Investigating Gene and MicroRNA Expression in Glioblastoma

Hua Dong¹², Hoicheong Siu¹, Li Luo², Xiangzhong Fang³, Li Jin¹ and Momiao Xiong²¹

1. State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, 200433, China
   E-mail: hdong0425@gmail.com
2. Human Genetics Center, University of Texas School of Public Health, Houston, TX 77030, USA
3. School of mathematical sciences, Peking University, Beijing, 100871, China
   E-mail: momiao@gmail.com

Abstract—Glioblastoma multiforme (GBM) is the most common primary brain tumor in adults. Here we present an integrated analysis of microRNA expression and gene expression in 237 tumor tissues and 10 normal tissues. We indentified 1,236 genes, and 131 pathways significantly differentially expressed between tumor and normal tissues. We also indentified 98 differentially expressed microRNAs, 22 of which have been reported to affect glioblastoma and 73 of which to affect cancers and brain diseases. We found one experimentally validated microRNA target gene and 1,094 miRNA-target gene pairs in our datasets which were predicted by miRanda algorithm, 16 of the 661 target genes are tumor suppressor genes and 5 are oncogenes. These findings gave important clues to study of the carcinogenic process in glioblastomas.

Keywords - glioblastoma; microRNA; gene expression; pathway;

I. INTRODUCTION

Glioblastoma multiforme (GBM) is the most common and most aggressive type of primary brain tumor, accounting for 52% of all primary brain tumor cases and 20% of all intracranial tumors [1]. Primary GBM arise de novo, without any history of preexisting lower-grade tumor, whereas secondary GBMs have clinical, radiologic, or histopathologic evidence of malignant progression from preexisting lower-grade tumor [2]. In the past two decades, the molecular mechanisms, genetics and paths to treatment of Glioblastoma have extensively been studied [3]. However, the causes and pathogenesis of glioblastoma have not been indentified clearly. With the continuing improvement of high-throughput genomic technologies, it is now feasible to survey human cancer genomes comprehensively. The Cancer Genome Atlas (TCGA) aims to catalogue and discover major cancer-causing genome alterations in large cohorts of human tumors through integrated multi-dimensional analyses [4]. Glioblastoma is the first cancer studied by TCGA. To identify the genetic alterations in glioblastoma, we investigated the expression profiles of gene and microRNA.

MicroRNAs (miRNAs) are single-stranded short coding RNA molecules of about 22 nucleotides in length, which usually repress gene expression by binding at the 3'UTR region of target gene[5]. The expressions of microRNAs are found to be highly different in organ development and tissue differentiation [6]. Moreover, many microRNAs have been found to associate with apoptosis and cancer, suggesting they function as oncogene or tumor suppressor gene[7]. In our study, we examined the expression levels of 470 human miRNAs in glioblastoma and indentify a group of microRNAs whose expression is significantly altered in this tumor. We also indentified the significantly altered gene expression and pathways related to glioblastoma.

II. METHODS

A. DAVID bioinformatic resources

The Database for Annotation, Visualization and Integrated Discovery (DAVID) provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes [8,9] (http://david.abcc.ncifcrf.gov/). After submission large gene lists, it automatically calculates and identifies enriched biological themes, particularly GO terms; discovers enriched functional-related gene groups and clusters redundant annotation terms.

B. Pathway-based differential expression analysis

For pathway analysis, we used algorithm proposed in TAPPA (Topological Analysis of Pathway Phenotype Association)[10]. It calculated a Pathway Connectivity Index for each pathway and then evaluates its correlation to the phenotype variation. Totally 501 pathways from KEGG[11] and Biocarta[12] were assembled in our analysis. Bonferroni correction was used for adjusting multiple tests.

C. Statistical analysis

T-test and Mann-Whitney test were used to test the differential expression of gene and microRNA. Linear regression was used to investigate the relationships among microRNA and gene expressions. The linear model took its common form: \( y = \beta_0 + \epsilon \) where \( y \) is an n-by-1 vector of observations, such as gene expression. \( X \) is an n-by-p matrix of regressors, such as microRNA expression, \( \beta \) is a p-by-1 vector of parameters, known as regression coefficient and \( \epsilon \) is an n-by-1 vector of random disturbances.

III. RESULTS

All types of data were acquired from TCGA project [4] (http://cancergenome.nih.gov/dataportal/data/about/). Gene expression microarrays were performed on Affymetrix HT
The pathway analysis. A total of 1236 genes were significantly related to this tumor. The 131 pathways belonged to 33 functional groups, among which Cell Signaling, Neuroscience, Immunology and Expression were the most enriched pathway groups. This suggested that the differentially expressed genes were most involved in signal and neuroscience pathway. Glioma pathway was the only significant pathway in the cancer functional group with $P$-value $= 5.75 \times 10^{-7}$. The pathway functional groups and the number of pathways in the groups were shown in Table 2. Some pathways may belong to different functional groups. The 16 significant pathways related to Neuroscience, Nervous System and brain diseases were shown in Table 3 in page 5.

### Table 1 The Top Ten GO Terms Most Enriched in Our Significant Gene List

| GO Term | Annotation                                      | P-Value |
|---------|-------------------------------------------------|---------|
| GO:0045202 | synapse                                         | 1.39E-15 |
| GO:0006810 | transport                                       | 3.10E-14 |
| GO:0007268 | synaptic transmission                            | 8.42E-14 |
| GO:0019226 | transmission of nerve impulse                    | 1.05E-13 |
| GO:0051234 | establishment of localization                    | 6.89E-13 |
| GO:0043005 | neuron projection                                | 8.64E-13 |
| GO:0051179 | localization                                     | 2.36E-12 |
| GO:0016020 | membrane                                        | 8.36E-10 |
| GO:0015075 | ion transmembrane transporter activity           | 1.98E-09 |
| GO:0007269 | neurotransmitter secretion                       | 2.62E-09 |

### B. Pathway analysis

We used TAPPA method [10] to identify pathways relevant to glioblastoma. The results revealed that 131 pathways are significantly related to this tumor ($P$-values $< 1.00 \times 10^{-4}$). The 131 pathways belonged to 33 functional groups, among which Cell Signaling, Neuroscience, Immunology and Expression were the most enriched pathway groups. This suggested that the differentially expressed genes were most involved in signal and neuroscience pathway. Glioma pathway was the only significant pathway in the cancer functional group with $P$-value $= 5.75 \times 10^{-7}$. The pathway functional groups and the number of pathways in the groups were shown in Table 2. Some pathways may belong to different functional groups. The 16 significant pathways related to Neuroscience, Nervous System and brain diseases were shown in Table 3 in page 5.

### Table 2 Significant Pathway Function Distribution

| Pathway Functional Group       | Count |
|--------------------------------|-------|
| Cell Signaling                 | 59    |
| Neuroscience                   | 15    |
| Immunology                     | 13    |
| Expression                     | 12    |
| Apoptosis                      | 11    |
| Cytokines/Chemokines           | 10    |
| Cell Cycle Regulation          | 7     |
| Developmental Biology          | 7     |
| Signal Transduction            | 7     |
| Amino Acid Metabolism          | 6     |
| Cell Activation                | 4     |
| Metabolism                     | 4     |
| Endocrine System               | 3     |
| Energy Metabolism              | 3     |
| Glycan Biosynthesis and Metabolism | 3   |
| Hematopoiesis                  | 3     |
| Lipid Metabolism               | 3     |

18
C. Analysis of differential expression of microRNA

As we did for gene expression data, t test was used to test the differential expression of miRNA. After Bonferroni correction for multiple tests (P <9.36×10^-5), 98 microRNA were significantly differentially expressed (Mann-Whitney Test got the same result). To check whether these microRNAs were associated with glioblastoma, we used miR2Disease[14] to validate our results(Updated Date: Dec.19, 2008). miR2Disease provides a comprehensive literature reported resource of microRNA deregulation in various human diseases. From the data in miR2Disease, 82 of the 98 significant microRNAs have been reported to associate with 84 diseases, among which, 73 microRNAs are related to 59 cancers and brain diseases. 22 of those microRNAs have been reported to induce glioblastoma/ glioblastoma multiforme(GBM)/neuroblastoma (NB) and the expression pattern of microRNA is exactly the same as that in our data. Table 4 gave the p-value, expression pattern, disease and references for the 22 microRNAs. We infer that the other 51 microRNAs related to cancers and brain diseases may also be important for carcinogenesis in brain. However, further experiment validations are required to confirm our results.

Among the 98 significant microRNAs, 30 microRNA were up-regulated and 68 down-regulated. We hypothesis that in brain tumor tissues, the up-regulated microRNAs function as oncogene and target tumor suppressor gene in brain, while down-regulated microRNAs function as tumor suppressor gene and repress the expression of oncogene related to brain tumor. On the contrary, in the brain tissue of normal people, those up-regulated microRNAs should express much less in brain than in other tissues and those down-regulated microRNAs should express much higher in normal brain tissue than in other normal tissues.

To prove our hypothesis, fist of all, we checked the expression of microRNAs in normal tissues using Yu Liang’s data[15], they provided expression data of 345 microRNAs in 40 normal human tissues, of which 81 microRNAs can be matched with our significant microRNAs. Among the 30 brain tumor tissue up-regulated microRNAs, 87.5% were under-expressed in normal brain tissue. Among the 68 brain tumor tissue down-regulated microRNAs, 75% were over-expressed in normal brain tissue. For the 22 brain tumor related microRNAs, 9 of the 10 up-regulated microRNAs were under-expressed in normal brain tissue, the left 1 had no obvious change (fold change =1.051). For the 12 down-regulated microRNAs, 10 of them were over expressed in normal brain tissue and 2 were also under-expressed. The fold changes of the 22 microRNAs in normal tissues were shown in table 4.

To check whether up-regulated microRNAs function as oncogene and down-regulated microRNAs function as tumor suppressor gene, we need to find the target gene of microRNA associated with glioblastoma. So we did the regression analysis for microRNA and gene expression.

*miRNA: all 22 miRNAs are human miRNA, “hsa-” was omitted. *FC is the fold change of expression value in brain tissue versus the average expression value in other 39 tissues of normal people in Yu Liang’s data. Our microRNA expression data is quartile normalized and we could not calculate the fold change. *Exp means whether the expression pattern of microRNA is up-regulated or down-regulated.

**Table 4 22 microRNAs related to glioblastoma/ GBM/ NB**

| *miRNA  | Pval  | *FC   | Disease     | *Exp   | Ref.  |
|---------|-------|-------|-------------|--------|-------|
| mir-21  | 3.08E-24 | 0.136 | glioblastoma | up     | [25,26,27] |
| mir-23a | 1.45E-16 | 0.048 | glioblastoma | up     | [25]   |
| mir-93  | 5.54E-12 | 0.010 | NB           | up     | [28]   |
| mir-25  | 1.85E-11 | 0.834 | glioblastoma | up     | [25]   |
| mir-155 | 1.40E-10 | 0.188 | GBM          | up     | [27]   |
| mir-92  | 2.52E-08 | 0.953 | NB           | up     | [28]   |
| mir-210 | 1.74E-07 | 0.588 | GBM          | up     | [27]   |
| mir-130a| 3.84E-07 | 0.333 | glioblastoma | up     | [25]   |
| mir-106a| 6.38E-06 | 0.574 | NB           | up     | [28]   |
| mir-17-5p| 2.03E-06 | 0.666 | NB           | up     | [28,29] |
| mir-323 | 5.78E-36 | 24.013 | GBM          | down   | [27]   |
| mir-137 | 1.86E-31 | 32.851 | GBM          | down   | [27]   |
| mir-128a| 5.47E-26 | 21.499 | GBM          | down   | [27]   |
| mir-154*| 7.62E-23 | 13.758 | GBM          | down   | [27]   |
| mir-153 | 1.06E-21 | 23.286 | GBM          | down   | [27]   |
| mir-132 | 6.84E-21 | 7.362 | GBM          | down   | [27]   |
| mir-7  | 2.53E-18 | 3.044 | glioblastoma | down   | [30]   |
| mir-124a| 1.76E-17 | 137.31 | GBM          | down   | [27]   |
| mir-133b| 1.36E-11 | 0.012 | GBM          | down   | [27]   |
| mir-29b | 1.42E-10 | 1.201 | GBM          | down   | [27]   |
| mir-149 | 2.87E-08 | 13.178 | GBM          | down   | [27]   |
| mir-133a| 9.15E-08 | 0.01   | GBM          | down   | [27]   |

| *miRNA  | Pval  | *FC   | Disease     | *Exp   | Ref.  |
|---------|-------|-------|-------------|--------|-------|
D. The regulation of gene expression by microRNA

miRNA has been thought to promote degradation of target mRNA or suppress translation of corresponding protein by matching with mRNA in the 3'UTR region.[16-19] For the shared 247 samples in gene expression and microRNA data for 12043 genes and 470 human microRNAs, we did linear regression to search the possible target genes of microRNAs in glioblastomas. For the original result, we did the following four step filtering: 1) Bonferroni correction: we chose all the record with regression p-value < 1.064×10^-4. 2) As microRNA represses the expression of gene, if a gene is the direct target of microRNA, the regression coefficient β should be negative; otherwise it’s not direct targeted by microRNA. 3) As microRNA always represses the expression of gene, the expression pattern of microRNA and its target gene should be opposite. 4) As we want to see the microRNAs and genes relevant to glioblastomas, both the differential expression p-values of microRNAs and genes should be significant. So we filtered the miR-Gene pairs with regression p-value < 1.064×10^-4, β<0, opposite expression pattern, gene differential expression p-value < 4.15×10^-6 and microRNA differential expression p-value < 1.064×10^-4. This reduced the original 5,660,210 records to 31624 records.

To check whether these genes were the real targets of microRNAs, we compared our results with microRNAs databases: miR2Disease[14] and miRBase[20]. MiR2Disease provides experimentally verified microRNA target genes and literature reference (updated on Dec.19, 2008), which contains more plentiful and updated records than that in TarBase[21]. MiRBase predict the target gene of microRNA by miRanda algorithm [22], where the predicted target genes and microRNAs could be downloaded (updated on: Oct.31, 2007). Finally we found one matched experiment validated result. The literature reported that the in nasopharyngeal carcinomas hsa-mir-29c (expression fold change(tumor/normal)=0.20) target COL4A1 expression fold change(tumor/normal)=5.24) [23]. In our result, down-regulated hsa-mir-29c (differentially expressed P-value < 5.11×10^-12 ) targets over-expressed gene COL4A1 (differentially expressed P-value < 3.58×10^-6 ) with regression β = -389.02 and P = 1.35×10^-8. We conclude that hsa-mir-29c is also an important miRNA in glioblastomas. For predicted targets in miRBase, we found 1,094 matched miR-gene pairs including 70 microRNAs and 661 genes. 44 down-regulated microRNAs target 202 over expressed genes while 26 up-regulated microRNAs target 459 under expressed genes. To get a clear observation of the microRNA-gene target relationship, we showed that the up and down-regulated microRNA-gene pairs in Figure 1(a) and (b).

IV. Discussion

In this report, we performed detailed analysis of differential expression of gene and microRNA between glioblastomas tumor tissues and normal brain tissues. We also conducted the pathway-based analysis. However, we did not perform genetic analysis. As the controls are from normal tissues of patients too, we could only investigate the influence of somatic mutation to disease. In future, we will further investigate how somatic mutation influence gene expression and disease phenotype. If there are normal people used as control, we could find more genetic factors and germlne mutation related to carcinogenic process in glioblastomas.

Acknowledgment

We thank the Cancer Genome Atlas Research Network for providing data and the members of TCGA’s External Scientific Committee and the Glioblastoma Disease Working Group (http://cancergenome.nih.gov/components). Li Jin, Hua Dong and Hoicheong Siu are supported by grants from the National Outstanding Youth Science Foundation of China (30625016), National Science Foundation of China (30890034), Shanghai Commission of Science and Technology (04dz14003) and 863 Program (2007AA02Z312). Momiao Xiong and Li Luo are supported by Grant from National Institutes of Health NIAMS P01 AR052915-01A1, NIAMS P50 AR054144-01 CORT, HL74735, ES09912, and Grant from Hi-Tech Research and Development Program of China(863)(2007AA02Z300).

References

[1] wikipedia: http://en.wikipedia.org/wiki/Glioblastoma_multiforme#Caususes

[2] F.B. Furnari, T. Fenton, R.M. Bachoo, A. Mukasa, J.M. Stommel, A. Steghl, W.C. Hahn, K.L. Ligon, D.N. Louis, C. Brennan, L. Chin, R.A. DePinho, and W.K. Cavenee, Malignant astrocytic glioma: genetics, biology, and paths to treatment. Genes Dev 21 (2007) 2683-710.

[3] P.S. Mischel, S.F. Nelson, and T.F. Cloughesy, Molecular analysis of glioblastoma: pathway profiling and its implications for patient therapy. Cancer Biol Ther 2 (2003) 242-7.

[4] The Cancer Genome Atlas Research Network, Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455 (2008) 1061-8.

[5] S. Sassen, E.A. Miska, and C. Caldas, MicroRNA: implications for cancer. Virchows Arch 452 (2008) 1-10.
Table 3 Significant pathways related to Neuroscience

| Pathway Symbol | Function | Gene No. | p-value |
|----------------|----------|----------|---------|
| Long-term depression | Nervous System | 78 | 2.61E-08 |
| Long-term potentiation | Nervous System | 71 | 2.74E-08 |
| Endocytotic role of NDK, Phosphins and Dynamin pathway | Neuroscience | 21 | 4.01E-08 |
| Phosphorylation of MEK1 by cdk5/p35 down regulates the MAP kinase | Cell Signaling/Neuroscience | 11 | 5.21E-08 |
| Reelin Signaling Pathway pathway | Cell Signaling/Neuroscience | 7 | 6.28E-08 |
| Blockade of Neurotransmitter Release by Botulinum Toxin pathway | Neuroscience | 5 | 7.23E-08 |
| Repression of Pain Sensation by the Transcriptional Regulator DREAM | Cell Signaling/Expression Neuroscience | 14 | 2.03E-07 |
| Lissencephaly gene (LIS1) in neuronal migration and development | Developmental Biology Neuroscience | 6 | 2.32E-07 |
| Synaptic Proteins at the Synaptic Junction pathway | Neuroscience | 4 | 4.23E-07 |
| Gamma-aminobutyric Acid Receptor Life Cycle pathway | Neuroscience | 10 | 4.94E-07 |
| Glioma | Cancers | 65 | 5.75E-07 |
| Rac1 cell motility signaling pathway pathway | Cell Signaling/Neuroscience | 23 | 3.17E-06 |
| TrkA Receptor Signaling Pathway pathway | Cell Signaling/Neuroscience | 12 | 5.74E-06 |
| Role of Erk5 in Neuronal Survival pathway | Cell Signaling/Neuroscience | 18 | 7.80E-06 |
| Amyotrophic lateral sclerosis (ALS) | Neurodegenerative Diseases | 19 | 1.13E-05 |
| Deregulation of CDK5 in Alzheimer's Disease pathway | Neuroscience | 11 | 1.75E-05 |
Figure 1. microRNAs and target gene relationships: (a) down-regulated microRNA and their targets. (b) up-regulated microRNA and their targets.