

# Identification of noninvasive imaging surrogates for brain tumor gene-expression modules

Maximilian Diehn<sup>\*†</sup>, Christine Nardini<sup>\*</sup>, David S. Wang<sup>\*</sup>, Susan McGovern<sup>‡</sup>, Mahesh Jayaraman<sup>§</sup>, Yu Liang<sup>¶</sup>, Kenneth Aldape<sup>‡</sup>, Soonmee Cha<sup>||</sup>, and Michael D. Kuo<sup>\*,\*\*\*††</sup>

<sup>\*</sup>Department of Radiology and <sup>\*\*</sup>Center for Translational Medical Systems, University of California San Diego Medical Center, San Diego, CA 92103; <sup>†</sup>Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA 94305; <sup>‡</sup>Department of Neuropathology, University of Texas M. D. Anderson Cancer Center, Houston, TX 77030; <sup>§</sup>Department of Radiology, Brown University, Providence, RI 02912; <sup>¶</sup>Department of Neurological Surgery, Brain Tumor Research Center, University of California San Francisco Medical Center, San Francisco, CA 94143; and <sup>||</sup>Department of Radiology, University of California San Francisco Medical Center, San Francisco, CA 94143

Communicated by Helen M. Ranney, University of California at San Diego, La Jolla, CA, February 9, 2008 (received for review October 16, 2007)

**Glioblastoma multiforme (GBM) is the most common and lethal primary brain tumor in adults. We combined neuroimaging and DNA microarray analysis to create a multidimensional map of gene-expression patterns in GBM that provided clinically relevant insights into tumor biology. Tumor contrast enhancement and mass effect predicted activation of specific hypoxia and proliferation gene-expression programs, respectively. Overexpression of EGFR, a receptor tyrosine kinase and potential therapeutic target, was also directly inferred by neuroimaging and was validated in an independent set of tumors by immunohistochemistry. Furthermore, imaging provided insights into the intratumoral distribution of gene-expression patterns within GBM. Most notably, an “infiltrative” imaging phenotype was identified that predicted patient outcome. Patients with this imaging phenotype had a greater tendency toward having multiple tumor foci and demonstrated significantly shorter survival than their counterparts. Our findings provide an *in vivo* portrait of genome-wide gene expression in GBM and offer a potential strategy for noninvasively selecting patients who may be candidates for individualized therapies.**

cancer | genomics | glioblastoma multiforme | radiogenomics

Recent advances in the molecular analysis of brain tumors have led to an improved understanding of glioblastoma multiforme (GBM) tumor biology and the genomic heterogeneity that typifies the disease (1–7). However, the diagnosis and treatment of GBM is still largely guided by histopathology and immunohistochemistry, approaches that group histologically similar tumors that can often demonstrate markedly distinct clinical behaviors. Overall survival remains poor, with most patients succumbing to their disease within 15 months of diagnosis. Methods that assess molecular differences between GBMs hold promise for improving outcome by potentially allowing for individualized patient management.

Magnetic resonance imaging (MRI) is routinely used in the diagnosis, characterization, and clinical management of GBM (8). It is a powerful and noninvasive diagnostic imaging tool that allows global assessment of GBMs and their interaction with their local environment. In its ability to extract structural, compositional, physiological, and functional information, MRI captures multidimensional, *in vivo* portraits of GBMs. Interestingly, histologically similar tumors often demonstrate highly distinct imaging profiles on MRI (9). Recently, several studies have attempted to correlate imaging findings with molecular markers, but no consistent associations have emerged, and many of the imaging features that characterize tumors currently lack biological or molecular correlates (10–15). Much of the information encoded within neuroimaging studies therefore remains unaccounted for and incompletely characterized at the molecular level. We reasoned that the phenotypic diversity of GBM captured by neuroimaging reflects underlying inter- and intratumoral gene-expression differences and that these relationships could be uncovered by combining genome-scale gene expression

and MRI. Here, we show that this approach reveals previously unknown associations between molecular properties of tumors and their appearance by radiologic imaging. Furthermore, we identified an imaging phenotype that is associated with overall survival of GBM patients.

## Results

### Identification of Imaging Surrogates for Gene-Expression Modules.

To test the hypothesis that phenotypic diversity of GBM captured by neuroimaging reflects underlying inter- and intratumoral gene-expression differences, we created a radiogenomic map using an integrative analysis of microarray gene-expression patterns and imaging profiles from pretreatment MRI studies for 22 GBMs. Each MRI was evaluated by two expert radiologists across 10 distinct radiophenotypes representing a spectrum of imaging characteristics seen in GBM. These imaging phenotypes captured aspects of tumor physiology, morphology, cellularity, and composition as well as the interaction of the tumors with their local environment [see supporting information (SI) *Methods*]. We then determined the relationship between each imaging phenotype and the respective gene-expression profiles previously assessed by DNA microarray analysis (3) using a two-step algorithm that accounted for multiple hypothesis testing. Because the imaging data were of significantly lower dimensionality than the gene-expression data, we focused on seven previously defined gene-expression modules related to known biologic processes (see *Methods*).

A high-level view of the data revealed a dense and robust association between radiophenotypes and GBM gene expression. Expression of 1,089 of the 2,188 cDNA clones previously shown to be differentially expressed among these GBMs significantly correlated with at least 1 of the 10 imaging phenotypes. Furthermore, 5 of the 10 imaging phenotypes were significantly associated with at least one of the seven gene-expression modules (Fig. 1A). The link between radiophenotype and gene expression appeared relatively insensitive to gene-expression module size, because we identified significant associations with modules as small as 10 clones (EGFR) and as large as 224 clones (immune cell infiltration). Most of the gene-expression signatures were best captured by unique imaging traits, suggesting specificity in the association between radiophenotypes and the

Author contributions: M.D. and M.D.K. designed research; M.D., C.N., D.S.W., S.M., M.J., Y.L., K.A., S.C., and M.D.K. performed research; C.N. contributed new reagents/analytic tools; M.D., C.N., M.J., and M.D.K. analyzed data; and M.D. and M.D.K. wrote the paper.

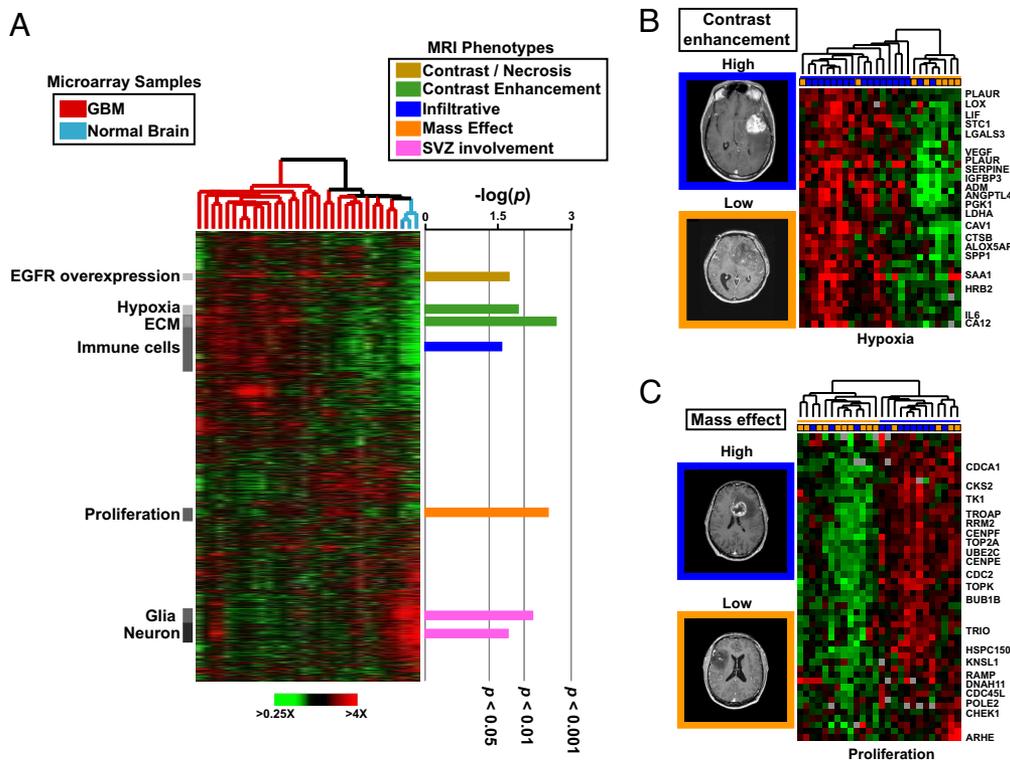
Conflict of interest statement: M.D.K. has filed a patent application in an area related to this field of work.

Freely available online through the PNAS open access option.

<sup>††</sup>To whom correspondence should be addressed. E-mail: mkuo@ucsd.edu.

This article contains supporting information online at [www.pnas.org/cgi/data/0801279105/DCSupplemental](http://www.pnas.org/cgi/data/0801279105/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA



**Fig. 1.** Gene-expression surrogates for MRI traits. (A) Unsupervised hierarchical clustering of 32 samples including GBMs and normal brain specimens was performed as described (3). Gene clusters with common functional or biological themes were identified and are indicated to the left of the expression map. Rows represent genes (unique cDNA elements), and columns represent experimental samples. By using a correlation-based approach, statistically significant positive associations between MR imaging phenotypes and the gene clusters were identified (see *Methods*). The imaging trait displaying the best fit for each gene-expression module is indicated to the right of the expression map. Permutation-based  $P$  values are shown by using colored bars. Statistical significance is represented in terms of  $-\log_{10}(p)$ , and several cut-off values are indicated. (B) Expanded view of the association between the hypoxia gene-expression module and contrast enhancement. Tumor arrays were clustered by using only cDNA clones contained within the gene module. The value of the imaging trait for each tumor is indicated by the colored box above the expression map. Representative MR images are depicted on the left. A subset of named genes is labeled. (C) Expanded view of the association between the proliferation gene-expression module and mass effect. Data are displayed as in B.

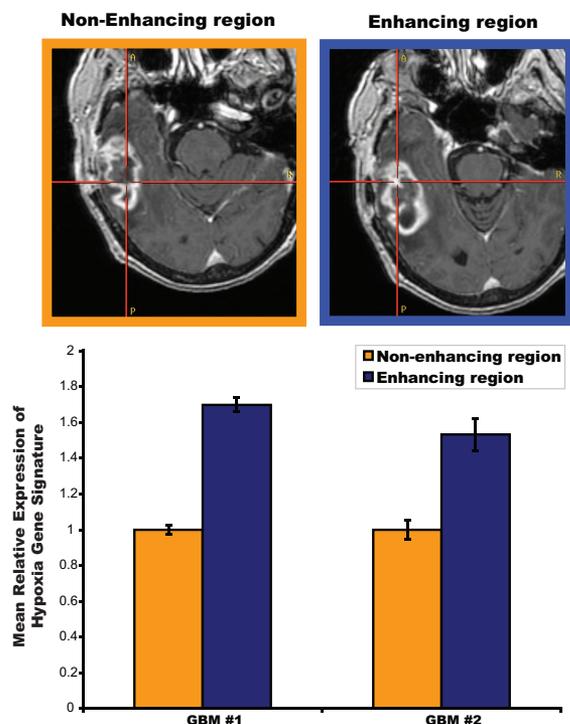
underlying gene-expression signatures. Moreover, when one imaging trait was associated with more than one gene module, the modules displayed very similar gene-expression patterns (e.g., association of SVZ involvement with neuronal and glial gene clusters.) Thus, a relatively large fraction of the molecular heterogeneity of GBM could be captured by using only a small number of MR imaging features.

A more detailed view of our map revealed biologically interesting insights between the imaging phenotypes and gene-expression modules. Activation or repression of specific gene-expression programs in GBM could be inferred from their characteristic radiophenotypes. For example, activity of the hypoxia module, which contains genes implicated in angiogenesis and tumor hypoxia (e.g., *VEGF*, *ADM*, *PLAUR*, *SERPINE1*, *CA12*) (16), was associated with the contrast enhancement (CE) imaging phenotype ( $P = 0.012$ ) (Fig. 1B). Given that hypoxia is closely linked to angiogenesis and inadequate blood flow through poorly formed tumor neovasculature (17), the correlation of CE with the hypoxia module likely reflects the presence of dysfunctional vasculature in hypoxic regions of tumors (18). Given the emerging role of targeted antiangiogenic drugs in the treatment of gliomas (19), this radiophenotype could potentially be used as an imaging biomarker for selecting patients for antiangiogenic therapy.

Similarly, a strong association between the mass effect radiophenotype and the proliferation gene-expression signature ( $P = 0.0017$ ) was uncovered. The proliferation cluster contains genes involved in proliferation and cell-cycle pro-

gression (e.g., *TOP2A*, *CDC2*, and *BUB1B*) (20). As such, it serves as a molecular marker of the proliferation rate of tumor cells. As seen in Fig. 1C, hierarchical clustering of tumors based solely on the expression of this module resulted in segregation of the majority of low- and high-mass-effect tumors to opposite branches of the dendrogram. Interestingly, tumor size alone did not significantly correlate with the proliferation cluster (data not shown). Mass effect may be a better surrogate for tumor proliferation rate than tumor size because it incorporates the response of surrounding brain tissue to a rapidly growing tumor. Thus, clinically and biologically relevant gene-expression signatures can be assessed noninvasively by using specific imaging phenotypes.

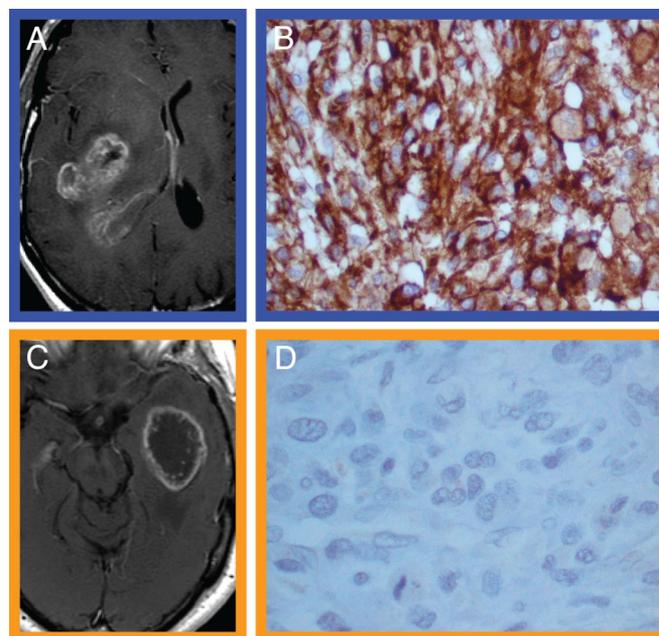
**Analysis of Spatial Gene-Expression Differences Within Tumors.** Although intertumoral gene-expression differences among GBMs are significantly greater than intratumoral differences, we aimed to test whether we could image spatial differences in gene expression within a tumor using MRI. We therefore evaluated the imaging and gene-expression patterns from paired stereotactic biopsy samples from several GBMs. Tissue sample selection was guided by imaging, with samples obtained from CE and non-CE regions within the same tumor (3). Consistent with our previous analyses, genes within the hypoxia cluster were differentially expressed within the same tumor according to the relative extent of CE within each region (Fig. 2). This indicates that spatial gene-expression differences within tumors can be visualized by radiologic imaging.



**Fig. 2.** Intratumoral contrast enhancement variability reflects spatial expression of the hypoxia gene-expression module. For two tumors, stereotactic biopsies were performed from contrast enhancing and nonenhancing regions, and gene-expression profiles were generated separately for each biopsy. The bar graphs (Lower) display the mean relative expression of the hypoxia gene-expression signature in each of the biopsied regions. The expression differences between the two regions were statistically significant for both tumors ( $P < 0.002$ ). Representative MR images of the biopsied regions are depicted (Upper).

**Validation of the Association of EGFR Overexpression with Contrast-to-Necrosis Ratio.** Given the emerging role of targeted molecular therapies in the treatment of GBM, we sought to determine whether protein expression of a therapeutic target could be predicted based on its imaging-gene-expression association. Because several drugs targeting EGFR are in clinical use, and because these may produce reduction of tumor size in select GBM patients (21, 22), we focused on the association between the overexpression of EGFR and the ratio of the contrast-enhancing volume to the necrotic tumor volume (C:N) found in our initial analysis ( $P = 0.019$ ). Specifically, a high C:N ratio correlated with EGFR overexpression. To confirm this association, we evaluated an independent set of 49 GBMs for EGFR expression by immunohistochemistry. The C:N imaging trait predicted EGFR protein overexpression with a sensitivity of 0.92 and specificity of 0.50 in the validation dataset ( $P < 0.002$  by  $\chi^2$ ) (Fig. 3). This finding confirms the link between the C:N imaging trait and EGFR expression, indicates that this imaging phenotype is a surrogate for EGFR overexpression, and suggests that it may be possible to develop imaging-based predictors of treatment response.

**Infiltrative Imaging Phenotype Predicts Patient Outcome.** Having demonstrated the ability of imaging phenotypes to reflect underlying gene-expression programs, we next sought to identify imaging surrogates for gene-expression profiles with prognostic implications. Initial inspection of our radiogenomic map revealed a significant overlap between a survival-associated gene signature and an infiltrative pattern of T2 edema ( $P = 0.007$ ). The infiltrative pattern was differentiated from an edematous



**Fig. 3.** C:N ratio is associated with EGFR protein expression in an independent group of GBMs. EGFR protein levels were assessed by immunohistochemistry (IHC) in an independent set of 49 GBMs for which preoperative MRI scans were available. C:N ratio was scored for each MRI and was found to be significantly associated with EGFR expression ( $P < 0.004$ ) (A) Representative MRI depicting high C:N ratio. (B) EGFR IHC for the patient in A. (C) Representative MRI depicting low C:N ratio. (D) EGFR IHC for the patient in C.

pattern based on the appearance of hyperintense signal on T2-weighted images and reflects the interface between a tumor and the adjacent normal brain. To our knowledge, this radiophenotype has not been previously characterized in this manner in the context of GBM. Gene Ontology term analysis of the infiltrative radiophenotype-associated genes (Fig. 4A) revealed a significant enrichment of terms involved in CNS development, including nervous system development ( $P < 0.003$ ) and gliogenesis ( $P < 0.034$ ). This finding suggests that tumors displaying the infiltrative phenotype may share gene-expression programs with glial progenitors or CNS stem cells.

To verify the prognostic association of the infiltrative radiophenotype suggested by the radiogenomic map, we performed Kaplan–Meier analysis on the original set of tumors. This demonstrated that patients with the infiltrative radiophenotype had significantly worse survival than their edematous counterparts ( $P < 0.02$ ). Other clinical parameters, including age, performance status, and extent of resection, were not significantly different between the groups (Table S1).

Given the potential clinical utility of this finding, we tested our prognostic imaging marker on an independent set of 110 GBMs. Survival analysis confirmed the association of the infiltrative phenotype with poor prognosis ( $P < 3.1 \times 10^{-7}$ , Fig. 4C), with a median survival of 216 days for patients with the infiltrative phenotype compared with 390 days for those with the edematous phenotype. Further, tumors with the infiltrative phenotype were more likely to have multiple tumor foci compared with their edematous counterparts ( $P = 0.009$ ). Our findings suggest that overexpression of the infiltrative radiophenotype-associated genes increases tumor cell invasion into the surrounding brain parenchyma, which is reflected by an infiltrative pattern of T2 hyperintense signal on MR images.

## Discussion

In summary, our results demonstrate that the integration of functional genomic datasets and medical imaging produces *in*



as those on clinical trials) to assess whether the new biomarker predicts outcome or treatment response.

Although our approach offers a number of inherent advantages over current diagnostic molecular and imaging approaches, there are a number of important caveats. First, the dimensionality of imaging, although rapidly increasing over time, is still orders of magnitude lower than that of whole genome gene-expression profiling. Thus, the molecular detail of this approach will likely never match that of conventional molecular profiling approaches. However, the ability of this technique to identify imaging biomarkers for molecular features of tumors should scale as a function of the resolving capability of imaging. Although we uncovered a number of molecular associations with conventional MR images, we expect that application of more advanced imaging techniques (including new contrast agents, pulse sequences, and molecular imaging tools) (8, 32) should facilitate the identification of additional clinically relevant radiophenotypic surrogates for molecular signatures within GBM. Furthermore, the application of more quantitative image analysis tools should also allow for richer image feature extraction and should facilitate the standardization and adoption of these types of imaging biomarkers by decreasing the potential for interobserver bias. A second caveat is that our study was retrospective in nature and relied on previously acquired gene-expression data in a relatively small patient cohort. Thus, our findings will need to be further characterized and validated in future studies, ideally with uniformly treated patients for which gene-expression data are being prospectively generated.

In conclusion, this study demonstrates a simple, widely applicable method for discovering imaging biomarkers that are associated with underlying gene-expression signatures. Our approach facilitates the association of complex molecular signatures with readily identifiable imaging characteristics. Given the noninvasive nature of medical imaging and its wide use in clinical practice, there are a wide range of human disease processes to which this method could be applied. It is likely that identification of imaging phenotypes tied to distinct molecular phenotypes will help to advance individualized patient care.

## Methods

**Patient Samples and Gene-Expression Data.** The details of the tissue samples and GBM gene-expression data and their analysis have been reported (3). All samples used for microarray analysis in the initial publication were from the Brain Tumor Tissue Bank at the University of California (San Francisco). Briefly, 22 GBMs and 3 normal brain samples were analyzed on cDNA microarrays containing  $\approx 23,000$  elements (representing  $\approx 18,000$  unique UniGene clusters) by using standard methods. Array elements that varied at least 2-fold from the median on at least five microarrays were included in Fig. 1 (2,188 cDNA elements representing  $\approx 1,800$  UniGene clusters). Gene-expression clusters were extracted from the hierarchical cluster as described (3).

**Annotated Gene Modules.** Gene-expression modules were curated from the 2,188-gene dataset as described (3). In total, seven annotated gene-expression modules were used in this study: epidermal growth factor receptor (EGFR) overexpression, hypoxia, extracellular matrix (ECM), immune cells, proliferation, glial, and neuronal.

**MRI Image Analysis.** All MRI exams were performed on a 1.5 T Signa Echospeed scanner (GE Medical Systems). The following 10 binary imaging traits were scored for each MRI study after IRB: contrast enhancement, necrosis, mass effect,

pattern of T2 edema (infiltrative/edematous), cortical involvement, SVZ involvement, C:N ratio, contrast/T2 ratio, degree of T2 edema, and T2 heterogeneity. The edematous T2 pattern was defined as classic vasogenic edema of the nonenhancing portion of the tumor where the T2 abnormality has a "pseudopod" appearance and a generally uniform T2 hyperintense signal. The infiltrative pattern was defined as a diffuse, expansile, T2 abnormality beyond the solid enhancing portion of the tumor that is slightly brighter than normal brain and does not have the classic appearance of vasogenic edema. Complete definitions for each of the traits can be found in *SI Methods*. All images were evaluated by consensus in a blinded fashion by two board-certified radiologists (S.C. and M.D.K.) across each of the 10 imaging traits.

**Bioinformatic Analyses.** To assess the association between the imaging traits and gene modules, we used a two-step algorithm. In the first step, we identified sets of cDNA elements (i.e., genes) that were statistically significantly correlated with each imaging trait (hereafter referred to as trait-associated genes.) We thus evaluated the correlation of the  $\log_2(\text{Cy5/Cy3})$  expression ratio for every cDNA clone with each of the imaging traits using a Spearman rank-correlation coefficient. To assess statistical significance and to control for multiple hypothesis testing, we generated 1,000 random permutations of the imaging trait values and recalculated the correlation coefficient for each cDNA element. The distribution consisting of these values was used to determine the correlation coefficient corresponding to a  $P$  value of 0.05. All cDNA elements with absolute real correlation coefficients greater than this cut-off were included in the trait-associated gene sets. In the second step, we determined the overlap of these trait-associated gene sets with the seven previously identified gene-expression clusters (or modules) using the hypergeometric distribution. To evaluate for statistical significance and to control for multiple hypothesis testing, we used the 1,000 random permutations from the above and identified 1,000 "random" trait-associated gene sets for each trait and permutation using the correlation coefficient cut-off identified in the first step. We then generated a null distribution of hypergeometric scores that was used to determine the final  $P$  value for the association between each of the seven gene modules and the trait-associated gene sets (*Dataset S1*).

**Gene Ontology Enrichment Analysis.** Gene Ontology enrichment analysis was performed for each imaging trait-associated gene set by using GoMiner (33).

**EGFR Immunohistochemistry Validation Analysis.** Paraffin-embedded tissues and MRI datasets were from the M. D. Anderson Cancer Center, University of Texas and were obtained with informed consent after approval of the Human Research Committee. To be included, tumors had to have preoperative, diagnostic-quality MRI scans displaying appreciable contrast-enhancing and necrotic components on postcontrast enhanced T1-weighted sequences. Antigen retrieval and immunostaining of EGFR in paraffin samples were performed as described (34) by using the mouse monoclonal antibody clone 528 (1:50 dilution; Oncogene Science). Scoring was semiquantitative and was performed by a single neuropathologist (K.D.A.).

**Infiltrative/Edematous Validation Analysis.** The independent validation dataset consisted of 110 GBM patients from Stanford University and the M. D. Anderson Cancer Center with clinical follow-up and appropriate MRI studies. All data were obtained with approval of the relevant Institutional Review Boards. MR images were scored for the infiltrative/edematous imaging trait as described above. Kaplan–Meier survival analysis was carried out by using WINSTAT ([www.winstat.com](http://www.winstat.com)).

**ACKNOWLEDGMENTS.** We thank R. Tibshirani, W. Cavenee, C. Wixom, and J. Diehn for their helpful discussion and comments. This work was supported by National Institutes of Health Grant NS045013 (to S.C.). M.D. is a recipient of the Leonard B. Holman Research Pathway Fellowship, and M.D.K. is the Bracco Diagnostics RSNA Research Scholar.

- Nutt CL, et al. (2003) Gene-expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 63:1602–1607.
- Freije WA, et al. (2004) Gene-expression profiling of gliomas strongly predicts survival. *Cancer Res* 64:6503–6510.
- Liang Y, et al. (2005) Gene-expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci USA* 102:5814–5819.
- Bredel M, et al. (2005) Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas. *Cancer Res* 65:8679–8689.
- Phillips HS, et al. (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9:157–173.
- Kitange GJ, Templeton KL, Jenkins RB (2003) Recent advances in the molecular genetics of primary gliomas. *Curr Opin Oncol* 15:197–203.
- Mischel PS, Nelson SF, Cloughesy TF (2003) Molecular analysis of glioblastoma: pathway profiling and its implications for patient therapy. *Cancer Biol Ther* 2:242–247.
- Henson JW, Gaviani P, Gonzalez RG (2005) MRI in treatment of adult gliomas. *Lancet Oncol* 6:167–175.

