Predicting and Identifying Human Glioblastoma MiRNA Targets Using RRSM and qPCR Methods

Jian Shi*1 and Shiwei Huang2
1Department of Neurology, Department of Veterans Affairs Medical Center, San Francisco and University of California, San Francisco, CA 94121, USA; 2Drexel University College of Medicine, Philadelphia, PA 19102, USA.

*Corresponding author(s): J. Shi, Department of Neurology, Department of Veterans Affairs Medical Center, San Francisco and University of California, San Francisco, CA 94121, USA
Email: jian.shi@ucsf.edu
Tel: +1-415-221-4810; Fax: +1-415-750-2273

Published: 2017.02.10.

ABSTRACT

The pathogenesis and progression of glioblastoma multiforme (GBM) have been investigated extensively, but the genetic factors involved in the development of the disease remain poorly understood. The outcome of GBM is still grim. Recently, numerous microRNA (miRNA) mediated gene regulation and interactions have been found to be involved in the development of the disease, making the disease more difficult to understand using the traditional methods. This study used a bioinformatics pipeline—the relative R-squared method (RRSM)—to predict cancer related miRNA targets using gene expression profiles and motif complementary sequences, prior to quantitative real-time PCR experiments in human GBM tissues and cancer cell lines. This study predicted and confirmed 25 miRNA candidates associated to GBM in a comprehensive and non-biased manner, and ten miRNA candidates were also investigated as potential GBM biomarkers. The combination of bioinformatics algorithm and molecular techniques may yield clues to the mechanism of transcriptional and post-transcriptional regulation in the development of glioblastoma and new information about the genetic networks related to the disease.

Keywords: miRNA, GBM, bioinformatics, real-time PCR, biomarker

INTRODUCTION

Glioblastoma multiforme (GBM) is the most aggressive form of brain tumor, causing more than 13,000 deaths annually in the United States. Despite intensive efforts to optimize the treatment of glioblastoma, the outcomes of GBM patients are still disappointingly poor with the median life expectancy after diagnoses of 15 months. The major reason contributing to this poor outcome is that the genetic factors involved in the development of the disease remain poorly understood despite extensive investigations of the pathogenesis of GBM. Based on the work of several groups, an estimated 2,000 mRNAs, 400 long non-coding RNAs, and 300 microRNAs (miRNAs) have been identified as distinctively expressed genes in GBM [1-4]. Considering the complex regulation and interaction among these genes, the amount of research required to investigate their relationship with GBM using traditional methods appears overwhelming. For example, over 248,000 miRNA-mediated interactions were identified to be involved in the post-transcriptional regulation of mRNAs when analyzing gene expression data in GBM in combination with matched miRNA profiles [5]. Therefore, it remains a significant challenge
to comprehensively investigate these genes, as well as their regulation mechanisms and interactions, in a detailed and non-biased manner.

The discovery of miRNAs represents a very important landmark of advancement in understanding the full scope of the post-transcriptional gene regulation in cancer [6]. These short~22 nt RNAAs suppress protein synthesis from specific transcripts that contain antisense target sequences so the miRNAs can hybridize with complete or partial complementarity. miRNAs have emerged as an important component in regulating the growth of cells, with current estimating that the human genome contains as many as 2000 miRNAs, which post-transcriptionally regulate approximately one third or more of all transcripts. In addition, each targeted transcript may have multiple miRNAs that regulate genes in combination, and each miRNA may target several transcripts [7].

Accurate miRNA target prediction can help gain a better understanding of miRNA regulatory mechanisms. Several target prediction computational algorithms for complementary motif predictions, such as miRanda [8] and TargetScanS [9], have been developed for this purpose. However, these algorithms show very poor overlap in their predicted results. Since computational sequence analysis methods for finding targets and expression profiling methods each have their own respective limitations, it is important to combine the two approaches to best identify the miRNA targets from both sequence and expression datasets.

Recently, some statistical methods have been established to explore miRNA targets, including GenMiR++ [7] and the relative R-squared method (RRST) [10], in which RRST is easy to interpret and less computationally expensive for predicting high confidence targets of miRNAs. Given the thousands of miRNA targets generated by target finding programs and the varied expression of thousands of miRNAs and miRNAs obtained by microarray techniques, the RRST algorithm was developed to predict the miRNA target genes. This learning algorithm can be used for understanding miRNA regulation, which will allow us to obtain a set of miRNA targets from both sequences and expression datasets. Using the RRST pipeline, the high confidence miRNAs and their targets associated with several cancers, such as colon cancer, prostate cancer, pancreatic cancer, lung cancer, breast cancer, bladder cancer and kidney cancer, were predicted [11].

In this study, we applied the RRST bioinformatics pipeline to existing miRNAs and miRNAs microarray data of GBM as well as the motif complementary sequences of miRNAs in order to select the candidates of miRNAs and miRNAs that regulate the development of GBM. Then, the predicted candidates were further identified by quantitative real-time PCR in GBM tissues and cell line. Therefore, this study is anticipated to provide a valuable foundation for future studies examining novel clinical targets for the treatment and diagnosis of GBM.

METHODS

Collected Expression Profiles in GBM Tissues and in Normal Tissues

For this study, we collected all miRNA and mRNA microarray data from EMBL gene bank. 23 miRNA microarray data were from E-GEOD-32466, in which samples of GSM803543-803554 were from primary GBM, and samples of GSM803555-803566 were from recurrent GBM, and the data were collected from Human miRNA Microarray 3.0 (Agilent Technologies, USA). 23 mRNA microarray data were from E-GEOD-60184, in which the grade of tumors was IV for all samples, and the data was collected from HG-U133_PLUS_2 (Affymetrix, USA). Eight normal miRNA data were from E-GEOD-37981, in which the tissues were from pyramidal neurons layer 3 of the superior temporal gyrus (STG) as controls for schizophrenia, and the data were collected from HG-U133-X3P (Affymetrix, USA). In addition, eight normal mRNA data were from E-GEOD-17612, in which the tissues were from pyramidal neurons layer 3 of the superior temporal gyrus (STG) as controls for schizophrenia, and the data were collected from HG-U133_PLUS_2 (Affymetrix, USA).

Selected Parameters of RRST

To predict high confidence miRNAs associated with GBM development, we applied the RRST bioinformatics pipeline separately to the expression profiles in tumor tissues and in normal tissues to obtain two groups of miRNA targets. The high confidence miRNAs related to GBM are those miRNAs present in the group of candidates from the expression profiles in GBM tissues but not in the group of candidates from the expression profiles in normal tissues. In this study, the statistical analysis used in the RRST is the linear regression model using R-project. The programs for the RRST are available on http://www.stat.nctu.edu.tw/~hwang/website_wang%20new.htm [12].

The original study considered the expression data for 114 human miRNAs and 41,699 mRNAs. After using the target-finding program TargetScanS [9], 6387 potential target pairs that covered 890 unique mRNAs were identified. In this study, we only considered those potential target pairs with scores larger than 50% for complementary sequences from miRanda [8]. Thus, 3195 potential target pairs were considered, and the number of unique mRNAs was 584. To perform the RRST, two thresholds p0=0.01 and s=0.95 were used as the criterion parameters after comparing several groups of parameters for running RRST to obtain high confidence candidates from potential targets. In addition, there are many other choices of setting p0 and s to accommodate different requirements. We downloaded the original R code of RRST and digested the code line by line based on the author’s advice. This allowed us to use our own collected data to run the program on our computer for GBM study.

Cell culture and GBM tissues

Human U87 GBM cell line was cultured in 10% FBS in DMEM (Invitrogen, USA). The prostate cancer PC3 cell line was cultured in 10% FBS in F-12 (Invitrogen, USA), and the prostate cancer Du145 cell line was cultured in 10% FBS in DMEM (Invitrogen, USA). 0.25% Trypsin-EDTA was used for passage, and DMSO was used for cell frozen in liquid nitrogen. All cell lines were bought from the American Type Culture Collection (ATCC).

Two fresh human glioblastoma multiform specimens were collected from patients, including GBM10302 and GBM 9746. Samples were submitted to the study under a protocol approved by the Review Board of the University of California, San Francisco. Tissue samples were obtained at primary resection, and none of the patients had undergone prior chemotherapy or radiation therapy. The samples were immediately cut into parts estimated at 100mg per piece, and stored in -80°C freezer.
Real-time PCR

Total RNA was obtained using mirVana™ miRNA Isolation Kit (ThermoFisher, USA), and the RNA quality and quantity were evaluated using the NanoDrop 2000 (Thermo Scientific, Waltham, MA). Reverse transcription (RT) was performed using first-strand cDNA synthesis system (Clontech, USA), according to the manufacturer’s instructions. For each RT reaction, 0.5μg of total RNA was used to synthesize the cDNA in 10μl volume which was diluted to 100μl volume after the reaction. Real-time PCR amplification was performed using 1μl of cDNA, 2xSYBR green supermix (Clontech, USA), and 5pmol of each miRNA 5’ primer and common primer in 10 μl total volume. Samples were subjected to 40 cycles of PCR amplification. In each PCR cycle, denaturing was at 95°C for 5 sec, and annealing at 60°C for 20 sec [12].

Because no normal tissue was available, we used human PC3 prostate cancer cells as a control reference and U6 as an internal control, and set \( \Delta \Delta C_t = (\text{GBM or cell lines} - \text{PC3}) \). The expression levels of miRNAs should be 2\(^{-\Delta \Delta C_t} \). In this study, we used two GBM samples: G10302 and G9746, one GBM cell-line: U87, and two prostate cancer cell-lines: PC3 and Du145 for real-time PCR confirmation. All experiments were performed in duplicate and repeated three times.

RESULTS

Predicted high-confident miRNAs in GBM tissues

The discovery of miRNAs provides an opportunity in understanding the full scope of the post-transcriptional gene regulation in cancer. Therefore, it is critical to predict and identify GBM associated miRNAs in a global and unbiased manner. The RRSM is a bioinformatics algorithm that is used for investigating the interactions between miRNAs and mRNAs from their expression profiles and motif complementary sequences. In this study, we considered the miRNA and mRNA expression data for 114 human miRNAs and 41,699 mRNAs across a mixture of 23 GBM samples and eight normal samples. We focused on the 3195 miRNA–mRNA potential target pairs with high scores selected from 6387 [7], determined each miRNA and its target, and used the results corresponding to the microarray expression 41,699 x 23 data matrix or 41,699 x 8 data matrix and114 x 23 data matrix or 114 x 8 data matrix to fit the regression model. One dataset was filtered from the data to include 3195 potential target pairs, including 584 unique transcripts. In order to select estimated 20 to 30 high confidence miRNA targets, we set \( p_0 = 0.01 \) and \( s = 0.95 \) in RRSM.

After running RRSM with the selected parameters, 86 high confidence miRNA candidates with the expression profiles in GBM tissue were found, and 61 miRNA candidates were selected from the expression profiles in the normal tissues [11]. After comparing the results from the expression profiles in the tumor tissue to those from expression profiles in the normal tissue, 25 high confident miRNAs (Table 1) were selected from those 86 miRNA candidates. In addition, 119 transcripts as those miRNA targets were also selected (not shown). In Table 1, we found 20 candidates that have been studied for glioblastoma from the literature, but the five candidates, miR-200b, miR-30d, miR-27b, miR-302a, and miR-99b, in the first column have not been studies for GBM yet.

<table>
<thead>
<tr>
<th>Table 1. The predicted high confidence miRNAs in GBM tissues</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-200b</td>
</tr>
<tr>
<td>hsa-miR-30d</td>
</tr>
<tr>
<td>hsa-miR-27b</td>
</tr>
<tr>
<td>hsa-miR-302a</td>
</tr>
<tr>
<td>hsa-miR-99b</td>
</tr>
</tbody>
</table>

Identified high confidence miRNAs in GBM tissues

The efficiency and accuracy of results from RRSM algorithm have to be confirmed by experiments using GBM tissues or cultured cells. Quantitative real-time PCR is often used to identify and confirm the miRNA results obtained from bioinformatics pipelines. After searching literature, we found that 20 predicted candidates were reported as GBM associated miRNAs, and five candidates were not reported in any literature. To confirm all these predicted miRNAs, we performed quantitative real-time PCR experiments to GBM tissues and cancer cell lines. Fig. 1A shows the confirmation of five miRNA candidates that were not confirmed by any literature, in which four miRNAs expressed in GBM tissues and U87 cell line except miR-302a. All expression of these four miRNAs was down-regulated in GBM tissues and U87 cell line and up-regulated in prostate DU145 cell line. It suggested that the expression patterns were different between GBM cell line and prostate cell line. When the expression levels of potential miRNA candidates in GBM or U87 cell line were at least two folds (\( |\Delta \Delta C_t| \geq 1 \)) greater or less than that in the PC3 cell line, we deemed the significant expression. Fig.1B shows several
significant expressions of miRNA candidates predicted by RRSM while compared with the PC3 cell line. According to the quantitative real-time PCR experiments, miR-148 and miR-99a were up-regulated dramatically in GBM tissues or cell line, in which the increased expression of miR-99a was also supported by previous literature [13]. In addition, there were other five significant expressions of miRNAs, including miR-29b, miR-25, miR-130a, miR-27a and miR-106b, as cancer biomarker candidates, showed in Fig. 2A and listed in the second column of Table 1. It suggests that those significantly expressed miRNAs might be GBM specific miRNAs because their expressions were significantly different with the expression in PC3 prostate cancer cells. In all these candidates, only miR-302a, miR-26b, and miR-181c could not be detected in GBM and U87 cell line, and miR-26b [14] and miR-181c [15] were reported down-regulated in GBM or GBM cell lines, and miR-302a could be detected in prostate cancer cell lines. Importantly, all other miRNA candidates were expressed in two GBM samples (G10302 and G9746) from the tissue bank of UCSF and U87 cell line. Their expression was not showed because it was not significantly changed compared to the expression in the PC3 cell line; therefore, those miRNAs might be general cancer associated miRNAs. Since no normal tissue was available, human PC3 prostate cancer cells were used as a control reference, and the expression of every miRNA was compared to the expression in the PC3 cell line. Even so, these findings indicate that the candidates from RRSM were highly specific for GBM, and the expression patterns in different prostate cell lines were similar.

![Fig 1.](image)

**Table 2.** MiRNAs as Cancer Potential Biomarkers

<table>
<thead>
<tr>
<th>Name</th>
<th>Cancer Type</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRNA-29b</td>
<td>Gastric cancer</td>
<td>Qiu, et al. 2016 [22]</td>
</tr>
<tr>
<td></td>
<td>Ovarian cancer</td>
<td>Meng, et al. 2015 [24]</td>
</tr>
<tr>
<td>miRNA-130a</td>
<td>Gastric cancer</td>
<td>Jiang, et al.2015 [25]</td>
</tr>
<tr>
<td></td>
<td>Breast cancer</td>
<td>Stuckrath, et al. 2015 [26]</td>
</tr>
<tr>
<td>miRNA-27a</td>
<td>Glioblastoma multiform</td>
<td>Rivera-Diaz, et al. 2014 [27]</td>
</tr>
<tr>
<td>miRNA-106b</td>
<td>Glioblastoma</td>
<td>Zhao, et al. 2013 [28]</td>
</tr>
<tr>
<td></td>
<td>Colorectal cancer</td>
<td>Koga, et al. 2013 [29]</td>
</tr>
<tr>
<td></td>
<td>Gastric cancer</td>
<td>Wang, et al. 2014 [31]</td>
</tr>
<tr>
<td></td>
<td>Head and neck cancer</td>
<td>Mitra, et al. 2014 [33]</td>
</tr>
<tr>
<td>miRNA-215</td>
<td>Colon cancer</td>
<td>Chen, et al. 2016 [34]</td>
</tr>
<tr>
<td>miRNA-203</td>
<td>Cervical cancer</td>
<td>Zhao, et al. 2013 [35]</td>
</tr>
<tr>
<td></td>
<td>Cutaneous squamous cell carcinoma</td>
<td>Canueto, 2016 [36]</td>
</tr>
<tr>
<td>miRNA-10b</td>
<td>Breast cancer</td>
<td>Eissa, 2015 [37]</td>
</tr>
<tr>
<td></td>
<td>Glioblastoma</td>
<td>Teplyuk, 2016 [38]</td>
</tr>
</tbody>
</table>


Identified miRNA candidates as biomarkers for GBM

GBM is a very aggressive brain tumor, but early diagnosis of GBM is the most effective method to achieve higher survival rates for patients. Recently, using miRNAs as potential cancer biomarkers was widely reported [16]. Several predicted miRNAs in Table 1 have been reported as potential cancer biomarkers, including miR-29b, miR-25, miR-130a, miR-27a and miR-106b, shown in upper of Table 2.

According to literature, these five miRNAs were also associated with GBM. In addition, we randomly chose five cancer-related miRNAs from literature, including miR-100, miR-107, miR-215, miR-203 and miR-10b shown in Table 2, in which these five miRNAs were also reported as other cancer potential biomarkers. Table 2 shows these candidates that had already been reported from literature to be related to several types of cancer, such as gastric cancer, pancreatic cancer, ovarian cancer, breast cancer, GBM, glioblastoma, colorectal cancer, bladder cancer, head and neck cancer, cervical cancer, and cutaneous squamous cell carcinoma.

Using quantitative real-time PCR, we compared their expression levels in GBM tissues and cell lines. miR-25, miR-130a, and miR-106b were up-regulated in GBM tissues and U87 cell line [Fig. 2A]. four miRNAs were demonstrated to be down-regulated in GBM tissues and cell line except for miR-10b that was up-regulated [Fig. 2B]. We focused on miRNAs whose expression levels were up-regulated because those candidates may be detected in blood or cerebrospinal fluid and could be served as potential biomarkers for GBM early detection.

DISCUSSION

In this study, we predicted and identified 25 GBM associated miRNA candidates using RRSM algorithm and quantitative real-time PCR technique from the profiles of miRNAs and mRNAs, in which we considered high scores of motif complementary sequences in those miRNA and mRNA interacting pairs. In this way, 22 miRNAs have been confirmed to express in GBM tissues and U87 cell line. Meanwhile, we also studied several candidates as potential GBM biomarkers. In this study, we initially combined advances in the bioinformatics pipeline with existing datasets, several target prediction computational algorithms, and further cellular and molecular studies in GBM tissues and cell lines in a comprehensive and non-biased manner. The results may reveal clues to the mechanism of transcriptional and post-transcriptional regulation in the development of glioblastoma and new information about the genetic networks related to the disease.

Bioinformatics pipelines and algorithms have been used to comprehensively and unbiasedly study interactions between miRNAs and mRNAs. Currently, researchers need a method to precisely predict the potential miRNA targets, before using experimental approaches that can not be skipped because they allow a better functional characterization of miRNAs in biological and disease processes. Computational prediction pipelines provide a rapid method to identify potential miRNA targets. RRSM has successfully discovered many high confidence human miRNA targets from the microarray expression data of the miRNAs and the mRNA in cancer studying [11].

Here, we demonstrated that the RRSM pipeline provided an accurate target prediction, which was confirmed by quantitative real-time PCR experiments and literature. In fact, it is now possible to identify endogenous target sites by sequencing this co-immunoprecipitate with miRISC factors using techniques such as cross-linking immunoprecipitation (CLIP) coupled with high-throughput sequencing (CLIP–seq), or high-throughput sequencing together with CLIP (HITS–CLIP) [17,18]. However, one limitation of CLIPs analysis is that alarge number of potential targets generated that cannot be...
confirmed without conducting extensive experiments. Thus, the combination of informatics pipelines and molecular and cellular experiments is more efficient at selecting and confirming miRNA targets, especially in complex diseases such as GBM.

Given that different complementary sequences use the different mechanism to suppress target genes, we should consider them differently. Generally, the more extensive of the complementarity is to the seed region, the stronger is the response of the target mRNA level to miRNA expression changes [19]. In this study, we only considered those miRNA and mRNA sequence complementary pairs that have high scores from miRanda [8] because those transcripts usually use the similar mechanism to regulate target genes. The lower score pairs were excluded because the effects of the complementary motifs with low score might be much lower in the regulation of mRNA expression. Indeed, six-nucleotide sites that pair strictly with the miRNA seed region typically have a smaller effect on target mRNA expression compared to eight or seven-nucleotide sites, and have been termed marginal sites [20].

Those bioinformatics pipelines did not take tissue specificities and different complementary motifs into consideration, and some seed-related miRNAs have distinct domains of expression and thereby repress their common targets in different tissues. For example, in mice, the retinoblastoma-like 2 protein, which has a crucial role in cell division, is regulated by the miR-290-295 cluster in embryonic stem cells, and by the miR-17-92a cluster during adipogenesis [20]. Therefore, we still need to develop better bio-informatics algorithms to consider more biological details in order to obtain more accurate results.

The results of this study should provide some clues in miRNAs post-transcriptional regulation in the pathogenesis of GBM and new information about GBM biomarkers although we used PC3 prostate cancer cell line as the reference. In fact, it is very difficult to obtain suitable brain tissues as references regardless their locations. Indeed, we provided several facts associated to GBM: 1) 22 predicted miRNAs expressed in GBM and U87 cell line confirmed by quantitative real-time PCR experiments; 2) More than ten miRNAs from 25 predicted candidates could be considered as GBM specific miRNAs because their expressions were significantly different compared to the expression in PC3 cell line; 3) Ten miRNA candidates were investigated and compared as GBM potential biomarkers.

In this study, the results should be examined carefully since prostate PC3 cell line was used as the tissue reference while calculating ΔΔCt. We discovered that two miRNAs, miR-26b and miR-181c, were not detectable in GBM tissues and all three cancer cell lines while miR-302a was not detectable only in GBM tissues and cell line but detectable in prostate cell lines. In order to determine the accurate expression levels of each miRNA in GBM tissues, proper normal tissues should be sued as references to detect these miRNAs. In addition, let-7e had a lower expression in Du145 cell line but not in GBM tissues and U87 cell line compared to the expression in the PC3 cell line, but it was reported to have a lower expression in GBM tissues and GBM stem cells in recently published studies [21].

Further, false positive and false negative should always be considered when using and studying bioinformatics algorithms. In this study, false positive results were unlikely since all results were confirmed by real-time PCR. However, since we chose certain parameters to limit the number of predicted candidates that was suitable for quantitative real-time experiments, this process might cause some false negative results. Indeed, we found that miR-142 (data not shown) and miR-10b were up-regulated in GBM tissues, but not predicted by RRSM. Furthermore, we need to enhance our miRNA studying pool to broaden the research capacity and to increase GBM samples to further confirm the results.

REFERENCE

9. Lewis, B.P.; Burge, C.B.; Bartel, D.P. Conserved seed pairing, oftenflanked by adenosines, indicates that thousands of human genes are microrna targets. Cell 2005, 120, 15


