers and serve as potential targets for treatment.

Acknowledgements

Not applicable.

Funding

The present study was supported by the National Natural Science Foundation of China (grant no. 81402521).

Availability of data and materials

All data generated or analysed during the present study are included in this published article.

Authors' contributions

MZ and QQ conceived and designed the study, and drafted the manuscript. XL, GL and YJ performed the data acquisition, analysed the data and interpreted the results. MZ and GL wrote and edited the manuscript. QQ contributed to critical revisions of the text. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

- Anjum K, Shagufta BI, Abbas SQ, Patel S, Khan I, Shah SAA, Akhter N and Hassan SSU: Current status and future therapeutic perspectives of glioblastoma multiforme (GBM) therapy: A review. *Biomed Pharmacother* 92: 681-689, 2017.
- Thakkar JP, Dolecek TA, Horbinski C, Ostrom QT, Lightner DD, Barnholtz-Sloan JS and Villano JL: Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiol Biomarkers Prev* 23: 1985-1996, 2014.
- Alexander BM and Cloughesy TF: Adult glioblastoma. *J Clin Oncol* 35: 2402-2409, 2017.
- Batash R, Asna N, Schaffer P, Francis N and Schaffer M: Glioblastoma multiforme, diagnosis and treatment; Recent literature review. *Curr Med Chem* 24: 3002-3009, 2017.
- Aldape K, Zadeh G, Mansouri S, Reifenberger G and von Deimling A: Glioblastoma: Pathology, molecular mechanisms and markers. *Acta Neuropathol* 129: 829-848, 2015.
- Kanwal R, Gupta K and Gupta S: Cancer epigenetics: An introduction. *Methods Mol Biol* 1238: 3-25, 2015.
- Romani M, Pistillo MP and Banelli B: Epigenetic targeting of glioblastoma. *Front Oncol* 8: 448, 2018.
- Werner RJ, Kelly AD and Issa JJ: Epigenetics and precision oncology. *Cancer J* 23: 262-269, 2017.
- Bender J: DNA methylation and epigenetics. *Annu Rev Plant Biol* 55: 41-68, 2004.
- Berghoff AS, Hainfellner JA, Marosi C and Preusser M: Assessing MGMT methylation status and its current impact on treatment in glioblastoma. *CNS Oncol* 4: 47-52, 2015.
- Skiriute D, Vaitkiene P, Saferis V, Asmoniene V, Skauminas K, Deltuva VP and Tamasauskas A: MGMT, GATA6, CD81, DR4, and CASP8 gene promoter methylation in glioblastoma. *BMC Cancer* 12: 218, 2012.
- Etcheverry A, Aubry M, de Tayrac M, Vauleon E, Boniface R, Guenot F, Saikali S, Hamlat A, Riffaud L, Menei P, *et al*: DNA methylation in glioblastoma: Impact on gene expression and clinical outcome. *BMC Genomics* 11: 701, 2010.
- Sang L, Wang XM, Xu DY and Zhao WJ: Bioinformatics analysis of aberrantly methylated-differentially expressed genes and pathways in hepatocellular carcinoma. *World J Gastroenterol* 24: 2605-2616, 2018.
- Nazarov PV, Muller A, Kaoma T, Nicot N, Maximo C, Birembaut P, Tran NL, Dittmar G and Vallar L: RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient derived samples. *BMC Genomics* 18: 443, 2017.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, *et al*: NCBI GEO: Archive for functional genomics data sets-update. *Nucleic Acids Res* 41: D991-D995, 2013.
- Edgar R, Domrachev M and Lash AE: Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-210, 2002.
- Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooij JJ, Stubbs AP, Duijm JE, Daemen A, Bleeker FE, Bralten LB, Kloosterhof NK, *et al*: Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res* 69: 9065-9072, 2009.
- Griesinger AM, Birks DK, Donson AM, Amani V, Hoffman LM, Waziri A, Wang M, Handler MH and Foreman NK: Characterization of distinct immunophenotypes across pediatric brain tumor types. *J Immunol* 191: 4880-4888, 2013.
- Lai RK, Chen Y, Guan X, Noursome D, Sharma C, Canoll P, Bruce J, Sloan AE, Cortes E, Vonsattel JP, *et al*: Genome-wide methylation analyses in glioblastoma multiforme. *PLoS One* 9: e89376, 2014.
- Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57, 2009.
- Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, *et al*: The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45: D362-D368, 2017.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
- Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D and Zhu J: The UCSC cancer genomics browser: Update 2015. *Nucleic Acids Res* 43: D812-D817, 2015.
- Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, Koelsche C, Sahm F, Chavez L, Reuss DE, *et al*: DNA methylation-based classification of central nervous system tumours. *Nature* 555: 469-474, 2018.
- Aoki K and Natsume A: Overview of DNA methylation in adult diffuse gliomas. *Brain Tumor Pathol* 36: 84-91, 2019.
- Wen WS, Hu SL, Ai Z, Mou L, Lu JM and Li S: Methylated of genes behaving as potential biomarkers in evaluating malignant degree of glioblastoma. *J Cell Physiol* 232: 3622-3630, 2017.
- Wang W, Zhao Z, Wu F, Wang H, Wang J, Lan Q and Zhao J: Bioinformatic analysis of gene expression and methylation regulation in glioblastoma. *J Neurooncol* 136: 495-503, 2018.
- Chi J, Gu B, Zhang C, Peng G, Zhou F, Chen Y, Zhang G, Guo Y, Guo D, Qin J, *et al*: Human herpesvirus 6 latent infection in patients with glioma. *J Infect Dis* 206: 1394-1398, 2012.
- Cobbs CS, Harkins L, Samanta M, Gillespie GY, Bharara S, King PH, Nabors LB, Cobbs CG and Britt WJ: Human cytomegalovirus infection and expression in human malignant glioma. *Cancer Res* 62: 3347-3350, 2002.
- Akhtar S, Vranic S, Cyprian FS and Al Moustafa AE: Epstein-barr virus in gliomas: Cause, association, or artifact? *Front Oncol* 8: 123, 2018.
- Tani E, Takeuchi J and Ametani T: Virus-like particles in cultured human glioma. *Acta Neuropathol* 16: 266-270, 1970.
- Lynch JR and Wang JY: G protein-coupled receptor signaling in stem cells and cancer. *Int J Mol Sci* 17: E707, 2016.
- O'Hayre M, Degese MS and Gutkind JS: Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Curr Opin Cell Biol* 27: 126-135, 2014.
- Miller G: Brain cancer. A viral link to glioblastoma? *Science* 323: 30-31, 2009.

35. Nitta T, Allegretta M, Okumura K, Sato K and Steinman L: Neoplastic and reactive human astrocytes express interleukin-8 gene. *Neurosurg Rev* 15: 203-207, 1992.
36. Florio T and Schettini G: Somatostatin and its receptors. Role in the control of cell proliferation. *Minerva Endocrinol* 26: 91-102, 2001 (In Italian).
37. Theodoropoulou M and Stalla GK: Somatostatin receptors: From signaling to clinical practice. *Front Neuroendocrinol* 34: 228-252, 2013.
38. Leiszter K, Sipos F, Galamb O, Krenács T, Veres G, Wichmann B, Fűri I, Kalmár A, Patai ÁV, Tóth K, *et al*: Promoter hypermethylation-related reduced somatostatin production promotes uncontrolled cell proliferation in colorectal cancer. *PLoS One* 10: e0118332, 2015.
39. Liu J, Li H, Sun L, Wang Z, Xing C and Yuan Y: Aberrantly methylated-differentially expressed genes and pathways in colorectal cancer. *Cancer Cell Int* 17: 75, 2017.
40. Li H, Liu JW, Liu S, Yuan Y and Sun LP: Bioinformatics-based identification of methylated-differentially expressed genes and related pathways in gastric cancer. *Dig Dis Sci* 62: 3029-3039, 2017.
41. Han M, Xu R, Wang S, Yang N, Ni S, Zhang Q, Xu Y, Zhang X, Zhang C, Wei Y, *et al*: Six-transmembrane epithelial antigen of prostate 3 predicts poor prognosis and promotes glioblastoma growth and invasion. *Neoplasia* 20: 543-554, 2018.
42. Alaminos M, Davalos V, Ropero S, Setién F, Paz MF, Herranz M, Fraga MF, Mora J, Cheung NK, Gerald WL and Esteller M: EMP3, a myelin-related gene located in the critical 19q13.3 region, is epigenetically silenced and exhibits features of a candidate tumor suppressor in glioma and neuroblastoma. *Cancer Res* 65: 2565-2571, 2005.
43. Taylor V and Suter U: Epithelial membrane protein-2 and epithelial membrane protein-3: Two novel members of the peripheral myelin protein 22 gene family. *Gene* 175: 115-120, 1996.
44. Ben-Porath I and Benvenisty N: Characterization of a tumor-associated gene, a member of a novel family of genes encoding membrane glycoproteins. *Gene* 183: 69-75, 1996.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.