Comparative microarray data analysis for the expression of genes in the pathway of glioma

Pramod Katara*, Neeru Sharma, Sugandha Sharma, Indu Khatri, Akansha Kaushik, Lalima Kaushal, Vinay sharma

Department of Bioscience & Biotechnology; Banasthali University, Bansthali 304022, India; Pramod Katara Email id: pmkatara@gmail.com, *Corresponding author

Abstract:
Our present work focuses on the set of genes, which are involved in primary brain tumors – the glioma pathway. These gliomas are mostly malignant (cancerous) in nature and are difficult to be cured and that's why they attract the attention of all the workers. To understand the relative functionality of these genes, we analyzed the expression pattern of all genes, using gene expression data, at genomic level, and then to check their universality in all other cancers, we compared their expression levels and patterns in all other types of cancers by using gene expression graphs, and observed their expression levels in all these cancers, whether they are over or under expressed. We found that every gene has its own unique expression pattern and level and on that basis it can be classified. We also found that oncogenes and tumor suppressor genes that were involved in the glioma pathway were showing similar expression patterns in other cancers too but their expression level is low.

Keywords: cancer, co-expression, gene expression pattern, glioma pathway.

Background:
Gliomas are among the most aggressive malignant tumors and the most refractory to therapy, in part due to the propensity for malignant cells to disseminate diffusely through out the brain. Glioma is a type of cancer that starts in brain and spine. It arises from glial cells. The most common site of gliomas is brain [1].Various workers have used the microarray technology to analyze the differential expression of the genes involved in the various cancers and related pathways, especially in the case of human [2] Gene expression profiling is proven useful in sub-classification and outcome prognostication for human glial brain tumors. The analysis of biosignificance of the 100s and 1000s of alteration in gene expression found in genomic profiling remains a major challenge. Moreover, it is increasingly evident that genes do not act as an individual unit, but collaboration in overlapping network, the deregulation which is a hallmark of cancer [3].

Over the last few years, the routine use of microarrays has made possible the creation of large datasets of molecular information characterizing complex biological systems. A single sample for microarray contains measurements for around 10,000 genes and hence the amount of data in each microarray is too overwhelming for manual analysis [4]. The true power of microarray analysis does not come from the analysis of single experiment but rather from the analysis of many hybridization to identify common pattern of gene expression and hence based on the available understanding of the cellular processes, the genes that are contained in a particular pathway or that responds to a common environmental challenge should be co-regulated and consequently should show similar patterns of expression [5].

Modern experimental techniques such as microarray analysis of gene expression, are improving our understanding of both the classification and biological basis of complex diseases. While there has been an explosion in the volume of raw data available for analysis, there is a widening gap between statistically compelling results and their biological interpretation. Research oftentimes becomes bogged down in an analytical maze of spreadsheets and arbitrary statistical significance thresholds. Tools are needed that can efficiently summarize huge amounts of information within a biological context. Novel combination of microarray clustering features supports such a comprehensive analysis. Combining flexibility, speed, and

visualization of both statistical and annotative information into a single package, hierarchical clustering and k-means clustering fulfills a crucial role in comprehensive microarray analysis [6].

The biological interpretation of gene expression microarray results is a daunting challenge. For complex diseases such as cancer, wherein the body of published research is extensive, the incorporation of expert knowledge provides a useful analytical framework. However, unexpected differences in survival time of various tumors have generated attempts to search for more precise parameters [6]. It has been clear that tumor behavior depends mostly on gene expression alterations of various genes at genomic level [7], thus the knowledge of single gene alterations failed to accurately define pattern and survival time of various malignant tumors; however, the gene expression profiling, based on microarray technology has raised hopes [8]. Our main objective is to perform a comparative analysis of expression pattern of various genes involved in the glioma pathway of *Homo sapiens* to understand their relative functionality in this pathway and also in various other cancers. Such analysis could be used in regulation of the disease by altering the expression pattern of responsible genes. This would further provide guidance for drug target.

Methodology:
We first of all retrieved the genes that are involved in pathway of glioma in Homo sapiens from the KEGG database and found that 65 genes are involved in the glioma pathway [9]. For the analysis of gene expression profile of these all genes we downloaded the microarray data from SMD database related to glioma and for their comparative expression analysis we also downloaded the data for other cancers i.e. breast, lung, Prostate, Liver, Pancreas, and Miscellaneous (various Acute Myeloid Leukemia, Cell line cancers, Ovarian Cancers and Skin Cancers) [10]. Before the analysis, we processed the whole data by taking gene expression ratio ($\log_2$ values) from the downloaded files. Then the data was normalized for missing values, by following the neutral method, which was proposed by Alizadeh *et al.* (2000) for the analysis of diffused large B-cell lymphoma [11] and for that we replaced the missing (empty) values with zero. After preparation of the working data, we carried the whole analysis in two different parts. For comparative analysis of various cancer patterns along with the glioma with reference to the known glioma suppressor gene i.e. PTEN, p53 and Rb1 and oncogenes i.e. EGFR, MDM2, PDGF family,

PDGFR family and CDK4 and analyzed their relative role in all the cancers and for that we created gene and sample expression profile and converted them in tabular form where each column represents a single gene and each row represents a single cancer type (Table 1 in supplementary material). Further, for the analysis of comparative expression of glioma genes within glioma, we analyzed their gene expression profile using all available data for glioma, and then for the analysis of comparative expression of these genes in other cancer, we analyzed the expression profile of these all genes using the data of all cancers except glioma and clumped them. For better and simplified understanding, we converted overall expression level (over/under/mix) in percentage (**Figure 1a and 1b).**
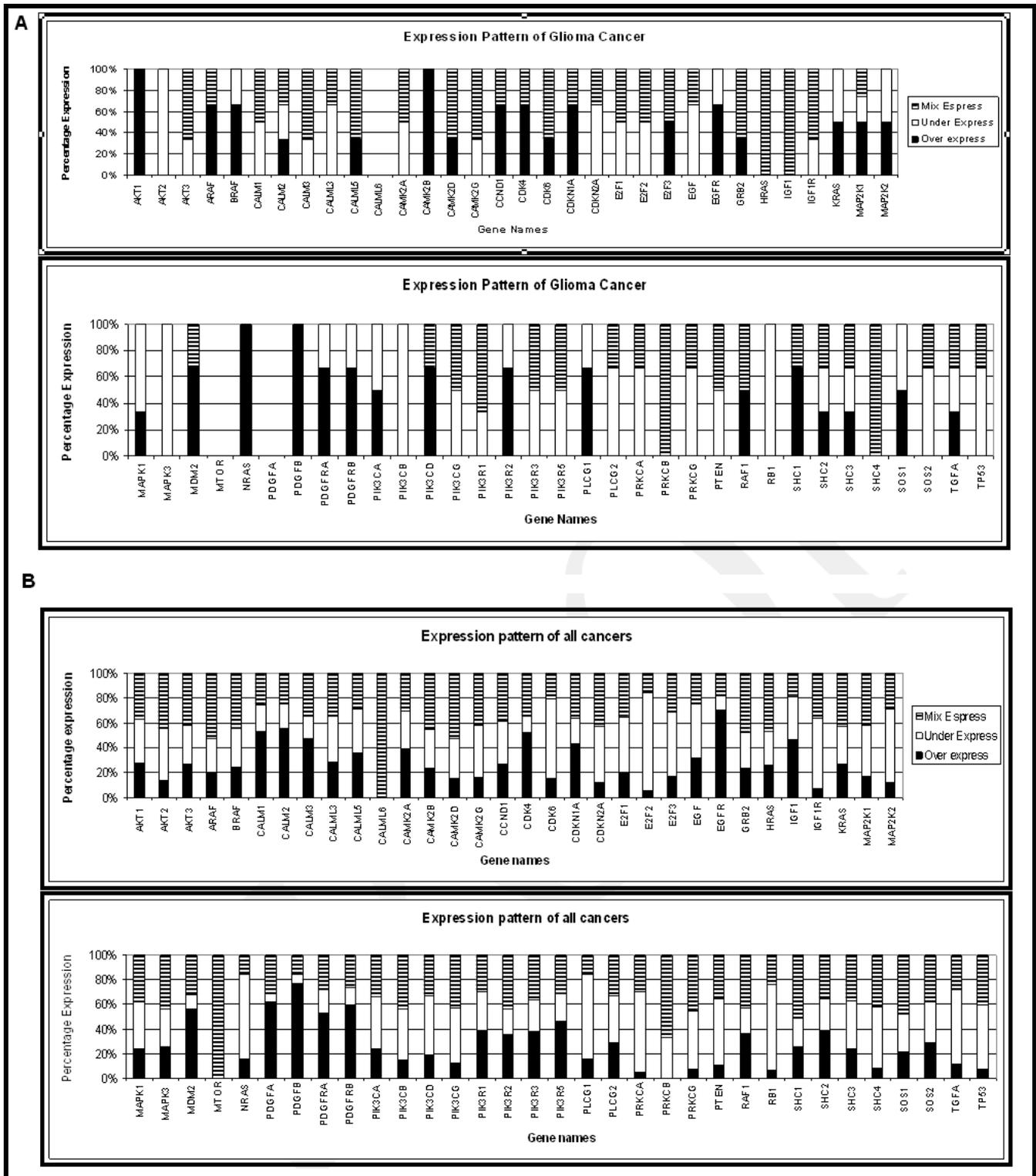

Figure 1: (A) Gene expression level of all 65 glioma genes in glioma tumors; **(B)** gene expression level of all 65 glioma genes in overall cancers

Bioinformation

www.bioinformation.net

Volume 5

Issue 1

open access

Hypothesis

Discussion:

In this study we concentrated on the differentially expressed genes that are involved in the pathway of glioma and might be involved in the other cancers. From the results of this study, expression of glioma pathway genes is characterized as given in **Table 1 (see supplementary material).** It depicts that the genes CDKN2A, PTEN, RB1 and TP53 are expressing at a lower level than the normal and are also an important cause in causing the cancers. This means that these genes are functioning as the tumor suppressor genes and whenever they are present in the tissues that we have taken into consideration; they are likely to cause the cancer. In contrast, the genes CDK4, EGFR, MDM2, PDGFA, PDGFB, PDGFRA and PDGFRB are showing an enhanced expression in the cancers and hence show that the genes are functioning as the oncogenes in the various types of cancers. It means, whenever these genes are over expressing in the tissues type we studied, they are expected to cause the cancer in these tissues. Even these features were compared in the various other cancers whose data has been downloaded from SMD itself **[10]**. The percentage expression of the oncogenes and tumor suppressor genes is almost same in the other cancers also, like breast cancer, liver cancer, prostate cancer, lung cancer, pancreatic cancer and miscellaneous cancers. Miscellaneous cancers involve various Acute Myeloid Leukemia, Cell line cancers, Ovarian Cancers and Skin Cancers. By the analysis of relative expression patterns of the glioma suppressor and oncogenes in all cancers, we observed that CDK4, MDM2, EGFR, PDGFA, PDGFB and PDGFRA have almost shown a similar pattern and their percentage for over expression was observed to be very high and it was found that they have barely shown under expression. Here it was also observed that except the PFGFA for which we did not found sufficient data, on an average all these genes showed the same expression percentage in all the cancers that they have shown in glioma. Genes CDKN2A, PTEN, RB1, and TP53 were found to be very close to each other and shared a comparatively very high percentage for under expression pattern and very low percentage for over expression in glioma and all other cancers. The percentage expression of PDGFRB is quite high for over expression in breast and glioma but for other cancers (on an average) it has shown the mixed expression.

These observations are almost similar as observed by wet lab experiments, thus they supported the earlier suggested views that their over expression made positive effects on various tumors growth **[12, 13, 14]**, in spite of all this, intensive research and clinical trials are already going on PDGF receptors to test it as the therapeutic targets **[12]**. Here we suggest that CDK4, MDM2, EGFR, PDGFA, PDGFB and PDGFRA genes can be used as the marker because they are robust and can be predicted by any method (computational and wet lab).

Detection of comparative expression level of all 65 genes involved in glioma pathway (**Figure 1a and 1b**) suggests that, in glioma cancer most of these genes show mixed type of expression levels. Some of them show 100% over expression and some shows 100% under expression. Expression level of mTOR, CALML6 and PDGFA was not mentioned in the figure because of the non availability of the sufficient amount of the data. It has been shown that to maintain a particular process or biological pathway it is essential that genes which are related to that process remains active, suppression of such genes results in lost of such activity. On the basis of this concept, essential genes can be used as the target to control pathway. AKT1, CAMK2B and NRAS genes shows 100% over expression in all tumors, it suggests that these three genes work as essential genes and might be playing some important role to maintain the glioma pathway.

Here, essentiality of all these three genes suggests that they could act as very good drug targets **[15]**. Further analysis suggests that ARAF, CCND1, CDK4, CDKN1A, MDM2, PIK3CD and SHC1 genes show more than 60% over expression and rest of the time a mixed expression , thus they also can be checked for being probable drug targets. The expression levels of these 65 genes were also observed in different cancers (Breast, Lung, Prostate, Liver and Pancreas) (**Figure 1b**) and suggests that, the genes of glioma pathway are also active in most of the cancers but their expression levels remain very low, thus their contributions in these all cancers seems little but positive.

Conclusion:

Analysis of comparative expression patterns of the genes of glioma pathway suggests that the genes CDKN2A, PETN, RB1 and TP53 are acting as the tumor suppressor genes, the under expression of which are a cause of cancer that we have analyzed and the genes CDK4, PDGFA and PDGFB, PDGFRA and PDGFRB, MDM2 and EGFR are over expressing at a rate higher than the normal and are acting as oncogenes in these cancers. The analyses of gene expression further provide us with the knowledge of oncogenes and tumor suppressor genes that can be further used in disease profiling. It suggests that the use of gene expression data can improve the functional gene annotation of the genes that were previously unknown. Further the detailed knowledge of oncogenes and tumor suppressor genes can be extended to get some potential drug targets against which drugs can be designed.

References:

[1] A Chakravarti *et al. Journal of Clinical Oncology* (2002) **20**:1063 [PMID: 11844831]

[2] CR Englert *et al. Curr Opin Mol Ther* (1999) **1(6):**712 [PMID: 19629868]

[3] M Bredel *et al. Cancer Research* (2005) **65**:4088 [PMID: 15899798]

[4] O Persson *et al. Journal of Neuro-Oncology* (2007)**85**:11 [PMID: 17634744]

[5] MB Eisen *et al. Proc Natl Acad Sci U S A* (1998)**95 (25):**14863 [PMID: 9843981]

[6] DM Reif *et al. Cancer Inform* (2007) **5**:19 [PMID: 19390666]

[7] I Shmulevich *et al. Cancer* (2002) **94(7)**:2069 [PMID: 11932911]

[8] T Czernicki *et al. Int J Oncol* (2007) **30**:55 [PMID: 17143512]

[9] http://www.genome.jp/kegg/pathway/hsa/hsa05214.html

[10] http://smd.stanford.edu/

[11] AA Alizadeh. *et al. Nature* (2000) **403:**503 [PMID: 10676951]

[12] A Ostman & CH Heldin, *Adv Cancer Res.* (2007) **97**:247 [PMID: 17419949]

[13] ML Rodriguez-Puebla *et al. Am J Pathol* (2002) **161(2):**405 [PMID: 12163365]

[14] K Pietras *et al. Cancer Res* (2002) **62 (19):**5476 [PMID: 12359756]

[15] W Hu *et al. PLoS Pathog* (2007) **3**:e24. [PMID: 17352532]

Supplementary materials:

Table 1: Percentage expression of glioma genes in various cancers

		CDK4	CDKN2A	EGFR	MDM2	PDGFA	PDGFB	PDGFRA	PDGFRB	PTEN	RB1	TP53
Cancer	Glioma	68,0,32	0,65,35	68,32,0	68,0,32	0,0,0	100,0,0	68,32,0	68,32,0	0,50,50	0,100,0	0,68,32
	Breast	75,10,15	25,15,60	78,14,8	60,10,30	68,32,00	90,10,0	50,25,25	90,0,10	20,50,30	0,70,30	0,55,45
	Lung	50,25,25	25,50,25	75,25,0	75,0,25	68,0,32	75,20,5	100,0,0	50,50,0	0,75,25	0,100,0	0,75,25
	Prostate	60,0,32	0,68,32	68,0,32	100,0,0	50,0,50	50,0,50	35,25,40	50,0,50	0,50,50	0,100,0	0,35,65
	Liver	60,20,20	20,40,40	80,0,20	80,0,20	100,0,0	100,0,0	0,25,75	40,20,40	0,100,0	0,75,25	0,40,60
	Pancreas	0,0,100	0,100,0	100,0,0	0,0,100	100,0,0	100,0,0	100,0,0	0,0,100	0,50,50	0,100,0	0,50,50
	Miscellaneous	35,20,45	8,22,70	60,8,32	35,35,30	0,0,100	62,8,32	50,0,50	48,22,30	15,40,45	20,50,30	25,50,25
	Overall %	51,11,38	12,49,39	75,11,14	59,7,34	64,6,30	82,6,12	57,16,27	51,17,32	6,58,36	4,83,13	4,53,43

(Note: % expression = Over, Under, Mixed expression, respectively)