We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Integrated Network Analysis of Genetic and Epigenetic Factors in Glioblastoma Multiform

Hua Dong[1] and MomiaoXiong[2,1]
[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,
[2]Human Genetics Center, University of Texas School of Public Health,
Houston,TX
[1]China
[2]USA

1. Introduction

Glioblastoma is the most common and aggressive type of primary brain tumor in humans. It is located preferentially in the cerebral hemispheres. Glioblastoma arises from complex interactions between a variety of genetic, epigenetic alterations and environmental perturbations. However, the precise mechanism of glioblastoma is unknown and its survival rate is very low. The Cancer Genome Atlas (TCGA) generates large-scale multi-dimensional genetic and epigenetic data to catalogue and identify cancer causing alterations (Kuhn, et al., 2008). Glioblastoma (GBM) is the first cancer studied by TCGA. In TCGA glioblastoma pilot project, a total of 601 genes were sequenced for detection of somatic mutations in 179 tumor and matched normal tissues pairs; expressions of 12,042 genes were measured in 243 tumor tissue samples and 10 normal tissue samples and 1 cell line; expressions of 534 miRNAs were profiled in 240 tumor tissue samples and 10 normal tissue samples and a total of 2,994 genes were examined for methylation in 239 tumor tissue samples and 1 cell line. This dataset will be used as an example for developing system biology and network approach as a general framework for integrated analysis of genetic and epigenetic alternations in cancer studies.

Biological functions and mechanisms are encoded in network properties. An important strategy for unraveling the mechanisms of initiation and progression of cancer is to conduct analysis of complex genetic and epigenetic networks and study their behaviors under genetic and epigenetic perturbations. Robustness of a biological network, ability to retain much of its functionality in the face of perturbation (Dartnell, et al., 2005), has emerged as a fundamental concept in the study of network topological properties (Demetrius and Manke, 2005). The locations of the DNA variants, mRNA, miRNA, and methylation in the genetic and epigenetic networks are likely to affect the phenotypes. We use network structural analysis as a tool to identify a set of key cancer causing genome alternations and core modules of biological networks that play an essential role in the development of cancer.

Purpose of this report is to use system biology approaches to develop novel analytic strategies for systematically integrating genetic and epigenetic data. To achieve this, we

reconstruct genetic and epigenetic networks involved in tumorigenesis and study how these networks respond to perturbation such as somatic mutations and methylation. Information that defines the path from genomic variations to tumors flows from mutations and methylation through mRNA and miRNA expressions to tumor formation. Therefore, the components of complex genetic and epigenetic networks that determine cell function and response to perturbation of external stimuli include genes harboring DNA variations, mRNAs, miRNAs, and methylation connected by their co-expressions or interactions. These networks consist of three levels. The subnetworks in the first level consist of mRNA and miRNA co-expression networks, mRNA and miRNA interaction networks. The subnetworks in the second level are miRNA target gene networks. The subnetworks in the third level are eQTL networks which connect mutations to gene and miRNA expression.

Somatic mutations that are likely to cause tumorigenesis are traditionally identified by comparing differences in mutation frequencies between tumor and normal tissues individually. However, most somatic mutations are rare mutations. Due to their rarity, the frequencies of rare alleles may be compatible with sequencing errors. As a consequence, individual tests of association of rare variants with disease have little power and may not be robust. To overcome this limitation, we develop group association tests in which a group of rare genetic variants are jointly tested for assessing association of rare mutations and LOH with cancer. Multiple rare mutations, each with a minor marginal genetic effect, but collectively may make big contributions.

We reconstructed two types of networks that connect mRNAs or miRNAs. One is mRNA or miRNA co-expression networks. To take inherent sparse structure of co-expression networks, we use partial correlation method and sparse regression techniques to infer co-expression networks. Although co-expression networks are very useful for revealing transcriptional regulatory actions, they do not directly provide information on differences in regulatory effects between tumor and normal tissues. It is differences in regulatory effects that cause cancer. Therefore, we propose a new concept of a regulatory difference network that is formally referred to as mRNA or miRNA interaction networks. We detect interaction between mRNAs or miRNAs by measuring differences in their mutual information between tumor and normal tissues and develop a novel MI-based statistic to test interaction between two mRNAs or miRNAs.

miRNAs are short endogenous noncoding RNAs of 22 nt that negatively regulate gene expression through base pairing with target mRNAs (Huang, et al., 2007; Huang, et al., 2007). It is increasingly recognized that miRNAs have emerged as an important component in the regulation of gene expression, with imperfect base pairing, to target sites in the 3' UTR of messenger RNAs (Gennarino, et al., 2009). Three types of methods have been used to identify potential target genes: sequence analysis, miRNA-mRNA regression analysis, and machine learning (Huang, et al., 2007; Maziere and Enright, 2007; Yang, et al., 2008). Sequence analysis methods rely on knowledge of the base pairing between and second structure of the miRNA and the target gene. However, the sequence approach has limited specificity due to imperfect miRNA-target pairing (Huang, et al., 2007; Huang, et al., 2007). Regressing the expression of a target gene on the expression of miRNA for identifying miRNA target is based on a linear relationship between miRNA and its target mRNA. However, the relationship between miRNA and target mRNA may be nonlinear, which leads to inaccurate prediction of the miRNA target. The precision of miRNA target prediction by machine learning depends on the number of experimentally validated miRNA targets; however, the number of experimentally

validated targets is limited. To improve the accuracy of target prediction, we will combine sequence analysis with regression analysis for target prediction.

2. Materials and methods

2.1 Test association of somatic mutations and LOH with glioblastoma

Cancers arise from mutations that confer growth advantage on cells (Greenman, et al., 2007).The somatic mutations in cancers can be classified either as "drivers" or "passengers" (Wood, et al., 2007). As the number of tumor tissues and normal tissues increases we can observe somatic mutations in both tumor and normal tissues. The current popular method for identifying driver mutations is to compare the difference in the mutation rates (Kuhn, et al., 2008; Parsons, et al., 2008). However, there is debate about how to assess a significant excess of mutations in tumors (Rubin and Green, 2007). We need to develop formal tests to detect differences in mutation rates between tumor and normal tissues. Most traditional statistical methods that often test the association of genetic variants individually were designed for testing association of common alleles with common diseases and are inappropriate for testing the association of rare somatic mutations. A feasible approach is to record rare sequence variants at different genome positions and to collectively test association of a set of rare variants. It has been shown that the number of rare alleles in large samples is approximately distributed as a Poisson process with its intensity depending on the total mutation rate(Joyce and Tavare, 1995). The intensity of the Poisson process within a segment of genome can be interpreted as the mutation rate. Similar to standard χ^2 test for association of SNPs which compare the difference in allele frequencies between cases and controls, the proposed statistics are to compare difference in the mutation rates between tumor and normal samples. Specifically, let $\bar{U}$ and $\bar{V}$ be the average number of rare mutations which is the intensity of the Poisson process underlying the rare variants, in the tumor and normal samples, respectively. Let S_{uv} be the pooled sample variance of the rare variants. Define the test statistic:

$$T_G = \frac{(\bar{U} - \bar{V})^2}{(\frac{1}{n_A} + \frac{1}{n_G})S_{uv}}, \tag{1}$$

where n_A and n_G are the number of sampled tumor tissues and normal tissues, respectively. Under the null hypothesis of no association of the set of rare variants with the disease, the average number of rare alleles in cases and controls should be equal and the statistic T_G is asymptotically distributed as a central $\chi^2_{(1)}$ distribution. In some cases, we may have a homozygous genotype of rare mutations. To improve the power, in this case, we can count it twice. Instead of defining statistics in terms of genotype, we can similarly define the test statistics in terms of the rare alleles, which is denoted as T_a.

$$T_a = \frac{(\bar{U} - \bar{V})^2}{(\frac{1}{2n_A} + \frac{1}{2n_G})S_{uv}}$$

To examine the validity of the test statistics, we performed a series of simulation studies. We used infinitely many allele models and software (GENOME) to generate rare variants. Suppose that the mutation rate per generation per base pair is 1.00×10^{-8}, the recombination rate between consecutive fragments is 0.0001, the migration rate per generation per individual is 0.00025, we have simulated 100 fragments with the length of each fragment equals 10k in base pair. Total of 100,000 individuals who were equally divided into cases and controls were generated in the general population. 500 to 2,000 individuals were randomly sampled from each of the cases and controls. 10,000 simulations were repeated. Table 1 summarized the type I error rates of two statistics. Table 1 showed that the estimated type I error rates of the statistics for testing association of a set of rare variants with the disease were not appreciably different from the nominal levels $\alpha = 0.05$, $\alpha = 0.01$ and $\alpha = 0.001$.

A LOH mutation was recorded when the genotype in blood or normal tissue is heterozygous, and in the tumor tissue, the reference allele loses normal function and the genotype becomes homozygous. The statistic T_G can be used to test association of LOH with glioblastoma.

Sample Sizes		T_G	T_a
2,000	α =0.001	0.0012	0.001
	α =0.01	0.0093	0.0096
	α =0.05	0.0489	0.0494
1,500	α =0.001	0.0015	0.001
	α =0.01	0.0106	0.0092
	α =0.05	0.0521	0.0500
1,000	α =0.001	0.0012	0.0007
	α =0.01	0.0106	0.0097
	α =0.05	0.0522	0.0459
500	α =0.001	0.0009	0.0008
	α =0.01	0.0102	0.0117
	α =0.05	0.0504	0.0548

Table 1. Type 1 error rates of the statistics T_G, T_a.

2.2 Lasso for co-expression networks

A co-mRNA expression or co-miRNA expression network can be constructed by joint sparse regression for estimating the concentration matrix in which off-diagonal elements represents the covariance between the corresponding variables given all other variables in the network (Peng, et al., 2009) . Sparse regression for reconstruction of co-expression network is briefly introduced here. (For details, please see(Peng, et al., 2009)). Denote the mRNA or miRNA expression levels as variables $y_1, ... y_q$. A variable is represented by a node. An edge

connecting two nodes indicates that the connected two variables are conditionally dependent, given all other variables. Assume that the vector of q variables $Y = [y_1, ... y_q]^T$ follow a normal distribution $N(0, \Sigma)$. Denote the partial correlations as $\rho^{ij} = Corr(y_i, y_j | y_{-(i,j)})$ for $1 \le i < j \le q$ and $-(i,j) \equiv \{k : 1 \le k \ne i, j \le q\}$. If we assume the normality of the variables, then two variables y_i and y_j are conditionally dependent, given all other variables if and only $\rho^{ij} \ne 0$. Let $\Sigma^{-1} = (\sigma^{ij})$ be the concentration matrix. Then,

$$\rho^{ij} = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}} \tag{2}$$

When sample size is much larger than the number of variables, the concentration matrix can be directly estimated from the inverse of the sampling covariance matrix. However, when the number of variables in the network is larger than the sample size, the inverse of sampling covariance matrix may not exist. We use sparse regression for network modeling via estimating sparse concentration matrix.

2.3 Functional module identification and gene set enrichment analysis

Co-expression networks are usually organized into functional modules that perform specific biological tasks. Genes within coexpression modules often share conserved biological functions. A dynamic tree cut procedure was used to identify modules(Langfelder, et al., 2008). A co-expression network was clustered using hierarchical clustering. Modules are defined as braches of dendrogram. One sided Fisher exact test that calculates the probability of seeing observed number of genes within a pathway or a GO category in the module by chance was used to test for enrichment of a pathway or a GO category in the module. We assembled 465 pathways from KEGG(Hashimoto, et al., 2006) and Biocarta (http://www.biocarta.com).

2.4 miRNA interaction and mRNA Interaction networks

Modern complex theory assumes that the complexity is the degree to which components engage in organized structured interactions. Intuitively, complexity can also be viewed as the degree of dependence among components. Therefore, interaction has been considered as a sensible measure of complexity of biological systems. The more interactions between components, the more complex the system is. We identify mRNA or miRNA interaction networks via detecting pair-wise interaction between mRNAs or miRNAs. We view mRNA or miRNA expressions as continuous variables. Formally, interactions can generally be defined as a stochastic dependence between two random variables. The concept of mutual information proposed by Shannon(Cover and Thomas, 1991) can serve as a general measure of interaction (dependence) between two random variables(Jakulin, et al., 2003; Matsuda, 2000; Nakahara, et al., 2003). It quantifies the stochastic dependence between two random variables. An additional asset is that mutual information measures more than linear dependence(Brillinger, 2004; McGill, 1954). Therefore, information theory will provide an unified framework for developing statistical methods to detect interaction between mRNAs or miRNAs. Similar to the detection of interaction between two SNPs where interaction between two SNPs is measured and tested as the difference in linkage disequilibrium(LD) between cases and controls(Wu, et al., 2008; Zhao, et al., 2006), we detect interaction

between mRNAs or miRNAs by measuring differences in mutual information between tumor and normal tissues.

The expression of mRNAs and miRNAs are continuous variables. Usually, we assume that they follow a normal distribution. Let random variables X and Y denote the expression of a pair of mRNAs or miRNAs. Let $f(x)$, $f(y)$ and $f(x,y)$ be the distribution functions of X and Y and their joint distribution function, respectively. The mutual information between two continuous variables is defined as

$$I(X,Y) = E\left[\log \frac{f(x,y)}{f(x)f(y)}\right] \tag{3}$$

This formula shows(Cover and Thomas, 1991) that mutual information is always nonnegative and is equal to zero if and only if two random variables are independent. Therefore, mutual information can be used to measure the degree of stochastic dependence. Assuming the normal distribution of two random variables, mutual information can easily be calculated by

$$I(X,Y) = -\frac{1}{2}\log(1-\rho^2), \tag{4}$$

where ρ is a canonical correlation. Similar to using LD to test interaction between two SNPs where tests for interaction between unlinked and linked loci are developed separately, we consider two cases in testing interaction between mRNAs or miRNAs: (1) a pair of mRNAs or miRNAs in the normal tissues are independent and (2) a pair of mRNAs or miRNAs in the normal tissues are dependent.

2.4.1 A pair of mRNAs or miRNAs in the normal tissues is independent

Let $I_T(X,Y)$ and $I_N(X,Y)$ be mutual information between x_1 and x_2 in tumor and normal tissues, respectively. Let n_A and n_G be the number of tumor and normal samples, respectively. Then, $2n_A\hat{I}_T(X,Y)$ and $2n_G\hat{I}_N(X,Y)$ are asymptotically distributed as a $\chi^2_{(1)}$ distribution (Brillinger, 2004). Define the statistic:

$$T_I = 2n_A\hat{I}_T(X,Y) - 2n_G\hat{I}_N(X,Y) \tag{5}$$

It has been shown(Holm and Alouini, 2004) that under the null hypothesis of no interaction, T_I has a distribution with the following probability density function

$$f(x) = \frac{1}{2\pi}K_0(\frac{|x|}{2}) \tag{6}$$

where $K_0(\cdot)$ is the modified Bessel function of the second kind of order 0 (Jeffrey and Zwillinger, 2000).

2.4.2 A pair of mRNAs or miRNAs in the normal tissues is dependent

When a pair of mRNAs or miRNAs in the normal tissues is dependent, we can define a test statistic that has a much simpler distribution than that of test statistic T_I. Let $\hat{\rho}_A$ and $\hat{\rho}_G$ be

the canonical correlation in tumor and normal tissues, respectively. We propose a novel statistic defined as

$$T_D = \frac{[I_T(X,Y) - I_N(X,Y)]^2}{(1/n_A + 1/n_G)\hat{\rho}^2} \qquad (7)$$

where $I_T(X,Y) = -\frac{1}{2}\log(1-\hat{\rho}_A^2)$ and $I_N(X,Y) = -\frac{1}{2}\log(1-\hat{\rho}_G^2)$. Then T_D follows a central $\chi^2_{(1)}$ distribution under the null hypothesis of no difference in mutual information between tumor and normal tissues. To evaluate the validity of the test statistic T_D, we calculated its type 1 error rates. We simulated a population of size 1,000,000 from a bivariate normal distribution with mean $[0,0]^T$. Then, we equally sampled 100, 200 and 400 subjects as tumor and normal samples and calculated the test statistic T_D. We repeated this with 10,000 simulations. The type 1 error rates for different samples and correlation coefficients were listed in Table 2. Type I error rates for the statistics T_D to test for the interactions between two mRNAs or miRNAs were not appreciably different from the nominal levels $\alpha = 0.05$, $\alpha = 0.01$ and $\alpha = 0.001$.

ρ	Sample Sizes	Nominal Level		
		$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.05$
0.2	100	0.0023	0.0113	0.0552
	200	0.0010	0.0086	0.0494
	400	0.0011	0.0115	0.0529
0.5	100	0.0008	0.0103	0.0532
	200	0.0011	0.0108	0.0528
	400	0.0012	0.0104	0.0496

Table 2. Type 1 error rates of T_D for testing interaction between two continuous variables

2.5 Ranking of the nodes in the network

Biological functions and mechanisms are encoded in network properties. An important strategy for unraveling the mechanisms of initiation and progression of cancer is to conduct analysis of complex biological networks and study their behaviors under genetic and epigenetic perturbations. Robustness of a biological network, ability to retain much of its functionality in the face of perturbation(Dartnell, et al., 2005), has emerged as a fundamental concept in the study of network topological properties(Demetrius and Manke, 2005). Widely used measures of network robustness include ranking importance of the nodes in the network. One of the most efficient measures of importance in robustness analysis of the network is the damage value of a node which quantifies the effect of the removal of that particular node from the network. Formally, we define the damage value of a node as follows. Let $G = (V, E)$ be the connected component that contains node $h \ni V$, and

let $\tilde{G} = (\tilde{V}, \tilde{E})$ be the largest connected component of $\tilde{G}$, after the removal of node h. Then, the value $D(h) = |V| - |\tilde{V}|$ is the damage value of node h.

2.6 Identify the genetic variants that have cis or trans regulatory effects on miRNA or mRNA expressions

Loci that are significantly associated with expression are called expression quantitative trait loci (eQTL). Somatic mutations may directly or indirectly regulate the expression of mRNAs or miRNAs. The traditional statistical methods to regress the expression levels on the individual genetic variant for identifying eQTL is inappropriate for studying the regulatory effect of somatic mutations due to their low allele frequencies. An alternative approach to current variant-by-variant regression method is groupwise regression methods in which a group of rare genetic variants are jointly analyzed. It is well known that eQTL include *cis*-eQTL in which an association exists between the expression of a specific gene (mRNA) and the genetic variants at that gene's locus, or between the expression of miRNA and the genetic variants at its host gene, and *trans*-eQTL in which there is an association between the expression of a gene or a miRNA and the genetic variants at a non-local genomic locus. Regression methods that regress the expression of a mRNA or a miRNA on the number of all mutated alleles across the region of interest to identify *cis*- or *trans*-eQTL .

2.7 miRNA target networks

miRNAs down regulate gene expressions by base-pairing with the 3′-noncoding region of the target mRNAs. It is estimated that up to 30% of genes might be regulated by miRNAs(Sassen, et al., 2008). It is hypothesized that miRNAs and their targets form complex networks to perform various biological functions. To reveal mechanisms of the GBM, we identified target genes of miRNAs and constructed miRNA target networks. Procedures for discovering the target genes consisted of two steps. The first step was to conduct sequence analysis that used sequence complementarities of miRNA and its target site to predict potential miRNA target genes. We searched the predicted potential miRNA targets in miRGen(Megraw, et al., 2007), which integrated animal miRNA targets according to combinations of four widely used target prediction programs: miRanda, PicTar, TargetScan, DIANA-microT, and experimentally supported targets from TarBase (Sethupathy, et al., 2006) and miR2Disease(Jiang, et al., 2009). Since miRNAs repress the expression of their target genes, the second step was to test the inverse relationship between the expression profile of miRNA and that of its potential targets. To achieve this, we regressed the expression of target mRNAs on the expression of miRNAs and select mRNAs with significant negative regression coefficients as miRNA targets. P-value for declaring significant evidence of miRNA target was 1.00×10^{-4} .

3. Results

3.1 Test association of somatic mutations and LOH mutations with glioblastoma

The first step for deciphering the path from somatic mutations to the GBM is to test association of somatic mutations with the GBM. The statistic T_a and T_G was applied to glioblastoma in TCGA dataset. The tumor tissues of 179 glioblastoma patients and 179 matched normal tissues were sequenced. A somatic mutation was recorded when the mutation was detected only in the tumor tissue. There were 306 genes in which at least one

somatic mutation was detected. A LOH mutation was recorded when the genotype in blood or normal tissue is heterozygous, and in the tumor tissue, the reference allele loses normal function and the genotype becomes homozygous. There were 124 genes in which at least one LOH mutation was detected. We identified association of somatic mutations in 14 genes by the statistics T_a (Table 3), and association of LOH in 11 genes by the statistic T_G with glioblastoma with false discovery rate (FDR) less than 0.05 (Table 4). Genes TP53, PTEN, EGFR, NF1, RB1 and ERBB2 were reported to be associated with GBM in the previous TCGA data analysis (The Cancer Genome Atlas Research Network, 2008). The remaining 8 somatic mutated genes and 10 LOH mutated genes were newly identified by the statistics T_G or T_a. NCBI Entrez gene database (Maglott, et al., 2007) reported that : CHEK2 is

Gene	P-value	FDR	Mutation frequency
TP53	3.46E-11	5.90E-10	0.1453
PTEN	2.25E-07	1.92E-06	0.0698
EGFR	1.22E-06	6.92E-06	0.0587
FKBP9	1.38E-04	5.89E-04	0.0363
CHEK2	1.39E-03	4.73E-03	0.0475
GSTM5	2.41E-03	6.86E-03	0.0251
DST	4.28E-03	8.12E-03	0.0223
RB1	3.54E-03	8.62E-03	0.0251
NF1	4.17E-03	8.89E-03	0.0363
BCL11A	1.36E-02	2.33E-02	0.0168
ERBB2	1.57E-02	2.43E-02	0.0307
PIK3C2G	2.45E-02	3.49E-02	0.0140
FN1	3.30E-02	4.33E-02	0.0168
COL3A1	4.46E-02	4.75E-02	0.0112

Table 3. P-values for testing association of somatic mutations with the GBM

Gene	P-value	FDR	Mutation frequency
NRAP	5.71E-07	6.24E-06	0.1173
MKI67	8.19E-07	4.47E-06	0.1229
C10orf54	1.02E-04	3.70E-04	0.0391
C9orf66	1.63E-04	4.46E-04	0.0419
MYO3A	6.40E-04	1.40E-03	0.0307
PRAME	2.14E-03	3.90E-03	0.0251
EGFR	3.20E-03	4.99E-03	0.0279
IL1RL1	7.11E-03	9.71E-03	0.0196
HLA-DOA	1.30E-02	1.57E-02	0.0168
ABCA13	2.37E-02	2.59E-02	0.0140
CYP1B1	2.37E-02	2.59E-02	0.0140

Table 4. P-values for testing association of LOH with the GBM

a cell cycle checkpoint regulator and putative tumor suppressor, and associated with GBM; GSTM5 was reported to be involved in cancer, BCL11A is a proto-oncogene, and FN1 is involved in tumor metastasis and angiogenesis. Gene PRAME was reported to be associated with melanoma and acute leukemias. Association of the other 9 genes such as NRAP, MK167, C10orf54 and C9orf66 with GBM was first reported here

3.2 Network analysis of gene expressions
3.2.1 Differential expression analysis

Comparative studies of gene expression between normal and tumor tissues is one of the most widely used strategies for unraveling the molecular circuitry underlying cancer (Liang and Pardee, 2003). To uncover the mechanisms of glioblastoma, expressions of 12,042 genes were measured in 243 tumor tissue samples and 10 normal tissue samples and 1 cell line by Affymetrix HT Human Genome U133 Array Plate Set. A total of 1,697 genes were differentially expressed between tumor and normal tissues by Wilcoxon rank-sum test (P-value for declaring significance after Bonferroni correction is 4.15×10^{-6}). Of the 1,697 genes, 72 genes were cancer genes or cancer candidate genes, 25 of which were GBM related genes including *TCF12, TP53, COL4A1, COL3A1* and *COL5A2*, 11 of them were oncogenes including *CDK4* and *RAF4*; 21 of them were tumor suppressor genes including *TP53* and *RB1* (We got the oncogene and tumor suppressor gene list from database: TSGDB(Yang and Fu, 2003) and UNSW Embryology DNA Tumor Suppressor and Oncogene Database(Ackermann and Strimmer, 2009)), 242 genes were in signal transduction pathways, 908 genes were down regulated and 789 genes were up regulated. From 1,697 significantly differentially expressed genes, we identified 97 genes that were reported to be cancer related in the literatures (Table 5).

Gene	P-value	Up-down	Cancer Gene	TSG	Cancer Signaling
ICAM5	2.52E-08	Down	CAN-gene		
TCF12	2.84E-08	Up	Cancer Gene		regulation
THRB	3.2E-08	Down		oncogene	
ABR	3.2E-08	Down		TSG	
BTG1	3.28E-08	Up	Cancer Gene	TSG	
KIAA0774	3.97E-08	Down	CAN-gene		
TMEM123	3.97E-08	Up	CAN-gene		
IGFBP7	4.57E-08	Up		TSG	
MAP2K4	4.68E-08	Down	Cancer Gene	TSG	Kinase
ELK1	4.68E-08	Down		oncogene	TF
COL4A1	4.91E-08	Up	GBM		
CD93	5.14E-08	Up	CAN-gene		
CDK2AP1	6.21E-08	Up		TSG	
ITGAV	7.49E-08	Up	GBM		
EDIL3	8.81E-08	Down	GBM		
RB1	1.24E-07	Up	Cancer Gene	TSG	TF
GRM1	1.33E-07	Down	CAN-gene		
TP53	1.40E-07	Up	Cancer Gene	TSG	TF
CDH11	1.46E-07	Up	Cancer Gene		

Gene	P-value	Up-down	Cancer Gene	TSG	Cancer Signaling
CDKN2C	1.49E-07	Up	GBM		CDK inhibitors
SNAP25	1.53E-07	Down	GBM		Vesicle
LDOC1	1.67E-07	Down		TSG	
IGFBP2	1.67E-07	Up	GBM		
BCL11B	1.75E-07	Down	Cancer Gene		
FABP3	1.77E-07	Down		TSG	
RARG	1.88E-07	Down	GBM		
PTPN13	2.3E-07	Up		TSG	cytoskeleton
GABRB2	2.40E-07	Down	GBM		
EIF4A2	2.46E-07	Down	Cancer Gene		
SP110	2.69E-07	Up	CAN-gene		
GAS1	2.75E-07	Up		TSG	
RIMS2	3.59E-07	Down	CAN-gene		
VSNL1	3.83E-07	Down	GBM		Adapter
CDH12	4.01E-07	Down	Cancer Gene		
RAP1GDS1	4.38E-07	Down	CAN-gene		
LMO7	4.57E-07	Down	CAN-gene		
ATM	4.99E-07	Up	Cancer Gene	TSG	Kinase
PDE4DIP	5.21E-07	Down	Cancer Gene		
HIP1	5.69E-07	Up	Cancer Gene		cytoskeleton
PRPF4B	5.94E-07	Up	CAN-gene		
NUP214	6.34E-07	Up	Cancer Gene		Transporter
HSP90AB1	6.77E-07	Down	Cancer Gene		
LYN	7.07E-07	Up		oncogene	Kinase
COL5A2	7.54E-07	Up	GBM		
BCL10	8.04E-07	Up	Cancer Gene		apoptosis
NCKIPSD	8.04E-07	Down	Cancer Gene		
TRIM24	8.40E-07	Up	Cancer Gene		
CLTC	8.95E-07	Down	Cancer Gene		SP
RAF1	9.55E-07	Up		oncogene	Kinase
SYK	1.02E-06	Up	Cancer Gene		Kinase
PIM1	1.03E-06	Up	Cancer Gene	oncogene	
SYT1	1.04E-06	Down	CAN-gene		Vesicle
MAPK8IP2	1.08E-06	Down	CAN-gene		
CHI3L2	1.13E-06	Up	GBM		
YES1	1.21E-06	Up		oncogene	Kinase
ATP8A2	1.31E-06	Down		TSG	Molecule
GABRA1	1.31E-06	Down	GBM		Receptor
RASSF1	1.34E-06	Up		TSG	
CHIC2	1.43E-06	Up	CAN-gene		
MYT1L	1.55E-06	Down	GBM		
COL3A1	1.55E-06	Up	GBM		
NBL1	1.62E-06	Down		TSG	
NONO	1.80E-06	Up	Cancer Gene		
TPR	1.80E-06	Up	Cancer Gene		

Gene	P-value	Up-down	Cancer Gene	TSG	Cancer Signaling
NOV	1.87E-06	Down		oncogene	GF
CDK4	1.87E-06	Up	GBM	oncogene	Kinase
TCF3	1.95E-06	Up	Cancer Gene		TF
NEFM	1.95E-06	Down	GBM		Cytoskeleton
SYN2	2.08E-06	Down	GBM		
FLT3	2.17E-06	Down	Cancer Gene		Generics
CCK	2.17E-06	Down	GBM		
IDH1	2.26E-06	Up	GBM		
CDC23	2.31E-06	Up		TSG	
CHI3L1	2.31E-06	Up	GBM		
BCL6	2.35E-06	Up	Cancer Gene		
RTN1	2.40E-06	Down	GBM		
DLEU2	2.43E-06	Up		TSG	
SH3GL2	2.50E-06	Down	GBM		Vesicle
BLM	2.56E-06	Up	Cancer Gene		
SMAD4	2.56E-06	Up	Cancer Gene		TF
CACNA2D2	2.95E-06	Down		TSG	
PRRX1	2.95E-06	Up	Cancer Gene		
SV2B	2.95E-06	Down	GBM		
NEFL	3.01E-06	Down	GBM		
MCF2	3.14E-06	Down		oncogene	Cytoskeleton
CTNNA1	3.14E-06	Up		TSG	Cytoskeleton
BMPR1A	3.27E-06	Up	Cancer Gene		Receptor
MSN	3.40E-06	Up	Cancer Gene		
RPN1	3.54E-06	Up	Cancer Gene		
LPP	3.69E-06	Up	Cancer Gene		
GABRG2	3.77E-06	Down	GBM		
TRIM8	3.77E-06	Down		TSG	
COX6C	3.84E-06	Down	Cancer Gene		
FBXW7	4.00E-06	Down	Cancer Gene		Generics
ETS2	4.34E-06	Down		oncogene	TF
EWSR1	4.8E-06	Up		oncogene	
ST5	4.99E-06	Up		TSG	

*TSG: tumor suppressor gene; TF: transcription factor; SP: Structural Protein, GF: Growth Factor

Table 5. A list of 97 genes was differentially expressed.

3.2.2 Gene co-expression network analysis

To investigate the functions of the genes at the system-level and uncover the mechanism of GBM, we used partial correlation method and sparse regression techniques to infer gene co-expression networks. The largest connected coexpression network with the average shortest path 20.4 and diameter 74 had 2,115 genes and 2,276 edges (Figure 1).

To explore the relationship between co-expressions of the genes and physical interaction between their corresponding proteins, we compared the constructed co-expression gene network with the protein-protein interaction (PPI) network (Chuang, et al., 2007),

comprising 57, 235 interactions among 11, 203 proteins. The human PPI network was integrated from yeast two-hybrid experiments, predicted interactions based on orthology and co-citation, and curation of literature. We restricted our analysis to 3,991 genes present in both co-expression and PPI networks, consisting of 2, 131 and 12, 182 edges, respectively. The two networks shared 85 edges involving 133 genes in common. The 133 genes and their connections were shown in Figure 2. We performed functional enrichment analysis of the 133 genes using DAVID(Huang da, et al., 2009). Enriched gene functional categories include protein biosynthesis/ribosome, development processes, metabolic processes and RNA splicing. It is noticeable that the largest connected component, which consists of 29 genes, in Figure 2 corresponds to a ribosomal protein complex.
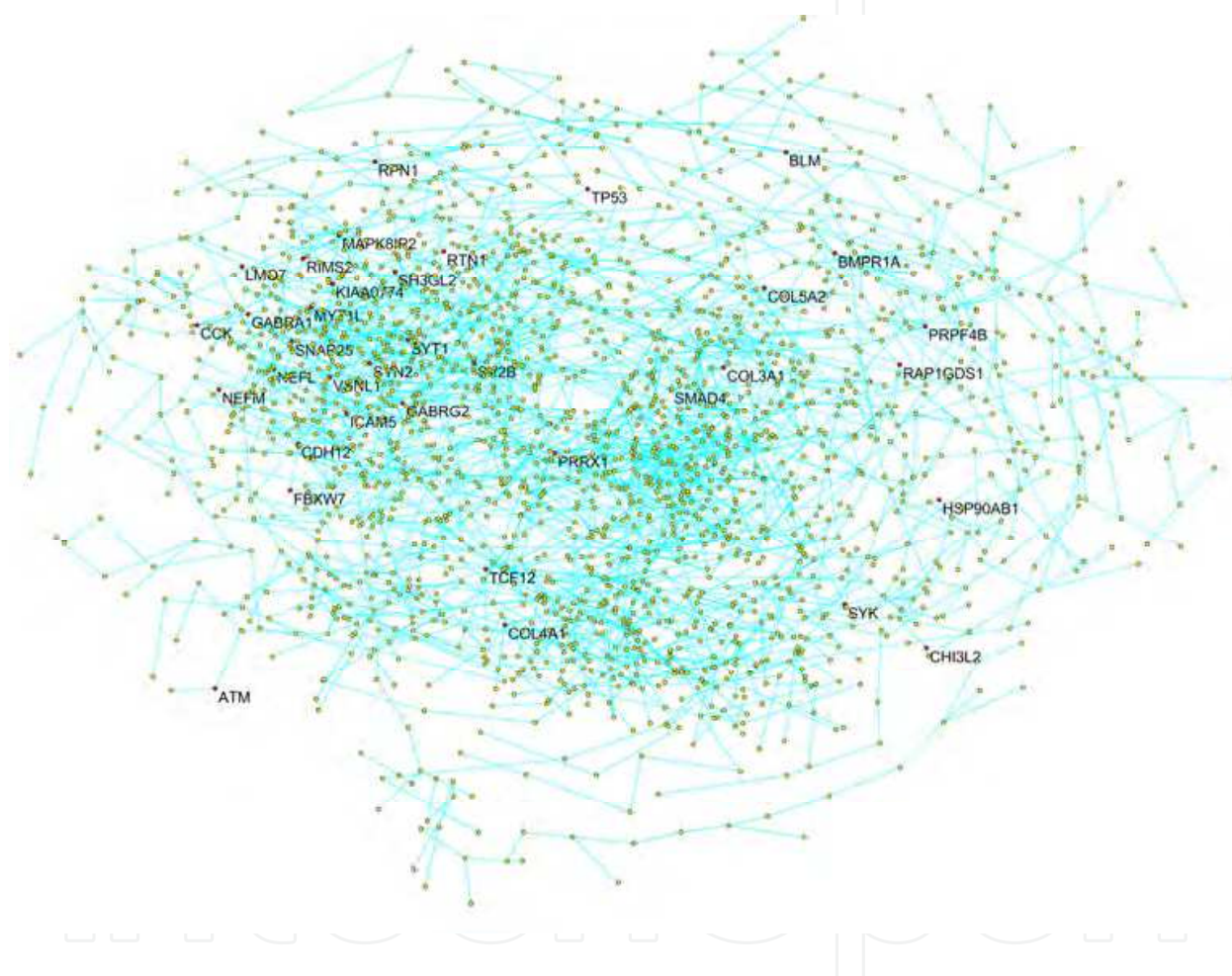
Fig. 1. The largest connected gene co-expression network. The network had 2,115 genes and 2,276 edges. Genes related to GBM were highlighted in red.

The network had 133 genes and 700 edges. This figure gives the sub-network of gene co-expressions in which a node contains both gene co-expressions and protein-protein interactions with other nodes in this sub-network.

Co-expression networks are usually organized into modules that perform specific biological process. 13 modules were significantly enriched for at least one pathway(Figure 3 and Table 6), indicating that co-expression network was organized into functional units. Enriched pathways in the modules were involved in neurodegenerative diseases, development

processes, cancer related signaling pathways and metabolism. Parkinson's disease pathway was enriched in module 11 with olive green color. Genes UQCRC1, ATP5H, COX7C, NDUFA1, COX5R and COX4I1 in the module 11 are involved in mitochondria dysfunction (http://www.genome.jp/kegg/pathway/hsa/hsa05012.html). Cell cycle pathway was enriched in module 7 with yellow color. We can see that co-expressed genes CHEK2, CCNB2, CCNA2, CCNE2, MAD2L1 and BUB1 in module 7 are directly or indirectly connected in cell cycle pathway (http://www.genome.jp/dbget-bin /show_ pathway? hsa04110).
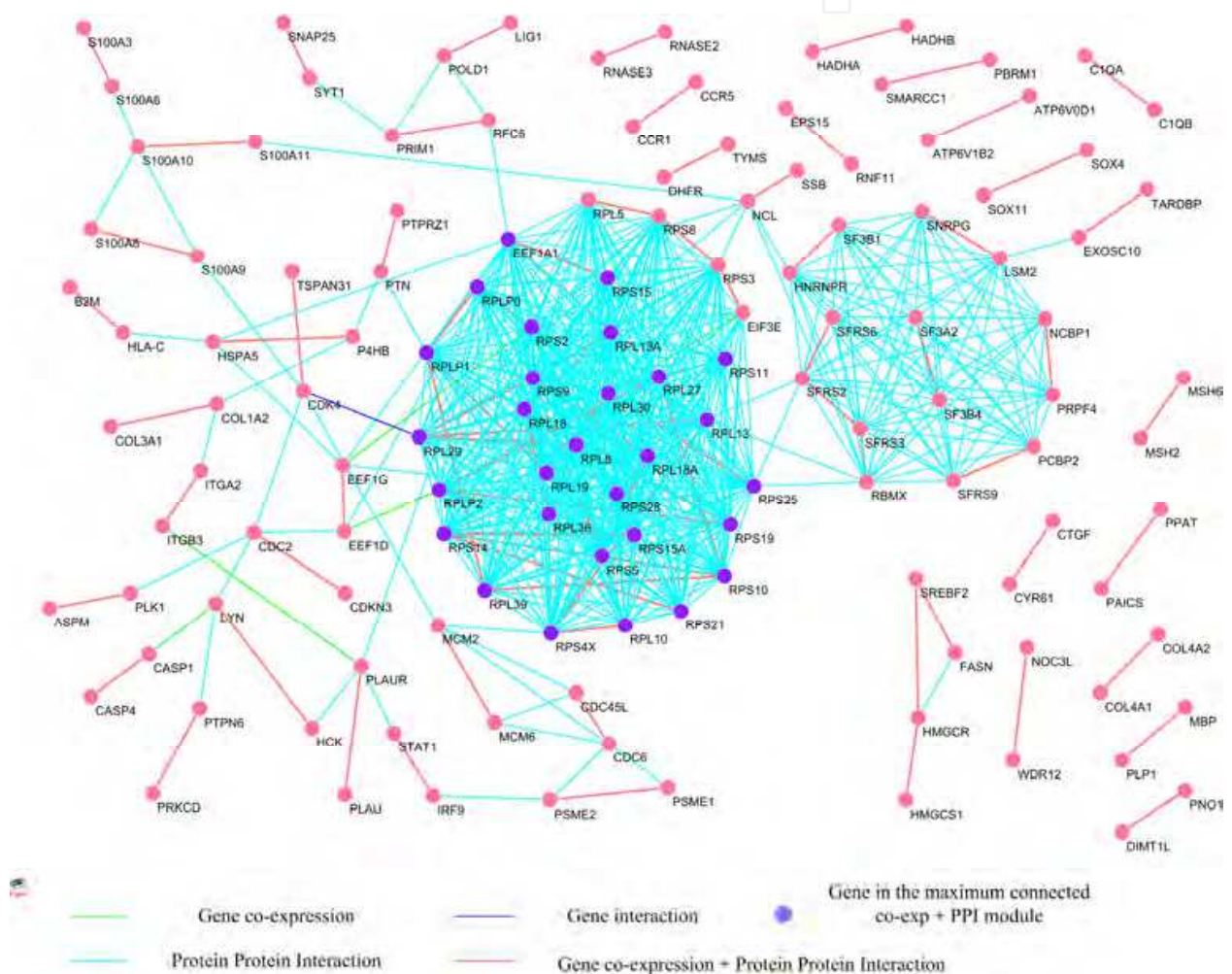
Fig. 2. Gene co-expressions and interactions supported by protein-protein interactions.

The architecture of a co-expression network is important for uncovering the genes which are involved in cancer. To identify the most important genes in the co-expression network, we used the damage value of a node as a measure to rank the importance of a node. We ranked all differentially expressed genes in the largest co-expression sub-network according to their

damage values. Top 5% genes in the ranked list with the damage values greater than 15 were summarized. These genes were essential to the function of co-expression network. We suspect that these genes may be involved in the development of GBM. For example, T1A1 that had the largest damage value 38 and was over expressed in the GBM tissues plays a role in apoptosis (Forch and Valcarcel, 2001). KIAA1279 was reported to be associated with the nervous systems (Brooks, et al., 2005). It was also reported that CACYBP participates in p53-induced beta-catenin degradation and can suppress proliferation and tumorigenesis of renal cancer cells (Sun, et al., 2007). CACYBP was under expressed in gastric cancer (Ning, et al., 2007) and renal cancer (Sun, et al., 2007).
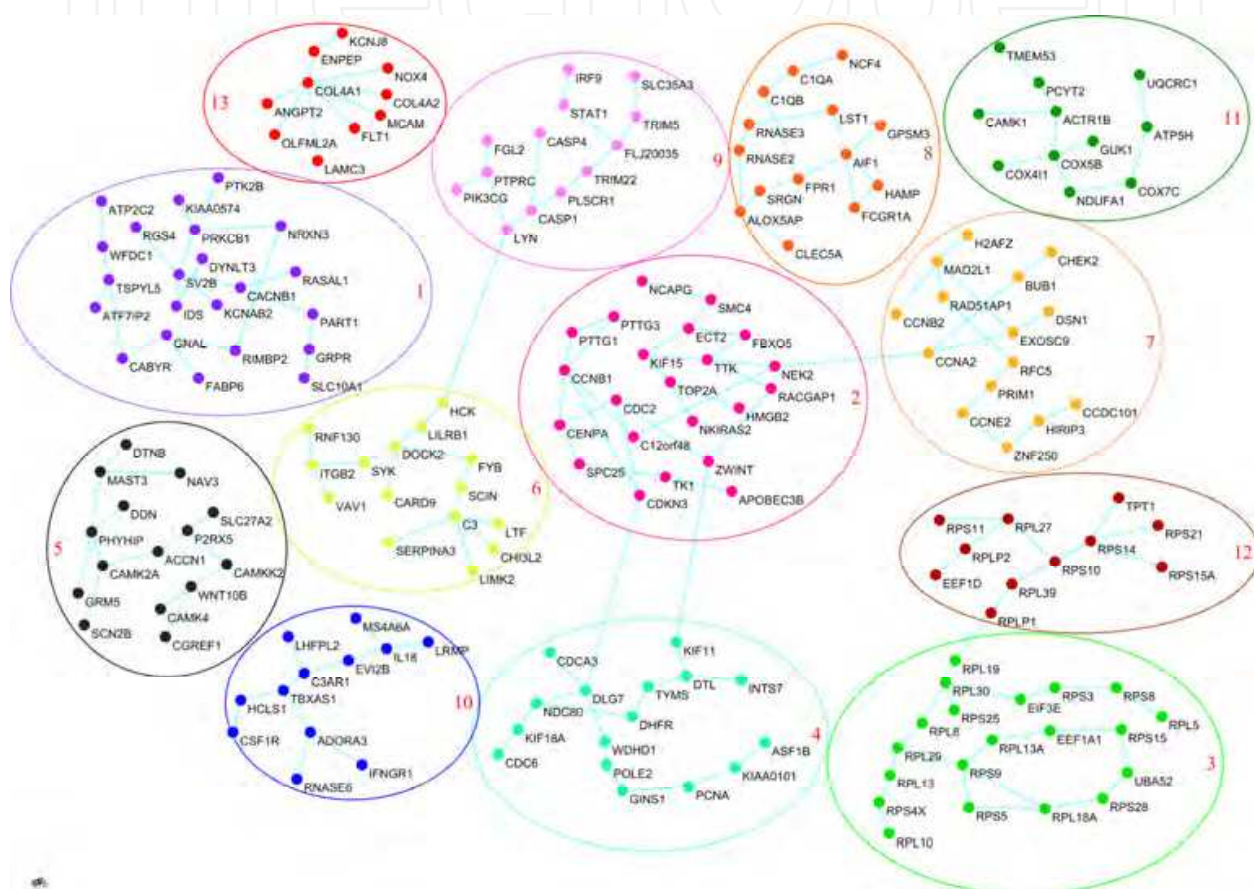
Fig. 3. 13 functional modules in the largest connected gene-coexpression network in Figure 1. The modules were indexed by the number and also represented by color. Enriched pathways in the modules were listed in Table 6.

There were three genes: TP53, COL3A1, and RAP1GDS1 whose damage values ranked the top among the cancer and cancer candidate genes (Table 5). Their removal from the network may disconnect some components in the network and hence compromise the functions of the genes in coexpression networks. TP53 and COL3A1 are GBM related genes and RAP1GDS1 is a cancer candidate gene. The p53 pathway is central to oncogenesis (El Hallani, et al., 2009). TP53, which was up regulated in GBM tissue samples, particularly, was highly over-expressed in CEREBRUM tissue samples. It was observed that TP53 had damage value 23, but was only connected with two up regulated genes: TGIF2 and EIF4A1. TGIF2 was differentially expressed (P-value< 1.43×10^{-7}) and had damage value 24. TGIF2 is a novel TALE superclass homeodomain protein and a transcriptional co-repressor that

interacts with Smad2 and Smad3 to negatively regulate the TGFb/Smad response in a cell, which in turn affects apoptosis and leads to cellular proliferation and differentiation. TGIF2 plays an oncogenic role through inhibition of TGFb (Imoto, et al., 2000). EIF4A1 was over expressed in tumor tissue (P-value$< 5.09 \times 10^{-6}$) and had damage value 22. EIF4A1 is ATP-dependent RNA helicase and required for mRNA binding to ribosome. EIF4A1 inhibit translation initiation and act as a tumor suppressor by forming complex with Programmed Cell Death 4 (PDCD4) (Yang, et al., 2004). COL3A1 was over expressed in GBM tissue samples (P-value$< 1.6 \times 10^{-6}$) and had damage value 21. It was also reported to be over

#	color	Category	Name	P-value
1	Purple	Cell Signaling	Ion Channel and Phorbal Esters Signaling	8.60E-03
2	Wind-Stawberry	Apoptosis	Apoptotic DNA fragmentation and tissue homeostasis	3.52E-02
		Cell Cycle Regulation	Activation of Src by Protein-tyrosine phosphatase alpha	4.23E-02
		Cell Cycle Regulation	AKAP95 role in mitosis and chromosome dynamics	4.98E-02
		Cell Cycle Regulation	Sonic Hedgehog (SHH) Receptor Ptc1 Regulates cell cycle	4.23E-02
3	Forest-Green	Translation	Ribosome	3.32E-09
4	TealBlue	Metabolism of Cofactors& Vitamins	One carbon pool by folate	4.72E-02
5	Black	Cell Signaling	Ca++/ Calmodulin-dependent Protein Kinase Activation	3.59E-03
6	Green-Yellow	Immunology	Cells and Molecules involved in local acute inflammatory response	4.68E-02
7	Yellow	Cell Growth and Death	Cell cycle	2.62E-02
8	Orange	Immunology	Classical Complement	2.85E-02
9	Lavender	Cell Signaling	IFN alpha signaling	1.03E-02
10	Blue	Cell Activation	Th1/Th2 Differentiation	3.79E-02
11	OliveGreen	Metabolism	Electron Transport Reaction in Mitochondria	1.31E-02
		Energy Metabolism	Oxidative phosphorylation	9.00E-03
		Neurodegenerative Diseases	Parkinson's disease	8.67E-03
12	Maroon	Translation	Ribosome	1.13E-03

#	color	Category	Name	P-value
13	Red	Metabolism|Neuroscience	Vitamin C in the Brain	9.11E-03
		Cell Activation	Angiotensin-converting enzyme 2 regulates heart function	1.27E-02
		Metabolism	Intrinsic Prothrombin Activation	3.83E-02
		Metabolism	Platelet Amyloid Precursor Protein	1.47E-02
		Hematopoiesis	Regulators of Bone Mineralization	2.58E-03

Table 6. 13 Modules in gene co-expression network in Figure 3 with their enriched pathways.

expressed in ovarian cancer and breast cancer (Helleman, et al., 2006; Turashvili, et al., 2007). COL3A1 encodes a fibrillar collagen which is a major component of the extracellular matrix protein surrounding cancer cells. Presence of ECM protein prevents apoptosis of cancer cells. COL3A1 plays an important role in apoptosis, proliferation regulation and anticancer drug resistance (Sethi, et al., 1999). RAP1GDS1 was under expressed in GBM tissue samples (P-value < 4.4×10^{-7}) and had damage value 18 in the co-expression network. RAP1GDS1 is a transcription factor (Cimino, et al., 2001). It was reported that translocation fusion of the *NUP98* and *RAP1GDS1* genes was recurrent in T-cell acute lymphocytic leukemia (Hussey, et al., 1999).

3.2.3 Gene interaction network analysis

To further uncover the mechanism of the GBM, the proposed MI-based statistic was applied to expressions of 12,042 genes which were measured in 243 tumor tissue samples and 10 normal tissue samples and 1 cell line to detect pair-wise mRNA interaction between the genes. We detected 967 interaction with theoretic P-values< 1.0×10^{-15} accounting for 0.1% proportion of total interactions. Pair-wise mRNA interactions were assembled into interaction networks defined by representing genes as nodes, with an edge between two nodes if an interaction was observed between the corresponding genes. The resulting largest connected mRNA interaction network consists of 187 genes and 196 edges (Figure 4) . Its average shortest path length was 5.88 and diameter was 23.

Top 10 genes with the largest damage value were SIVA1, SLC27A5, PFDN5, GBAS, COX6A1, AKR1B1, NPM3, UBE2W, CDK4 and SGCG. The gene with both the largest degree (32) and damage value (62) in the mRNA interaction network was SIVA1, an apoptosis-inducing factor. SIVA1 was over expressed in the GBM tissue samples (P-value < 0.00061). Although it was not significantly differentially expressed, regulatory relationship between SIVA1 and each one of 32 genes was significantly different between the GBM tumor tissue and normal tissue samples. This indicated that SIVA1 can act synergistically to influence the development of the GBM. SIVA1 is a pro-apoptotic protein. Function of SIVA1 is to promote DNA damage induced apoptosis and inhibit NF-κB (Barkinge, et al., 2009). SIVA1 plays an important role in cell proliferation and death (Spinicelli, et al., 2002). It was reported that SIVA1 was involved in breast cancer (Chu, et al., 2005). GBAS named as Glioblastoma-amplified sequence, was located in Chromosomal region 7p12, which contains the EGFR gene, and was reported to be co-amplified with EGFR in glioblastomas (Wang, et al., 1998). It was over expressed with p-value 1.01×10^{-5} in our expression data.CDK4 with degree (8)

and damage value (13) was up regulated in the GBM tumor tissue samples and significantly differentially expressed (P-value < 1.87×10^{-8}). CDK4 was a member of the Ser/Thr protein kinase family and involved in Glioma pathway(Zhu and Parada, 2002). CDK4 was an oncogene and was reported to be associated with multiple tumors(Kim and Diehl, 2009).

3.3 Network analysis of miRNA expressions
3.3.1 Differential expression analysis

To unravel the pattern of differential regulation of miRNA, expressions of 534 microRNAs including 470 human microRNAs were profiled in 240 tumor tissue samples and 10 normal tissue samples by Agilent 8 x 15K Human microRNA-specific microarray. A total of 149 miRNAs were differentially expressed between the GBM tumor tissue and normal tissue samples which were identified by Wilcoxon rank-sum test (P-value for declaring significance after Bonferroni correction is 9.36×10^{-5}). Of 149 differentially expressed miRNAs, 73 miRNAs were up-regulated and 76 down-regulated. Among them, 21, 81 and 15 miRNAs were reported to be associated with the GBM, other cancers and other diseases, respectively, in the literatures(Jiang, et al., 2009) .
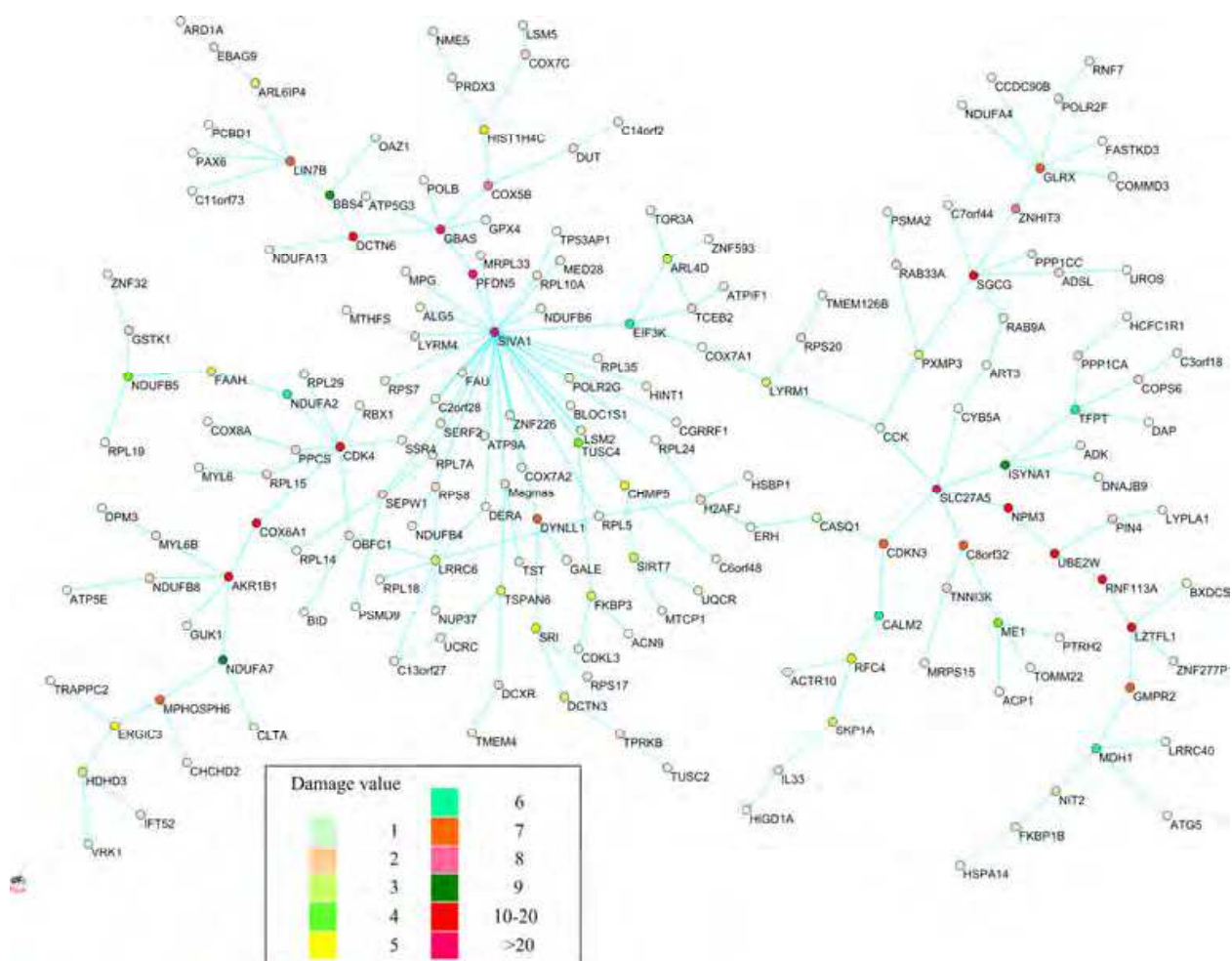

Fig. 4. Gene interaction network. The largest connected gene interaction network consists of 187 genes and 196 edges. Genes with different damage value were marked in different colors.

3.3.2 MiRNA co-expression network analysis

Similar to genes, miRNAs are not isolated, instead they act together to perform biological functions. Initiation and progression of cancer are influenced not only by individual miRNAs, but also by the coordinated effect of many miRNAs. To understand how miRNAs regulate biological processes at a system level, we reconstruct miRNA co-expression network. The largest connected miRNA coexpression network with the average shortest path 10.75 and diameter 49 had 385 miRNAs and 451 edges (Figure 5). One main feature of this miRNA coexpression network is that mir-770-5p and mir-329 divided the whole network into three subnetworks: the left, the middle and the right subnetwork. The middle subnetwork was densely connected.
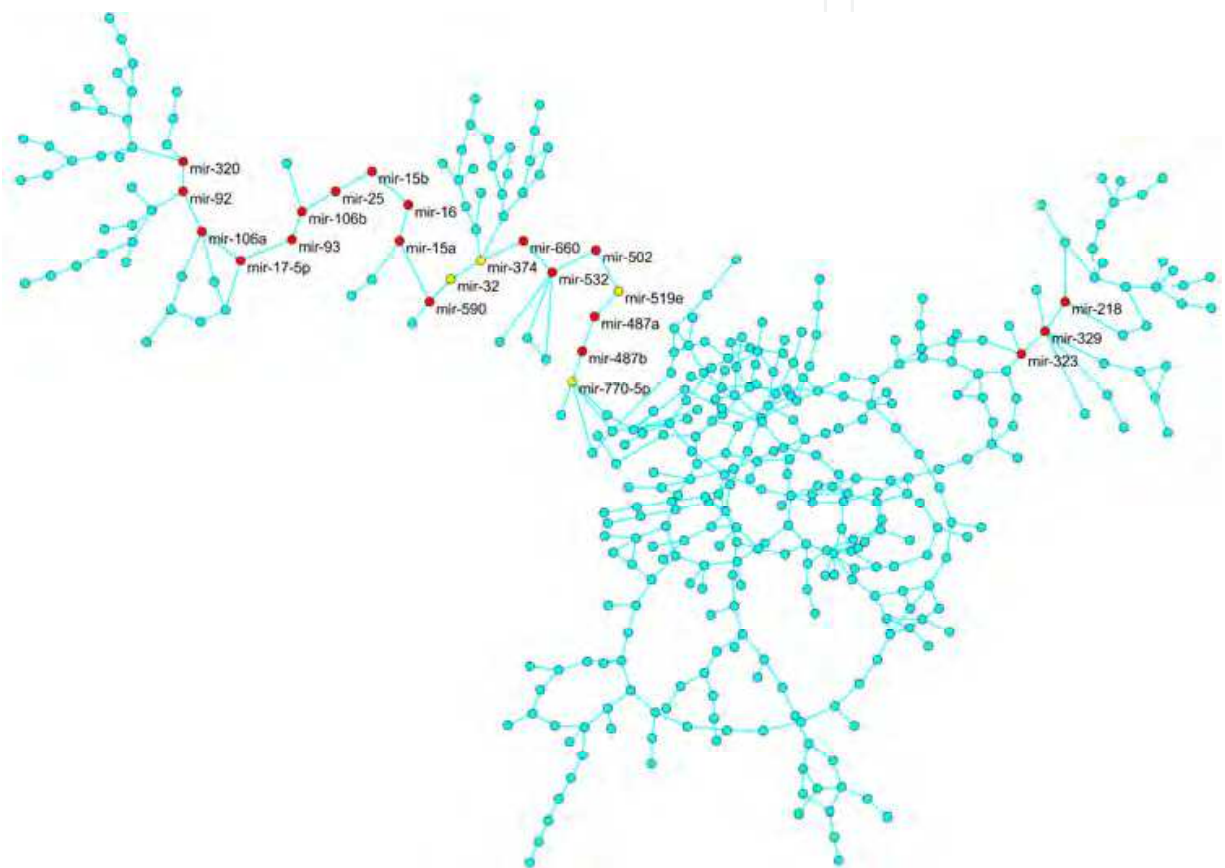

Fig. 5. MiRNAs coexpression network. MiRNAs with damage value larger than 20 were highlighted in red and yellow, where red nodes denoted significantly differentially expressed miRNAs (p-value< 9.36×10^{-5}) and yellow nodes denoted they were not significant.

Similar to gene co-expression network, we used the damage value of a node as a robustness measure to rank the importance of a miRNA in miRNA co-expression network. We ranked all differentially expressed miRNAs in the largest miRNA co-expression network according to their damage values. There were 19 differentially expressed miRNAs with the largest damage values (>20) which we referred to as fragile miRNAs. Among the 19 fragile miRNAs, 16 miRNAs (14 miRNAs were over expressed in the GBM tissues) were in the left coexpression subnetwork, and 3 under expressed miRNAs were in the right coexpression subnetwork. The middle coexpression subnetwork contains no fragile miRNAs. Figure 5 showed that miRNAs in the middle coexpression subnetwork were densely connected. The

middle coexpression network with large number of redundant miRNAs was highly robust in response to perturbation of external forces. From Figure 5, we can see that mir-487a, mir-487b, mir-502 and mir-532 were the most important components in the miRNA co-expression network. Removal of one of them would cause disconnection between the left part and right part of miRNA co-expression network and hence lead to dysfunction of the whole miRNA co-expression network. We observed that mir-487a and mi-487b were down-regulated, and mir-502 and mir-532 were up-regulated in the GBM tissues. We searched miR2disease database (Jiang, et al., 2009) and found that 7 of the 19 miRNAs were associated to GBM/glioma/ neuroblastoma (mir-15b, mir-25, mir-93, mir-17-5p, mir-323, mir-106a, mir-92), 6 others were associated with cancer and neuro-diseases(mir-487b, mir-15a, mir-16, mir-106b, mir-218, mir-320). Take

3.3.3 MiRNA interaction network analysis

We also used MI to quantify interaction between miRNAs and MI-based statistic to detect pair-wise interaction between miRNAs. We assembled the interacted miRNAs into miRNA interaction network. Significant P-values for declaring an interaction between miRNAs were 8.51×10^{-7} accounting for 1% proportion of total interactions. An assembled largest connected miRNA interaction network was shown in Figure 6. The network included 333 miRNAs and 563 interactions. Its average shortest path length was 5.10281 and diameter was 17.
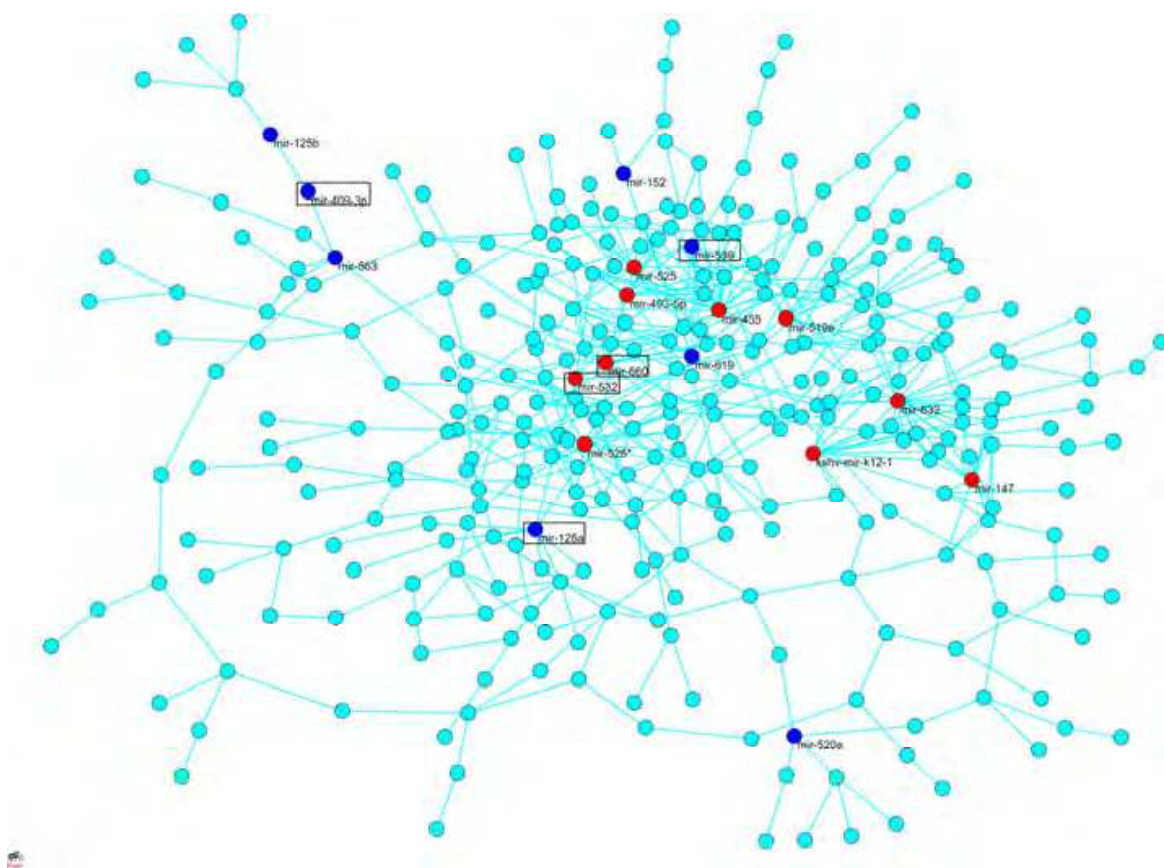

Fig. 6. MiRNA interaction network. miRNAs with damage value larger than 4 were marked in blue and miRNAs with degree larger than 9 were marked in red. Significantly differentially expressed miRNAs (p-value< 9.36×10^{-5}) were marked in black box.

To examine topological network properties of miRNA interaction networks, we also used the damage value and the degree of a node to identify cancer relevant miRNAs. Among all the miRNAs with either damage values greater than 4 or degrees greater than 9, miR-539 was associated with autism spectrum disorder (ASD), a brain development disorder(Abu-Elneel, et al., 2008). It was reported that mir-563 is involved in modulating tumors in variable microenvironments (Hebert, et al., 2007), miR-125a and miR-125b were indentified to control human neuroblastoma cell proliferation by repressing a common target gene NTRK3 (neurotrophic tyrosine kinase, receptor, type 3)(Laneve, et al., 2007). They were found to be down-modulated in primary neuroblastoma tumors and were under expressed with p-value 4.83×10^{-5} and 0.03 respectively in our dataset.

3.4 miRNA target networks

We compiled miRNA and mRNA expression data in 237 tumor tissue samples and 10 normal tissue samples for 1,697 differentially expressed mRNAs and 149 differentially expressed miRNAs. This resulted in extremely complex miRNA target networks. We found 3,953 matched miRNA-mRNA pairs for 127 differentially expressed miRNAs and 1, 089 differentially expressed genes. Of the 3,953 target-pairs, 65 down-regulated miRNAs targets 468 over expressed genes while 62 up-regulated miRNAs target 621 under expressed genes.

A total of 14 experimentally verified targets of 7 miRNAs were listed in Table 7. Of 8 miRNAs, 4 under expressed miRNAs (mir-124a, mir-29b, mir-29c and mir-33) function as tumor-suppressors and 4 over expressed miRNAs (mir-155, mir-16, mir-21 and mir-210) function as oncogenes. It has been reported that CTDSP1 was a validated target gene of mir-124a (Karginov, et al., 2007), RTN4 and SLC25A22 was validated targets of mir-16 (Selbach, et al., 2008). mir-21 was found to be over expressed in multiple cancers and down regulated tumor suppressor genes: TPM1, PTEN, PDCD4 BASP1 and RTN4 in invasion and metastasis of cancer(Yang, et al., 2009). BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1, thus it can regulate WT1 transcriptional activity (Carpenter, et al., 2004). Nogo-A , one protein isoforms encoded by RTN4, had turned out to be a neuronal protein involved in diverse processes that go from axonal fasciculation to apoptosis(Mingorance, et al., 2004). Mir-155 targeted a regulator of apoptosis gene: LDOC1 (Skalsky, et al., 2007). Up-regulation of miR-210 directly targeted gene EFNA3, which is crucial for endothelial cell response to hypoxia, affecting cell survival, migration, and differentiation (Fasanaro, et al., 2008). Mir-29c was reported to be under expressed in Nasopharyngeal Carcinomas and up-regulated genes COL4A1, COL4A2 and TDG (Sengupta, et al., 2008). COL4A1, COL4A2 were genes encoding extracellular matrix proteins, as we discussed previously, they played an important role in apoptosis, proliferation regulation and anticancer drug resistance (Sethi, et al., 1999) . COL4A2 was validated to be also targets of mir-29b in another research group (Li, et al., 2009). TDG was involved in DNA repair, a process frequently dysregulated in many cancers (Sengupta, et al., 2008). In mouse and human cells, miR-33 inhibits the expression of ABCA1, thereby attenuating cholesterol efflux to apolipoprotein A1 (Rayner, et al., 2010). It was also reported that the role of miR-33 controlling the hematopoietic stem cells self-renewal through p53 may lead to the prevention and treatment of hematopoietic disorders (Herrera-Merchan, et al., 2010).

The resulting miRNA target networks have several remarkable features. First, many important genes in the mRNA co-expression networks were targets of differentially expressed miRNAs. Top 17 genes with damage values greater than 19 in the gene co-

expression network were negatively regulated by 34 differentially expressed miRNAs. All 17 genes and many miRNAs were crucial components in the gene and miRNA

miRNA	P-value for differential expression	Up or down regulated	Gene	P-value for differential expression	Up or down regulated	Regression coefficient	P-value for regression	Coefficient	Refer
mir-124a	8.62E-08	Down	CTDSP1	8.4E-07	Up	-17.7	2.22E-16	0.244	Karginov, et al., 2007
mir-155	1.31E-07	Up	*LDOC1	1.67E-07	Down	-197	4.00E-15	0.223	Skalsky, et al., 2007
mir-155	1.31E-07	Up	SCAMP1	3.27E-06	Down	-30.9	6.06E-10	0.145	Selbach, et al., 2008
mir-16	0.0000178	Up	*RTN4	9.44E-08	Down	-766	2.66E-08	0.119	Selbach, et al., 2008
mir-16	0.0000178	Up	SLC25A22	3.36E-08	Down	-37.8	1.97E-11	0.168	Selbach, et al., 2008
mir-16	0.0000178	Up	VTI1B	7.15E-08	Down	-122	1.03E-05	0.0765	Selbach, et al., 2008
mir-21	9.98E-08	Up	BASP1	2.98E-08	Down	-743	0.00E+00	0.349	Yang, et al., 2009
mir-21	9.98E-08	Up	RTN4	9.44E-08	Down	-246	1.16E-05	0.0756	Yang, et al., 2009
mir-210	1.48E-06	Up	EFNA3	4.67E-07	Down	-10.4	4.61E-05	0.0657	Fasanaro, et al., 2008
mir-29b	2.07E-07	Down	COL4A2	1.83E-07	Up	-164	8.92E-05	0.0609	Li, et al., 2009
mir-29c	3.66E-07	Down	COL4A2	1.83E-07	Up	-260	1.99E-07	0.105	Sengupta, et al., 2008
mir-29c	3.66E-07	Down	COL4A1	4.91E-08	Up	-389	1.35E-08	0.124	Sengupta, et al., 2008
mir-29c	3.66E-07	Down	TDG	5.21E-07	Up	-95.5	3.22E-14	0.21	Sengupta, et al., 2008
mir-33	9.6317E-06	Down	ABCA1	1.0396E-06	Up	-114.5	5.941E-12	0.176	Rayner, et al., 2010

* LDOC1 and RTN4 are tumor suppressor genes

Table 7. Experimentally verified targets.

coexpression networks. Their altered expressions played an important role in tumorigenesis. Among them, overexpressed RBBP4 (P-value 1.8×10^{-8}) with the largest damage value (37) in the table was negatively regulated by under expressed mir-29b (P-value 2.07×10^{-7} and mir-29c (P-value 3.66×10^{-7}), underexpressed TRIM8 (P-value 3.77×10^{-6}) with the damage value 29 was negatively regulated by overexpressed mir-629 (P-value 2.02×10^{-7}), overexpressed TP53 ((P-value 1.40×10^{-7}) with damage value 23 was negatively regulated by underexpressed mir-485-5p. RBBP4 has been implicated in chromatin remodeling and regulation of cell proliferation. It has been reported that RBBP4 is overexpressed in different human tumors, such as lung, liver and thyroid cancer, acute myelocytic leukemia, and acute lymphoblastic leukemia (Pacifico, et al., 2007). TRIM8 is thought of as a new tumor suppressor gene (Caligo, et al., 1997).

Second, many genes targeted by critical components in miRNA co-expression network, played important roles in regulation of neural process and tumor genesis. To study the function of the top 19 differentially expressed miRNAs with the largest damage values, we constructed mRNA target network of these 19 miRNAs with 476 nodes and 1,128 arcs and 174 edges shown in Figure 7A, where 5 underexpressed miRNAs negatively regulated 85 over expressed genes and 14 overexpressed miRNA negatively regulated 372 underexpressed genes, 1,128 arcs were miRNA-mRNA pairs, 15 edges connected co-expressed miRNAs, and remaining 159 edges linked co-expressed genes. The genes with large damage value or related to cancer in the gene co expression network were highlighted and those gene were marked in the Figure 7A. 8 genes with the damage values greater than 15 were regulated by 11 miRNAs with the damage values greater than 20. 14 of the15 cancer genes in Figure 7A were under-expressed FBXW7, GABRA1, MYT1L, NEFL, RTN1, SH3GL2, SNAP25, SV2B, SYN2, KIAA0774, VSNL1, NEFM, RIMS2, SYT1. These genes were regulated by 10 overexpressed miRNAs: mir-15b, mir-25, mir-16, mir-92, mir-15a, mir-320, mir-106b, mir-93, mir-106a, mir-17-5p, which were connected in the mRNA co-expression network (Figure 7B). They are involved in synaptic transmission process. Neuron communication occurs at the synapse via neurotransmitters. MYT1L (myelin transcription factor 1-like) regulates nervous system development. Both SYT1 and VSNL1 serve as Ca(2+) sensors in synaptic transmission(Maglott, et al., 2007). SNAP25 is a presynaptic plasma membrane protein and regulates neurotransmitter release. It is reported that SNAP25 is implicated in neuritogenesis in human neuroblastoma(Heraud, et al., 2008). GABRA1 encodes a gamma-aminobutyric acid (GABA) receptor and GABA is the major inhibitory neurotransmitter in the mammalian brain. Both NEFL and NEFM are Neurofilaments. They play a role in intracellular transport to axons and dendrites. NRGN encodes a postsynaptic protein kinase substrate that binds calmodulin in the absence of calcium and is a direct target for thyroid hormone in human brain(Maglott, et al., 2007). FBXW7 serves as a negative regulator of oncoprotein and is a general tumor suppressor. It regulates cell differentiation and proliferation. It has been reported that FBXW7 is implied in various cancers including glioblastoma(Hagedorn, et al., 2007). Remaining genes that were involved in neuron communication include SLC6A15, SYN2, SV2B and RIMS2. SLC6A15 is a neurotransmitter transporter, SYN2 is a member of the synapsin gene family, SV2B is synaptic vesicle glycoprotein and RIMS2 regulates synaptic membrane exocytosis(Maglott, et al., 2007).

Table 8 summarizes the major functions of target genes of miRNAs in Figure 7A where right-railed Fisher's exact test were used to test for significantly enriched GO categories or pathways. Table 8 shows that the target genes of these miRNAs were mainly involved in cancer related signaling pathways and nerves systems process including synapse,

synaptic transmission, neurotransmitter transport, nervous system development and neurological system process. It is interesting to note that co-expressed mir-15a, mir-15b, mir-16, mir-25 and mir-92 were major miRNAs that targeted genes in both cancer related signaling pathways and nerves systems process.

Gene Ontology (GO) category or pathway	Enrichment P-values								
	mir-15b	mir-15a	mir-16	mir-25	mir-106a	mir-93	mir-106b	mir-92	mir-323
Glioma		1.9E-02							
Regulation of neurotransmitter levels		1.3E-02	1.1E-02						
Neurotransmitter transport	6.8E-03	1.5E-02	3.3E-02	3.9E-02					
Long-term potentiation	4.2E-03	3.7E-05	2.4E-03	3.1E-02		1.4E-02	5.3E-03		
Synaptic transmission	2.4E-04	4.0E-04	5.3E-04	3.0E-06		4.4E-02		1.6E-03	
Neurotransmitter secretion		1.7E-02	1.0E-02	2.9E-02					
Nervous system development	3.2E-02	1.3E-02	5.5E-03				2.7E-02		
Central nervous system development			2.5E-02						
Transmission of nerve impulse	2.3E-04	9.6E-04	4.8E-04	1.8E-06				6.0E-04	
Regulation of synaptic plasticity				4.6E-02					
Synaptic vesicle transport		2.8E-02						3.1E-02	
Regulation of synapse structure and activity				4.6E-02					
Neurological system process	7.1E-03	4.3E-03	5.1E-03	1.5E-04				3.1E-03	
Learning and/or memory	3.7E-02								
Learning	4.0E-02								
Neurofilament				2.1E-02					
Synaptic vesicle	1.1E-02		8.0E-03	2.7E-02					
Synaptosome								1.4E-02	
Synaptic vesicle membrane				3.4E-02					
Neuron projection			4.5E-02						

| Function	1	2	3	4	5	6	7	8	9
Synapse part				4.7E-05	1.5E-02	2.5E-03	1.6E-02	1.2E-02	
Synapse	4.9E-03	1.5E-02	2.1E-02	1.1E-07	6.2E-03	8.4E-04	5.3E-03	2.3E-04	
Axon guidance						2.1E-02			
Antigen processing and presentation									4.0E-02
Ras protein signal transduction					1.9E-02		2.3E-02		
Calcium ion transport	1.8E-03	3.6E-04	7.8E-03				4.2E-02		
Calcium signaling pathway	2.7E-04	7.5E-06	1.7E-03	1.7E-02		2.0E-04	2.2E-05		
Cell communication	1.9E-02	2.2E-02	5.6E-03	2.3E-03		1.8E-02	2.6E-02	2.5E-02	
Cell-cell signaling	1.4E-03	8.1E-04	9.5E-04	4.2E-06				7.0E-03	
Wnt signaling pathway		4.0E-02				2.9E-03	7.6E-03		
MAPK signaling pathway		1.5E-02	8.4E-03			3.8E-02	2.6E-02		
GnRH signaling pathway	1.0E-02	8.7E-05	6.6E-03						
ErbB signaling pathway		2.2E-02	3.0E-02						
B cell receptor signaling pathway						6.3E-03			
VEGF signaling pathway						8.4E-03			
T cell receptor signaling pathway						2.1E-02			
Apoptosis						2.1E-02			

Table 8. Function of target genes of 9 miRNAs.
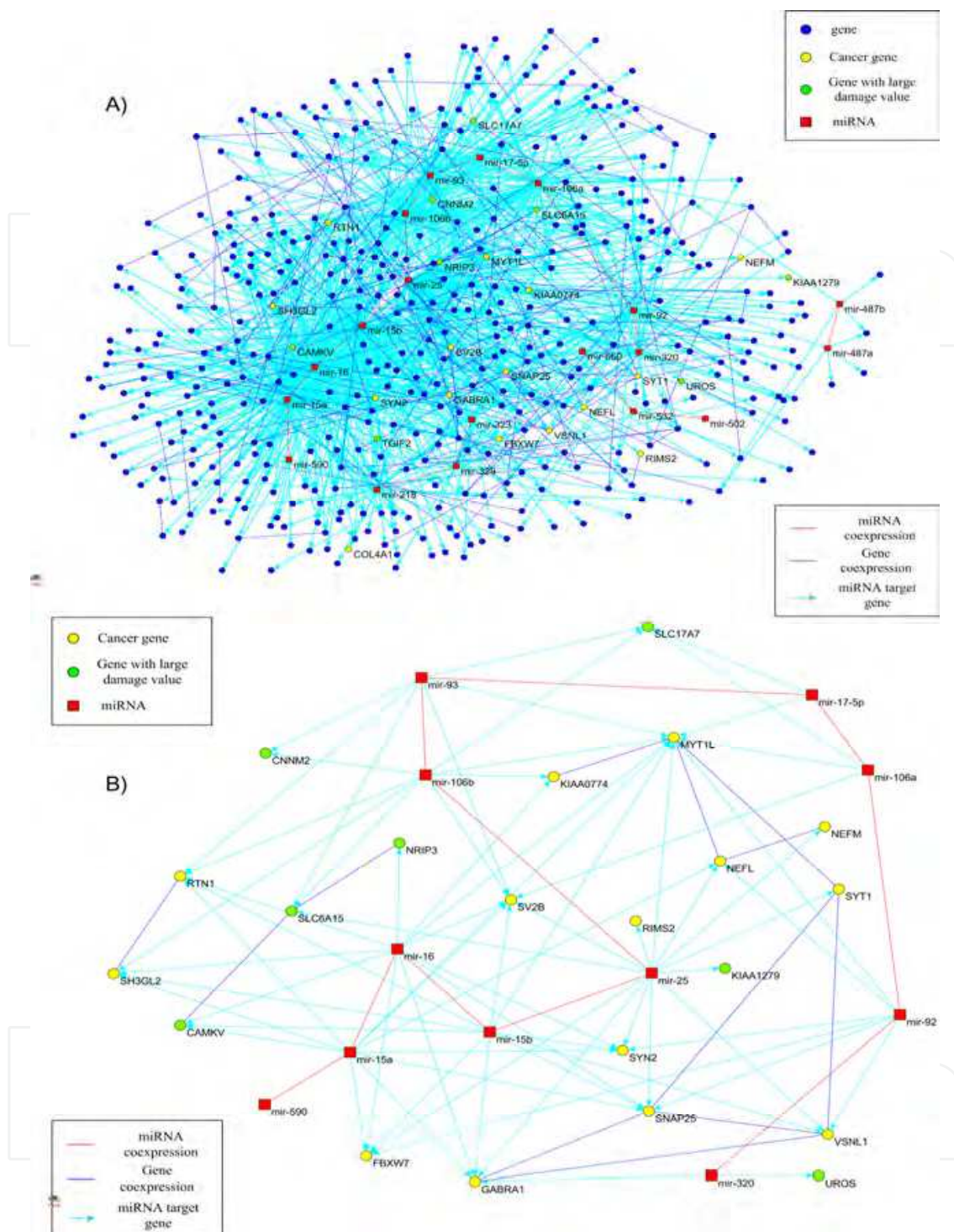
Fig. 7. MiRNA target (sub)network. (A) mRNA target networks of the 19 miRNAs with 476 nodes and 1,128 arcs and 174 edges, where 1128 arcs were miRNA-mRNA pairs, 15 edges connected coexpressed miRNAs, and remaining 159 edges linked coexpressed genes. The genes with large damage value or related to cancer in the co expression network genes were highlighted and those gene names were marked in the figure. (B) MiRNA target subnetwork. 7 genes with the damage values greater than 15 and 14 cancer related genes were regulated by 11 miRNAs with the damage values greater than 20. Those nodes and edges were extracted from Figure 7A mRNA target network.

Third, contribution of miRNAs to expression variation of some genes is very high. The proportion of expression variation of target gene explained by the linear influence of miRNA variation can be measured by the coefficient of determination (R^2). There were a total of 78 under-expressed genes with the coefficient of determination greater than 20% which were negatively regulated by 7 overexpressed miRNAs: mir-15a, mir-15b, mir-16, mir-25, mir-93, mir-106a and mir-106b. Surprisingly, up to 33% of expression variations of underexpressed gene GCC2 were explained by single overexpressed mir-25. Functional roles of GCC2 in cancer are unknown. The gene with the second largest coefficient of determination was CAMK2N1. Its 28% of expression variations were explained by overexpressed mir-106b. It was reported that CAMK2N1 is a candidate tumor suppressor (Wang, et al., 2008). Function of CAMK2N1 is to inhibit MEK/ERK activity and induce p27 accumulation which negatively regulates cell-cycle progression. Reducing expression of CAMK2N1 will accelerate tumor growth. 26% of expression variation of CPEB1 that regulates synaptic plasticity and is implied in cancer development (Hansen, et al., 2009) and GRM1 that is involved in neurotransmitter in the central nervous system and implicated in Melanoma (Namkoong, et al., 2007) was explained by overexpressed mir-25 and mir-15b, respectively. Expressions of genes SYT1, SNAP25, GABRA1, VSNL1, MYT1L, SV2B, SYN2 and SLC12A5 which were involved in synaptic transmission, were also largely regulated by miRNAs.

Fourth, we observed that one miRNA may target multiple genes (up to 179 mRNAs) directly or indirectly, and multiple miRNAs (up to 9 miRNAs) can target a given gene. More than 40% of expression variation of each of six genes (YWHAH, GABRA1, DYNC1I1, ERC2, PEX5L and CAMK2G) was explained by several miRNAs. Expressions of these genes were largely regulated by multiple differentially expressed miRNAs. YWHAH and GABRA1 are implied in neural diseases, and CAMK2G is involved in regulation of plasticity at glutamatergic synapses(Maglott, et al., 2007). Functions of other three genes are unknown.

3.5 DNA methylation analysis

To study how genes are regulated by DNA methylation in GBM, two methylation datasets: OMA-002 dataset and OMA-003 dataset were generated by TCGA project. OMA-002 dataset used the Illumina GoldenGate Methylation Cancer Panel 1 platform, which contains 1,505 CpG loci selected from 813 genes. OMA-002 dataset included 239 GBM tumor tissue samples and 1 cell line, 228 of which were shared among the miRNA expression, gene expression datasets and thus used for Methylation-miRNA regression and Methylation-gene regression analysis. OMA-003 dataset used the Custom Illumina Goldengate Methylation Cancer Panel, which contains 1,489 gene promoters. 245 GBM tumor samples and 1 cell line were used for experimented, 234 of which were shared among the miRNA expression, gene expression datasets and used for Methylation-miRNA regression and Methylation-gene regression analysis.

A total number of 341 genes were hypomethylated genes (average methylation ratio < 0.25 and fold changes <0.5) in two datasets and 113 hypermethylated genes (average methylation ratio > 0.75 and fold changes > 1.5) in glioblastomas. For OMA002 dataset we averaged the methylation ratio for several probes in one gene. To investigate the effects of aberrant

methylation on gene expressions, we regressed gene expression on the methylation profile. We found that 120 hypermethylated genes negatively regulated 404 under expressed genes, and 41 hypomethylated genes negatively regulated 85 overexpressed genes. One methylated gene might influence expressions of up to 142 genes and multiple methylated genes (up to 32) might regulate a given gene. Interestingly, we found that 56.3% of expression variation of underexpressed tumor-suppressor gene BCL11B (P-value 1.75×10^{-7}) is explained by hypermethylated gene HLA-DPB1 with Average methylation 0.9631 and fold change 7.4. We also found that a total of 24 hypermethylated genes that silenced expression of BCL11B (Table 9). These hypermethylated genes include PMP22, a major component of myelin in the peripheral nervous system, GPR116, involved in neuropeptide signaling pathway, LTA, involved in apoptosis. Another gene with the largest coefficient of determination was ANXA5(P-value 3.07×10^{-6}). ANXA5 binds to phosphatidylserine, which is one of the "eat me" signals at the surface of the apoptotic cell, thus involved in apoptosis and cancers (Boersma, et al., 2005). Up to 11 hypormethylated genes were negatively correlated with overexpression of ANXA5 (Table 9).

Methylation also directly or indirectly regulates expression of miRNAs. We observed that 29 underexpressed and 26 overexpressed miRNAs were directly or indirectly regulated by 82 hypermethylated genes. One hypermethylated gene might significantly influence expressions of up to 25 miRNAs and multiple hypermethylated genes (up to 30) might influence expression of one miRNA. We also found that a total of 67 hypomethylated genes directly or indirectly induced expression variation of 38 down regulated and 23 up regulated miRNAs. One hypomethylated gene might influence expression of up to 25 miRNAs and multiple hypomethylated genes (up to 22) might influence expression of one miRNA.

A total of 542 underexpressed and 218 overexpressed genes were directly or indirectly regulated by 61 underexpressed and 62 overexpressed miRNAs, and 300 hypomethylated and 149 hypermethylated genes. In general, multiple miRNAs and methylated genes jointly contribute (in many cases, indirectly) to the variation of gene expressions. In many cases, both miRNAs and methylation repressed gene expression. However, surprisingly we observed that indirect effect of methylation on the regulation of gene expression was to increase gene expression rather than repressing expression. Functions of the many genes regulated by both miRNAs and methylation such as ANXA5, CACYBP, SLC17A7, SLC6A15, INA, SHANK2, PAK3, RIMS2, NBEA, VSNL1, GABRG2, PCLO, VAMP2, PLCB1, GRM7 and SNAP25 were involved in neuron communication and in nervous system. We also found that expressions of the cancer relevant genes TRIM8 (which has been tested for association with Glioblastoma), TP53, FGF13, ENC1, FBXW7, BCL2L2 were influenced by both miRNAs and methylation.

Methylated gene	Average Methylation	Fold Changes	Regu-lated	Regression coefficient	P-value	Coeffi-cient
HLA-	0.963075314	7.408272	BCL11B	-2.01E+02	0.00E+00	5.63E-01
TNF	0.954752441	5.967203	BCL11B	-1.30E+02	2.22E-16	2.59E-01
PMP22	0.912672293	40.56321	BCL11B	-5.31E+01	6.94E-13	2.06E-01
GPR116	0.945111576	7.270089	BCL11B	-7.32E+01	6.84E-11	1.72E-01

Methylated gene	Average Methylation	Fold Changes	Regu-lated	Regression coefficient	P-value	Coeffi-cient
FLJ25084	0.881295918	2.037678	BCL11B	-6.56E+01	5.84E-11	1.69E-01
EGF	0.870425384	2.451902	BCL11B	-3.42E+01	2.84E-09	1.45E-01
FCRLM1	0.819928571	46.85306	BCL11B	-2.57E+01	4.03E-09	1.39E-01
LTA	0.887046722	18.67467	BCL11B	-5.02E+01	9.74E-09	1.36E-01
TRPM5	0.84431311	2.447284	BCL11B	-4.26E+01	8.54E-08	1.19E-01
MPL	0.841936275	7.165415	BCL11B	-3.40E+01	9.62E-08	1.19E-01
BLK	0.954205021	42.40911	BCL11B	-1.01E+02	5.84E-07	1.05E-01
IL4	0.846998588	2.540996	BCL11B	-3.64E+01	8.17E-07	1.03E-01
HLA-DOB	0.886698047	39.4088	BCL11B	-4.82E+01	9.44E-07	1.01E-01
CD1D	0.955290456	3.746237	BCL11B	-6.56E+01	1.09E-06	9.92E-02
CLDN4	0.843246165	16.06183	BCL11B	-3.49E+01	1.29E-06	9.87E-02
CXCL9	0.76874477	2.520475	BCL11B	-2.14E+01	1.36E-06	9.83E-02
KISS1	0.925806122	4.571882	BCL11B	-5.68E+01	1.59E-06	9.47E-02
APOC2	0.88015625	1.614966	BCL11B	-4.65E+01	1.33E-05	9.29E-02
TNFRSF10	0.929268207	2.323171	BCL11B	-4.22E+01	4.64E-06	8.92E-02
C18orf2	0.755978723	4.142349	BCL11B	-1.77E+01	1.24E-05	8.26E-02
PLA2G2A	0.785191771	31.40767	BCL11B	-2.25E+01	1.11E-05	8.21E-02
EPHX1	0.900292887	1.809634	BCL11B	-5.92E+01	1.16E-05	8.17E-02
PLA2G2A	0.809780335	2.816627	BCL11B	-2.74E+01	2.04E-05	7.74E-02
KLK11	0.873908647	2.39427	BCL11B	-3.44E+01	2.33E-05	7.63E-02
MMP14	0.214993026	0.217165	ANXA5	-2.11E+03	0.00E+00	2.91E-01
MST150	0.230010204	0.2788	ANXA5	-1.44E+03	1.09E-13	2.12E-01
COL5A2	0.201823129	0.260417	ANXA5	-2.06E+03	1.55E-12	1.94E-01
FRZB	0.152287308	0.491249	ANXA5	-1.62E+03	2.91E-11	1.78E-01
PAX6	0.094173729	0.396521	ANXA5	-2.43E+03	3.41E-10	1.62E-01
MMP14	0.165882148	0.402139	ANXA5	-2.45E+03	3.66E-09	1.43E-01
DES	0.192259414	0.359363	ANXA5	-1.37E+03	1.66E-07	1.14E-01
LOX	0.125711297	0.177058	ANXA5	-1.69E+03	2.65E-07	1.11E-01
NEBL	0.06257483	0.161483	ANXA5	-2.18E+03	2.55E-06	9.12E-02
SPARC	0.150543933	0.174039	ANXA5	-1.99E+03	6.49E-06	8.62E-02
FGFR2	0.048397711	0.116621	ANXA5	-2.32E+03	1.94E-05	7.98E-02

Table 9. 24 hypermethylated genes negatively regulated expression of gene BCL11B and 11 hypomethylated genes regulated expression of gene ANXA5.

3.6 Deciphering the path connecting mutations to gene and miRNA expression.

To study function of somatic mutations and LOH and connect them with disease through gene expressions and miRNAs, we study their *cis* or *trans* regulatory effects on gene or miRNA expression traits. We applied the group regression method to expression profiles of 12,042 genes and expression profiles of 470 human microRNAs in 169 tumor tissue samples from glioblastomas patients, which were shared among the gene expression, miRNA expression and mutation datasets.

Fig. 8. Somatic and LOH mutation eQTL network. (A) Somatic mutation eQTL network. A network that connects 14 genes with somatic mutations associated with GBM, their regulated mRNAs and miRNA expressions was shown. (B) LOH mutation eQTL network. A network that connects 11 genes with LOH associated with GBM, their regulated mRNAs and miRNA expressions was shown.

Mutation	Gene expression				miRNA expression			
Gene	Gene	Damage value	P-value for gene expression	Up or down regulate	miRNA	Damage value	P-value for miRNA expression	Up or down regulated
RB1	LAMP2	0	8.22E-07	Up	hsa-mir-340	0	1.16E-06	Down
DST	OTUB1	0	5.33E-07	Down	hsa-mir-15b	45	1.30E-06	Up
FN1	SSR2	0	3.67E-07	Up	hsa-mir-125a	2	4.83E-05	Down
FN1	TDG	3	5.21E-07	Up	hsa-mir-125a	2	4.83E-05	Down
TP53	TP53	23	1.40E-07	Up	hsa-mir-504	1	1.42E-06	Down

Mutation-gene regression			miRNA-gene regression			Mutation -miRNA regression		
Regression coefficient	P-value	Coefficient of determination	Regression coefficient	P-value	Coefficient of determination	Regression coefficient	P-value	Coefficient of determination
1.03E+00	7.04E-04	6.47E-02	-2.77E+02	3.17E-06	8.50E-02	1.81E+00	4.10E-09	1.85E-01
1.23E+00	5.91E-04	6.65E-02	-2.28E+01	9.77E-07	9.34E-02	1.27E+00	3.80E-04	7.18E-02
1.54E+00	1.02E-05	1.07E-01	-2.33E+02	4.92E-05	6.52E-02	-1.18E+00	8.72E-04	6.33E-02
1.39E+00	7.60E-05	8.72E-02	-6.91E+01	4.77E-05	6.54E-02	-1.18E+00	8.72E-04	6.33E-02
2.81E-01	3.30E-02	2.62E-02	-4.14E+01	2.72E-06	8.61E-02	3.43E-01	9.10E-03	3.93E-02

Table 10. Genes with somatic mutations regulate both mRNA and miRNA expressions, and formed triangle regulation cycles.

We first studied *cis* regulatory effects of somatic and LOH mutations on mRNA or miRNA expression traits. For somatic mutation, we found four cis-eQTL: TP53, EGFR, NF1and PIK3C2G. P-values for testing association of somatic mutations in TP53 , EGFR, NF1 and PIK3C2G with their expressions were 0.033, 0.019, 0.028 and 0.006, respectively. Fold

changes (defined as the ratio of their average expressions of the samples with somatic mutations over the average expressions of the samples without somatic mutations) for the 4 genes were 1.10,1.73, 0.86,1.23, respectively. Although regression analysis did not show significant association of somatic mutations in BCL11A, FN1 and COL3A1 with their expressions due to very low frequencies of mutations, the fold changes were 1.45, 0.60 and 0.20, respectively. We still observed some regulatory effects of somatic mutations on these two genes. For LOH mutation, two cis-eQTL: NRAP and EGFR were detected with P-values 0.030, 0.034 and fold change 0.90 and 2.32, respectively. Regression analysis did not show significant association of LOH mutations in gene CYP1B1 with their expressions, but the fold change were 0.48.

Next we identified *trans*-regulatory effects of somatic and LOH mutations on mRNA or miRNA expression traits. The thresholds for declaring significant association of the set of somatic mutation and LOH in the gene with mRNA and miRNA expression after Bonferroni correction for multiple tests were 1.63×10^{-4} and 4.03×10^{-4}, respectively. A network that connects 14 genes with somatic mutations associated with GBM, and their regulated mRNAs and miRNA expressions was shown in Figure 8A. Somatic mutations in these 14 genes were strongly correlated with expressions of 177 significantly differentially expressed genes, 45 of which were under-expressed and 132 were over-expressed in tumor tissue samples, a total of 262 trans gene eQTL were found. These mutations also significantly affected expressions of 23 miRNAs, 11 of which were underexpressed and 12 were overexpressed and a total of 26 trans miRNA eQTL were found. Remarkably, we found 5 paths: (1) RB1 with association of somatic mutations with GBM regulated expressions of gene LAMP2 and mir-340, and mir-340 in turn targeted gene LAMP2; (2) DST with association of somatic mutations regulated expressions of both gene OTUB1 and mir-15b and mir-15b in turn targeted gene OTUB1; (3) FN1 with association of somatic mutations regulated expressions of both SSR2 and mir-125a and mir-125a in turn targeted gene SSR2; (4) FN1 with association of somatic mutations regulated expressions of both gene TDG and mir-125a and mir-125a in turn targeted gene TDG; and (5) TP53 with association of somatic mutations regulated expressions of both gene TP53 and mir-504 and mir-504 in turn targeted TP53 (Table 10).

Similarly, Figure 8B showed a network that connects 11 genes with LOH mutations associated with GMB, their regulated mRNAs and miRNA expressions. These 11 genes with LOH as *trans*-eQTL affected 323 differentially expressed or interacted genes, 190 of which were underexpressed genes and 133 were overexpressed, a total of 409 *trans gene*-eQTL were found. The LOH also affected expressions of 19 miRNAs, 9 of which were underexpressed and 10 were overexpressed, 2 of them were differentially expressed and a total of 27 *trans miRNA*-eQTL were found.

Sources for regulating gene expressions include DNA variations, miRNAs, and methylation. To provide a comprehensive view of how DNA variations, miRNA and methylation directly or indirectly regulate gene expression, we found a total of 87 underexpressed and 48 overexpressed genes which were commonly regulated by 38 underexpressed and 28 overexpressed miRNAs, 14 hypomethylated and 107 hypemethylated genes, and 9 significant somatic mutation and 10 significant LOH mutation. Surprisingly, multiple genetic perturbations such as multiple genes with somatic mutations or LOH, multiple miRNAs and multiple methylation may directly or indirectly regulate expression of a single gene. For example, expression levels of BCL11b which is a haploinsufficient tumor suppressor gene were positively regulated by genes DST and BCL11A, with associated somatic mutations, and were negatively affected by over-expressed mir-155, mir-23a, mir-92b and mir-30a-5p, and 24 hypermethylated genes.

4. Discussion and conclusion

Genetic and epigenetic alternations that are likely to cause tumor formation are often organized into complex biological networks. Genetic and epigenetic risk factors individually cannot fully explain initiation and progression of cancer. The purpose of this report is (1) to explore the possibility of integrating altered DNA variations, mRNA and miRNA expression variation, and methylation patterns into multi-level complex genetic and epigenetic networks that contribute to tumorigenesis; (2) decipher paths from somatic mutations and LOH to tumor formation through genetic and epigenetic networks; and (3) identify key genetic and epigenetic alternations causing tumor formation by network analysis.

The approach developed here for integrating genetic and epigenetic alternations into molecular networks consists of three steps. The first step is to reconstruct single type molecular networks whose components are of the same type of molecules, either mRNAs or miRNAs. We reconstructed mRNA co-expression networks and miRNA co-expression networks for glioblastoma using partial correlation approach. The co-expression networks attempt to uncover the regulatory relationships among mRNAs or miRNAs. However, many examples show that the regulatory relationships may be changed when normal tissues become tumor tissues. To identify the alternated regulations we developed a novel mutual information-based statistic for testing interactions between mRNAs or miRNAs, which allow detecting differential regulations between tumor and normal tissues, and reconstructed mRNA and miRNA interaction networks. The second step is to identify miRNA target and reconstruct miRNA target networks that connect mRNA and miRNA co-expression networks or mRNA and miRNA interaction networks.

The third step is to decipher the paths from somatic and LOH mutations to tumor formation. The first issue is how to test association of somatic mutations or LOH with glioblastoma. We developed a group association test that is based on population genetics for assessing association of somatic mutations and LOH with cancer. We identified 14 genes harboring somatic mutations associated with glioblastoma and 11 genes harboring LOH associated with glioblastoma. The second issue is how to uncover the components of mRNA or miRNA co-expression networks and interaction networks that respond to somatic mutations and LOH associated with glioblastoma. Traditionally, when alleles are common, these components are identified by mapping *cis*-eQTL or *trans*-eQTL. However, when alleles are rare, these components are hard to find by individually mapping *cis*-eQTL or *trans*-eQTL. The approach developed here is to extend group tests for a qualitative trait to a quantitative trait. The components that respond to perturbation of rare somatic genetic variants were identified by regressing the expression of an mRNA or an miRNA on the number of all mutated alleles across the gene. We discovered large comprehensive genetic and epigenetic networks that connect genes harboring associated mutations or LOH, mRNAs and miRNAs. Interestingly, we found five triangle cycles in the networks which indicated that significantly associated somatic mutations or LOH regulated both differentially expressed mRNA and miRNA and the miRNA in turn also affected expression levels of the mRNA. The approach presented here has two remarkable features. First, it offered a powerful tool for differentiating driver mutations from passenger mutations. Second, it provides functional information on how somatic or LOH mutations lead to tumorigenesis through complex genetic and epigenetic networks. Similar to mutations,

methylation is another source for causing tumorigenesis. By regressing expression levels of mRNA or miRNA on methylation level, we can incorporate methylation into network. Finally we identified large genetic and epigenetic networks including associated mutations, LOH, methylation, differentially expressed mRNAs and miRNAs to explain how mutations and methylation cause tumorigenesis through molecular networks.

Our studies support the hypothesis that cancer may be the emergent properties of large genetic and epigenetic networks that are highly interconnected. Genetic, epigenetic and environmental stimuli can be viewed as random attacks to genetic and epigenetic networks. Topological properties of the genetic and epigenetic networks are closely related to the function of cells. Cancer arises from the failure of genetic and epigenetic networks to respond to attacks. In other words, the attacked networks are unable to return to their normal states and remain functional in the face of random perturbations. We suspect that key components of network contributing to the robustness of the network also play an important role in the function of cells. We modeled over- or under-expression of mRNA and miRNA, methylation, and genetic alternation as deletion of a node in the network which will cause dynamical changes in the network and used a damage value of the node to measure its contribution to the robustness of the network. We found the several cancers related genes such as TP53 have large damage values in the genetic and epigenetic networks. These key components in the network may serve as therapeutic intervention points.

Our results are preliminary. Although network analysis may have the potential to unravel the mechanism of tumor initiation and progression, the presented network structures and their properties in this report may depend on sample size. Whether the structures of our reconstructed genetic and epigenetic network can be replicated in other samples or not is a key to the success of network analysis in cancer studies.

5. Acknowledgments

6. References

Abu-Elneel, K., *et al.* (2008) Heterogeneous dysregulation of microRNAs across the autism spectrum, *Neurogenetics*, 9, 153-161.

Ackermann, M. and Strimmer, K. (2009) A general modular framework for gene set enrichment analysis, *BMC Bioinformatics*, 10, 47.

Barkinge, J.L., *et al.* (2009) The p53-induced Siva-1 plays a significant role in cisplatin-mediated apoptosis, *J Carcinog*, 8, 2.

Boersma, H.H., *et al.* (2005) Past, present, and future of annexin A5: from protein discovery to clinical applications, *J Nucl Med*, 46, 2035-2050.

Brillinger, D.R. (2004) Some data analyses using mutual information., *Brazilian J Probability Statistics*, 18, 163.

Brooks, A.S., *et al.* (2005) Homozygous nonsense mutations in KIAA1279 are associated with malformations of the central and enteric nervous systems, *Am J Hum Genet*, 77, 120-126.

Caligo, M.A., *et al.* (1997) NM23 gene expression in human breast carcinomas: loss of correlation with cell proliferation in the advanced phase of tumor progression, *Int J Cancer*, 74, 102-111.

Carpenter, B., *et al.* (2004) BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1, *Mol Cell Biol*, 24, 537-549.

Chu, F., *et al.* (2005) Expression of Siva-1 protein or its putative amphipathic helical region enhances cisplatin-induced apoptosis in breast cancer cells: effect of elevated levels of BCL-2, *Cancer Res*, 65, 5301-5309.

Chuang, H.Y., *et al.* (2007) Network-based classification of breast cancer metastasis, *Mol Syst Biol*, 3, 140.

Cimino, G., *et al.* (2001) Molecular evaluation of the NUP98/RAP1GDS1 gene frequency in adults with T-acute lymphoblastic leukemia, *Haematologica*, 86, 436-437.

Cover, T. and Thomas, J. (1991) *Elements of information theory.* . John Wiley and Sons Inc, New York.

Dartnell, L., *et al.* (2005) Robustness of the p53 network and biological hackers, *FEBS Lett*, 579, 3037-3042.

Demetrius, L. and Manke, T. (2005) Robustness and network evolution-an entropic principle., *Physica A: Statistical Mechanics and its Applications.* , 346, 682

El Hallani, S., *et al.* (2009) TP53 codon 72 polymorphism is associated with age at onset of glioblastoma, *Neurology*, 72, 332-336.

Fasanaro, P., *et al.* (2008) MicroRNA-210 modulates endothelial cell response to hypoxia and inhibits the receptor tyrosine kinase ligand Ephrin-A3, *J Biol Chem*, 283, 15878-15883.

Forch, P. and Valcarcel, J. (2001) Molecular mechanisms of gene expression regulation by the apoptosis-promoting protein TIA-1, *Apoptosis*, 6, 463-468.

Gennarino, V.A., *et al.* (2009) MicroRNA target prediction by expression analysis of host genes, *Genome Res*, 19, 481-490.

Greenman, C., *et al.* (2007) Patterns of somatic mutation in human cancer genomes, *Nature*, 446, 153-158.

Hagedorn, M., *et al.* (2007) FBXW7/hCDC4 controls glioma cell proliferation in vitro and is a prognostic marker for survival in glioblastoma patients, *Cell Div*, 2, 9.

Hansen, C.N., *et al.* (2009) Expression of CPEB, GAPDH and U6snRNA in cervical and ovarian tissue during cancer development, *APMIS*, 117, 53-59.

Hashimoto, K., *et al.* (2006) KEGG as a glycome informatics resource, *Glycobiology*, 16, 63R-70R.

Hebert, C., *et al.* (2007) High mobility group A2 is a target for miRNA-98 in head and neck squamous cell carcinoma, *Mol Cancer*, 6, 5.

Helleman, J., *et al.* (2006) Molecular profiling of platinum resistant ovarian cancer, *Int J Cancer*, 118, 1963-1971.

Heraud, C., *et al.* (2008) Vasoactive intestinal peptide-induced neuritogenesis in neuroblastoma SH-SY5Y cells involves SNAP-25, *Neuropeptides*, 42, 611-621.

Herrera-Merchan, A., *et al.* (2010) miR-33-mediated downregulation of p53 controls hematopoietic stem cell self-renewal, *Cell Cycle*, 9, 3277-3285.

Holm, H. and Alouini, M.-S. (2004) Sum and difference of two squared correlated Nakagami variates in connection with the McKay distribution. , *IEEE Trans. Commun.*, 52, 10.

Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat Protoc*, 4, 44-57.

Huang, J.C., *et al.* (2007) Using expression profiling data to identify human microRNA targets, *Nat Methods*, 4, 1045-1049.

Huang, J.C., Morris, Q.D. and Frey, B.J. (2007) Bayesian inference of MicroRNA targets from sequence and expression data, *J Comput Biol*, 14, 550-563.

Hussey, D.J., *et al.* (1999) The (4;11)(q21;p15) translocation fuses the NUP98 and RAP1GDS1 genes and is recurrent in T-cell acute lymphocytic leukemia, *Blood*, 94, 2072-2079.

Imoto, I., *et al.* (2000) Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines, *Biochem Biophys Res Commun*, 276, 264-270.

Jakulin, A., *et al.* (2003) Attribute interactions in medical data analysis., *Proceedings of the 9th Conference on Artificial Intelligence in Medicine in Europe*. Protaras, Cyprus.

Jeffrey, A. and Zwillinger, D. (2000) Table of Integrals, Series, and Products. In Gradshteyn and Ryzhik (eds). pp. 1163.

Jiang, Q., *et al.* (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease, *Nucleic Acids Res*, 37, D98-104.

Joyce, P. and Tavare, S. (1995) The distribution of rare alleles, *J Math Biol*, 33, 602-618.

Karginov, F.V., *et al.* (2007) A biochemical approach to identifying microRNA targets, *Proc Natl Acad Sci U S A*, 104, 19291-19296.

Kim, J.K. and Diehl, J.A. (2009) Nuclear cyclin D1: an oncogenic driver in human cancer, *J Cell Physiol*, 220, 292-296.

Kuhn, D.E., *et al.* (2008) Experimental validation of miRNA targets, *Methods*, 44, 47-54.

Laneve, P., *et al.* (2007) The interplay between microRNAs and the neurotrophin receptor tropomyosin-related kinase C controls proliferation of human neuroblastoma cells, *Proc Natl Acad Sci U S A*, 104, 7957-7962.

Langfelder, P., Zhang, B. and Horvath, S. (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, *Bioinformatics*, 24, 719-720.

Li, Z., *et al.* (2009) Biological functions of miR-29b contribute to positive regulation of osteoblast differentiation, *J Biol Chem*, 284, 15676-15684.

Liang, P. and Pardee, A.B. (2003) Analysing differential gene expression in cancer, *Nat Rev Cancer*, 3, 869-876.

Maglott, D., *et al.* (2007) Entrez Gene: gene-centered information at NCBI, *Nucleic Acids Res*, 35, D26-31.

Matsuda, H. (2000) Physical nature of higher-order mutual information: intrinsic correlations and frustration, *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 62, 3096-3102.

Maziere, P. and Enright, A.J. (2007) Prediction of microRNA targets, *Drug Discov Today*, 12, 452-458.

McGill, W. (1954) Multivariate information transmission., *Psychometrika*, 19, 97.

Megraw, M., *et al.* (2007) miRGen: a database for the study of animal microRNA genomic organization and function, *Nucleic Acids Res*, 35, D149-155.

Mingorance, A., Soriano-Garcia, E. and del Rio, J.A. (2004) [Nogo-A functions during the development of the central nervous system and in the adult], *Rev Neurol*, 39, 440-446.

Nakahara, H., *et al.* (2003) Gene interaction in DNA microarray data is decomposed by information geometric measure, *Bioinformatics*, 19, 1124-1131.

Namkoong, J., *et al.* (2007) Metabotropic glutamate receptor 1 and glutamate signaling in human melanoma, *Cancer Res*, 67, 2298-2305.

Ning, X., *et al.* (2007) Calcyclin-binding protein inhibits proliferation, tumorigenicity, and invasion of gastric cancer, *Mol Cancer Res*, 5, 1254-1262.

Pacifico, F., *et al.* (2007) RbAp48 is a target of nuclear factor-kappaB activity in thyroid cancer, *J Clin Endocrinol Metab*, 92, 1458-1466.

Parsons, D.W., *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme, *Science*, 321, 1807-1812.

Peng, J., *et al.* (2009) Partial correlation estimationby joint sparse regression model, *Journal of the American Statistical Association*, 104 735.

Rayner, K.J., *et al.* (2010) MiR-33 contributes to the regulation of cholesterol homeostasis, *Science*, 328, 1570-1573.

Rubin, A.F. and Green, P. (2007) Comment on "The consensus coding sequences of human breast and colorectal cancers", *Science*, 317, 1500.

Sassen, S., Miska, E.A. and Caldas, C. (2008) MicroRNA: implications for cancer, *Virchows Arch*, 452, 1-10.

Selbach, M., *et al.* (2008) Widespread changes in protein synthesis induced by microRNAs, *Nature*, 455, 58-63.

Sengupta, S., *et al.* (2008) MicroRNA 29c is down-regulated in nasopharyngeal carcinomas, up-regulating mRNAs encoding extracellular matrix proteins, *Proc Natl Acad Sci U S A*, 105, 5874-5878.

Sethi, T., *et al.* (1999) Extracellular matrix proteins protect small cell lung cancer cells against apoptosis: a mechanism for small cell lung cancer growth and drug resistance in vivo, *Nat Med*, 5, 662-668.

Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets, *RNA*, 12, 192-197.

Skalsky, R.L., *et al.* (2007) Kaposi's sarcoma-associated herpesvirus encodes an ortholog of miR-155, *J Virol*, 81, 12836-12845.

Spinicelli, S., *et al.* (2002) GITR interacts with the pro-apoptotic protein Siva and induces apoptosis, *Cell Death Differ*, 9, 1382-1384.

Sun, S., *et al.* (2007) Overexpressed CacyBP/SIP leads to the suppression of growth in renal cell carcinoma, *Biochem Biophys Res Commun*, 356, 864-871.

The Cancer Genome Atlas Research Network, T.C.G.A.R.N. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways, *Nature*, 455, 1061-1068.

Turashvili, G., *et al.* (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis, *BMC Cancer*, 7, 55.

Wang, C., *et al.* (2008) A novel endogenous human CaMKII inhibitory protein suppresses tumor growth by inducing cell cycle arrest via p27 stabilization, *J Biol Chem*, 283, 11565-11574.

Wang, X.Y., *et al.* (1998) GBAS, a novel gene encoding a protein with tyrosine phosphorylation sites and a transmembrane domain, is co-amplified with EGFR, *Genomics*, 49, 448-451.

Wood, L.D., *et al.* (2007) The genomic landscapes of human breast and colorectal cancers, *Science*, 318, 1108-1113.

Wu, X., Jin, L. and Xiong, M. (2008) Composite measure of linkage disequilibrium for testing interaction between unlinked loci, *Eur J Hum Genet*, 16, 644-651.

Yang, H.S., *et al.* (2004) A novel function of the MA-3 domains in transformation and translation suppressor Pdcd4 is essential for its binding to eukaryotic translation initiation factor 4A, *Mol Cell Biol*, 24, 3894-3906.

Yang, Y., *et al.* (2009) Identification of miR-21 targets in breast cancer cells using a quantitative proteomic approach, *Proteomics*, 9, 1374-1384.

Yang, Y. and Fu, L.M. (2003) TSGDB: a database system for tumor suppressor genes, *Bioinformatics*, 19, 2311-2312.

Yang, Y., Wang, Y.P. and Li, K.B. (2008) MiRTif: a support vector machine-based microRNA target interaction filter, *BMC Bioinformatics*, 9 Suppl 12, S4.

Zhao, J., Jin, L. and Xiong, M. (2006) Test for interaction between two unlinked loci, *Am J Hum Genet*, 79, 831-845.

Zhu, Y. and Parada, L.F. (2002) The molecular and genetic basis of neurological tumours, *Nat Rev Cancer*, 2, 616-626.
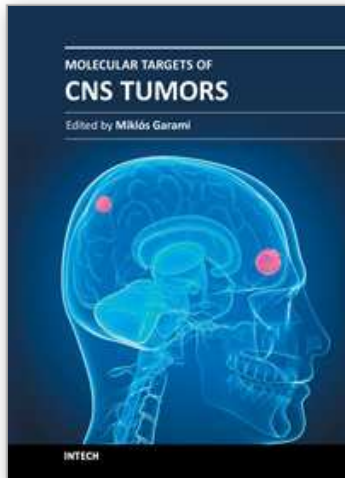
Molecular Targets of CNS Tumors

Edited by Dr. Miklos Garami

Molecular Targets of CNS Tumors is a selected review of Central Nervous System (CNS) tumors with particular emphasis on signaling pathway of the most common CNS tumor types. To develop drugs which specifically attack the cancer cells requires an understanding of the distinct characteristics of those cells. Additional detailed information is provided on selected signal pathways in CNS tumors.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hua Dong and MomiaoXiong (2011). Integrated Network Analysis of Genetic and Epigenetic Factors in Glioblastoma Multiform, Molecular Targets of CNS Tumors, Dr. Miklos Garami (Ed.), ISBN: 978-953-307-736-9, InTech, Available from: http://www.intechopen.com/books/molecular-targets-of-cns-tumors/integrated-network-analysis-of-genetic-and-epigenetic-factors-in-glioblastoma-multiform

INTECH
open science | open minds