MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma

Brian J. Gillb,1, David J. Pisapiaa,1, Hani R. Malonea, Hannah Goldsteinb, Liang Ley, Adam Sonabendc, Jonathan Yunb, Jorge Samanamud, Jennifer S. Simsa, Matei Banu, Athanassios Dovasb, Andrew F. Teichb, Sameer A. Shetha, Guy M. McKannah, Michael B. Sistia, Jeffrey N. Bruce,a,2, Peter A. Simsc,d,2, and Peter Canollb,2

Departments of *Neurological Surgery, †Pathology and Cell Biology, †Systems Biology, and ‡Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032

Edited by Webster K. Cavenee, Ludwig Institute for Cancer Research, University of California, San Diego, La Jolla, CA, and approved July 11, 2014 (received for review March 29, 2014)

Glioblastomas (GBMs) diffusely infiltrate the brain, making complete removal by surgical resection impossible. The mixture of neoplastic and nonneoplastic cells that remain after surgery forms the biological context for adjuvant therapeutic intervention and recurrence. We performed RNA-sequencing (RNA-seq) and histological analysis on radiographically guided biopsies taken from different regions of GBM and showed that the tissue contained within the contrast-enhancing (CE) core of tumors have different cellular and molecular compositions compared with tissue from the nonenhancing (NE) margins of tumors. Comparisons with the The Cancer Genome Atlas dataset showed that the samples from CE regions resembled the proneural, classical, or mesenchymal subtypes of GBM, whereas the samples from the NE regions predominantly resembled the neural subtype. Computational deconvolution of the RNA-seq data revealed that contributions from nonneoplastic brain cells significantly influence the expression pattern in the NE samples. Gene ontology analysis showed that the cell type-specific expression patterns were functionally distinct and highly enriched in genes associated with the corresponding cell phenotypes. Comparing the RNA-seq data from the GBM samples to that of nonneoplastic brain revealed that the differentially expressed genes are distributed across multiple cell types. Notably, the patterns of cell type-specific alterations varied between the different GBM subtypes: the NE regions of proneural tumors were enriched in oligodendrocyte progenitor genes, whereas the NE regions of mesenchymal GBM were enriched in astrocytic and microglial genes. These subtype-specific patterns provide new insights into molecular and cellular composition of the infiltrative margins of GBM.

Glioma | tumor heterogeneity | microenvironment

GBM typically appears as a contrast-enhancing mass, which represents the highly cellular core of the tumor with vascular proliferation and blood–brain barrier breakdown. This contrast-enhancing (CE) region is typically surrounded by a diffuse, nonenhancing (NE) region of abnormal T2/FLAIR signal, which represents edematous brain tissue with varying numbers of infiltrating glioma cells. The primary treatment of GBM is surgical resection, during which the surgeon removes as much of the CE mass as possible. Thus, molecular and genetic profiling of GBM, including the TCGA effort, has predominantly used samples from the CE regions of tumor. However, it is the NE regions of glioma that are left behind after surgery, which neurooncologists must treat and which inevitably give rise to recurrence. Thus, there is immense prognostic and therapeutic significance to understanding the cellular and molecular features of the NE regions of tumor, yet often these areas are not resected and, therefore, have not been directly studied.

There are two major obstacles to this goal. The first is the surgical challenge of radiographically localized sampling of the tumors. MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma...

Significance

Molecular analysis of surgically resected glioblastomas (GBM) samples has uncovered phenotypically and clinically distinct tumor subtypes. However, little is known about the molecular features of the glioma margins that are left behind after surgery. To address this key issue, we performed RNA-sequencing (RNA-seq) and histological analysis on MRI-guided biopsies from the contrast-enhancing core and nonenhancing margins of GBM. Computational deconvolution of the RNA-seq data revealed that cellular composition, including nonneoplastic cells, is a major determinant of the expression patterns at the margins of GBM. The different GBM subtypes show distinct expression patterns that relate the contrast enhancing centers to the nonenhancing margins of tumors. Understanding these patterns may provide a means to infer the molecular and cellular features of residual disease.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE59612).

1J.G. and D.I.P. contributed equally to this work.

2To whom correspondence may be addressed. Email: pc561@columbia.edu, jnb2@columbia.edu, or pas2182@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1405839111/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1405839111
NE tumor margins. The second is the issue of the complex cellular composition that characterizes these regions of diffuse infiltration. In this study, we have addressed both challenges, and associated distinct molecular and cellular features of the NE regions of GBM with the molecular subtype, as defined by the resected CE regions of the tumor.

Results

Image-Guided Biopsies Reveal Distinct Molecular and Cellular Composition in the CE and FLAIR^NERegions of GBM. We collected radiographically localized biopsies from the CE and FLAIR^NE regions of GBM from 69 patients. Histological and immunohistochemical analysis of the samples showed significant differences in the cellular density and cellular composition between the CE and NE samples (Fig. 1). The CE samples had significantly higher cellularity (P < 0.00001) than NE samples and were significantly more likely to contain the histological hallmarks of GBM, including glomeruloid-type vascular proliferation and necrosis (P < 0.00001 for each feature). Conversely, the NE regions showed the histological features of diffusely infiltrating glioma with neoplastic glial cells intermingled with nonneoplastic and reactive cells (SI Appendix, Fig. S1). The cellular composition of the NE samples was variable but showed significantly higher numbers of NeuN^+ neurons compared with CE samples (Fig. 1). In addition to demonstrating that cellular composition varies across different radiographically localized regions of the tumor, these results show that nonneoplastic cells are a major component of the NE regions of GBM and highlight the importance of considering this complex cellular composition when analyzing the expression pattern of diffusely infiltrating gliomas.

We performed RNA-seq analysis on 75 glioma samples from 27 different glioma patients including 39 samples from the CE regions and 36 samples from the NE regions (SI Appendix, Table S1). We also performed RNA-seq on 17 nonneoplastic/normal brain (NB) samples acquired from 11 patients with no oncological history undergoing ventriculoperitoneal shunt placement or resection for seizures. For each CE and NE tumor sample in the RNA-seq dataset, we determined the GBM subtype as reported (4). Briefly, we calculated the Spearman correlation between each RNA-seq profile and the subtyped TCGA profiles from Verhaak et al. (2). We used the median value of the correlation between an RNA-seq profile and the TCGA microarray data for a given subtype as a similarity score for that subtype (Fig. 2 and SI Appendix, Table S1). The majority of CE samples showed the highest correlation with the proneural, classical, or mesenchymal subtypes. In contrast, the majority of samples from NE regions showed the highest correlation with the neural subtype. A heatmap shows the expression level of the GBM subtype classifier genes across all samples (including the 17 NB samples), with the majority of CE samples clustered into three groups showing enrichment in proneural, classical, or mesenchymal genes. In contrast, most NE samples were clustered with the NB samples and were classified as neural or proneural (Fig. 2 and SI Appendix, Table S1). Notably, the NE and NB samples tended to show higher expression in only a subset of the neural classifier genes (SI Appendix, Fig. S2). This subset of neural genes is selectively up-regulated in “neural GBM” and contains genes that are normally expressed by mature neurons and oligodendrocytes (2). Together with the histological and immunohistochemical data showing that the NE samples represented infiltrated brain, these results suggest that the composite gene expression profile of the NE samples contains a significant contribution from nonneoplastic brain cells such as neurons and oligodendrocytes.

Computational Deconvolution of RNA-Seq Data Identifies Cell-Type Specific Expression Profiles. As glioma cells infiltrate the brain, they encounter a uniquely complex and interconnected environment consisting of several different cell types, including astrocytes, neurons, microglia, and oligodendrocytes. Most efforts at expression profiling in glioma, including a large-scale microarray study by TCGA (1), have not addressed the complex cellular composition of the tumor. However, efforts to physically separate the various cell types by flow sorting or other methods face technical challenges and run the risk of perturbing RNA expression levels.

To address these issues, we used a computational approach to deconvolve the composite RNA-seq expression data from homogenized NE samples (5–8). A report in which expression profiles of multiple neural cell types were deconvolved by microarray analysis of brain tissue is particularly relevant to our application (7). Our approach is based on the algorithm described by Kuhn et al. (7), where constrained, least-squares fitting is used to estimate average expression profiles of each cell type based on
multiple homogenous expression profiles. The expression levels of marker genes are used to approximate the fractional composition of each cell type in each sample, allowing us to estimate the average expression profile of each cell type for a set of samples. We deconvolved the NE and nonneoplastic brain RNA-seq profiles into average profiles for each of six cell types, which represent the major lineages present in brain tissue. Specifically, we seeded the algorithm with OLIG2 for oligodendrocyte progenitor-like (OPC-like) cells, CD44 for reactive astrocytes, AQP4 for unreactive gray matter astrocytes, MAL and MOG for mature oligodendrocytes, RBFOX3 (NEUN) and NEUROD6 for neurons, and AIF1 (IBA1) and CD68 for microglia. Our decision to distinguish between two different astrocyte populations (CD44+ and CD44−) was motivated by recent studies showing that astrocytes in normal and pathological brain tissue can be subdivided into CD44+ and CD44− populations, and that these cell types have very different characteristics (9). Furthermore, it has been shown that reactive astrocytes associated with a variety of pathological conditions, including GBM, up-regulate CD44+ populations (10, 11). The expression profile predicted by our deconvolution algorithm, and the neuron-specific expression profiles derived from the deconvolution of the NE glioma samples are remarkably consistent with the cell type-specific expression patterns reported in previous studies and with the expected functionality of these cell types as independently determined by gene ontology analysis. As further validation of the deconvolution analysis, immunohistochemical analysis with antibodies against three neural classifier genes that are predicted to be predominantly expressed in neurons (HPCA, CRYM, and CHN1) showed selective staining of neurons in both NB and glioma samples (SI Appendix, Fig. S3).

**Differentially Expressed Genes Are Distributed in Multiple Cell Types in a GBM Subtype-Specific Pattern.** Although the above analysis demonstrated that nonneoplastic brain cells contribute significantly to the NE expression data, histological analysis showed that the NE glioma samples are distinct from normal brain. Not only do they contain infiltrating glioma cells, but they also show reactive astrogliosis and microgliosis (SI Appendix, Fig. S1). Therefore, to characterize the changes in expression patterns associated with these histological alterations, we performed differential gene expression analysis between NE glioma samples and nonneoplastic brain specimens into the different subtypes, based on the GBM subtype assignment of the CE portion of the tumor and compared the nonneoplastic brain and the NE glioma from each group. We then queried the NE deconvolution data to assess the cellular distribution of the differentially expressed genes in each sample group (Fig. 4). Although the deconvolution algorithm receives no input from the differential expression analysis, we consistently observe a sharp transition in cellular distribution between overexpressed and underexpressed genes. Genes that are highly expressed in the NE tumor tissue relative to nonneoplastic brain are expressed mainly in OPC-like cells, astrocytes, or microglia. Conversely, genes that are highly expressed in nonneoplastic brain were localized primarily to neurons and oligodendrocytes. Furthermore, the cellular distributions of differentially expressed genes in NE tissue were subtype-specific. We found that genes that were increased in NE regions of proneural tumors predominately distributed to OLIG2+ OPC-like cells, whereas the genes that were increased in the NE regions of mesenchymal tumors predominately distributed to CD44+ astrocytes and microglia (Fig. 4). We also performed differential expression analysis...
between nonneoplastic brain and the NE samples from recurrent mesenchymal GBM (the majority of the recurrent tumors were subtyped as mesenchymal). These results revealed a prominent expression of genes mapped to CD44+ astrocytes and microglia and a marked decrease in genes that mapped to OLG2+ OPC-like cells (Fig. 4).

To complement our deconvolution analysis of the differentially expressed genes, we also conducted a pathway analysis for each of the four sample groups. Using iPAGE, a mutual information-based algorithm for gene ontology enrichment analysis, we determined the key pathways that are enriched among differentially expressed genes. The results, shown in Fig. 5, are remarkably consistent with our assessment of the cellular distributions for these genes. For example, iPAGE shows that cellular proliferation ontologies like DNA replication and M phase are enriched among highly expressed genes in proneural NE tissue, consistent with the predominance of proliferating OPC-like cells. However, immune response and inflammatory ontologies are preferentially enriched in the primary mesenchymal NE tissue, consistent with the microglial cellular distribution described above. In recurrent mesenchymal NE tissue, which we find to be largely devoid of OPC-like expression, we do not observe enrichment of cell cycle- and proliferation-related ontologies.

Discussion

We used radiographically localized biopsies and RNA-seq to show that the molecular and cellular composition of NE regions differ significantly from those of the CE regions of a GBM. We also used a computational strategy for deconvolving composite expression profiles into cell type-specific profiles to identify GBM subtype-specific relationships between CE and NE regions. One important implication of these results is that the molecular and cellular characteristics of NE regions of a glioma, which are often left behind after surgery, may be inferred from the molecular subtype of the CE component of the tumor, which is resected during surgery. Understanding these relationships could lead to development of new therapies that target these subtype-specific alterations, which are expressed in residual tumor cells or reactive cells in the tumor microenvironment. For example, our results show that the NE regions of proneural GBM are enriched in genes expressed by OLG2+/OPC-like cells, which most likely represent infiltrating glioma cells. Conversely, the NE regions of mesenchymal GBMs were highly enriched in genes expressed by microglia and CD44+ astrocytes. These findings are consistent with previous studies showing that proneural GBM is highly enriched in OPC genes (17, 18), whereas mesenchymal GBM harbors an inflammatory signature (13, 19, 20). However, our
results provide new insight into how the expression of mesenchymal and inflammatory genes are distributed among different cell types. Notably, the deconvolution analysis shows that within the NE regions of GBM, the majority of classifier genes used to define the mesenchymal subtype are predominantly expressed by microglia, whereas a smaller subset of mesenchymal genes are expressed in CD44\(^+\) astrocytes (SI Appendix, Fig. S4 and Table S4). Similarly, differential gene expression analysis of our samples identified a set of CD44\(^+\) astrocytic and microglial genes that are up-regulated in the NE regions of mesenchymal GBM compared with nonneoplastic brain. Gene ontology analysis shows that these genes are associated with inflammatory and immune response, and include cytokines and cytokine receptors that may mediate reciprocal signaling between glioma cells and microglia in the infiltrative margins of mesenchymal GBM.

Fig. 5. iPAGE gene ontology analysis of differentially expressed genes comparing normal brain (NB) to NE of proneural GBMs (A), NE of classical GBMs (B), NE of Primary mesenchymal GBMs (C), and NE of Recurrent mesenchymal GBMs (D). The gene ontology categories highlighted in red are associated with genes that are highly expressed in the NE tissue relative to normal brain, whereas those in green are associated with genes that are highly expressed in normal brain relative to NE tissue. Gene ontology categories in black are associated with genes that are not differentially expressed. Although cell cycle and proliferation-related pathways dominate the proneural and classical NE, an immune response/inflammatory signature dominates the mesenchymal NE. These expression signatures are consistent with the cellular distributions that we estimated for differentially expressed genes by deconvolution analysis.

Gill et al.
In further support of these subtype specific differences in lineage relationships of the transformed cell populations, the deconvolution showed that EGFR, which is frequently amplified in the classical subtype of GBM, is predominantly expressed by CD44+ astrocytes, whereas PDGFRa, which is amplified or overexpressed in proneural GBM, is predominantly expressed by OPC-like cells (SI Appendix, Tables S3 and S4). These two populations are likely to have different susceptibilities to chemotherapy and radiation, and what is therapeutically effective for one population may be relatively ineffective for the other. Consistent with this idea, the deconvolution analysis also revealed a marked loss of the OLG2+ population in GBM samples that have undergone treatment (Fig. 4). Previous studies have shown that OPCs are highly sensitive to chemotherapy and radiation (24–26). OLG2+ OPC-like glioma cells are more sensitive to radiation than CD44+ astrocytic tumor cells (13). It is also possible the OPC-like glioma cells undergo mesenchymal transformation to CD44+ glioma cells. In either case, such a shift in cellular composition might explain the previously reported tendency for recurrent tumors to acquire a mesenchymal phenotype (20, 27). Future studies with greater numbers of primary and recurrent tumor specimens will be needed to address this possibility.

The computational approach used in this study could be used to characterize the cell type-specific expression patterns of any sample with a heterogeneous cellular composition. However, the samples taken from the margins of glioma, which contain a variable mixture of infiltrating glioma cells and nonneoplastic brain cells, provide a particularly robust and clinically relevant dataset for this type of analysis. The deconvolution of cell type-specific expression profiles offers a means to disentangle the molecular signature of nonneoplastic cells, which can be a minority of the total cell population, from the nonneoplastic populations, which may have significant molecular alterations of their own. In future applications, this approach can be used to refine the analyses of intracellular signaling pathways and transcription networks, which presume expression in the same cell. Conversely, identifying signaling molecules (such as ligand-receptor partners) that are distributed in different cell types may give new insights into paracrine signaling interactions. This type of analysis can be used to identify and access new therapeutic targets that may be expressed by infiltrating glioma cells or by nonneoplastic/reactive cells that also play a role in disease progression. Finally, understanding common patterns that relate the center of the tumor, which is commonly resected during surgery, to the infiltrative margins, which are often left behind, will greatly facilitate the design of clinical trials that target residual disease.

Materials and Methods

This study included 69 adult patients presenting for open surgical resection of glioma. MRI-localized biopsies measuring ~1.0 cm × 0.5 cm × 0.5 cm were obtained before surgical debulking. Sampled regions included areas within the gadolinium enhancing core of the tumors (CE) and areas of non-enhancing, FLAIR hyperintense tissue at the margins of the tumors (NE). Radiographic localization was confirmed by using intraoperative stereotaxis with T1+Gadolinium and FLAIR sequences using the Brainlab Neuro-navigation interface (Brainlab). Samples from each region were divided into two pieces in the operating suite: One piece was immediately flash frozen in liquid nitrogen; the other piece was fixed in 10% (vol/vol) formalin and used for histological and immunohistochemical analysis. A total of 17 nonneoplastic brain tissue samples were collected from 11 patients, with no oncological history, who were undergoing ventriculoperitoneal shunt placement for normal pressure hydrocephalus or surgical resection for seizure control. A small biopsy was obtained at the cortical entry point before passing the ventricular catheter. These samples are referred to as nonneoplastic/normal brain (NB). A detailed description of the methods used for immunohistochemistry, microscopy, RNA extraction, sequencing, computational deconvolution of RNA-seq data, gene ontology, and differential expression analysis are available in SI Appendix, SI Materials and Methods. All RNA-seq data is available on the Gene Expression Omnibus (accession no. GSE59612).

ACKNOWLEDGMENTS. P.A.S. is supported by National Institutes of Health (NIH)/National Institute of Biomedical Imaging and Bioengineering Grant 1K01EB016071. P.C. is supported by NIH/National Institute of Neurological Disorders and Stroke Grant 1R01NS066955. P.A.S. and P.C. are part of the Brain Tumor Ecology Collaborative supported by the James S. McDonnell Foundation. The project was also supported by the Collaborative and Multidisciplinary Pilot Research Award from the Irving Institute for Clinical and Translational Research, Columbia University Medical Center.


Gill et al.