Title page

Comparative analyses of gene copy number and mRNA expression in GBM tumors and GBM xenografts

J. Graeme Hodgson [#] *, Ru-Fang Yeh [#], Amrita Ray, Nicholas J Wang, Ivan Smirnov, Mamie Yu, Sujatmi Hariono, Joachim Silber, Heidi S. Feiler, Joe W. Gray, Paul T. Spellman, Scott R. Vandenberg, Mitchel S. Berger, C. David James

Departments of Neurological Surgery (JGH, IS, MY, SH, JS, SRV, MSB, CDJ), Pathology (SRV), and Epidemiology and Biostatistics (RY) University of California, San Francisco, CA 94143. Life Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA 94720 (AR, NJW, HSF, JWG, PTS)

[#] These authors contributed equally to this work

Running title: Genomic analyses of a GBM xenograft tumor panel

* Corresponding Author

J. Graeme Hodgson

Dept. Neurological Surgery

UC San Francisco

San Francisco, CA

94143-0808

Phone: 415-476-3630

Fax: 415-476-8218

e-mail: ghodgson@cc.ucsf.edu

Abstract

Development of model systems that recapitulate the molecular heterogeneity observed amongst GBM tumors will expedite the testing of targeted molecular therapeutic strategies for GBM treatment. In this study, we profiled DNA copy number and mRNA expression in 21 independent GBM tumor lines maintained as subcutaneous xenografts (GBMX), and compared GBMX molecular signatures to those observed in GBM clinical specimens derived from The Cancer Genome Atlas (TCGA). The predominant copy number signature in both tumor groups was defined by chromosome-7-gain/chromosome-10-loss, a poor prognosis genetic signature. We also observed, at frequencies similar to that detected in TCGA GBMs genomic amplification and overexpression of known GBM oncogenes such as EGFR, MDM2, CDK6 and MYCN, and novel genes including NUP107, SLC35E3, MMP1, MMP13 and DDX1. The transcriptional signature of GBMX tumors, which was stable over multiple subcutaneous passages, was defined by overexpression of genes involved in M-phase, DNA Replication, and Chromosome organization (MRC) and was highly similar to the poor-prognosis mitosis-and-cell-cycle-module (MCM) in GBM. Assessment of gene expression in TCGA-derived GBMs revealed overexpression of MRC cancer genes AURKB, BIRC5, CCNB1, CCNB2, CDC2, CDK2, and FOXM1, which form a transcriptional network important for G2/M- progression and/or -checkpoint activation. In conclusion, our study supports propagation of GBM tumors as subcutaneous xenografts as a useful approach for sustaining key molecular characteristics of patient tumors, and highlights therapeutic opportunities conferred by this GBMX tumor panel for testing targeted therapeutic strategies for GBM treatment.

Keywords: GBM, xenograft, comparative genomics

Introduction

Glioblastoma multiforme (GBM; WHO grade IV) is the most common type of CNS tumor, the prognosis for which remains dismal despite intervention with surgery, radiation, and chemotherapy [1]. A large number of genetic and epigenetic alterations have been identified in GBMs [2, 3], many of which enhance the ability of tumor cells to proliferate, invade surrounding brain tissue, and evade therapeutic treatments. Whereas the mRNA and protein products of these genes are attractive candidates for targeted therapeutics, realization of the potential of targeted therapeutics for improved treatment of GBM will require more extensive understanding of the molecular pathways that underlie tumorigenesis and preclinical models that closely recapitulate the human disease.

Ex vivo cell culture models of GBM have provided valuable insights into the mechanisms by which oncogene and tumor suppressor dysfunction promote GBM development. However, it is well known that GBM cell lines cultured ex vivo lack amplification- and associated over-expression- of EGFR, which occur in 40-50% of primary tumors. In addition, tumors that develop in rodents following intracranial implantation of cultured GBM cells often lack key phenotypes observed in patient tumors such as angiogenesis and infiltrative growth. In contrast, GBM tumors maintained as subcutaneous xenografts in nude mice demonstrate maintenance of EGFR amplification through serial in vivo propagation [4], and additionally recapitulate the invasive growth pattern of patient tumors when transplanted intracranially in rodents [5]. A GBM-xenograft (GBMX) tumor panel has enabled studies aimed at directly assessing the effect of EGFR amplification on GBM radiation response [6], and correlating tumor PTEN and EGFR status with response to the EGFR kinase inhibitor Erlotinib [7]. While these studies suggest that subcutaneously-propagated GBMX tumors more accurately model GBM molecular biology and therapeutic responses than permanent cell lines, it remains unclear the extent to which this GBMX tumor panel represents the molecular subtypes of patient GBMs.

In this study we assessed DNA copy number aberrations and mRNA transcript levels in a 21 member GBMX tumor panel, and compared these molecular datasets with datasets derived from GBM clinical specimens. This comparative genomic approach enabled the identification of a number of aberrantly- overexpressed transcripts in GBM- and GBMX-tumors for which targeted inhibition may prove efficacious for disease treatment.

Materials and Methods

Samples

Subcutaneous GBM xenograft tumors were surgically removed in accordance with IACUC approved procedures and snap frozen in liquid nitrogen. Snap frozen non-neoplastic brain tissues were derived from the temporal lobes of epileptic patient surgeries and were comprised primarily of cortex with mild to moderate reactive astrocytosis and neurons. Non-neoplastic controls were obtained from the BTRC tissue core at UCSF in accordance with CHR approved procedures. All samples were ground to a powder using a liquid nitrogen-cooled pestle and mortar, and DNA and RNA were extracted from separate aliquots of ground tissue.

DNA copy number analyses

DNA extractions [8] and hybridizations to Affymetrix 50K Xba SNP chip arrays [9] were performed as described. SNP array data was pre-processed as follows: PM probe intensities of the 21 GBMX tumors were quantile normalized with those of 90 normal tissue controls (Hapmap trios, Affymetrix). The total hybridization intensities, PMA+PMB (in logarithm base two), were median-summarized over the 5-7 probe quartets for each SNP, followed by a fragment length adjustment using cubic splines [10]. We then calculated the log2 copy number (CN) ratio for each SNP by subtracting the mean SNP log2 intensity of all 90 Hapmap reference samples from the per-SNP intensity within each GBMX tumor to remove SNP-

specific effects. CN was segmented along each chromosome into regions of equal copy number changes with a circular binary segmentation algorithm [11] implemented in the DNAcopy package of R/Bioconductor [12], using the NCBI Build 36.1 annotation from Affymetrix (dated July 12, 2007). Neighboring genomic segments were merged if their estimated copy numbers did not differ by more than one standard deviation. Frequencies of CN gain or loss were calculated using segment mean thresholds of +/- 0.3. For analyses of GBM clinical specimens segment mean thresholds of +/- 0.3 were used on data generated from SNP arrays [13], and thresholds of +/- 0.1 were used on data generated from BAC arrays [14, 15], and Agilent 244K oligonucleotide arrays [2].

mRNA expression analyses

Total RNA was extracted from GBMX tumors and non-neoplastic control brain using the mirVana RNA isolation system (Ambion), further purified using RNeasy columns (Qiagen), and RNA integrity assessed using a bioanalyzer (Aglient). RNA from all samples was hybridized in parallel to Human U133A GeneChip™ arrays on the Affymetrix HTA system (HT_HG-U133A). CEL files were read into R/Bioconductor using the Affy/affyPLM package [16], and RMA (robust multi-array average) intensity in log2 scale was generated for each probe set (gene). The 11 perfect match (PM) intensities per probe set were (i) background corrected; (ii) quantile-normalized (to make the distribution of intensities the same for all arrays); and (iii) summarized for each probe set using a robust fit of linear models as described [17].

Unsupervised hierarchical clustering based on the most variably expressed genes, defined by the medium absolute deviation (MAD), was conducted using the Pearson's correlation coefficient or Euclidean distance as the similarity metrics, and the Ward's linkage method or the complete linkage method as the between-cluster distance metrics. Separate analyses were conducted for the top N (N=50,200,500) variably expressed genes amongst all samples (non-neoplastic control brain and GBMX tumors), and amongst GBMX tumors only. Supervised clustering analyses based on the proneural-mesenchymal-proliferative

GBM gene classes [15] were conducted using 24 U133A genes from the 35-signature-gene set, and 478 U133A genes from the 725-survival-associated all-marker gene set.

For comparisons of gene expression between GBMX tumors and non-neoplastic controls, paired t-tests were performed on the average log2 intensity of each probe set in GBMX tumors and non-neoplastic controls. P-values were adjusted for multiple comparisons using Bonferroni-corrected p-values of the moderated t-statistics. GO stat analysis [18] was conducted using 607 genes that were significantly up-regulated at least 2-fold in GBMX tumors. This gene list was searched against the AFFY_HG_U133A GO gene-association database, the maximal p-value in GO output list was set to 1e-10, and the minimal length of considered GO paths was set to 5. GOs were merged if the indicating gene lists were inclusions or differed by less than 10 genes.

Assessment of gene expression from The Cancer Genome Atlas [2] was conducted on GBM tumors (n=201) and 100% non-tumor controls (n=5) analyzed with the Affymetrix exon array 1.0 platform. Raw data was pre-processed with RMA [17] and aroma.affymetrix [19]. The average and maximum fold changes in the GBMs were calculated relative to the median log2 expression value of the non-tumor samples.

Results

Molecular sub-classification of GBM xenografts

Prior global assessments of DNA copy number and mRNA expression suggest the presence of distinct molecular subsets of GBM [14, 15, 20-22]. To assess molecular subclass representation among GBMX tumors, we examined DNA copy number aberrations (CNAs) and mRNA expression profiles in 21 distinct xenografts (Supplementary table 1). For CNA assessment, we used the Affymetrix 50K Xba SNP chip, which enabled identification of the expected EGFR-amplifications [4] and CDKN2A-homozygous deletions (CD James, personal communication; supplementary figure 1). We next compared the frequencies of genomic CNA between GBMX tumors and series' of de novo GBM specimens [2, 13, 14]. This comparison revealed that the patterns of recurrent CNA were highly similar between both tumor groups (Figure 1); the most frequently observed CNAs were whole chromosome 7 gains, whole chromosome 10 losses, CDKN2A homozygous deletions, and EGFR amplifications.

To assess transcriptional heterogeneity in the GBMX tumor panel, we next determined mRNA expression profiles of GBMX tumors and non-neoplastic control brain tissues using Human U133A GeneChip™ arrays on the Affymetrix HTA system (HT_HG-U133A). As expected, unsupervised hierarchical clustering of the most variably expressed genes amongst all samples segregated tumors and non-neoplastic controls into distinct classes (Figure 2A). Also, GBMX tumors from within a tumor line, but from distinct tumor passages, showed a higher extent of identity than when compared to any other tumor line (Figure 2A), suggesting that GBMX transcriptional signatures are stable in association with subcutaneous propagation. However, unlike expression profiling studies of patient tumors [15, 21] GBMX tumors did not reliably segregate into 2 or 3 distinct subclasses.

We next performed a supervised classification of the GBMX tumors with respect to the proneural-mesenchymal-proliferative signature gene classification scheme of high-grade

astrocytomas [15]. This revealed that all GBMX tumors invariably contained a strong proliferative expression signature, whereas no GBMX tumors contained evidence of a proneural signature (Figure 2B). Expression of the mesenchymal signature was variable across GBMX tumor lines (Figure 2B), consistent with observations in primary tumors [15].

Cell cycle gene expression networks in GBMX tumors

To further investigate the proliferative signature in GBMX tumors, we compared the average expression of all HT_HG-U133A probe sets between GBMX tumors and non-neoplastic control brain samples. This analysis revealed 809 probe sets (607 unique genes) that were significantly ($P<0.01$, adjusted for multiple comparisons) over expressed at least 2-fold on average in GBMX tumors. To determine the biological processes associated with these 607 genes, we assessed their Gene-Ontology (GO) classifications using GOstat [18]. This analysis revealed 4 main GO clusters that were significantly overrepresented in this gene list (Supplementary table 2), comprising highly significant enrichment for genes associated with mitosis ($P = 0$), DNA replication ($P = 1.3x10^{-51}$), RNA splicing ($P = 7.9x10^{-28}$), and chromosome organization and biogenesis ($P = 9.6x10^{-21}$). Interestingly, these GO biological processes closely resembled the mitosis and cell cycle module (MCM) gene expression signature previously identified in GBM and breast cancer [23]. Indeed, analyses of MCM hub genes (genes that show high intermolecular physical- and/or functional-interactions) revealed that 27/35 (77%) of the hub genes were amongst the most highly overexpressed genes (99^{th} percentile) in GBMX tumors relative to non-neoplastic controls; each of the 27 hub genes were overexpressed >8-fold on average in GBMX tumors relative to non-neoplastic controls. Therefore, the predominant gene expression signatures observed in GBMX tumors significantly overlap with signatures observed in human GBM clinical specimens.

To determine which of the Mitosis-, DNA-Replication-, Chromosome-organization (MRC) genes overexpressed in GBMX tumors (n=389) were also overexpressed in GBM clinical samples, we analyzed expression data from 201 GBM clinical specimens derived

from The Cancer Genome Atlas (TCGA). This revealed that 41/389 (11%) of these genes were over-expressed at least 2-fold on average in GBM clinical specimens, as well as in GBMX tumors, relative to non-tumor controls (Table 1). Characterization of these genes using Ingenuity Pathways Analysis (IPA) revealed 2 main cell cycle expression networks: 1) Cellular assembly and organization (Figure 3A), and 2) DNA-replication, -recombination and -repair (Figure 3B). Within the cellular assembly and organization network Aurora Kinase B (AURKB), Cyclin B1 (CCNB1), Cyclin-dependent kinase 1 (CDC2), Cyclin-dependent kinase 2 (CDK2), and FOXM1 were the principal hub genes, showing the highest degree of intranetwork connectivity. These genes have been implicated in the development of multiple malignancies, including GBM, and play important roles in ensuring appropriate progression through mitosis. Within the DNA-replication, recombination, and repair network the main hub gene was the tumor suppressor TP53, suggesting that loss of TP53 function in GBM results in transcriptional upregulation of a gene expression network important for transition through S-phase of the cell cycle.

Expression of genomically-amplified genes in GBMX tumors

As expected from previous FISH studies [4], our microarray analyses revealed high-level EGFR amplification (log2ratio > 4; 32 copies) in a significant proportion (8/21) of GBMX tumors. This frequency is very consistent with frequencies reported in association with the analysis of large series' of patient tumors (Libermann et al., 1985; Wong et al.; 1987; Ekstrand et al., 1991), suggesting that there is no selection bias for establishing xenografts based on patient tumor EGFR amplification status. Amongst GBMX tumors, EGFR-transcript levels were highly correlated with genomic amplification (Figure 4A). Because of this, we identified all genomic loci for which the segment mean log2 copy number ratio was greater than 4 in at least 1 GBMX tumor line to identify additional amplification-copy number relationships. We identified 15 such amplicons (including the EGFR amplicon) which ranged in size from 215kb to 3.0Mb (Table 2). Within these amplicons, we identified a total of 58 RefSeq genes, 51 of which contained probe sets on the U133A gene-expression arrays

(supplementary table 3). Of these, 33 were selectively overexpressed >5-fold in the amplicon-bearing tumor relative to the average expression in non-amplified tumors (supplementary table 3).

We next assessed whether DNA copy number and mRNA expression were increased for the 33 amplified-and-overexpressed GBMX genes in 228 human GBMs from The Cancer Genome Atlas (TCGA) [2]. We first compared the amplicons observed in GBMX tumors with the amplicons detected in TCGA GBM copy number datasets; a locus was considered amplified in GBMs if it was observed on at least 1 of 3 microarray platforms and with at least 1 of 3 analytical algorithms [2]. This reveled that 7/15 of the GBMX amplicons were focally amplified in TCGA GBMs (Table 2). Further, a majority of the amplified-and-overexpressed GBMX genes (19/33) showed evidence for high overexpression (4-to-321 fold) in a subset of TCGA GBMs (supplementary table 3). Of particular interest, in the 12p13 amplicon expression of MDM2, SLC35E3, and NUP107 clearly separated GBMs into 2 distinct expression groups (Figure 4B). In the 11q22 amplicon, the collagenases MMP1 and MMP13 were transcriptionally overexpressed at least 10-fold in ~5% of GBMs relative to the median expression in all tumors (Figure 4C); MMP13 was overexpressed over 100-fold in multiple GBMs. Finally, in the 2p24 amplicon, both DDX1 and MYCN were clearly overexpressed in a small proportion (1%, 2/228) of GBMs (Figure 4D).

Discussion

Molecular sub-classification of GBMX tumors

Results from the microarray analyses conducted in this study revealed that GBMX tumors recapitulate many of the key molecular features described in GBM clinical samples. For example, DNA copy number aberrations in GBMX tumors showed significant similarity with results from previously-published studies of patient tumors, with chromosome 7 gains, chromosome 10 losses, EGFR amplifications, and CDKN2A homozygous deletions representing the most frequent alterations in both tumor groups (Figure 1). With respect to mRNA expression, the strongest expression signature in GBMX tumors was defined by genes that promote transition through S-phase and mitosis during the cell cycle, and was highly similar to the mitosis and cell cycle module (MCM) previously described in patient GBMs [23]. However, our study also revealed differences between GBMX tumors and sets of patient GBMs. The most prominent example is the evident lack of the proneural GBM expression signature and over-representation of the proliferative expression signature [15] (Figure 2). This discrepancy suggests a selection bias in xenograft establishment (i.e., preferential successful engraftment of patient tumors with proliferative signatures), or that GBMs which successfully engraft in nude mice adopt a proliferative gene expression signature, irrespective of the classification signature of the engrafted patient tumor. With regard to this latter possibility, it is important to note that 3 stable expression subclasses of high-grade astrocytomas could be established only if the gene list for clustering was weighted to include fewer proliferative markers [15]. This gene weighting requirement argues that most, if not all, GBMs harbor a strong proliferative component that must be computationally masked to permit the unveiling of additional expression signatures (e.g. Mes, PN). Extending this line of reasoning, it is reasonable to hypothesize that it is this proliferative component/signature of GBM that emerges during subcutaneous xenograft growth. The consequence of this type of selection, as concerns the ability of xenograft

panels to recapitulate the variability of patient tumor therapeutic response, has yet to be extensively investigated, although results from preliminary reports indicate significant differences are evident between xenografts in regard to their inherent radiation sensitivity, and their response to the EGFR inhibitor erlotinib (Sarkaria et al., 2006; Sakaria et al., 2007).

The MRC expression signature in GBMX tumors

Previous molecular profiling analyses of GBM tumors have defined expression signatures comprised of genes that promote G1/S and G2/M cell cycle progression [15, 23]. We observed significant overlap of these expression signatures with the predominant expression signatures observed in GBMX tumors, defined by genes that drive M-phase, DNA Replication and Chromosome Organization (MRC). Analysis of TCGA exon array expression data revealed that a subset (11%) of MRC signature genes were overexpressed greater than 2-fold on average in GBMs compared to non-neoplastic controls (Table 1). Network analysis of the 41 MRC signature genes revealed 2 principal cell cycle networks - cellular assembly and organization (Figure 3A) and DNA-replication, -recombination and -repair (Figure 3B). The cellular assembly and organization network was primarily comprised of genes that promote mitotic progression, many of which have been implicated in cancer etiology including AURKB, BIRC5 (Survivin), CCNB1 (Cyclin B1), CCNB2 (Cyclin B2), CDC2 (CDK1), CDK2, and FOXM1. Activation of these genes induces tumorigenic phenotypes in a number of cancers, whereas their inhibition, such as has been shown for FOXM1, abrogates tumor growth and invasion [24-27], and induces genomic instability [28, 29]. Further, many of G2/M genes identified in GBMX tumors play a role in mediating the DNA damage response in cancer cells. For example, inhibition or loss of BIRC5 sensitizes GBM cells [30] and pancreatic cancer cells [31, 32] to ionizing radiation, and CHK2 mediates stabilization of FOXM1 to stimulate expression of DNA repair genes [33]. Therefore the GBMX tumor panel should enable investigations of therapeutics that specifically inhibit genes that promote G2/M cell cycle progression in the face of DNA damage (genomic instability, ionizing radiation, Temozolomide) and other cellular stresses.

DNA amplifications in GBMX tumors and primary GBMs

Our studies revealed a number of genomically-amplified, highly-overexpressed genes in GBMX tumors that are similarly amplified and/or overexpressed in GBM clinical specimens (Table 2, Figure 4A, Supplementary table 3). We identified well-established GBM oncogenes such as EGFR, MDM2, and CDK6, and identified additional amplified and/or overexpressed genes of potential biological or therapeutic interest in GBM. For example, NUP107 and SLC35E3 were co-amplified and overexpressed with MDM2 in the 12p13 amplicon in GBMX tumor line 5, and the expression patterns of each gene clearly separated TCGA GBM clinical specimens into 2 distinct groups (Figure 4B). NUP107 is a nuclear pore protein essential for kinetocore function and spindle assembly during mitosis [34, 35]. SLC35E3 (solute carrier family 35, member E3; UniProt ID Q7Z769) is a predicted multi-pass membrane protein [36] that may enable targeted delivery of therapeutic agents to SLC35E3/MDM2-amplified GBMs. In the second example, we identified a cluster of MMP genes on chromosome 11q22 that were upregulated in an 11q22-amplified GBMX tumor line (GBMX- 22). Within this amplicon, the collagenases MMP1 and MMP13 were transcriptionally upregulated 10- to 100-fold in ~5% of GBM clinical specimens (Figure 4C); both collagenases are known to promote growth and invasion of cancer cells [37-40]. In the final example, we identified co-amplification and overexpression of NMYC and DDX1 at chromosome locus 2p34 in GBMX line 28, as well as in a small proportion (1%) of TCGA GBMs (Figure 4D). The MYCN transcription factor is a well known oncogene in neuroblastoma [41] and MYCN amplifications have been previously observed in GBM [42, 43]. DDX1 is a member of the DEAD box protein family of RNA helicases that play important roles in RNA metabolism through modulation of inter- or intra-molecular RNA structures or dissociation of RNA–protein secondary structures [44]. Recent data suggest that through interaction with ATM, DDX1 plays an RNA clearance role at ionizing radiation induced DNA double strand break sites thereby facilitating template-guided repair of transcriptionally active regions of the genome [45].

Collectively, the GBMX tumor panel provides a valuable resource with which to dissect the biology of the amplicons described herein, as it is becoming clearer that multiple genes encoded within amplicons play important roles in driving tumor biology [46, 47]. This in turn may lead to the development of novel therapeutic agents and strategies for disease treatment.

Targeted molecular therapeutics and personalized medicine

We have identified number of candidate therapeutic targets in GBM, comprised of genes that are genomically amplified and/or overexpressed in clinical specimens and xenografted tumors. The GBMX tumor panel provides an important resource with which to develop and test the efficacy of targeted molecular therapeutics such as novel small molecule inhibitors and RNA interference therapeutics [48, 49] as monotherapies or in combination with DNA damaging agents such as Temozolomide and ionizing radiation. RNA interference (RNAi) utilizes small double stranded RNA-based molecules such as small interfering RNAs (siRNAs) and microRNAs (miRNAs) to inhibit gene expression in a nucleic-acid-sequence specific manner. The principal advantage of siRNAs over small molecule inhibitors and antibodies is that all genes are potential targets for inhibition; drug targeting is not limited to kinases and cell surface proteins. This dramatically expands the repertoire of candidate therapeutic targets in GBM, to include so-called 'undrugable' targets like transcription factors and oncogenes that have not been amenable to direct inhibition with small molecule inhibitors. Examples of such genes identified in this study include FOXM1, MYCN, and BIRC5.

While delivery of therapeutics to the central nervous system is particularly challenging because of the blood brain barrier, a number of promising strategies have recently been developed that may circumvent this problem. These include intranasal delivery of oligonucleotides [50], lipid encapsulation and targeted delivery of nucleic acids [51, 52], and direct administration of therapeutic agents to brain tumor tissues by convection enhanced delivery [53, 54]. Because the GBMX tumor lines described in this study form invasive GBMs

when implanted intracranially in rodents, they enable development and testing of novel strategies for targeted delivery of therapeutics to intracranial GBM xenografts in a pre-clinical setting.

Development of panels of tumor lines that closely model the molecular-heterogeneity and -biology of patient tumors will be invaluable for developing and testing personalized molecular therapeutic strategies. The GBMX tumor panel described in this study and tumor panels described in other cancers [55] constitute an important component of realizing the long-term goal of personalized medicine in cancer, wherein molecular diagnostics is closely coupled to therapeutic intervention. Future preclinical efficacy studies in the GBMX tumor panel will enable the development of predictive markers of response to a variety of inhibitory therapeutics, and may also provide insights into the mechanisms of acquired resistance to these agents.

Acknowledgements

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 (AR, JWG, PTS), by the U54 CA 112970 (AR, JWG, PTS), by the National Institutes of Health grants CA101777 (JGH), CA097257 (SRV, MSB, CDJ), NSO49720 (CDJ); and by the Center for Bioinformatics and Molecular Biostatistics grant (RY).

References

1. Chang, S. M., Butowski, N. A., Sneed, P. K. & Garner, I. V. Standard treatment and experimental targeted drug therapy for recurrent glioblastoma multiforme. Neurosurg Focus 20, E4 (2006).
2. McLendon, R. et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature (2008).
3. Parsons, D. W. et al. An integrated genomic analysis of human glioblastoma multiforme. Science 321, 1807-12 (2008).
4. Pandita, A., Aldape, K. D., Zadeh, G., Guha, A. & James, C. D. Contrasting in vivo and in vitro fates of glioblastoma cell subpopulations with amplified EGFR. Genes Chromosomes Cancer 39, 29-36 (2004).
5. Giannini, C. et al. Patient tumor EGFR and PDGFRA gene amplifications retained in an invasive intracranial xenograft model of glioblastoma multiforme. Neuro-oncol 7, 164-76 (2005).
6. Sarkaria, J. N. et al. Use of an orthotopic xenograft model for assessing the effect of epidermal growth factor receptor amplification on glioblastoma radiation response. Clin Cancer Res 12, 2264-71 (2006).
7. Sarkaria, J. N. et al. Identification of molecular characteristics correlated with glioblastoma sensitivity to EGFR kinase inhibition through use of an intracranial xenograft test panel. Mol Cancer Ther 6, 1167-74 (2007).
8. Hodgson, G. et al. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. Nat Genet 29, 459-64 (2001).
9. Zhao, X. et al. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 65, 5561-70 (2005).
10. Ishikawa, S. et al. Allelic dosage analysis with genotyping microarrays. Biochem Biophys Res Commun 333, 1309-14 (2005).

11. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5, 557-72 (2004).
12. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80 (2004).
13. Kotliarov, Y. et al. High-resolution global genomic survey of 178 gliomas reveals novel regions of copy number alteration and allelic imbalances. Cancer Res 66, 9428-36 (2006).
14. Nigro, J. M. et al. Integrated array-comparative genomic hybridization and expression array profiles identify clinically relevant molecular subtypes of glioblastoma. Cancer Res 65, 1678-86 (2005).
15. Phillips, H. S. et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 9, 157-73 (2006).
16. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185-93 (2003).
17. Irizarry, R. A. et al. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31, e15 (2003).
18. Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics 20, 1464-5 (2004).
19. Bengtsson, H., Irizarry, R., Carvalho, B. & Speed, T. P. Estimation and assessment of raw copy numbers at the single locus level. Bioinformatics 24, 759-67 (2008).
20. Misra, A. et al. Array comparative genomic hybridization identifies genetic subgroups in grade 4 human astrocytoma. Clin Cancer Res 11, 2907-18 (2005).
21. Mischel, P. S. et al. Identification of molecular subtypes of glioblastoma by gene expression profiling. Oncogene 22, 2361-73 (2003).

22. Maher, E. A. et al. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. Cancer Res 66, 11502-13 (2006).
23. Horvath, S. et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc Natl Acad Sci U S A 103, 17402-7 (2006).
24. Liu, M. et al. FoxM1B is overexpressed in human glioblastomas and critically regulates the tumorigenicity of glioma cells. Cancer Res 66, 3593-602 (2006).
25. Chan, D. et al. Over-expression of FOXM1 transcription factor is associated with cervical cancer progression and pathogenesis. J Pathol 215, 245-52 (2008).
26. Dai, B. et al. Aberrant FoxM1B expression increases matrix metalloproteinase-2 transcription and enhances the invasion of glioma cells. Oncogene 26, 6212-9 (2007).
27. Wang, Z., Banerjee, S., Kong, D., Li, Y. & Sarkar, F. H. Down-regulation of Forkhead Box M1 transcription factor leads to the inhibition of invasion and angiogenesis of pancreatic cancer cells. Cancer Res 67, 8293-300 (2007).
28. Laoukili, J. et al. FoxM1 is required for execution of the mitotic programme and chromosome stability. Nat Cell Biol 7, 126-36 (2005).
29. Schuller, U. et al. Forkhead transcription factor FoxM1 regulates mitotic entry and prevents spindle defects in cerebellar granule neuron precursors. Mol Cell Biol 27, 8259-70 (2007).
30. Chakravarti, A. et al. Survivin enhances radiation resistance in primary human glioblastoma cells via caspase-independent mechanisms. Oncogene 23, 7494-506 (2004).
31. Guan, H. T. et al. Down-regulation of survivin expression by small interfering RNA induces pancreatic cancer cell apoptosis and enhances its radiosensitivity. World J Gastroenterol 12, 2901-7 (2006).
32. Kami, K. et al. Downregulation of survivin by siRNA diminishes radioresistance of pancreatic cancer cells. Surgery 138, 299-305 (2005).

33. Tan, Y., Raychaudhuri, P. & Costa, R. H. Chk2 mediates stabilization of the FoxM1 transcription factor to stimulate expression of DNA repair genes. Mol Cell Biol 27, 1007-16 (2007).

34. Orjalo, A. V. et al. The Nup107-160 nucleoporin complex is required for correct bipolar spindle assembly. Mol Biol Cell 17, 3806-18 (2006).

35. Zuccolo, M. et al. The human Nup107-160 nuclear pore subcomplex contributes to proper kinetochore functions. Embo J 26, 1853-64 (2007).

36. Clark, H. F. et al. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. Genome Res 13, 2265-70 (2003).

37. Pulukuri, S. M. & Rao, J. S. Matrix metalloproteinase-1 promotes prostate tumor growth and metastasis. Int J Oncol 32, 757-65 (2008).

38. Stojic, J. et al. Expression of matrix metalloproteinases MMP-1, MMP-11 and MMP-19 is correlated with the WHO-grading of human malignant gliomas. Neurosci Res 60, 40-9 (2008).

39. Chu, C. Y. et al. Involvement of matrix metalloproteinase-13 in stromal-cell-derived factor 1 alpha-directed invasion of human basal cell carcinoma cells. Oncogene 26, 2491-501 (2007).

40. Ala-aho, R. et al. Targeted inhibition of human collagenase-3 (MMP-13) expression inhibits squamous cell carcinoma growth in vivo. Oncogene 23, 5111-23 (2004).

41. Grimmer, M. R. & Weiss, W. A. Childhood tumors of the nervous system as disorders of normal development. Curr Opin Pediatr 18, 634-8 (2006).

42. Collins, V. P. Amplified genes in human gliomas. Semin Cancer Biol 4, 27-32 (1993).

43. Hui, A. B., Lo, K. W., Yin, X. L., Poon, W. S. & Ng, H. K. Detection of multiple gene amplifications in glioblastoma multiforme using array-based comparative genomic hybridization. Lab Invest 81, 717-23 (2001).

44. Rocak, S. & Linder, P. DEAD-box proteins: the driving forces behind RNA metabolism. Nat Rev Mol Cell Biol 5, 232-41 (2004).

45. Li, L., Monckton, E. A. & Godbout, R. A role for DEAD box 1 at DNA double-strand breaks. Mol Cell Biol (2008).

46. Guan, Y. et al. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. Clin Cancer Res 13, 5745-55 (2007).

47. Holtkamp, N. et al. Characterization of the amplicon on chromosomal segment 4q12 in glioblastoma multiforme. Neuro Oncol 9, 291-7 (2007).

48. Aagaard, L. & Rossi, J. J. RNAi therapeutics: principles, prospects and challenges. Adv Drug Deliv Rev 59, 75-86 (2007).

49. Mathupala, S. P., Guthikonda, M. & Sloan, A. E. RNAi based approaches to the treatment of malignant glioma. Technol Cancer Res Treat 5, 261-9 (2006).

50. Hashizume, R. et al. New therapeutic approach for brain tumors: Intranasal delivery of telomerase inhibitor GRN163. Neuro Oncol 10, 112-20 (2008).

51. Hayes, M. E. et al. Increased target specificity of anti-HER2 genospheres by modification of surface charge and degree of PEGylation. Mol Pharm 3, 726-36 (2006).

52. Hayes, M. E. et al. Genospheres: self-assembling nucleic acid-lipid nanoparticles suitable for targeted gene delivery. Gene Ther 13, 646-51 (2006).

53. Saito, R. et al. Convection-enhanced delivery of Ls-TPT enables an effective, continuous, low-dose chemotherapy against malignant glioma xenograft model. Neuro-oncol 8, 205-14 (2006).

54. Yamashita, Y. et al. Convection-enhanced delivery of a topoisomerase I inhibitor (nanoliposomal topotecan) and a topoisomerase II inhibitor (pegylated liposomal doxorubicin) in intracranial brain tumor xenografts. Neuro-oncol 9, 20-8 (2007).

55. Neve, R. M. et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515-27 (2006).

Captions for all illustrations

Figure 1

Frequencies of genomic copy number gains and losses in GBMX tumor lines and GBM clinical samples. (A) GBMX tumors (N=21) analyzed on the Affymetrix Xba 50k SNP array platform; (B) GBM tumors (N=82) analyzed on the Affymetrix Xba 50k SNP array platform [13]; (C) GBM tumors (N=56) analyzed on the BAC array platform [14]; (D) GBM tumors (N=221) analyzed on the Agilent 244K oligonucleotide array platform [2].

Figure 2

(A) Unsupervised hierarchical clustering dendrogram (Pearson-Ward) of GBMX tumors and non-neoplastic controls based on the 100 most variably expressed genes amongst the samples. Colored boxes represent tumors from the same line, but from distinct tumor passages (the number after the period represents the generation number). (B) Expression of proneural, mesenchymal, and proliferative genes in GBMX tumors, and GBM clinical samples [15].

Figure 3

Ingenuity networks identified from MRC-gene list. (A) Cellular assembly and organization network. (B) DNA replication, recombination and repair network. Genes overexpressed in GBMX and GBM tumors (Table 1) are shown by red symbols.

Figure 4

Expression of amplified genes in GBMX and GBM tumors. (A) Correlation of EGFR expression versus EGFR copy number in GBMX tumors. (B) Expression of amplified genes at 12q15 in GBMX tumors and TCGA GBMs. (C) Expression of amplified genes at 11q22 in GBMX tumors and TCGA GBMs. (D) Expression of amplified genes at 2p24 in GBMX and TCGA GBMs. For B-D: open circle represents expression in the respective amplified GBMX

line; expression for non-neoplastic controls and GBMX tumors has been normalized to the median expression of non-neoplastic controls; expression for TCGA GBMS has been normalized to the median expression of all tumors.

Figures 1-4

Figure 1

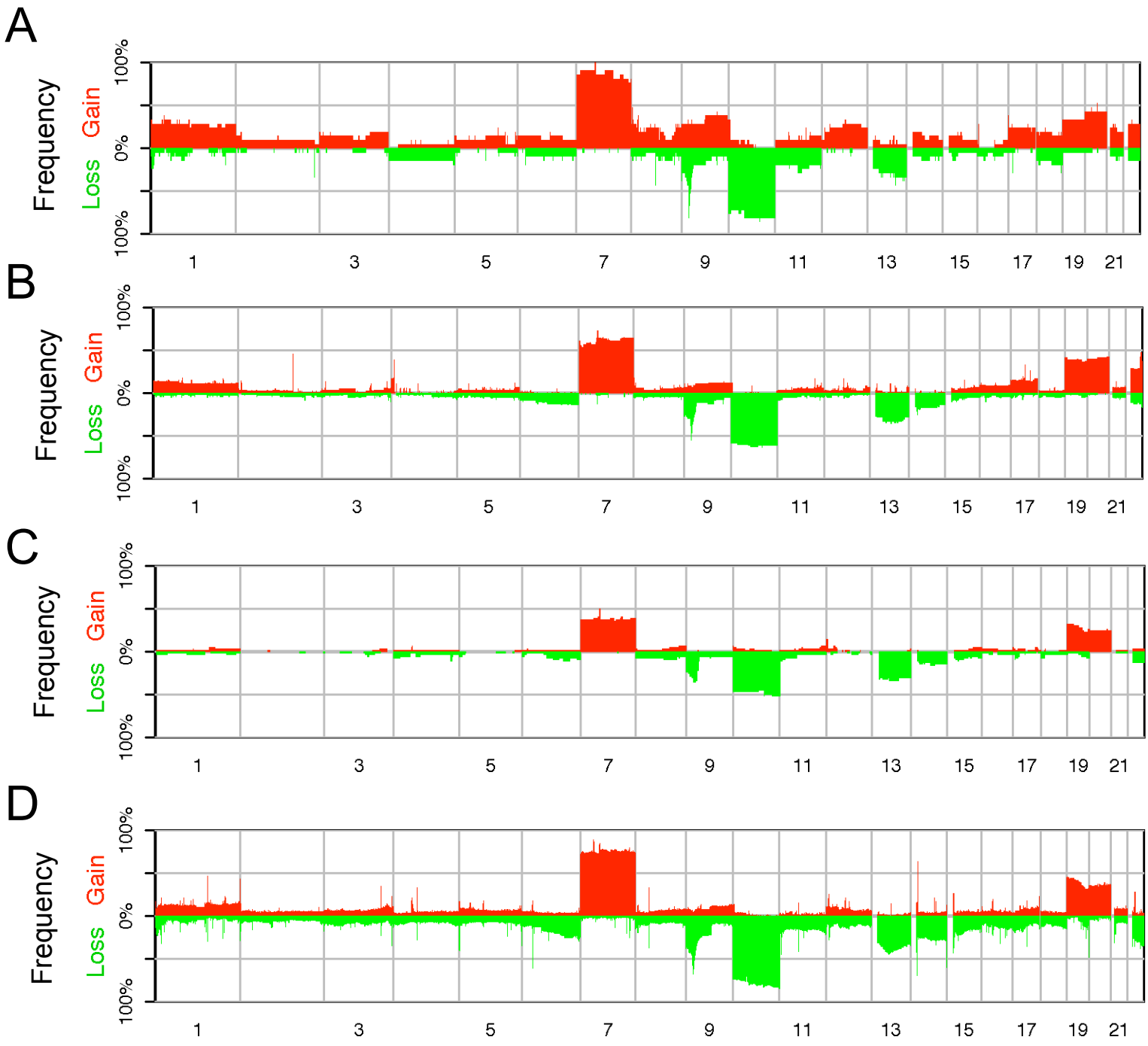


Figure 2

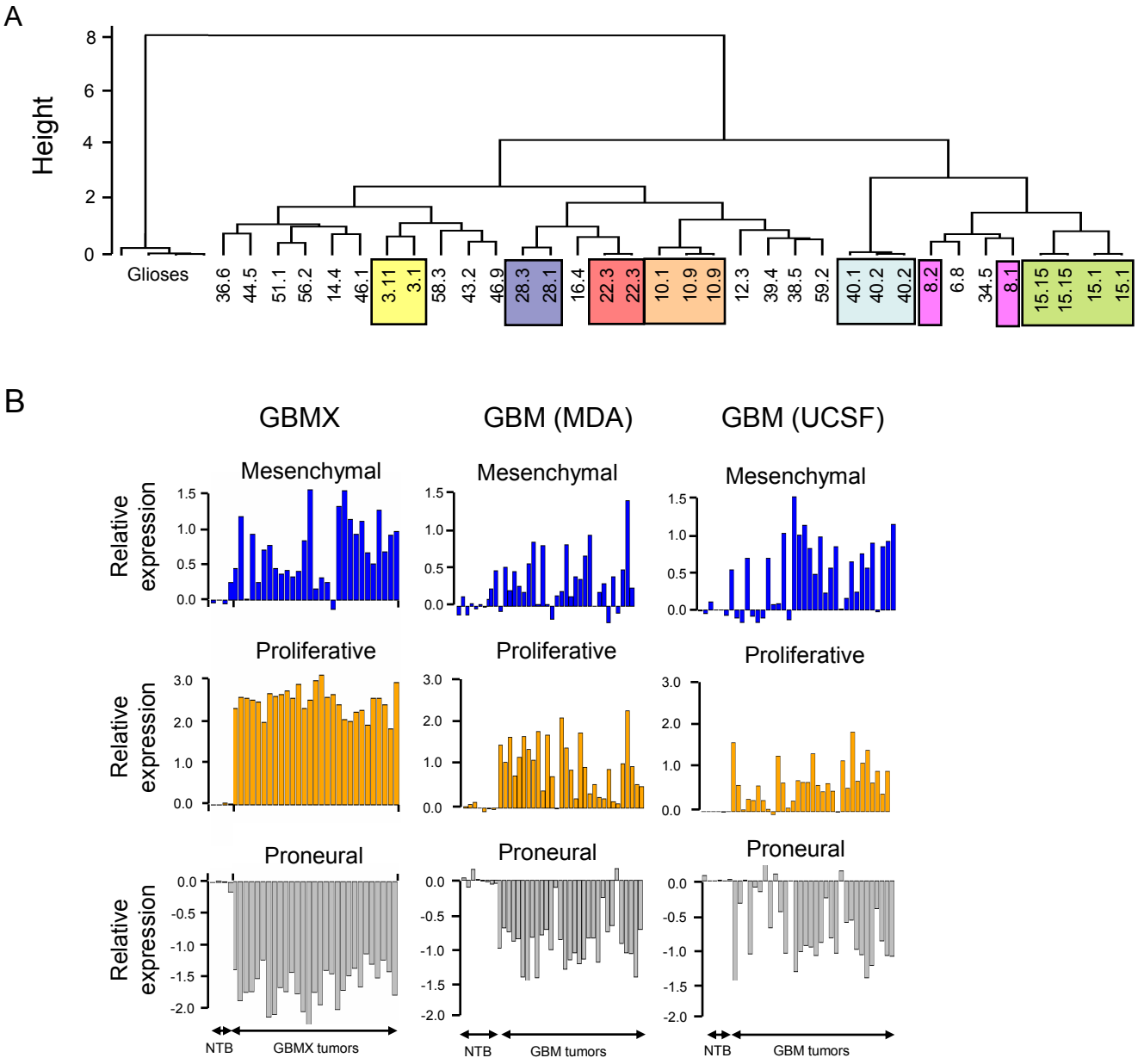
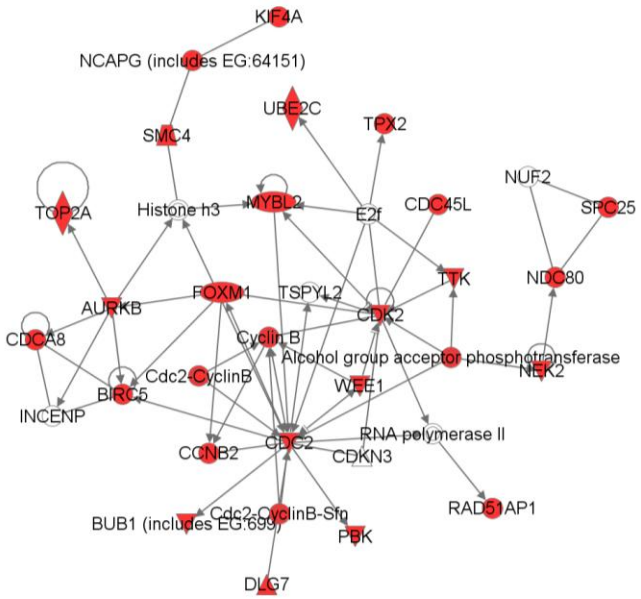

A
Height
8
6
4
2
0
Glioses
36.6
44.5
51.1
56.2
14.4
46.1
3.11
3.1
58.3
43.2
46.9
28.3
28.1
16.4
22.3
22.3
10.1
10.9
10.9
12.3
39.4
38.5
59.2
40.1
40.2
40.2
8.2
6.8
34.5
8.1
15.15
15.15
15.1
15.1
B
GBMX
GBM (MDA)
GBM (UCSF)
Mesenchymal
Proliferative
Proneural
Relative expression
NTB
GBMX tumors
GBM tumors

Figure 3

A Cellular assembly and organization network



B DNA replication, recombination and repair network

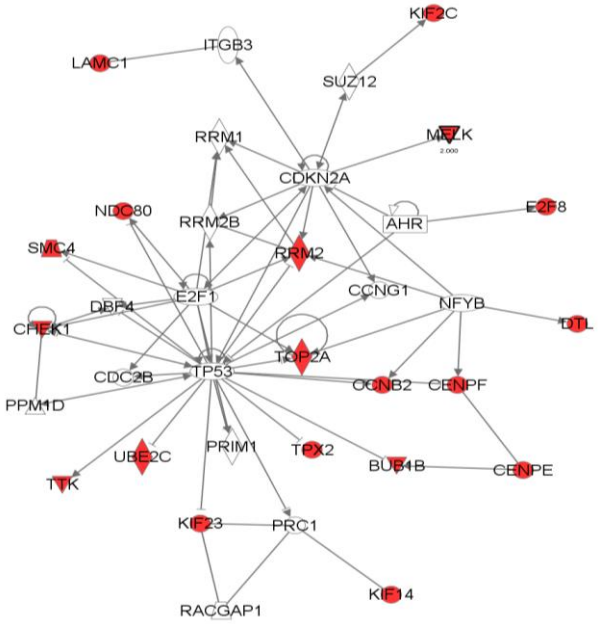


Figure 4

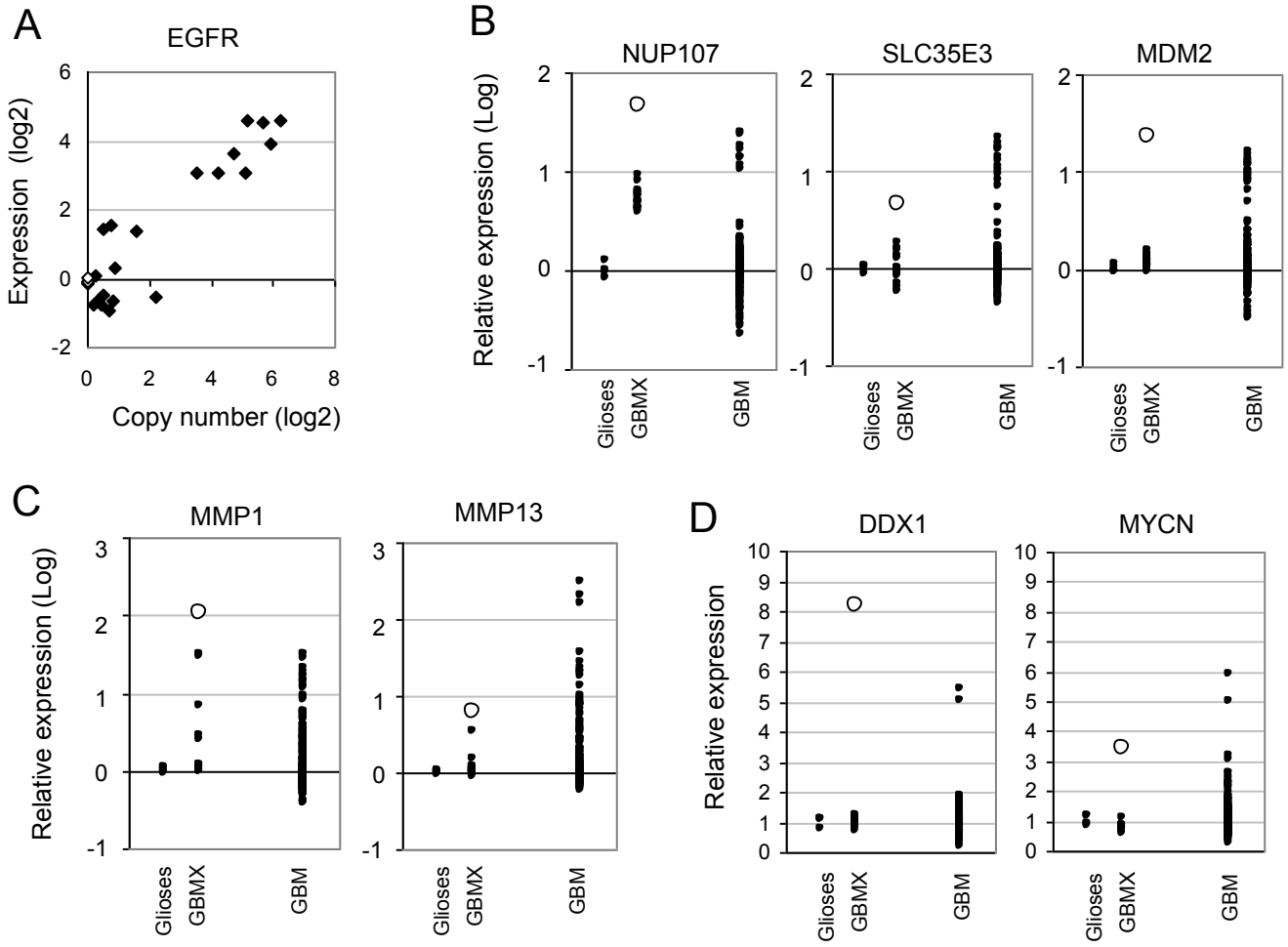

A
EGFR
Expression (log2)
Copy number (log2)
B
NUP107
SLC35E3
MDM2
Relative expression (Log)
Glioses
GBMX
GBM
C
MMP1
MMP13
Relative expression (Log)
D
DDX1
MYCN
Relative expression

Table1

Table 1: Mitosis, DNA replication, and chromosome organization (MRC) genes over-expressed in GBMX tumors and de novo GBMs

Gene	Gene symbol	Fold change in GBMX tumors		Fold change in GBM tumors	
		Average	Max	Average	Max
1	ASPM	71	140	13	28
2	AURKB	11	23	3	6
3	BIRC5	42	109	3	6
4	BUB1	22	45	5	9
5	BUB1B	26	53	6	13
6	CCNB2	48	103	3	7
7	CDC2	50	126	5	17
8	CDC45L	10	30	6	10
9	CDCA8	9	22	3	7
10	CDK2	16	31	3	5
11	CENPE	17	32	8	18
12	CENPF	33	67	16	22
13	CHEK1	14	40	3	5
14	DLG7	38	85	11	16
15	DTL	41	74	4	11
16	E2F8	7	18	6	9
17	FOXM1	18	32	3	6
18	GINS2	11	20	3	6
19	HELLS	5	16	3	9
20	IGF2BP3	64	167	2	3
21	KIF14	13	33	8	14
22	KIF18A	9	18	6	17
23	KIF20A	32	70	5	9
24	KIF23	21	42	6	11
25	KIF2C	33	81	3	6
26	KIF4A	24	50	7	11
27	MLF1IP	57	140	6	20
28	MYBL2	5	13	3	7
29	NCAPG	22	37	6	13
30	NDC80	30	73	8	11
31	NEK2	16	31	2	4
32	PBK	106	252	7	14
33	RAD51AP1	21	46	2	6
34	RRM2	112	285	5	13
35	SMC4	28	59	3	6
36	SPC25	16	30	3	13
37	TOP2A	83	152	10	14
38	TPX2	47	113	6	10
39	TTK	25	50	10	19
40	UBE2C	49	112	3	4
41	WEE1	13	24	4	6

Table 2: High-level genomic amplicons in GBMX tumors.

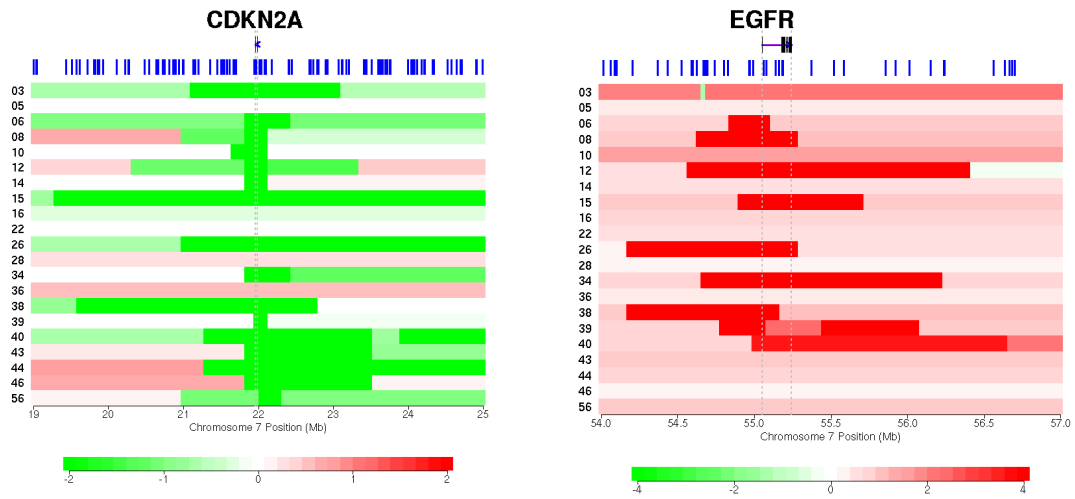
GBMX amplicon (log2rat>4)	Chr.	Minimal amplicon (NCBI build 36.1)	Max Copy No. (log2)	Width (kb)	GBMX line	Frequency of focal amplifications (GBM)		Focal CNA region (GBM)	Amplified and overexpressed genes in GBMX tumors
						GISTIC	RAE	GTS	
1	**2**	**chr2:14372951-15050902**	**5.7**	**678**	**28**	**2.4**	**-**	**-**	
2	**2**	**chr2:15614969-16256395**	**6.2**	**641**	**28**	**2.4**	**-**	**-**	**DDX1, MYCN**
3	**2**	**chr2:120958377-121603516**	**6.3**	**645**	**28**	**-**	**-**	**Y**	**GLI1**
4	**7**	**chr7:51987798-53893473**	**5.0**	**1906**	**26**	**-**	**-**	**Y**	
5	**7**	**chr7:54970126-55186653**	**6.2**	**215**	*****	**43.2**	**44.0**	**Y**	**EGFR**
6	7	chr7:90939401-91161022	6.5	222	34	-	-	-	
7	**7**	**chr7:91236115-94061316**	**5.0**	**2825**	**34**	**3.4**	**3.7**	**-**	**BET1,CASD1,CDK6,DKFZP564O0523, KRIT1,MTERF,PEX1,SAMD9**
8	7	chr7:150881861-150906739	5.7	347	12, 34	-	-	-	
9	8	chr8:36998497-37472939	4.2	474	22	-	-	-	
10	8	chr8:129006828-131970985	4.8	2964	5	-	-	-	DDEF1,FAM49B
11	8	chr8:137526870-137892295	4.7	365	22	-	-	-	
12	11	chr11:101575571-102626961	5.1	1051	22	-	-	-	BIRC2,BIRC3,MMP1,MMP10,MMP12,MMP13,MMP27,MMP3MMP8,TMEM123,YAP1
13	**12**	**chr12:66409069-67695739**	**4.9**	**1287**	**5**	**10.7**	**10.2**	**Y**	**CPM,IFNG,IL26,MDM1,MDM2,NUP107, SLC35E3**
14	13	chr13:23598616-24008514	5.3	410	6	-	-	-	PARP4
15	13	chr13:66424482-67844382	4.5	1420	6	-	-	-	

* EGFR amplification (log2>4) was observed in 8/21 independent GBMX lines: 6, 8, 12, 15, 26, 34, 38, 39. Bolded amplicons are those present in both GBMX tumors (this study) and TCGA-derived GBM tumors.

Supplementary Figure

Supplemental Figure 1

A



Supplementary figure 1 (A) Segmented means of chromosome 9 copy number (log2) around CDKN2A in GBMX lines. The extent of homozygous deletion ranged from 130 Kb (GBMX39) to 8 Mb (GBMX15). (B) Segmented means of chromosome 7 copy number (log2) around EGFR in GBMX lines. whereas the extent of amplification at the EGFR locus ranged from 250 Kb (GBMX06, GBMX39) to 2.7 Mb (GBMX40). The SNPs (blue bars) and gene structure are displayed on top. The GBMX line numbers are displayed to the left of each panel.

Supplementary table 1: GBM xenograft- and non-neoplastic-samples profiled in this study.

Sample ID*	Tissue type	DNA copy number	mRNA expression
GBMX_3.1	Xenograft	-	Y
GBMX_3.11	Xenograft	Y	Y
GBMX_5.1	Xenograft	Y	Y
GBMX_6.8	Xenograft	Y	Y
GBMX_8.1	Xenograft	-	Y
GBMX_8.2	Xenograft	Y	Y
GBMX_10.1	Xenograft	-	Y
GBMX_10.9	Xenograft	Y	Y
GBMX_12.3	Xenograft	Y	Y
GBMX_14.4	Xenograft	Y	Y
GBMX_15.1	Xenograft	-	Y
GBMX_15.15	Xenograft	Y	Y
GBMX_16.4	Xenograft	Y	Y
GBMX_22.3	Xenograft	Y	Y
GBMX_26.7	Xenograft	Y	-
GBMX_28.1	Xenograft[&]	-	Y
GBMX_28.3	Xenograft[&]	Y	Y
GBMX_34.5	Xenograft	Y	Y
GBMX_36.6	Xenograft	Y	Y
GBMX_38.5	Xenograft	Y	Y
GBMX_39.4	Xenograft	Y	Y
GBMX_40.1	Xenograft	-	Y
GBMX_40.2	Xenograft	Y	Y
GBMX_43.2	Xenograft	Y	Y
GBMX_44.5	Xenograft	Y	Y
GBMX_46.1	Xenograft	-	Y
GBMX_46.9	Xenograft	Y	Y
GBMX_56.2	Xenograft	Y	Y
GBMX_58.3	Xenograft	-	Y
GBMX_59.2	Xenograft	-	Y
SF4916	Gliosis	-	Y
SF6637	Gliosis	-	Y
SF6700	Gliosis	-	Y
SF7178	Gliosis	-	Y

* Nomenclature for GBM xenograft (GBMX) tumors: the number preceding '.' is the tumor line number, the number after '.' is the xenograft generation/passage number. [&]Line 28 was derived from a grade IV gliosarcoma.

Supplementary Table 2

Supplementary table 2: Gene ontology (GO) classification of genes significantly up-regulated greater than 2-fold on average in GBMX tumors relative to glioses		
GO term*	P value*	GO numerical identifiers
M phase	0	**GO:0000279**; GO:0022403; GO:0000087; GO:0007067; GO:0022402; GO:0007017; GO:0000074; GO:0051726; GO:0000070; GO:0000819; GO:0007051; GO:0051325
DNA replication	1.27E-51	**GO:0006260**; GO:0006259; GO:0006139; GO:0016071; GO:0006974; GO:0006281; GO:0006396; GO:0009719
RNA splicing	7.85E-28	**GO:0008380**; GO:0006397; GO:0043283; GO:0043170
Chromosome organization and biogenesis	9.59E-21	**GO:0051276**
*GO term and P value correspond to the most significant GO numerical identifier (bold) contained within each GO cluster		

Supplementary Table 3

Supplementary table 3: Expression of genes that map within a high-level amplicon in GBMX tumors

Gene	Amplicon	Fold change in GBMX tumors*	Fold change in TCGA GBMs**
ADCY8	10	0.90	not determined
AKAP9	7	0.91	not determined
ANKIB1	7	no U133A probe set	not determined
BET1	7	8.10	3.63
BIRC2	12	9.46	2.06
BIRC3	12	89.74	16.03
C1QTNF9	14	no U133A probe set	2.39
CALCR	7	0.91	not determined
CASD1	7	10.61	2.71
CCDC132	7	no U133A probe set	10.07
CDK6	7	26.69	6.70
COL1A2	7	0.07	not determined
CPM	13	22.85	12.46
CYP51A1	7	2.48	not determined
DDEF1	10	6.50	2.45
DDX1	2	8.63	5.43
DKFZp564N2472	4	no U133A probe set	2.40
DKFZP564O0523	7	9.89	7.38
EGFR	5	13.96	9.51
FAM133B	7	no U133A probe set	4.50
FAM49B	10	4.81	2.85
GATAD1	7	1.21	not determined
GLI2	3	10.35	2.12
GNGT1	7	0.93	not determined
HEPACAM2	7	no U133A probe set	not determined
IFNG	13	15.04	4.31
IL22	13	2.22	not determined
IL26	13	4.78	2.97
KRIT1	7	4.61	2.44
MDM1	13	48.86	3.63
MDM2	13	19.76	16.39
MLZE	10	no U133A probe set	2.95
MMP1	12	56.93	31.48
MMP10	12	11.75	12.99
MMP12	12	14.86	48.49
MMP13	12	5.62	308.41
MMP20	12	1.69	not determined
MMP27	12	37.14	2.26
MMP3	12	21.75	321.39
MMP7	12	1.83	not determined
MMP8	12	54.79	8.46
MTERF	7	14.27	3.63
MYCN	2	4.32	5.92
NAG	2	0.54	not determined
NUP107	13	9.11	24.37
PARP4	14	15.57	3.48
PCDH9	15	0.75	not determined
PEX1	7	13.04	6.67
PRKAG2	8	2.99	not determined
RAP1B	13	0.60	not determined
SAMD9	7	6.24	5.56
SAMD9L	7	no U133A probe set	4.47
SGCE	7	0.10	not determined
SLC35E3	13	4.69	21.39
SPATA13	14	no U133A probe set	2.78
TFPI2	7	0.30	not determined
TMEM123	12	8.68	2.39
YAP1	12	40.03	3.38

No RefSeq genes were observed in amplicons 1, 6, 9, 11.

* Expression in amplified tumor relative to average expression in non-amplified tumors

** Maximum expression observed in tumors realtive to the median expression observed in tumors