Genome-Wide Identification and Analysis of the Methylation of lncRNAs and Prognostic Implications in the Glioma

Yijie He¹, Lidan Wang¹,², Jing Tang¹,²* and Zhijie Han¹*

¹ Department of Bioinformatics, School of Basic Medicine, Chongqing Medical University, Chongqing, China, ² Joint International Research Laboratory of Reproduction & Development, Chongqing Medical University, Chongqing, China

Glioma is characterized by rapid cell proliferation and extensive infiltration among brain tissues, but the molecular pathology has been still poorly understood. Previous studies found that DNA methylation modifications play a key role in contributing to the pathogenesis of glioma. On the other hand, long noncoding RNAs (lncRNAs) has been discovered to be associated with some key tumorigenic processes of glioma. Moreover, genomic methylation can influence expression and functions of lncRNAs, which contributes to the pathogenesis of many complex diseases. However, to date, no systematic study has been performed to detect the methylation of lncRNAs and its influences in glioma on a genome-wide scale. Here, we selected the methylation data, clinical information, expression of lncRNAs, and DNA methylation regulatory proteins of 537 glioma patients from TCGA and TANRIC databases. Then, we performed a differential analysis of IncRNA expression and methylated regions between low-grade glioma (LGG) and glioblastoma multiform (GBM) subjects, respectively. Next, we further identified and verified potential key lncRNAs contributing the pathogenesis of glioma involved in methylation modifications by an annotation and correlation analysis, respectively. In total, 18 such lncRNAs were identified, and 7 of them have been demonstrated to be functionally linked to the pathogenesis of glioma by previous studies. Finally, by the univariate Cox regression, LASSO regression, clinical correlation, and survival analysis, we found that all these 18 lncRNAs are high-risk factors for clinical prognosis of glioma. In summary, this study provided a strategy to explore the influence of IncRNA methylation on glioma, and our findings will be benefit to improve understanding of its pathogenesis.

Keywords: glioma, methylation modification, long non-coding RNAs, clinical prognosis, the cancer genome atlas (TCGA)
INTRODUCTION

Glioma is the most common and highly malignant tumor in the intraparenchymal central nervous system (CNS) tumors (1). It is characterized by the rapid and extensive proliferation among brain tissues (2, 3). The high grade glioma subtype, glioblastoma multiform (GBM), could cause the significant mortality that are disproportionate to their relatively rare incidence (4). Even under the best treatment, the median survival time is just over a year, and the few GBM patients survive more than 3 years (1). The etiology and pathogenesis of GBM have been extensively investigated, but the epigenetic mechanisms contributing to its pathogenesis were much less understood (2, 3, 5).

To date, DNA methylation is the most widely studied epigenetic mechanisms (6). Tremendous evidences shows that the DNA methylation is involved in tumorigenesis and development of the GBM (1, 7). For example, the promoter DNA methylation pattern of genes involved in RB1 and TP53 signaling pathways were identified in GBM patients (7). The promoter methylation of the DNA repair enzyme (O6-methylguanine-DNA methyltransferase) was discovered as a significant prognostic factor for temozolomide resistance in GBM patients (8).

The long non-coding RNAs (lncRNAs), a kind of non-protein coding transcripts of >200 nucleotides (9–11), has been reported to be a key regulator in a broad range of biological and cellular processes of GBM, including cell proliferation, motility, hypoxia response, and apoptosis (12–14). The expression levels and functions of lncRNAs could be significantly affected by the genomic methylation in many complex diseases (15–18). Moreover, there is increasing evidences that the methyltransferase, demethylase, and binding protein dynamically regulate the methylation level of the lncRNAs, which influences their expression in specific biological processes (18, 19). However, to date, no systematic study has been conducted to discovery the methylation of lncRNAs and its influences in the glioma on a genome-wide scale.

Herein, to address this lack of knowledge, we used a cohort of low-grade glioma (LGG) and GBM from The Cancer Genome Atlas (TCGA) database to investigate the contribution of lncRNA methylation to tumorigenesis and development in glioma. Specifically, we first downloaded the expression data of lncRNAs from The Atlas of Noncoding RNAs in Cancer (TANRIC) database, and then implemented a differential expression analysis between the LGG and GBM subjects. Second, we obtained the glioma-related methylation array data and the protein-coding gene expression data of the same samples from TCGA database, and then identified the differentially methylated regions of the differentially expressed lncRNAs according the GENCODE reference annotation for human genomes. Third, we conducted a correlation analysis between methylation level and expression of the lncRNAs and the genes involving in the three kinds of methylation regulatory proteins, and identified the potential key lncRNAs contributing the pathogenesis of glioma. Finally, we conducted the univariate Cox regression, least absolute shrinkage and selection operator (LASSO) regression, clinical correlation, and survival analysis based on the clinical data of these samples to explore the influence of these methylated and potentially disease-related lncRNAs on clinical prognosis of glioma. The flow chart was shown in Figure 1.

MATERIALS AND METHODS

Data Collection and Preprocessing

The clinical information and the methylation information of patients with glioma were downloaded from the TCGA database (http://cancergenome.nih.gov), a comprehensive resource for investigating the molecular basis in various cancers. According to TCGA annotation, glioma is classified as the LGG and the GBM. The Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/) was used to access these TCGA data. Particularly, we selected “DNA methylation” in the Data Category, “Illumina human methylation 450” in the Platform, “brain” in the Primary Site, and “gliomas” in the Disease Type to screen out the methylation information of patients. Then, we selected “clinical” in the Data Category, “brain” in the Primary Site, and “gliomas” in the Disease Type to screen out the clinical information of patients. Next, we removed the samples which lack the methylation or clinical information. Finally, the lncRNA expression data of the same patients was downloaded from the TANRIC database, which quantified the expression profiles of lncRNAs in Ensembl using the TCGA data (20).

Moreover, we further searched all the possible studies in PubMed database (http://www.ncbi.nlm.nih.gov/pubmed) using the keywords to “methylase gene,” “methyltransferase gene,” “binding protein gene,” “demethylase gene” to identify the DNA methylation regulatory proteins. The search was performed before the last update of this database on May 13, 2020. The gene expression of the methylation regulatory proteins was obtained from TCGA database. The expression data of lncRNAs and methylation regulatory proteins have been normalized as reads per kilobase of exon per million mapped reads (RPKM) and RNA-Seq by Expectation-Maximization (RSEM), respectively. The DNA methylation values were normalized using the “betaqn” function of the R package “wateRmelon” (http://bioconductor.org/packages/release/bioc/html/wateRmelon.html) (21).

Differential Expression Analysis of lncRNAs Between Low-Grade Glioma and Glioblastoma Multiform

To identify the key lncRNAs which are potentially associated with the gliomas progression, we performed a differential expression analysis of all the lncRNAs obtained from TANRIC database between LGG and GBM subjects using the R package “limma” with its default parameter settings (i.e. link.function = “log,” simulated.pvalue = FALSE, permutation = 100) (https://
CRAN.R-project.org/package=lncDIFF). It is a powerful differential analysis tool by the normalized expression data (e.g. RPKM values) as the input, and has high sensitivity to identify the low abundant differentially expressed genes, as commonly observed in lncRNAs. This package adopts the generalized linear model with zero-inflated exponential quasi-likelihood to estimate group effect on normalized counts, and employs the likelihood ratio test to detect differential expressed genes. The proposed method and tool are applicable to data processed with standard RNA-seq preprocessing and normalization pipelines (22). We first removed 21 lncRNAs whose expression is zero in all the LGG or GBM subjects. Then, we set significance level according to the common threshold of the absolute value of fold change ($FC \geq 2$ and false discovery rate (FDR) $p < 0.05$. The $p$ values are corrected for multiple testing by Benjamini–Hochberg method. Finally, we used a volcano plot to describe the profile of whole lncRNA expression by the R package “ggplot2” (https://CRAN.R-project.org/package=ggplot2), and used a heatmap to visualize the cluster pattern of the differentially expressed lncRNAs based on Manhattan distance by the R package “gplots” (https://CRAN.R-project.org/package=gplots).

**Differential Methylation Analysis and lncRNA Annotation**

To identify the glioma-related methylation positions and regions, and the differentially expressed lncRNAs located in these regions, we performed a differential methylation analysis and lncRNA annotation. The differential methylation analysis was conducted by the R package “minfi” which is a specialized tool designed to process the Illumina methylation 450 array data (http://bioconductor.org/packages/release/bioc/html/minfi.html). It used the Subset-quantile Within Array Normalization method to preprocess data and the bump-hunting algorithm to discover the differential methylation information (23). Firstly, we used the
“densityBeanPlot” function of this package to conduct the quality control for each array. The qualified samples should have the characteristic that the methylation levels (beta values) of CpG positions are distributed around 0 and 1, respectively. Then, we used the “dmpFinder” function (type = “categorical”) of this package to identify the differentially methylated positions between LGG and GBM subjects based on the methylation array. The significance level was set according to a common threshold of the absolute intercept $\geq 0.2$ (i.e. 20% difference on the beta values) and $p < 1 \times 10^{-3}$ (24). Next, based on these differentially methylated positions, we further used the “bumphunter” function ($B = 10$, type = “Beta”) of this package to look for the differentially methylated regions between LGG and GBM subjects with the common threshold of average methylation level difference $\geq 0.2$ (25, 26). The differentially methylated regions are the consecutive genomic locations containing a battery of differentially methylated positions in the same direction. Finally, we download the ensGene annotation file (hg19) from the Ensembl (release 75) which stores the location information of lncRNA transcripts and exons in human genome. Based on this file, the ANNOVAR software was used to perform an lncRNA annotation and identify the differentially expressed lncRNAs located in the differentially methylated regions. ANNOVAR is a Perl command-line tool for rapidly and efficiently annotating the genomic variants, including gene-based, region-based and filter-based annotations on a variant call format (VCF) file generated from human genomes (27).

**Correlation Analysis Between Methylation and Expression of lncRNAs**

To explore the influence of methylation on the corresponding lncRNAs, and identify the potential key lncRNAs contributing the pathogenesis of glioma, we performed a correlation analysis between methylation and expression of lncRNAs. Particularly, we first selected the differentially methylated positions to be included in each of the identified lncRNAs in the previous step, and calculated the average values of these methylation positions for each lncRNAs, respectively. Then, we calculated the Pearson’s correlation coefficient between the expression of these lncRNAs and their average methylation level using the R function “cor.test.” The threshold of significance was set at the absolute value of $r > 0.6$ and FDR $p < 0.05$. The p values are corrected for multiple testing by Benjamini–Hochberg method. Finally, it is reported that the methylation regulatory proteins (including methyltransferase, demethylase, and binding protein) dynamically regulate the methylation level of lncRNAs, which influences their expression in specific biological processes (18, 19). Therefore, to explore which methylation regulatory proteins are involved in the methylation modification of the potential key lncRNAs and further increase the reliability of our findings, we selected the known methylation regulatory proteins and calculated the Pearson’s correlation coefficient between their expression and the average methylation level of these lncRNAs using the same significance threshold.

**Influence of the Methylated lncRNAs on Clinical Prognosis of Glioma**

We further analyzed the influence of these identified methylated lncRNAs potentially contributing the pathogenesis of glioma on the clinical prognosis of glioma. First, we calculated the average expression of the key lncRNAs obtained above in each patient and get the median of these average expressions. According to the median, the patients were separated into the lncRNAs low and high expression groups. We compared prognosis between the high expression and low expression subjects using a Kaplan-Meier overall survival curves. Then, we performed a univariate Cox regression analysis to assess the association between these methylated lncRNAs and the prognosis of glioma. The threshold of significance was set at 95% confidence interval (CI) of hazard ratio (HR) $\geq 1$ and $p < 0.05$. The R package “survival” (https://CRAN.R-project.org/package=survival) was used for these analyses. Next, based on the results of univariate Cox regression analysis, we further used the LASSO regression algorithm to identify the key lncRNAs whose methylation and expression impact on the prognosis of glioma by R package “glmnet.” It is a pathwise algorithm for the Cox proportional hazards model, regularized by convex combinations of $\ell_0$ and $\ell_1$ penalties (elastic net). The algorithm fits via cyclical coordinate descent, and employs warm starts to find a solution along a regularization path (28). The parameter family, maxit, and alpha were set to Cox, 1000 and 1, respectively (others were set by their default values). And then we calculated the risk score of each subject using them through the “survival” package. Finally, we used a receiver operator characteristic (ROC) curve to verify the reliability of the risk score by the R package “survivalROC” (https://CRAN.R-project.org/package=survivalROC). In addition, we also assessed the association between these lncRNA expressions and other clinical features of the patients (including age at initial pathologic diagnosis, vital status, and gender) using the chi-square test. The threshold of significance was set at $p < 0.05$.

**RESULTS AND DISCUSSION**

**Methylation, Expression, and Clinical Information of 537 Glioma Samples**

After the data collection, we found a total of 537 glioma samples (including 486 LGG and 51 GBM patients from TCGA) with the DNA methylation values, expression levels of protein-coding genes, and clinical information. Particularly, according to the annotation of Illumina human methylation 450 array, a total of 369,531 CpGs methylation positions were quantified after removing the missing values. We normalized the CpGs methylation values for the subsequent analyses. The results were shown in **Supplementary Figure S1**. The lncRNA expression data of the 537 glioma samples were obtained from TANRIC database. A total of 12,727 lncRNAs of these samples were quantified as RPKM values. Through the keyword search and the title/abstract screening, 23 articles containing genes for methylation-related enzymes were obtained from PubMed.
In total, we identified 32 DNA methylation regulatory proteins (including methyltransferase, demethylase, and binding protein) from the 23 articles (Table 1). We extracted the expression data of these 32 methylation regulatory proteins for each sample (quantified as RSEM values) from the TCGA database. The clinical information of these samples contains age, gender, survival time, and vital status. The summary of these glioma samples was listed in Table 2.

**Differential Expression Analysis of lncRNAs Between Low-Grade Glioma and Glioblastoma Multiforme**

We used the R package “lncDIFF” to perform the differential expression analysis of lncRNAs between LGG and GBM subjects according to the significance threshold of $|FC| \geq 2$ and FDR $p < 0.05$. In total, we identified 1,988 significantly differentially expressed lncRNAs, which include 1,284 highly expressed (i.e. $FC \geq 2$) and 704 lowly expressed lncRNAs ($FC \leq -2$) in the GBM subjects. The details are described in Supplementary Table S1. We used a volcano plot to describe the profile of whole lncRNA expression (Figure 2A). Then, to verify these findings, we contrasted these identified differentially expressed lncRNAs with another independent study. This study used 19 glioblastoma and 9 control brain samples to perform the differential expression analysis of 30,856 lncRNA transcripts (Arraystar Human lncRNA Microarray V3.0, nearly 30% of them overlap with our study). According to its results, about 71.5% differentially expressed lncRNAs overlapped with our findings (52). Further, we selected the top most 100 differentially expressed lncRNAs to visualize the cluster pattern of their expression by a heatmap. As Figure 2B shows, the GBM and LGG subjects are mainly grouped under a cluster according to high

### TABLE 1 | The information of the 32 DNA methylation regulatory proteins.

| Gene   | ID   | Description                      | Type             | Reference |
|--------|------|----------------------------------|------------------|-----------|
| DNMT3A | 1788 | DNA methyltransferase 3 alpha    | Methyltransferase | (29, 30, 31) |
| DNMT3B | 1789 | DNA methyltransferase 3 beta     | Methyltransferase | (29, 30, 31) |
| DNMT3L | 29947| DNA methyltransferase 3 like     | Methyltransferase | (30)      |
| DNMT1  | 1786 | DNA methyltransferase 1          | Methyltransferase | (31–33)   |
| DMAP1  | 55929| DNA methyltransferase 1 associated protein 1 | Binding protein | (33)      |
| SUV39H1| 6839 | Suppressor of variegation 3-9 homolog 1 | Methyltransferase | (34)      |
| MESP2  | 4204 | Methyl-CpG binding protein 2     | Binding protein   | (35)      |
| MB1D1  | 4152 | Methyl-CpG binding domain protein 1 | Binding protein   | (35)      |
| MB2D2  | 8932 | Methyl-CpG binding domain protein 2 | Binding protein   | (35)      |
| MB3D3  | 53615| Methyl-CpG binding domain protein 3 | Binding protein   | (35)      |
| MB4D   | 8930 | Methyl-CpG binding domain 4, DNA glycosylase | Binding protein | (35)      |
| SETDB1 | 9869 | SET domain bifurcated histone lysine methyltransferase 1 | Methyltransferase | (31, 33) |
| MGMT   | 4255 | O-6-methylguanine-DNA methyltransferase | Methyltransferase | (36)      |
| TET1   | 80312| Tet methylcytosine dioxygenase 1  | Demethylase       | (37)      |
| TET2   | 54790| Tet methylcytosine dioxygenase 2  | Demethylase       | (37)      |
| TET3   | 200424| Tet methylcytosine dioxygenase 3  | Demethylase       | (37)      |
| JMJD6  | 23210| Jumonj domain containing 6, arginine demethylase and lysine hydroxylase | Demethylase       | (38)      |
| KDM5A  | 55818| Lysine demethylase 5a            | Demethylase       | (39)      |
| KDM5C  | 8242 | Lysine demethylase 5c            | Demethylase       | (39)      |
| KDM1A  | 23028| Lysine demethylase 1a            | Demethylase       | (40)      |
| KDM5B  | 10765| Lysine demethylase 5b            | Demethylase       | (41)      |
| KDM5A  | 5927 | Lysine demethylase 5a            | Demethylase       | (42)      |
| KDM5D  | 8284 | Lysine demethylase 5d            | Demethylase       | (42)      |
| KDM3B  | 51780| Lysine demethylase 3b            | Demethylase       | (43)      |
| KDM4A  | 9682 | Lysine demethylase 4a            | Demethylase       | (44, 45) |
| KDM4B  | 23030| Lysine demethylase 4b            | Demethylase       | (46)      |
| KDM4C  | 23081| Lysine demethylase 4c            | Demethylase       | (47)      |
| KDM4D  | 55693| Lysine demethylase 4d            | Demethylase       | (48)      |
| KDM6A  | 7403 | Lysine demethylase 6a            | Demethylase       | (42, 49) |
| KDM6B  | 23135| Lysine demethylase 6b            | Demethylase       | (42, 49) |
| KDM2A  | 22992| Lysine demethylase 2a            | Demethylase       | (48)      |
| KDM2B  | 84678| Lysine demethylase 2b            | Demethylase       | (50, 51) |

### TABLE 2 | Summary of the 537 individuals studied in this work.

| Individuals | Sample Type     | Sample Size | Mean Age (SD) | Male/Female (%) | Death Rates (%) |
|-------------|-----------------|-------------|---------------|-----------------|-----------------|
| GBM subjects| Primary Tumor   | 51          | 61.54 (13.41) | 56.00 / 44.00   | 66.00           |
| LGG subjects| Primary Tumor   | 486         | 42.91 (13.42) | 54.64 / 45.36   | 25.15           |
| Total       |                 | 537         | 44.66 (14.48) | 54.77 / 45.23   | 28.97           |
and low expression of the lncRNAs, respectively, and most of these lncRNAs (83%) are significantly highly expressed in the GBM than LGG subjects. These differentially expressed lncRNAs may contribute to the progress of glioma. Thus, we used these significantly differentially expressed lncRNAs to conduct the subsequent analysis.
Differential Methylation Analysis and lncRNA Annotation

We performed a differential methylation analysis and lncRNA annotation to identify the glioma-related DNA methylation and the differentially expressed lncRNAs located in the regions. The results of array quality control showed that the beta values of DNA methylation positions are mainly distributed around 0 and 1, respectively, for each sample (Supplementary Figure S2). Then, we used these arrays to identify the differentially methylated positions and regions, respectively. The results showed that there are a total of 208,138 positions and 13,227 corresponding regions with a significantly differential methylation level between LGG and GBM subjects (Supplementary Tables S2, S3). The methylation array GSE90496 (contains 347 GBM and 301 LGG subjects) was used to verify these findings. These results showed that about 71.1% differential methylation positions are consistent with our findings (53). Finally, based on these differentially methylated regions, we performed the location annotation of the differentially expressed lncRNAs. In total, we identified 744 lncRNAs which are located in the differentially methylated regions. According to the results of annotation, these differentially methylated regions are at five categories of different genomic locations, i.e. intergenic, ncRNA_exonic, ncRNA_intronic, upstream and downstream, and the proportion of intergenic areas is significantly increased compared with other types (Supplementary Table S4). We further calculated the proportion of them in the different genomic locations. We found that these lncRNAs are mainly distributed in chromosome 1, 2, 7, and 12 (11.42, 10.48, 8.06, and 8.06%, respectively) (Supplementary Table S5). Moreover, as the Supplementary Table S1 shown, these identified lncRNAs include 1,284 highly and 704 lowly expressed ones in the GBM subjects. But not all of these highly and lowly expressed lncRNAs have significantly reduced and increased methylation levels in the GBM subjects, respectively, which imply that not all of DNA methylation changes can affect the expression of lncRNAs in the corresponding genomic regions.

Correlation Analysis Between Methylation and Expression of lncRNAs

We first conducted a Shapiro-Wilk normality test for each vector by the R function “shapiro.test.” According to the threshold $P > 0.05$, 11 lncRNAs that do not obey the normal distribution were removed. Then, to identify the differentially expressed lncRNAs affected by the glioma-related DNA methylation, we performed a Pearson’s correlation analysis between methylation and expression of lncRNAs. The results revealed that there are a total of 18 lncRNAs (including 16 highly and 2 lowly expressed ones) whose expression is significantly associated with their DNA methylation level, and all of them show a significant negative correlation ($r < -0.6$ and FDR $p < 0.05$) (Table 3). It is consistent with the common understanding that DNA methylation inhibits the corresponding gene expression in a variety of tissues and cell lines (54, 55). Next, we further measured the association between the expression of the 32 methylation regulatory proteins and methylation level of these lncRNAs. The results showed that there is a significantly negative correlation between TET1 expression and most of the 18 lncRNAs’ methylation level. TET1 is a demethylase which can catalyze the conversion of 5-methylcytosine to 5-hydroxymethylcytosine and maintains hypomethylation status of the corresponding regions (37). Besides this gene, the expression of KDM4B and MBD2 also show a significantly negative and positive correlation ($r > 0.6$ and FDR $p < 0.05$) with a part of the 18 lncRNAs’ methylation level, respectively. KDM4B is also a demethylase of histone lysine by a hypoxia-induced pathway, and an important epigenetic modifer in cancer (46). MBD2 is a methyl-CpG binding protein which binds and maintains methylated gene promoter to repress its transcriptional activity (35). However, this significant correlation is not observed in the other 29 genes, which imply that the DNA methylation of lncRNAs in glioma may be influenced predominantly by some specific methylation regulatory proteins (Table 3 and Supplementary Table S5). Moreover, the mean differential methylation of the 18 lncRNAs was calculated. We found that all of the lowly expressed lncRNAs have significantly reduced methylation levels and most of them have significantly reduced methylation levels in the GBM subjects. It is also consistent with the understanding that DNA methylation inhibits gene expression (56, 57). Therefore, we considered them as the potential key lncRNAs contributing the pathogenesis of glioma involved in methylation modifications. Finally, we further queried PubMed for the functions of the 18 potential key lncRNAs contributing the pathogenesis of glioma. We found that seven of them have been demonstrated to be functionally linked to the pathogenesis of glioma by the previous studies. For example, the overexpression of the ENSG00000222401 (CYTOR) partially reversed the inhibitory effects of UPF1 on proliferation and invasion abilities in glioma (58, 59). The more details were described in the Supplementary Table S6.

Influence of the Methylated lncRNAs on Clinical Prognosis of Glioma

We further analyzed the influence of the 18 identified lncRNAs, which potentially contribute the glioma pathogenesis by methylation modifications, on the clinical prognosis of glioma. For the 16 lncRNAs highly expressed in GBM patients, we found that the overall survival curve of the subjects with high lncRNA expression is significantly longer than the subjects with low lncRNA expression ($p = 1.38 \times 10^{-10}$) (Figure 3A). On the contrary, the overall survival curve of the subjects with low lncRNA expression is significantly longer than the subjects with high lncRNA expression for the two lowly expressed lncRNAs ($p = 3.11 \times 10^{-10}$) (Figure 3B). It reflects an association between the dysregulation of lncRNA expression and a bad prognosis of glioma patients. To avoid dependence on the tumor grade, we performed the univariate Cox regression analysis of the 18 lncRNAs in GBM and LGG subjects, respectively. We did not find a significant association between lncRNA expression and poor patient outcomes in GBM subjects. However, the results showed that all of the 18 identified methylated lncRNAs are high-risk factors for the prognosis of glioma in LGG subjects (i.e. 95% CI HR $\geq 1$ and $p < 0.001$) (Figure 3C). This suggests that both over-expression of those 16 lncRNAs and under-expression of other 2 ones can
We found that the GBM subjects are mainly distributed in the key lncRNAs is significantly more than that of LGG in the low-risk group. This demonstrates the consistency between high-risk group, while the LGG subjects are mainly distributed in low-risk group. In total, we identified these 18 lncRNAs which has the following four characteristics: 1) they are located in the specific sequences; 2) at least one of the differentially methylated regions, significantly differentially expressed between the LGG and GBM subjects; 3) there is a strong correlation between the methylation level of these lncRNAs and the expression of lncRNAs contributing the pathogenesis of glioma involved in the clinical prognosis of glioma. In this study, we used the TCGA data to identify potential key lncRNAs contributing the pathogenesis of glioma involved in methylation modifications and further explore the influence of them on the clinical prognosis of glioma. In total, we identified 18 such lncRNAs which has the following four characteristics: 1) they are significantly differentially expressed between the LGG and GBM subjects; 2) at least one of the differentially methylated regions, which cover the contiguous differentially methylated positions, is located in these lncRNA sequences; 3) there is a strong correlation between the methylation level of these lncRNAs and the expression of lncRNA regulatory proteins; 4) the expression of these lncRNAs is significantly associated with the methylation level. Further, the results of clinical data analysis show that all these 18 lncRNAs are high-risk factors for the clinical prognosis of glioma, and four of them (i.e. ENSG00000256802, ENSG00000232333, ENSG00000227372, ENSG00000222041) are selected when the cross-validated partial likelihood deviance reaches its minimum value, and the coefficients of all these lncRNAs are positive (i.e. increase risk of disease). The area under the curve (AUC) of the ROC is 0.903, which shows the reliability of the risk score (Figure 3F). According to the median of risk scores, the patients were separated into low and high-risk groups. We found that the GBM subjects are mainly located in high-risk group, while the LGG subjects are mainly distributed in low-risk group. This demonstrates the consistency between the sample risk score by the key lncRNAs and the severity of glioma. Moreover, as Figure 4A shows, the risk classification by the key lncRNAs is significantly associated with the age at initial pathologic diagnosis ($p = 1.32 \times 10^{-2}$) and vital status ($p = 1.72 \times 10^{-8}$). But we observed no association with the gender of the patients ($p = 1.97 \times 10^{-3}$). The similar results were also observed for all the 18 identified methylated lncRNAs ($p$ value of age, vital status and gender is $2.87 \times 10^{-14}$, $2.01 \times 10^{-9}$, and $1.45 \times 10^{-1}$, respectively) (Figure 4B).

CONCLUSIONS

In this study, we used the TCGA data to identify potential key lncRNAs contributing the pathogenesis of glioma involved in methylation modifications and further explore the influence of them on the clinical prognosis of glioma. In total, we identified 18 such lncRNAs which has the following four characteristics: 1) they are significantly differentially expressed between the LGG and GBM subjects; 2) at least one of the differentially methylated regions, which cover the contiguous differentially methylated positions, is located in these lncRNA sequences; 3) there is a strong correlation between the methylation level of these lncRNAs and the expression of lncRNA regulatory proteins; 4) the expression of these lncRNAs is significantly associated with their methylation level. Further, the results of clinical data analysis show that all these 18 lncRNAs are high-risk factors for the clinical prognosis of glioma, and four of them (i.e. ENSG00000256802, ENSG00000232333, ENSG00000227372, ENSG00000222041) are most important for the severity of glioma. All in all, we performed a strategy to explore the influence of the lncRNA methylation on the pathogenesis of glioma, and these findings will be benefit to further glioma research in the future.
FIGURE 3 | The influence of the methylated IncRNAs on the prognosis of glioma. **(A)** For the 16 IncRNAs highly expressed in GBM patients, the Kaplan-Meier overall survival curve of the subjects with high IncRNA expression is significantly longer than the subjects with low IncRNA expression. **(B)** For the two IncRNAs lowly expressed in GBM patients, the Kaplan-Meier overall survival curve of the subjects with low IncRNA expression is significantly longer than the subjects with high IncRNA expression. **(C)** The forest plot for the results of univariate Cox regression analysis in LGG. **(D)** The relationship between the partial likelihood deviance and the penalty coefficient λ value. The log(λ) is equal to about 2.5 when the partial likelihood deviance reaches its minimum value. **(E)** The LASSO regression for calculating the coefficient of each IncRNA. There are four IncRNAs with non-zero coefficients when the log(λ) is equal to 2.5. **(F)** The ROC curve shows the reliability of the risk score.

FIGURE 4 | The association between the methylated IncRNAs and the clinical features of glioma patients. **(A)** The risk classification by the four key methylated IncRNAs is significantly associated with the age at initial pathologic diagnosis and vital status, but not with the gender of the patients. **(B)** The 18 glioma-related IncRNAs are significantly differentially expressed between GBM and LGG groups, which are also significantly associated with the age and vital status, but not with the gender.
DATA AVAILABILITY STATEMENT
Publiquely available datasets were analyzed in this study. This data can be found here: TCGA (http://cancergenome.nih.gov), GDC Data Portal (https://portal.gdc.cancer.gov/), and PubMed (http://www.ncbi.nlm.nih.gov/pubmed).

AUTHOR CONTRIBUTIONS
ZH designed the research. ZH, YH, JT, and LW collected the data. ZH and YH performed the research and analyzed data. ZH, JT, and YH wrote the paper. ZH and JT reviewed and modified the manuscript. All authors discussed the results and contributed to the final manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES
1. Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning KH, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nat Med* (2018) 24:1611–24. doi: 10.1038/s41591-018-0156-x
2. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M. Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neurol* (2006) 2:249–503. doi: 10.1038/nccneuro2899
3. Weller M, Wick W, Aldape K, Brada M, Berger M, Pfister SM, et al. Glioma. *Nat Rev Dis Primers* (2015) 1:1–18. doi: 10.1038/nrdp.2015.17
4. Ostrom QT, Gittleman H, Stetson L, Virk SM, Barnholtz-Sloan JS. Epidemiology of gliomas. *Cancer Treat Res* (2015) 163:1–14. doi: 10.1007/978-3-319-12048-5_1
5. Tang J, He D, Yang P, He J, Zhang Y. Genome-wide expression profiling of glioblastoma using a large combined cohort. *Sci Rep* (2018) 8:15104. doi: 10.1038/s41598-018-33323-z
6. Hwang JY, Aromolaran KA, Zukin RS. The emerging field of epigenetics in neurodegeneration and neuroprotection. *Nat Rev Neurosci* (2017) 18:347–61. doi: 10.1038/nrn.2017.46
7. Etchevery A, Aubry M, de Tayrac M, Vauleon E, Boniface R, Guenet F, et al. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics* (2010) 11:701. doi: 10.1186/1471-2164-11-701
8. Chen WJ, Zhang X, Han H, Ly JN, Kang EM, Zhang YL, et al. The different role of YKLI-40 in glioblastoma is a function of MGMT promoter methylation status. *Cell Death Dis* (2020) 11:668. doi: 10.1038/s41419-020-02909-9
9. Rion N, Ruegg MA. LncRNA-encoded peptides: More than translational noise? *Cell Res* (2017) 27:604–05. doi: 10.1038/cr.2017.35
10. Tang J, Wu X, Mou M, Wang C, Wang L, Li F, et al. GIMICA: host genetic and immune factors shaping human microbiota. *Nucleic Acids Res* (2020). doi: 10.1093/nar/gkaa851
11. Han Z, Xue W, Tao L, Lou Y, Qiu Y, Zhu F. Genome-wide identification and analysis of the eQTL lncRNAs in multiple sclerosis based on RNA-seq data. *Briefings Bioinf* (2020) 21:1023–37. doi: 10.1093/bib/bbz036
12. Li Q, Jia H, Li H, Dong C, Wang Y, Zou Z. LncRNA and mRNA expression profiles of glioblastoma multiforme (GBM) reveal the potential roles of IncRNAs in GBM pathogenesis. *Tumour Biol* (2016) 37:14537–52. doi: 10.1007/s13277-016-5289-0
13. Zhang X-Q, Leung GK-K. Long non-coding RNAs in glioma: functional roles and clinical perspectives. *Neurochem Int* (2014) 77:78–85. doi: 10.1016/j.neuint.2014.05.008
14. Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell* (2018) 172:393–407. doi: 10.1016/j.cell.2018.01.011
15. Wu Z, Liu X, Liu L, Deng H, Zhang J, Xu Q, et al. Regulation of IncRNA expression. *Cell Mol Biol Lett* (2014) 19:561. doi: 10.2478/cmbl-2014-0021-12
16. Heilmann K, Toth R, Bossmann C, Klimo K, Plass C, Gerhauser C. Genome-wide screen for differentially methylated long noncoding RNAs identifies Espr2 and IncRNA Espr2-as regulated by enhancer DNA methylation with prognostic relevance for human breast cancer. *Oncogene* (2017) 36:6446–61. doi: 10.1038/onc.2017.246
17. Dong Z, Zhang A, Liu S, Lu F, Guo Y, Zhang G, et al. Aberrant methylation-mediated silencing of IncRNA MEG3 functions as a ceRNA in esophageal cancer. *Med Cancer Res* (2017) 15:890–10. doi: 10.1185/1541-7786.MCR-16-0385
18. Wu SC, Kallin EM, Zhang Y. Role of H3K27 methylation in the regulation of IncRNA expression. *Cell Res* (2010) 20:1109–16. doi: 10.1038/cr.2010.114
19. Lui F, Shiekhattar R. Where long noncoding RNAs meet DNA methylation. *Cell Res* (2014) 24:263–4. doi: 10.1038/cr.2014.13
20. Li J, Han L, Roebuck P, Diao L, Liu L, Yuan Y, et al. TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer Res* (2015) 75:3728–37. doi: 10.1158/0008-5472.CAN-15-0273
21. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* (2015) 16:303. doi: 10.1186/s13059-015-0503-2
22. Li Q, Yu X, Chaudhary R, Siebos RJ, Chung CH, Wang X. *IncDiff*: a novel quasi-likelihood method for differential expression analysis of non-coding RNA. *BMC Genomics* (2019) 20:539. doi: 10.1186/s12864-019-5926-4
23. Aryee MJ, Iafrate AJ, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* (2014) 30:1363–9. doi: 10.1093/bioinformatics/btu049
24. Guo X, Xu Y, Zhao Z. In-depth genomic data analyses revealed complex transcriptional and epigenetic dysregulations of BRAFV600E in melanoma. *Mol Cancer* (2015) 14:60. doi: 10.1186/s12943-015-0328-y
25. Lu T, Klein KO, Colmenga I, Lora M, Greenwood CMT, Hudson M. Whole-genome bisulﬁte sequencing in systemic sclerosis provides novel targets to understand disease pathogenesis. *BMC Med Genomics* (2019) 12:144. doi: 10.1186/s12864-019-0662-9
26. Torabi Moghadam B, Etemadkhah M, Rajkovska G, Stockmeier C, Grabbhem M, Komorowski J, et al. Analyzing DNA methylation patterns in subjects diagnosed with schizophrenia using machine learning methods. *J Psychiatr Res* (2019) 114:41–7. doi: 10.1016/j.jpsychires.2019.04.001
27. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* (2015) 10:1556–66. doi: 10.1038/nprot.2015.105
28. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent. *J Stat Software* (2011) 39:1–13. doi: 10.18637/jss.v039.i01
29. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* (1999) 99:247–57. doi: 10.1016/s0092-8674(00)81656-6
30. Apola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, et al. Isolation and initial characterization of a novel zinc finger gene, DNM3, on 2q12.3, related to the cytokine-5-methyltransferase 3 gene family. *Genomics* (2000) 65:293–9. doi: 10.1006/geno.2000.6168
31. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* (2010) 464:927–31. doi: 10.1038/nature08858
32. Bewick AJ, Schmitz RJ. Gene body DNA methylation in plants. *Carr Opin Plant Biol* (2017) 36:103–10. doi: 10.1016/j.cob.2016.12.007

33. Gegner J, Gegner T, Vogel H, Vlichos A. Silencing of the DNA methyltransferase 1 associated protein 1 (DMAP1) gene in the invasive ladybird Harmonia axyridis implies a role of the DNA methyltransferase 1-DMAP1 complex in female fecundity. *Insect Mol Biol* (2020) 29:148–59. doi: 10.1111/imb.12616

34. Vandel L, Nicolas E, Vautre R, Ait-Si-Ali S, Trouche D. Transcriptional repression by the retinoblastoma protein at the recruitment of a histone methyltransferase. *Mol Cell Biol* (2001) 21:6484–94. doi: 10.1128/mcb.21.19.6484-6494.2001

35. Chakraborty A, Viswanathan P. Methylation-Demethylation Dynamics: Implications of Changes in Acute Kidney Injury. *Anal Cell Pathol* (2018) 2018:8764384. doi: 10.1155/2018/8764384

36. Verbeek B, Southgate TD, Gilham DE, Margison GP. O6-Methylguanine-DNA methyltransferase inactivation and chemotherapy. *Mol Carcinog* (2011) 50:931–8. doi: 10.1002/mc.20758

37. Ross SE, Bogdanovic O. TET enzymes, DNA demethylation and pluripotency. *Nat Rev Drug Discov* (2013) 12:917–21. doi: 10.1038/nrd4154

38. Hojfeldt JW, Agger K, Hedin K. Histone lysine demethylation represses transcription in vivo. *Nat Genet* (2006) 38:123–28. doi: 10.1038/ng1801

39. Kayser M, de Knijff P. Improving human forensics through advances in DNA methyltransferase 1 associated protein 1 (DMAP1) gene in the invasive ladybird Harmonia axyridis implies a role of the DNA methyltransferase 1-DMAP1 complex in female fecundity. *Insect Mol Biol* (2020) 29:148–59. doi: 10.1111/imb.12616

40. Han M, Xu W, Cheng P, Jin H, Wang X. Histone demethylase lysine demethylase 3B in development and cancer. *Mol Cell Biol* (2011) 31:4115–25. doi: 10.1016/j.mcb.2011.03.008

41. Jeffrey R, Robinson BD, Downes MJ, Powell LG, Lee MM, Scherr DS, et al. Role of androgen receptor and associated lysine-demethylase coregulators, LSD1 and JMJD2A, in localized and advanced human bladder cancer. *Mol Carcinog* (2011) 50:931–4. doi: 10.1002/mc.20758

42. Yang J, Harris AL, Davidoff AM. Hypoxia and Hormone-Mediated Pathways Converge at the Histone Demethylase KDM4B in Cancer. *Int J Mol Sci* (2018) 19:240. doi: 10.3390/ijms19010240

43. Liu G, Bollig-Fischer A, Kreike B, van de Vijver MJ, Abrams J, Ethier SP, et al. Genomic amplification and oncogenic properties of the GASC1 histone demethylase gene in breast cancer. *Oncogene* (2009) 28:4491–500. doi: 10.1038/onc.2009.297

44. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med* (2017) 49:e324. doi: 10.1038/emplm.2017.11

45. Jiang W, Wang J, Zhang Y. Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. *Cell Res* (2013) 23:122–30. doi: 10.1038/ cr.2012.119

46. Kotakis F, Polyarchou C, Foltopoulou P, Sanidas I, Kampranis SC, Tschilis PN, FGF-2 regulates cell proliferation, migration, and angiogenesis through an NDY1/KDM2B-miR-101-152Z2 pathway. *Mol Cell Biol* (2011) 43:285–98. doi: 10.1128/jmcb.2011.06.020

47. He J, Nguyen AT, Zhang Y. KDM2b/JHDH1, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* (2011) 117:3869–80. doi: 10.1182/blood-2010-10-312736

48. Capper D, Jones DTW, Still M, Hovestadt V, Schirmpl D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. *Nature* (2018) 555:469–74. doi: 10.1038/nature26000

49. Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H. DNA methylation represses transcription in vivo. *Nat Genet* (1999) 22:203–6. doi: 10.1038/79727

50. Azar A, Cedar H. DNA methylation and gene expression. *Microbiol Rev* (1991) 55:451–8. doi: 10.1128/MMBR.55.3.451-458.1991

51. Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H. The role of DNA methylation in setting up chromatin structure during development. *Nat Genet* (2003) 34:187–92. doi: 10.1038/ng1158

52. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* (2011) 12:179–92. doi: 10.1038/nrg2952

53. Liang J, Wei X, Liu Z, Cao D, Tang Y, Zou Z, et al. Long noncoding RNA (lncRNAs) in glioblastoma: identification of complex lncRNA regulatory networks and clinically relevant lncRNAs in glioblastoma. *Oncotarget* (2018) 9:29548. doi: 10.18632/oncotarget.25434

54. Patani N, Jiang WG, Newbold RF, Mokbel K, Heinecke AS, et al. Genetic landscape of long noncoding RNA (lncRNAs) in glioblastoma: identifications of complex lncRNA regulatory networks and clinically relevant lncRNAs in glioblastoma. *Oncotarget* (2018) 9:29548. doi: 10.18632/oncotarget.25434

55. Capper D, Jones DTW, Still M, Hovestadt V, Schirmpl D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. *Nature* (2018) 555:469–74. doi: 10.1038/nature26000

56. Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H. DNA methylation represses transcription in vivo. *Nat Genet* (1999) 22:203–6. doi: 10.1038/79727

57. Razin A, Cedar H. DNA methylation and gene expression. *Microbiol Rev* (1991) 55:451–8. doi: 10.1128/MMBR.55.3.451-458.1991

58. Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H. The role of DNA methylation in setting up chromatin structure during development. *Nat Genet* (2003) 34:187–92. doi: 10.1038/ng1158

59. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* (2011) 12:179–92. doi: 10.1038/nrg2952

60. Liang J, Wei X, Liu Z, Cao D, Tang Y, Zou Z, et al. Long noncoding RNA (lncRNAs) in cancer: A TCGA data review. *Clin Chim Acta* (2018) 483:227–33. doi: 10.1016/j.cca.2018.05.010

61. Zou SF, Yang XY, Li JB, Ding H, Bao YY, Xu J. UPF1 alleviates the progression of glioma via targeting lncRNA CYTOR. *Eur Rev Med Pharmacol Sci* (2019) 23:10005–12. doi: 10.26355/eurrev_201911_19567

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2021 He, Wang, Tang and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*