Research Paper

Integrative analysis of DNA methylation and gene expression to identify key epigenetic genes in glioblastoma

Danyun Jia[1,*], Wei Lin[2,*], Hongli Tang[1,*], Yifan Cheng[3], Kaiwei Xu[1], Yanshu He[1], Wujun Geng[1], Qinxue Dai[1]

[1]Department of Anesthesiology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang, China
[2]Department of Pediatric Intensive Care Unit, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325027, Zhejiang, China
[3]Department of Neurology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang, China
*Equal contribution

Correspondence to: Wujun Geng, Qinxue Dai; **email:** gengwujun@126.com, daiqinxue@wzhospital.cn
Keywords: glioblastoma, TCGA, methylation, gene expression, biomarker
Received: April 2, 2019 **Accepted:** July 29, 2019 **Published:** August 8, 2019

Copyright: Jia et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Glioblastoma (GBM) ranks the most common and aggressive primary brain malignant tumor worldwide. However, the survival rates of patients remain very poor. Therefore, molecular oncology of GBM are urgently needed. In this study, we performed an integrative analysis of DNA methylation and gene expression to identify key epigenetic genes in GBM. The methylation and gene expression of GBM patients in The Cancer Genome Atlas (TCGA) database were downloaded. After data preprocessing, we identified 4,881 differentially expressed genes (DEGs) between tumor and normal samples, including 1,111 upregulated and 3,770 downregulated genes. Then, we randomly separated all samples into training set (n = 69) and testing set (n = 69). We next obtained 11,269 survival-methylation sites by univariate and multivariate Cox regression analyses. In the correlation analysis, we defined 198 low promoter methylation with high gene expression as epigenetically induced (EI) genes and 111 high promoter methylation with low gene expression as epigenetically suppressed (ES) genes. Key markers including C1orf61 and FAM50B were selected with a Pearson correlation coefficient greater than 0.75. Further, we chose the 20 CpG methylation sites of above two genes in unsupervised clustering analysis using the Euclidean distance. We found that the prognosis of the hypomethylated group was significantly better than that in the hypermethylated group (log-rank test *p*-value = 0.011). Based on the validation in the TCGA testing set and GEO dataset, we validated the prognostic value of our signature (*p*-value = 0.02 in TCGA and 0.012 in GEO). In conclusion, our findings provided predictive and prognostic value as methylation-based biomarkers for the diagnosis and treatment of GBM.

INTRODUCTION

Glioblastoma (GBM) ranks the most common and aggressive primary brain tumor worldwide [1]. It is a fast-growing malignant tumor that arises from multiple cell types with neural stem-cell-like properties. Besides conventional therapy, current approaches such as small molecules and gene therapy are developed in recent years [2, 3]. New synthetic small molecules were discovered as promising anti-GBM agents [3]. Although with various treatments, patient outcomes remain between 12 and 15 months survival rate, and with five-year survival rates at only 10% [4]. Therefore, advances in the field of molecular oncology of GBM are urgently needed.

The major factors contributing to the pathogenesis of human cancers were epigenetic molecular mechanisms, including GBM [5]. With the help of gene microarray and RNA-seq, the aberrant expression profiles of GBM in genome and transcriptome level were increasing reported. Using the gene expression data from Gene Expression Omnibus (GEO) database, Bo et al. [6] identified a total of 431 differentially expressed genes (DEGs) between GBM and normal samples. After various bioinformatics analysis, 69 DEGs were identified significantly associated with GBM prognosis. Another study found 486 DEGs based on the gene expression profile of GSE50161 [7]. CDK1, CCNB1 and CDC20 were selected in survival analysis and high expression was significantly associated with poor survival in GBM. However, numerous identified DEGs will not contribute to the clear understandings of biological pathogenesis of GBM.

DNA methylation was found in the dinucleotides of nearly eighty percent of the CpG islands in the genome [8]. It was catalyzed by DNA methyltransferases that controls various cell activities such as proliferation, apoptosis, and differentiation. As for human cancers, methylation was known to be abnormal in all forms of cancers [9] and abnormal methylation of promoters could lead to silence of tumor suppressor genes, affecting transcriptional pathways and resulting in the cancer development [10]. In addition, targeted drugs about DNA methyltransferase inhibitors have been approved for the treatment of chronic myelomonocytic leukemia and acute myelogenous leukemia as well as a second generation of DNA methyltransferase inhibitors [11]. Intra-tumor DNA methylation heterogeneity has been proved a feature of GBM [12]. Moreover, the promoter methylation status of the O^6-methylguanine-DNA methyltransferase (MGMT) gene has been described as the predictor of chemotherapeutic response and patients' survival in GBM [13]. Wang et al. [14] developed a signature with three genes (FPR3, IKBIP and S100A9) signature for prognosis in patients with MGMT promoter-methylated GBM using data from Chinese Glioma Genome Atlas (CGGA) and TCGA. In another study, Wen et al. [15] performed analysis of methylated genes as potential biomarkers in evaluating malignant degree of GBM. In this study, they found a total of 668, 412, 470, and 620 methylation or demethylation genes associated with the degree of GBM from grades 1 to 4. Therefore, abnormal methylation genes can act as potential oncogenes or anti-oncogenes in the development and progression of cancers, suggesting their potential roles as biomarkers.

In the present study, we performed an integrative analysis of DNA methylation and gene expression identified key epigenetic genes in GBM. The methylation and gene expression of GBM patients in TCGA database were downloaded. After data preprocessing, we identified 4,881 DEGs between tumor and normal samples, including 1,111 upregulated and 3,770 downregulated genes. Then, we randomly separated all samples into training set and testing set. We next obtained 11,269 survival-methylation sites by univariate and multivariate Cox regression analyses. In the correlation analysis, we defined 198 low promoter methylation with high gene expression as EI genes and 111 high promoter methylation with low gene expression as ES genes. Key markers including C1orf61 and FAM50B were selected with a Pearson correlation coefficient greater than 0.75. Further, we chose the 20 CpG methylation sites of above two genes in unsupervised clustering analysis using the Euclidean distance. We found that the prognosis of the hypomethylated group was significantly better than that in the hypermethylated group (log-rank test p-value = 0.011). Based on the validation in the TCGA testing set and GEO dataset, we validated the prognostic value of our signature (p-value = 0.02 in TCGA and 0.012 in GEO). In conclusion, our findings provided predictive and prognostic value as methylation-based biomarkers for the diagnosis and treatment of GBM.

RESULTS

DNA methylation data selection and characteristics

In this study, we performed an integrative analysis of DNA methylation and gene expression identified key epigenetic genes in GBM (Figure 1). We used the gene expression and DNA methylation profiles from TCGA database. A total of 138 GBMs and normal samples with clinical information data were obtained. Moreover, there were 20,530 genes were downloaded from the TCGA database for subsequent analysis. Because DNA methylation in promoter regions strongly influences gene expression, we selected CpGs in promotor regions that were defined as 2 kb upstream to 0.5 kb downstream from TSS. After preprocessing data, we finally obtained 145,907 methylation sites for downstream analysis.

Clinical patient characteristics

We obtained the clinical information including sample ID, vital status, age at initial pathologic diagnosis, days to death, days to last follow up, and grade. All samples were randomly divided into two groups: the training set (n = 69) and the testing set (n = 69). The training set and test set are required to meet the following criteria: first, samples are randomly assigned to training set and testing set; second, the age distribution, follow-up time and patient death rate should be similar in these two groups. The expression profiles and clinical information of training set were shown in Supplementary Tables 1 and 2, respectively. In addition, the expression profiles and

clinical information of testing set were shown in Supplementary Tables 3 and 4, respectively.

Determining DEGs of GBM

According to the screening criteria, a total of 4,881 significant DEGs were obtained from all the tumor and normal samples (Supplementary Table 5). There were 1,111 genes were upregulated and 3,770 genes downregulated. The expression profiles of the most significant 100 genes were shown in Figure 2.

Survival analysis of methylation sites in the training set

In order to determine methylation sites associated with survival outcomes, we performed univariate and

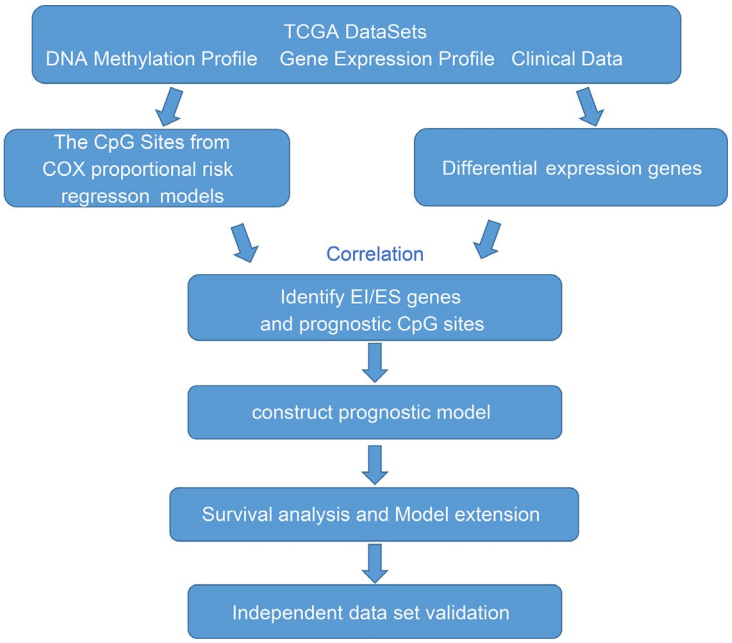

Figure 1. The workflow of the present study.

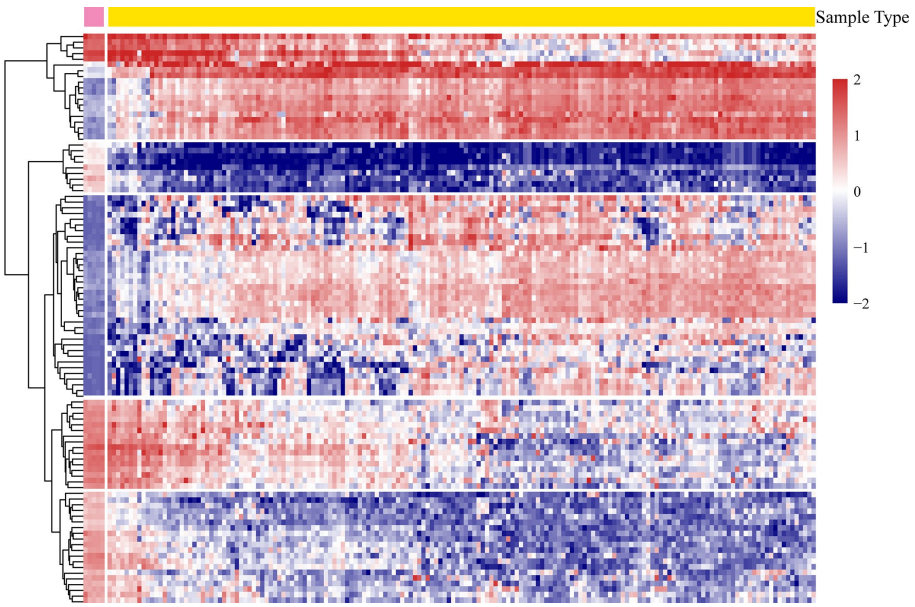

Figure 2. The heatmap expression profiles of the most significant 100 genes.

multivariate Cox regression analyses of the obtained methylation sites of GBM. There were a total of 11,269 methylation sites and we generated a new survival-methylation expression profiles for further analysis (Supplementary Table 6).

Correlation analysis of DEGs and survival-methylated genes

DNA methylation level can affect the gene expression. High methylation expression often inhibits downstream gene expression, and low methylation level tends to increase the downstream gene expression. The correlation analysis steps for calculating differentially expressed genes and differentially methylated genes were as follows: 1) Calculating the intersection of differentially methylated genes and DEGs. 2) Identifying the number of genes whose differential expression was up-regulated and differentially methylated was down-regulated. In addition, identifying the number of genes whose differential expression was down-regulated and differential methylation was up-regulated. Therefore, a total of 324 up-regulated genes, 162 down-regulated genes, 249 methylated down-regulated genes, and 237 methylated up-regulated genes were obtained (Supplementary Table 7).

We then analyzed the Pearson correlations between upregulated DEGs and downregulated survival-methylated genes, as well as downregulated DEGs and upregulated survival-methylated genes. As shown in Figure 3A, we found that there were a total of 198 genes between upregulated DEGs and downregulated survival-methylated genes. In addition, 111 genes were selected between downregulated DEGs and upregulated survival-methylated genes. Next, we performed analysis of the promoter methylation distribution of DEGs between tumor samples and normal samples. The results showed that highly expressed genes in tumors had lower promoter methylation in normal samples, indicating a negative correlation between promoter DNA methylation and gene expression in normal and tumor tissues (Figure 3B).

Pathway enrichment analysis of EI and ES genes

We found a total of 198 low promoter methylation with high gene expression (EI genes), as well as a total of 111 high promoter methylation with low gene expression (ES genes) (Supplementary Table 8).

Next, we used online tools "Metascape" to performed pathway enrichment analysis. As shown in Figure 4A, we found that EI and ES genes were significantly enriched in pathways including Signaling by WNT, negative regulation of cell differentiation, regulation of extracellular matrix organization, and cellular response to cAMP. The "Metascape" also provided the interactions of genes based on these pathways (Figure 4B). These results suggested that EI and ES genes screened in our study were involved in the biological process of the occurrence and development of GBM.

Construction of the prognosis risk model based on methylation genes

In order to further screen potential EI and ES genes, Pearson correlation analysis was used to calculate the correlation between promoter methylation and gene expression of EI and ES genes. There were 16 key genes with negative correlations. Next, we selected genes with a correlation coefficient greater than 0.75 as key markers. They were C1orf61 and FAM50B.

Further, we chose the 20 CpG methylation sites of above two genes (Table 1) in unsupervised clustering analysis. Using the Euclidean distance to calculate the similarity between samples, we found that all samples can be divided into two groups Cluster 1 and Cluster 2 according to the 20 CpG methylation sites. Moreover, the samples in Cluster 1 were with high methylation level, but samples in Cluster 2 were with low methylation level (Figure 5A). Further analysis was performed to explore the prognosis between two groups. As shown in Figure 5B, we found that the prognosis of the hypomethylated group was significantly better than that in the hypermethylated group (log-rank test p-value = 0.011). Moreover, we compared the ages of patients in these two groups and found that the age distribution of patients in the hypomethylated group was lower than that in the hypermethylated group (Figure 5C).

IDH1 mutation and DNA methylation in GBM

IDH mutation is a phenomenon that occurs in the early stage of tumor and IDH mutation is considered as an important marker of low-grade glioma and GBM. IDH mutation can promote the hypermethylation of CpG in the promoter of most genes which contributes to the epigenetic instability of tumor cells. To explore the association between IDH1 mutation and DNA methylation in GBM, all samples were divided into IDH mutation group (n = 7) and IDH non-mutation group (n = 131) according to the IDH1 gene mutation. As shown in Figure 6, samples in IDH mutation group exhibited lower methylation level than that in IDH non-mutation group.

Then, we compared the expression of each methylation site in two groups. As shown in Figure 7, we found that 19 of the 20 sites were significantly expressed between IDH mutation and IDH non-mutation groups

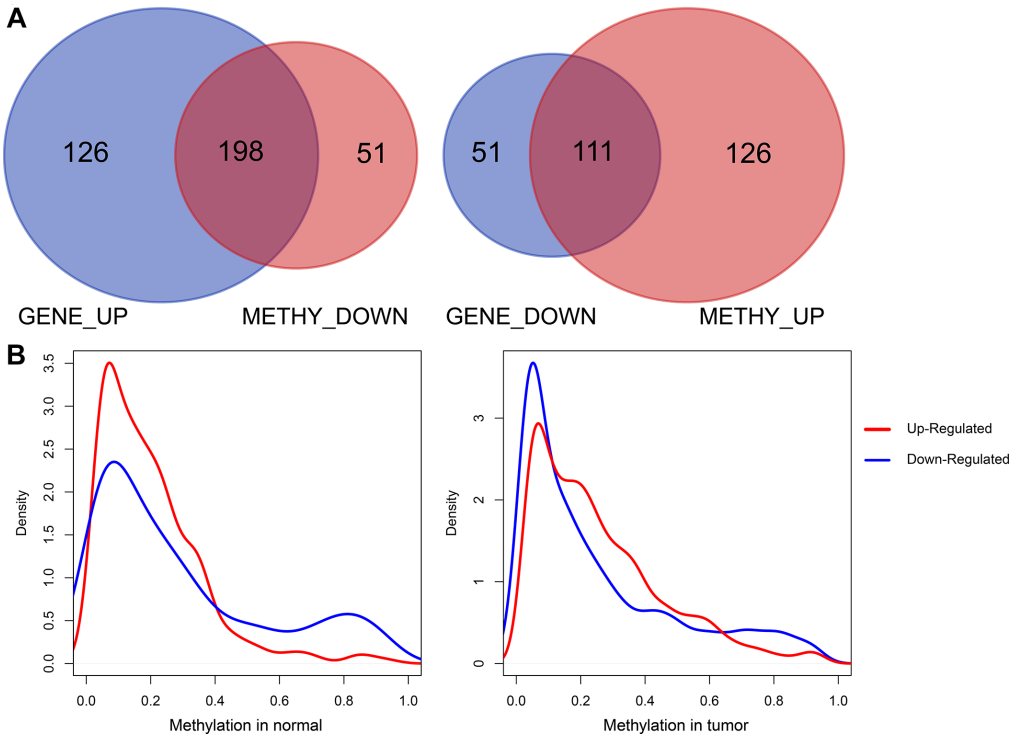

Figure 3. Correlation analysis of DEGs and survival-methylated genes. (**A**) The intersection results of DEGs and survival-methylated genes. (**B**) Distribution of promoter methylation levels in tumor and normal samples.

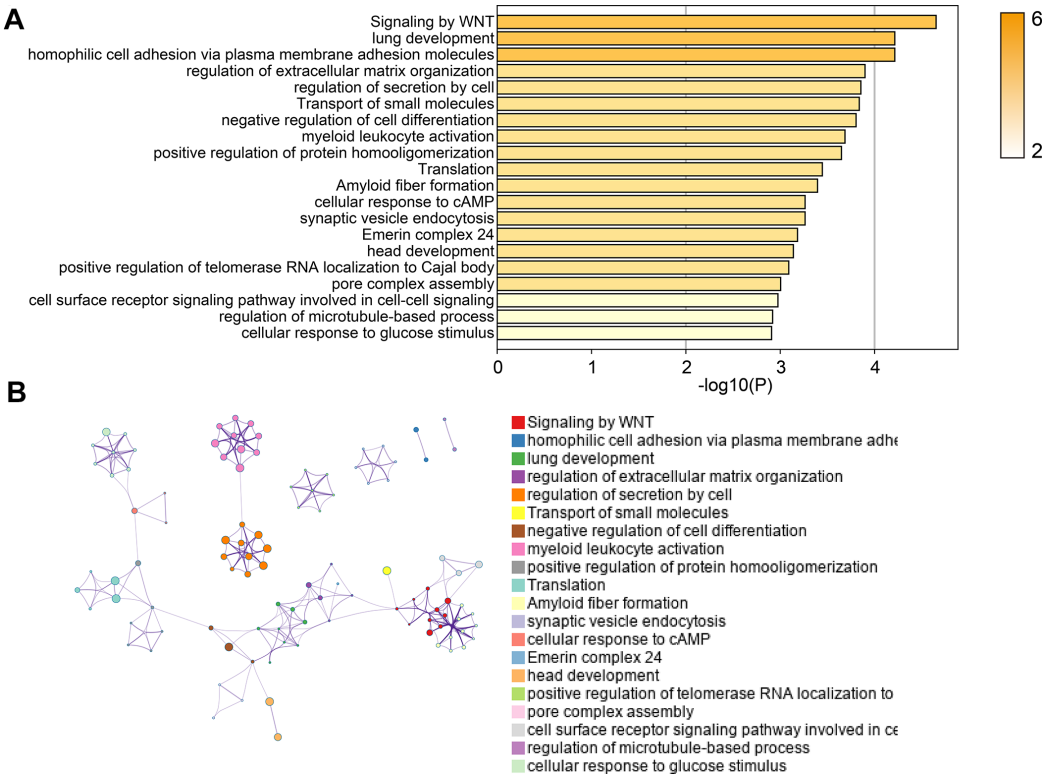

Figure 4. Pathway enrichment analysis of EI and ES genes. (**A**) The pathway enrichment results of EI and ES genes. (**B**) The network diagram of interacting genes.

Table 1. The annotation of 20 CpG sites.

cg probe	Gene	Chrom	Site
cg09938227	C1orf61	1	156390124
cg18197332	FAM50B	6	3849458
cg01570885	FAM50B	6	3849272
cg04447621	FAM50B	6	3849475
cg21740964	FAM50B	6	3849331
cg07898446	FAM50B	6	3849294
cg18487516	FAM50B	6	3849542
cg18872973	FAM50B	6	3849095
cg25195497	FAM50B	6	3849327
cg13101072	FAM50B	6	3849818
cg21177626	FAM50B	6	3849411
cg18656763	FAM50B	6	3849235
cg27445347	FAM50B	6	3849801
cg03954573	FAM50B	6	3849434
cg01905633	FAM50B	6	3849391
cg23835083	FAM50B	6	3849536
cg12840312	FAM50B	6	3849381
cg12497786	FAM50B	6	3849577
cg13289019	FAM50B	6	3849350
cg17739279	FAM50B	6	3849190

(p-value < 0.01). Above results suggested that these methylation sites were closely associated with IDH1 mutation.

Validation in the TCGA testing set and GEO dataset

To validate the results of our methylation data and prognostic model, we used the testing set (n = 69) based on TCGA data. We used the expression of 20 methylation sites and further used hierarchical cluster analysis. We found that the 20 CpG methylation sites can also clearly divide all samples into two groups (Figure 8A). The methylation levels of Cluster 1 group were significantly higher than Cluster 2. Moreover, the prognosis of samples in the hypomethylated group was significantly better than that in the hypermethylated group (log-rank test p-value = 0.02) (Figure 8B). It can also be seen that the age distribution in hypomethylated group was lower than that in the hypermethylated group, which was consistent with the results of the training set (Figure 8C).

In addition, the DNA methylation (GSE36278) [16] of GBM was downloaded with a total of 142 patients. First, we selected the expression profiles of 20 methylation sites (Supplementary Table 9) and clinical information (Supplementary Table 10). Next, we divided all samples into two groups using hierarchical cluster method (Figure 9A). Results showed that significant survival difference was found in two groups (log-rank test p-value = 0.012) (Figure 9B). Moreover, we compared the age distribution between two groups and found that high methylation group was higher than low methylation group (Figure 9C). These results were consistent with TCGA dataset, suggesting that this model can be applied to other samples.

DISCUSSION

In the present study, we performed an integrative analysis of DNA methylation and gene expression identified key epigenetic genes in GBM. We obtained 11,269 survival-methylation sites by univariate and multivariate Cox regression analyses. In the correlation

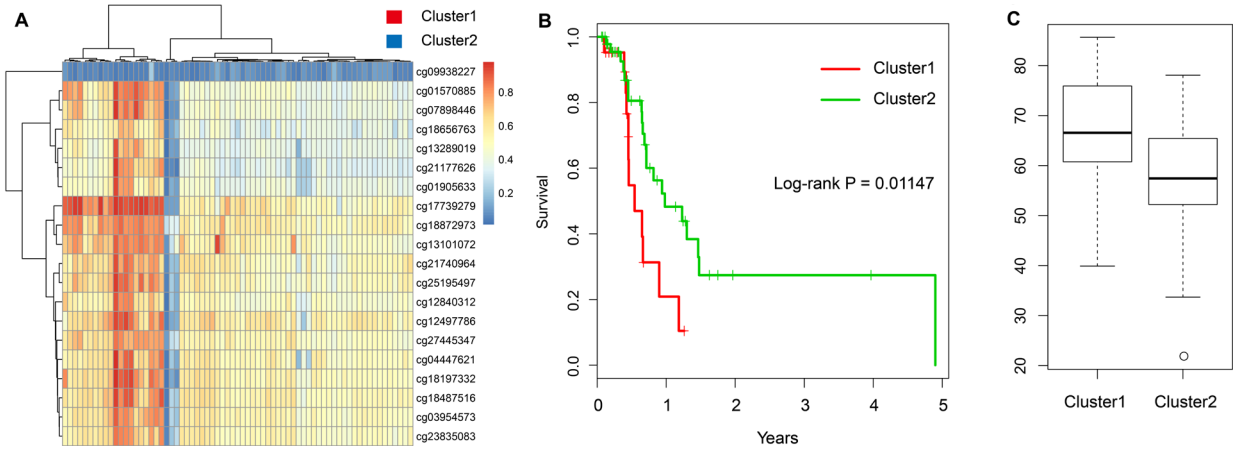

Figure 5. Construction of the prognosis risk model based on methylation genes. (**A**) The heatmap of 20 methylation sites in the training set. (**B**) The K-M plot of the hypomethylated and hypermethylated groups. (**C**) The age distribution of patients in the hypomethylated and hypermethylated groups.

analysis, we defined 198 low promoter methylation with high gene expression as EI genes and 111 high promoter methylation with low gene expression as ES genes. Further, we chose the 20 CpG methylation sites of above two genes in unsupervised clustering analysis using the Euclidean distance. We found that the prognosis of the hypomethylated group was significantly better than that in the hypermethylated group. Based on the validation in the TCGA testing set and GEO dataset, we validated the prognostic value of our signature. Our findings provided predictive and prognostic value as methylation-based biomarkers for the diagnosis and treatment of GBM.

The occurrence and proliferation of cancer is regulated by epigenetic and genetic events, as well as epigenetic modifications. They are increasingly identified as important targets for cancer research [10]. DNA methylation catalyzed by DNA methyltransferases (DNMTs) is one of the important epigenetic mechanisms for controlling cell proliferation, apoptosis, differentiation, cell cycle and transformation in eukaryotes. Abnormal DNA methylation in cancer can be produced by mutation before or after cell transformation [9]. Moreover, it can regulate normal gene expression and facilitate chromatin organization within cells, which are accompanied by alterations in chromatin structure at gene regulatory regions [17]. Also, there were many literatures about the use of DNA methylation measurements for cancer diagnosis through examples of methylated genes [18].

In GBM, there are several studies about the molecular roles of DNA methylation. For examples, Wang et al. [19] used the gene expression and methylation profiles from TCGA as well as the Chinese Glioma Genome Atlas (CGGA) database. A total of 3,365 DEGs were identified with 2,940 genes expressed hypomethylation and high expression, while 425 genes showed hypermethylation and low expression in GBM. The eight genes (C9orf64, OSMR, MDK, MARVELD1, PTRF, MYD88, BIRC3, RPP25) were characterized to divide GBM patients into two groups with different survival outcomes. In addition, different clinical and

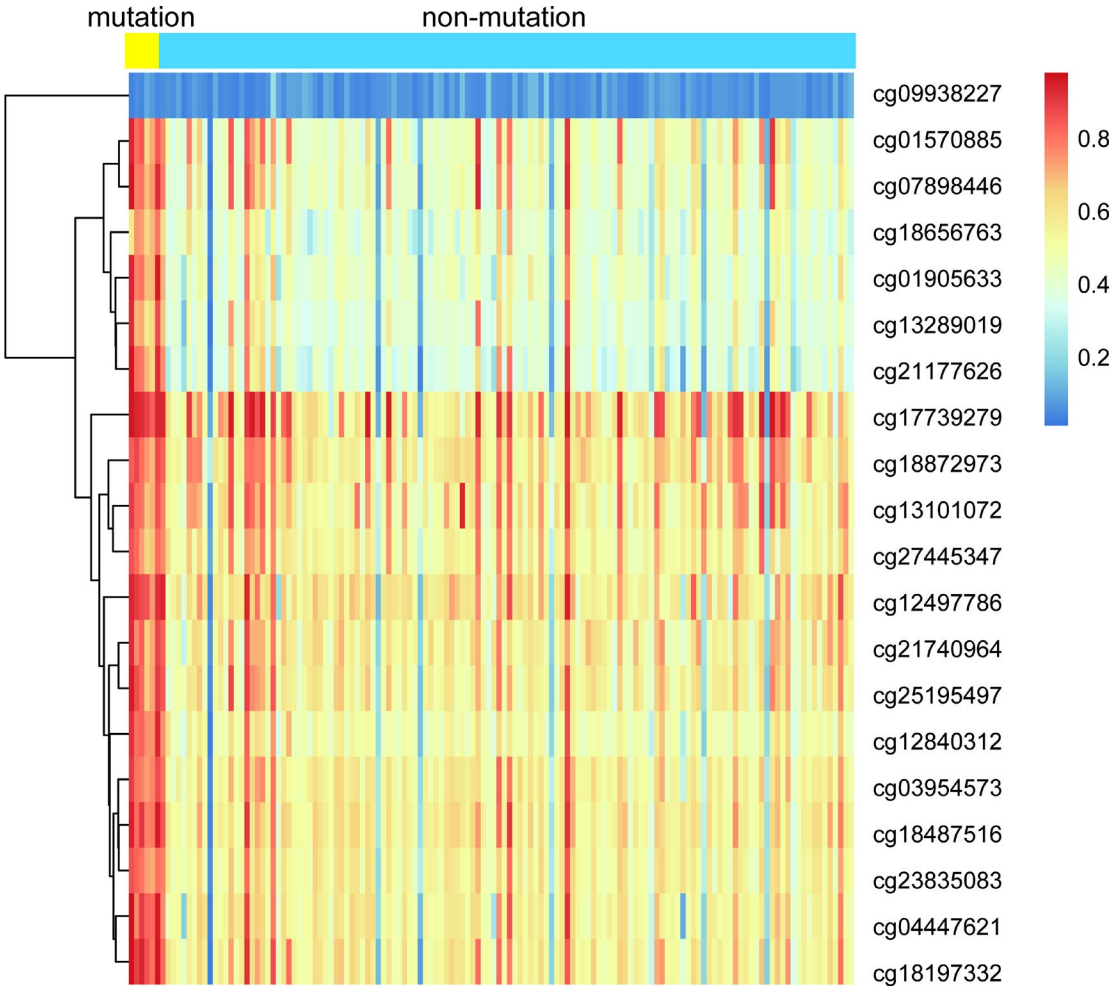

Figure 6. The heatmap of IDH1 mutation and DNA methylation in GBM.

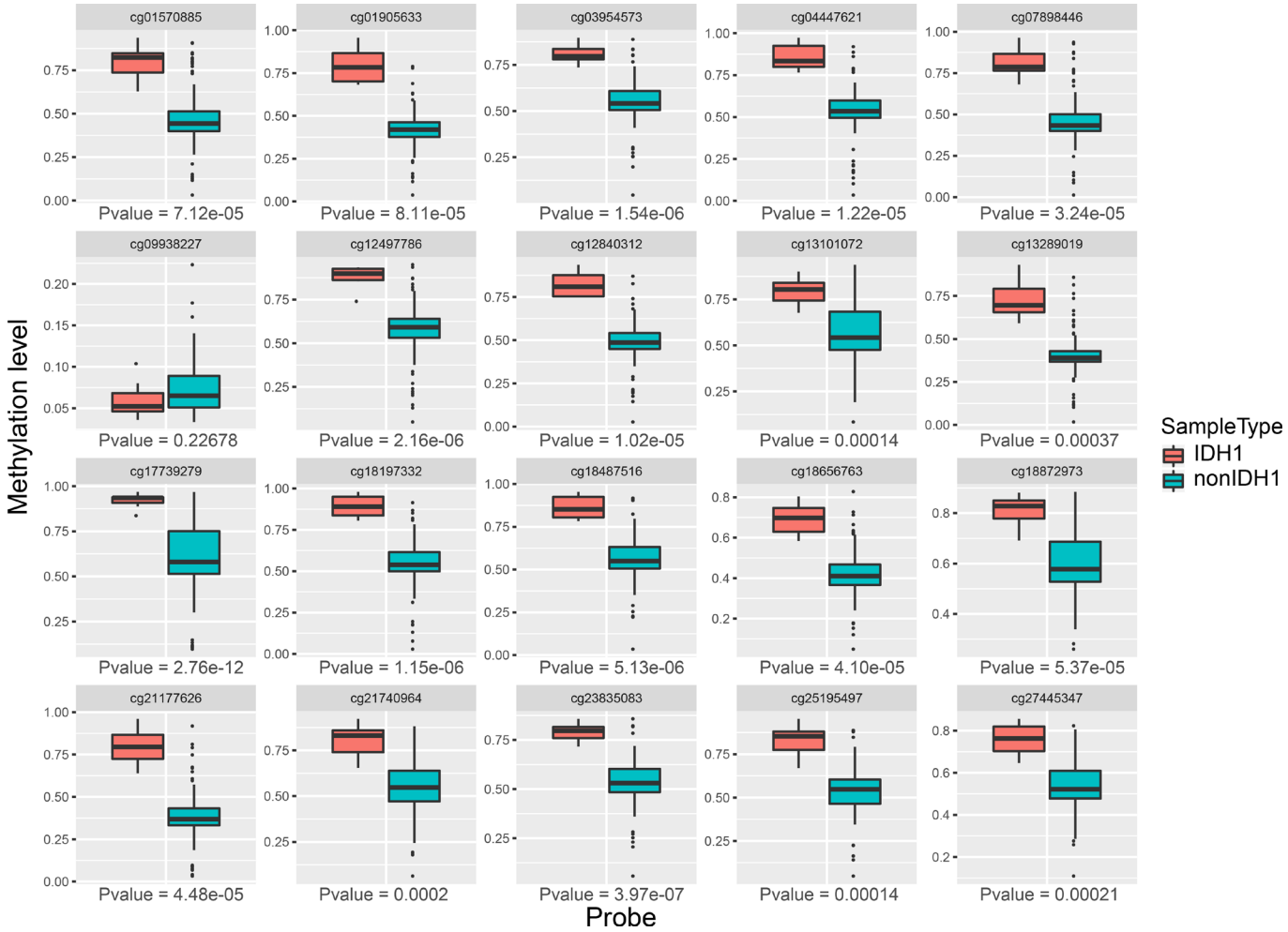

Figure 7. The expression profiles of 20 methylation sites between IDH1 mutation and non-mutation groups.

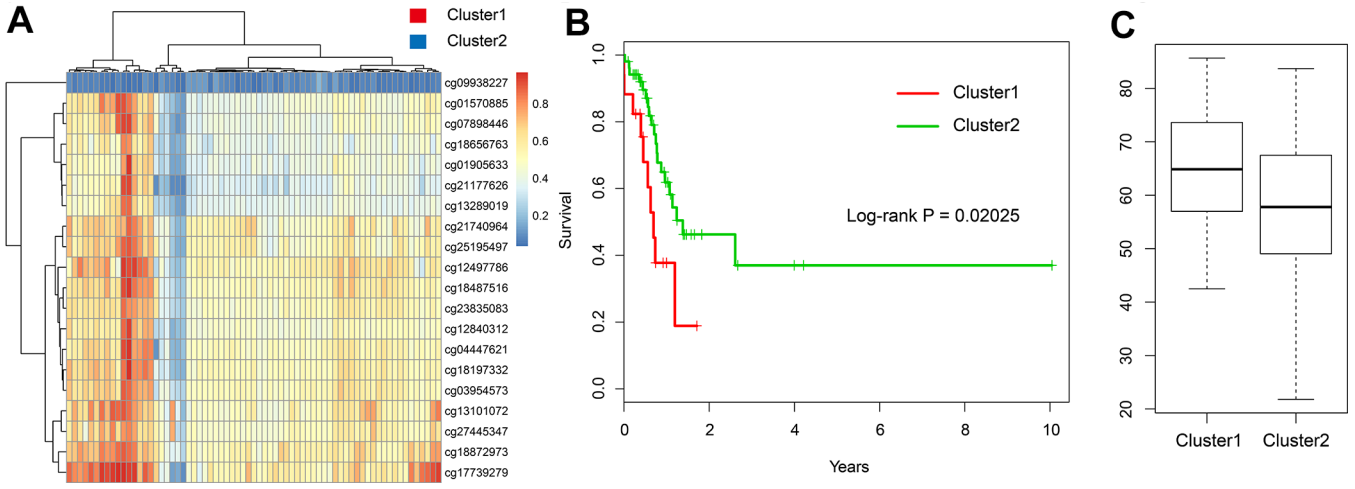

Figure 8. Validation in the TCGA testing set. (**A**) The heatmap of 20 methylation sites in the testing set. (**B**) The K-M plot of the hypomethylated and hypermethylated groups in the testing set. (**C**) The age distribution of patients in the hypomethylated and hypermethylated groups in the testing set.

molecular characteristics were also shown between the two groups. In another study, the positive prognostic value of MGMT promoter hypermethylation has been demonstrated in adult GBM, and the MGMT promoter methylation status is a clinically relevant predictor of the newly diagnosed GBM elderly population [19]. The roles of MGMT promoter methylation in GBM were also reported in various studies [20–22]. Ma et al. [23] reported that the hypermethylation of CXCR4 can predict patients' OS in GBM. Besides, the methylation of AURKA, KIF4A, and NUSAP1 in GBM was also investigated [24].

In our study, key markers including C1orf61 and FAM50B were selected with a Pearson correlation coefficient greater than 0.75. Chromosome 1 open reading frame 61 (C1orf61) was reported to be up-regulated in hepatic cirrhosis tissues and up-regulated in primary hepatocellular carcinoma. Moreover, hepatitis B virus (HBV)-positive patients exhibited significantly higher levels of C1orf61 expression than HBV-negative patients. The overexpression of C1orf61 promoted cell proliferation and colony formation, as well as cell cycle progression. In addition, the overexpression of C1orf61 facilitated cellular invasion and metastasis. The overexpression of C1orf61 induced the epithelial-mesenchymal transition (EMT) that is linked to metastasis [25]. FAM50B (family with sequence similarity 50, member B) was shown that average methylation level of FAM50B was lower in asthenozoospermia group than in control group [26]. CpG sites (mapped to gene FAM50B) were also reported to be differentially expressed in the study of 24-hour exposure to air pollution [27]. DNA methylation changes of FAM50B in individuals with developmental delay/intellectual disability were observed [28]. How-ever, these two genes were not reported in GBM.

DNA methylation patterns can predict prognosis and survival of human cancers [29]. The utility of methylation biomarkers for the molecular characterization of cancer with implications for patients' prognosis. In one study, researchers identified and validated biomarkers for melanoma development (HOXA9 DNA methylation) and tumor progression (TBC1D16 DNA methylation). In addition, this study determined a prognostic signature with potential clinical value [30]. Gastric cancers showed significantly lower LINE-1 methylation levels compared to matched normal gastric mucosa and hypomethylation of LINE-1 was significantly associated with shorter overall survival [31]. Moreover, in the study of esophageal squamous cell carcinoma, LINE-1 hypomethylation is associated with a poor prognosis among patients [32]. Its methylation level was also associated with hepatocellular carcinomas [33]. Based on TCGA methylation expression profiles of gastric cancer, Hu et al. performed a DNA methylation gene signature consisting of five genes (SERPINA3, AP000357.4, GZMA, AC004702.2, and GREB1L) [34]. In addition, in other human cancers, there were also various studies about DNA methylation and prognostic signature, such as head and neck squamous cell carcinoma [35], cutaneous melanoma [36], glioma [37], and lung cancer [38]. Above results suggested that significant DNA methylation genes may be a new predictor and prognostic biomarker for cancers.

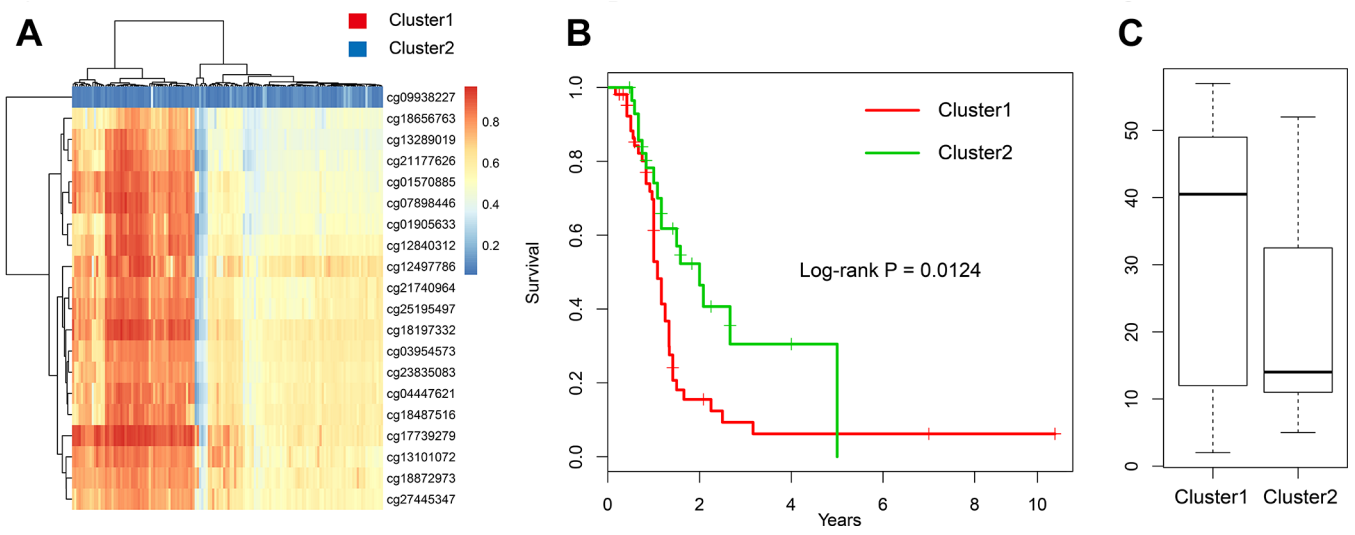

Figure 9. Validation in the GEO dataset. (**A**) The heatmap of 20 methylation sites in GSE36278. (**B**) The K-M plot of the hypomethylated and hypermethylated groups in GSE36278. (**C**) The age distribution of patients in the hypomethylated and hypermethylated groups in GSE36278.

The prognostic ability of this methylation signature may improve the risk stratification of patients with GBM. In the future clinical application, this methylation signature may help people accurately guide clinical treatments and determine prognosis of patients. However, whether this signature can improve GBM diagnosis or treatment, it still remains unknown and this is what we will study in our future work. Besides, considering that C1orf61 was closely associated with cell proliferation, colony formation, cell cycle progression, and EMT, we assumed that this gene can participate in the occurrence and development of GBM through the above pathways. However, how methylation impacts the two key genes and their downstream effects are still the work we need to explore in the future.

In our study, we established a prognosis risk model based on methylation genes in GBM using the 20 CpG methylation sites of above two genes for GBM. In conclusion, our findings provided predictive and prognostic value as methylation-based biomarkers for the diagnosis and treatment of GBM.

MATERIALS AND METHODS

Data selection from TCGA database and preprocessing

All data were downloaded from the TCGA database (https://cancergenome.nih.gov/) [39] based on RNA-seq including DNA methylation, gene expression and IDH1 mutation expression profiles. The methylation data generated with the Illumina Infinium HumanMethylation 450 BeadChip array. The methylation level of each probe was represented by the β-value (from 0 to 1). First, the CpG sites with missing value > 70% of all samples were removed. Then, we used *impute* R package by k-nearest neighbors (KNN) method for the missing values of methylation data. We further removed the genomic unstable sites including CpGs in sex chromosomes and single nucleotide polymorphisms. We selected CpGs in promotor regions, which were defined as 2 kb upstream to 0.5 kb downstream from transcription start sites (TSS) [40]. Finally, we selected samples with gene expression profiles including a total of 138 tumor and normal samples.

All samples were separated into two cohorts: a training set (n = 69) and a testing set (n = 69). The methylation data of training set and clinical information (survival status, time, and age) was used to select CpG sites with prognostic value by univariate and multivariate COX proportional risk regression models. Last, according to the relationship between CpG sites and genes, we obtained key genes that significantly associated with survival.

Determining DEGs of GBM and methylated sites

We used paired T-test as a statistical method to screen DEGs and methylated sites between tumor and normal samples, and multiple tests were performed for *p*-value correction. Finally, genes with false discovery rate (FDR) < 0.01 were screened as significant DEGs and methylated sites.

Correlation analysis of DEGs and survival-methylated genes

To explore the association between DEGs and methylation, first, we used univariate Cox proportional risk regression model to analyze each methylation site and survival data. Then, clinical factors including grade and age were added as covariables for multivariate Cox regression analyses. Finally, the intersection results of univariate and multivariate Cox regression (*p*-value < 0.05) were obtained. Here, we defined genes with downregulated methylation in promoter region as downregulated survival-methylated genes, and genes with upregulated methylation in promoter region as upregulated survival-methylated genes.

We next performed the correlation analysis between upregulated DEGs and downregulated survival-methylated genes, as well as downregulated DEGs and upregulated survival-methylated genes. We used *Venny* software to screen the intersected genes. The average expression level of all methylated sites associated with survival represented the final expression level of this survival-methylated gene.

Pathway enrichment analysis of epigenetically induced and epigenetically suppressed genes

In order to further identify the mutex genes, we defined low promoter methylation with high gene expression as EI genes. High promoter methylation with low gene expression as ES genes. Then, we used online tools "Metascape" (http://metascape.org) to performed pathway enrichment analysis of EI and ES genes.

Construction of the prognosis risk model based on methylation genes

In order to further screen potential EI and ES genes, Pearson correlation analysis was used to calculate the correlation between promoter methylation and gene expression of EI and ES genes. We selected genes with a correlation coefficient greater than 0.75 as key markers. Hierarchical clustering algorithm was used to cluster the samples of the training set, and Euclidean distance was used to calculate the similarity between the samples. We used *survival* R package to observe whether the survival

difference between the high-risk and low-risk groups by K-M survival analysis.

IDH1 mutation and DNA methylation in GBM

To explore the association between IDH1 mutation and DNA methylation in GBM, according to the IDH1 gene mutation, all samples were divided into IDH mutation group (n = 7) and IDH non-mutation group (n = 131). 20 methylation sites were used to compare methylation differences between two groups.

Validation in the TCGA testing set and GEO dataset

To validate the results of our methylation data and prognostic model, we used the testing set (n = 69) based on TCGA data. In addition, the DNA methylation (GSE36278) [18] of GBM was downloaded from NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/). A total of 142 patients with DNA methylation profiling were included for further validation. This dataset was carried on Illumina HumanMethylation450 BeadChip platform.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

FUNDING

This work was supported by the National Natural Science Foundation of China, No. 81704180, 81774109 and science and technology plan project of Wenzhou science and technology bureau (Y20170143, Y20180496, Y20190511), and Major Science and Technology Innovation Project of Wenzhou (2019ZY0012).

REFERENCES

1. Holland EC. Glioblastoma multiforme: the terminator. Proc Natl Acad Sci USA. 2000; 97:6242–44. https://doi.org/10.1073/pnas.97.12.6242 PMID:10841526
2. Caffery B, Lee JS, Alexander-Bryant AA. Vectors for Glioblastoma Gene Therapy: Viral & Non-Viral Delivery Strategies. Nanomaterials (Basel). 2019; 9:9. https://doi.org/10.3390/nano9010105 PMID:30654536
3. Fernandes GF, Fernandes BC, Valente V, Dos Santos JL. Recent advances in the discovery of small molecules targeting glioblastoma. Eur J Med Chem. 2019; 164:8–26. https://doi.org/10.1016/j.ejmech.2018.12.033 PMID:30583248
4. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, et al, and European Organisation for Research and Treatment of Cancer Brain Tumour and Radiation Oncology Groups, and National Cancer Institute of Canada Clinical Trials Group. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol. 2009; 10:459–66. https://doi.org/10.1016/S1470-2045(09)70025-7 PMID:19269895
5. Lee DH, Ryu HW, Won HR, Kwon SH. Advances in epigenetic glioblastoma therapy. Oncotarget. 2017; 8:18577–89. https://doi.org/10.18632/oncotarget.14612 PMID:28099914
6. Bo L, Wei B, Li C, Wang Z, Gao Z, Miao Z. Identification of potential key genes associated with glioblastoma based on the gene expression profile. Oncol Lett. 2017; 14:2045–52. https://doi.org/10.3892/ol.2017.6460 PMID:28789435
7. Zhang Y, Xia Q, Lin J. Identification of the potential oncogenes in glioblastoma based on bioinformatic analysis and elucidation of the underlying mechanisms. Oncol Rep. 2018; 40:715–25. https://doi.org/10.3892/or.2018.6483 PMID:29901201
8. Craig JM, Bickmore WA. The distribution of CpG islands in mammalian chromosomes. Nat Genet. 1994; 7:376–82. https://doi.org/10.1038/ng0794-376 PMID:7920655
9. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. Cancer Res. 2016; 76:3446–50. https://doi.org/10.1158/0008-5472.CAN-15-3278 PMID:27256564
10. Pan Y, Liu G, Zhou F, Su B, Li Y. DNA methylation profiles in cancer diagnosis and therapeutics. Clin Exp Med. 2018; 18:1–14. https://doi.org/10.1007/s10238-017-0467-0 PMID:28752221
11. Da Costa EM, McInnes G, Beaudry A, Raynal NJ. DNA Methylation-Targeted Drugs. Cancer J. 2017; 23: 270–76. https://doi.org/10.1097/PPO.0000000000000278 PMID:28926427
12. Wenger A, Ferreyra Vega S, Kling T, Bontell TO, Jakola AS, Carén H. Intratumor DNA methylation heterogeneity in glioblastoma: implications for DNA

methylation-based classification. Neuro-oncol. 2019; 21:616–27.
https://doi.org/10.1093/neuonc/noz011
PMID:30668814

13. Kanazawa T, Minami Y, Jinzaki M, Toda M, Yoshida K, Sasaki H. Predictive markers for MGMT promoter methylation in glioblastomas. Neurosurg Rev. 2019. [Epub ahead of print].
https://doi.org/10.1007/s10143-018-01061-5
PMID:30673883

14. Wang W, Zhang L, Wang Z, Yang F, Wang H, Liang T, Wu F, Lan Q, Wang J, Zhao J. A three-gene signature for prognosis in patients with MGMT promoter-methylated glioblastoma. Oncotarget. 2016; 7:69991–99.
https://doi.org/10.18632/oncotarget.11726
PMID:27588397

15. Wen WS, Hu SL, Ai Z, Mou L, Lu JM, Li S. Methylated of genes behaving as potential biomarkers in evaluating malignant degree of glioblastoma. J Cell Physiol. 2017; 232:3622–30.
https://doi.org/10.1002/jcp.25831
PMID:28145562

16. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, Pfaff E, Tönjes M, Sill M, Bender S, Kool M, Zapatka M, Becker N, et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell. 2012; 22:425–37.
https://doi.org/10.1016/j.ccr.2012.08.024
PMID:23079654

17. Taberlay PC, Jones PA. DNA methylation and cancer. Prog Drug Res. 2011; 67:1–23.
https://doi.org/10.1007/978-3-7643-8989-5_1
PMID:21141722

18. Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis. Int J Mol Sci. 2013; 14:15029–58.
https://doi.org/10.3390/ijms140715029
PMID:23873296

19. Berghoff AS, Hainfellner JA, Marosi C, Preusser M. Assessing MGMT methylation status and its current impact on treatment in glioblastoma. CNS Oncol. 2015; 4:47–52.
https://doi.org/10.2217/cns.14.50 PMID:25586425

20. Binabaj MM, Bahrami A, ShahidSales S, Joodi M, Joudi Mashhad M, Hassanian SM, Anvari K, Avan A. The prognostic value of MGMT promoter methylation in glioblastoma: A meta-analysis of clinical trials. J Cell Physiol. 2018; 233:378–86.
https://doi.org/10.1002/jcp.25896 PMID:28266716

21. Rapkins RW, Wang F, Nguyen HN, Cloughesy TF, Lai A, Ha W, Nowak AK, Hitchins MP, McDonald KL. The MGMT promoter SNP rs16906252 is a risk factor for MGMT methylation in glioblastoma and is predictive of response to temozolomide. Neuro Oncol. 2015; 17:1589–98.
https://doi.org/10.1093/neuonc/nov064
PMID:25910840

22. Trabelsi S, Mama N, Ladib M, Karmeni N, Haddaji Mastouri M, Chourabi M, Mokni M, Tlili K, Krifa H, Yacoubi MT, Saad A, H'Mida Ben Brahim D. MGMT methylation assessment in glioblastoma: MS-MLPA versus human methylation 450K beadchip array and immunohistochemistry. Clin Transl Oncol. 2016; 18:391–7.
https://doi.org/10.1007/s12094-015-1381-0
PMID:26289551

23. Ma X, Shang F, Zhu W, Lin Q. CXCR4 expression varies significantly among different subtypes of glioblastoma multiforme (GBM) and its low expression or hypermethylation might predict favorable overall survival. Expert Rev Neurother. 2017; 17:941–46.
https://doi.org/10.1080/14737175.2017.1351299
PMID:28685624

24. Zhong S, Jiang S, Peng Y, Chen Y. Further Investigation About Copy Number Variations and Methylation of AURKA, KIF4A, and NUSAP1 in Glioblastoma. World Neurosurg. 2018; 110:513–14.
https://doi.org/10.1016/j.wneu.2017.11.180
PMID:29433173

25. Hu HM, Chen Y, Liu L, Zhang CG, Wang W, Gong K, Huang Z, Guo MX, Li WX, Li W. C1orf61 acts as a tumor activator in human hepatocellular carcinoma and is associated with tumorigenesis and metastasis. FASEB J. 2013; 27:163–73.
https://doi.org/10.1096/fj.12-216622
PMID:23012322

26. Xu J, Zhang A, Zhang Z, Wang P, Qian Y, He L, Shi H, Xing Q, Du J. DNA methylation levels of imprinted and nonimprinted genes DMRs associated with defective human spermatozoa. Andrologia. 2016; 48:939–47.
https://doi.org/10.1111/and.12535 PMID:26804237

27. Mostafavi N, Vermeulen R, Ghantous A, Hoek G, Probst-Hensch N, Herceg Z, Tarallo S, Naccarati A, Kleinjans JC, Imboden M, Jeong A, Morley D, Amaral AF, et al. Acute changes in DNA methylation in relation to 24 h personal air pollution exposure measurements: A panel study in four European countries. Environ Int. 2018; 120:11–21.
https://doi.org/10.1016/j.envint.2018.07.026
PMID:30055357

28. Kolarova J, Tangen I, Bens S, Gillessen-Kaesbach G, Gutwein J, Kautza M, Rydzanicz M, Stephani U, Siebert R, Ammerpohl O, Caliebe A. Array-based DNA

methylation analysis in individuals with developmental delay/intellectual disability and normal molecular karyotype. Eur J Med Genet. 2015; 58:419–25. https://doi.org/10.1016/j.ejmg.2015.05.001 PMID:26003415

29. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, Flagg K, Hou J, Zhang H, Yi S, Jafari M, Lin D, Chung C, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci USA. 2017; 114:7414–19. https://doi.org/10.1073/pnas.1703577114 PMID:28652331

30. Wouters J, Vizoso M, Martinez-Cardus A, Carmona FJ, Govaere O, Laguna T, Joseph J, Dynoodt P, Aura C, Foth M, Cloots R, van den Hurk K, Balint B, et al. Comprehensive DNA methylation study identifies novel progression-related and prognostic markers for cutaneous melanoma. BMC Med. 2017; 15:101. https://doi.org/10.1186/s12916-017-0851-3 PMID:28578692

31. Shigaki H, Baba Y, Watanabe M, Murata A, Iwagami S, Miyake K, Ishimoto T, Iwatsuki M, Baba H. LINE-1 hypomethylation in gastric cancer, detected by bisulfite pyrosequencing, is associated with poor prognosis. Gastric Cancer. 2013; 16:480–87. https://doi.org/10.1007/s10120-012-0209-7 PMID:23179365

32. Iwagami S, Baba Y, Watanabe M, Shigaki H, Miyake K, Ishimoto T, Iwatsuki M, Sakamaki K, Ohashi Y, Baba H. LINE-1 hypomethylation is associated with a poor prognosis among patients with curatively resected esophageal squamous cell carcinoma. Ann Surg. 2013; 257:449–55. https://doi.org/10.1097/SLA.0b013e31826d8602 PMID:23023202

33. Harada K, Baba Y, Ishimoto T, Chikamoto A, Kosumi K, Hayashi H, Nitta H, Hashimoto D, Beppu T, Baba H. LINE-1 methylation level and patient prognosis in a database of 208 hepatocellular carcinomas. Ann Surg Oncol. 2015; 22:1280–87. https://doi.org/10.1245/s10434-014-4134-3 PMID:25319577

34. Hu S, Yin X, Zhang G, Meng F. Identification of DNA methylation signature to predict prognosis in gastric adenocarcinoma. J Cell Biochem. 2019; 120:11708–15. https://doi.org/10.1002/jcb.28450 PMID:30775802

35. Ma J, Li R, Wang J. Characterization of a prognostic four-gene methylation signature associated with radiotherapy for head and neck squamous cell carcinoma. Mol Med Rep. 2019; 20:622–32. https://doi.org/10.3892/mmr.2019.10294 PMID:31180552

36. Guo W, Zhu L, Zhu R, Chen Q, Wang Q, Chen JQ. A four-DNA methylation biomarker is a superior predictor of survival of patients with cutaneous melanoma. eLife. 2019; 8. https://doi.org/10.7554/eLife.44310 PMID:31169496

37. Wang Q, He Z, Chen Y. Comprehensive Analysis Reveals a 4-Gene Signature in Predicting Response to Temozolomide in Low-Grade Glioma Patients. Cancer Contol. 2019; 26:1073274819855118. https://doi.org/10.1177/1073274819855118 PMID:31167546

38. Wang Y, Deng H, Xin S, Zhang K, Shi R, Bao X. Prognostic and Predictive Value of Three DNA Methylation Signatures in Lung Adenocarcinoma. Front Genet. 2019; 10:349. https://doi.org/10.3389/fgene.2019.00349 PMID:31105737

39. Cancer Genome Atlas Research N. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013; 45:1113–20. https://doi.org/10.1038/ng.2764 PMID:24071849

40. Zhang S, Wang Y, Gu Y, Zhu J, Ci C, Guo Z, Chen C, Wei Y, Lv W, Liu H, Zhang D, Zhang Y. Specific breast cancer prognosis-subtype distinctions based on DNA methylation patterns. Mol Oncol. 2018; 12:1047–60. https://doi.org/10.1002/1878-0261.12309 PMID:29675884

SUPPLEMENTARY MATERIALS

Please browse Full Text version to see the data of Supplementary Tables 1–10:

Supplementary Table 1. The clinical information of training set.

Supplementary Table 2. The clinical information of testing set.

Supplementary Table 3. The results of significant DEGs of glioblastoma.

Supplementary Table 4. Survival-methylation expression profiles for further analysis.

Supplementary Table 5. The results of significant DEGs of glioblastoma.

Supplementary Table 6. Survival-methylation expression profiles for further analysis.

Supplementary Table 7. The information of differentially expressed genes and differentially methylated genes.

Supplementary Table 8. The results of EI and ES genes.

Supplementary Table 9. The expression profiles of 20 methylation sites in GSE36278.

Supplementary Table 10. The clinical information of GSE36278 and testing set.