Abstract

We used microarray analysis to investigate associations between genotypic expression profiles and survival phenotypes in patients with primary glioblastoma (GBM). Tumor samples from 7 long-term glioblastoma survivors (>24 months) and 13 short-term survivors (<9 months) were analyzed to detect differential patterns of gene expression between these groups and to identify genotypic subclasses of glioblastomas that correlate with survival phenotypes. Five unsupervised and three supervised clustering algorithms consistently and accurately grouped the tumors into genotypic subgroups corresponding to the two clinical survival phenotypes. Three unique prospective mathematical classification algorithms were subsequently trained to use expression data to stratify unknown glioblastomas between survival groups and performed this task with 100% accuracy in validation studies. A set of 1478 genes with significant differential expression \( p < 0.01 \) between long-term and short-term survivors was identified, and additional mathematical filtering was used to isolate a 43-gene “fingerprint” that distinguished survival phenotypes. Differential regulation of a subset of these genes was confirmed using RT-PCR. Gene ontology analysis of the fingerprint demonstrated pathophysiologic functions for the gene products that are consistent with current models of tumor biology, suggesting that differential expression of these genes may contribute etiologically to the observed differences in survival. These results demonstrate that unique expression profiles characterize genotypic subsets of primary GBMs associated with differential survival phenotypes, and these profiles can be used in a prospective fashion to assign unknown tumors to survival groups. Future efforts will focus on building more robust classifiers and identifying additional subclasses of gliomas with phenotypic significance.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Glioma; Glioblastoma; Genomics; Microarray; Classification
subcellular level. The consequences of this suboptimal system for classification can be witnessed clinically in the variable outcomes and response to therapy characteristic of patients whose gliomas are assigned a WHO grade of IV (GBM). While prognosis and response are poor for the majority of these patients, a small but discrete subgroup of therapeutic responders and subsequent long-term survivors exists. The inadequacy of histopathologic grading is evidenced, in part, by the inability to recognize these patients prospectively.

Over the past several years the concept of using molecular characteristics to predict clinical differences among glioma patients better has surfaced, and genomic studies of the gliomas have begun to appear in the neurooncology literature. Initial efforts focused on examining one or a few genetic alterations and analyzing their association with major clinical features. Such work demonstrated that primary or de novo GBMs typically are found in older patients with alterations in the epidermal growth factor receptor (EGFR) but without mutations in p53, while secondary GBMs tend to arise from gradual progression of lower-grade lesions with primary p53 mutations but without changes in EGFR.

DNA microarray techniques and other high-throughput technologies for genomic analysis permit a more intensive interrogation of the molecular pathogenesis of glioblastoma. Initial microarray investigations support the belief that gliomas in general and GBMs in particular are molecularly heterogeneous tumors and that molecular approaches to classification and grading may have distinct diagnostic and prognostic advantages [13–23]. The most fundamental of these studies have demonstrated the ability to group gliomas into mathematically cohesive classes based solely upon similarities in gene expression patterns [13–17]. Additionally, these classes can be used to train mathematical algorithms for class discrimination of novel tumors based upon expression profiles [18]. Subsequent investigations [19] focused more specifically on class discrimination among high-grade gliomas and suggested the existence of distinct molecular subclasses within this one histologic group. All of these studies were essentially proof-of-principle investigations demonstrating the potential role for expression profiles as a tool for glioma classification, but clinical correlations for the identified subgroups were largely absent.

The vanguard of this research is the search for distinct molecular subclasses of gliomas that have clinical or prognostic significance. Correlating clinical data with genomic expression data has demonstrated that expression profiles may be predictive of both tumor progression [20] and patient survival [21]. This is particularly evident when examining expression profiles of high-grade gliomas, for which molecular subgroups correlate with differential survival despite the similar histologic grade [22–24]. These data strengthen the belief that it is possible to use molecular expression profiles to construct subgroups of high-grade gliomas that have clinical and prognostic significance.

Phenotype generally reflects genotype, and thus it is logical that a classification scheme based directly upon the genotype could more accurately predict survival than one that relies on surrogate histological markers reflective of the underlying biology. Similarly, it is not fundamentally surprising that comparing the transcriptomes of gliomas known to behave differently under similar extrinsic conditions reveals a subset of differentially expressed genes. Accordingly, it appears likely that molecular methodologies of glioma identification and grading based upon differential gene expression profiles reflective of clinically significant phenotypic differences will progressively replace the traditional histopathologic methods of tumor grading. As described above, current glioma genomics research has demonstrated not only that classification based upon expression profiles is technically possible [18], but also that clinical behavior may be more accurately predicted by molecular classes than by the current histological (WHO) grades [21–23]. Classification schemes based upon expression profiling therefore have the potential to improve tumor grading, prognostication, and patient management, and our goal is to help translate this technology into the clinical arena.

This study examines the expression profiles of 20 primary GBMs. Thirteen of these tumors were collected from patients who ultimately died early from their disease (“short-term survivors” (STS)), while the remaining 7 were resected from patients who subsequently demonstrated prolonged survival (“long-term survivors” (LTS)). We examined the expression data from these tumors using two distinct analytic models. First, we used a robust combination of unsupervised and supervised array analysis techniques to identify differential patterns of gene expression that distinguish the LTS from the STS. The set of differentially expressed genes has potential mathematical significance as the substrate for prospective class discrimination algorithms and has potential clinical significance as a reflection of tumor biology. Second, we used a subset of the expression profiles to train a variety of prospective mathematical classifiers and then tested the ability of these tools to use tumor genotype to predict clinical phenotype. We believe that the analytic model that we have implemented represents the most comprehensive strategy employed to date for identification of differential expression patterns and subsequent phenotypic class discrimination between long-term and short-term GBM survivors.

Results

Using expression profiles for phenotypic class discovery

Selection of genes differentially expressed between long-term and short-term survivors

Genes that are differentially expressed between short-term and long-term survivors serve as the basis for class discovery, while genes without statistically significant differential expression do not make meaningful contributions to class discrimination and interfere with clustering algorithms. Accordingly, the Student t test (with \(p<0.01\)) was used to identify a population of 1478 genes with statistically significant differential expression between long-term and short-term survivors to be used as the basis for subsequent class discovery.

Unsupervised analysis

Unsupervised analysis algorithms employ unbiased searches for patterns of expression that can be used to develop hypotheses
regarding the mechanistic associations between genotype and phenotype. These approaches are useful for identifying expression patterns within the set of differentially regulated genes that may have phenotypic significance before any additional clinical data are factored into the analysis. Five independent unsupervised clustering algorithms successfully separated long-term from short-term survivors with 100% accuracy. Average linkage hierarchical clustering of samples divided the LTS group from the STS group at the most proximal bifurcation of the dendrogram. Support trees confirmed the LTS vs STS split held in 99% of the 100 iterations (i.e., 99% support for this distinction, Fig. 1). The self-organizing tree algorithm (SOTA), which clusters samples through a deconstructive, “splitting” process rather than an agglomerative process, also made the LTS versus STS distinction with 100% accuracy. Correspondence analysis (CoA) (Fig. 2) and principal components analysis (PCA) (data not shown), which resolve the majority of the variability between samples into three composite vectors and represent the differences in three-dimensional space, clearly distinguish the LTS from the STS group.

Supervised analysis

Supervised analysis methods exploit known phenotypic relationships between samples as well as information about known molecular pathways to identify expression “fingerprints” that can be used in tumor classification. The simplest supervised algorithms use basic phenotypic information for discovery of genotypic classes with phenotypic significance. Our study population contains two distinct phenotypic subgroups (LTS and STS), and several minimally supervised algorithms can use this a priori knowledge of two distinct phenotypic groups to guide class discovery. k-means clustering is an iterative process in which the expression profile of each sample is tested against a specified number of centroids, and the process repeats until a “best fit” is achieved and the groups are optimally separated. k-means clustering with \( k=2 \) separates LTS from STS samples with 100% accuracy. This result is supported by the \( k \)-means support algorithm, which runs the \( k \)-means clustering process multiple times from random starting points to verify the clustering results. Self-organizing mapping (with a two-class, \( 2 \times 1 \) matrix) and terrain mapping (Fig. S1), neighbor-based unsupervised algorithms, also demonstrate an STS group and an LTS group with all samples correctly assigned.

Building a prospective classifier—using genotype to predict phenotype

The ability to identify a set of genes associated with a particular phenotype does not necessarily imply that this set will allow accurate and reliable classification of unknown tumors into appropriate clinical groups, because the difference between deconstructing a known expression profile and effectively searching for patterns within a novel profile is paramount. Moreover, a small set of “highly significant” genes is not necessarily the best pattern upon which to build a classifier, as this set has likely undergone multiple iterations of filtering, which tends to introduce compounding biases. For this reason, algorithms have been created that use expression data first to define a composite classification set and subsequently to assign an unknown sample to a particular group.

A relatively simple example of this type of classifier is the \( k \)-nearest neighbors algorithm. This strategy is essentially an extension of the \( k \)-means approach to clustering, in which a neighbors-based algorithm using distance metrics assigns an unknown to one of \( k \) (user defined) groups based upon its relative distance from each group’s expression centroid. This
method has been used to distinguish glial from nonglial brain tumors [18], and here we have applied it in a more focused fashion to classify subgroups of GBMs. A training set of 7 samples was used to construct the LTS and STS expression centroids, and the remaining 13 samples, whose expression profiles were not used to train the classifier, were treated as unknowns and were prospectively classified. The algorithm correctly classified 100% of the unknowns into the appropriate phenotypic group.

A second strategy for prospective classification is discriminant analysis (DAM). This is a multistep algorithm that first applies an ANOVA filter to select genes that should be near optimal for partitioning the unknown samples based upon permutations of gene expression in the training set. Next, a multivariate partial least squares method is used for gene dimensional reduction. This is followed by a quadratic discriminant analysis. The result is not only classification of the unknowns, but also identification of the subset of genes whose expression patterns were most useful for classification. When DAM was applied to the sample population, again using the 7-member training set and the 13 unknowns, samples were classified into the LTS and STS phenotypic groups with 100% accuracy. Additionally, the resulting list of useful genes generated by this algorithm was used later to investigate the “survival fingerprint” further (see below).

A third strategy for classification is the recently described uncorrelated shrunken centroids (USC) algorithm. This method was designed with the specific goal of allowing prediction of phenotypic category of a tissue sample based upon its expression profile as well as identification of genes relevant to this classification. The USC method is based upon a shrunken centroids algorithm but modifies the approach by removing genes with highly similar expression patterns. This therefore exploits the functional interdependence of genes to reduce the number of genes used to classify, producing a smaller set of important genes and improving the accuracy of classification. USC was applied to our sample population using the 7-member training set and the 13 unknowns and again resulted in classification of the unknowns into the LTS or STS group with 100% accuracy. The list of genes used to make these distinctions was also used in investigations of the survival fingerprint (see below).

Virtual logic machines offer a robust approach to class discrimination based upon expression profiles. One such algorithm is the support vector machine (SVM), which uses known relationships between members of a subset of elements coupled with the expression patterns of all of the elements to create a training set of numeric weights for each element. In a second phase, the weights are used to assign discriminators to the elements, and a binary output results in elements being considered either “in or out of” the group defined in the initial training process. The pattern of weights can be retained after training and used to classify novel elements. We used SVM with the 7-member training set and the 13 unknowns. Using radial metrics with a constant of 2 and a width factor of 2, the SVM was able to classify the 13 unknowns into the appropriate survival groups with 100% accuracy.

### Identifying genes that distinguish survival groups

Seven distinct clustering algorithms and two support algorithms successfully and accurately distinguished between LTS and STS samples based on expression profiles. Additionally, four prospective clustering algorithms exploited known genotypic–phenotypic relations to build classifiers that use expression profiles to classify unknown samples into phenotypic (survival) groups with 100% accuracy. The next objective was to determine what patterns of differential gene expression enabled class discrimination. Identification of this “molecular fingerprint” is essential in the development of clinically useful tools for molecular classification of GBM.

This was accomplished by first constructing five independent lists of differentially expressed genes, each derived from a different analytical model. These models included both unsupervised class discovery strategies and prospective molecular classifiers. As each of these algorithms is optimized to identify genes with unique patterns of differential expression, constructing an expression fingerprint by focusing on the overlap these lists could exclude potentially important genes that may appropriately appear on only one list. Instead, the molecular signature that we describe is a composite constructed by capturing genes identified as most significant from each of the five lists.

The first list was constructed by applying a two-tailed \( t \) test \((p<0.01)\) to the total population of 24,421 elements to identify a group of 1478 genes with statistically significant differential expression between the LTS and the STS groups (as described above). Second, the significance analysis of microarrays (SAM) algorithm, an interactive algorithm that is more versatile and more robust that the Student \( t \) when used with array data, was also employed to discover genes in the total population with significant differential expression. More than just a test of significance, SAM allows for operator control of the false discovery rate and is applicable even in data sets with missing data points. Setting the false discovery rate at 2 (genes) identified a group of 53 genes with highly significant differential expression between the two groups. These genes comprised the second gene list. Two additional lists were created by applying the DAM and USC algorithms to the total population. As described above, each of these algorithms accurately separated the LTS from the STS group, and each generates a list of genes that were most useful for class discrimination. Student’s \( t \) test \((p<0.05)\) was subsequently applied to both lists, generating lists of 8 (from USC) and 28 (from DAM) genes whose differential expression patterns were both statistically significant and useful for supervised class discrimination. Six genes failed to achieve statistical significance but appeared on both the USC and the DAM list, and these were retained in subsequent analyses as the fifth and final list.

The five gene lists were combined into a composite gene list, and this list was then further refined. Redundant entries were deleted, and genes that appeared only on the \( p<0.01 \) list and not on any additional lists were removed. Next, the mean, standard deviation, and 95% confidence interval for each
gene’s log₂ expression ratio in the LTS and STS groups were computed. Genes were deleted (1) if they represented only ESTs, (2) if the 95% confidence intervals for mean expression value in the LTS and STS groups intersected, or (3) if the absolute difference in average expression groups was <0.35 log₂ units (~0.27-fold change). Variance of the mean was calculated as a measure of dispersion for the expression of each gene within each group, but this value was not used as an absolute exclusion criteria. The result was a list of 43 genes that are both differentially regulated and useful for class discrimination (Table S1, Fig. 3).

Verifying expression data

Reverse transcriptase polymerase chain reaction (RT-PCR) was performed to validate the 43-gene survival fingerprint described above. Six housekeeping genes were tested empirically. GAPDH was determined to be the most consistent control and was used as the reference for all subsequent RT-PCR. Fourteen genes from the survival fingerprint were selected for further study (see Table S2). These genes had one or more of the following attributes: (1) large magnitude of differential regulation between survival groups, (2) narrow variance of magnitude of relative expression within a given survival group, (3) representation in multiple GO categories and terms (see below), and/or (4) known or theoretical pathophysiological significance in GBM. These criteria were used in an attempt to select high-quality genes with both biological and statistical significance for separating the LTS and STS groups.

Expression of the 14 genes was assayed in all 20 glioblastoma samples, in the pooled LTS and STS samples (two technical replicates of each, constructed as described above), and in the control normal brain RNA sample. A total of 280 data points representing the expression of each gene in each sample (relative to control) were generated. Internal controls in the RT-PCR analysis system indicated that 266 of these points represented reliable data, and these were used for subsequent investigation. Analysis of the average expression data for each of the 14 genes across all 20 glioblastoma samples indicated that the direction of expression of a given gene assayed by RT-PCR (upregulation or downregulation relative to control) matched the direction of expression measured by microarray analysis in a mean of 78.7% of samples. Additionally, the RT-PCR and microarray data concurred on the direction of regulation in at least 50% of samples for all 14 assayed genes. Analysis of these data by sample demonstrated concurrence of RT-PCR and microarray data for the direction of regulation in a mean of 74.9% of the genes in each sample. Overall, the RT-PCR and microarray data demonstrated regulation in the same direction for a total of 197 of the 266 data points assayed (74.1%). Eighty-three (31.2%) of these data points demonstrated <50% difference in the measured magnitude of differential gene expression (relative to control), although comparisons of magnitude of differential regulation across platforms are subject to multiple sources of error and must be interpreted in that context.

Comparison of the differential expression of the assayed genes between the LTS and the STS groups demonstrated a high degree of correlation between the RT-PCR (pooled and non-pooled samples) and the microarray data (Table S3). More than half of the analyzed genes had differential regulation in the same direction between phenotypic groups in all three assays (8 of 14; 57.1%). The semiquantitative data also showed a high degree of correlation between assays. When the mean \( \log_2( \text{LTS/STS} ) \) was calculated, the standard deviation, expressed as a percentage of the mean, was less than 75% for 8 of 14 (57.1%) genes. Similarly, the variance of the mean was <0.1 for 8 of 14 (57.1%) genes. These data suggest that RT-PCR data, whether derived from pooled samples or from the average expression of non-pooled samples, support the qualitative and semiquantitative microarray data.

Fig. 3. Hierarchical tree of survival fingerprint genes only. STS, short-term survivor group; LTS, long-term survivor group. Unique sample identifiers have been removed.
Functional characteristics of genes that distinguish survival groups

The purpose of this study is to demonstrate the ability to classify GBMs into subgroups with prognostic significance based upon the expression profile, and functional analysis of the expression fingerprint that distinguishes the LTS from the STS groups is useful for the purpose of further exploring the molecular biology of the genes that facilitate such classification. If the fingerprint reflects transcriptome-level changes that serve as the biological basis for the observed genotypic–phenotypic correlation, then it would be expected to contain genes having some pathophysiologic significance. Accordingly, functional classification of the genes comprising the GBM fingerprint was performed using EASE [45] and DAVID [46] for Gene Ontology (GO) [47,48] annotation of the gene list. The resulting data provide insight into the functional significance of the transcriptome-level changes in cellular processes and macromolecular localizations that contribute to the survival fingerprint.

Table 1 lists the 15 most commonly occurring GO terms in the GO molecular function, biological processes, and cellular component classes as well as the number of genes matching each term in the GO annotation of our 43-gene survival fingerprint.

Additionally, we used the relative expression values of the genes annotated to each GO term to investigate further the net trend in expression for that term. This was accomplished by first averaging the log2 expression ratio relative to control for each gene and then calculating the mean expression for that gene in both the LTS and the STS populations. Next, the properties of log subtraction were used to calculate the relative expression of each gene in the LTS group relative to its expression in the STS group. Finally, the mean log2(FiLTS/FiSTS) for the genes annotated to each GO term was calculated. The statistical significance of the differential expression reflected by this mean value was calculated using the Student t test to compare the mean expression values for each gene (relative to control) in the LTS and STS groups.

The results of this analysis demonstrate that genes coding for cellular elements involved in protein binding, receptor binding, protein metabolism, and cell–cell signaling have statistically significant upregulation in the LTS group relative to the STS group. Differential regulation of additional structural and functional elements is suggested by the additional GO categories listed in Table 1, although the number of genes assigned to each category is insufficient to achieve statistical significance.

Discussion

Patient selection and inclusion criteria

Microarray studies conducted for the purpose of class discovery are sensitive to the effects of bias introduced prior to the data analysis. Erroneous classification often results when assumptions are made in advance regarding the relative importance of a particular gene or group of genes in defining a functional class or when such assumptions are made part of the inclusion criteria in a class discovery study. Because of the large amount of data generated in microarray analyses, class discovery algorithms can often generate groups with distinct and statistically significant differences in expression profiles, but these groupings may nonetheless be inaccurate or invalid because the classification is predicated upon faulty assumptions. This is a well-known problem in molecular class discovery research, and the solution is to control potential biases aggressively prior to analysis and to limit or eliminate preconceived assumptions when designing a class discovery experiment.

This study was constructed with these considerations in mind, and both the study design and the analytic model were engineered to limit such biases. Inclusion criteria for this study were made as broad as possible to limit selection bias. Samples of primary (de novo) GBMs from adult patients resected prior to adjuvant therapy were included without regard to the molecular characteristics of the tumor. This was done in an attempt to eliminate a priori conclusions regarding clinical significance of select genes, as the results of EGFR amplification [49] and p53 mutational analyses [1], for example, have proven inconclusive as to their exact effect in terms of response to therapy or prognosis. The distribution of EGFR and p53 alterations in our population of STS and LTS patients appears to be random and not likely to be determinative of the microarray patterns.
In addition, inclusion of only primary GBMs minimizes the potential effect of prior treatments that may have been previously administered to patients with secondary GBMs. Finally, we are also aware that limiting the group to only those tumors histologically classified as GBMs is a potential source of bias, as histopathologic classification of gliomas is itself subjective. A more extensive investigation including all grades of glioma would be better suited for unbiased class discovery, but such an analysis was beyond the scope of this proof-of-principle study. Given that GBM is the most histologically distinct tissue diagnosis among the gliomas, we believe the chance of including a “lower grade” tumor is minimal. We therefore believe that these inclusion criteria limit selection bias, but we are aware that the chance of such bias is not zero.

Differences in extent of initial resection between survival groups

Examination of the demographic data for the LTS and STS groups demonstrates that the only statistically significant difference between groups was the extent of the initial tumor resection. One hundred percent of patients in the LTS group had surgical intervention resulting in either gross total resection or near-total resection, but this result was achieved in only 46% of STS patients. Extent of resection with survival in patients with malignant gliomas [50], but the pathophysiologic nature of this relationship remains unclear. While one interpretation of this finding is that an increased number of residual tumor cells leads to earlier recurrence, another possibility is that the phenotypic correlations of the tumor genotype (i.e., growth rate, location, degree of microinvasion) ultimately affect the resectability of the tumor. The relatively small sample sizes in each group should also be considered when interpreting these data, and the extent to which this trend is apparent in the overall populations of long-term and short-term survivors cannot necessarily be extrapolated from these data. We are optimistic that future, large-scale investigations will help to clarify further the true nature of this observation.

Without more extensive knowledge of the quantitative relationships of both extent of resection and tumor genotype on overall survival, an appropriately constructed multivariate survival analysis cannot be performed. Nonetheless, we felt that it was important to investigate the degree to which extent of resection affected our class discovery, our prospective tumor classifiers, and our survival fingerprint. We therefore repeated the analysis outlined above after subtracting data from the seven STS patients who underwent subtotal resection. The population for this redo analysis therefore consisted of samples from seven LTS and six STS patients ($n = 13$).

The class discovery results were identical to those described above, with hierarchical clustering (HCL), support tree algorithm (STA), SOTA, PCA, CoA, $k$-means clustering (KMC), and terrain mapping (TM) algorithms all separating the LTS from the STS group with 100% accuracy. The prospective classifiers performed nearly as well, with $k$-nearest neighbor clustering (KNNC), DAM, and USC classifying unknowns with 100% accuracy and SVM misclassifying only 1 of 7 unknowns.

We believe that this slight increase in the SVM error rate may be attributable to the smaller number of samples available for construction of the training set. The survival fingerprint was also similar in composition, but contained approximately half as many genes. The genes that were absent after this redo analysis were generally those in the original survival fingerprint with smaller relative expression differences between LTS and STS. This effect is not surprising, given that data are available from fewer samples to contribute to the two-sided $t$ test. The relative differences in these “borderline” genes may therefore fail to achieve statistical significance after expression data from six samples are subtracted, and thus the genes are excluded from the fingerprint. We believe that the high degree of correlation between class discovery, prospective classification, and gene identification in the original and the redo analyses argues for at least some degree of pathophysiologic contribution to the differential survival, rather than simply reflecting the effects of extent of tumor resection.

Class discovery

A two-tiered data analysis strategy was used for class discovery. The initial analysis employed unsupervised class discovery algorithms, including hierarchical clustering, the support tree algorithm, the self-organizing tree algorithm, correspondence analysis, and the analysis of principal components. Because the initial two-tailed $t$ test that defined the population of genes used as the basis for this class discovery was performed such that genes with differential expression between two known phenotypic groups were selected as the substrate for subsequent class discovery, we are aware that this initial analysis is not absolutely unsupervised in the strictest sense. It is, however, unsupervised in that no information regarding the phenotype of any individual sample was used to influence class discovery.

The unsupervised clustering algorithms used unbiased searches for patterns of expression to cluster samples based purely upon their genotype. Each of the five algorithms independently identified a fundamental distinction between two unique genotypic subgroups within the sample population. The final assignment of samples into these two subgroups was identical for all five algorithms, supporting the hypothesis that the sample population contains at least two unique genotypic groups. Subsequent phenotypic analysis demonstrates that one of the primary genotypic subgroups comprises the samples from all 7 long-term survivors, while the other genotypic subgroup contains the samples from all 13 short-term survivors. The unsupervised analysis therefore demonstrates the presence of two unique genotypic groups within the sample population and suggests a phenotypic correlation for these groups.

The supervised analysis was predicated upon the knowledge that two distinct phenotypic groups existed in the sample population. Using only the a priori knowledge of the number of subgroups, the KMC, self-organizing maps (SOM), and TM again demonstrated phenotypically similar samples clustered together into two distinct genotypic groups with 100% accuracy. This supports the hypothesis that there are fundamental differences in the transcriptomes of GBMs from patients in the
LTS and STS groups, and it again demonstrates that statistical algorithms can use these genomic expression differences to cluster samples into groups with phenotypic significance.

Finally, we demonstrated, using four different algorithms, that the genomic expression patterns that characterize the two survival groups can be used to train mathematical classifiers that are subsequently able to classify unknowns into genotypic groups with phenotypic correlations. KNNC is the least user-dependent algorithm, having a minimal number of parameters to adjust prior to training and classification. This is attractive because it limits operator bias, but it also minimizes the extent to which the algorithm can be customized. While the DAM algorithm is optimized for class discrimination among three or more classes, the results were favorable and the gene list generated by the algorithm was useful in constructing the expression fingerprint. The USC and the SVM methods can be adjusted to allow accurate classification of unknowns while minimizing overclassification, and the SVM classification table can be saved for future use on a novel population of unknown samples. These strategies represent the newest direction in genome-based prospective classification strategies, and we have demonstrated that they are potentially useful for constructing prospective, molecular classifiers capable of distinguishing subgroups of gliomas with phenotypic significance.

Differential genomic variability between phenotypic groups

An interesting effect observed in the course of analyzing the expression profiles of samples in the LTS and STS groups is the differential variability of the expression profiles within each subgroup. The PCA and CoA plots of the LTS group demonstrate a high degree of similarity in the overall expression signature of these samples, with most of the variability represented along a single axis of the 3D plots (Fig. 2). The opposite is true of the STS group, in which there is significant spread along all three axes. This effect is also noted in the SOTA analysis, by which the LTS samples cluster early into one large group, while the STS samples stratify into multiple subgroups of one or two samples each. These findings suggest that the members of the STS group may have a high degree of individual genomic variability, while the LTS group demonstrates a more cohesive overall pattern of expression suggestive of fundamental genomic similarity among members of the group.

This finding is also significant considering the mechanism by which class assignments are made for novel samples. It appears that the LTS phenotype correlates with a distinct genotypic signature, whereas the STS phenotype reflects expression patterns that are “everything else.” If this is the case, then the ability of the prospective classifiers to assign a novel sample to a survival group may be predicated upon a simple “in-class” versus “out-of-class” call for the unknown expression profile, relative to the LTS expression fingerprint. This concept is similarly illustrated by the CoA plot, in which the expression differences distinguishing the survival groups are reflected purely by position along the x axis (axis 1). The remainder of the genotypic variability does not contribute to the in-class or out-of-class distinction.

The physiologic correlation of this observation is also important when interpreting these findings. One interpretation of the above observations is that the LTS profile reflects a unique subclass of GBMs that are somehow inherently less aggressive (slower replication, less invasive, etc.). These tumors would therefore be expected to grow and to recur more slowly, and the survival fingerprint reflects expression patterns that facilitate this inherently more gradual tumor progression. An alternate interpretation of the data suggests that the survival fingerprint is characteristic of a GBM subgroup that has improved response to initial adjuvant therapy and thus enjoys prolonged survival. This conclusion is based upon the observation that all of the patients included in this study underwent adjuvant therapy following the initial resection, usually consisting of XRT + Temodar. Overall, patients in the LTS group responded well to adjuvant therapy, as evidenced by an increased time to progression as well as a prolonged survival time. Patients in the STS group may have received relatively less benefit from similar adjuvant therapy, reflected by relative decreases in the same clinical parameters. This suggests that the survival fingerprint may reflect an expression pattern that makes the tumor cell relatively more sensitive to radiation damage and alkylators and therefore ultimately prolongs survival. This study has not been specifically designed to investigate this question further, but we intend to pursue future investigations that further investigate the existence and nature of such “response to therapy” expression profiles.

Functional significance of the survival fingerprint

The 43 differentially regulated genes in the survival fingerprint code for proteins involved in signal transduction, second-messenger pathways, extracellular matrix interaction, cellular growth and metabolism, and stress response. Subjective review of this gene list suggests that differences in response to immunologic signaling, extracellular interactions, and cell metabolism may contribute to the differences in clinical outcomes observed in these phenotypic subsets of GBM patients. Relative upregulation in the LTS group of genes coding for proteins involved in cell–cell signaling and in second-messenger pathways may suggest that these cells are more responsive to extrinsic signaling, more susceptible to immune modulation, or have less dysregulation of normal cellular processes. This, in turn, may contribute to the relative increase in survival observed in the LTS group.

A more objective analysis of the functional significance of the survival fingerprint was performed using EASE and DAVID to assign GO annotations to the 43 genes in the profile. The degree of net differential regulation of the genes associated with each GO term was calculated (as described above) and tested for statistical significance. This analysis demonstrated statistically significant upregulation of the receptor binding, protein binding, cellular protein metabolism, and cell–cell signaling GO terms in the LTS subgroup. While the magnitude of the net differential regulation of these terms is prone to biases introduced by overclassification, the trends illustrated by this approach lend further support to the hypothesis that GBM cells
in the LTS group may be more responsive to external modulation and may have less dysregulation of normal cellular physiologic processes.

We note, however, that this experiment was designed to maximize the potential for clinically relevant, phenotypic tumor classification and that it has not been optimized for the study of tumor biology. As such, it is possible that functionally significant genes do not appear in this profile, and it is also possible that some genes in this profile reflect expression in the tumor microenvironment rather than in the transformed astrocyte. Additionally, the nature of the algorithms used to identify genes for this list may have excluded some genes with physiologic significance in exchange for genes with more a statistically significant contribution to class discovery. We therefore have been careful to avoid drawing excessive conclusions regarding the role of the specific genes comprising the profile, and we do not suggest that the presence or absence of any particular gene in the fingerprint be overly scrutinized. The genes in this molecular signature seem to be useful for classification, and their products have functions that fit with current paradigms of tumor cellular biology. However, detailed conclusions regarding their functional significance are premature.

Conclusions

We used microarray analysis to investigate the association between genotypic expression patterns and survival phenotype in patients with a primary glioblastoma. We demonstrate that GBM tumor genotype corresponds with survival phenotype and that expression data can be used to classify GBMs into genomic subgroups with phenotypic significance. Class discrimination can be performed prospectively, using previously trained classification algorithms, and this strategy can be used to assign unknown tumors into genotypic subgroups that associate directly with the survival phenotypes. A population of at least 43 genes is differentially expressed between long-term and short-term glioma survivors, and the expression patterns of these genes contribute to these class distinctions. In addition, these 43 genes code for proteins that may be functionally significant in the molecular genesis of GBMs. Future work is focused on building more robust classifiers and identifying additional sub-classes of gliomas with phenotypic significance from a larger, unknown series of GBMs. Such classification tools can be used both as a method to discriminate LTS and STS patients at the time of diagnosis—which may prove useful in future human trials—and to explore the significance of select gene products as potential biological targets for novel therapeutics.

Materials and methods

Inclusion criteria

Patients selected for inclusion in this study had a confirmed diagnosis of primary (de novo) glioblastoma and were at least 18 years of age at the time of diagnosis. Tissue samples were obtained during initial resections of newly diagnosed GBMs performed as part of standard medical management of this disease between December 2002 and January 2006. Patients were eligible for this study if their documented survival was either ≤9 months (short-term survivor) or ≥24 months (long-term survivor) from initial diagnosis. Patients were excluded if they had undergone prior radiation, chemotherapy, or surgical resection for their GBM. Patients were also excluded if the available clinical information was insufficient to verify study eligibility or if the pathologic specimen was inadequate for microarray analysis.

We chose patients with a primary GBM to avoid the effects of prior treatment (chemotherapy, radiation), which can influence tumor behavior at both the gene and the protein expression level [25,26].

Sample selection

The Brain Tumor Institute database was queried to generate a list of all patients meeting the inclusion criteria. The electronic medical records of these patients were then reviewed to verify study eligibility. Dates of death were confirmed by querying the Social Security Death Index. Seven long-term survivors meeting all inclusion criteria were included in this study. Eligible short-term survivors were randomized using a random number generator, and tissue from 13 was subsequently selected for inclusion. This study was approved by the Cleveland Clinic Institutional Review Board.

Tissue samples from the 20 selected GBMs were originally collected at the time of surgery following the Brain Tumor Institute’s standard operating protocols. All samples used in this study were flash frozen as quickly as possible after resection and were subsequently stored at −80 °C until RNA extraction was performed.

Demographics

Demographic information for the LTS group and the STS group are summarized in Table 2. The mean ages of patients comprising the LTS and STS

<table>
<thead>
<tr>
<th>Demographics and treatment</th>
<th>LTS</th>
<th>STS</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>7.00</td>
<td>13.00</td>
<td></td>
</tr>
<tr>
<td>Age (mean)</td>
<td>57.71</td>
<td>64.00</td>
<td>0.2259</td>
</tr>
<tr>
<td>Age standard deviation</td>
<td>8.50</td>
<td>13.85</td>
<td></td>
</tr>
<tr>
<td>Age 95% CI</td>
<td>6.30</td>
<td>7.53</td>
<td></td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>0.57</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Sex (% male)</td>
<td>0.43</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Survival (mean) (months)</td>
<td>29.09</td>
<td>5.53</td>
<td>0.0000</td>
</tr>
<tr>
<td>Survival standard deviation</td>
<td>4.29</td>
<td>3.13</td>
<td></td>
</tr>
<tr>
<td>Survival 95% CI</td>
<td>3.18</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>No. (% GTR)</td>
<td>4 (57%)</td>
<td>3 (23%)</td>
<td></td>
</tr>
<tr>
<td>No. (% NTR)</td>
<td>3 (43%)</td>
<td>3 (23%)</td>
<td></td>
</tr>
<tr>
<td>No. (% STR)</td>
<td>0 (0%)</td>
<td>7 (54%)</td>
<td></td>
</tr>
<tr>
<td>Resections (mean)</td>
<td>1.43</td>
<td>1.15</td>
<td>0.2567</td>
</tr>
<tr>
<td>Resections standard deviation</td>
<td>0.53</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>Resections 95% CI</td>
<td>0.40</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Initial intervention was resection</td>
<td>7 (100%)</td>
<td>13 (100%)</td>
<td></td>
</tr>
<tr>
<td>Postresection XRT</td>
<td>7 (100%)</td>
<td>11 (85%)</td>
<td></td>
</tr>
<tr>
<td>Total dose (mean)</td>
<td>53.30</td>
<td>49.28</td>
<td>0.5704</td>
</tr>
<tr>
<td>Dose fractions (mean)</td>
<td>34.25</td>
<td>24.80</td>
<td>0.3845</td>
</tr>
<tr>
<td>Adjuvant chemotherapy</td>
<td>7 (100%)</td>
<td>9 (69%)</td>
<td></td>
</tr>
<tr>
<td>Additional chemotherapy</td>
<td>5 (71%)</td>
<td>6 (46%)</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy modalities (mean)</td>
<td>2.86</td>
<td>1.00</td>
<td>0.0325</td>
</tr>
<tr>
<td>Chemotherapy modalities standard deviation</td>
<td>1.77</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy modalities 95% CI</td>
<td>1.31</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Recurrence/progression (radiographic)</td>
<td>6 (86%)</td>
<td>9 (69%)</td>
<td></td>
</tr>
<tr>
<td>Recurrence/progression unknown/not documented</td>
<td>1 (14%)</td>
<td>4 (31%)</td>
<td></td>
</tr>
<tr>
<td>Time to recurrence/progression (mean) (months)</td>
<td>7.28</td>
<td>2.40</td>
<td>0.0414</td>
</tr>
<tr>
<td>Time to recurrence/progression standard deviation (months)</td>
<td>4.15</td>
<td>2.80</td>
<td></td>
</tr>
<tr>
<td>Time to recurrence/progression 95% CI (months)</td>
<td>3.26</td>
<td>1.04</td>
<td></td>
</tr>
</tbody>
</table>

CI, confidence interval; GTR, gross total resection; NTR, near-total resection; STR, subtotal resection; XRT, radiotherapy.
groups were 57.7 and 64.0 years, respectively (p = 0.23). The LTS group was 57% female/43% male, and the STS group was 77% female/23% male. The mean survival was 29.1 months (±3.1, 95% CI) for the LTS group and 5.5 months (±1.7, 95% CI) for the STS group (p = 0.0001).

All patients except two underwent initial attempted gross total resection at our institution as the initial therapeutic intervention. Both of the exceptions were patients in the STS group. One underwent open biopsy followed by gross total resection, and the other had an incomplete resection at an outside institution and underwent reoperation with gross total resection at our institution approximately 2 weeks later. The extent of resection in all patients was assessed on postoperative, contrast-enhanced MRI or CT imaging. We defined gross total resection (GTR) as absence of residual loselomatous enhancement, near-total resection (NTR) as trace amounts of residual enhancement without obvious residual tumor mass, and subtotal resection (STR) as obvious residual, enhancing tumor mass, with resection of >88% of original tumor [4]. Using this definition, GTR or NTR was achieved in seven (100%) patients in the LTS group and in six (46%) patients in the STS group. STR was performed in the remaining seven (54%) STS patients.

All patients included in this study were initially managed with the goal of receiving maximal medical and surgical management for their tumors, and an intent-to-treat model was followed. The mean number of total surgical resections performed per patient was 1.43 (±0.4, 95% CI) in the LTS group and 1.15 (±0.4, 95% CI) in the STS group (p = 0.26). Postoperative radiation therapy (XRT) was used in 7 (100%) LTS patients and 11 (85%) STS patients. Two STS patients did not receive XRT; one ultimately refused and the other died prior to initiation of therapy. XRT dosimetry data demonstrate that LTS patients received a mean total dose of 53.3 Gy in a mean of 34.3 fractions, and STS patients received a mean of 49.3 Gy in a mean of 24.8 fractions (p = 0.57 and p = 0.38). Adjuvant chemotherapy was used in 7 (100%) LTS patients and 9 (69%) STS patients. Two of the 4 STS patients who did not receive adjuvant chemotherapy refused this modality, and 1 underwent adjuvant local radiotherapy with a glueSite device. Chemo-therapy data were unavailable for the fourth patient.

Follow-up imaging was available for six (86%) LTS patients and nine (69%) STS patients. Radiographic evidence of recurrence or progression was demonstrated in these patients at a mean of 7.3 months (±3.3, 95% CI) for patients in the LTS group and at a mean of 2.4 months (±1.0, 95% CI) for patients in the STS group (p = 0.04). With the exception of one current survivor, all patients ultimately died of their disease. Progression is therefore assumed to have occurred in the remaining one LTS patient and four STS patients despite the unavailability of imaging.

Data from molecular genetic assays routinely performed on brain tumor specimens at our institution were collected from the immunohistochemistry results reported in the clinical pathology reports for all tumors included in this study. Loss of chromosome 1p, loss of chromosome 19q, EGFR amplification status, and maximum Ki-67 index were not significantly different between the STS and the LTS groups. TP53 immunopositivity, a marker for the presence of TP53 mutations (for which 0STS and the LTS groups. TP53 immunopositivity, a marker for the presence of TP53 mutations (for which 0 alleles) [27], suggested a predominance of wild-type p53 in both survival groups (Table 3).

Histologic examination

Flash-frozen tissue samples were removed from −80 °C storage and were immediately embedded in OCT medium. Representative 16-μm sections (two replicates) were cut from the embedded tumor block. Each tissue block was returned to −80 °C storage immediately after sectioning to minimize any potential RNA degradation that may result from tissue warming. Frozen sections were then stained with hematoxylin and eosin following standard protocols and were examined using transmitted light microscopy. Samples were reviewed prior to RNA extraction and confirmed to contain 90% or more viable (nonnecrotic) tumor cells.

RNA extraction and purification from tissue samples

RNA extraction was performed using a modification of the TRIzol method optimized to maximize the extraction of high-quality RNA from mammalian brain tissue (data not shown). Briefly, mechanical homogenization in TRIzol reagent was used to disrupt and lysse small surgical biopsy samples of GBM tissue. This was followed by purification with serial extractions and Qiagen RNeasy spin column cleanup. The quantity, quality, and integrity of the RNA were verified using the NanoDrop spectrophotometer and the Agilent Bioanalyzer 2100. The detailed protocol is described in the supplemental material.

Control RNA

Ambion’s FirstChoice human brain reference RNA was used as the experimental control. This RNA represents normal, whole-brain, total RNA extracted from 26 unique brains and pooled after extraction. Control RNA was analyzed using the NanoDrop spectrophotometer and the Bioanalyzer 2100 as described above.

Microarrays

RNA (3 μg) from the GBM samples and the controls was used as the starting material for synthesis of cDNA target. Target preparation, labeling, and hybridization were performed according to standard Affymetrix protocols by the Gene Expression and Genotyping Facility of the Comprehensive Cancer Center of Case Western Reserve University and University Hospitals of Cleveland. All samples were hybridized to the Affymetrix Human Genome U133A Plus 2.0 array, which returns expression data for 24,421 unique genes. Two replicate hybridizations were performed for the control RNA, giving a total of 22 micro-arrays performed for this study (13 STS, 7 LTS, 2 control).

Real-time RT-PCR

Real-time RT-PCR was performed to validate the relative expression levels of a subset of 15 genes (ANK1, APP, BMX, CD34, GBAS, HSPA1B, IL13, IL17, ITGA6, JAG2, MMP4, NOS2, RPL10, SOX4, UCP3) included in the survival fingerprint. The TaqMan assay (Applied Biosystems) was used in conjunction with the ABI Prism 7900HT sequence detection system (Applied Biosystems) to measure the expression of each of the 15 genes in each of the 20 tumor samples and in two technical replicates of the control. All assays were performed using 3 μg of total RNA from each sample. Additionally, a pooled sample of LTS RNA was constructed by combining 1 μg of RNA from each of the 7 LTS samples, and a pooled sample of STS RNA was constructed in a similar fashion using RNA from the 13 STS samples. Expression levels of each of the 15 genes were also tested in these pooled samples, as an additional method for examining the mean expression of each gene in each phenotypic subclass. Relative expression for each gene in each sample was calculated relative to the expression in the control, and differential expression was calculated by comparing the mean relative expression for each gene in the LTS versus the STS group.

Data analysis

Following hybridization, fluorescence for each array element was quantified and the resulting data were examined using Affymetrix GCS. Quality control checks were verified and the array data were normalized using the Affymetrix GCS software. The average value of the normalized fluorescence intensity for each gene in the technical replicates of the control sample was computed, and the ratio of the normalized fluorescence intensity for each gene in each GBM sample was calculated relative to the corresponding average control intensity. This ratio

| Table 3
| Summary of molecular genetics of samples by survival group |
| Mean STS | Mean LTS | p value (STS vs LTS) |
| 1p loss | 0.00% | 0.00% |
| 1q loss | 15.40% | 0.00% |
| EGFR amplified | 33.33% | 14.29% |
| Ki-67 index (max) | 33.78 | 12.20 | 0.13 |
| p53 (max) | 17.78 | 49.17 | 0.11 |
| STS, short-term survivor group; LTS, long-term survivor group; 1p, chromosome 1p; 19q, chromosome 19q; EGFR, epidermal growth factor receptor. |
was log2 converted, and the log2(FI_sample/FI_average control) for each gene in each sample was used in subsequent analyses.

Expression data were analyzed using TIGR Multiple Experiment Viewer [28]. Euclidean distance metrics were used for all calculations, and default algorithm parameters were used unless otherwise indicated. Genes for which expression data were available in at least 94% of samples were included for subsequent analysis. Filtering for statistically significant genes was performed using the Student t test and the SAM algorithm [29]. HCL of experiments using average linkage clustering was used to group samples in an unsupervised fashion [30]. The STA uses bootstrapping of samples (resampling with replacement) to assign relative likelihoods to hierarchical clustering dendrograms and was used to validate the HCL results [31]. Additional unsupervised clustering was performed using the SOTA [32,33], PCA [34], and CoA [35]. Supervised analysis for class discrimination was conducted using KMC with k-means support [36], SOM [37,38], and TM [39]. Supervised prospective class assignment was performed using KNNC [40], DAM [41], USC [42], and an SVM [43,44]. A random-number generator was used to select 3 members of the LTS group and 4 members of the STS group as the training set for all prospective classification algorithms, and the remaining 13 samples (4 LTS, 9 STS) were treated as unknowns and were used to test the classifiers.

Acknowledgments

This research was generously supported by the Research Projects Committee, the Karen Colina Wilson research endowment within the Brain Tumor and Neuro-Oncology Center, and the Melvin Burkhart Chair in Neurosurgical Oncology at the Cleveland Clinic Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2008.01.002.

References
