Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities

Haruka Itakura,1 Achal S. Achrol,2 Lex A. Mitchell,3 Joshua J. Loya,2 Tiffany Liu,1 Erick M. Westbroek,4 Abdullah H. Feroze,2 Scott Rodriguez,2 Sebastian Echegaray,5 Tej D. Azad,2 Kristen W. Yeom,3 Sandy Napel,3 Daniel L. Rubin,1,13 Steven D. Chang,2 Griffith R. Harsh IV,2* Olivier Gevaert1*†

Glioblastoma (GBM) is the most common and highly lethal primary malignant brain tumor in adults. There is a dire need for easily accessible, noninvasive biomarkers that can delineate underlying molecular activities and predict response to therapy. To this end, we sought to identify subtypes of GBM, differentiated solely by quantitative magnetic resonance (MR) imaging features, that could be used for better management of GBM patients. Quantitative image features capturing the shape, texture, and edge sharpness of each lesion were extracted from MR images of 121 single-institution patients with de novo, solitary, unilateral GBM. Three distinct phenotypic “clusters” emerged in the development cohort using consensus clustering with 10,000 iterations on these image features. These three clusters—pre-multifocal, spherical, and rim-enhancing, names reflecting their image features—were validated in an independent cohort consisting of 144 multi-institution patients with similar tumor characteristics from The Cancer Genome Atlas (TCGA). Each cluster mapped to a unique set of molecular signaling pathways using pathway activity estimates derived from the analysis of TCGA tumor copy number and gene expression data with the PARADIGM (Pathway Recognition Algorithm Using Data Integration on Genomic Models) algorithm. Distinct pathways, such as c-Kit and FOXA, were enriched in each cluster, indicating differential molecular activities as determined by the image features. Each cluster also demonstrated differential probabilities of survival, indicating prognostic importance. Our imaging method offers a noninvasive approach to stratify GBM patients and also provides unique sets of molecular signatures to inform targeted therapy and personalized treatment of GBM.

INTRODUCTION

Glioblastoma (GBM) is the most common and lethal primary malignant brain tumor in adults. Upon patient presentation with subacute and progressive neurologic signs and symptoms, gadolinium-enhanced cranial magnetic resonance imaging (MRI) is used as the main diagnostic modality for brain abnormalities (1). Characteristic hypointensity on T1-weighted images and heterogeneous enhancement after contrast infusion strongly suggest GBM. MR images demonstrate the extent and location of tumor involvement, which can determine the feasibility of, and approach used in surgical intervention. Although recent clinical trials are evaluating advanced MRI techniques to improve the assessment of treatment response in GBM (2) or to evaluate changes in tumor blood flow after treatment (3) in known GBM cases, MR images are not currently being used to subclassify GBM risk groups. Moreover, regardless of the imaging findings, a tissue diagnosis is ultimately required for definitive histopathologic confirmation and to distinguish from other primary and metastatic brain tumors. Factors currently known to be associated with survival include age and Karnofsky performance status (KPS) (4), as well as O6-methylguanine–DNA methyltransferase (MGMT) promoter hypermethylation (5) and mutations in isocitrate dehydrogenase 1 (IDH1) or IDH2 (6, 7). Furthermore, gene expression–based molecular classification of GBM (8), epidermal growth factor receptor (EGFR) amplification (9), and CpG island methylator phenotype (CIMP) status (10) have emerged as potential, additional predictors of treatment response and outcome. Although such genomic characterization that encompasses descriptions of gene expression profiles, underlying genomic abnormalities, and epigenetic modification has improved the clinical assessment of GBM (8, 10–12), there remains an unmet clinical need for easily accessible, surrogate biomarkers able to delineate accurately underlying molecular activities and predict response to therapy.

Tumor molecular heterogeneity poses a challenge to the accurate understanding of the underlying molecular activities in GBM (13, 14). Substantial intratumoral heterogeneity requires analysis of multiple regions of a tumor to capture its full clonal history. Recent advances in imaging analysis permit three-dimensional (3D) quantitative characterization of the imaging phenotype of GBM tumors (15–18) that includes this heterogeneity. The emerging field of imaging genomics involves mapping image features to molecular data. In pioneering work, investigators have linked quantitative computed tomography (CT) image features to gene expression data of non–small cell lung cancer to predict survival (19, 20). Similarly, a handful of groups have discovered associations between imaging and gene expression modules in GBM (15), and built models predicting survival by correlating qualitative imaging phenotypes with gene expression data alone (9) or with the addition of microRNA data (21).

In this study, we sought to establish image-based biomarkers of GBM subtypes, ultimately enabling imaging to substitute for intensive molecular
analysis. Such an image-based approach would avoid the risks of biopsy and more comprehensively assess intratumoral heterogeneity. Here, we identify three GBM subtypes differentiated solely by quantitative MRI features and show that these subtypes have prognostic relevance and reflect distinct molecular pathways.

RESULTS

Three MRI GBM subtypes exist

MRI data were obtained in 265 GBM patients, split into two different cohorts: the development cohort for subtype identification and a validation cohort. The development cohort consisted of 121 patients from the Stanford University Medical Center with solitary unilateral tumors evident on MRIs. The validation cohort, which was used to validate findings from the development cohort, was composed of 144 subjects with solitary, unilateral brain lesions from The Cancer Imaging Archive (TCIA). Table 1 summarizes the baseline characteristics of the two cohorts. The selection process did not materially alter the clinical characteristics (age, sex, and KPS) of each cohort from those of the original cohorts (tables S1 and S2). For each subject in each cohort, we applied a quantitative imaging pipeline, extracting gray value histogram statistics, textures, sharpness of lesion boundaries, and metrics of compactness and roughness as described in Materials and Methods, and generated 388 image features representing both 2D and multislice 2D (aggregated 2D slices) characteristics of each lesion (table S3).

On the basis of consensus clustering of the patients’ quantitative imaging features, we chose the solution with $k = 3$ as optimal, using maximal consensus matrices, the consensus cumulative distribution function (CDF) curve, and the CDF progression graph when the Stanford and The Cancer Genome Atlas (TCGA) cohorts were used as development and validation cohorts, respectively (Fig. 1, A to C), and vice versa (Fig. 1, D to F). The three-cluster solution produced the largest $k$ that induced the smallest incremental change in the area under the receiver operating characteristic curve (AUC) (Fig. 1, C and F) while still maximizing the consensus within clusters (Fig. 1, A and D) and minimizing the rate of ambiguity in cluster assignments across 10,000 iterations (Fig. 1, B and E). This resulted in 36 development cohort patients in cluster 1 (30%), 51 patients in cluster 2 (42%), and 34 patients in cluster 3 (28%). When the development and validation cohorts were swapped, the consensus clustering for $k = 3$ yielded 25 patients in cluster 1 (18%), 107 patients in cluster 2 (74%), and 12 patients in cluster 3 (8%).

Next, we used significance analysis of microarrays (SAM) (22) to identify the quantitative image features significantly associated with each subtype. In SAM, each feature represented a statistical descriptor of tumor pixel intensities and characteristics, and together, they defined the multivariate image phenotypes (Fig. 2A and table S4). The top 24 2D and multislice 2D imaging features are in fig. S1. Cluster 1 was characterized by quantitative features that described the high irregularity of tumor shapes and many concavities along the tumor outlines. In addition, we observed that the excluded cases from the development cohort with multifocal lesions ($n = 76$) appeared to resemble features that characterized cluster 1. Therefore, we defined the phenotype of cluster 1 as the “pre-multifocal GBM cluster” (Fig. 2A). Cluster 2 was characterized by spherical tumors with regular edges that were well circumscribed. We defined the phenotype of cluster 2 as the “spherical GBM cluster” (Fig. 2A). Cluster 3 was distinguished by prominence of central hypointensity encompassed by a hyperintense rim. We defined the phenotype of cluster 3 as the “rim-enhancing GBM cluster” (Fig. 2A). The aggregated multislice 2D renditions of the three clusters are shown in Fig. 2B.

In a separate analysis, we evaluated fluid-attenuated inversion recovery (FLAIR) data that we previously analyzed on 30 subjects (15). We determined that one quantitative FLAIR feature (“histogram kurtosis”), which characterizes image homogeneity, was statistically significantly different among the three clusters (Kruskal-Wallis, $P = 0.01575$).

GBM imaging subtypes are confirmed in a heterogeneous validation cohort

We used the in-group proportion (IGP) statistic (23) to validate clusters 1 to 3 in an independent, heterogeneous cohort. The IGP is a measure of

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stanford development cohort ($n = 121$)</th>
<th>TCGA validation cohort ($n = 144$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>64.6 (14.1)</td>
<td>58.5 (15.1)</td>
</tr>
<tr>
<td>Sex, n male (%)</td>
<td>69 (57)</td>
<td>73 (64)</td>
</tr>
<tr>
<td>KPS, n (%)</td>
<td>65 (54)</td>
<td>78 (68)</td>
</tr>
<tr>
<td>&gt;70%</td>
<td>44 (37)</td>
<td>19 (69)</td>
</tr>
<tr>
<td>50–70%</td>
<td>44 (37)</td>
<td>19 (69)</td>
</tr>
<tr>
<td>&lt;50%</td>
<td>11 (9)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>71.7 ± 17</td>
<td>78.6 ± 12.3</td>
</tr>
<tr>
<td>Location, n (%)</td>
<td>Frontal 42 (35)</td>
<td>Parietal 23 (19)</td>
</tr>
<tr>
<td></td>
<td>Basal ganglia 3 (3)</td>
<td>Temporal 49 (41)</td>
</tr>
<tr>
<td></td>
<td>Occipital 4 (3)</td>
<td>EGFR amplification, n (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21 (17)</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100 (69)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27 (22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25 (23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proneural 32 (30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neural 18 (17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mesenchymal 31 (29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Classical 25 (23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing 1</td>
</tr>
</tbody>
</table>
cluster homogeneity and prediction accuracy and indicates whether a cluster in one data set is similar to a cluster in another data set; if the clusters are similar, the IGP approaches 100%. In IGP, the Pearson’s centered correlation coefficient between each datum and each centroid is calculated. The datum is then classified to the group whose centroid has the highest correlation with the datum. The groups are ordered by consensus-clustered groups, depicted as a dendrogram above the heat map. (B and E) The CDF curve was one diagnostic tool used to select the optimal number of clusters in consensus clustering. The bottom left of the graph represents sample pairs rarely clustered together, whereas the upper right contains those almost always paired together. The middle segment represents sample pairs with ambiguous assignments across different clustering runs. The goal was to identify the lowest rate of ambiguous assignments (flat middle segment). (C and F) The delta curve depicts the CDF progression graph, plotting the relative change in area under the CDF curve, comparing k with k + 1. The goal was to select the largest k that induced the smallest incremental change in the AUC.

Secondary analysis including midline-crossing lesions does not change clusters
To address the impact of having excluded midline-crossing lesions, we performed a secondary analysis that included these lesions. The addition of midline-crossing lesions (n = 19) did not alter the distribution of clusters. Cluster assignment remained grossly unchanged in 116 (97%) of the original 121 subjects in the development cohort (table S5). Of the additional 19 subjects, 3 (16%) were assigned to cluster 1, 8 (42%) to cluster 2, and 8 (42%) to cluster 3. The additional samples increased the total number of members in cluster 1 to 39 (28% of 140), cluster 2 to 59 (41%), and cluster 3 to 43 (31%) from 30, 42, and 29%, respectively. Thus, midline-crossing lesions do not form their own cluster but are subsumed under the existing three clusters.

GBM imaging subtypes are prognostic of survival, independent of established or putative clinical and molecular markers
We correlated the imaging-based GBM subtypes with overall survival of patients in the validation cohort treated on the Stupp protocol (25), which is standard first-line therapy consisting of concomitant...
Three distinct image-based subtypes were derived from a development cohort (Stanford cohort) and validated in an independent validation cohort (TCGA cohort). Imaging phenotypes are illustrated as simplified, representative pictograms for each cluster, although the multivariate combination of quantitative images features that characterize each cluster (table S4) cannot be fully visually exemplified. Aggregate multislice 2D renditions of the three imaging subtypes (clusters). Kaplan-Meier survival curves (solid lines) with 95% confidence intervals (CIs) (dotted lines) derived from TCGA survival data are shown for each cluster in the TCGA validation cohort. Survival differences across the clusters: $P = 0.004$, log-rank test ($n = 37$ subjects across the clusters who underwent the same Stupp protocol treatment regimen; cluster 1, $n = 6$; cluster 2, $n = 22$; cluster 3, $n = 9$). Molecular changes associated with each cluster. Arrows indicate up- or down-regulation of sample pathways identified using PARADIGM. Table S6 provides a comprehensive list of significant regulatory pathways associated with each cluster at FDR <5%.

GBM subtypes map to canonical molecular pathways
To estimate various signaling pathway activities as they pertain to the three clusters, we used the Pathway Recognition Algorithm Using Data Integration on Genomic Models (PARADIGM) (26) to integrate gene expression and copy number variation (CNV) data of patients in the TCGA validation cohort. All three clusters were significantly associated with one or more regulatory pathways [all false discovery rates (FDRs) <5%] (Fig. 2D, Table 3, and table S6). Notably, the pre-multifocal cluster 1 was marked by only one association: the up-regulation of the c-Kit stem cell factor receptor pathway. The spherical cluster 2 was characterized by the down-regulation of 21 pathways, including c-Kit, VEGFR signaling, PDGFR-α signaling, FOXA transcriptional networks, and Ang/Tie2. Last, the rim-enhancing cluster 3 was differentiated by the up-regulation of 31 pathways, including canonical WNT and PDGFR-β signaling and many of the pathways down-regulated in cluster 2, such as VEGFR signaling, FOXA, and Ang/Tie2 (Table 3 and table S6).

DISCUSSION
We have identified three distinct clusters of unilateral, solitary GBM defined by quantitative MR image features. The clusters of GBM subtypes were discovered in one cohort and validated in a second cohort. In addition to distinguishing imaging phenotypes, or “clusters”—which we named pre-multifocal, spherical, and rim-enhancing—the three clusters demonstrated significant differences in survival probabilities and associations with canonical signaling pathways. Imaging-based markers of disease phenotype may therefore offer actionable knowledge for clinical decision making and therapeutic targeting.
The current clinical standard for disease relapse surveillance after chemoradiation is tracking changes clinically (symptomatic and neurologic) and with MRI (27–29). The initial treatment for newly diagnosed GBM is maximal surgical resection while preserving neurologic function. However, a major challenge in the management of GBM is the limited number of nonsurgical therapeutic options, such as targeted

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cluster 1, pre-multifocal</th>
<th>Cluster 2, spherical</th>
<th>Cluster 3, rim-enhancing</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>64.7 (17.7)</td>
<td>66.0 (11.7)</td>
<td>62.0 (13.0)</td>
<td>0.324*</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>24</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>22</td>
<td>27</td>
<td>20</td>
<td>0.736†</td>
</tr>
<tr>
<td>Mean KPS (SD)</td>
<td>67.7 (19.2)</td>
<td>73.7 (13.9)</td>
<td>72.9 (18.1)</td>
<td>0.439*</td>
</tr>
<tr>
<td>Tumor volume, cm³ (SD)</td>
<td>31.2 (30.1)</td>
<td>22.6 (18.9)</td>
<td>55.9 (32.7)</td>
<td>4.49 × 10⁻⁷*</td>
</tr>
<tr>
<td>MGMT hypermethylation, n</td>
<td>8</td>
<td>12</td>
<td>7</td>
<td>0.747†</td>
</tr>
<tr>
<td>Missing</td>
<td>25</td>
<td>29</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>EGFR amplification, n</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>0.702†</td>
</tr>
<tr>
<td>Missing</td>
<td>24</td>
<td>33</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Location, n</td>
<td></td>
<td></td>
<td></td>
<td>0.0532†</td>
</tr>
<tr>
<td>Basal ganglia</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Frontal</td>
<td>9</td>
<td>20</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Occipital</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Parietal</td>
<td>7</td>
<td>6</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Temporal</td>
<td>16</td>
<td>23</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>(B) Validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>58.6 (14.5)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>15 (37)</td>
</tr>
<tr>
<td>Male</td>
<td>15 (37)</td>
</tr>
<tr>
<td>Missing</td>
<td>11</td>
</tr>
<tr>
<td>Mean KPS ± SD</td>
<td>78.3 ± 10.3</td>
</tr>
<tr>
<td>Missing</td>
<td>7</td>
</tr>
<tr>
<td>Molecular subgroups, n (%)</td>
<td></td>
</tr>
<tr>
<td>Classical</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Mesenchymal</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Neural</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Proneural</td>
<td>7 (17)</td>
</tr>
<tr>
<td>Missing</td>
<td>16</td>
</tr>
<tr>
<td>CIMP, n (%)</td>
<td>4 (10)</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
</tr>
<tr>
<td>IDH1 mutation, n (%)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
</tr>
<tr>
<td>MGMT hypermethylation, n (%)</td>
<td>6 (21)</td>
</tr>
<tr>
<td>Missing</td>
<td>17</td>
</tr>
<tr>
<td>EGFR amplification, n (%)</td>
<td>27 (66)</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
</tr>
</tbody>
</table>
therapies against molecular derangements. Our results show the potential of imaging features to infer up-regulated molecular activities for which targeted therapies exist. If the effectiveness of such therapies can be demonstrated, there may be a role of such agents as adjuvant therapy in the postoperative setting, primary treatment in surgically unsectable cases, and neoadjuvant treatment for the preoperative reduction of tumor volumes.

On the basis of unique associations with canonical signaling pathways, the imaging subtypes have the potential to direct targeted therapy. Specifically, the pre-multifocal cluster 1 was associated with the up-regulation of the c-Kit pathway, whereas the rim-enhancing cluster 3 was characterized by the up-regulation of several pathways, including canonical Wnt, VEGFR, PDGFR, and Ang/Tie2. This suggests the potential use for patients in cluster 1 of targeting c-Kit with tyrosine kinase inhibitors, such as imatinib and dasatinib. Similarly, although it has long been described that GBM overexpresses VEGF-A and is thus amenable to treatment with a VEGF inhibitor (30–37), there is now a strong rationale for the use of bevacizumab—a VEGF inhibitor—or antiangiogenic multikinase inhibitors that target both VEGFR and PDGFR, such as sorafenib and sunitinib, in patients in cluster 3. Clinical trials are currently investigating inhibitors of the Ang/Tie2 pathway, potentially expanding therapeutic options for members of cluster 3 (38). These results provide intriguing data for retrospective validation and eventually for designing prospective clinical trials to confirm these hypotheses.

Conversely, absence of correlation between a cluster and a signaling pathway suggests that the tumors in that cluster lack the potential therapeutic targets of the pathway. For instance, the treatment options described for clusters 1 and 3 would likely be ineffective in cluster 2. Such knowledge makes clinical trials more efficient and spares patients the time and risk of undergoing trials of limited likelihood of success. Recently, Chinot et al. reported the findings of a randomized, placebo-controlled, double-blind, phase 3 trial of bevacizumab as a first-line therapy for patients with newly diagnosed GBM in concert with standard chemoradiotherapy (39) and identified no overall survival difference in response to bevacizumab (40). Because many new therapeutic approaches are often applied to all patients in a diagnostic category without individualization or with suboptimal individualization, potential beneficial effects may be masked. We hypothesize that the rim-enhancing cluster (cluster 3), which up-regulates the VEGFR signaling pathway, would be more likely to respond to bevacizumab.

The proposed imaging-based subtypes also stratified survival in GBM patients treated on the Stupp protocol, which entails concurrent chemoradiotherapy, followed by adjuvant chemotherapy. Moreover, the imaging subtypes conferred survival differences independent of established risk factors, thus offering additional independent prognostic information, similar to identifying significant factors while controlling for established risk factors. The quantitative MR image features used to define the three clusters thus constitute potential biomarkers for survival in GBM.
Our image-based clusters have the potential to be used for non-invasive treatment follow-up. The cluster phenotypes could be used as surrogate markers of underlying molecular activities for disease monitoring and lead to new pathophysiologic understanding. In addition, the multivariate pathway activity profile can potentially be linked to a drug response profile. The association of distinct molecular activities with specific image-based GBM clusters suggests that image features alone may be used to track changes in disease activity, including providing insight into the natural progression of disease. Such understanding currently requires invasive biopsies and intensive molecular assays. The main technical challenge is integrating our multivariate image profile into the clinical workflow to assign patients to the three subtypes. Ideally, our methodology and image profile would be incorporated into the radiologists’ toolbox. This would require manual image segmentation, which is currently labor-intensive.

Our study extended the earlier efforts in imaging genomics analysis of cancers. Previous studies have linked quantitative CT image features to gene expression data of lung cancer (19, 20) and MR image features with genomic data in GBM (15, 17, 21). Whereas previous work focused on genomic analysis before linking with an image phenotype, our study inverted this design by beginning with a characterization of imaging phenotypes of GBM tumors, followed by linking to signaling pathways. We believe that subtypes defined in this manner are more stable and less susceptible to sampling error resulting from assessing a limited portion of a molecularly heterogeneous tumor. Thus, our approach attempted to minimize the bias inherent in molecular sampling of a heterogeneous tumor and maximize the information captured from the entire 3D imaging phenotype of each tumor.

Here, we meticulously selected homogeneous study populations. We selected only those subjects with solitary, unilateral lesions and excluded patients with multifocal or midline-crossing lesions. We eliminated patients with the most advanced tumor growth and imaging characteristics overtly indicative of poor prognosis, such as lesions crossing the midline or multifocal lesions. Extensive tumor growth was also assumed to signify a confluence of molecular activities with less amenability to therapy. Instead, we focused on the subset of patients with the greatest clinical need for imaging biomarkers that could differentiate prognostic groups that is, those with earlier stages of disease and, thus, for whom interventions may have the greatest impact on the outcome and personalization of treatment might be most beneficial. Moreover, the homogeneity of study subjects was designed to maximize signal detection. However, we measured the impact of having excluded midline-crossing lesions by performing a secondary analysis that included these lesions. Subject cluster assignments remained largely unchanged, with a 97% retention of the original assignment (table S5), suggesting that quantitative image features for midline-crossing lesions do not differ materially enough to generate their own separate cluster. Conversely, the inclusion of multifocal lesions remains infeasible given that their analysis would necessitate biopsies at each tumor focus to overcome intratumoral molecular heterogeneity, and multifocal biopsies and molecular features are not routinely acquired or available through TCGA or in local data repositories.

Next, we used MR images from a heterogeneous validation cohort (TCGA) that were acquired on multiple scanner models from three different manufacturers at four different institutions and successfully validated the imaging subtypes in a real-world setting. Swapping the development and validation cohorts did not result in the identification of robust cluster generation (fig. S1), indicating that a homogeneous cohort is necessary for learning subtypes but not for assigning subjects to imaging subtypes. Nevertheless, the successful validation of the clusters in a more heterogeneous environment highlights the robustness of the model.

We also focused only on T1 MRI, the most widely used MR modality for GBM, to allow the broad applicability and use of the model in the clinical setting. We identified one quantitative FLAIR feature that characterizes image homogeneity as significantly different among the three clusters. Larger studies will become possible as more FLAIR data become available in the future. Nevertheless, it is possible that quantitative features derived from other MRI modalities may further enhance the definition of our clusters or elucidate new associations with signaling pathways. The limited availability of multimodal MR data precluded a comprehensive analysis of imaging features at this time.

In summary, the three GBM subtypes discovered here were distinguished by different expressions of molecular pathways and survival probabilities, which open possibilities for target identification and therapeutic development unique to each subtype. Quantitative imaging features may therefore serve as potential biomarkers for defining subtypes of GBM and potentially other cancers.

MATERIALS AND METHODS

Study design

We sought to identify novel subtypes of GBM, differentiated solely by quantitative MRI features, and discover signature canonical signaling pathways that are molecularly associated with each subtype. To do this, we separately analyzed two cohorts of subjects with GBM (Table 1). This study analyzed data that had already been collected and did not involve new patients randomized to different interventions or treatments. Thus, there was no blinding. Also, we analyzed as many subjects as possible, applying stringent inclusion and exclusion criteria to homogenize our study cohorts as was most scientifically reasonable. Because this was not a prospective study, there were no a priori power calculations performed.

In selected patient images, quantitative image features capturing the shape, texture, and edge sharpness of each lesion for each subject were extracted from T1 post-contrast MRI. We used consensus clustering to discover subtypes in the development cohort and performed 10,000 iterations to achieve high robustness and to minimize cluster selection bias resulting from a few iterations. Regions of interest (ROIs) on the MRI study that were too small (<10 pixels) to interpret were excluded. PAM and the IGP statistic were used to validate the reproducibility and robustness of our subtype classification in a second cohort of subjects (validation cohort). To map imaging subtypes to particular molecular signaling pathways, we performed SAM on pathway activity estimates derived from the analysis of TCGA tumor CNV and gene expression data with the PARADIGM algorithm. We used SAM to correct for multiple-hypothesis testing. We mitigated against biases in our analyses by the performance of numerous iterations of consensus clustering to ensure cluster stability, the use of an external validation cohort to validate our cluster findings, and correction for multiple hypothesis testing in SAM. Imaging subtypes were then correlated with survival and molecular characteristics, as well as traditional risk factors of GBM.

For survival analysis, we focused on the TCGA cohort patients treated according to the Stupp protocol. Unlike the TCGA cohort, the Stanford cohort did not have a well-characterized subgroup of...
patients who had undergone a homogeneous treatment plan (Stupp protocol), making the comparison of survival curves less interpretable. Because the Stanford cohort had undergone heterogeneous treatments due to diagnosis over a long period (2001–2010), meaningful survival curves could not be depicted.

**Patient selection**

The study was approved by Stanford’s Institutional Review Board. For the first cohort, we identified 364 single-institution subjects diagnosed with and treated for de novo GBM at the Stanford Medical Center between 2001 and 2010 and selected 121 that met the inclusion criteria below; this was defined as the development cohort. The second was a multi-institutional cohort from the TCGA project, a federally funded cancer-specific repository of clinical, molecular, and imaging data, in which we selected 144 subjects of 575 that met the inclusion criteria; this was defined as the validation cohort. Selection criteria for both cohorts included (i) diagnosis of de novo GBM, (ii) minimum age of 18 years, (iii) availability of preoperative MRI data, and (iv) presence of solitary, unilateral tumors. For the development cohort alone, we further selected patients on the basis of similarity of their MR images with respect to machine model and slice thickness.

**Quantitative imaging pipeline**

MR images for the development cohort were obtained from the Stanford University Medical Center, whereas images for the validation cohort were obtained from TCGA, TCGA’s companion repository of imaging data for cases donated to the TCGA (www.cancerimagingarchive.net). The images from both cohorts underwent the same preprocessing procedures. Two blinded, board-certified neuroradiologists reached consensus regarding a 3D tumor volume by delineating ROIs around areas of enhancement in each T1 post-contrast MR slice. We computed a set of quantitative image features, as previously reported in lung cancer and brain tumors, with respect to machine model and slice thickness.

**Subtype discovery**

We used k-means consensus clustering for unsupervised class discovery on the development cohort to define imaging subtypes (41, 42). We performed 10,000 bootstraps with 80% item resampling on the quantitative image features and used the k-means clustering algorithm with the Euclidean distance metric to examine all resulting clusters from 2 to 10. We selected the number of clusters that yielded the most stable consensus matrices and the most unambiguous cluster assignments across permuted clustering runs. This established the optimal number of intrinsic unsupervised classes as defined by image features in the development cohort. We analyzed the discovered clusters using SAM, where its application on each discovered cluster identified specific image features defining each subtype. Significant associations were corrected for multiple testing using an FDR threshold of ≤15%.

**Validation of clusters**

To validate the reproducibility of the clusters derived from consensus clustering in the development cohort, we used the IGP analysis to demonstrate the existence of these clusters in the validation cohort, as described in Supplementary Methods.

**Mapping canonical pathways to imaging subtypes**

We used the matched molecular data in the TCGA validation cohort to assign canonical signaling pathways to the discovered imaging subtypes, as described in Supplementary Methods.

**Evaluating imaging subtypes for traditional risk factors**

We examined the subtypes for statistically significant differences in historically associated risk factors, as described in Supplementary Methods.

**Statistical analysis**

To ensure cluster stability in our unsupervised analysis, we performed consensus clustering with 10,000 iterations and used CDF and CDF progression graphs to select the optimal number of clusters. Statistical analysis for determining the validity and reproducibility of the three clusters identified in the development cohort entailed the IGP statistic in the validation cohort (Supplementary Methods). When comparing across-cluster clinical and molecular feature differences in each cohort, we performed the Kruskal-Wallis test for continuous variables and the Fisher’s exact test for categorical variables (Tables 1 and 2). Parametric assumptions were not made for continuous variables. [The overall sample size and counts as low as 0 in some cells necessitated the use of the Fisher’s exact test rather than the Pearson χ2 test (43, 44)]. For evaluation of statistically significant imaging and signaling pathway features associated with each cluster, we used SAM, which adjusts for multiple comparisons (Tables S4 and S5). We reported results for FDR <15% for imaging features and FDR <5% for signaling pathways. The Kaplan-Meier survival curves were compared using the log-rank test.

**SUPPLEMENTARY MATERIALS**

www.sciencetranslationalmedicine.org/cgi/content/full/7/303/303ra138/DC1

Methods

Fig. S1. Relative contributions of the top 24 imaging features in characterizing each of the clusters.

Table S1. Clinical characteristics of the Stanford cohort before and after selection of the validation cohort.

Table S2. Clinical characteristics of the TCGA cohort before and after selection of the development cohort.

Table S3. All 2D and multislice 2D quantitative MR image features used for analysis.

Table S4. Two-dimensional and multislice 2D quantitative MR image features significantly associated with each cluster.

Table S5. Cluster assignment by subjects in the development cohort with and without midline-crossing lesions.

Table S6. Regulatory signaling pathways significantly associated with each cluster. References (45–47)

**REFERENCES AND NOTES**

2. Vanderbilt-Ingram Cancer Center, “3T MRI Biomarkers of Glioma Treatment Response,” ClinicalTrials.gov (Vanderbilt-Ingram Cancer Center, National Cancer Institute, 2013).
3. Massachusetts General Hospital, “Bevacizumab with temozolomide PET and vascular MRI for GBM,” ClinicalTrials.gov (Massachusetts General Hospital, 2013).


Acknowledgments: We thank R. Tibshirani for reviewing the statistical methodologies used in our study. Funding: Research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the NIH under award no. R01EB020527, and by the National Cancer Institute under award no. R01CA160251. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. H.I. was supported by the Advanced Residency Training at Stanford (ARTS) Fellowship. S.D.C. acknowledges support for this study from C. and K. Darian and C. Bade. Author contributions: H.I. and O.G. conceived and designed the study; G.R.H., S.D.C., and D.L.R. created the Stanford GBM imaging database; A.S.A., L.A.M., J.J.L., T.L., E.M.W., A.H.F., S.R., S.E., T.D.A., and S.D.C. acquired imaging data and performed annotations and preprocessing; S.N. provided the software for generating quantitative image features; K.W.Y., S.D.C., and G.R.H. provided the single-institution patient cohort; H.I. and O.G. analyzed and interpreted the data; H.I. wrote the manuscript; H.I. and O.G. edited the manuscript; G.R.H., S.N., D.L.R., L.A.M., K.W.Y., T.D.A., S.E., A.H.F., T.L., and S.D.C. provided edits; and all authors read and approved the manuscript. Competing interests: O.G. and H.I. have filed a provisional patent on the method described in this manuscript. The authors declare that they have no competing interests. Data and materials availability: The TCGA molecular data are available at http://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM. Corresponding imaging data are available at www.cancerimagingarchive.net.

Submitted 12 February 2015
Accepted 4 August 2015
Published 2 September 2015
10.1126/scitranslmed.aaa7582

Editor's Summary

Brain images create cancer clusters

When directing therapies toward tumors, a sample of the cancerous tissue is needed to identify molecular targets. For patients with glioblastoma, however, it is invasive to biopsy the brain. Itakura et al. sought to identify noninvasive determinants of tumor phenotype that would potentially correlate with molecular pathways, thus allowing for targeted therapy without such brain invasion. The authors used magnetic resonance imaging to look at solitary, unilateral tumors from 121 glioblastoma patients and then generated nearly 400 unique image features that could be used to describe each tumor. The tumors could be grouped into three different phenotypes or "clusters": pre-multifocal cluster, with highly irregular tumor shapes; spherical cluster, with defined edges; and rim-enhancing cluster, with a hypointense center ringed by hyperintensity. The distinct clusters were further validated in a separate cohort of 144 patients. These clusters could be used to stratify patients not only according to molecular pathways for targeted therapy but also by survival, indicating the potential for such noninvasive image-based quantitative biomarkers to be used for patient prognosis.