Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280

# Genomic Changes and Gene Expression Profiles Reveal That Established Glioma Cell Lines Are Poorly Representative of Primary Human Gliomas

**Aiguo Li,[1] Jennifer Walling,[1] Yuri Kotliarov,[1] Angela Center,[1] Mary Ellen Steed,[1] Susie J. Ahn,[1] Mark Rosenblum,[2] Tom Mikkelsen,[2] Jean Claude Zenklusen,[1] and Howard A. Fine[1]**

*[1]Neuro-Oncology Branch, National Cancer Institute, National Institutes of Neurological Disorder and Stroke, NIH, Bethesda, Maryland and [2]Neurology and Neurosurgery, Hermelin Brain Tumor Center, Henry Ford Hospital, Detroit, Michigan*

## Abstract

**Genetic aberrations, such as gene amplification, deletions, and loss of heterozygosity, are hallmarks of cancer and are thought to be major contributors to the neoplastic process. Established cancer cell lines have been the primary *in vitro* and *in vivo* models for cancer for more than 2 decades; however, few such cell lines have been extensively characterized at the genomic level. Here, we present a high-resolution genome-wide chromosomal alteration and gene expression analyses of five of the most commonly used glioma cell lines and compare the findings with those observed in 83 primary human gliomas. Although genomic alterations known to occur in primary tumors were identified in the cell lines, we also observed several novel recurrent aberrations in the glioma cell lines that are not frequently represented in primary tumors. Additionally, a global gene expression cluster distinct from primary tumors was identified in the glioma cell lines. Our results indicate that established cell lines are generally a poor representation of primary tumor biology, presenting a host of genomic and gene expression changes not observed in primary tissues, although some discrete features of glioma biology were conserved in the established cell lines. Refined maps of genetic alterations and transcriptional divergence from the original tumor type, such as the one presented here, may help serve as a guideline for a more biologically rational and clinically relevant selection of the most appropriate glioma model for a given experiment. (Mol Cancer Res 2008;6(1):21–30)**

## Introduction

Changes to the normal DNA content, including copy number alterations (CNA) and allelic imbalances, and their resultant effects on gene transcription have long been recognized as hallmarks of cancer. Due to the inherent difficulty in establishing and maintaining primary tumor cell cultures, established cell lines have been traditionally used to characterize the biological significance of specific genomic aberrations identified in primary tumors. It has therefore been assumed that the major genomic alterations found in the primary tumor specimens are represented in the corresponding cell lines and that the biological consequences of these alterations in primary tumors are mirrored by those observed in the established cell lines.

Evidence for genomic instability is seen in malignant gliomas, as it is in other tumors, and may be a contributor to gliomagenesis (1). The resulting chromosomal copy number changes and loss of heterogeneity can lead to alteration of oncogene and tumor suppressor gene dosage (2-4). Detection of specific chromosomal alterations can therefore aid in the identification of genes whose aberrant function may contribute to the tumorigenic process. We recently published a high-resolution genomic survey of 178 primary human gliomas using Affymetrix 100K single nucleotide polymorphism (SNP) arrays and identified many novel regions of gene copy number and allelic alterations (5), greatly expanding the public glioma genomics database that had previously been generated using comparative genomic hybridization (CGH; refs. 2, 6-8).

Few studies, however, have attempted a systematic genomic survey of the most commonly used glioma cell lines for the purpose of evaluating their similarity to primary gliomas. Kubota et al. (2) studied the chromosomal translocation and abnormalities of nine glioma cell lines and identified seven recurrent marker chromosomes with similar chromosome rearrangements and 12 amplified loci in five of these cell lines. It was subsequently shown, however, that several of the cell lines studied were subclones of the same cell line (U251, U373, and SNB-19), casting uncertainty on the identified recurrent

Received 6/15/07; revised 8/21/07; accepted 8/21/07.
**Grant support:** Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
**Note:** Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).
J.C. Zenklusen and H.A. Fine contributed equally to this work.
**Requests for reprints:** Howard A. Fine, Neuro-Oncology Branch, National Cancer Institute, National Institutes of Neurological Disorder and Stroke, NIH, 9030 Old Georgetown Drive, MSC8200, Building 82, Room 225, Bethesda, MD 20892. Phone: 301-402-6383. E-mail: hfine@mail.nih.gov

doi:10.1158/1541-7786.MCR-07-0280

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



alterations (9). Moreover, the study by Kubota el al. used a then standard metaphase-spread CGH platform, allowing only low-level genomic resolution. Finally, the previously published work did not include the U87 glioma cell line, one of the most commonly used cell lines and one of the few consistently found to be tumorigenic in both s.c. and intracranial animal models, thereby making it a standard preclinical model for screening therapeutic agents. Thus, it is still unclear how closely established glioma cell lines mimic primary gliomas at the genomic level.

To address the genetic similarities between cell lines and primary gliomas, we did a high-resolution genomic survey of frequently used glioma cell lines using a high-density oligonucleotide array-based, 10K SNP array to identify genome-wide chromosomal abnormalities, including loss of heterozygosity (10, 11). The advantage of high-density SNP arrays for genomic surveys is that it allows one to identify regions of copy number neutral loss of heterozygosity at high resolution, which cannot be done by traditional and array-based CGH. In addition to evaluating the similarity in genomic alterations, we also did genome-wide mRNA expression profiling of the glioma cell lines compared with 74 primary gliomas to evaluate the similarities and differences in gene expression between glioma cell lines and primary gliomas.

As described below, our findings show that established glioma cell lines and primary tumor have significant differences in both genomic alterations and gene expression, leading us to conclude that glioma cell lines may not be an accurate representation or model system for primary gliomas.

## Results

### *Copy Number Aberrations in Established Glioma Cell Lines*

The SNP array profiles of the five established glioma cell lines presented a large number of CNAs as shown in Fig. 1 and Supplementary Table S1. In some cases, amplifications and deletions seemed to be focal changes, restricted to small regions of a chromosome, whereas in the other cases, CNAs extend to large regions (whole arms or even entire chromosomes). As an example of the first kind of alterations, the p arm of chromosome 9 exhibited deletions in four of five cell lines, encompassing the region containing known tumor suppressor genes *CDKN2A* and *CDKN2B*. Some of these deletions are homozygous in nature and the extent of the alteration varies from 0.4 million to 27 million bp. More extensive CNAs are exemplified by trisomies or polysomies of chromosome 7 that were identified in all cell lines, with the exception of A172 that bears partial amplification of the p arm of the chromosome 7. Both of these alterations have been reported in primary tumors. However, many of the alterations found have not been reported as common events in tumor samples, yet they seem to be very common in the long established cell lines, as in the case of polysomies for chromosome 5 found in all cell lines, with the exception of U87. Examples of other nonrandom events (occurring in more than one cell line) included amplifications of 1q21-44 and 2q34-37.1 (40%) and deletions of 11q14-32.3 and 18q11-23 (40%). A host of other nonrecurrent alterations was found in individual cell lines. Among these were complex rearrangements of chromosome 6 in U87 (with two distinct amplification and one deletion region), similar to the situation found in chromosomes 2 and 3 on the Hs683 cell line and chromosome 2 in T98G. There was a tendency for some individual chromosomes to be more frequently altered in the cell lines (all cell lines had chromosome 6 aberrations), although the specific aberration differed from one cell line to another, suggesting perhaps an unstable genotype rather than a biological selection for a specific genetic gain and/or loss.

### *Loss of Heterozygosity in Established Glioma Cell Lines*

One of the advantages of using high-density oligonucleotide SNP arrays is that loss of heterozygosity studies can be determined along with CNA identification. In our analysis, loss of heterozygosity of both chromosomes 10 and 14 appears as the most frequent event in the cell lines studied. Chromosomes 6, 9, and 11 also present loss of heterozygosity with relatively high frequency, whereas chromosomes 5, 8, 16, 20, and 21 presented loss of heterozygosity only in a fraction of the cell lines characterized. In most cases, loss of heterozygosity highly correlated with areas found to be deleted by signal intensity (copy number) analysis. Some areas, however, did present allelic imbalances in a copy number neutral fashion, suggesting a secondary recombination event after deletion (i.e., gene conversion). The results of this analysis are presented in Fig. 2 and Supplementary Table S2.

### *Genomic Comparison between Established Cell Lines and Primary Tumors*

To investigate the biological relevance of the cell line genomic alterations, we compared the profiles of glioma cell lines with those obtained from 83 primary glioblastomas (Fig. 3A; ref. 5). Although some of the alterations identified in the primary gliomas (such as chromosome 7 amplification) are found in the glioma cell lines, the overall spectrum of genomic alterations differs substantially between primary tumors and established glioma cell lines. For example, chromosome 13 deletions commonly found in primary glioblastomas are not found in cell lines, whereas the frequent chromosome 5 amplifications and chromosome 18 deletion seen in cell lines are not present in tumor samples. Additionally, amplification of chromosome 20 and deletions of chromosomes 21, commonly found in glioblastomas, were not found in glioma cell lines. Beyond these commonly found alterations, cell lines also presented a host of nonrecurrent changes in areas that rarely are seen altered in primary tumors, such as the CNAs seen in chromosomes 2, 3, 6, and 8. This wide range and diversity of alterations suggest a genomic instability phenotype in established cell lines not observed in the tumor tissue of origin.

These striking discrepancies lead us to question if some of the commonly seen alterations in the glioma cell lines were not, in fact, general cancer cell line *in vitro* artifacts secondary to the selection pressure for optimal *in vitro* cell growth following years of cell passage. To address this possibility, an additional 10 established nonglioma cancer cell line genomes were evaluated using the same high-resolution SNP arrays. Our analysis indicated that 9 of these 10 cell lines contained the fragmental or entire chromosome 5 amplifications seen in our glioma cell lines but not seen in glioma tumors, suggesting that

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



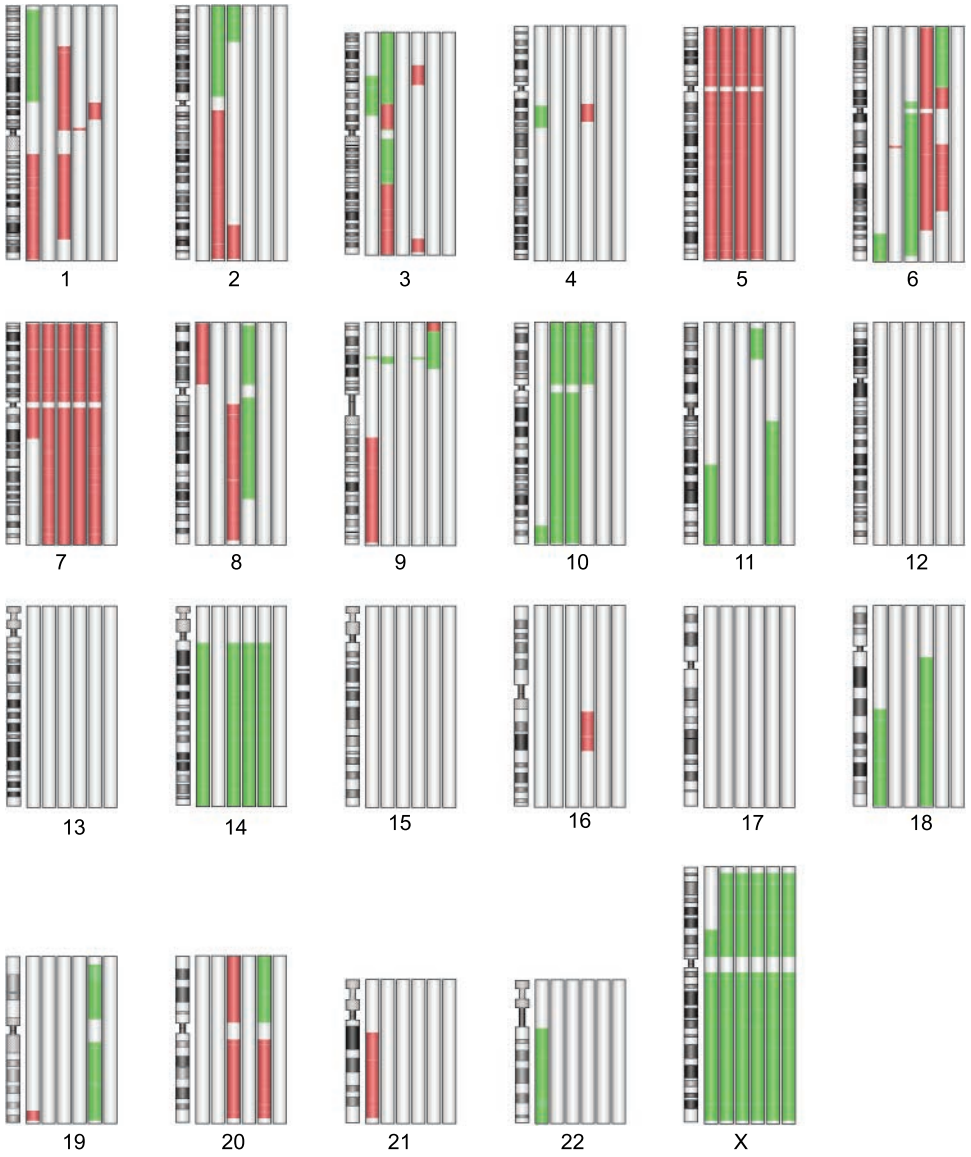


**FIGURE 1.** Regions of chromosome amplification and deletion detected using DNAcopy package and CGHAnalytics software for (in order from *left* to *right*) A172, Hs683, T98G, U251, and U87 glioma cell lines across all chromosomes. Red, amplifications; green, deletions.

this change is an *in vitro* cell line rather than glioma-specific change. Likewise, 8 of 10 nonglioma cell lines and 2 of 5 glioma cell lines showed chromosome 18 deletions, a phenomenon rarely seen in primary glioblastomas.

### *Transcriptomic Comparison between Established Cell Lines and Primary Glial Tumors*

To further characterize the degree of similarity between established glioma cell lines and primary gliomas, we compared the mRNA expression profiles between these two groups. The global gene expression of 73 primary glioblastoma specimen and all cell lines grown *in vitro* was profiled using Affymetrix U133 Plus 2.0 chips. To partially account for the significant effects that the local microenvironment can have on tumor cell gene expression, we included U87 and U251 cell line xenografts grown in the brains of immunodeficient mice. Unsupervised principal component analysis (PCA) and hierarchical clustering (HC) of all probe sets clearly segregated the primary tumors from established cell lines (Fig. 3B and C). The first of the three dimensions of the PCA explained 36.9% of the variations inherent in the data. The PCA and HC analysis also clearly distinguishes the glioma cell lines from the nonglioma cancer cell lines, revealing tissue of origin differences. All established glioma and nonglioma cell lines grown either *in vitro* or *in vivo*, however, had gene expression profiles closer to each other than to the primary tumors, suggesting that gene expression was more a function of common culture conditions than to shared histologic lineage.

### *Gene Set Enrichment Analysis of Glioblastoma and Cell Lines*

It is well established that most cellular processes are not a function of a single gene but rather are a result of a well-orchestrated network of multiple gene interactions. Regular class comparison methodologies do not allow one to probe these complex relationships but only single gene expression

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



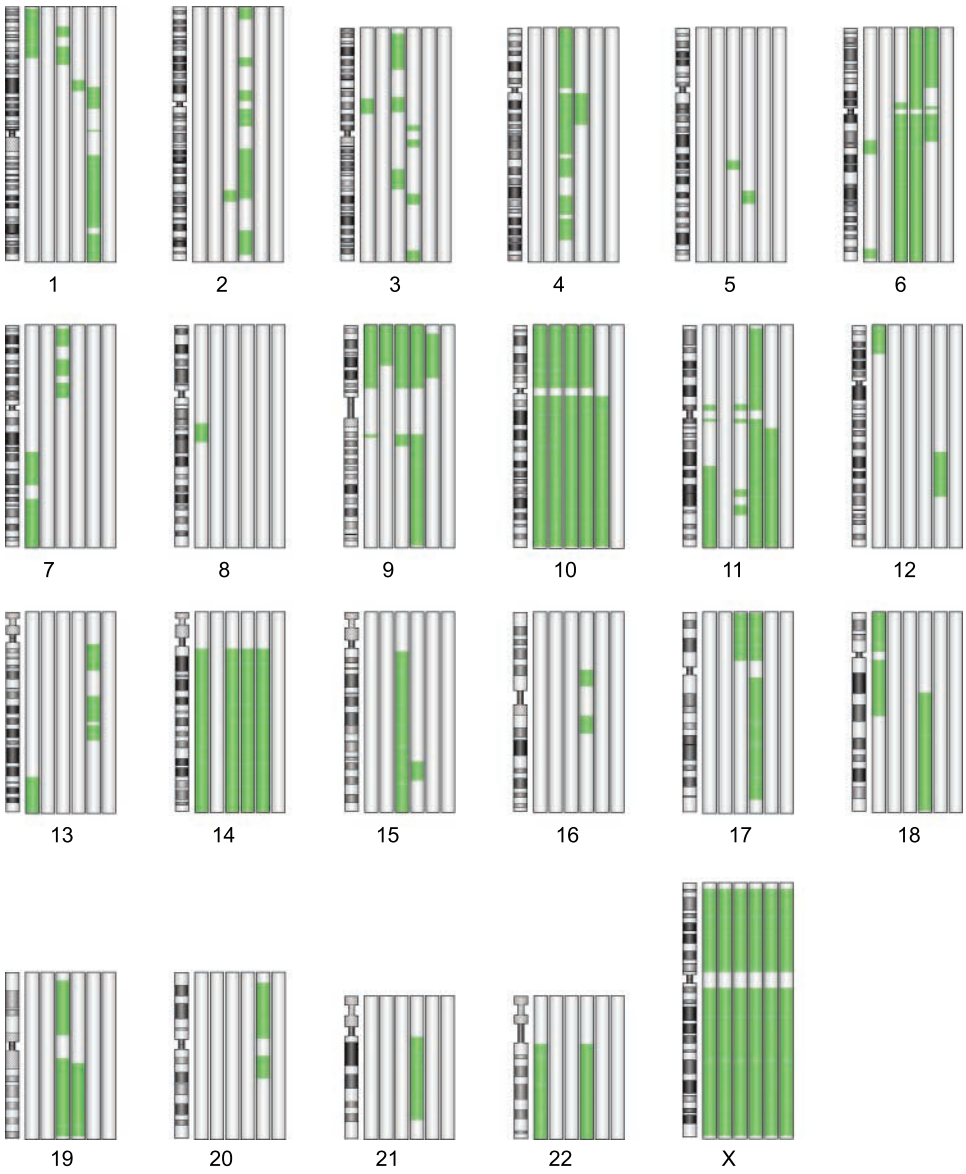


**FIGURE 2.** Regions of loss of heterozygosity detected according to the genotype calls from GDAS software with a threshold of 6% for the probability of being homozygous for a given region for (in order from *left* to *right*) A172, Hs683, T98G, U251, and U87 glioma cell lines across all chromosomes. Green, loss of heterozygosity regions.

differences between groups. Gene set enrichment analysis (GSEA) is an innovative analytic methodology that allows one to assess the differences between phenotypes of interest in biologically relevant functional gene units rather than as individual genes (12). We used GSEA to evaluate the distribution of overexpressed and underexpressed gene sets by examining all 526 functional gene sets (475 metabolic and signaling pathways and 51 expression signature sets) in MSigDB v1.0 against four pairwise phenotypes: glioblastomas versus glioma cell lines, glioblastomas versus all cell lines, glioma cell lines versus nonglioma tumor cell lines, and *in vivo* versus *in vitro* glioma cell lines.

Analysis of significantly up-regulated gene sets in primary glioblastomas compared with established cell lines revealed overlap in several gene sets representing canonical, oncogenic-associated signaling pathways, such as those related to *p53*, hypoxia, vascular endothelial growth factor, *Wnt*, platelet-derived growth factor, *PTEN*, transforming growth factor, and phosphatidylinositol 3-kinase signaling pathways (Fig. 4C). Despite the similarities in general oncogenic signaling pathways, there were very large differences in gene expression sets between primary tumors and *in vitro* cell lines in many areas, such as the overrepresentation of expression in primary gliomas of genes related to cell adhesion, G protein–coupled receptor signaling (B secretin-like signaling, par1 signaling, and Nfat signaling pathways), and fatty acid metabolism (Table 1; Supplementary Tables S3 and S4; Fig. 4B). By contrast, cell lines showed up-regulated expression of gene sets related to the cell cycle, purine metabolism, proteasome activity, mitochondrial activity, and oxidative phosphorylation, indicating a selection process for pathways essential for high replicative potential (Table 2; Supplementary Tables S5 and S6; Fig. 4A). Consistent with immortalization, we also found that telomerase-related (*hTERT*) genes are up-regulated in the cell lines.

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



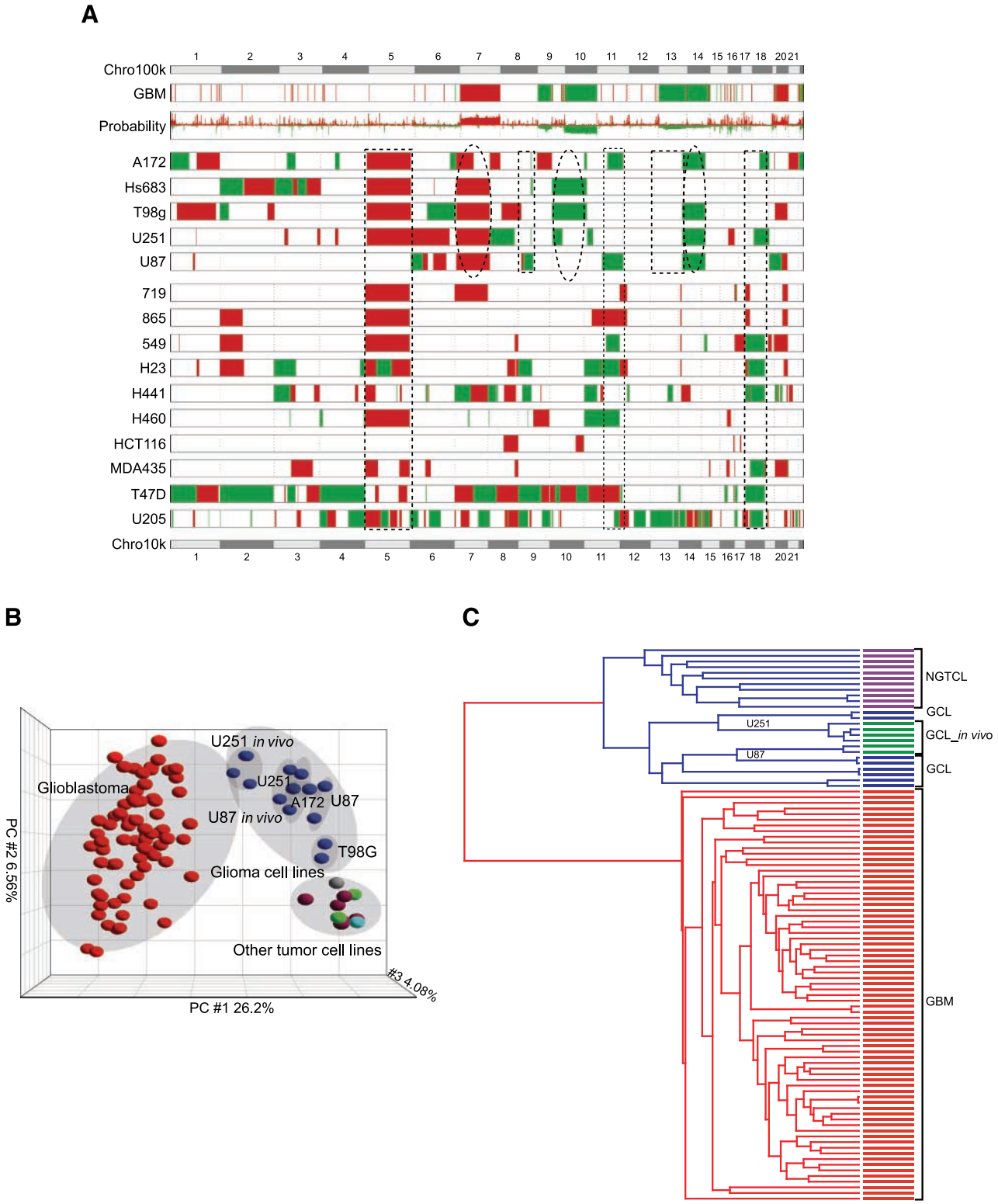


**FIGURE 3.** Differences in genomic alterations and transcriptomic profiles between established cell lines and tumor samples. **A.** The spectrum of genomic aberrations found in a consensus (10% or more) of 83 glioblastoma samples, in 5 glioma cell lines, and in 10 nonglioma tumor cell lines. Red, amplifications; green, deletions. Chromosome labels are indicated using two shades of gray bars on the top of **A** for 100K SNP arrays and on the bottom of **A** for 10K SNP arrays. Eclipse, alterations commonly present in both glioma cell lines and in tumor samples; rectangular shape, unique alterations present only in cell lines or only in tumor samples. **B.** PCA of glioblastoma, established glioma cell lines, and nonglioma tumor cell lines. **C.** HC of glioblastoma, established glioma cell lines, and nonglioma tumor cell lines. GBM, primary gliomas; GCL, glioma cell lines (*in vitro* cultured); GCL_*in vivo*, U87 and U251 intracranial xenografts; NGTCL, nonglioma tumor cell lines.

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



Our analysis indicates that the gene sets significantly underrepresented (Table 1) or overrepresented (Table 2) in our glioma cell lines compared with primary glioblastomas largely overlapped with the expression of gene sets in other *in vitro* propagated nonglioma cancer cell lines. These data are once again consistent with the hypothesis that established cancer cell lines share an underlying molecular similarity more closely related to their *in vitro* culture conditions than to their original tumor type of origin.

To assess if tissue culture conditions alone could explain these large mRNA expression differences, we compared the expression profiles of glioma cell lines grown as i.c. xenografts in immunodeficient mice with that of the same cell lines grown *in vitro*. U87 and U251 glioma cell–derived i.c. xenografts did show some altered gene set expression compared with their *in vitro* grown counterparts, including down-regulation of the interleukin signaling pathways, insulin-like growth factor signaling pathway, and tRNA synthesis pathways, gene sets overexpressed in *in vitro* glioma cell lines (Supplementary Tables S7, S8, and S9). Alternately, intracranially grown glioma cells up-regulated expression of some gene sets that are also seen overexpressed in primary tumors, such as cell adhesion and G protein–coupled receptor signaling genes. Nevertheless, the overall expression profiles of the xenografts were overwhelmingly similar to the cell lines grown *in vitro*, showing overrepresentation of proliferative gene sets (i.e., cell cycle and *hTERT*-associated genes), with little resemblance to primary human gliomas (Table 1; Supplementary Table S7). These data, taken together with the observation that there were few significant differences in gene set expression between glioma and nonglioma cell lines grown under the same *in vitro* conditions, bring into question the biological relevance of standard glioma cell lines as models for primary human gliomas (Supplementary Table S10).

## Discussion

Established cancer cell lines have long been regarded as the primary reagent for elucidating the complex molecular pathways that contribute to tumorigenesis secondary to their immortal nature, ability to generate tumors in immunodeficient animals, and their relative ease of propagation and genetic manipulation. The underlying assumption behind their use is that these cell lines represent reasonable model systems for the biology of the primary human tumors from which they were originally derived. It has been known for some time, however, that most of the long-term established glioma cell lines are nontumorigenic. Furthermore, the few glioma lines that are tumorigenic form i.c. tumors composed of monomorphic, noninfiltrating cells that bare little resemblance to the highly heterogeneous and infiltrating cells prototypic for human malignant gliomas. Concerns for the meaning of the differences in the biological phenotype of these cell lines compared with primary tumors are further confounded by the documented lack of overexpression of the epidermal growth factor receptor in most glioma cell lines, a prototypic finding in ~80% of primary human glioblastomas. Other individual genetic changes, such as *p16* silencing and loss of *PTEN* in serially

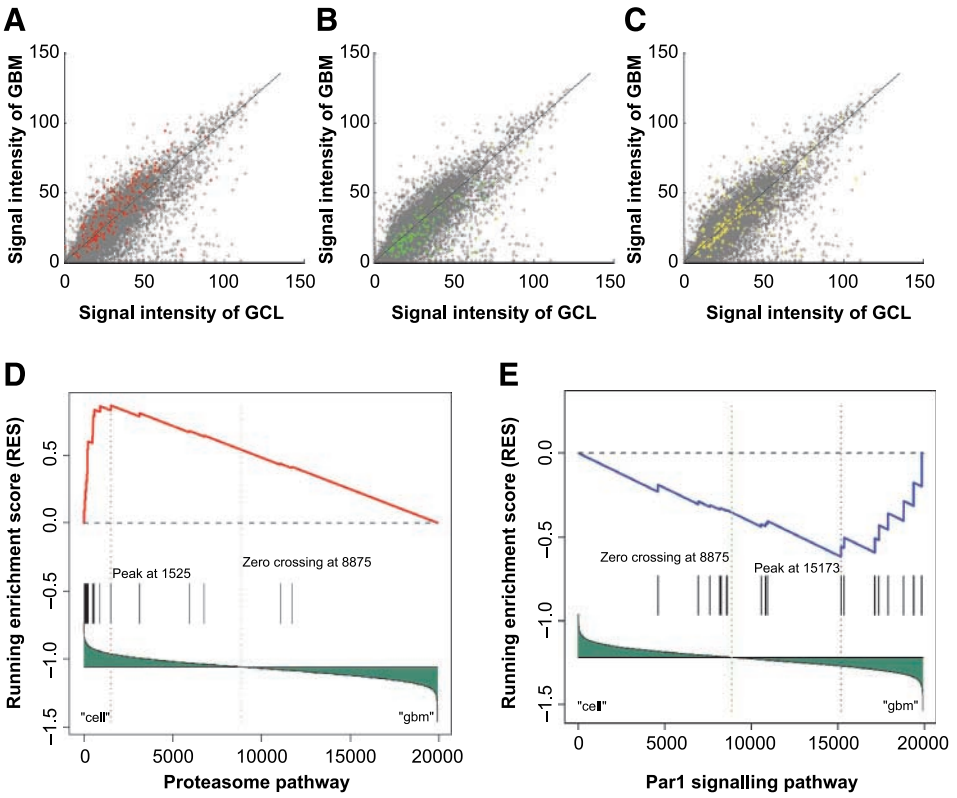


**FIGURE 4.** Results of GSEA for the phenotypes of glioblastomas and cell lines. **A.** Genes in significantly up-regulated gene sets in cell lines (*red*) in all data (*gray*). **B.** Genes in significantly up-regulated gene sets in glioblastoma (*green*) in all data (*gray*). **C.** Genes in significantly up-regulated gene sets in glioma representative pathways (*yellow*) in all data (*gray*). Data are transformed to reflect the expression value differences. **D.** Running enrichment scores for proteasome pathway in cell lines, the genes in the set as "hits" against the ranked list of genes (*bottom*) in the expression data set. **E.** Running enrichment scores for par1 signaling pathway in glioblastoma, the genes in the set as hits against the ranked list of genes (*bottom*) in the expression data set.

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



**Table 1. Summary of GSEA of Glioblastomas Compared with Established Cell Lines**

| Gene sets | Genes | All cell lines | | Glioma cell lines | | *In vivo* | *In vitro* |
|---|---|---|---|---|---|---|---|
| | | ES | Nom p | ES | Nom p | Nom p | Nom p |
| BioCarta | | | | | | | |
| fatty_acid_metabolism | 25 | −0.586 | 0.008* | −0.568 | 0.013* | No † | 0.724 |
| badPathway | 22 | −0.496 | 0.019 ‡ | −0.458 | 0.034 ‡ | No † | 0.933 |
| dcPathway | 20 | −0.525 | 0.028 ‡ | −0.474 | 0.095 | No † | 0.505 |
| par1Pathway | 18 | −0.530 | 0.106 | −0.620 | 0.011* | 0.766 | |
| calcineurinPathway | 18 | −0.539 | 0.047 ‡ | −0.493 | 0.145 | No † | 0.337 |
| nfatPathway | 52 | −0.397 | 0.048 ‡ | −0.358 | 0.144 | No † | 0.490 |
| Gene Ontology | | | | | | | |
| GPCRs_Class_B_Secretin-like | 24 | −0.467 | 0.026 ‡ | −0.480 | 0.016 ‡ | 0.027 ‡ | No |
| Exp. sig. | | | | | | | |
| cell_adhesion | Exp. sig. | −0.436 | 0.002* | −0.385 | 0.009* | 0.042 ‡ | No |
| ANTI_CD44_up | 23 | −0.483 | 0.049 ‡ | −0.353 | 0.397 | 0.023 ‡ | No |
| STKE and others | | | | | | | |
| ST_adrenergic | 31 | −0.447 | 0.068 | −0.458 | 0.034 ‡ | No † | 0.21 |
| SIG_PIP3_signaling_in_B_lymphocytes | 35 | −0.491 | 0.037 ‡ | −0.428 | 0.1371 | No † | 0.133 |

Abbreviations: ES, enrichment score; Nom p, nominal *P* value; Exp. sig., expression signatures.
*Nominal *P* value of <0.01.
†Gene set is not in the list.
‡Nominal *P* value of <0.05.

passaged cell lines, add further to the concern that glioma cell lines may not optimally represent primary human gliomas (11, 13).

The tumorigenic phenotype is a result of interplay between numerous genetic and epigenetic changes to an original tissue-specific target cell. Because the majority of biological properties of a primary human tumor cannot be readily modelled *in vitro*, cancer cell–specific genomic changes and gene expression profiles should represent reasonable surrogate markers for biological properties of a primary tumor. A detailed survey of the genomic aberrations in glioma cells compared with primary gliomas has not been possible until now, secondary to the lack of high-resolution genomic data on a large number of primary gliomas. Although Kubota et al. (2) had attempted to produce a genomic atlas of some of these cell lines using a standard metaphase CGH methodology, their study was limited by inclusion of redundant cell lines (U373 and SNB-19), the failure to include the most commonly used glioma cell line (U87), and most importantly a low-resolution genomic survey methodology (standard CGH) with only a small reference primary glioma database for comparison.

We have extended the work of Kubota et al. by doing a high-resolution genomic survey of the most commonly used and publicly available glioma lines. We used a high-density SNP array for its high-resolution and ability to discriminate more detailed genomic alterations than can any standard CGH method while also providing information on copy number neutral allelic imbalances, which cannot be seen using CGH methods. The power of our analysis to compare and contrast established glioma cell lines to primary tumors is enhanced by

**Table 2. Summary of GSEA of Established Cell Lines Compared with Glioblastomas**

| Gene sets | Sources | Genes | All cell lines | | Glioma cell lines | |
|---|---|---|---|---|---|---|
| | | | ES | Nom p | ES | Nom p |
| Proteasome activities | | | | | | |
| Proteasome_Degradation | BioCarta | 31 | 0.787 | 0* | 0.787 | 0* |
| proteasomePathway | BioCarta | 21 | 0.902 | 0* | 0.863 | 0.002 † |
| Cell cycle activities | | | | | | |
| CR_cell_cycle | Exp. sig. | 77 | 0.562 | 0.002 † | 0.510 | 0.051 ‡ |
| Cell_cycle | na | 77 | 0.615 | 0* | 0.523 | 0.091 |
| cellcyclePathway | na | 21 | 0.560 | 0.011 † | 0.413 | 0.283 |
| Purine_metabolism | GenMapp | 81 | 0.477 | 0* | 0.414 | 0.029 ‡ |
| HTERT_up | Exp. sig. | 98 | 0.563 | 0.017 ‡ | 0.558 | 0.036 ‡ |
| Mitochondrial activities | | | | | | |
| PGC | Exp. sig. | 339 | 0.427 | 0.037 ‡ | 0.413 | 0.052 ‡ |
| Mitochondrial | Exp. sig. | 397 | 0.456 | 0.054 ‡ | 0.402 | 0.141 |
| Others | | | | | | |
| RAP_down | Exp. sig. | 187 | 0.606 | 0.015 ‡ | 0.570 | 0.05 ‡ |
| LEU_down | Exp. sig. | 138 | 0.609 | 0.016 ‡ | 0.532 | 0.105 |
| GLUT_down | Exp. sig. | 249 | 0.595 | 0.01 † | 0.542 | 0.078 |

Abbreviation: na, data source information not available.
*Nominal *P* value of <0.001.
†Nominal *P* value of <0.01.
‡Nominal *P* value of <0.05.

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



our recently published high-resolution SNP array genomic survey of 178 primary gliomas (5). The results obtained from our study correlate to a large extent with those reported by Kubota et al. while providing a higher level of resolution, allowing us to further define areas of interest and discovering small, novel regions of alteration (Fig. 1; Table 1) that could not be detected by the low-resolution CGH technology (e.g., chromosome 1p12 amplification in U251 and chromosome 9p21.3 deletion in A172 and U251).

In addition to finding novel CNA in glioma cell lines, SNP analysis allowed us to define new areas of glioma cell loss of heterozygosity (Fig. 2; Table 2), revealing that the regions of deletion correlate quite well with those showing allelic imbalance (e.g., chromosomes 6, 10, and 14). Nevertheless, the loss of heterozygosity analysis revealed regions where heterozygosity was lost without compromising the DNA copy number (copy number neutral events and gene conversion), suggesting an active recombination machinery in these cell lines. Several of these gene conversions are found in each cell line, some of which are recurrent as in the case of chromosome 9p21.1-24.1 and chromosome 6 qcen-q27.

If genomic alterations do represent surrogate markers for biological relatedness, then the changes found in our glioma cell lines should be commonly represented in primary human gliomas. Indeed, many of the genomic alterations found in primary glioblastomas were also found in our cell lines, consistent with the glioma cell lines having originated from gliomas. Nevertheless, we found numerous genomic alterations in the glioma cell lines that are not found in primary glioblastomas. These alterations included extremely common amplifications of chromosome 5 (80% of glioma cell lines), amplifications of 1q21-44 and 2q34-37.1 (40%), and deletions of 11q14-32.3 and 18q11-23 (40%), none of which is seen in primary glioblastomas. Moreover, a wide variety of large-scale, nonrecurrent alterations are seen in the cell lines in chromosomes that rarely are represented in the primary tumors. Striking among these is a complex spectrum of alterations seen in chromosomes 2, 3 (three of five cell lines), and 6 (100% of the glioma cell lines), where large amplifications mix with deletions, whereas the same chromosome area is never substantially altered in primary gliomas.

Similarly, frequent alterations encountered in GBMs are absent in the established cell lines, as is the case for deletion of chromosomes 13 and 22 and amplification of chromosome 20. Moreover, even when alterations are found in both cell lines and primary gliomas, their characteristic of these alterations tends to differ substantially as in the case of chromosome 7 amplification. In primary gliomas, this amplification is known to result in a large increase of the DNA copy number around the *epidermal growth factor receptor* gene (six copies or more in 40% of our primary samples) with a concomitant increase in mRNA levels. In the established cell lines, however, amplification is more modest (three or four copies) and is generally not associated with a substantial increase in mRNA (Supplementary Fig. S1). Taken together, all these data suggest that established cell lines do not faithfully represent the genomic profiles of primary gliomas.

To help determine whether the genomic divergence of glioma cell lines from primary gliomas is a feature unique to gliomas or is a more general property imposed by *in vitro* culture conditions, we profiled 10 additional nonglioma tumor cell lines using SNP arrays and found that 9 of the 10 nonglioma cell lines showed fragmental or whole chromosome 5 amplification like those found in the glioma cell lines. Likewise, chromosome 18 deletions were also frequently seen in both glioma and nonglioma cell lines, suggesting that the cell lines undergo a set of stereotypic genomic alterations attributable to *in vitro* culture conditions rather than to primary tumor biology.

Genomic changes are static and are only an indirect indicator of cellular biology. Gene expression is a more direct and dynamic process for assessing cell function secondary to both genetic and epigenetic influences. Thus, in addition to the comparative genomic surveys, we chose to examine the gene expression profiles of all the cell lines and the primary gliomas. Unsupervised PCA and HC of global gene expression profiles confirmed the divergence of all cultured cell lines from primary tumors on the transcriptomic level, suggesting that *in vitro* cultured nonglioma tumor cell lines are more similar to glioma cell lines than are primary gliomas.

We used GSEA, an analytic method for comparative analysis of enrichment in functional relevant gene sets over multiple data sets, to better characterize the biological significance of the observed divergence in gene expression between cell lines and primary tumors. Gene sets representing canonical oncogenic signaling pathways (*vascular endothelial growth factor receptor* and *epidermal growth factor receptor*) were unregulated in primary tumors as well as in both glioma and nonglioma cell lines consistent with a "cancer signature". Established cell lines, however, had a large number of unregulated gene sets not found in primary tumors, such as those involved in proteasome signaling and degradation activities, nucleotide metabolism, and oxidative phosphorylation pathways (14). These are gene sets that are involved in high cellular turnover, consistent with a highly replicative phenotype as one would expect from rapidly growing cancer cell lines where ~100% of the cells are undergoing active cell cycle at any given time. Glioblastomas, on the other hand, generally have proliferative indices of only about 7% to 14% *in situ*, and thus, it is not surprising that gliomas do not exhibit the same replicative expression profiles seen in the cancer cell lines. By contrast, certain gene expression patterns, such as those representing tissue structure (i.e., cell adhesion), are up-regulated in primary gliomas but not in cell lines. Interestingly, the differences in gene set expression between glioblastomas and cell lines are generally independent of the cell line tissue of origin. This suggest that much of the biology of established cell lines has been subverted by *in vitro* selection pressure for proliferative capacity and in so doing has become less representative of the original tumor of origin. A proteomic analysis, looking specifically for active signaling pathways, would further enrich the type of analysis presented here.

We believe that these data should give pause to investigators who routinely use established glioma cell lines to explore basic glioma biology and for the use of these cells in preclinical screens for therapeutic agents. It is of interest that the replicative phenotype is the predominant expression pattern in these cell lines and that most drug screens have historically been

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



established to identify agents with cytotoxicity against rapidly dividing cells. Thus, it is conceivable that drugs have historically been selected largely for a biological property (replication) that is more a function of cell culture than of the primary tumor being targeted. It, therefore, may not be all that surprising that most chemotherapeutic agents with *in vitro* activity against glioma cell lines cause toxicity to normal highly proliferative cells in patients (i.e., bone marrow and gastrointestinal epithelium) but have minimal clinical activity against gliomas. Increasingly, novel therapeutic agents are being developed to inhibit specific molecular targets and signaling pathways. Given the differences we have shown between glioma cell lines and primary tumors, it is highly unclear whether glioma cell lines will have any predictive value as screens for these newer molecularly targeted agents.

In conclusion, the strong divergence at genomic, transcriptomic, and functional pathway levels suggests that the prevailing view that glioma cells can be used as a reasonable biological representation of primary human gliomas must be viewed with caution and newer models of human cancer, such as genetic mouse models and tumor stem cells, will need to be carefully considered when designing experiments to ask specific biological and clinically related questions.

## Materials and Methods

### *Cell Lines and Tumor Samples*

Five commonly used glioma cell lines (A172, Hs683, T98G, U251, and U87) and 10 nonglioma tumor cell lines, including 2 small cell lung cancer lines (NCI-H719 and NCI-H865), 4 non–small cell lung cancer cell lines (NCI-H23, NCI-H441, A549, and NCI-H460), 2 breast cancer cell lines (T47D and MDA-435), 1 osteosarcoma cell line (U205), and 1 melanoma cell line (LOX 1MV1), were used for this study. All of the glioma cell lines were obtained from the American Type Culture Collection repository, whereas the nonglioma cell lines were provided by the Laboratory of Human Carcinogenesis at the National Cancer Institute. U87, A172, and T98G were originally described as glioblastomas, whereas U251 and Hs683 were considered as astrocytomas. The cell lines were cultured in the DMEM containing 10% fetal bovine serum. Eighty-three glioblastomas were provided by the Hermelin Brain Tumor Center, Departments of Neurology and Neurosurgery at the Henry Ford Hospital, along with snap-frozen sections of areas immediately adjacent to the region used for the histopathologic diagnosis.

### *DNA Extraction and Hybridization*

Genomic DNA from all cultured cells was prepared using the QIAmp DNA kit (Qiagen) following the manufacturer's instructions. Mapping 10K Xba Array 131 (GeneChip Mapping 10K Array, Affymetrix, Inc.) was used for cell lines in this study, whereas GeneChip Human Mapping 100K (15) arrays (Affymetrix) were used for the primary tumor samples. The mapping array hybridization was done according to the manufacturer's instruction with minor modifications (5). After hybridization, the chips were processed using fluidics station 450, high-resolution microarray scanner 3000, and GeneChip Operating Software workstation version 1.3.

### *RNA Extraction and Hybridization*

Approximately $10^6$ cells or 50 mg of tissue were used to extract total RNA using the Trizol reagent (Invitrogen) following the manufacturer's instructions. The quality of RNA obtained was verified with the Bioanalyzer System (Agilent Technologies; ref. 16) using the RNA Pico Chips. RNA (6 μg) was processed for hybridization on the GeneChip Human Genome U133 Plus 2.0 Expression arrays (Affymetrix; ref. 17) according to the manufacturer's recommendations. After hybridization, the chips were processed using fluidics station 450, high-resolution microarray scanner 3000, and GeneChip Operating Software workstation version 1.3.

### *Genomic Aberration Profiling Data Analysis*

The quality of the 10K and 100K SNP arrays was controlled by the genotype call rates from GeneChip DNA Analysis Software outputs. The call rates across all cell lines ranged from 85% to 98%, whereas tissues ranged from 93% to 99%. The genotype calls for each SNP were generated using dynamic model mapping algorithm from Affymetrix GeneChip DNA Analysis Software (version 3.0).

*Copy Number Determination.* The individual chromosome copy number for each SNP was determined using chromosomal copy number tool (version 2.0) software from Affymetrix (18). The significance of copy number for a given SNP was determined by comparison with a reference pool of normal samples within the same region and same genotype. The chromosome copy number outputs for 10K SNP arrays were smoothed using a circular binary segmentation algorithm implemented in DNAcopy package in R. The circular binary segmentation is a modified version of binary segmentation by introducing an "undo" function. Hence, it identifies the genomic break points recursively and transforms the individual signal intensity measurement into regions of equal copy number relevant to regions of genomic amplification or deletion (19). Additionally, CGHAnalytics software (version 3.1) from Agilent Technologies was used to help identifying CNAs in areas in which the circular binary segmentation did not do well. This package uses an aberration detection method 1 (or "adam-one") algorithm for finding changes that identifies the DNA CNAs according to the derivative log ratio spread. For the tumor tissues, the chromosomal copy number tool output was then smoothed using hidden Markov model algorithms implemented in R package of array CGH (20, 21).

*Loss of Heterozygosity Analysis of Cell Lines.* The loss of heterozygosity was determined using the log probability of a continuous genomic region of homozygous calls generated by chromosomal copy number tool and applied a threshold of 6% to the probability of being homozygous (18). All the annotations were established according to the human genome build 34.

### *Gene Expression Profiling Data Analysis*

The .cel files and detection calls from MAS5 files were imported into dChip and normalized using the invariant set algorithm and the perfect match/mismatch difference model was used for determining model-based expression values. Arrays with percentage of outliers >5% were excluded for data

Published OnlineFirst January 9, 2008; DOI: 10.1158/1541-7786.MCR-07-0280



analysis. Additionally, two independent filters were applied to the normalized and cleaned data set (i.e., filtering off the probe sets with "present" calls <10% of the total number of samples or probe sets containing >10% of the zeros across samples that flag the probe sets with signal intensities of mismatch > perfect match). This step reduced the number of probe sets from 54,000 to 34,482. The unsupervised PCA and HC were used to reduce the dimensionality of the data and detect the dominated structure of the data variation. HC was done using distance matrix of Pearson correlation with average linkage method and PCA with correlation dispersion matrix and normalized eigenvector scaling using Partek software version 6.2 (Partek, Inc.). Additionally Matlab scripts and stand-alone Perl scripts were written to facilitate the data analysis and filtering.

### *GSEA for Functional Annotation*

The GSEA algorithm maps an independent gene set with known function to a ranked list of genes ordered according to the correlation to the phenotypes of interest and determines whether a defined gene set is concordantly located to the extreme top or bottom in the given data set measured with running enrichment scores. GSEA was done according to the procedures previously described (12) by contrasting the phenotypes of glioblastomas versus all cell lines, glioblastomas versus glioma cell lines, glioma cell lines versus nonglioma tumor cell lines, and *in vivo* versus *in vitro* glioma cell lines separately. Five hundred and twenty-six gene sets (475 gene sets for metabolic and signaling pathways and 51 gene expression signatures) in MSigDB v1.0 were used for GSEA analysis (12).

## Acknowledgments

We thank Drs. Ana Robles and Jeongwu Lee for the generous provision of cell lines and tumor material for this work.

## References

1. Bredel M, Bredel C, Juric D, et al. High-resolution genome-wide mapping of genetic alterations in human glial brain tumors. Cancer Res 2005;65:4088–96.

2. Kubota H, Nishizaki T, Harada K, et al. Identification of recurrent chromosomal rearrangements and the unique relationship between low-level amplification and translocation in glioblastoma. Genes Chromosomes Cancer 2001;31:125–33.

3. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998;396:643–9.

4. Barbashina V, Salazar P, Holland EC, Rosenblum MK, Ladanyi M. Allelic losses at 1p36 and 19q13 in gliomas: correlation with histologic classification, definition of a 150-kb minimal deleted region on 1p36, and evaluation of CAMTA1 as a candidate tumor suppressor gene. Clin Cancer Res 2005;11:1119–28.

5. Kotliarov Y, Steed ME, Christopher N, et al. High-resolution global genomic survey of 178 gliomas reveals novel regions of copy number alteration and allelic imbalances. Cancer Res 2006;66:9428–36.

6. Maruno M, Yoshimine T, Muhammad AK, Ninomiya H, Kato A, Hayakawa T. Chromosomal aberrations detected by comparative genomic hybridization (CGH) in human astrocytic tumors. Cancer Lett 1999;135:61–6.

7. Arslantas A, Artan S, Oner U, et al. The importance of genomic copy number changes in the prognosis of glioblastoma multiforme. Neurosurg Rev 2004;27:58–64.

8. Weber RG, Sabel M, Reifenberger J, et al. Characterization of genomic alterations associated with glioma progression by comparative genomic hybridization. Oncogene 1996;13:983–94.

9. MacLeod RA, Dirks WG, Drexler HG. Persistent use of misidentified cell lines and its prevention. Genes Chromosomes Cancer 2002;33:103–5.

10. Matsuzaki H, Loi H, Dong S, et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res 2004;14:414–25.

11. Ishii N, Maier D, Merlo A, et al. Frequent co-alterations of TP53, p16/CDKN2A, p14ARF, PTEN tumor suppressor genes in human glioma cell lines. Brain Pathol 1999;9:469–79.

12. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.

13. Schmidt EE, Ichimura K, Goike HM, Moshref A, Liu L, Collins VP. Mutational profile of the PTEN gene in primary human astrocytic tumors and cultivated xenografts. J Neuropathol Exp Neurol 1999;58:1170–83.

14. Mootha VK, Bunkenborg J, Olsen JV, et al. Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. Cell 2003;115:629–40.

15. Matsuzaki H, Dong S, Loi H, et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nat Methods 2004;1:109–11.

16. Miller CL, Diglisic S, Leister F, Webster M, Yolken RH. Evaluating RNA status for RT-PCR in extracts of postmortem human brain tissue. Biotechniques 2004;36:628–33.

17. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. Nat Genet 1999;21:20–4.

18. Huang J, Wei W, Zhang J, et al. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Hum Genomics 2004;1:287–99.

19. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004;5:557–72.

20. Fridlyand J. SA, Pinkel D, Albertson DG, Jain A. Application of hidden Markov models to the analysis of the array CGH data. J Multivariate Anal 2004;90:132–53.

21. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.

AACR American Association for Cancer Research

Molecular Cancer Research

# Genomic Changes and Gene Expression Profiles Reveal That Established Glioma Cell Lines Are Poorly Representative of Primary Human Gliomas

Aiguo Li, Jennifer Walling, Yuri Kotliarov, et al.

*Mol Cancer Res* 2008;6:21-30. Published OnlineFirst January 9, 2008.