Integrated Array-Comparative Genomic Hybridization and Expression Array Profiles Identify Clinically Relevant Molecular Subtypes of Glioblastoma


Departments of Neurological Surgery (Brain Tumor Research Center) and Laboratory Medicine, and San Francisco General Hospital General Clinical Research Center, University of California, School of Medicine, San Francisco, California; and Departments of Pathology, Neuro-Oncology, Biostatistics, and Brain Tumor Center, University of Texas M.D. Anderson Cancer Center, Houston, Texas

Abstract

Glioblastoma, the most aggressive primary brain tumor in humans, exhibits a large degree of molecular heterogeneity. Understanding the molecular pathology of a tumor and its linkage to behavior is an important foundation for developing and evaluating approaches to clinical management. Here we integrate array-comparative genomic hybridization and array-based gene expression profiles to identify relationships between DNA copy number aberrations, gene expression alterations, and survival in 34 patients with glioblastoma. Unsupervised clustering on either profile resulted in similar groups of patients, and groups defined by either method were associated with survival. The high concordance between these separate molecular classifications suggested a strong association between alterations on the DNA and RNA levels. We therefore investigated relationships between DNA copy number and gene expression changes. Loss of chromosome 10, a predominant genetic change, was associated not only with changes in the expression of genes located on chromosome 10 but also with genome-wide differences in gene expression. We found that CHI3L1/YKL-40 was significantly associated with both chromosome 10 copy number loss and poorer survival. Immortalized human astrocytes stably transfected with CHI3L1/YKL-40 exhibited changes in gene expression similar to patterns observed in human tumors and conferred radioresistance and increased invasion in vitro. Taken together, these results indicate that integrating DNA and mRNA-based tumor profiles offers the potential for a clinically relevant classification more robust than either method alone and provides a basis for identifying genes important in glioma pathogenesis. (Cancer Res 2005; 65(5): 1678-86)

Introduction

Glioblastoma is the most common intrinsic brain tumor in adults and is one of the most aggressive human neoplasms. The median survival for a patient with newly diagnosed glioblastoma is ~1 year, despite maximal surgical and medical intervention (1). Less than 5% of patients are alive 3 years following diagnosis (2). Older age at diagnosis is a negative prognostic factor (3), and molecular markers differ between older and younger patients (4, 5), suggesting that glioblastoma has age-associated molecular sub-classes. With expression profiling, larger groups of genes have been probed in human gliomas. More detailed molecular subtypes of malignant gliomas are now evident (6, 7) and genes specifically associated with tumor grade and progression have come to light (8, 9), including those between glioblastoma and anaplastic oligodendroglioma (10). However, relationships between DNA copy number, gene expression, and clinical behavior of glioblastoma glioma subtypes are not well understood.

Using conventional CGH, we have identified copy number aberrations (CNA) associated with survival in a global genomic screen of typical and long-term glioblastoma survivors (11). Here we evaluate previously untreated frozen human tumor samples from patients with known clinical outcomes using array-comparative genomic hybridization (aCGH) combined with gene expression array analysis. Together, these approaches identify specific DNA copy number aberrations which may be used for clinically relevant and robust molecular classification, and point to candidate genes associated with these aberrations which may influence biologic and clinical behavior of these tumors.

Materials and Methods

Case Selection. We obtained 34 snap-frozen glioblastoma (WHO grade 4 astrocytoma) samples (enriched for long-term survivors) from the tissue bank at the Brain Tumor Research Center in the Department of Neurological Surgery, University of California, San Francisco. All cases were newly diagnosed, and patients received no therapy before sample collection. Long-term survivors (LTS) were alive >2 years following diagnosis, and short-term survivors (STS) were cases with <2 years of survival. All survival analysis used this cutoff, except subgroups defined by aCGH, where we analyzed survival data based on median survivals. All cases were from patients who underwent gross total resection as assessed by post-operative imaging studies. All patients completed conventional external beam radiation therapy at a single institution (University of California San Francisco) assuming sufficient survival time after surgery. The median Karnofsky performance score was 90 in each group and did not differ between the LTS and STS groups (P = 0.7). To eliminate age as a prognostic factor in these studies, we matched the median age of STS and LTS. For each case, histologic assessment of the tumor tissue used was done by a neuropathologist (K. Aldape). Glioblastoma tumor samples contained at least 90% tumor as assessed by a section cut from the tissue block used for RNA and DNA isolation. Five samples of nonneoplastic brain taken from patients undergoing temporal lobe epilepsy surgery were used for comparison. All patients signed consent for specimens to be used for research purposes. Overall survival was determined as the interval from initial surgery to death and was obtained from either the patient’s medical...
RNA Isolation. Total RNA was prepared from tumor samples immediately stored in liquid nitrogen following surgical resection. One-hundred milligrams of tissue were ground into a fine powder with a mortar and pestle and remained cool on dry ice. We isolated RNA using TRIzol reagent (Life Technologies, Gaithersburg, MD) to lyse the tissue. Four milliliters of solution per 100 mg of tissue were necessary for adequate aqueous and organic phase separation. The mixture was homogenized by passing it through a 21-gauge needle three to four times. RNA was precipitated from the aqueous phase with an equal volume of isopropanol, and the dried pellet was resuspended in RNase-free water. The quality of the RNA was assessed on 1% agarose, MOPS-formaldehyde gels. Fifty micrograms of total RNA from each tissue sample were then passed over an RNeasy column (Qiagen, Chatsworth, CA) for further cleanup.

DNA Isolation. We isolated DNA either from TRIzol fractions remaining after the RNA isolation procedure or from frozen tumor treated with proteinase K. DNA was extracted with phenol/chloroform/isooamyl alcohol, precipitated, and resuspended in 10 mmol/L Tris-HCl, 0.1 mmol/L EDTA (pH 8.0).

cDNA and cRNA Synthesis for Affymetrix Expression Arrays. First strand cDNA synthesis was done on 15 μg of total RNA in a 20-μl reaction with 400 units of Superscript II (Life Technologies) and 100 pmol of a primer containing a promoter sequence for T7 RNA polymerase joined to a string of 24 dTTPs. Reaction components for second strand synthesis were added directly to the first strand reaction. Double stranded cDNA was purified by a phenol/chloroform/isooamyl alcohol extraction followed by precipitation. The pellet was resuspended in 12 μl of RNase-free water. Biotinylated cRNA was synthesized employing the Enzo BioArray kit (Enzo) from 5 μl of cDNA, and unincorporated nucleotides were removed by passing reactions over RNeasy columns (Qiagen). The cRNA was quantified after removal from the column and then fragmented at a concentration of 0.5 μg/μl at 94 °C in buffer containing 40 mmol/L Tris-acetate (pH 8.1), 100 mmol/L KOAc, and 30 mmol/L MgOAc.

Hybridization and Data Analysis. Test arrays were used first to screen for cDNA quality, and 15 μg of cRNA were used in the hybridizations to U95Av2 human GeneChip expression arrays done according to the specifications of the manufacturer (Affymetrix). Intensity data were obtained from array images using Perfect Match as previously described (12). Expression data for ~5% of the probe sets were related to the batch in which the samples were processed, and these were excluded from further analysis. Data were also analyzed using Affymetrix Microarray Suite (MAS 5.0). Results were similar for the two methods. The primary data (.CEL files) obtained from the hybridization images are located at http://www.mrdanderson.org/labs/adape.

Array-Comparative Genomic Hybridization, DNA Labeling, and Image Analysis. Whole genome arrays of 2,464 mapped bacterial artificial chromosome (BAC) clones were hybridized simultaneously with ~500 ng of tumor DNA and reference DNA from normal tissue (13). Tumor DNAs were labeled with Cy3-dCTP and reference DNAs with Cy5-dCTP by random priming (14). The spotted BACs were counterstained with 4′,6′-diamino-2-phenylindole hydrochloride. Arrays were scanned and images processed using custom software (15). We normalized relative ratios of tumor and normal signals by setting the value of the median relative ratio equal to 1. The data were then transformed into log 2 space and plotted as a histogram to determine cutoffs for scoring loss or gain. Three Gaussian distribution curves were fitted to the histogram, and values >3 SD from the central Gaussian were scored as losses or gains for that tumor.

Statistical Analysis. Unsupervised hierarchical clustering was done on both aCGH and expression microarray data using Cluster and TreeView software (16). aCGH data were Z-transformed using characteristics of the central peaks of the histograms described above. Significance Analysis of Microarrays (SAM) was used to identify statistically significant correlations between genetic/expression data and clinical variables (17). For aCGH data analysis, the raw data were scaled as loss (~1), no change (0), or gain (+1) by the rules described above, and loci without CNAs among all 34 samples were excluded due to the SAM algorithm requirement of variation for each locus. Prognostic significances of copy number gain and loss were assessed separately. For expression array data, prior experience suggested that gene expression measurements were least reliable in genes with lower expression (12). For supervised analysis of the expression data set, we therefore included only probe sets where median expression level was in the top 50th percentile (6,226/12,453). Two-class analysis of this subset identified genes associated with either STS or LTS. SAM output data are presented along with the false discovery rate. Where possible, the lowest false discovery rate was used, but in all cases a false discovery rate ≤20% was required to report a gene output list.

Array-Comparative Genomic Hybridization Expression Correlations. We assigned unambiguously mapped BACs and their associated relative DNA copy number to 5 MB bins (n = 554) along each chromosome (University of California, Santa Cruz human genome database July 2003 freeze [http://genome.ucsc.edu/]). If there were multiple BACs in a bin, their copy number values were averaged for each sample. The final list of 4,132 genes was derived from the 6,226 genes in the upper 50th percentile of overall expression by ensuring that each gene was represented only once and unambiguously mapped to a nucleotide position on a chromosome. The correlation between each CGH bin and the expression value for every gene was determined using Pearson’s correlations to identify relationships between CNA and gene expression. Kendall’s τ correlations were also calculated and gave qualitatively similar results.

We estimated relationships between CNA and gene expression by mapping each of the 4,132 genes to one of the 554 CGH bins and correlating the DNA copy number with expression for each gene localized to the bin. We estimated correlations of CNAs with global gene expression correlations using the least-squares ($R^2$) method. $R^2$ values were obtained by averaging the square of the Pearson’s coefficients for each bin to determine the variability in gene expression explained by CNAs.

Real-time Quantitative PCR for Expression. cDNA synthesis for real-time quantitative PCR was done in a 100-μl reaction containing 1× PCR buffer (Applied Biosystems, Foster City, CA), 7.5 mmol/L MgCl2, 1 mmol/L each deoxynucleotide triphosphate, 5 μmol/L hexamers (Life Technologies), RNase inhibitor (0.4 units/μl; Roche, Indianapolis, IN), and MultxRT (2.5 units/μl; Life Technologies). Reactions were incubated at 25 °C for 10 minutes, 45 °C for 40 minutes, and 95 °C for 5 minutes. To show that cDNA synthesis from tumor RNAs was quantitative, input RNA concentrations were varied (500, 250, 125, and 62.5 ng), and 5 μl were used in triplicate in real-time amplification reactions using control primers and a probe for glucuronidase B. The PCR profile was as follows: 10 minutes at 95 °C for 1 cycle; 15 seconds at 95 °C, 1 minute at 60 °C for 40 cycles. Primers and probes for glioblastoma growth factor receptor, aquaporin 1, epithelial membrane protein 3, and glucuronidase B were purchased from Applied Biosystems.

Immunohistochemistry. Validation of protein expression was done using immunohistochemistry as described previously (18). Briefly, 5 μmol/L sections were cut from paraffin blocks, deparaffinized and hydrated through an ethanol series. Antibodies against caveolin-1 (1:1,000; Sigma, St. Louis, MO), aquaporin 1 (1:3,000; Upstate Biotechnology, Lake Placid, NY), and CHI3L1/YKL-40 (1:1,000; Quiel, San Diego, CA) were incubated with slides overnight at 4 °C following microwave antigen retrieval. Staining was visualized using the DAKO Envision kit according to the instructions of the manufacturer (DAKO, Carpinteria, CA).

Subcloning of CHI3L1/YKL-40. CHI3L1/YKL-40 cDNA was amplified from total RNA made from glioblastoma tissues by reverse transcription-PCR (RT-PCR) using forward primer 5′-GGAGAGTCCCTGCTCGTGACCAGCA-3′ and reverse primer 5′-GGTCCGACCTGCTGCTGAC-3′ containing BanHI and SalI restriction sites, respectively. The RT-PCR product was double-digested with BanHI and SalI and ligated into pLNCX-2 mammalian expression vector linearized with the same double digestion. The plasmid was sequence-verified and was transfected into the DAKO Envision kit according to the instructions of the manufacturer (DAKO, Carpinteria, CA).
Invasion, Zymography, and Serum Starvation Assays. We measured invasion by counting the number of cells traversing Matrigel-coated filters in triplicate modified Boyden chambers as previously described (20). The number of cells that traversed the filter was quantified and the average of triplicate wells determined. Gelatin zymography to detect metalloproteinase activity was done as previously described (20).

To compare the ability YKL-40-transfected versus mock-transfected cells to survive following withdrawal of serum, a flow cytometric assay for DNA content was used. Subconfluent cells (500,000 cells per 100-mm dish) were washed thrice in serum-free medium (DMEM) after which the cells were placed in the tissue culture incubator. To analyze DNA content, the cells were harvested by trypsinization, fixed in ice-cold 70% ethanol, and stored at −20°C. They were then washed in PBS stained with propidium iodide, and analyzed in a flow cytometer as previously described (21). The proportion of cells with a sub-G1 DNA content was determined using CellQuest software.

Clonogenic Survival Following Radiation. Clonogenic survival of cultured cells following radiation was done as previously described (22). Briefly, subconfluent cells were harvested and plated in triplicate on 60-mm dishes at varying densities, treated with γ-rays using a 137Cs source (3.7 Gy/min) at doses of 2, 4, and 6 Gy. After 14 days of incubation, the contents of the dishes were stained with 0.5% crystal violet in absolute methanol and colonies with >50 cells were counted. Three vector-transfected clones and three CHI3L1/YKL-40-transfected clones were tested, each clone in triplicate. The results for each clone were averaged among triplicate plates and were normalized for plating efficiency.

Results

Array-Comparative Genomic Hybridization Data and Clinical Outcome. Figure 1A illustrates the two groups that resulted from unsupervised clustering (25) of unfiltered aCGH data. The group in the left branch of the dendrogram (Fig. 1A) was composed predominately of STS cases (15/16), whereas the group in the right branch of the dendrogram was equally populated with STS (n = 9) and LTS (n = 9) cases. The relationship between survival time (STS versus LTS) and cluster group was highly significant (P = 0.008, Fisher’s exact test, odds ratio = 15.0), indicating a close relationship between CNA and survival. Owing to case 103.6 being below but very near our cutoff of 2 years, we also analyzed the data after assigning it to LTS, and the results remained statistically significant (data not shown). The heat map (Fig. 1A) also revealed the genetic patterns of gain and loss that defined the two groups. Cases in the left dendrogram branch had primarily sustained gains involving most of chromosome 7 and/or losses of chromosome 10 (“altered 7/10 cases”). Cases in the right branch did not show these changes and, with the exception of chromosome 13 loss, were without an evident pattern (“intact 7/10 cases”). Correlations among CNAs by chromosomal location (Fig. 1B) indicated the expected associations of specific CNAs with additional CNAs in neighboring regions along the same chromosome (boxes along diagonal line) as well as the inverse relationships of chromosome 10 status with chromosomes 7, 19, and 20 (loss of 10 versus gains of 7, 19, and 20).

We used SAM software to compare CNAs in 24 STS with those in 10 LTS. As expected from inspection of the unsupervised clustering data, most survival-associated CNAs were located on chromosomes 7 and 10 (Supplementary Table 1). There were 126 BACs of which gain was associated with the STS and 100 of these were localized to chromosome 7, with the remaining BACs mapping to chromosomes 2, 19, and 20 (Supplementary Table 1, left). A similar analysis of losses revealed 75 BACs associated with the STS, of which 60 mapped to chromosome 10, with the remaining BACs mapping to chromosomes 2, 4, 5, 9, 13, and 14 (Supplementary Table 1, right). No CNAs were positively associated with the LTS.

Gene Expression Changes in Glioblastoma versus Non-Neoplastic Brain. We analyzed the set of genes comprising the upper 50th percentile in average gene expression across the tumors for overexpressed genes compared with nonneoplastic
brain. As expected, SAM analysis identified overexpression of the genes for epidermal growth factor receptor, vascular endothelial growth factor, and tenasin C (Supplementary Table 2). CHI3L1/YKL-40 was the most highly expressed gene in glioblastoma relative to nonneoplastic brain tissue (15-fold increase).

Gene Expression Changes and Clinical Outcome. To evaluate relationships between gene expression and survival, we chose probes of which expression was both in the highest 50th percentile (log 2 value ≥ 9.0) and highly variable (SD > 90) for unsupervised clustering. Analysis yielded two major branches of a dendrogram (Fig. 2). As with aCGH data, there was a significant association between dendrogram branch and survival. Seventeen of the 18 cases in the left branch were from the STS, whereas the right branch contained nine cases from the LTS and seven cases from the STS (P = 0.002, Fisher's exact test, odds ratio = 21.9). The distribution of the cases across these two groups was remarkably similar to that obtained from the aCGH data clustering (Fig. 1A). Comparing the two dendrograms, 28 of 34 cases were concordantly distributed (P = 0.002, Fisher's exact test, odds ratio = 24.5).

The SAM algorithm identified gene expression changes associated with survival. The genes of which expression was significantly increased in the STS compared with the LTS are shown in Supplementary Table 3 (left). Some of these genes were previously associated with glioblastoma pathogenesis, including vascular endothelial growth factor (23), tenascin-C (24), insulin-like growth factor binding protein 2 (25), and cysteine-rich angiogenic inducer-61 (26) as well as other genes not previously implicated in glioblastoma. Genes underexpressed in STS are shown in Supplementary Table 3 (right).

Validation of Expression Array Results by Real-time Reverse Transcription-PCR and Immunohistochemistry. We verified our microarray expression results by assaying 16 tumor RNAs from our original set with quantitative real-time PCR. We evaluated two genes overexpressed in poorer survivors, aquaporin 1 and epithelial membrane protein 3, as well as epidermal growth factor receptor, a gene known to be amplified/overexpressed in these tumors, and normalized the data to glucuronidase B. Supplementary Fig. 1 shows the relationships between expression array and real-time RT-PCR data. Immunohistochemistry verified and localized protein expression of selected candidates. CHI3L1/YKL-40, caveolin-1, and aquaporin 1 were overexpressed in glioblastoma relative to normal tissue, and the overexpressed proteins were localized in the neoplastic cells (Fig. 3).
In contrast, analysis of the intact 7/10 cases (right dendrogram branch) identified 32 genes associated with survival (Supplementary Table 5). Forty percent (13/32) of these were identical to genes associated with survival in the entire group of 34 cases (Supplementary Table 3).

**Overexpression of CHI3L1/YKL-40 in Cultured Cells.** Because CHI3L1/YKL-40 was highly overexpressed in glioblastoma relative to nonneoplastic brain, strongly associated with survival, and associated with genetic subtype, we investigated its biologic properties in vitro. We compared expression profiles of three clones of immortalized human astrocytes stably transfected with full-length CHI3L1/YKL-40 cDNA with three mock-transfected clones. Table 1 shows genes that differentiate transfectants from controls and genes that differentiate between high- and low-expressing CHI3L1/YKL-40 tumors. There were nine genes common to the list of the top 100 genes with the greatest differences (based on fold change) in CHI3L1/YKL-40 versus mock transfecants and the list of the top 100 genes with greatest differences in CHI3L1/YKL-40 high-expressing versus CHI3L1/YKL-40 low-expressing tumor samples. This result was highly significant \( P < 0.001, \chi^2 \) test. The list of overlapping genes includes several identified as related to patient survival.

**Figure 3.** Validating array data with protein expression. Immunohistochemical staining in human glioblastoma samples was done using antibodies directed against CHI3L1/YKL-40 (A), caveolin-1 (B), and aquaporin 1 (C). Representative images of positively stained tumors are shown (magnification, ×400).

**Figure 4.** Correlation between DNA copy and gene expression. aCGH data were condensed into 5 MB bins (see Materials and Methods) and correlated with expression of 4,132 highly expressed unique genes. A and B, correlations of DNA copy number with global gene expression. A, matrix indicating correlations between each aCGH bin and expression and all 4,132 genes. The correlations are plotted as a function of chromosome position. Shown are Pearson's correlations >0.4 (red) and < -0.4 (blue). B, \( R^2 \) values representing relationships between relative DNA copy number and global gene expression plotted by chromosome position.
depriving cells of serum. The proportion of vector-transfected death following growth factor deprivation was evaluated by increased resistance to radiation as compared with controls. Cell death after serum starvation, on clonogenic survival after X- and resistance to therapy are associated with high grade glioblastoma.

Table 1. CHI3L1/YKL-40-associated genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cells</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>72.8</td>
</tr>
<tr>
<td>Nicotinamide N-n-methyltransferase</td>
<td>11.3</td>
</tr>
<tr>
<td>Collagen, type VI, α1</td>
<td>4.7</td>
</tr>
<tr>
<td>Collagen, type VI, α2</td>
<td>4.7</td>
</tr>
<tr>
<td>Tenascin C (hexabrachion)</td>
<td>3.9</td>
</tr>
<tr>
<td>Basic helix-loop-helix domain containing, class B, 2</td>
<td>2.4</td>
</tr>
<tr>
<td>Epithelial membrane protein 1</td>
<td>2.4</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 5</td>
<td>2.3</td>
</tr>
<tr>
<td>Collagen, type XVI, α1</td>
<td>2.1</td>
</tr>
<tr>
<td>Likely orthologue of rat brain specific binding protein/Ras-induced senescence 1</td>
<td>1.9</td>
</tr>
<tr>
<td>Transforming growth factor-β receptor II</td>
<td>1.8</td>
</tr>
<tr>
<td>Vascular endothelial growth factor</td>
<td>1.9</td>
</tr>
<tr>
<td>Transforming growth factor-β-induced A kinase (PKA) anchor protein 11</td>
<td>1.8</td>
</tr>
<tr>
<td>Cadherin 11, type 2, OB-cadherin (osteoblast)</td>
<td>1.7</td>
</tr>
<tr>
<td>Pre-B-cell colony-enhancing factor</td>
<td>1.7</td>
</tr>
<tr>
<td>ADP-ribosylation factor-like 7</td>
<td>1.6</td>
</tr>
<tr>
<td>KIAA0062 protein</td>
<td>1.6</td>
</tr>
<tr>
<td>Caldesmon 1</td>
<td>1.5</td>
</tr>
<tr>
<td>Plasminogen activator inhibitor type 1</td>
<td>1.5</td>
</tr>
<tr>
<td>Paternally expressed 10</td>
<td>0.5</td>
</tr>
</tbody>
</table>

NOTE: In vivo and in vitro expression patterns for cells overexpressing CHI3L1/YKL-40 were compared. In vivo expression patterns were derived from glioblastoma tumor samples that were divided into two groups based on CHI3L1/YKL-40 expression levels, whereas in vitro profiles were obtained with CHI3L1/YKL-40 and mock-transfected cells. Genes that differed by 1.5-fold in these comparisons and were common to both in vitro and in vivo profiles are shown.

We have profiled changes in genome copy number and gene expression in glioblastoma tumor samples from patients with differing clinical outcomes. aCGH (Fig. 1A) and expression array (Fig. 2) data both indicated that a relationship exists between molecular profiles and survival in glioblastoma patients. The fact that both data sets clustered cases similarly suggested that expression patterns are related to specific CNAs. A correlation matrix between the two data sets helped to define these relationships. As expected, changes in DNA copy number were directly associated with changes in the corresponding genes, and we estimated that this accounted for 6% of the variation of gene expression in these tumors. Surprisingly, specific CNAs were also associated with expression of genes at other loci to a comparable degree. Chromosome 10 status emerged as a primary influence on gene expression across the tumor genome. Whereas the finding of frequent chromosome 10 loss in glioblastoma is well established, evidence to suggest that it is a marker for global changes in gene expression has not been previously reported. These data support the hypothesis that subgroups of glioblastoma, as defined by signature chromosomal aberrations, are also clinically and biologically diverse.

Unsupervised hierarchical clustering of aCGH data revealed two groups of glioblastoma. One group (Fig. 1A, altered 7/10 cases, left dendrogram branch) was composed of patients with almost uniformly poor (<2 year) survival. It contained the most common CNA, loss of chromosome 10. This loss was associated with gain of chromosome 7, consistent with previous data (11, 27–30). Losses of chromosome 10 and gains of 7p were identified as poor prognostic factors in grade 3 astrocytomas, and gain of 7q in the absence of a gain on 7p portended intermediate survival in a previous study (31). Other relatively frequent CNAs in this cluster, such as gains on chromosome 19 and 20, have been associated with poorer survival (11). Thus, our data are consistent with the idea that specific CNAs, and in particular loss of chromosome 10, signify poorer survival. Whereas the finding of chromosome 10 loss as a frequent CNA was largely corroborative, the finding that chromosome 10 loss is a strong predictor of global gene expression signature is a novel finding, suggesting that loss of chromosome 10 is a key step towards defining distinct pathways in the molecular pathogenesis of glioblastoma.

The proportion of glioblastoma cases with loss of chromosome 10 in our set (≈50%) was lower than we previously reported (28). The reasons for this difference are likely related to our selection criteria: the cases reported here had a wide range of survival times to facilitate identification of molecular alterations associated with survival. Given the strong relationship...
between older age and worse survival in malignant gliomas, our set of patients was younger than the median age of patients with glioblastoma. Ongoing studies on a larger set of tumors suggest that a relationship exists between loss of chromosome 10 and older age, which may in part account for the age-dependent survival association in these tumors.

Unsupervised hierarchical clustering of expression microarray data divided glioblastoma tumors into two groups largely concordant with aCGH data (Fig. 2), suggesting that the subclasses identified by CNA signatures extend to the mRNA level. We generated a matrix of Pearson’s correlations across our data sets to examine relationships between DNA and mRNA changes (Fig. 4A and B). This matrix indicated that CNAs on chromosome 10 were most highly associated with gene expression, suggesting that chromosome 10 status is related to distinct expression profiles. As expected from the relationship between chromosome 10 status and survival, many genes most closely associated with chromosome 10 were also associated with survival (Supplementary Tables 3 and 4).

The intriguing result was that for most of the correlations, genomic and expression elements did not map to the same chromosomal loci. Mapping those regions of loss on chromosome 10 which were most closely associated with global gene expression changes was not possible in this study given that most cases which showed loss included nearly all loci on chromosome 10. The mechanisms of relationships between genetic changes on chromosome 10 and changes in expression of genes mapping to regions outside of chromosome 10 remain to be determined, but at a minimum the data indicate widespread differences in expression of genes in glioblastoma based on aCGH profiles.

Genes that were associated with survival included tenascin C (24), vascular endothelial growth factor (32), and insulin-like growth factor binding protein 2 (25). These genes have been implicated in the molecular pathogenesis of glioblastoma and provide a measure of validation for the study. We confirmed previous reports that CHI3L1/YKL-40 (33) and aquaporin 1 (34) were highly expressed in glioblastoma, but our data add to these findings by identifying an association between higher expression levels and poor clinical outcome. Genes reduced in expression in glioblastoma from STS included the glutamate receptor subunit APMA 2, a finding that is consistent with published data which indicated that overexpression of this gene results in loss of glioma cell migration and cell viability (35). Delineating the function of these genes in pathogenesis and resistance to radiotherapy and chemotherapy characteristic of glioblastoma are potential areas of interest for future investigation.

The apparent relationships between genome copy number, gene expression, and survival suggested that subgroups defined by CNA profile might be further characterized by differences in gene expression. We analyzed gene expression differences of LTS and STS within groups identified through unsupervised clustering of CNA data (Fig. 1). We were unable to identify genes associated with survival in the 7/10-altered cases (left dendrogram branch of Fig. 1) even after we modified the survival cutpoints, suggesting that gene expression differences within this group do not account for variation in survival. However, the sample size in this subgroup analysis is small and requires confirmation in a larger study. In addition, the range of survival was relatively small, with the exception of one outlier with long-term survival. Patient survival was more variable in the right dendrogram branch (7/10 intact cases) of Fig. 1A. Although this group of patients was also small, their varied CNAs and survivals suggest greater genetic and clinical heterogeneity than the 7/10-altered cases. This heterogeneity is evident when comparing expression of specific genes to chromosome 10 status as shown in Supplementary Fig. 2. For each of these chromosomes 10–associated genes, expression is uniformly high in the 10-loss cases. In contrast,
some of the 10-intact cases have lower expression, whereas others have levels comparable to those seen in the 10-loss cases. The relative heterogeneity in the 7/10-intact cases may facilitate the detection of genes associated with a specific end point, such as aggressive biologic behavior and patient survival.

Comparison of the analysis in the 7/10 intact subset with the analysis using the entire sample set (Supplementary Table 3 versus Supplementary Table 5) identified 13 genes in common, suggesting that the use of DNA-based genetic signature subsets can be used to refine the prognostic significance of gene expression changes operative in specific subsets. In essence, whereas these 13 genes seemed to be prognostic in the entire group, closer examination indicated that they are related to clinical outcome only in one of the two aCGH-defined subsets. Our data support the concept and utility of using aCGH changes to identify robust molecular subsets, where patterns of gene expression may help identify specific genes and mechanisms underlying the behavior of malignant gliomas.

Several lines of investigation provided the rationale to examine the potential role of CHI3L1/YKL-40. First, it was the most highly expressed gene in glioblastomas relative to nonneoplastic brain samples. Second, its expression was associated with survival and it was also highly associated with chromosome 10 status. Third, previous studies suggest that this protein is overexpressed in and is a serum/plasma marker for epithelial tumors such as colon, ovarian, and breast carcinomas (36–38). CHI3L1/YKL-40 was also elevated in blood from patients with glioblastoma (39). These data suggested that CHI3L1/YKL-40 is a promising marker and plays a role in glioblastoma pathogenesis. In contrast to a recent report suggesting that elevated CHI3L1/YKL-40 in glioblastoma is due to expression in macrophages rather than tumor cells (39), our immunohistochemistry data indicated positive staining in the tumor cells themselves (Fig. 3). We identified candidate genes altered by CHI3L1/YKL-40 by comparing gene expression in CHI3L1/YKL-40-overexpressing transfectants versus mock-transfected cells. The genes that differed between the two groups in this experiment were then compared with CHI3L1/YKL-40-associated genes in glioblastoma tumors. This approach led to a set of candidate genes that are potentially downstream of CHI3L1/YKL-40 and relevant in human tumors (Table 1). Among the top 100 CHI3L1/YKL-40-associated genes in the tumor samples and transfected cell lines, 9 of 100 were common to both lists, a result which was much greater than expected by chance. This strongly suggests that CHI3L1/YKL-40 directly affects the expression of downstream genes in glioblastoma in vivo. One of these genes, nictinamide N-methyltransferase, was validated as a CHI3L1/YKL-40 responsive gene by real-time RT-PCR (Supplementary Fig. 3). The gene was associated with patient survival in our data (Supplementary Table 3) and previously to radiation response in bladder cancer (decreased expression was associated with increased radioresistivity; ref. 40). This latter association may be relevant to the clinical behavior of glioblastoma, where radiation is the primary treatment modality following surgery. Another study that directly tests the hypothesis that CHI3L1/YKL-40 is associated with clinical radiation response and genetic subtype supports the data in our present studies by showing that higher CHI3L1/YKL-40 levels portend radiation resistance and are associated with chromosome 10 status (41). Whereas further studies are required to define the precise mechanisms by which CHI3L1/YKL-40 contributes to tumor behavior and the relationship between loss of chromosome 10 and increases in CHI3L1/YKL-40, these data support an important role for this gene in glioblastoma.

In summary, combined aCGH and expression profiling analyses defined clinically relevant molecular subtypes of glioblastoma. Unsupervised analysis of the aCGH data suggested that chromosome 10 status was the most important CNA for glioblastoma classification. In turn, chromosome 10 was the DNA copy number aberration most highly associated with global changes in gene expression. Tumor subsets predominately defined by chromosome 10 status represented two groups of tumors distinguishable by both gene expression and clinical behavior. On a practical level, analyzing genetic signatures on both DNA and RNA levels may result in more robust molecular classification schemes than those derived from either experimental method alone. The combined data provide an opportunity to identify DNA copy number aberrations of which strong relationships with gene expression implicate them as genetic signatures that can be used to define tumor subtypes, which may have different biological and clinical behavior. Further studies should yield useful insights into the molecular subclassification of current morphology-based definitions of neoplastic entities.

Acknowledgments

Received 8/12/2004; revised 12/22/2004; accepted 12/29/2004.

Grant support: San Francisco General Hospital General Clinical Research Center grant MO1RR0083-41; NIH grants CA85799, NS42927, and CA90291; and National Brain Tumor Foundation, NIH grant CA97257.

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.

We thank Dolores Dougherty of the University of California San Francisco Brain Tumor Research Center tissue bank for assistance in providing tissue samples, Joan Aaron for editorial assistance, and Peng Chen for computational analyses.

References

15. Jain AK, Tokuyasu TA, Snijders AM, Segreves R, Albertson DG, Pinkel D. Fully automatic quantification

www.aacrjournals.org 1685 Cancer Res 2005; 65: (5). March 1, 2005

Downloaded from cancerres.aacrjournals.org on July 13, 2017. © 2005 American Association for Cancer Research.


Integrated Array-Comparative Genomic Hybridization and Expression Array Profiles Identify Clinically Relevant Molecular Subtypes of Glioblastoma

Janice M. Nigro, Anjan Misra, Li Zhang, et al.


Updated version Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/65/5/1678

Supplementary Material Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2005/04/07/65.5.1678.DC2

Cited articles This article cites 39 articles, 19 of which you can access for free at: http://cancerres.aacrjournals.org/content/65/5/1678.full#ref-list-1

Citing articles This article has been cited by 57 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/65/5/1678.full#related-urls

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.