



LncRNAs Predicted to Interfere With the Gene Regulation Activity of *miR-637* and *miR-196a-5p* in GBM

Jingfang Zheng¹, Zhiying Su², Yang Kong^{3,4}, Qingping Lin¹, Hongli Liu¹, Yanlong Wang^{1*} and Jian Wang^{3,4*}

¹ Department of Gynecology, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, ² Department of Reproductive Medicine, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, ³ Department of Neurosurgery, Qilu Hospital of Shandong University and Institute of Brain and Brain-Inspired Science, Shandong University, Jinan, China, ⁴ Translational Cancer Research Group, Department of Biomedicine, University of Bergen, Bergen, Norway

OPEN ACCESS

Edited by:

Liam Chen,
Johns Hopkins University,
United States

Reviewed by:

Maria Caffo,
University of Messina, Italy
Arif Ozgun Harmanci,
University of Texas Health Science
Center at Houston, United States

*Correspondence:

Yanlong Wang
xmfy_wyl@126.com
Jian Wang
jian.wang@uib.no

Specialty section:

This article was submitted to
Neuro-Oncology and Neurosurgical
Oncology,
a section of the journal
Frontiers in Oncology

Received: 03 November 2019

Accepted: 20 February 2020

Published: 09 March 2020

Citation:

Zheng J, Su Z, Kong Y, Lin Q, Liu H,
Wang Y and Wang J (2020) LncRNAs
Predicted to Interfere With the Gene
Regulation Activity of *miR-637* and
miR-196a-5p in GBM.
Front. Oncol. 10:303.
doi: 10.3389/fonc.2020.00303

Rigorous molecular characterization of biological systems has uncovered a variety of gene variations underlying normal and disease states and a remarkable complexity in the forms of RNA transcripts that exist. A recent concept, competitive endogenous RNA, suggests that some non-coding RNAs can bind to miRNAs to modulate their role in gene expression. Here, we used several platforms, integrating mRNA, non-coding RNAs and protein data to generate an RNA-protein network that may be dysregulated in human glioblastoma multiforme (GBM). Publicly available microarray data for mRNA and miRNA were used to identify differentially expressed miRNAs and mRNAs in GBM relative to non-neoplastic tissue samples. Target miRNAs were further selected based on their prognostic significance, and the intersection of their target gene set with the differentially expressed gene set in Venn diagrams. Two miRNAs, *miR-637* and *miR-196a-5p*, were associated with poor and better prognosis, respectively, in GBM patients. Non-coding RNAs, *ENSG00000203739/ENSG00000271646* and *TPTEP1*, were predicted to be miRNA target genes for *miR-637* and *miR-196a-5p* and positively correlated with the selected mRNA, *CYBRD1* and *RUFY2*. A local protein interaction network was constructed using these two mRNAs. Predictions based on the *ENSG00000203739/ENSG00000271646-miR-637-CYBRD1* and *TPTEP1-miR-196a-5p-RUFY2* regulation axes indicated that the two proteins may act as an oncogene and tumor suppressor, respectively, in the development of GBM. These results highlight competitive endogenous RNA networks as alternative molecular therapeutic targets in the treatment of the disease.

Keywords: lncRNA, mRNA, GBM, molecular datasets, network

INTRODUCTION

Glioblastoma multiforme (GBM) is the most aggressive central nervous system tumor in adults, and the prognosis is bleak. Conventional therapies, including surgery, radiotherapy, and chemotherapy with temozolomide have not resulted in significant improvement in the survival outcomes of patients with GBM. Median overall survival is 15–23 months, and 5-year survival is <6% (1). Causes of poor prognosis include invasive tumor growth in an essential organ that cannot be

thoroughly removed, the presence of the blood-brain barrier (BBB), their intrinsic resistance to the induction of cell death, tumor heterogeneity and a complex pathogenesis. The basis for this behavior is the simultaneous corruption of many genes which results in the lack of a single, targetable oncogenic pathway, and thus, significant challenges in systemic therapy.

Answers for treatment are thought to be buried in the molecular datasets accumulating since 2006 when The Cancer Genome Atlas (TCGA) team, sponsored by the National Cancer Institute (NCI), published DNA copy number, gene expression, and DNA methylation analysis for 206 GBMs. Rigorous analysis of these data has led to some critical insights in the development of human gliomas. GBMs can now be classified as one of four molecular subtypes based on transcriptome expression data: classical, neural, mesenchymal, and proneural. Subsequent analysis of the methylation status of DNA promoter regions in 272 GBMs revealed two major glioma-CpG island methylation phenotypes (glioma-CpG island methylator phenotype, or G-CIMP and non-G-CIMP types). Finally, a total 1,122 gliomas samples were divided into *IDH* mutated and wild type tumors based on analysis of the multi-dimensional histological data.

Non-coding RNAs have also become part of the story. A collection of dysregulated lncRNAs, including hundreds of candidate onco- and tumor-suppressor lncRNAs, have been identified in the context of 14 different tumor types (2). Recurrent hypomethylation of 1,006 lncRNA genes in cancer, including *EPIC1* (epigenetically-induced lncRNA1) has also been described (3). *EPIC1* promotes cell-cycle progression by interacting with MYC, enhancing luminal B breast cancer cell growth *in vitro* and *in vivo*.

The expanding landscape for RNA transcript types has triggered additional theories about gene regulation. A recent concept, competitive endogenous RNA (ceRNA), represents a novel regulatory mechanism between non-coding and coding RNAs. The theory suggests that lncRNAs, circRNAs, and pseudogenes can act as “molecular sponges” to compete for miRNAs and effectively modulate their functions. The competition for miRNAs is mediated by miRNA binding sites or miRNA response elements (4). This creative hypothesis is supported by an increasing number of experimental results.

With the development of high-throughput gene sequencing and chip technology, analyzing molecular data has become an extremely meaningful but challenging task. An increasing number of R language packages and bioinformatics analysis tools have become more user friendly for a broader range of investigators. Here, we used some of these tools to analyze miRNA and mRNA datasets to determine where they might converge in the development of human GBM. Our results led us to two miRNAs, *miR-196a-5p* and *miR-637*, their target mRNAs encoding a putative oncogene and tumor suppressor, and non-coding RNAs regulating the miRNA activity. We show how a fundamental biological question, which genes and miRNAs are differentially expressed in human GBM, can provide the basis for the construction of a molecular network including RNAs and proteins that might drive aspects of GBM development. Such “excavation” of molecular datasets is the key for the advancement of novel therapies in the treatment of the disease.

MATERIALS AND METHODS

Microarray Data

MicroRNA expression profiles in GSE25631 from the publicly available NCBI GEO database, which had been collected using the Illumina GPL8179 platform (Human v2 microRNA Expression Beadchip), were analyzed. The GSE25631 dataset includes 82 primary GBM surgical specimens and 5 non-neoplastic brain tissue samples from areas surrounding arteriovenous malformations as controls. mRNA expression profiles were obtained from the GSE4290 dataset (5), which is based on the Affymetrix GPL570 platform (HG U133 Plus 2.0 Array). The GSE4290 dataset includes 79 GBM samples and 23 non-neoplastic brain tissue samples from epilepsy patients as controls. To validate our results, GSE90604 and GSE65626, two microRNA expression datasets, were also analyzed.

Analysis to Identify Differentially Expressed microRNAs and Differentially Expressed Genes (DEGs)

Analysis using GEO2R, a webtool available from the NCBI, was performed to detect differentially expressed microRNAs and DEGs between GBM and non-neoplastic control samples. Details of the R script of GEO2R are provided in the **Supplementary Material**. Adjusted *P*-values were used to reduce the false positive rate using the Benjamini and Hochberg false discovery rate method by default (6, 7). $P < 0.05$ and $|\log_{2}FC| \geq 2$ were set as the cutoff values.

Identification of Target Genes of Candidate microRNAs

Cytoscape, open-source software for the integration of molecular interaction network data, was used to visualize the relationship between microRNAs and differentially expressed genes (DEGs). CyTargetLinker (8), a plug-in for Cytoscape, was used to identify microRNA-target genes (MTGs), based on experimentally validated microRNA-target interaction (MTIs) files stored in miRTarBase (9), a database containing miRNA-target interactions. In general, the collected MTIs in miRTarBase have been validated experimentally using luciferase assays, western blots, microarrays and next-generation sequencing.

GO and KEGG Pathway Enrichment Analysis for MTGs of Candidate microRNAs and DEGs

Kyoto Encyclopedia of Genes and Genomes (10) (KEGG) pathway analysis was performed to identify potential functions of the MTGs of the candidate microRNAs and DEGs. Gene ontology analysis (GO), a common useful method for annotating genes and identifying characteristic biological attributes, including biological processes, molecular functions, and cellular components, for high-throughput genome or transcriptome data (11), was performed on DEGs. Metascape (<http://metascape.org>), a web-based online bioinformatics resource that aims to provide tools for the functional interpretation of large lists of genes or proteins (12), was also used to identify function of

MTGs and to conduct GO and KEGG pathway enrichment (13) on DEGs derived in our analysis. The enriched KEGG pathways of MTGs were visualized using ClueGO+Cluepedia, a plug-in that visualizes the non-redundant biological terms for large clusters of genes in a functionally grouped network (14). For DEGs, visualization of the biological processes, molecular functions, cellular components and pathways was performed using Excel and R ggplot2 packages.

Identification of Hub Genes Among DEGs

Protein names encoded by DEGs were imported into STRING (<https://string-db.org/>) to obtain a protein-protein interaction (PPI) network (15). CentiScaPe 2.2 was used to analyze nodes in the network (16). Genes with the highest degrees of connectivity were selected as hub genes. Analysis of the core genes can represent whether the chip results are consistent with GBM.

Identification of Candidate Genes Regulated by DEGs and MTGs

Venn diagrams (17) were used to identify the intersection between *miR-196a-5p* target and GBM down-regulated genes, as well as between *miR-637* target and GBM up-regulated genes. Gene Expression Profiling Interactive Analysis (GEPIA; <http://gepia.cancer-pku.cn/index.html>), a newly developed interactive web server, was used to analyze differences in expression between tumor and normal samples using RNA sequencing data (18). A boxplot was generated to visualize the relationship.

Identification of Target Non-coding RNAs of Candidate microRNAs

Analysis using LncBase v.2 was performed to predict the target non-coding RNAs of differentially expressed microRNAs in GBM (19). To acquire high confidence target non-coding RNAs, the threshold was set at > 0.9 , and the tissue was confined to brain. Target non-coding RNAs of candidate genes, including lncRNAs, cirRNAs, and pseudogenes, were chosen based on a positive relationship with candidate genes in the data collected from TCGA GBMs on the Tanric website (20). The expression of target non-coding RNAs and candidate genes was set to a positive correlation above moderate levels (correlation coefficient > 0.4 , P -value < 0.01). The intersection between candidate genes and predicted target non-coding RNAs was made.

PPI Network Extension and Establishment of the Competitive Endogenous RNA (ceRNA) Hypothesis

Protein names encoded by candidate genes were imported into STRING, a database of known and predicted protein-protein interactions (<https://string-db.org/>) (21). PPI networks were extended until the proteins from our analysis connected with each other. Candidate microRNA and non-coding RNAs were then mapped to the network.

RESULTS

Identification of Differentially Regulated Candidate miRNAs in GBM

Analysis of the GSE25631 dataset yielded a total of 67 differentially expressed miRNAs ($P < 0.05$ and $|\log_{2}FC| \geq 2$) between GBM and non-neoplastic brain. Of these miRNAs, 27 were up-regulated and 40 were down-regulated in GBM relative to control samples (Figure 1A). We examined the prognostic value of 10 miRNAs with the most significant fold changes in expression (Table 1), using OncoLnc, a tool for interactively exploring survival correlations coupled to expression data for mRNAs, miRNAs, or lncRNAs. Two of these miRNAs, *miR-196a-5p* and *miR-637*, were associated with overall survival (OS; Figures 1B,C). High expression of *miR-196a-5p* in GBMs (HR = 0.196, $P = 0.000795$) was associated with worse OS in patients (Figure 1D), while high expression of *miR-637* (HR = -0.634 , $P = 0.045$) was associated with better OS in patients (Figure 1E). Because the statistical difference in OS using *miR-637* was not significant, we selected multiple cutoff values for verification. The results are shown in Supplementary Figure 1. The expression of *miR-196-5p* and *miR-637* was also verified in the TCGA database. The expression of *miR-196-5p* was consistent with the results obtained with the GSE25631 dataset. However, compared with the control group, *miR-637* was not significantly reduced in GBM from the TCGA (Supplementary Figures 2A,B).

KEGG Enrichment Analysis Links *miR-196-5p* and *miR-637* to Pathways Involved in Cancer

To understand the possible function of *miR-196-5p* and *miR-637* in the development of GBM, KEGG pathway enrichment analysis of their target genes was performed. KEGG analysis was performed on the *miR-196-5p* and *miR-637* target genes ($n = 356$) identified using Cytoscape (Figure 1F). The results revealed these genes to be associated with several pathways involved in disease development, including small cell lung cancer, proteoglycans in cancer, Parkinson's disease, viral carcinogenesis, prostate cancer, chronic myeloid leukemia, the Hedgehog signaling pathway, glioma, microRNAs in cancer, and the cell cycle (Figure 1G). We further validated the results using the open-source pathway database REACTOME (<https://reactome.org/>) (Supplementary Figures 3A,B) (22).

Identification of Differentially Expressed Genes (DEGs) and Enrichment Analysis

To identify differentially expressed genes in GBM relative to non-neoplastic brain, we performed analysis on the GSE4290 dataset containing mRNA expression profiles (5). A total of 1,170 differentially expressed genes were detected; 397 were up-regulated and 773 were down-regulated in GBM samples relative to non-neoplastic brain tissue (Figure 2A). To associate function with the DEGs, we performed GO (Figure 2B) and KEGG pathway analysis (Figure 2C). DEGs were found to be enriched in biological processes (BP, Figure 2B) involving the regulation of neurogenesis, plasma membrane

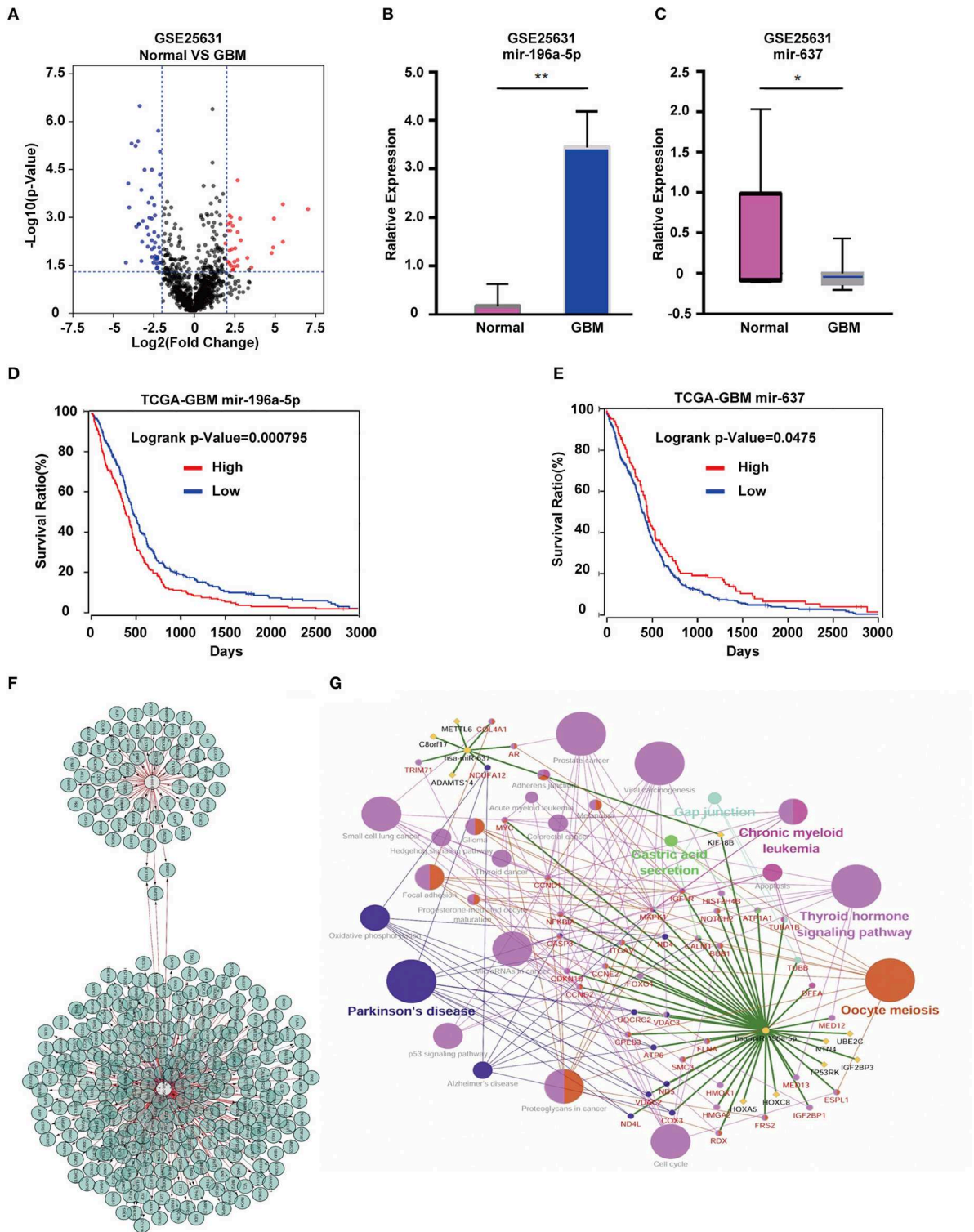


FIGURE 1 | Identification of candidate miRNAs and prediction of their target genes. **(A)** Volcano plot compiled using expression data obtained from the publicly available dataset GSE25631. Red and blue dots represent up-regulated and down-regulated differentially expressed microRNAs, respectively ($P < 0.05$, $|\log_{2}FC| \geq 2$). **(B)** The expression of *miR-196a-5p* and **(C)** *miR-637* in GBM relative to non-neoplastic brain tissue samples. **(D)** Prognostic value of *miR-196a-5p* **(E)** and *miR-637* in GBM based on the TCGA database. **(F)** Predicted target genes of *miR-196-5p* and *miR-637*. **(G)** KEGG pathway enrichment analysis of target genes of *miR-196-5p* and *miR-637*. Figures in **(F,G)** were designed using the open source software Cytoscape 3.6.1 and its plugin or app CyTargetLinker, ClueG+Cluepedia. Data are shown as the mean \pm standard deviation. * $P < 0.05$; ** $P < 0.01$ vs. control samples.

TABLE 1 | Top 10 differentially expressed miRNAs in GSE25631.

miRNA_ID	P-value	Log ₂ FC
hsa-miR-196a-5p	4.38E-04	5.599188
hsa-miR-558	6.37E-03	5.553814
hsa-miR-144	1.19E-03	5.045044
hsa-miR-106a	9.89E-03	4.979875
hsa-miR-637	2.93E-02	-4.14014
hsa-miR-876-3p	9.70E-05	-3.97221
hsa-miR-1224-5p	5.50E-06	-3.78824
hsa-miR-518e	4.09E-02	3.611791
hsa-miR-138-2-3p	6.31E-06	-3.54098
hsa-miR-203	2.13E-03	-3.45736

bounded cell projection morphogenesis, cell morphogenesis involved in neuron differentiation, neurotransmitter secretion, and regulation of neurogenesis. They were also enriched in molecular functions (MF, **Figure 2B**) involving ion channel activity, protein kinase activity, and cell adhesion molecule binding. Based on cellular components (CC, **Figure 2B**), DEGs were found to be located in the synapse, the vesicle membrane, the perinuclear region of the cytoplasm, and the extracellular matrix. KEGG pathway enrichment analysis linked DEGs to processes involving the synaptic vesicle cycle, pathways in cancer, PI3K-Akt signaling, proteoglycans in cancer, and Ras signaling. These results suggest that these DEGs may play a role in promoting tumor progression through their function.

Analysis of DEGs Interaction Network and Acquisition of Hub Genes

Using the STRING protein databases, we generated a PPI network for the top 20 hub genes with the highest degrees of connectivity (**Figure 2D**). The top 20 included genes known to promote the development of human cancer. *CDK1*, *CCNB2*, and *CDC20* (23) are involved in the regulation of the cell cycle. *EGFR* and *VEGFA* have been reported to promote GBM proliferation and invasion. *TOP2A*, *BUB1*, *NDC80*, and *TTK* participate in mitosis. *BIRC5* is a member of the inhibitor of apoptosis (IAP) gene family, which prevents apoptotic cell death. Finally, *HMMR* encodes a protein involved in cell motility.

Identification of Candidate Genes Regulated by *miR-196-5p/miR-637*

To identify DEGs that may be regulated by *miR-196-5p/miR-637*, we generated a venn diagram illustrating the intersection between the target genes of *miR-196-5p/miR-637* and the DEGs (**Figure 3A**). Among the genes appearing in the intersection of the two datasets were *CYBRD1* and *RUFY2* (**Table 2**). Expression levels were high and low for *CYBRD1* and *RUFY2*, respectively, in GBM, corresponding with activity as a putative oncogene or tumor suppressor gene. The expression of these two genes was also related to the OS of patients (**Figures 3B,C**). High expression of *CYBRD1* was related to poor survival whereas high expression of *RUFY2* was associated with better survival. The differential

mRNA expression of *CYBRD1* and *RUFY2* was consistent with results obtained using the expression data from the TCGA GBMs (**Figures 3D,E**). We also analyzed the relationship between the expression of *CYBRD1/RUFY2* and *MGMT/IDH* status within TCGA GBM samples. Neither mRNA exhibited significant differences between methylated/unmethylated *MGMT* and wild-type/mutant *IDH* in tumors (**Supplementary Figure 4**). Finally, a heatmap based on mRNA expression data of intersection genes from GSE4290 illustrates the differential expression of the two genes between GBM and non-neoplastic samples (**Figure 3F**).

Identification of Non-coding RNAs Involved in the Regulation of *miR-196-5p/miR-637* and *CYBRD1/RUFY2*

To determine whether any non-coding RNAs might be involved in the regulation of *CYBRD1* and *RUFY2*, we generated a diagram to reveal the intersection between predicted target ncRNAs of *miR-196a-5p/miR-637* and non-coding RNAs which are positively related to these two genes (**Figure 3G**). Three ncRNAs, *ENSG00000203739*, *ENSG00000271646*, and *ENSG00000100181*, emerged from the analysis, and high expression of *ENSG00000203739* and *ENSG00000271646* was associated with poor prognosis in GBM using Log-Rank models (**Figures 3H,I**). The survival curve for the third ncRNA, *ENSG00000100181*, was only statistically significant in the Cox model for GBM and low grade glioma (**Figures 3J,K**).

The official symbol of *ENSG00000100181* is *TPTEP1*, which is also known as *psiTPTE22*. *psiTPTE22-HERV* has been reported to be epigenetically silenced by DNA methylation in cancers of the kidney, liver, lung, and stomach (24–26). We were therefore interested in the possibility that *TPTEP1* expression might differ on the basis of GBM molecular subtype. Using the TCGA database, we found expression of *TPTEP1* to be significantly higher in the G-CIMP subtype relative to the other molecular subtypes. As patients with G-CIMP subtype tumors in general have a better prognosis, low-expression of *TPTEP1* in non-G-CIMP subtypes may be consistent with a role as a tumor suppressor gene (**Figure 3L**). Therefore, we believe the role of *TPTEP1* in the pathogenesis of GBM warrants further investigation.

A Competitive Endogenous RNA (ceRNA) Regulation Network Involving *TPTEP1*, *CYBRD1*, and *RUFY2* Built in GBM

Using the String database, we constructed a local protein network between the proteins *CYBRD1* and *RUFY2* (**Figure 4**). We then integrated the ncRNAs. In this network, high-expression of *ENSG00000203739/ENSG00000271646* was predicted to promote GBM proliferation and invasion by suppressing *miR-637* which leads to increased expression of *CYBRD1*, a putative oncogene. Loss of *TPTEP1* however leads to increased levels of *miR-196a-5p* to adsorpt/bind to *miR-196a-5p*. When overexpressed, *miR-196a-5p* impedes translation of *RUFY2*, a putative tumor suppressor protein. Thus, dysregulation of these ncRNAs can lead to progression of GBM through gain and loss of the *CYBRD1* and *RUFY2*, respectively.

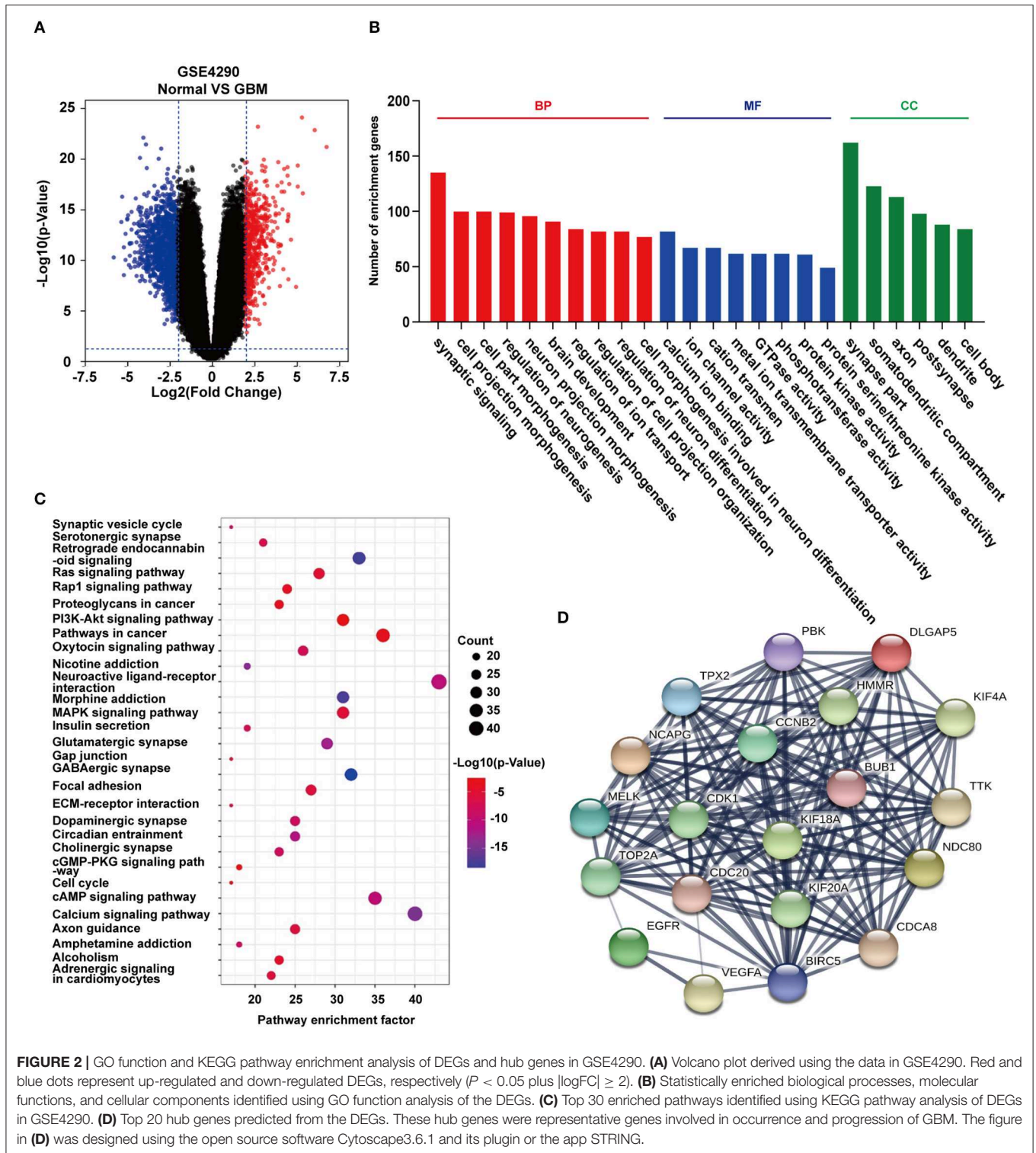


FIGURE 2 | GO function and KEGG pathway enrichment analysis of DEGs and hub genes in GSE4290. **(A)** Volcano plot derived using the data in GSE4290. Red and blue dots represent up-regulated and down-regulated DEGs, respectively ($P < 0.05$ plus $|\log_{2}FC| \geq 2$). **(B)** Statistically enriched biological processes, molecular functions, and cellular components identified using GO function analysis of the DEGs. **(C)** Top 30 enriched pathways identified using KEGG pathway analysis of DEGs in GSE4290. **(D)** Top 20 hub genes predicted from the DEGs. These hub genes were representative genes involved in occurrence and progression of GBM. The figure in **(D)** was designed using the open source software Cytoscape3.6.1 and its plugin or the app STRING.

The local PPI network also included the proteins EFNA5, EFNB2, ACTR2, EPHA3. Many of these proteins are implicated in GBM development. EFNA5, for example, is a member of the ephrin gene family, which participates in late stage nervous system development and differentiation. EFNB2 plays an important role in physiological and pathological angiogenesis,

and its role in tumor vessel development has been extensively studied. EFNB2 has been shown to mediate perivascular invasion of glioblastoma stem-like cells (27). ACTR2 is known to be a major constituent of the ARP2/3 complex, which is located at the cell surface and is essential for cell shape and motility through lamellipodial actin assembly and protrusion. Overexpression of

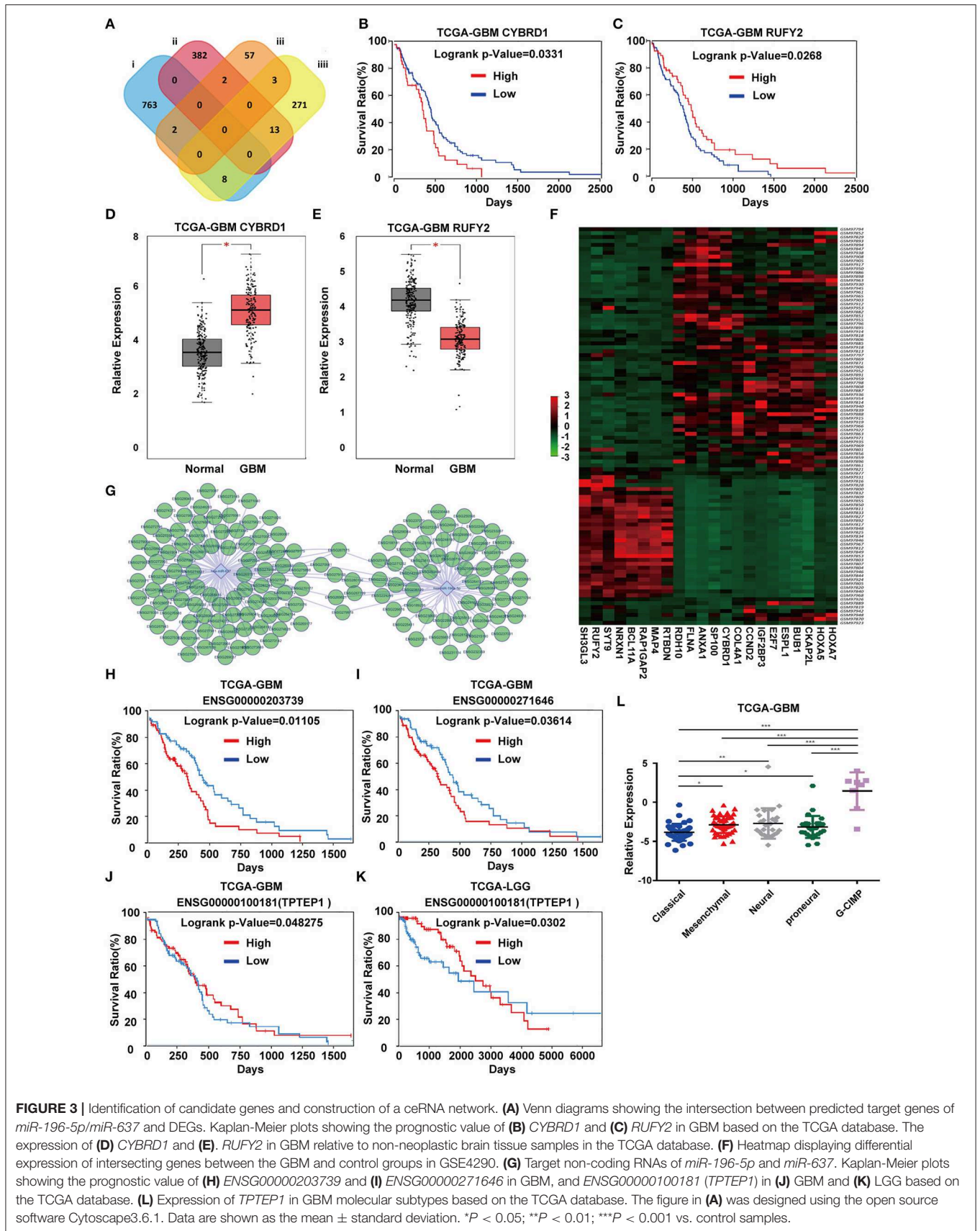
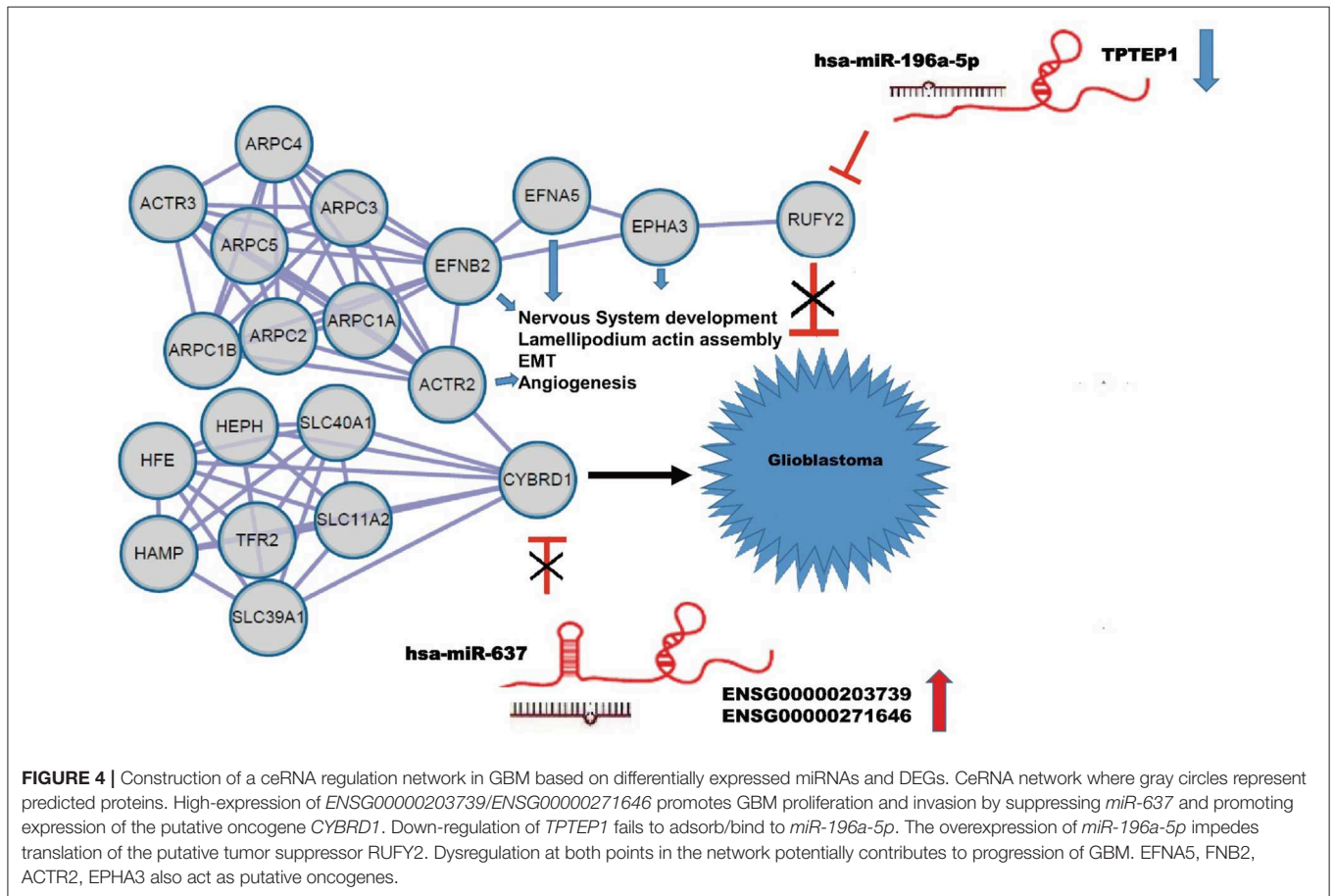


FIGURE 3 | Identification of candidate genes and construction of a ceRNA network. **(A)** Venn diagrams showing the intersection between predicted target genes of *miR-196-5p/miR-637* and DEGs. Kaplan-Meier plots showing the prognostic value of **(B)** *CYBRD1* and **(C)** *RUFY2* in GBM based on the TCGA database. The expression of **(D)** *CYBRD1* and **(E)** *RUFY2* in GBM relative to non-neoplastic brain tissue samples in the TCGA database. **(F)** Heatmap displaying differential expression of intersecting genes in the GBM and control groups in GSE4290. **(G)** Target non-coding RNAs of *miR-196-5p* and *miR-637*. Kaplan-Meier plots showing the prognostic value of **(H)** *ENSG00000203739* and **(I)** *ENSG00000271646* in GBM, and *ENSG00000100181 (TPTEP1)* in **(J)** GBM and **(K)** LGG based on the TCGA database. **(L)** Expression of *TPTEP1* in GBM molecular subtypes based on the TCGA database. The figure in **(A)** was designed using the open source software Cytoscape3.6.1. Data are shown as the mean ± standard deviation. **P* < 0.05; ***P* < 0.01; ****P* < 0.001 vs. control samples.

TABLE 2 | Intersection of non-coding RNAs involved in the regulation of miR-196-5p/miR-637 and CYBRD1/RUFY2.

MicroRNA	Prediction score	Non-coding RNA	Gene	Correlation coefficient	P-value
hsa-miR-637	1	ENSG00000203739	CYBRD1	0.401	3.17E-07
hsa-miR-637	0.999	ENSG00000246263	CYBRD1	0.425	5.35E-08
hsa-miR-637	0.999	ENSG00000254154	CYBRD1	0.417	9.96E-08
hsa-miR-637	0.999	ENSG00000271646	CYBRD1	0.445	1.09E-08
hsa-miR-637	0.998	ENSG00000272908	CYBRD1	0.403	2.89E-07
has-196a-5p	0.904	ENSG00000100181	RUFY2	0.421	2.78E-05



ARP2 has been shown to promote gastric cancer cell migration and invasion. In contrast, ARP2 knockdown suppressed cell motility (28). Finally, *EPHA3* is frequently overexpressed in GBM and in particular, in the mesenchymal molecular subtype, which also shows a more aggressive phenotype in patients (29). Importantly, *EPHA3* is highly expressed on the tumor-initiating cell population in glioma, which potentially maintains tumor cells in a less differentiated state by modulating mitogen-activated protein kinase signaling.

The results, therefore, could be consistent with a regulatory network composed of ceRNA plus the interactive proteins *CYBRD1* and *RUFY2* and involved in the proliferation and invasion of GBM.

DISCUSSION

The challenge today is in analyzing burgeoning molecular datasets in ways that will yield biologically and clinically meaningful insights into the development and treatment of human disease. In the current study, we explored the integration of RNA and protein datasets to identify pathways regulating the development of human GBM. We first analyzed standard array data to identify differentially expressed microRNAs and mRNAs in GBM relative to non-neoplastic brain tissue controls using the data contained in the StarBase, LncBase, Tanric, and TCGA databases. Based on the theory of ceRNA, we found potential ncRNA regulatory pathways involving an oncogene and a tumor

suppressor, *ENSG00000203739/ENSG00000271646-miR-637-CYBRD1* and *TPTEP1-miR-196a-5p-RUFY2*, and constructed a local PPI network which might contribute to proliferation and invasion of GBM.

Experimental results are consistent with some of our predictions. First, Boulton et al. have reported that CYBRD1 was overexpressed in the progression of Barrett's metaplasia to adenocarcinoma and this change are associated with increased iron deposition (30). Brookes et al. also confirmed that the increased expression in iron import proteins, including CYBRD1, was associated with progression to colorectal cancer (31). However, other studies showed that high CYBRD1 expression has been associated with increased metastasis- and/or relapse-free survival in breast cancer (32). In functional experiments, CYBRD1 inhibited phosphorylation of FAK and the focal adhesion pathway which is involved in migration and invasion. The reasons for these discrepancies include differences in pathogenesis and the shift between proliferation and migration in tumors. Second, *miR-196a* has been reported to function as a putative oncogene in many cancers. *MiR-196a* is significantly upregulated in GBM, and high levels of *miR-196a* are positively related to the malignant progression of gliomas (33). Third, *miR-196a-5p* promotes proliferation and suppresses apoptosis in GBM cells both *in vitro* and *in vivo* by targeting Ikb α (34). *MiR-196a-5p* also interacts directly with the 3'UTR of *ZMYND11* and promotes the growth of GBM cells (35). Finally, *miR-637* has been reported to be a tumor suppressor gene in diverse human cancers, such as gastric (36), ovarian (37) and colorectal cancers (38). Results in human glioma are consistent with a function as a tumor suppressor; expression levels of *miR-637* were significantly reduced in clinical glioma tissues compared with normal brain tissues (39). Moreover, these studies revealed that *miR-637* directly binds AKT1 and inhibits glioma cell growth, migration and invasion *in vitro* and *in vivo*. These data support the feasibility of our approach.

There are however deficiencies in our strategy. First, excluding function as a "molecular sponge," the mode of action of lncRNAs can be roughly divided into the following categories: signal, guide and scaffold. Second, due to the complexity of molecular mechanisms regulating disease development and the limitations of our analytical methods, many important molecules involved in GBM remain unidentified. Third, we input more processed TCGA data into our analysis, which may result in the loss of a more comprehensive perspective. Finally, because many algorithms are based on computer models, many prediction results cannot be achieved in real-life experiments.

Significant effort to analyze and integrate such large amounts of molecular data are taking place worldwide. Researchers funded by the National Institutes of Health (NIH), for example, have completed a comprehensive genomic analysis known as the

PanCancer Atlas. This project published a total of 27 top bioinformatics papers. These articles are undoubtedly innovative and encouraging. But such efforts are merely the beginning. Here, we laid the groundwork for a new strategy to explore the complicated molecular mechanisms underlying the development of GBM or other diseases. The next important step is to verify our approach using functional experiments to confirm our model in **Figure 4**.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GSE25631, GSE4290, GSE90604, GSE65626.

AUTHOR CONTRIBUTIONS

JZ, YK, and JW contributed to the conception of the study. JZ and YK contributed to experimental technology and experimental design. JZ, ZS, YK, QL, and HL performed the data analyses. JZ, YK, and JW wrote the manuscript. YW and JW supervised the study.

FUNDING

This work was supported by the Department of Science & Technology of Shandong Province (2017CXGC1502, GG201809170226, and ZR2017MH116), the Jinan Science and Technology Bureau of Shandong Province (2019GXRC006), Helse-Vest, the University of Bergen, the Norwegian Cancer Society, and the Norwegian Research Council.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2020.00303/full#supplementary-material>

Supplementary Figure 1 | Survival curves for the upper 20% of patients with high hsa-miR-637 expression in comparison with the lower 20%, 30%, 40%, 50%, 60%, 70%, and 80% of patients with low hsa-miR-637 expression (**A-G**). Survival curve for the upper 30% of patients with high hsa-miR-637 expression in comparison with the lower 30% of patients with low hsa-miR-637 expression (**H**).

Supplementary Figure 2 | The expression of (**A**) *miR-196a-5p*, (**B**) *miR-637*, (**C**) *CYBRD1* and (**D**) *RUFY2* in GBMs and normal tissues in the TCGA database.

*** $P < 0.001$ vs. control sample.

Supplementary Figure 3 | (**A**) PPI network generated using the REACTOME database. (**B**) The PPI network consists of top 300 DEGs in GSE4290.

Supplementary Figure 4 | Relationship between the expression of *CYBRD1/RUFY2* and *MGMT/IDH* status within TCGA GBM samples. The expression of *CYBRD1* in (**A**) methylated/unmethylated *MGMT* and (**B**) wild-type/mutant *IDH* tumors. The expression of *RUFY2* in (**C**) methylated/unmethylated *MGMT* and (**D**) wild-type/mutant *IDH* tumors.

REFERENCES

- Ostrom QT, Gittleman H, Xu J, Kromer C, Wolinsky Y, Kruchko C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2009-2013. *Neuro Oncol.* (2016) 18:v1-75. doi: 10.1093/neuonc/nov207
- Chiu HS, Somvanshi S, Patel E, Chen TW, Singh VP, Zorman B, et al. Pan-cancer analysis of lncRNA regulation supports their targeting of

- cancer genes in each tumor context. *Cell Rep.* (2018) 23:297–312.e12. doi: 10.1016/j.celrep.2018.03.064
3. Cui X, Morales RT, Qian W, Wang H, Gagner JP, Dolgalev I, et al. Hacking macrophage-associated immunosuppression for regulating glioblastoma angiogenesis. *Biomaterials.* (2018) 161:164–78. doi: 10.1016/j.biomaterials.2018.01.053
 4. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the rosetta stone of a hidden RNA language? *Cell.* (2011) 146:353–8. doi: 10.1016/j.cell.2011.07.014
 5. Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, et al. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell.* (2006) 9:287–300. doi: 10.1016/j.ccr.2006.03.003
 6. Feser WJ, Fingerlin TE, Strand MJ, Glueck DH. Calculating average power for the Benjamini-Hochberg procedure. *J Stat Theory Appl.* (2009) 8:325–52.
 7. Madar V, Batista S. FastLSU: a more practical approach for the Benjamini-Hochberg FDR controlling procedure for huge-scale testing problems. *Bioinformatics.* (2016) 32:1716–23. doi: 10.1093/bioinformatics/btw029
 8. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303
 9. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. Mirtarbase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* (2018) 46:D296–302. doi: 10.1093/nar/gkx1067
 10. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst.* (2014) 10:2441–7. doi: 10.1039/C4MB00287C
 11. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* (2015) 43:D1049–56. doi: 10.1093/nar/gku1179
 12. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, et al. Meta- and orthogonal integration of influenza “OMICs” data defines a role for UBR4 in virus budding. *Cell Host Microbe.* (2015) 18:723–35. doi: 10.1016/j.chom.2015.11.002
 13. Shirolkar A, Chakraborty S, Mandal T, Dabur R. Unbiased plasma metabolomics reveal the correlation of metabolic pathways and Prakritis of humans. *J Ayurveda Integr Med.* (2017) 9:113–22. doi: 10.1016/j.jaim.2017.05.002
 14. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and *in silico* data. *Bioinformatics.* (2013) 29:661–3. doi: 10.1093/bioinformatics/btt019
 15. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* (2017) 45:D362–8. doi: 10.1093/nar/gkw937
 16. Scardoni G, Tosadori G, Faizan M, Spoto F, Fabbri F, Laudanna C. Biological network analysis with CentiScaPe: centralities and experimental dataset integration. *F1000Res.* (2014) 3:139. doi: 10.12688/f1000research.4477.1
 17. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: an open access standalone functional enrichment and interaction network analysis tool. *Proteomics.* (2015) 15:2597–601. doi: 10.1002/pmic.201400515
 18. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* (2017) 45:W98–102. doi: 10.1093/nar/gkx247
 19. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.* (2016) 44:D231–8. doi: 10.1093/nar/gkv1270
 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. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015) 43:D447–52. doi: 10.1093/nar/gku1003
 22. Haw R, Stein LD. Using the reactome database. *Curr Protoc Bioinformatics.* (2004) Chapter 8:Unit 8.7. doi: 10.1002/0471250953.bi080757
 23. Belinky F, Nativ N, Stelzer G, Zimmerman S, Iny Stein T, Safran M, et al. PathCards: multi-source consolidation of human biological pathways. *Database.* (2015) 2015:bav006. doi: 10.1093/database/bav006
 24. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J Pathol.* (2002) 196:1–7. doi: 10.1002/path.1024
 25. Norwood MS, Lupo PJ, Chow EJ, Scheurer ME, Plon SE, Danysh HE, et al. Childhood cancer risk in those with chromosomal and non-chromosomal congenital anomalies in Washington state: 1984–2013. *PLoS ONE.* (2017) 12:e0179006. doi: 10.1371/journal.pone.0179006
 26. Liang Q, Ding J, Xu R, Xu Z, Zheng S. The novel human endogenous retrovirus-related gene, psiTPTE22-HERV, is silenced by DNA methylation in cancers. *Int J Cancer.* (2010) 127:1833–43. doi: 10.1002/ijc.25213
 27. Sun Y, Shang Y, Ren G, Zhou L, Feng B, Li K, et al. Coronin3 regulates gastric cancer invasion and metastasis by interacting with Arp2. *Cancer Biol Ther.* (2014) 15:1163–73. doi: 10.4161/cbt.29501
 28. Krusche B, Ottone C, Clements MP, Johnstone ER, Goetsch K, Lieven H, et al. EphrinB2 drives perivascular invasion and proliferation of glioblastoma stem-like cells. *Elife.* (2016) 5:e14845. doi: 10.7554/eLife.14845
 29. Day BW, Stringer BW, Al-Ejeh F, Ting MJ, Wilson J, Ensby KS, et al. EphA3 maintains tumorigenicity and is a therapeutic target in glioblastoma multiforme. *Cancer Cell.* (2013) 23:238–48. doi: 10.1016/j.ccr.2013.01.007
 30. Boulton J, Roberts K, Brookes MJ, Hughes S, Bury JP, Cross SS, et al. Overexpression of cellular iron import proteins is associated with malignant progression of esophageal adenocarcinoma. *Clin Cancer Res.* (2008) 14:379–87. doi: 10.1158/1078-0432.CCR-07-1054
 31. Brookes MJ, Hughes S, Turner FE, Reynolds G, Sharma N, Ismail T, et al. Modulation of iron transport proteins in human colorectal carcinogenesis. *Gut.* (2006) 55:1449–60. doi: 10.1136/gut.2006.094060
 32. Lemler DJ, Lynch ML, Tesfay L, Deng Z, Paul BT, Wang X, et al. DCYTB is a predictor of outcome in breast cancer that functions via iron-independent mechanisms. *Breast Cancer Res.* (2017) 19:25. doi: 10.1186/s13058-017-0814-9
 33. Huang F, Tang J, Zhuang X, Zhuang Y, Cheng W, Chen W, et al. MiR-196a promotes pancreatic cancer progression by targeting nuclear factor kappa-B-inhibitor alpha. *PLoS ONE.* (2014) 9:e87897. doi: 10.1371/journal.pone.0087897
 34. Zhang J, Zheng F, Yu G, Yin Y, Lu Q. MiR-196a targets netrin 4 and regulates cell proliferation and migration of cervical cancer cells. *Biochem Biophys Res Commun.* (2013) 440:582–8. doi: 10.1016/j.bbrc.2013.09.142
 35. Yang JP, Yang JK, Li C, Cui ZQ, Liu HJ, Sun XF, et al. Downregulation of ZMYND11 induced by miR-196a-5p promotes the progression growth of GBM. *Biochem Biophys Res Commun.* (2017) 494:674–80. doi: 10.1016/j.bbrc.2017.10.098
 36. Guan EC, Xu XG, Xue FX. Circ-NOTCH1 acts as a sponge of miR-637 and affects the expression of its target gene Apelin to regulate gastric cancer cell growth. *Biochem Cell Biol.* (2019). doi: 10.1139/bcb-2019-0079. [Epub ahead of print].
 37. Zhang M, Xia B, Xu Y, Zhang Y, Xu J, Lou G. Circular RNA (hsa_circ_0051240) promotes cell proliferation, migration and invasion in ovarian cancer through miR-637/CLK4 axis. *Artif Cells Nanomed Biotechnol.* (2019) 47:1224–33. doi: 10.1080/21691401.2019.1593999
 38. Wang L, Jiang F, Xia X, Zhang B. LncRNA FAL1 promotes carcinogenesis by regulation of miR-637/NUPR1 pathway in colorectal cancer. *Int J Biochem Cell Biol.* (2019) 106:46–56. doi: 10.1016/j.biocel.2018.09.015
 39. Que T, Song Y, Liu Z, Zheng S, Long H, Li Z, et al. Decreased miRNA-637 is an unfavorable prognosis marker and promotes glioma cell growth, migration and invasion via direct targeting Akt1. *Oncogene.* (2015) 34:4952–63. doi: 10.1038/onc.2014.419
- 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 © 2020 Zheng, Su, Kong, Lin, Liu, Wang and Wang. 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.