

Published in final edited form as:

Science. 2014 June 20; 344(6190): 1396–1401. doi:10.1126/science.1254257.

Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma

Anoop P. Patel^{1,2,3,4,†}, Itay Tirosh^{3,†}, John J. Trombetta³, Alex K. Shalek³, Shawn M. Gillespie^{2,3,4}, Hiroaki Wakimoto¹, Daniel P. Cahill¹, Brian V. Nahed¹, William T. Curry¹, Robert L. Martuza¹, David N. Louis², Orit Rozenblatt-Rosen³, Mario L. Suvà^{2,3,*;‡}, Aviv Regev^{3,4,5,*;‡}, and Bradley E. Bernstein^{2,3,4,*;‡}

¹Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114 USA

²Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114 USA

³Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

⁴Howard Hughes Medical Institute

⁵Department of Biology, MIT, Cambridge, MA 02139, USA

Abstract

Human cancers are complex ecosystems composed of cells with distinct phenotypes, genotypes and epigenetic states, but current models do not adequately reflect tumor composition in patients. We used single cell RNA-seq to profile 430 cells from five primary glioblastomas, which we found to be inherently variable in their expression of diverse transcriptional programs related to oncogenic signaling, proliferation, complement/immune response and hypoxia. We also observed a continuum of stemness-related expression states that enabled us to identify putative regulators of stemness *in vivo*. Finally, we show that established glioblastoma subtype classifiers are variably expressed across individual cells within a tumor and demonstrate the potential prognostic implications of such intratumoral heterogeneity. Thus, we reveal previously unappreciated heterogeneity in diverse regulatory programs central to glioblastoma biology, prognosis, and therapy.

Tumor heterogeneity poses a major challenge to cancer diagnosis and treatment. It can manifest as variability between tumors, wherein different stages, genetic lesions or

*Corresponding authors: Bernstein.bradley@mgh.harvard.edu (BEB), aregev@broadinstitute.org (AR), Suva.Mario@mgh.harvard.edu (MLS).

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

Supplementary Materials:

www.sciencemag.org

Materials and Methods

Tables S1–S3

Figures S1–S23

References (45–49)

expression programs are associated with distinct outcomes or therapeutic responses (1–3). Alternatively, cells from the same tumor may harbor different mutations or exhibit distinct phenotypic or epigenetic states (4–7). Such intratumoral heterogeneity is increasingly appreciated as a determinant of treatment failure and disease recurrence (8).

Glioblastoma is an archetypal example of a heterogeneous cancer and one of the most lethal human malignancies (9, 10). Intratumoral heterogeneity and redundant signaling routes likely underlie the inability of conventional and targeted therapies to achieve long-term remissions (11–13). These tumors contain cellular niches enriched for distinct phenotypic properties, including transient quiescence and self-renewal (14–16), adaptation to hypoxia (17), and resistance to radiation induced DNA damage (18, 19). DNA and RNA profiles of bulk tumors have enabled genetic and transcriptional classification of glioblastomas (20, 21). However, the relationships between different sources of intratumoral heterogeneity – genetic, transcriptional and functional – remain obscure.

Single cell transcriptome analysis by RNA-seq (22, 23) should in principle enable functional characterization from landmark genes and annotated gene sets, relate *in vivo* states to *in vitro* models, inform transcriptional classifications based on bulk tumors, and even capture genetic information for expressed transcripts. To interrogate intratumoral heterogeneity systematically, we isolated individual cells from five freshly resected and dissociated human glioblastomas and generated single cell full-length transcriptomes using SMART-seq (96–192 cells/tumor, total 672 cells; Fig. 1A). Prior to sorting, the suspension was depleted for CD45⁺ cells to remove inflammatory infiltrate. As a control, we also generated population (bulk) RNA-seq profiles from the CD45-depleted tumor samples. All tumors were *IDH1/2* wild type primary glioblastomas (Fig. S1) and three were *EGFR* amplified as determined by routine clinical tests (Table S1). We excluded genes and cells with low coverage (24), retaining ~6,000 genes quantified in 430 cells from five patient tumors and population controls (Table S1). The population level controls correlated with the average of the single cells in that tumor (Fig. S2), supporting the accuracy of the single cell data. Individual cells from the same tumor were more correlated to each other than cells from different tumors (Fig. S2). Nevertheless, correlations between individual cells from the same tumor showed a broad spread ($R \sim 0.2$ – 0.7) (Fig. S2), consistent with intratumoral heterogeneity.

Although our isolation procedures specifically targeted glioblastoma cells, we tested whether our sampling also included normal cells. To distinguish normal from malignant, we attempted to infer large-scale copy number alterations for each cell by averaging relative expression levels over large genomic regions (24). This allowed us to suppress individual gene-specific expression patterns and emphasize the signal of large-scale copy number variations (CNVs). As a control, we included RNA-seq profiles from (bulk) normal human brain (25). Hierarchical clustering of all single cells and normal brain samples identified seven groups with concordant CNV profiles (Fig. 1B). The normal brain sample clustered with 10 single cells that presumably have ‘normal’ copy number. In parallel, unsupervised transcriptional analysis identified 9 outlier cells with increased expression of mature oligodendrocyte genes and down-regulation of glioblastoma genes (Fig. S3,S4). Remarkably, all nine of these expression outliers clustered with the normal brain in the CNV analysis (Fig. 1B). The one additional ‘normal’ cell inferred from this CNV cluster

correlated with a monocytic expression signature (26) (Fig. 1B). None of the remaining 420 cells show similarity to the transcriptional programs of non-malignant brain or immune cell types (Fig. S5) (24). While non-malignant cells are critical components of the tumor microenvironment, the combination of dissociation methods, CD45+ depletion, flow cytometry gating and computational filtering used in this study largely excluded non-tumor cells.

Normalization of CNV profiles using signal from the ‘normal’ cluster revealed coherent chromosomal aberrations in each tumor (Fig. 1C). Gain of chromosome 7 and loss of chromosome 10, the two most common genetic alterations in glioblastoma (20), were consistently inferred in every tumor cell. Chromosomal aberrations were relatively consistent within tumors, with the exception that MGH31 appears to contain two genetic clones with discordant copy number changes on chromosomes 5, 13 and 14. While this data suggests largescale intratumoral genetic homogeneity, we recognize that heterogeneity generated by focal alterations and point mutations will be grossly underappreciated using this method. Nevertheless, such panoramic analysis of chromosomal landscape effectively separated normal from malignant cells.

To interrogate global transcriptional interrelationships, we used multi-dimensional scaling to represent the degree of similarity between the cells in the dataset (Fig. 1D, (24)). In contrast to the chromosome-scale analysis above, we observed extensive intratumoral heterogeneity at the transcriptional level. Although most cells grouped by tumor of origin, there were many examples of cells from one tumor crossing into the transcriptional space of another tumor. Moreover, the transcriptional diversity within each individual tumor was significantly greater than that observed for the normal oligodendrocytes (Fig. S4) or for an *in vitro* model of stem-like tumor propagating glioblastoma cells (GBM6, GBM8) (27, 28) (Fig. S2).

Cell-to-cell variability is also evident in the expression and splicing patterns of signaling molecules such as receptor tyrosine kinases (RTKs), which are important therapeutic targets (29). Mosaic RTK amplification and redundant signaling pathways contribute to targeted therapy resistance in glioblastoma (11, 30, 31). We found mosaic expression for *EGFR*, *PDGFRA*, *PDGFA*, *FGFR1*, *FGF1*, *NOTCH2*, *JAG1* and other surface receptors and ligands in pathways pertinent to glioblastoma (Figs. 1E,S6,S7). Notably, the transcripts encoding such genes are highly expressed in individual cells and in the aggregate profiles, increasing our confidence that their absence reflects true negatives (23). Additionally, multiple *EGFR* truncations and in-frame deletions have been described, including an oncogenic mutant form, *EGFRvIII*, which lacks the extra-cellular domain (de2–7) and is a putative target for immunotherapy (32). Of the three tumors with significant *EGFR* expression in our dataset (MGH26, MGH30, and MGH31), MGH30 expresses *EGFRvIII*. By examining junction-spanning ‘spliced’ reads at the single cell level, we identified cells expressing wild type *EGFR* (7%), *EGFRvIII* (19%), as well as a second oncogenic variant (*de4* (33); 25%) (Fig. S8). These variants were almost mutually exclusive, with just 1–2% of cells co-expressing wild type *EGFR* and *EGFRvIII*. Moreover, several cells lack *EGFR* but express other tyrosine kinase receptors, suggesting potential alternative pathways for proliferative signaling (Figs. 1E,S7). For example, *EGFR* expression is anti-correlated with *PDGFRA* and

PDGFC expression in cells from MGH30 (Fig. S9). These findings suggest that heterogeneous expression and/or mutational status of RTKs and other signaling molecules across individual glioblastoma tumor cells may compromise therapies targeting receptor immunogenicity or RTK signaling.

We next used hierarchical clustering and principal component analysis (PCA) to define four meta-signatures, each comprised of multiple related clusters that coherently vary across individual cells from a given tumor or the full dataset (24) (Fig. 2A). These four meta-signatures were respectively enriched for genes related to cell cycle (Fig. 2B), hypoxia (Fig. 2C), complement/immune response (Fig. S10), and oligodendrocyte function (demarcating the 9 normal oligodendrocytes). We validated the co-expression of meta-signature genes by single cell qPCR on another 91 cells from MGH26 and 76 cells from MGH30 using primers for 24 genes (Fig. S11). Although immune cells are an important component of the tumor microenvironment, expression of complement and immune genes by malignant cells was somewhat unexpected. We validated this result using two approaches. MGH29 harbors a previously described p53 mutation, R248L (C→T) (34) (Table S1). Although coverage of this transcript was relatively low, we identified three cells from MGH29 that clearly contain this oncogenic mutation. All of these cells also expressed C3 and other genes from the complement/immune module (Fig. S12). Moreover, direct examination of additional cells scoring for the complement/immune signature confirmed chromosomal aberrations characteristic of glioblastoma (Fig. 1C). Notably, the module may be more generally relevant as robust co-expression of these genes was detected in multiple cell lines derived from brain tumors in the Cancer Cell Line Encyclopedia (CCLE) (Fig. S13) (35).

Another notable feature of the module analysis is the activity of the cell cycle program in a relatively small proportion of cells from each tumor, ranging from just 1.4% in MGH31 to 21.9% in MGH30 (Table S2, Fig. 2B). These values contrast markedly with the *in vitro* glioblastoma models in which almost 100% of cells scored positively for the cell cycle module, but are relatively consistent with Ki67+ quantifications for these tumors (Fig. S14,S15). We investigated several markers previously linked to quiescence, including *HES1*, *TSC22D1*, and *KDM5B* (36–38). Transcripts for *HES1* were not well detected in our data due to low expression, but *TSC22D* and *KDM5B* showed significantly higher expression in non-cycling tumor cells (Fig. S16A). *KDM5B*, which has been implicated in quiescence and therapeutic resistance in melanoma, was detected in 10–20% of individual cells across all tumors and confirmed to anti-correlate with cell cycle by single cell qPCR in MGH26 (Fig. S16B).

Clustering of genes anti-correlated to the cell cycle meta-signature also revealed a group of 12 genes (Fig. S17A), nine of which were in the hypoxia meta-signature (Fig. 2C). In fact, these meta-signatures were diametrically opposed (Fig. S17B). Although this meta-signature might be influenced by tissue processing procedures, hypoxia has been studied extensively as a stimulus for angiogenesis (17) and transdifferentiation of glioblastoma stem cells into vascular endothelium (39–41). Reordering of the cells by hypoxic module score, which was pronounced in MGH28 and MGH31, demonstrated clear gradients in each sample (Fig. 2C), potentially reflecting variations in the tumor microenvironment that impact oxygen tension, blood supply or nutritional source (17, 39–41). Further studies are needed to understand the

spatial relationships of these transcriptional niches *in vivo*. Thus, *in vivo* microenvironment and genes linked to quiescence may affect dormant and possibly refractory compartments in glioblastoma.

Glioblastomas likely contain a primitive subpopulation of stem-like cells (GSC) with preferential resistance to existing therapies (16, 18). GSCs can be modeled *in vitro* as spherogenic cultures that potentially initiate tumors in mice (15, 27). Glioblastoma is also postulated to contain more differentiated cells (DGCs) that can be expanded as adherent monolayers in serum (27, 40). We established GSC and DGC cultures from three of our tumors (MGH26, MGH28, MGH31). As expected, the GSCs exhibit a stem-like phenotype, express the stemness marker CD133, and propagate tumors in xenotransplantation (Figs. 3A,S18). To identify *in situ* tumor cells with stem-like or differentiated phenotypes, we derived a stemness signature from a consensus set of genes differentially expressed between three respective GSC and DGC culture models (Fig. 3B).

Application of the stemness signature to the single cell transcriptional profiles revealed stemness gradients in all five tumors (Fig. 3D). The stemness gradient is modestly anti-correlated to the cell cycle meta-signature (Fig. 3C), consistent with the notion that stem-like cells divide at lower overall rates (16). Notably, the stemness-differentiation axis was occupied continuously rather than discretely, consistent with the notion that the respective *in vitro* models emulate phenotypic extremes but do not capture the full spectrum of cellular states within a primary tumor.

Genes correlated to the *in vivo* gradient include expected classifier genes from the *in vitro* analysis, as well as additional candidates that may reflect aspects of stemness not evident in the culture model (Fig. S19, red and blue, respectively). These include several transcription factors (TFs), including *POU3F2*, *NFIA* and *NFIB*, which have been implicated in tumor propagation, neural stem cell self-renewal and quiescence (42, 43). The *in vivo* stemness gradient also significantly correlated with the average expression of target genes for these TFs, which we predicted from ChIP-seq data (Fig. 3C). Thus, expression signatures and regulatory circuits derived from GSC and neural stem cell models converge to a coherent gradient of cells within primary glioblastoma, and identify TFs likely to promote stem-like regulatory programs *in vivo*.

We next considered the classification scheme established by the Cancer Genome Atlas (TCGA) (21) to distinguish four glioblastoma subtypes: proneural, neural, classical, and mesenchymal. Although these original definitions were established from bulk tumor profiles, we wanted to explore whether individual cells in a tumor vary in their classification. Based on population level (bulk) expression data, the tumors in our study scored as proneural (MGH26), classical (MGH30) or mesenchymal (MGH28, MGH29) subtypes (Fig. S20). To examine the distribution of subtype signatures across individual cells, we calculated subtype scores for each cell using the classifier gene sets.

All five tumors consist of heterogeneous mixtures with individual cells corresponding to different glioblastoma subtypes (Fig. 4A,B). All tumors had some cells conforming to a proneural subtype regardless of the dominant subtype of the tumor, whereas each of the

other subtypes was below detection in at least one tumor. Single cell qPCR of thirty classifier genes in 167 additional cells from MGH26 and MGH30 (Fig. S21) confirmed the presence of multiple subtypes within these tumors in proportions similar to those identified by single cell RNA-seq. Thus, while population level data detects the dominant transcriptional program, it does not capture the true diversity of transcriptional subtypes within a tumor.

Intratumoral subtype heterogeneity provides potentially important insights into tumor biology. The stemness signature is strongest in individual cells conforming to the proneural ($r = 0.12-0.68$, $p < 0.01$, t-test) and classical ($r = 0.26-0.64$, $p < 0.01$, t-test) subtypes, but underrepresented in cells of the mesenchymal subtype (Figs. 3C,S22), which has been correlated with astrocytic differentiation (21). In contrast, cells of the neural subtype do not correspond to either *in vitro* model (Fig. 3C), but are more similar to normal oligodendrocytes (Fig. 4B). These findings highlight parallels between intratumoral cellular heterogeneity in glioblastoma and cellular diversity in the developing brain, with respective subsets of tumor cells resembling a progenitor compartment, an astrocytic lineage, or an oligodendrocytic lineage. This analysis also revealed “hybrid” states in which a single cell scored highly for two subtypes, most commonly classical and proneural (progenitor states) or mesenchymal and neural (differentiated states). These hybrid states may reflect aberrant developmental programs and/or interconversion between phenotypic states.

Finally, we examined whether subtype heterogeneity is relevant to prognosis (24). We focused on tumors classified as proneural, controlling for *IDH1* status (3, 44) and binning them into three groups: pure proneural tumors without any transcriptional signal for other subtypes, low heterogeneity tumors with modest signal for other subtypes (defined as average expression of the alternative subtype genes $>50\%$ of the median value in the group), and high heterogeneity tumors with stronger signals for other subtypes (defined as $>85\%$ of the median value in the group). We also partitioned the proneural tumors according to the other detected subtype. We found that increased heterogeneity was associated with decreased survival (Fig. 4C,4D). This suggests that the clinical outcome of a proneural glioblastoma is influenced by the proportion of tumor cells of alternate subtypes, and emphasizes the clinical significance of intratumoral heterogeneity.

We have leveraged single cell transcriptomics to characterize heterogeneous gene expression programs within five glioblastoma tumors, and interrelate their transcriptional, functional and (to a limited extent) genetic diversity. These findings have fundamental implications for cancer biology and therapeutic strategies, as signaling molecules relevant to targeted therapy show cell-to-cell variability in expression and isoform selection. Moreover, *in vivo* tumor cells display a spectrum of stemness and differentiation states, variable proliferative capacity, and variable expression of quiescence markers, all of which may confound therapeutic strategies. While population-level methods for glioblastoma classification have provided important prognostic insights, they do not recapitulate the diversity of transcriptional programs present in an individual tumor. Our analysis reveals that tumors contain multiple cell states with distinct transcriptional programs and provides inferential evidence for dynamic transitions. A better understanding of the spectrum and

dynamics of cellular states in glioblastoma is thus critical for establishing faithful models and advancing therapeutic strategies that address the complexity of this disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Erica Shefler for project management, James Kim and Pavel Santos for histology, and Pinky Bautista and Yukako Yagi for slide scanning. A.P.P was supported by NIH/NINDS R25NS065743. I.T. was supported by a Human Frontier Science Program fellowship and a Rothschild fellowship. M.L.S was supported by a grant from Medic Foundation. This research was supported by funds from NIH R01 NS032677 (R.L.M), Howard Hughes Medical Institute, Burroughs Wellcome Fund, Harvard Stem Cell Institute, and Klarman Family Foundation. A.R. is a scientific advisory board member for Syros Pharmaceuticals and a consultant for Cancer Therapeutics Innovation Group. B.E.B. is a scientific advisory board member for Syros Pharmaceuticals and a founder and scientific advisor for HiFiBio SAS. RNA-seq data are deposited in Gene Expression Omnibus, accession number GSE57872.

References and Notes

1. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010; 467:1114–1117. [PubMed: 20981102]
2. Eppert K, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med*. 2011; 17:1086–1093. [PubMed: 21873988]
3. Parsons DW, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008; 321:1807–1812. [PubMed: 18772396]
4. Navin N, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011; 472:90–94. [PubMed: 21399628]
5. Gerlinger M, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012; 366:883–892. [PubMed: 22397650]
6. Driessens G, et al. Defining the mode of tumour growth by clonal analysis. *Nature*. 2012; 488:527–530. [PubMed: 22854777]
7. Schepers AG, et al. Lineage Tracing Reveals Lgr5+ Stem Cell Activity in Mouse Intestinal Adenomas. *Science*. 2012; 337:730–735. [PubMed: 22855427]
8. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. *Nature*. 2013; 501:355–364. [PubMed: 24048068]
9. Stupp R, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005; 352:987–996. [PubMed: 15758009]
10. Sottoriva A, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*. 2013; 110:4009–4014. [PubMed: 23412337]
11. Stommel JM, et al. Coactivation of Receptor Tyrosine Kinases Affects the Response of Tumor Cells to Targeted Therapies. *Science*. 2007; 318:287–290. [PubMed: 17872411]
12. Nathanson DA, et al. Targeted Therapy Resistance Mediated by Dynamic Regulation of Extrachromosomal Mutant EGFR DNA. *Science*. 2013 10.1126/science.1241328.
13. Gilbert MR, et al. A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma. *New England Journal of Medicine*. 2014; 370:699–708. [PubMed: 24552317]
14. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013; 501:328–337. [PubMed: 24048065]
15. Singh SK, et al. Identification of human brain tumour initiating cells. *Nature*. 2004; 432:396–401. [PubMed: 15549107]
16. Chen J, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*. 2012; 488:522–526. [PubMed: 22854781]

17. Li Z, et al. Hypoxia-Inducible Factors Regulate Tumorigenic Capacity of Glioma Stem Cells. *Cancer Cell*. 2009; 15:501–513. [PubMed: 19477429]
18. Bao S, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature*. 2006; 444:756–760. [PubMed: 17051156]
19. Bhat KP, et al. Mesenchymal differentiation mediated by NF-kappaB promotes radiation resistance in glioblastoma. *Cancer Cell*. 2013; 24:331–346. [PubMed: 23993863]
20. Brennan CW, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013; 155:462–477. [PubMed: 24120142]
21. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010; 17:98–110. [PubMed: 20129251]
22. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol*. 2012; 30:777–782. [PubMed: 22820318]
23. Shalek AK, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. 2013; 498:236–240. [PubMed: 23685454]
24. Materials and methods are available as supplementary material on Science Online.
25. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; 45:580–585. [PubMed: 23715323]
26. Novershtern N, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011; 144:296–309. [PubMed: 21241896]
27. Lee J, et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell*. 2006; 9:391–403. [PubMed: 16697959]
28. Wakimoto H, et al. Maintenance of primary tumor phenotype and genotype in glioblastoma stem cells. *Neuro Oncol*. 2012; 14:132–144. [PubMed: 22067563]
29. Cloughesy TF, Cavenee WK, Mischel PS. Glioblastoma: from molecular pathology to targeted treatment. *Annu Rev Pathol*. 2014; 9:1–25. [PubMed: 23937436]
30. Snuderl M, et al. Mosaic Amplification of Multiple Receptor Tyrosine Kinase Genes in Glioblastoma. *Cancer Cell*. 2011; 20:810–817. [PubMed: 22137795]
31. Nathanson DA, et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science*. 2014; 343:72–76. [PubMed: 24310612]
32. Choi BD, et al. Systemic administration of a bispecific antibody targeting EGFRvIII successfully treats intracerebral glioma. *Proc Natl Acad Sci U S A*. 2013; 110:270–275. [PubMed: 23248284]
33. Wang H, et al. Identification of an exon 4-deletion variant of epidermal growth factor receptor with increased metastasis-promoting capacity. *Neoplasia*. 2011; 13:461–471. [PubMed: 21532887]
34. Xu J, et al. Heterogeneity of Li-Fraumeni syndrome links to unequal gain-of-function effects of p53 mutations. *Sci Rep*. 2014; 4:4223. [PubMed: 24573247]
35. Cho MS, et al. Autocrine effects of tumor-derived complement. *Cell Rep*. 2014; 6:1085–1095. [PubMed: 24613353]
36. Sang L, Collier HA, Roberts JM. Control of the reversibility of cellular quiescence by the transcriptional repressor HES1. *Science*. 2008; 321:1095–1100. [PubMed: 18719287]
37. Shostak KO, et al. Downregulation of putative tumor suppressor gene TSC-22 in human brain tumors. *J Surg Oncol*. 2003; 82:57–64. [PubMed: 12501169]
38. Roesch A, et al. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell*. 2010; 141:583–594. [PubMed: 20478252]
39. Ricci-Vitiani L, et al. Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature*. 2010; 468:824–828. [PubMed: 21102434]
40. Wang R, et al. Glioblastoma stem-like cells give rise to tumour endothelium. *Nature*. 2010; 468:829–833. [PubMed: 21102433]
41. Soda Y, et al. Transdifferentiation of glioblastoma cells into vascular endothelial cells. *Proc Natl Acad Sci U S A*. 2011; 108:4274–4280. [PubMed: 21262804]
42. Martynoga B, et al. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev*. 2013; 27:1769–1786. [PubMed: 23964093]

43. Suva ML, et al. Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells. *Cell*. 2014; 157:580–594. [PubMed: 24726434]
44. Noushmehr H, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. 2010; 17:510–522. [PubMed: 20399149]
45. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
46. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12:323. [PubMed: 21816040]
47. Cahoy JD, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci*. 2008; 28:264–278. [PubMed: 18171944]
48. Butovsky O, et al. Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. *Nat Neurosci*. 2014; 17:131–143. [PubMed: 24316888]
49. Ho M, et al. Identification of endothelial cell genes by combined database mining and microarray analysis. *Physiol Genomics*. 2003; 13:249–262. [PubMed: 12644598]

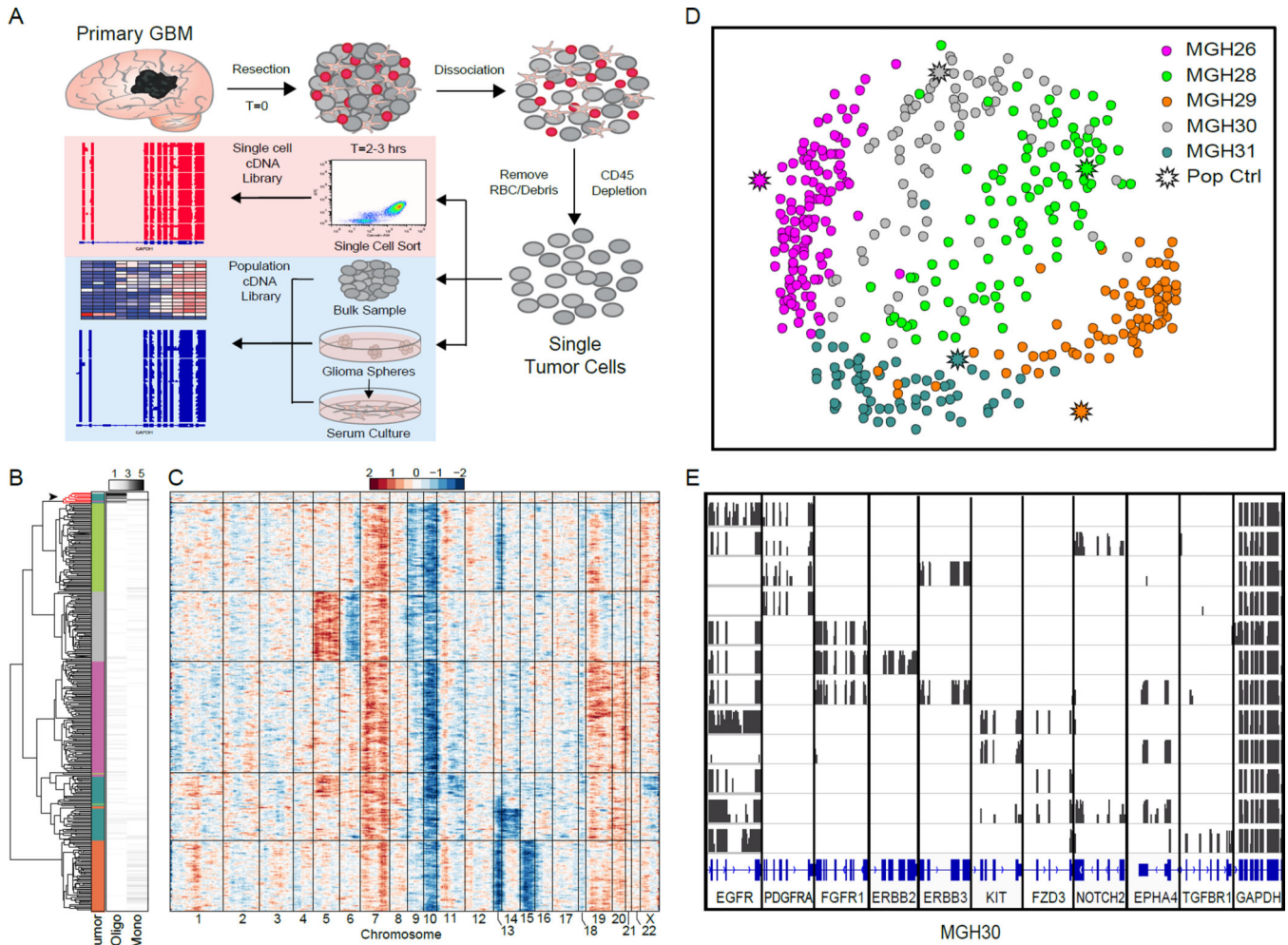


Figure 1. Intratumoral glioblastoma heterogeneity quantified by single cell RNA-seq
(A) Workflow depicts rapid dissociation and isolation of glioblastoma cells from primary tumors for generating single cell and bulk RNA-seq profiles and deriving glioblastoma culture models. **(B)** Clustering of copy number variation (CNV) profiles inferred from RNA-seq data for all single cells and a normal brain sample. Clusters (dendrogram) primarily reflect tumor-specific CNV (colored bar coded as in panel D). Topmost cluster (red, arrow) contains the normal brain sample and 10 single cells, nine of which correlate with normal oligodendrocyte expression profiles and one with normal monocytes ('Oligo' and 'Mono', black and white heatmap). **(C)** Heatmap of CNV signal normalized against the 'normal' cluster defined in (B) shows CNV changes by chromosome (columns) for individual cells (rows). All cells outside the normal cluster exhibit chromosome 7 gain (red) and chromosome 10 loss (blue), which are characteristic of glioblastoma. **(D)** Multidimensional scaling illustrates the relative similarity between all 430 single cells and population controls. The distance between any two cells reflects the similarity of their expression profiles. Cells group by tumor (color code), but each tumor also contains outliers that are more similar to cells in other tumors. **(E)** RNA-seq read densities (vertical scale of 10) over surface receptor genes are depicted for individual cells (rows) from MGH30. Cell-to-cell variability suggests a mosaic pattern of receptor expression, in contrast to constitutively expressed *GAPDH*.

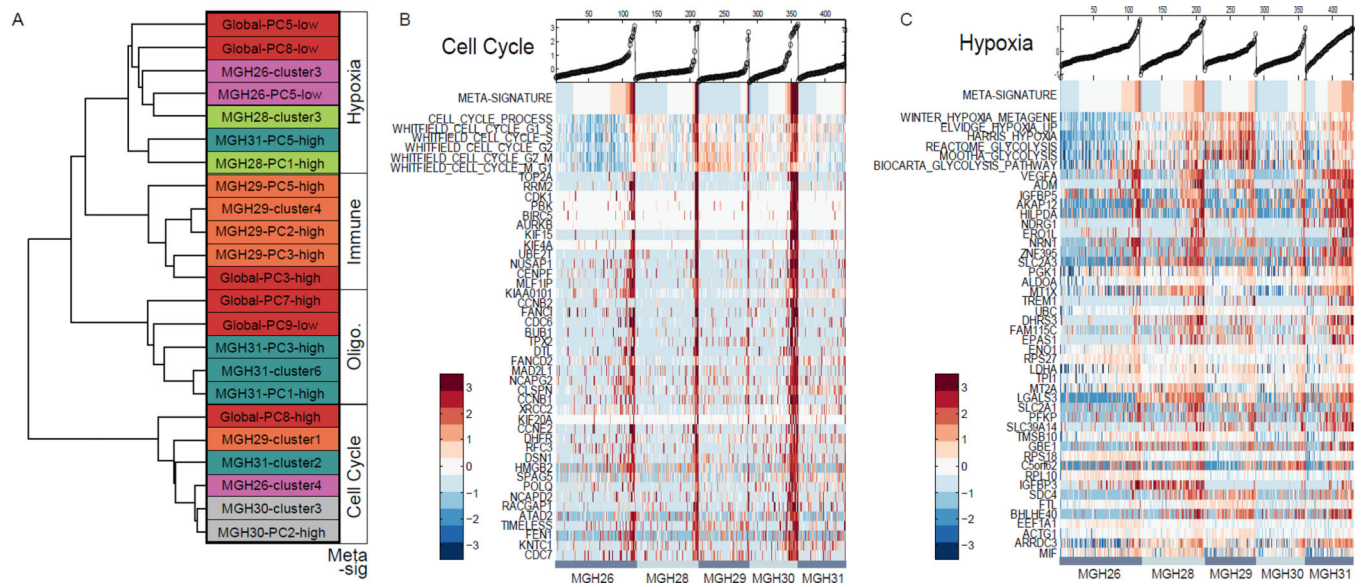


Figure 2. Unbiased analysis of intratumoral heterogeneity reveals coherent transcriptional modules

(A) Gene sets that vary coherently between cells in specific tumors or across the global dataset (colored boxes) were identified by principal component analysis or clustering (24). Hierarchical clustering of these gene sets across all cells (tree) reveals four meta-signatures related to hypoxia, complement/immune response, oligodendrocytes and cell cycle. (B) Heatmap shows expression of the cell cycle meta-signature, selected cell cycle gene sets and representative genes from the signature (rows) in individual glioblastoma cells (columns). Cells were grouped by tumor and ordered by meta-signature score. (C) Heatmap depicts hypoxia meta-signature as in (B).

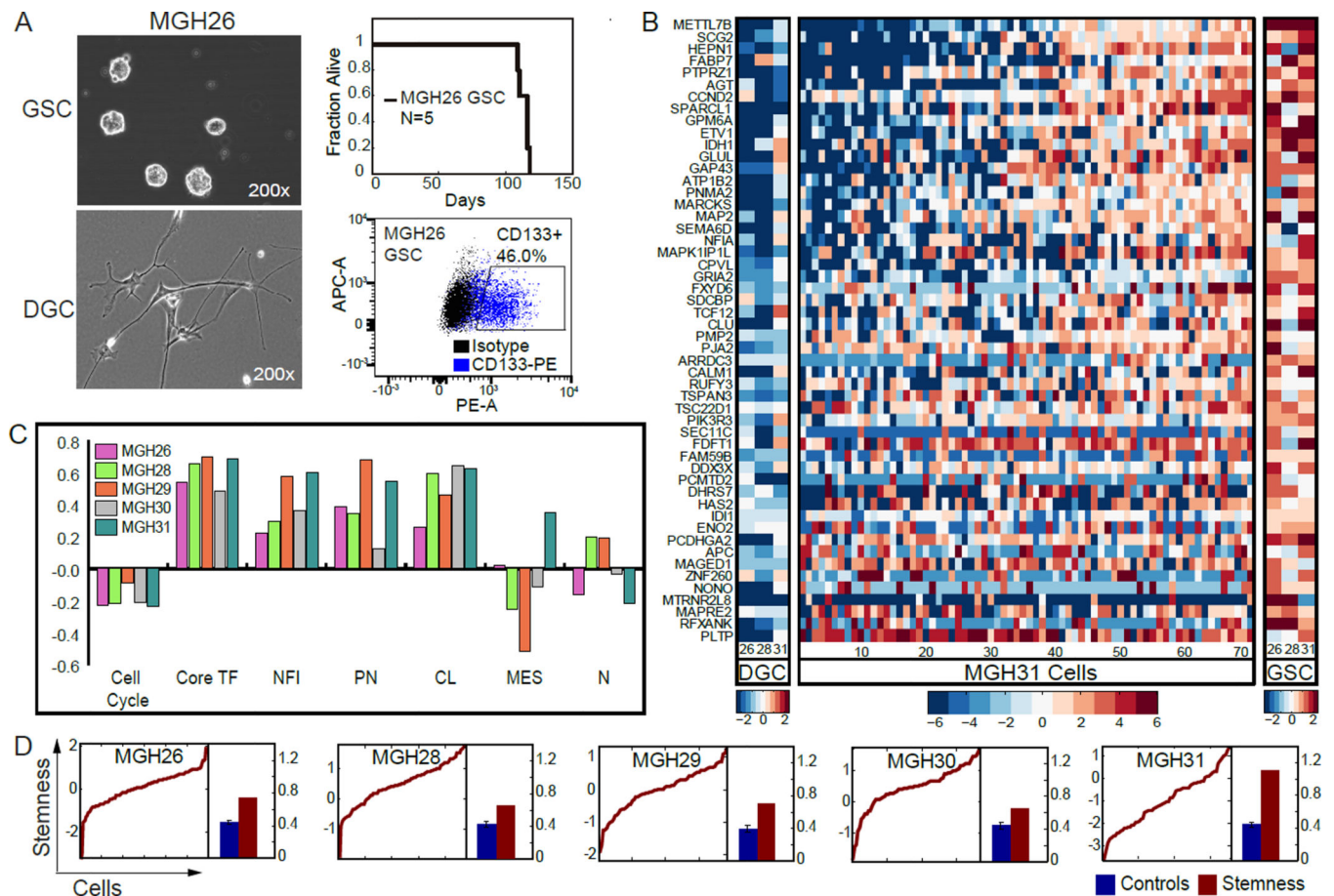


Figure 3. Transcriptional signatures of a stem-like compartment in primary glioblastoma
(A) Stem-like (GSC) and differentiated (DGC) culture models were derived from patient tumor MGH26. GSCs grow as spheres (left, top), initiate tumors in xenotransplantation (right, top), and express the stem cell marker CD133 (right, bottom). **(B)** Heatmap depicts expression of genes (rows) from a stemness signature in differentiated models (DGC, left columns), stem-like models (GSC, right columns) derived from 3 tumors, and in 70 individual cells from MGH31 (middle). **(C)** Bar plot depicts the Pearson correlation coefficient (y-axis) between the stemness signature and selected transcriptional modules in each tumor (colored bars): cell cycle, transcriptional targets of POU3F2, SOX2, SALL2, OLIG2 (core TF) (43), NFI transcriptional targets (NFI) (42), and the proneural (PN), classical (CL), mesenchymal (MES) and neural (N) subtypes defined by the Cancer Genome Atlas (21). **(D)** Plot depicts stemness score (y-axis) computed from stemness signature gene expression in individual cells from each tumor (x-axis) ordered by score. Bar plots depict the overall variance (y-axis, standard deviation) in the stemness score (red) and the average variance of simulated control gene sets (blue), confirming the significance of the gradient.

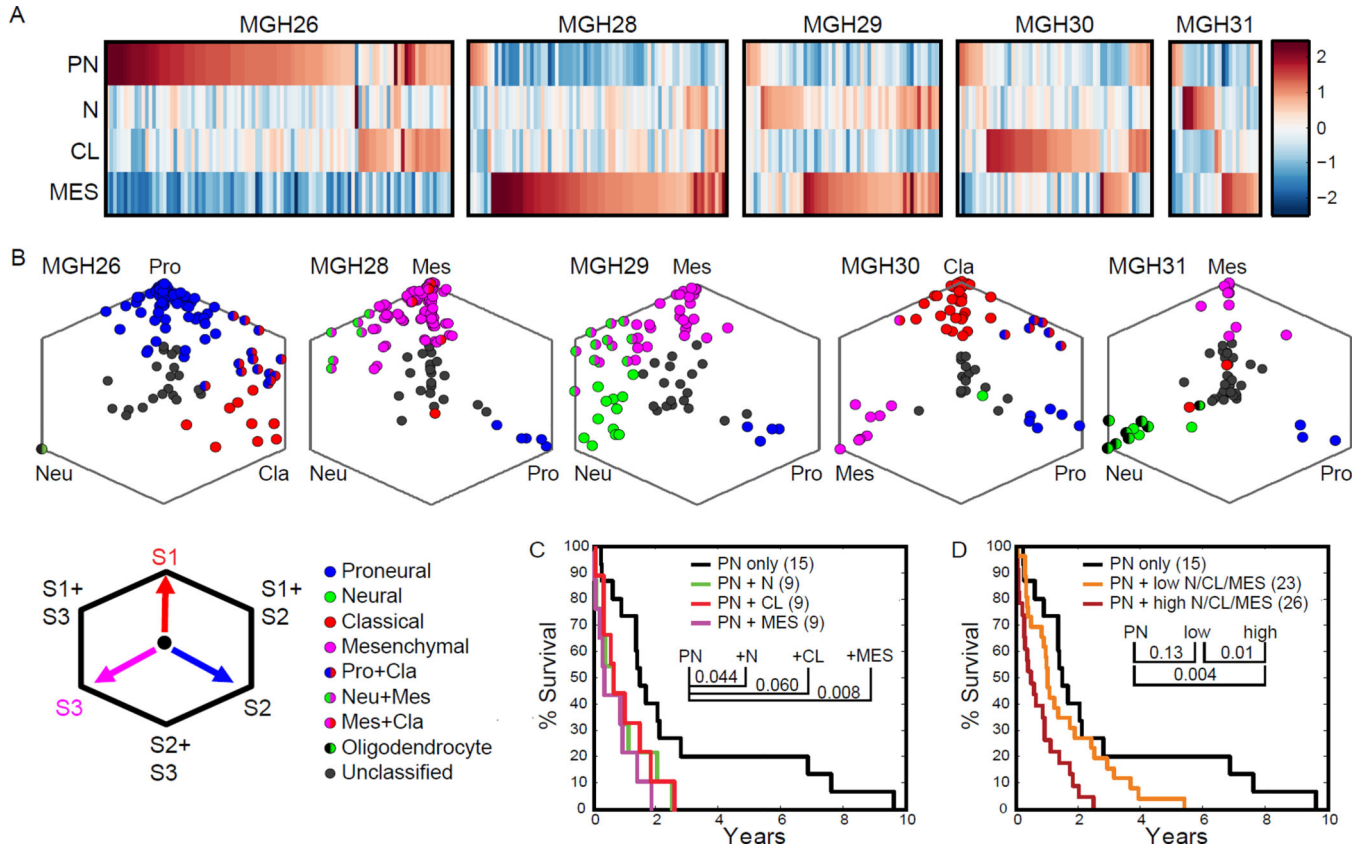


Figure 4. Individual tumors contain a spectrum of glioblastoma subtypes and hybrid cellular states

(A) Heatmap depicts average expression of classifier genes for each subtype (rows) across all classifiable cells grouped by tumor (columns). PN: proneural, CL: classical, MES: mesenchymal, N: neural. Each tumor contains a dominant subtype, but also has cells that conform to alternate subtypes. (B) Hexagonal plots depict bootstrapped classifier scores for all cells in each tumor. Each data point corresponds to a single cell and is positioned along three axes according to its relative scores for the indicated subtypes (Supplemental Methods). Cells corresponding to each subtype are indicated by solid color, while hybrid cells are depicted by two colors. (C) Kaplan-Meier survival curves are shown for proneural tumors from the Cancer Genome Atlas (21). Intratumoral heterogeneity was estimated based on detected signal for alternative subtypes, and used to partition the tumors into a pure proneural group and three groups with the indicated additional subtype. Tumors with mesenchymal signal had significantly worse outcome than pure proneural ($p < 0.05$). (D) Kaplan-Meier survival curves shown for proneural tumors partitioned based on the relative strength of alternative subtype signatures in aggregate (24). Tumors with high signal for alternative subtypes had significantly worse outcome than pure proneural ($p < 0.05$).