

A survey of glioblastoma genomic amplifications and deletions

Shailaja K. Rao · Jennifer Edwards ·
Avadhut D. Joshi · I-Mei Siu · Gregory J. Riggins

Received: 25 July 2008 / Accepted: 6 July 2009 / Published online: 17 July 2009
© Springer Science+Business Media, LLC. 2009

Abstract Glioblastoma Multiforme (GBM) is a malignant brain cancer that develops after accumulating genomic DNA damage that often includes gene amplifications and/or deletions. These copy number changes can be a critical step in brain tumor development. To evaluate glioblastoma genomic copy number changes, we determined the genome-wide copy number alterations in 31 GBMs. Illumina Bead Arrays were used to assay 22 GBMs and Digital Karyotyping was used on 8 GBM cell lines and one primary sample. The common amplifications we observed for all 31 samples was *GLI/CDK4* (22.6%), *MDM2* (12.9%) and *PIK3C2B/MDM4* (12.9%). In the 22 GBM tumors, *EGFR* was amplified in 22.7% of surgical biopsies. The most common homozygously deleted region contained *CDKN2A/CDKN2B* (p15 and p16) occurring in 29% of cases. This data was compiled and compared to published array CGH studies of 456 cases of GBMs. Pooling our Illumina data with published studies yielded these average amplification rates: *EGFR*—35.7%, *GLI/CDK4*—13.4%, *MDM2*—9.2%, *PIK3C2B/MDM4*—7.7%, and *PDGFRA*—7.7%. The *CDKN2A/CDKN2B* locus was deleted in 46.4% of the combined cases. This study provides a larger assessment of amplifications and deletions in glioblastoma patient populations and shows that several different copy number technologies can produce similar results. The main pathways known to be involved in GBM tumor formation such as p53 control, growth signaling, and cell cycle control are all represented by amplifications or deletions of critical pathway genes. This information is

potentially important for formulating targeted therapy in glioblastoma and for planning genomic studies.

Keywords Glioblastoma multiforme · Oncogenomics · Genomic amplifications · Homozygous deletions · SNP array · Digital karyotyping · Copy number alterations · Whole genome scans

Introduction

The ability to find mutations and other related genomic changes has greatly improved due to advances in cancer genome analysis. Of the different cancer-causing genomic changes that occur, copy number changes are technically the easiest to identify by whole genome scans. Comparative Genomic Hybridization (CGH) was first widely employed to find large amplifications and deletions, but as a microscopy-based technique, its resolution is limited. Newer high-resolution techniques such as array CGH [1], cDNA microarrays [2], Digital Karyotyping (DK) [3], Illumina Bead chip arrays [4], and SNP arrays [5] have been successful at finding important amplifications and deletions that contribute to tumor formation. For many cancers, the information on newly identified amplified and deleted regions is accumulating faster than the ability to fully explore this data. Additionally, reports on copy number changes in a particular cancer will have conflicting conclusions, or be based only on a small number of samples with potential selection bias.

Glioblastoma multiforme (GBM), the most common brain cancer, is the highest grade (WHO grade IV) astrocytic tumor and accounts for the majority of astrocytic tumors [6]. This deadly tumor accumulates numerous genetic changes during its development and is characterized

S. K. Rao · J. Edwards · A. D. Joshi · I.-M. Siu ·
G. J. Riggins (✉)
Department of Neurosurgery, Johns Hopkins University, CRB II
Rm. 257, 1550 Orleans Street, Baltimore, MD 21231, USA
e-mail: griggin1@jhmi.edu

by frequent DNA copy number imbalances [7]. Due to the interest in this tumor, there has been an increase in the number of genomics studies that now include The Cancer Genome Atlas (TCGA) of the United States National Institutes of Health (<http://cancergenome.nih.gov>). Given the amount of cancer genome information that has come from these genomic studies of GBM, it is useful to analyze and summarize what is currently known. Since the copy number changes have been assayed in many studies for GBM, a cumulative analysis of copy number changes can provide more accurate conclusions on the extent of these alterations and a further understanding of the genetic changes in glioblastoma.

To find genomic changes like amplifications and deletions in GBM, array CGH has been the most commonly employed technique [8–11]. Several of these studies also show that consistent changes can help classify GBM subtypes [11, 12]. SNP arrays using Affymetrix technology have also been used for the genomic survey of gliomas, including GBM [13]. Although the information from these studies individually is useful, a comparison of this data and the addition of new genomic data should provide better estimates of consistently amplified and deleted regions in GBM.

In this study, we generated new copy number analyses on 31 cases of GBMs. Nine genomes were analyzed by Digital Karyotyping (DK) and 22 genomes were analyzed by Illumina Bead arrays. We then compared and combined our data with those of five published GBM array CGH studies and the data from TCGA project to identify the most frequently occurring gene amplifications and homozygous deletions in a total of 456 cases. Functional classification of the likely oncogenes or tumor suppressors altered by these copy number changes cluster into three major cancer-causing pathways that regulate either p53 or the cell cycle, or signal malignant changes via receptor tyrosine kinases (RTKs). This compilation of the copy number changes in GBM will be a useful aid for identifying the most promising molecular targets for the most common brain cancer.

Materials and methods

Tissue samples and cell lines

A total of 31 DNA samples were used for this study. Fresh tissues from 22 GBM resections were snap-frozen and stored for later DNA purification. Consented patients ranged in age from 30 to 80 years. Eight cell lines and a bulk tumor sample were obtained from the Duke Brain Tumor Bank, DNA purified and used for DK analysis. Cultured cell lines were initially used with the intention of avoiding

contaminating normal cells, although carefully prepared bulk tumor samples later proved to not contain enough normal DNA to interfere with the assays, and avoided bias due to EGFR amplification loss during culturing. The 22 DNA samples for the Illumina Bead Array analysis were purified from bulk tumor samples ($n = 20$), a GBM adherent cell line grown in serum ($n = 1$) and a GBM spheroid line grown in stem cell media ($n = 1$).

Digital karyotyping

For each library, 1 μg of genomic DNA was digested with the mapping enzyme *Sac* I, ligated to biotinylated *Sac* I linkers (Integrated DNA Technologies, Coralville, IA, USA), and digested with the fragmenting enzyme *Nla* III. DNA fragments containing the biotinylated linkers were isolated using Streptavidin-coated magnetic beads (DynaL Biotech, Brown Deer, WI, USA) and ligated to linkers containing recognition sites for *Mme* I. The 21 bp sequence tags were released by digestion with *Mme* I as has been described for Long SAGE [14]. A detailed protocol can be obtained at www.digitalkaryotyping.org [3]. The isolated tags were ligated to form ditags, PCR amplified, concatenated and cloned into the pZero vector (Invitrogen, Carlsbad, CA, USA). DNA from the plasmid inserts containing serial genomic tags were purified and sequenced at Agencourt Bioscience Corporation (Beverly, MA, USA) as part of the Cancer Genome Anatomy Project. The genomic tags were extracted from the sequence files using the SAGE 2000 software. The genomic content of the tumor was compared with the genomic average of virtual tags in the human genome with a sliding window of 300. The virtual genomic tags were extracted from the human genome sequence (UCSC human genome assembly, July 2003 freeze, Build 34, <http://genome.ucsc.edu/>). The virtual tags can be downloaded from www.digitalkaryotyping.org.

DK data analysis

The genomic tags were extracted with the SAGE2002 software using a tag length of 17 and a ditag length of 38. The experimental tags were filtered to remove any repetitive elements and matched with virtual tag sequences extracted from the public genome sequence (UCSC human genome assembly, July 2003 freeze, Build 34) to remove any tags not found in the virtual tag list. The DK software package was used to align the filtered tags with each chromosome and calculate the tag densities. The densities were calculated for each window by summing the experimental tags observed and dividing the sum by the average tag count for all the same-sized windows across the genome. In the DK software, we set a window size of 300. The

genes in the amplified or deleted regions were determined by looking at the genomic coordinates using the UCSC Genome Browser. Regions of amplification are defined as 8 or more copies per diploid genome. Chromosomal loss regions are defined as 1 or fewer copies per diploid genome. Chromosomal deletions are defined as 0 copies per diploid genome.

Illumina Bead Array hybridization

The DNA extraction and purification was performed using a DNA purification column (Qiagen, Valencia, CA). Approximately 750 ng of DNA at a concentration of 50 ng/μl was fragmented, precipitated, re-suspended, and hybridized to the Illumina Hap317 chips. An extension discriminated the SNP base on the surface of the chip. These bases were then stained and the chip was washed, dried, and then scanned for the resulting 317,000 SNP calls and copy number response. Hybridization was done in a teflow cell constructed by the assembly of a chip and a capillary gap system incubated in an oven over night at 56°C. These samples were run at the JHU SNP Center on a particular whole genome array which provides genotype and copy number analysis information.

Illumina data analysis

Data was analyzed by using Illumina Bead Studio Software (Illumina, San Diego, CA, USA). Bead Studio provides visualization tools to analyze single samples for LOH or chromosomal aberrations. The Illumina Genome Viewer (IGV) can plot any portion of the full data table for any sample and any chromosome. The data of interest are B allele frequency, Log R ratio, LOH score and Copy Number Score. In the present analysis, the data used was the Log R ratio. The Log R Ratio for a sample is the log (base 2) ratio of the normalized R value for the particular SNP divided by the expected normalized R value. The red line in the log R plot indicates a smoothing series with a 200 kb moving average window. Thus, a Log R Ratio > 2 was considered to represent a true amplification and Log R Ratio < -1.5 was considered to represent a probable homozygous deletion.

Comparative studies

A comparative study was undertaken combining whole genome copy number analysis of GBMs. The criterion for inclusion in this analysis was at least 20 GBMs analyzed using a high resolution technique, with complete data in an accessible format. Five studies with more than 20 glioblastomas using genome wide scanning technology and data from Cancer Genome Atlas ([www.tcgadata.nci.nih.gov](http://www.tcgadata.nci.nih.gov/tcga)) with 206 studies to find amplifications and deletions were included for comparison [8–12]. Amplifications and homozygous deletions at defined genomic loci were calculated for their percentage of occurrence. These genome analyses were combined with the 31 GBMs analyzed by Illumina and DK to yield combined percentages in a total of 456 cases.

Statistical analysis of the combined dataset

Statistical analysis of the combined dataset

A comparative statistical analysis was also conducted to calculate the rates of amplifications and homozygous deletions in the above five studies. An overall Chi-square test was performed and a statistical significance was defined with a *P* value <0.05. Multiple comparisons were done using Fisher's exact test to check the similarity and deviations occurring amongst the six studies undertaken and a statistical significance was defined with a *P* value <0.01. These analyses were performed using PRISM software, version 5 (Graphpad, San Diego, CA, USA).

Results

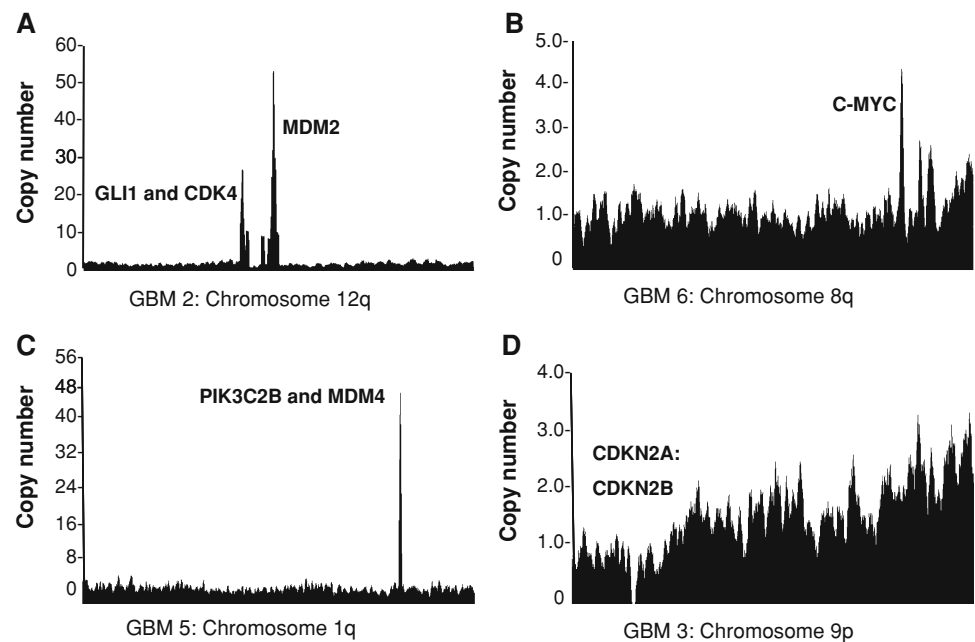
Genomic aberrations detected by DK

DK was used to search for genome-wide DNA copy number alterations in eight cell line samples and one bulk tumor sample. We found genomic amplification of the 12q13.3-14.2 locus, which contains the oncogenes *GLI1* and *CDK4*, in 33% of the samples (Table 1). Other amplifications, using a threshold of 5-fold amplification or more, were observed in 1q32.1, which contains *PI3KC2B*

Table 1 Amplifications and homozygous deletions analyzed by Digital karyotyping

Chr band	Target oncogene	Total found	Our DK %
Amplifications			
1q32.1	PIK3C2B, MDM4	2	22
7p11.2-12.1	EGFR	1	11
8q24.21	MYC	1	11
12q13.3-q14.1	GLI1, CDK4	3	33
12q14.1	Unknown	2	22
12q15	MDM2	1	11
Number of cases (n)		9	
Homozygous deletions			
1p36.31-p36.23	TP73, LRRC47, DFFB	3	33
9p21.3-22.3	CDKN2A, CDKN2B	4	44
9q34.3	CACNA1B	4	44
Number of cases (n)		9	

Fig. 1 Genomic aberrations detected by DK. These Digital Karyotyping examples show amplifications at 12q, 8q, 1q, and a homozygous deletion at 9p in the cell line samples. The values on the y-axis indicate genome copies per haploid genome, and values on the x-axis represent positions along the indicated chromosome arm. A sliding window of 300 virtual tags showed prominent peaks at (a) 12q13.3-14.2 locus with GLI, CDK4 and 12q15 with MDM2, (b) 8q24 locus with C-MYC, (c) 1q32.1 with PI3KC2B and MDM4, and (d) 9p21 deletion with CDKN2A and CDKN2B



and MDM4, in 2 samples (22% of samples), and in 12q15, which contains MDM2, in one sample (11%). In this analysis, GLI and CDK4 emerged as the most prominent amplifications seen in GBMs, likely since EGFR amplification is lost in serum cultured adherent cell lines. EGFR amplification on 7p11.2 was found only in the lone bulk tumor sample, consistent with previous observations that adherent GBM cell lines will consistently lose EGFR amplification during adherent culturing.

Four of nine (44%) of the samples showed homozygous deletions near 9p21 (Fig. 1), which contains CDKN2A and CDKN2B (p15 and p16). Other deletions were found in 33% of samples at 1p36, which contains CAMTA1. Forty-four percent of the samples also had a deletion containing the CACNA1B calcium ion channel gene that is found at 9q34.3, adjacent to the 9q telomere. This region was not described by others or found in the other techniques (described below), so it is not clear if this deletion is significant. No other novel deletions were found. Only completely deleted regions containing neither allele were considered.

Illumina Bead arrays

We also searched for amplifications and deletions in 22 tumor samples using the 550 k SNP Illumina Bead array (Table 2). The amplifications observed more than once were EGFR (4 of 21 non-cell line samples), the 12q14 locus containing both GLI and CDK4 (4 of 22), MDM2 (3 of 22), PI3KC2B and/or MDM4 (2 of 22), and MYCN (2 of 22). A single occurrence each of two unknown loci was also observed. The major homozygous deletions observed

in 22 tumor samples were the 9p deletions containing CDKN2A and CDKN2B. Single homozygous deletions were found at the RB1 and PTEN loci. Homozygous deletions were also observed on 6q, 10p, 10q, 11q, 13q, and Xp, largely in regions with no obvious tumor suppressor. The amplifications with the Log R ratio > 2 were considered to be true amplifications and samples with Log R ratio < -1.5 were considered to be probable homozygous deletions (Fig. 2).

When the DK and Illumina Bead Array datasets were combined, the most prominent alterations were the amplifications of EGFR and of GLI/CDK4 and the homozygous deletion at 9p21. An interesting observation in the combined Illumina Bead Array and DK datasets was not the appearance of the few unknown alterations observed, but the 22.6% occurrence for amplification at the GLI/CDK4 in the 31 patients.

Comparison to published data

We compiled and compared our results with six major whole genome copy number studies on GBM. The literature was searched for studies that used high resolution techniques, such as array CGH that can locate deletions and amplifications that span regions as small as several kilobases [8–12, 15]. Five studies were array CGH-based, using an arrayed normal genome representation of BACs or oligonucleotides as the arrayed substrate for hybridization with tumor DNA. One study was Illumina Bead Array based [15], similar to part of our data. The amplification and homozygous deletion rates for these studies were extracted and compiled in Tables 2 and 3.

Table 2 Genomic amplifications found in 456 cases of glioblastoma: combined totals from digital karyotyping, bead arrays, and published reports

Chr band	Target Oncogene	Illumina	Parson et al.	Ruano et al.	Maher et al.	Roerig et al.	Misra et al.	Korshunov et al.	TCGA	Total found	Our %	Combined %
1p12-13.3	NRAS				1			1		2		0.4
1p22.2-36.3	Unknown	1	1		3		1	1		7	4.5	1.5
1q21-q44	PIK3C2B,MDM4?	2	4	1	2	1	1	2	22	35	9.1	7.7
2p22.2-25.1	MYCN	2	3		1					6	9.1	1.3
2q31-36.3	Unknown		6		3			1		10		2.2
3p13-26.3	CTNNB1 (beta-catenin)?				2			2		4		0.9
3q21-q29	Unknown		2		2		1	2		7		1.5
4p12-16.1	Unknown		2	1	1	1				5		1.1
4q11-35.2	PDGFRA		3	1	2	4	1	2	22	35		7.7
5p16.3	Unknown		1							1		0.2
5q13-35.2	MAT2B		2							2		0.4
6p12-21.1	Unknown		1							1		0.2
6q16.3-27	MYB		2					2		4		0.9
7p11.2-22.1	EGFR	4	6	10	14	6	11	24	88	163	18.2	35.7
7q11.2-22.1	Unknown				4			1		5		1.1
7q21.1-22	CDK6			1			2	3	3	9		2.0
7q11.2-36.3	MET		1		7		1	2	6	17		3.7
8p21-23.3	Unknown									0		
8q11-q24.3	MYC		5		3			1		9		2.0
9p13.2-23	Unknown				2					2		0.4
9q33.2-34.2	Unknown		1		1	1				3		0.7
10p11.2-15.3	Unknown				3					3		0.7
10q11.2-12.2	Unknown		2							2		0.4
11p13-15.5	Unknown		1							1		0.2
11q13.1-24	Unknown		2		1			1		4		0.9
12p11-13.3	CCND2				5				5	10		2.2
12q13-14.1	GLI, CDK4	4	4	3	9	3	2	7	29	61	18.2	13.4
12q14.2-24	MDM2	3	1	3	8	2	2	1	22	42	13.6	9.2
13q31-34	TNFSF13B or COL4A2?		1	3	2	2		2		10		2.2
14q32.3	AKT1								4	4		0.9
14q tel	Unknown							2		2		0.4
15q13-26.2	Unknown				3			1		4		0.9
16p13.12	Unknown		1							1		0.2
16q24.1	Unknown		1		2					3		0.7
17q11.2-q25.3	Unknown		2		2					4		0.9

Table 2 continued

Chr band	Target Oncogene	Illumina	Parson et al.	Ruano et al.	Maher et al.	Roerig et al.	Misra et al.	Korshunov et al.	TCGA	Total found	Our %	Combined %
18p11-11.3	Unknown				1	3				4		0.9
18q21.2-23	Unknown		2		2			2		6		1.3
19p11.3-13.3	Unknown		1		3					4		0.9
19p tel	Unknown							2		2		0.4
19q12-13.4	Unknown	1	2		2					5	4.5	1.1
20p11-13	Unknown	22	22	20	37	29	50	70	206	456		0.4
Number of cases (n)												

We then compared our data with the compiled published data to determine the most frequent alterations in a larger population. The last column in Tables 2 and 3 show the combined rates. A total of 456 cases presented data suitable for comparison of amplifications, but only 366 cases could be compiled for homozygous deletions due to the format of the data presentation of some of the studies. Not surprisingly, EGFR was the most common amplification, occurring in 35.7% cases. GLI/CDK4 co-amplification occurs in 13.4% of the cases, while MDM2 amplification occurs in 9.2%, PI3KC2B/MDM4 occurs in 7.7%, MET occurs in 3.7% and PDGFRA occurs in 7.7% cases. The regions of 12p11-13.3, containing CCDN2, and 13q31-34 both had 2.2% amplifications: interestingly, none of the genes in these regions were previously implicated in GBM. There were other regions which had amplifications, but were all less than 2% incidence.

The combined CDKN2A/CDKN2B deletion was the most frequent deletion in the combined data, occurring in 46.4% of the cases. Deletions at PTEN and RB1 occurred in 10.9 and 6.3% of cases, respectively. Again, there were regions of deletion in common with several studies where no tumor suppressor for GBMs has been described with nearly 5% incidence which might play a role in GBM occurrence.

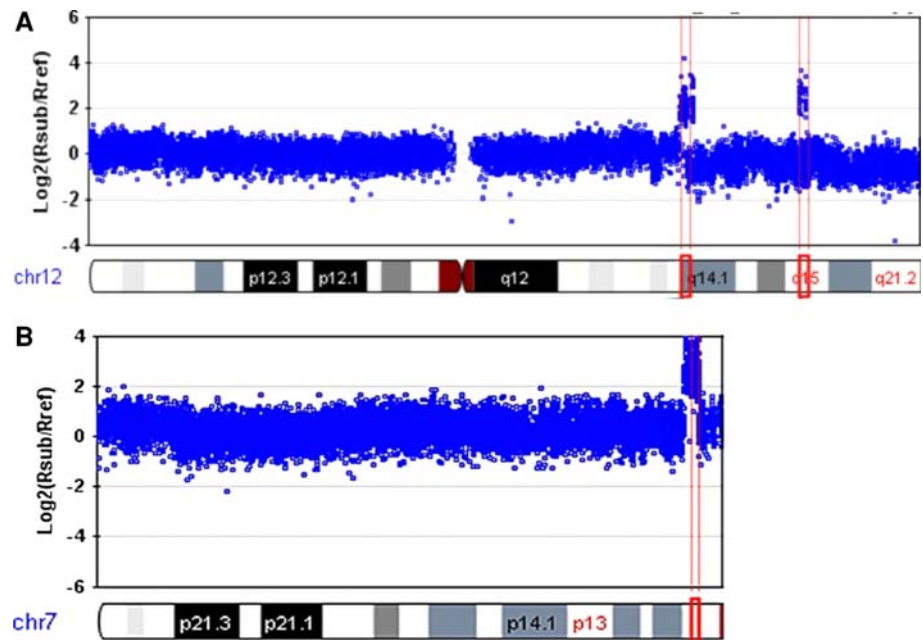
Statistical comparison of studies

One of the studies appeared to have a consistently higher rate of amplifications and deletions when compared to the other studies. A statistical analysis was performed to check if these rates were statistical outliers. A Chi-Square test indicated that Maher and coworkers data [12] significantly differed from the others. Each of the other study was compared to the rest of the studies by removing one at a time. A Fisher's exact study also indicated that the study by Maher and coworkers [12] was different when compared to the rest of the studies. None of the other studies, including our own data, exhibited statistically different rates of genomic alterations from the average. Further analysis suggested that the deviation of the Maher study was due to the threshold values used for calling amplifications and deletions; this study had used cutoff rates of >0.4 for amplifications and less than -0.4 for homozygous deletions for the Mean Log Ratios, whereas the other studies had taken Mean Log Ratios with cutoff rates ranging from >2 to >3 for amplifications and less than -1 to -2 for homozygous deletions.

Copy number alterations viewed in a pathway context

Figure 3 shows the major amplifications and homozygous deletions described in this study, based on the combined

Fig. 2 Illumina Bead arrays. Examples of common amplifications found using 550 K SNP Illumina Bead Arrays. The Log R Ratio > 2 was set as the threshold value for amplification and less than -1.5 as the threshold value for homozygous deletion. **a** A sample with significant amplifications at 12q14.1–12q15 with an amplification at CDK4 and MDM2 and a Log R Ratio ≥ 4 , and **b** A sample with a prominent EGFR amplification at 7p11.2 with a Log R Ratio > 3



data. The pathway interactions shown in the figure are combined and edited based on pathways displayed by Biocarta (<http://www.biocarta.com/>). We confined the presentation to those portions of the pathways with alterations in GBM. Figure 3 indicates that most of the alterations cluster into the well-described cellular functions of growth factor receptor signaling and cell cycle control. Many of the same genes shown here are also those that have point mutations at a much higher frequency than the amplifications/deletions shown here (e.g., PTEN, PIK3CA, and p53). The activation of these pathways is likely a combination of protein altering mutations, promoter mutations, and copy number changes, as well as epigenetic mechanisms.

Discussion

An in-depth analysis of a large number of glioblastoma multiforme (GBM) samples will provide the most accurate information on the amplifications and deletions present in GBMs. We have combined the results of our study with five array CGH studies [8–12] and TCGA data to determine the rate of gene amplifications and deletions in a larger patient population. The data assembled here has several possible uses. First, since the incidence of GBM amplifications and deletions calculated here are averaged over several study populations and different techniques, this composite data might be a more accurate prediction of true rates. Second, those consistently amplified or deleted loci found in the pooled data should be more likely to have an oncogene or tumor suppressor gene of pathological

significance. The data here may be helpful to point to those loci worthy of further investigation, if no cancer related gene is already known at that locus. Finally, the copy number alterations here can also be used to augment other genomic data such as large-scale sequencing data. The combination of copy number alterations, mutations, functional data and possibly epigenetic changes will help provide a more complete picture of the alterations that drive GBM formation.

The amplifications and deletions found in this study and the cited studies cluster into well-known oncogenic pathways, summarized in Fig. 3. The pathways implicated by copy number changes are also implicated by gene mutation data for GBM, such as mutations in p53 or PIK3CA [16]. It is clear from the combined data that growth signaling from the receptor tyrosine kinases is a significant characteristic of this malignancy. Likewise, inactivation of the p53 check point is critical, as with the majority of other cancers. In addition to p53 mutations, amplification of either MDM2 or MDM4 appears to be a common mechanism that may have implications for therapeutic strategies. Cell cycle control in GBM is also frequently driven by CDKN2B/CDKN2A (p15/p16) deletion or CDK4 amplification. The frequency at which certain amplicons were represented in the combined data is not obvious from individual studies; a good example is the *GLI/CDK4* amplification in 13% of GBMs, which suggests that certain potential therapeutic targets may be relatively understudied.

The most studied target in GBM is EGFR and in the present study EGFR represents an amplification rate of approximately 35.7%, which is less than the figure of 50% of GBMs exhibiting EGFR amplification frequently

Table 3 Homozygous deletions found in 366 cases of glioblastoma: combined totals from digital karyotyping, bead arrays, and published reports

Chr band	Target Oncogene	Illumina	Parson et al.	Maher et al.	Roerig et al.	Misra et al.	TCGA	Combined total	Our %	Combined %
1p13.3-36.3	CAMTA1, CDKN2C		3	5	1	1	7	17		4.6
1q21.1-44	Unknown	1	3	4				8	4.5	2.2
2q11.1-37.3	Unknown		3	2				5		1.4
2p16.3-22.3	Unknown		2					2		0.5
3p12.3-24.3	Unknown		1	3				4		1.1
3q12.3-29	Unknown		2	1		1		4		1.1
4p12-16	Unknown		1	2				3		0.8
4q11-34.1	Unknown		6	4		1		11		3.0
5p12-15.33	Unknown		1	3		1		5		1.4
5q13.1-35.3	Unknown		4	4	1	1		10		2.7
6p12.2-25.3	Unknown		5	2		1		8		2.2
6q13.1-27	PARK2	2	1	10		1		14	9.1	3.8
7p12-15.2	Unknown			1	3			4		1.1
7q21.1-35	Unknown		1					1		0.3
8p12-23.2	Unknown		1	3		1		5		1.4
8q11.2-24.2	Unknown		3	1				4		1.1
8q tel	Unknown							2		0.5
9p11.1-24.3	CDKN2B, CDKN2A	5	13	17	7	14	114	170	22.7	46.4
9q13-34.3	Unknown		3	4				7		1.9
10p11-15.3	Unknown	2		15				17	9.1	4.6
10q11.2-q26.3	PTEN	1	4	18		1	16	40	4.5	10.9
11p11.2-15.4	Unknown		1	6	1	1		9		2.5
11q13.4-24	Unknown	3		7		1		11	13.6	3.0
12p12.1-13.2	Unknown			2				2		0.5
12q12.3-24.3	Unknown		3	6				9		2.5
13q12.11-13.2	Unknown	1		7				8	4.5	2.2
13q13.2-33.1	SPRY2, RBI	1	3	10	1		8	23	4.5	6.3
14q11.2-13	Unknown			10				10		2.7
14q23.1-32.3	Unknown		2	6		1		9		2.5
15q11.2-15.3	Unknown		1	6				7		1.9
15q15.3-26.3	Unknown		3	4				7		1.9
16p12.3-13.3	Unknown		1					1		0.3
16q11.2-23.2	Unknown		1	2				3		0.8
17p11.2-13.2	TP53			5			2	7		1.9

Table 3 continued

Chr band	Target Oncogene	Illumina	Parson et al.	Maher et al.	Roerig et al.	Misra et al.	TCGA	Combined total	Our %	Combined %
17q11.2-24.3	NF1		1	3			4	8		2.2
18p11-11.3	Unknown			7				7		1.9
18q11.1-22	Unknown		1	3		2		6		1.6
19p12-13.1	Unknown									
19q13.12-13.3	Unknown			4				4		1.1
20p11.2-13	Unknown			2				2		0.5
20q13.2-13.3	Unknown		1					1		0.3
21p11-13	Unknown			3				3		0.8
21q11.2-22.3	Unknown		2	2				4		1.1
22p11.2	Unknown									
22q11.2-13.3	Unknown		3	4	1	2		10		2.7
Number of cases		22	22	37	29	50	206	366		

quoted. Although EGFR is critically important for the development of many GBMs, parallel signaling through MET has been implicated as a means for lung cancers and possibly GBMs to circumvent EGFR targeting therapy [17, 18]. Therefore, simply selecting the most commonly amplified gene or mutated gene as a therapeutic target may not be sufficient without a deeper understanding of the pathways in which they function and the possible mechanisms by which targeted treatment resistance can occur.

The differences in the copy number rates observed between the various studies may represent real differences between patient populations, non-statistically significant differences due to chance, bias due to technique, bias due to different copy number change definitions, or a combination of any/all of these. Although we combined our data from DK and Illumina Bead Array experiments with the results from five other studies and with the recently available TCGA data to survey over 400 cases of glioblastoma, we only analyzed those studies with greater than 20 cases and those with a data presentation format that allowed us to combine datasets. The six studies cited for this study [8–12, 15] were all array based studies. The main difference in these comparative genomic hybridization studies is in the selection of the probe library for the array: one used oligonucleotides [12]; two used BAC [9, 10]; two used cDNA libraries [8, 11] and one used Illumina SNP probes [15]. When comparing each of the studies to the others, one slight outlier was the oligonucleotide CGH study [12], which reported higher rates of amplifications and deletions than the average of all six studies. It appears that this difference may be due to a lower threshold in this study for calling a copy number change. The threshold for calling a biologically meaningful copy number change is in part arbitrary, and it is not clear what threshold reflects best a true physiological impact on gene function. Additionally, this threshold may vary from gene to gene. Therefore, we presented individual results as well as combined totals. It is hoped that the combined percentages will be more accurate than any individual study by averaging out differences or bias due to patient selection bias, threshold differences, experimental technique or data analysis. We suspect based on the trends observed in this study, but cannot prove, that differences in setting thresholds for amplification and deletion is the most significant contribution for the differences in alteration rates observed between studies.

In our data, we initially used DK, with genomic tags generated by Sanger Sequencing. The sensitivity of DK for detecting alterations is directly proportional to the number of tags sequenced. For our data, this was about 200,000 tags per library, which gave us a reasonable genomic resolution with a very quantitative assessment of the amplification size. However, we found that the Illumina data was more likely to detect amplifications or deletions

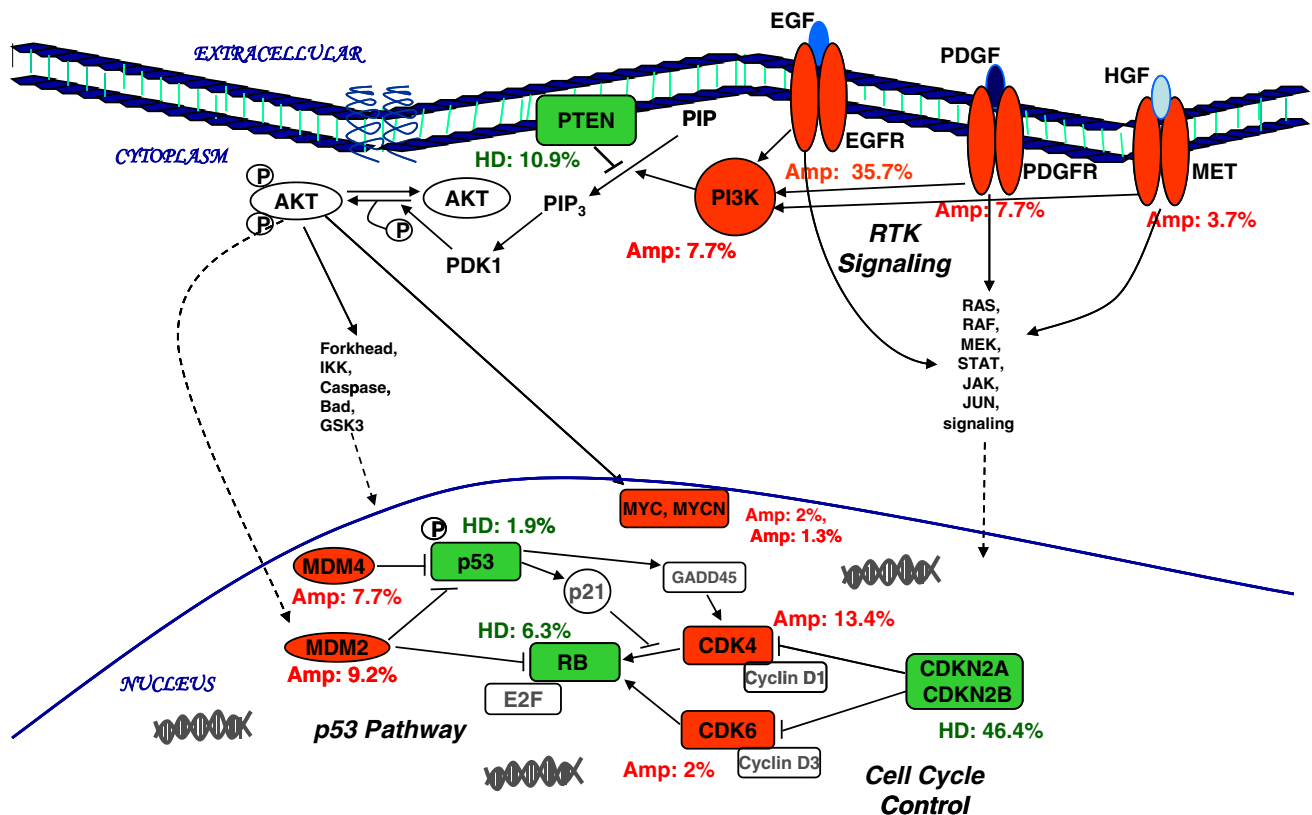


Fig. 3 Copy number alterations viewed in a pathway context. The major cell pathways related to glioblastoma were drawn in a condensed format and labeled with the combined rates of amplifications and deletions shown in Tables 1 and 2. Amplified oncogenes are

in red and deleted tumor suppressors are in green. Most all known amplified or deleted genes in glioblastoma cluster into growth factor signal transduction or cell cycle control all cluster into the pathways shown here

that were narrower in genomic distance. As next generation sequencing is applied to DK and DK-like methods it is expected that a high resolution and economical method of assessing copy number will be available.

There were very few instances of amplifications in regions where there was not a well described oncogene, suggesting that most amplified oncogenes have been identified in GBM. There are several altered loci observed in this data without an obvious oncogene or tumor suppressor, suggesting regions where research attention should be focused. A low frequency of loss or gains for GBM in general is motivation for examining these loci for possible new cancer-related genes. We believe that, more practically, combining this data with other genomic data will be a more efficient strategy for locating new cancer-related genes. One approach might be to combine copy number changes with new mutation screen data for these loci, with the hopes of finding a combined alteration rate greater than the few percent observed based on just copy number changes.

Molecular biology technology has advanced to the point where comprehensive mutational surveys can be undertaken

for human cancer genomes. Due to the relative ease at collecting copy number change data to detect amplifications and deletions, there are numerous individual copy number studies for GBM and other cancers. However, there is no efficient way to access compiled copy number data from the various studies. It is hoped that this study has provided more accurate rates of amplifications and homozygous deletions by generating new copy number data from different techniques and combining this data with compiled published results. This might provide motivation to compile and utilize genomic amplification and deletion data for other cancers. If done properly, average rates are likely to be better estimates of true rates, and lead to a more complete picture of the changes that drive tumor formation and growth. This study may also provide a means where an average rate of amplification or deletion in GBM for a particular gene might be cited more accurately based on a larger number of observations.

Acknowledgments This research was supported by NIH Grant NS052507 and the Virginia and D. K. Ludwig Fund for Cancer Research. G. J. R. is the Irving J. Sherman M.D. Professor of Neurosurgery Research.

References

- Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20(2):207–211
- Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 23(1):41–46
- Wang TL, Maierhofer C, Speicher MR, Lengauer C, Vogelstein B, Kinzler KW, Velculescu VE (2002) Digital karyotyping. *Proc Natl Acad Sci USA* 99(25):16156–16161
- Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E, Bierle J, Doucet D, Milewski M, Yang R, Siegmund C, Haas J, Zhou L, Oliphant A, Fan JB, Barnard S, Chee MS (2004) Decoding randomly ordered DNA arrays. *Genome Res* 14(5):870–877
- Mei R, Galipeau PC, Prass C, Berno A, Ghandour G, Patil N, Wolff RK, Chee MS, Reid BJ, Lockhart DJ (2000) Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res* 10(8):1126–1137
- Kleihues P, Burger PC, Plate KH, Ohgaki H, Cavenee WK (2000) Pathology and genetics: tumors of the nervous system. International Agency for Research on Cancer, Lyon
- Burton EC, Lamborn KR, Feuerstein BG, Prados M, Scott J, Forsyth P, Passe S, Jenkins RB, Aldape KD (2002) Genetic aberrations defined by comparative genomic hybridization distinguish long-term from typical survivors of glioblastoma. *Cancer Res* 62(21):6205–6210
- Korshunov A, Sycheva R, Golanov A (2006) Genetically distinct and clinically relevant subtypes of glioblastoma defined by array-based comparative genomic hybridization (array-CGH). *Acta Neuropathol (Berl)* 111(5):465–474
- Misra A, Pellarin M, Nigro J, Smirnov I, Moore D, Lamborn KR, Pinkel D, Albertson DG, Feuerstein BG (2005) Array comparative genomic hybridization identifies genetic subgroups in grade 4 human astrocytoma. *Clin Cancer Res* 11(8):2907–2918
- Roerig P, Nessling M, Radlwimmer B, Joos S, Wrobel G, Schwaenen C, Reifenberger G, Lichter P (2005) Molecular classification of human gliomas using matrix-based comparative genomic hybridization. *Int J Cancer* 117(1):95–103
- Ruano Y, Mollejo M, Ribalta T, Fiano C, Camacho FI, Gomez E, de Lope AR, Hernandez-Moneo JL, Martinez P, Melendez B (2006) Identification of novel candidate target genes in amplicons of Glioblastoma multiforme tumors detected by expression and CGH microarray profiling. *Mol Cancer* 5:39
- Maher EA, Brennan C, Wen PY, Durso L, Ligon KL, Richardson A, Khatry D, Feng B, Sinha R, Louis DN, Quackenbush J, Black PM, Chin L, DePinho RA (2006) Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res* 66(23):11502–11513
- Kotliarov Y, Steed ME, Christopher N, Walling J, Su Q, Center A, Heiss J, Rosenblum M, Mikkelsen T, Zenklusen JC, Fine HA (2006) High-resolution global genomic survey of 178 gliomas reveals novel regions of copy number alteration and allelic imbalances. *Cancer Res* 66(19):9428–9436
- Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE (2002) Using the transcriptome to annotate the genome. *Nat Biotechnol* 20(5):508–512
- Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321(5897):1807–1812
- Gallia GL, Rand V, Siu IM, Eberhart CG, James CD, Marie SK, Oba-Shinjo SM, Carlotti CG, Caballero OL, Simpson AJ, Brock MV, Massion PP, Carson BS, Sr, Riggins GJ (2006) PIK3CA gene mutations in pediatric and adult glioblastoma multiforme. *Mol Cancer Res* 4(10):709–714
- Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, Kosaka T, Holmes AJ, Rogers AM, Cappuzzo F, Mok T, Lee C, Johnson BE, Cantley LC, Janne PA (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* 316(5827):1039–1043
- Huang PH, Cavenee WK, Furnari FB, White FM (2007) Uncovering therapeutic targets for glioblastoma: a systems biology approach. *Cell Cycle* 6(22):2750–2754