www.sciencemag.org/cgi/content/full/science.1254257/DC1

Supplementary Materials for

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

Anoop P. Patel, Itay Tirosh, John J. Trombetta, Alex K. Shalek, Shawn M. Gillespie, Hiroaki Wakimoto, Daniel P. Cahill, Brian V. Nahed, William T. Curry, Robert L. Martuza, David N. Louis, Orit Rozenblatt-Rosen, Mario L. Suvà,* Aviv Regev,* Bradley E. Bernstein*

*Corresponding author. E-mail: bernstein.bradley@mgh.harvard.edu (B.E.B.); aregev@broadinstitute.org (A.R.); suva.mario@mgh.harvard.edu (M.L.S.)

Published 12 June 2014 on *Science* Express
DOI: 10.1126/science.1254257

This PDF file includes:

- Materials and Methods
- Tables S1 to S3
- Figs. S1 to S23
- References

Materials and Methods

Human Tumor Specimens

Surgically removed glioblastoma (GBM) specimens were collected at Massachusetts General Hospital with approval by the Institutional Review Board (Protocols: 2005P001609/16 and 2011P002334). Tissue was mechanically dissociated and then processed into single cell suspension using a papain-based brain tumor dissociating kit (Miltenyi Biotec). Single cell suspensions were run through Lympholyte-H separation (Cedar Lane Labs) to remove RBCs and debris according to manufacturer's specifications. CD45 depletion for inflammatory infiltrate was performed using MACS beads and LS columns (Miltenyi Biotec). H&E sections of tumors were obtained through the MGH tumor bank.

Flow Cytometry and Sorting

Cells were stained for viability (Calcein AM, TO-PRO-3, Life Technologies) and sorted into 96 well plates containing 10μl TCL buffer (Qiagen) + 1% β-mercaptoethanol using a MoFlo XDP or Astrios high speed flow sorter (DakoCytomation). Strict singlets were selected for using pulse area/pulse width gates, followed by selection of highly viable cells that were TO-PRO-3 (APC) negative and Calcein (FITC) positive. Plates were spun down briefly and snap frozen immediately until processing. For staining of cell lines for surface marker expression, CD133/1-PE antibody (Milentyi Biotec) was used according to manufacturer's specifications. Cells were analyzed using an LSRII cytometer (Beckton Dickinson). All analysis was done using FlowJo software (TreeStar).

cDNA Synthesis and Library Construction

SMART-seq protocol was performed on single sorted cells as described (*23*), with some modifications. Briefly, RNAClean XP beads (Agencourt) were used for RNA cleanup prior to reverse transcription and whole transcriptome amplification with the SMARTer Ultra Low RNA cDNA Synthesis Kit (Clontech). Full length cDNA libraries were barcoded using the Nextera XT Tagmentation protocol (Illumina). 96 pooled cells were sequenced as paired-end 25 base reads on a HiSeq2500 (Illumina). For population controls, RNA was isolated from 2,000-10,000 pooled cells using RNEasy Micro Kit (Qiagen) and carried through the same protocol.

Cell Culture and Cell Line Derivation

Single cell suspensions were grown as gliomaspheres in serum-free Neurobasal medium (Invitrogen) supplemented with 3mmol/L L-glutamine (Gibco), 1X B27 supplement (Invitrogen), 0.5X N2 supplement (Invitrogen), 20ng/mL recombinant human EGF (R & D systems), 20ng/mL recombinant human FGF2

(R & D systems), and 1X penicillin G/streptomycin sulfate. GBM6 and GBM8 have been characterized previously(*28*). Differentiated glioma cells (DGCs) were generated by transitioning gliomaspheres to Dulbecco's Modified Eagle's medium supplemented with 10% fetal bovine serum for 7 days.

Xenograft Model

Animal experiments were approved by the Institutional Animals Care and Use Committee (IACUC) at Massachusetts General Hospital. To verify the tumorigenicity of derived cell lines, 50,000 MGH26 GSCs were inoculated into the right basal ganglia of five severe combined immunodeficient (SCID) mice (NCI Frederick) using a stereotactic frame (Kopf Instruments) at coordinates +2.5 mm lateral, +1.0 mm anterior to bregma and -3.0 mm from cortical surface. When mice developed signs of illness, they were sacrificed, the brains removed and fixed in formalin prior to sectioning and staining with H&E.

Single Cell qRT-PCR

Additional banked plates of single sorted cells from MGH26 and MGH30 were thawed and subjected to reverse transcription using random hexamer priming and the SuperScript VILO cDNA synthesis kit (Invitrogen), followed by selective transcriptome amplification using gene specific primers (Table S3). Quantification was performed using EvaGreen SsoFast (Bio-Rad) assay in BioMark 96.96 dynamic arrays according to manufacturer's protocols and visualized using the BioMark HD system (Fluidigm).

$\log_2$ expression levels were normalized by centering the expression level of each gene across the single cells, and then subtracting the average expression of housekeeping genes (GAPDH, ACTB) from all other genes. Cells were then sorted by the average expression of the genes that were selected from each of the modules and subtype predictor sets.

Processing of RNA-Seq data

Paired-end 25bp reads were mapped to the UCSC human transcriptome (hg19) by Bowtie (version 1.4.1, with parameters -n 0 -e 99999999 -l 25 -I 1 -X 2000 -a -m 15 –S) (*44*) , and expression levels of all genes were estimated by RSEM (version 1.2.3, using the option estimate-rspd and default parameters) (*45*) using only the reads with exact matches. TPM values as defined by RSEM were added a value of 1 (to avoid zeros) and then $\log_2$-transformed. To compare population RNA-Seq data to single-cell RNA-Seq data, we aggregated the expression level estimates for all single cells from the same tumor, by averaging the TPM values of these cells, and only then adding 1 and $\log_2$-transforming. We focused on 5,948 genes with the highest composite expression either across all cells combined (average $\log_2(\mathrm{TPM})>4.5$) or within a single tumor (average $\log_2(\mathrm{TPM})>6$ in at least one tumor). We also excluded from our analysis any cells that did not express at least 2,000 of these 5,948 genes. After filtering, our dataset included 430 single cell transcriptomes distributed across five patient tumors, each represented by 70 to 118 cells. Finally, to examine relative expression levels, we centered the data by subtracting the average expression ($\log_2$(TPM+1)) of each gene from all cells.

Copy Number Variation Analysis

We reasoned that averaging the relative expression of a large number of genomically-adjacent genes would average out gene-specific expression patterns and yield profiles that primarily reflect chromosomal copy number variations (CNVs). We sorted all analyzed genes by their genomic locations, first by chromosome (from chromosome 1 to chromosome 22 and ending with chromosome X), and then by the gene start position. We used a moving average of 100 analyzed genes and the following equation to estimate of chromosomal CNVs in each cell and at each analyzed gene:

$CNV_k(i) = \sum_{j=i-50}^{i+50} E_k(o_j)/101$, where $CNV(i)$ is the estimated relative copy number of cell k at the i'th gene in the genomically-ordered list of genes, o_j is the j'th gene in the genomically-ordered list of genes, and $E_k(o_j)$ is the relative normalized expression of that gene in cell k.

Since our refined data includes only ~6,000 analyzed genes, windows of 100 analyzed genes in fact correspond, on average, to large genomic segments that include ~350 genes. We centered the CNV vector of each cell at zero to avoid bias towards high values for high quality cells and low values for low-quality cells. The CNV patterning is often consistent with chromosome borders, with increased or decreased values within specific chromosomes, suggesting that it indeed represents chromosomal changes.

To compare these patterns to an external reference of normal cells we downloaded RNA-Seq data from the GTEX portal (http://www.gtexportal.org/; gene read counts file from Jan. 2013), and estimated CNV values as above: we normalized the read counts into $\log_2$(TPM+1), averaged all brain samples, restricted the data to the ~6,000 analyzed genes, subtracted for each gene the average normalized expression from the GBM single-cell data (this step is comparable to the centering of the single cell data) and then used a moving average of 100 genes over the genomically-ordered list of genes to define CNV_{cont}.

We then clustered all *CNV* profiles (together with CNV_{cont}) using average-linkage hierarchical clustering and found seven prominent clusters. Cluster one contained CNV_{cont} and 10 additional single cells, which we postulate are non-malignant tumor-infiltrating cells. We also compared the expression patterns of all 430 single cells to gene expression signatures of various brain and immune cell types (see below, *Analysis* *of annotated gene sets to identify normal cells*). We found that nine of the cells in cluster one are the top scoring cells for the oligodendrocyte signature, and the tenth cell is the top scoring cell for the monocyte signature. The observation that CNV profiles cluster together human brain, nine (presumably) oligodendrocytes and one (presumably) monocyte, despite their variable gene expression patterns strongly supports that our moving-average approach largely eliminates gene-specific expression signals while maintaining signals of chromosome aberrations. We therefore used the average *CNV* profile of the 'normal' cluster one as a reference to normalize the entire *CNV* dataset.

After subtracting this reference from the CNV profiles, we observed deletions and amplifications of entire chromosomes for the six other clusters (2-7). In particular, all cells in these clusters share an inferred chromosome 7 amplification and an inferred chromosome 10 deletion, both of which are characteristic of glioblastoma (*20*). Notably, four of these clusters (2, 3, 4 and 7) each correspond exactly to a single tumor, suggesting a degree of clonality with respect to these large inferred chromosomal aberrations. Clusters 5 and 6 both corresponded to cells from MGH31, suggesting that this tumor is composed of at least two genetically distinct populations with distinct chromosomal aberrations (e.g., chromosome 5 inferred amplification and chromosome 14 inferred deletion).

Two statistics were used to validate this approach for inferring CNVs from RNA-seq data. First, we performed control analyses on a large number of cancer cell lines (CCLE) for which both RNA-seq and single nucleotide polymorphism (SNP) array data were available. The correlation between CNV as inferred by RNA-seq and the SNP data was 0.72, and the resulting normalized heatmaps are shown in Figure S23. Second, we verified that shuffling the order of genes in the genome before applying the moving average on the single cell RNA-seq data significantly decreased the observed signal of predicted CNVs ($p < 10^{-5}$).

Multi-dimensional scaling (MDS)

We used multi-dimensional scaling to define a two-dimensional representation of single cells and population controls, for which the set of all pairwise distances in that representation provide a good approximation for the "real" pairwise distances, as calculated by $1-r(i,j)$, where r is the Pearson correlation coefficient in comparison of the relative normalized expression of all analyzed genes between cells (or populations) i and j.

Transcriptional modules

An initial set of 122 modules of co-regulated genes were identified by principal component analysis

(PCA) and by hierarchical clustering with average linkage. For each PCA run, 10 PCA modules were defined as (1) the top 100 positively correlated and (2) the bottom 100 negatively correlated genes, for each of the first five PCs. These modules were defined both in analysis of the entire dataset (430 cells) and in separate analysis of cells from each tumor, for a total of 60 modules. Hierarchical clustering was performed separately in each tumor, only for genes that correlate ($R>0.4$) with at least two other genes. A clustering threshold (horizontal cut of the dendrogram) was determined such that the number of clusters would be equal to a third of the number of analyzed genes, and thus each cluster contained, on average, only three genes. The few clusters that were especially large (at least 8 genes) were retained for further analysis.

We used the 122 initial modules to score each cell by the average relative expression of the module's genes minus the average relative expression of all analyzed genes. The average expression of all genes was subtracted as a way to reduce the effects of cell quality, as in cells with higher-quality data we detect more genes and thus the average relative expression tends to be higher regardless of the gene-set (module) that is being used for averaging. We compared modules both by clustering the cell scores and by comparing their top enriched annotations as defined by MsigDB with a hypergeometric test. We identified four groups of modules that have correlated cell scores and similar enrichments. For each group of modules we defined a meta-module by taking all genes that are members of at least two different modules among the group. These meta-modules were used to score the cells by the average relative expression of all the genes in the meta- module minus the average of all analyzed genes, and cells were sorted according to these scores in Fig. 2. We also calculated the correlation of each gene with the metamodule scores, and the most correlated genes are shown in Fig. 2.

Analysis of annotated gene sets to identify normal cells

To examine if any of the single cells in our data correspond to normal (non-tumor) infiltrating cells, we compiled cell type-specific gene-sets for six neuronal cell types that may be expected to interact with a GBM tumor. Gene-sets for oligodendrocytes, neurons and astrocytes were taken from Cahoy *et al*. (*46*), microglia from Butovsky *et al.* (*47*), and endothelial cells from combining the genes from Ho *et al*. (*48*) and the ENDOTHELIAL_CELL_PROLIFERATION term from MSigDB; For immune gene sets, expression signatures were taken from the Differentiation Map portal (DMAP)(*26*). For each of these gene-sets we defined a signature score for each cell as the average relative expression (centered $\log_2(\text{TPM}+1)$) of the genes in the gene-set minus the average relative expression of all genes from the external gene-sets. We then centered and normalized the signature scores to define Z-scores and plotted the sorted Z-scores for the six cell types and 38 immune cell types (Figure S5). This analysis revealed nine outlier cells for the oligodendrocyte signature, one cell for the monocytic signature from DMAP, but no apparent outliers for any of the other cell types. As another test, we also queried the 122 transcriptional modules for annotations that were enriched out of MSigDB datasets that might correspond to other cell types. Multiple modules and the associated meta-module were enriched with oligodendrocyte-related gene sets, and these modules suggested the presence of nine oligodendrocytes in our data. Apart from this, no other associations were discovered.

Analysis of immune module in Cancer Cell Line Encyclopedia

We downloaded the CCLE gene-centric RMA-normalized Affymetrix data (http://www.broadinstitute.org/ccle/), and centered the expression of each gene across all cell lines at zero. We used all CCLE profiles (~1000 cell lines) to cluster all genes with immune-related annotations (based on the "Reactome Immune System" gene-set at MsigDB) as well as the genes in our immune module (Fig. S14A). Hierarchical clustering revealed one prominent cluster of 98 genes that included most of the genes in our immune module (Imm1). In addition to this cluster we observed several additional clusters and in particular examined the largest cluster (Imm2) composed of 236 immune genes.

Next, we scored each cell line by the average relative expression of the two immune clusters, Imm1 and Imm2. Remarkably, we observed almost a complete separation between hematopoietic and solid tumor cell lines, with hematopoietic cell lines scoring high for Imm2 and low for Imm1, and solid tumors scoring high for Imm1 and low for Imm2 (Fig. S14B). Moreover, we found that brain tumor cell lines in the CCLE have the highest expression of Imm1 genes (Fig. S14C). This supports the hypothesis that the module we observed in glioblastoma, which is closely related to Imm1, reflects an expression program in solid tumor cells, which may be linked to inflammation.

Stemness-differentiation gradient

We defined a stemness signature as the set of genes whose expression was higher in the spherogenic cultures compared with the serum-cultured lines, each of which were derived for MGH26, MGH28 and MGH31. We used a threshold of 1.5-fold difference in each of the tumor-derived pairs and 3-fold in at least one tumor, resulting in 52 stemness signature genes. The average relative expression of those genes minus the average relative expression of all genes was used to define a stemness score for each single cell from the tumors. To validate the significance of the stemness gradient we performed simulations in which we randomly selected sets of genes with the same size as the signature and used that random gene set to score the cells with the same procedure used to score stemness. This simulation was repeated a hundred times and we compared the standard deviation of the stemness gradient with the random gene sets.

Subtype analysis

The average relative expression of each set of subtype predictor genes (as defined by TCGA, (*21*)) minus the average relative expression of all genes was used to calculate an initial subtype score for each cell. To estimate the significance of the subtype scores we randomly selected a hundred sets of genes with the same size as each subtype predictor and used these random sets to define a 5% cutoff for the expected subtype scores. We then classified the cells based on the identity of the subtypes for which they passed the 5% threshold.

To evaluate the robustness of each subtype assignment we performed bootstrapping: for each subtype, we defined 1000 sets of predictor genes (samplings), each with equal size to the original set of predictors for that subtype, and derived by random sampling with replacement from the original set of predictors for that subtype (some predictor genes are chosen twice or more and others are not chosen at all in each sampling). We used the number of samplings in which each cell passed the 5% threshold as the confidence in the assignment to the subtype. These confidence scores were then used to visualize the spread of cells from each tumor in a space that reflects the three main subtypes of that tumor. Each cell was initially placed in the center of a triangle, and for each sampling in the bootstrapping process we moved the cells in the direction of the respective subtype, such that the cell would reach a vertex of the triangle only if it passed the 5% threshold in all samplings for one subtype but did not pass the 5% threshold in any of the samplings for the other two subtypes. In this way, three vertices reflect individual specific subtypes while the other three vertices reflect a combination of two subtypes (i.e. a cell would reach those vertices only by passing the threshold in all samplings of two different subtypes, and by not passing the threshold in any of the samplings for the third subtype).

This approach is supported by the observations that (1) in each tumor we find representation for more than one subtype but not for all subtypes - typically three subtypes are represented in a meaningful fraction of cells, and we can exclude the fourth subtype in each tumor; (2) individual cells often score highly for two subtypes but not for three subtypes; if each cell has at most two significant components, then only two of the three directions in the hexagon are relevant for each cell, and there will not be interference with the third direction; we can therefore encode three directions in a two-dimensional plot, since only two dimensions are relevant for each particular cell (while the identity of the relevant dimension differ between cells).

Relations between Heterogeneity and Prognosis

RNA-seq and clinical data were downloaded from the TCGA portal (https://tcga-data.nci.nih.gov/tcga/). We excluded samples with *IDH1* mutations, secondary glioma, prior treatment, or missing prognostic information. For the remaining samples, we centered the expression data of each gene and then calculated the average relative expression of each set of subtype predictor genes. These subtype scores were then used to classify the samples into five classes using K-means clustering. As initial centroids for the Kmeans clustering we used four centroids, each reflecting high expression of one subtype and low expression of all other subtypes, and an additional centroid with a similar expression of all subtypes. This resulted in classification of 64 samples as Proneural, and very similar results were obtained when we ran K-means clustering with random initial centroids.

We then focused on the Proneural-classified samples and ranked them by the score of each of the other three subtypes. Samples among the 50% lowest-scoring for each of the three subtypes were defined as pure proneural tumors. Samples among the 15% highest-scoring for each subtype were defined as high heterogeneity, with a secondary classification of the corresponding subtype. All remaining samples (not among the lowest 50% for all subtypes and not among the top 15% for any specific subtype) were defined as low-heterogeneity samples. We then used a Kaplan-Meier survival curve and a Cox proportional hazard regression model to compare the survival (days to death) between the different sets of Proneural samples.

Supplemental Tables

Sample number	Age	Final Path	IDH	EGFR	PDGFR	MET	MGMT	1p19q	Other Mutations	Number of Cells
MGH26	61	GBM (IV)	WT	++	NT	WT	M	WT	NT	118
MGH28	60	GBM (IV)	WT	WT	NT	WT	NT	NT	NT	94
MGH29	60	GBM (IV)	WT	WT	WT	WT	M	NT	p53 R248L	75
MGH30	55	GBM (IV)	WT	++	WT	WT	NM	NT	WT	73
MGH31	70	GBM (IV)	WT	++	WT	WT	NM	NT	WT	70

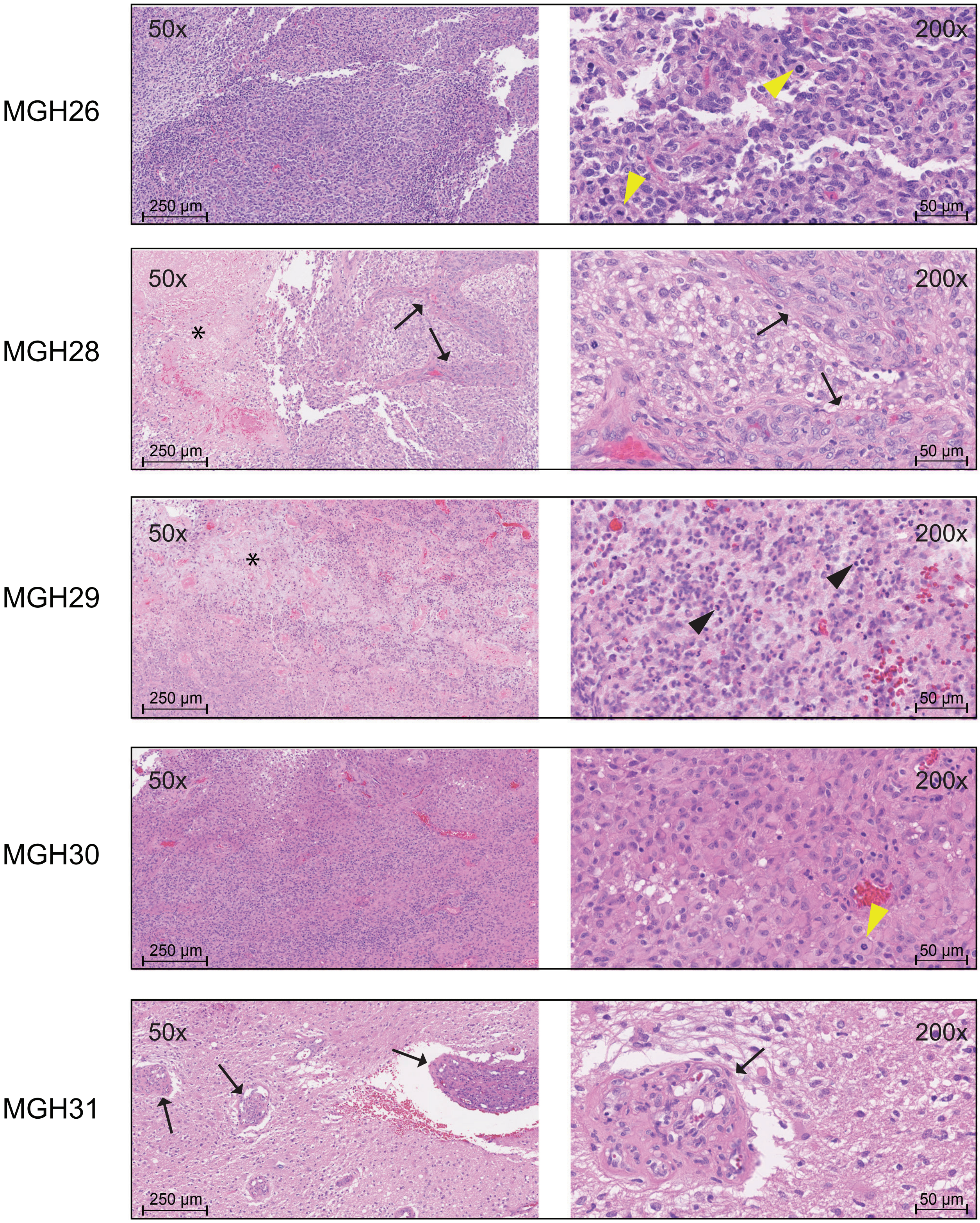
WT=wild type
NT=not tested
++=amplified
M=promoter methylated
NM=promoter not methylated
Number of cells=number of cells included in analysis (see **Supplemental Table 2**)

Supplemental Table 1: Patient demographics and clinical information.

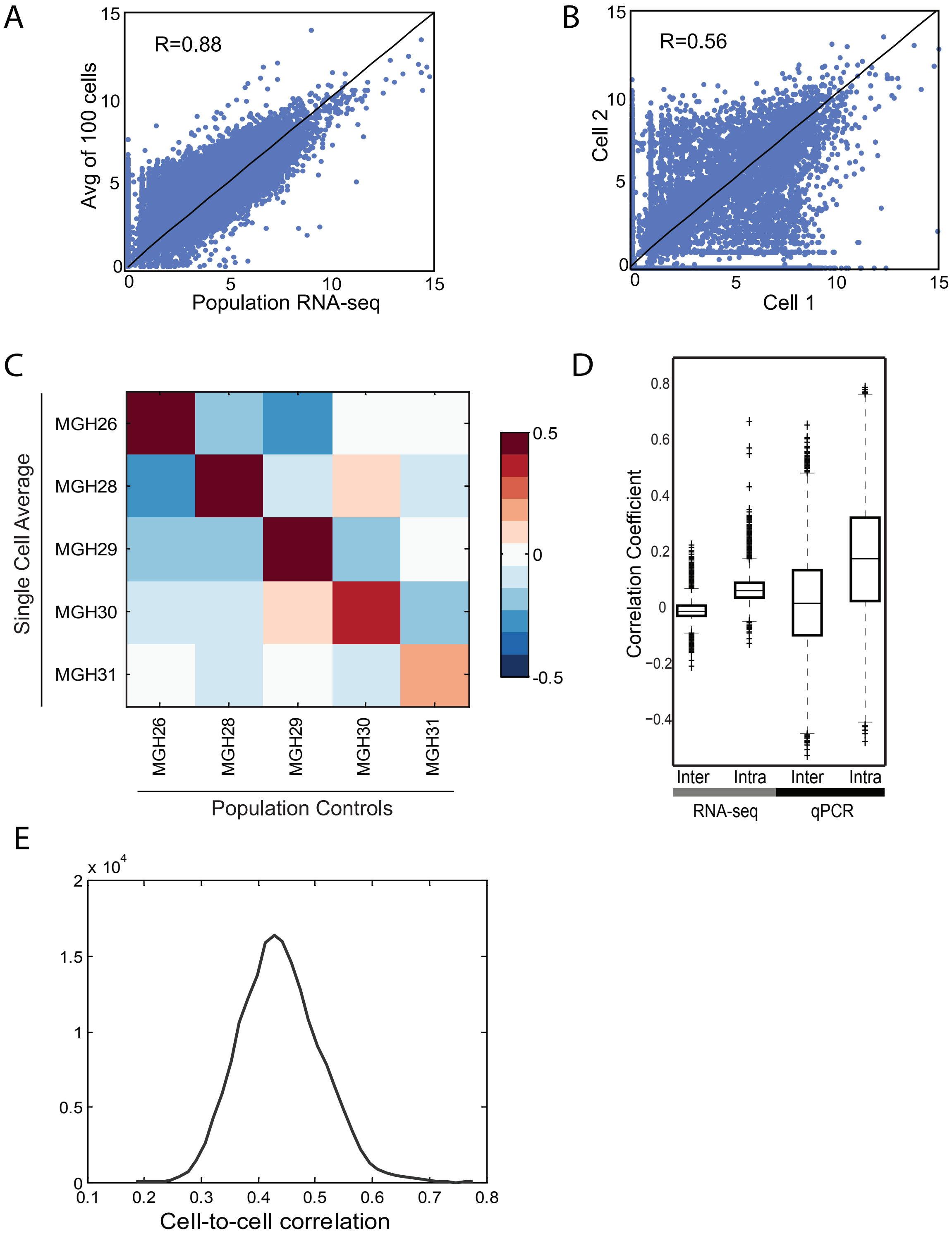
Sample number	Assayed cells	Included in analysis	Avg % aligned to transcriptome in included cells	# proliferating cells (%)
MGH26	192	118	27.16	10 (8.5%)
MGH28	192	94	30.28	5 (5.3%)
MGH29	96	75	45.83	2 (2.7%)
MGH30	96	73	31.45	16 (21.9%)
MGH31	96	70	16.17	1 (1.4%)

Supplemental Table 2: Statistics on sequenced single cells and percentage of cells showing high expression of proliferation module.

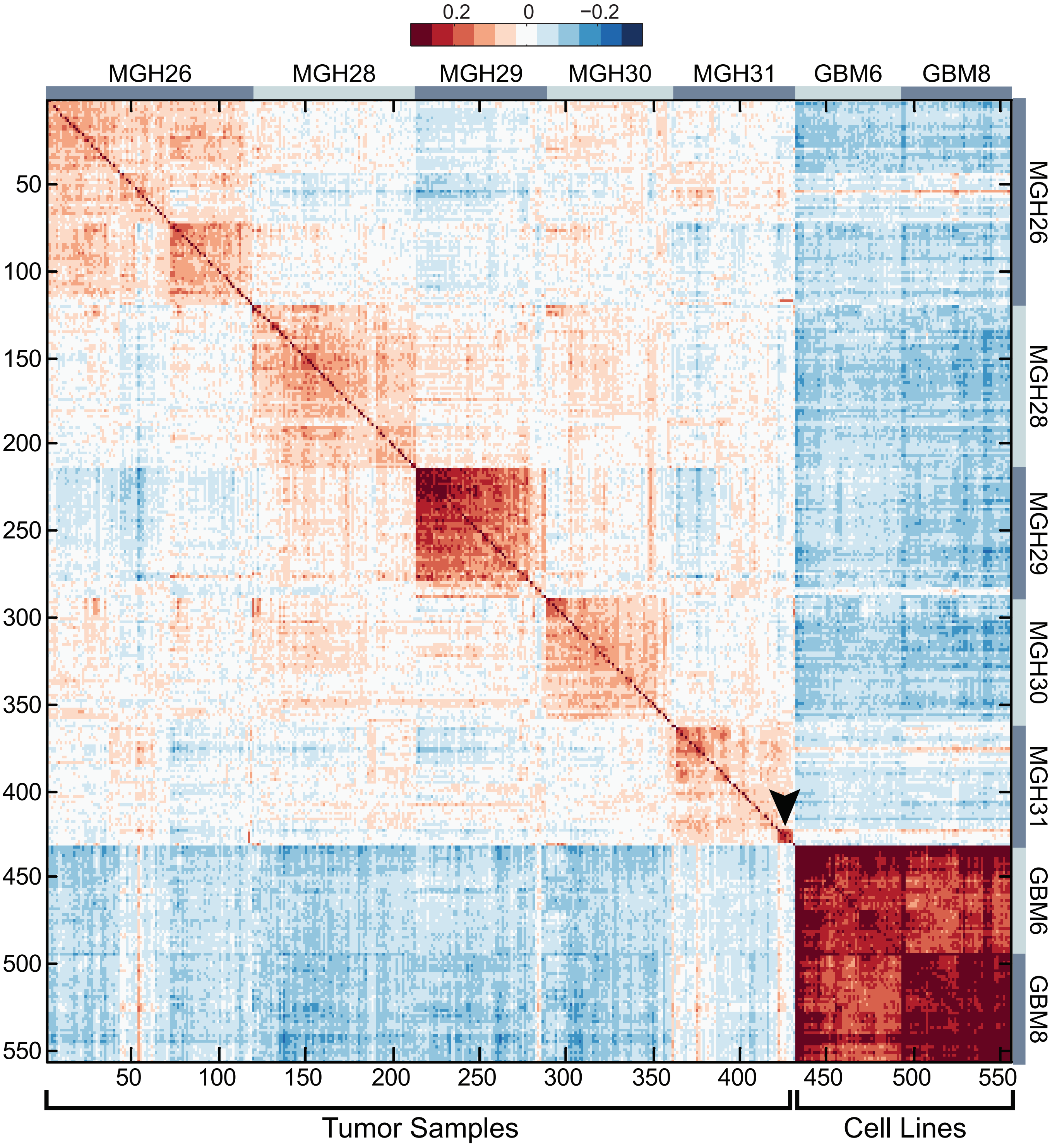
Supplemental Figures



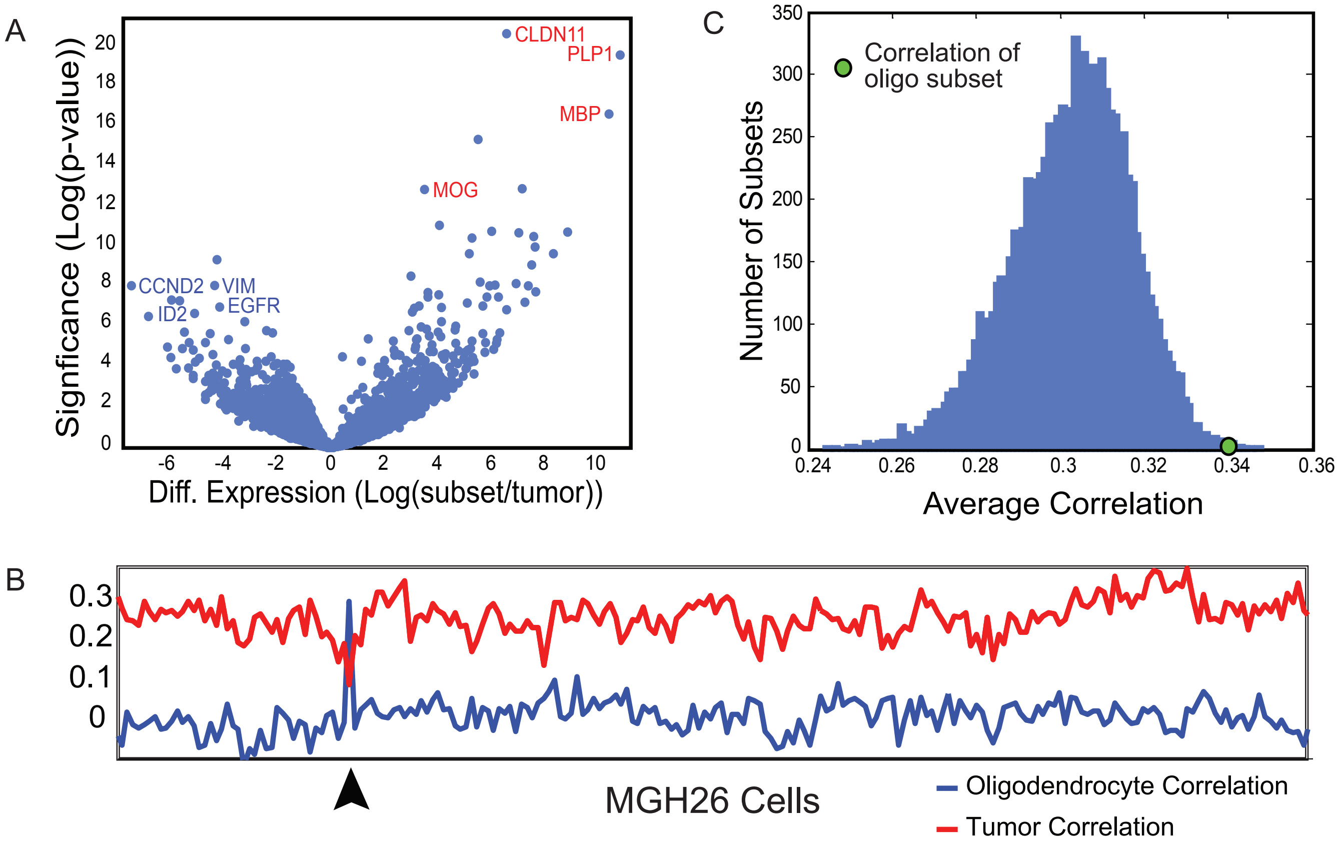
Supplemental Figure 1: Histology of primary glioblastoma cases used in this study. All tumors were confirmed to be primary glioblastoma, WHO grade IV/IV. Representative examples of morphology for each tumor at 50x (left) and 200x (right) magnification are shown. MGH26 and 30 show characteristic cytonuclear morphology, with atypical features and frequent mitotic figures (yellow arrowhead). MGH28 and 31 show florid vascular proliferation (arrows) and large areas of necrosis (*). In MGH29, abundant necrosis (*) and neutrophil infiltration (black arrowhead) is observed.



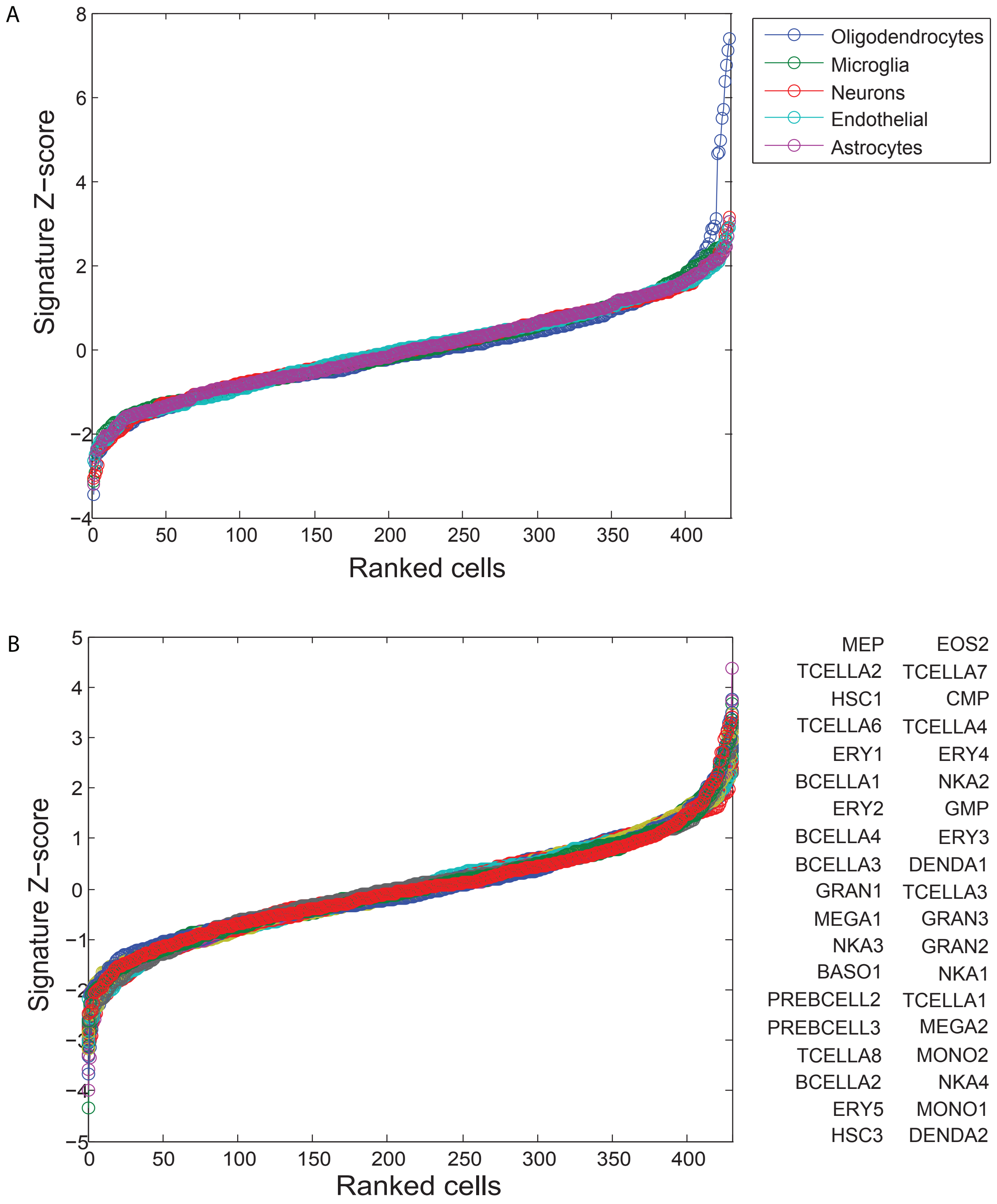
Supplemental Figure 2: Correlation between single cell profiles and (bulk) population profiles. (A) Scatter plot of gene expression values ($\log_2$(TPM+1)) for the population (bulk) RNA-Seq (*x*- axis) and the average $\log_2$(TPM+1) from 100 single cells for MGH26 (*y*-axis), showing high correlation (R=0.88). (B) Same plot for two randomly selected cells in MGH26 demonstrating lower correlation (R=0.56) suggestive of variable gene expression. (C) Centered correlation (color bar) between each single cell average (rows) and each population (bulk) profile (column), showing high correlation between single cell averages and their matching population (diagonal entries). (D) Box plots of the distribution of correlation coefficients between single cell expression profiles for cells from different tumors (inter) and the same tumor (intra) in single cell RNA-seq (left) and single cell qPCR (right) datasets. Intratumor pairs have a significantly higher ($p<10^{-5}$) correlation than intertumor pairs. (E) Distribution of correlation coefficients for all single cell pairs showing high degree of variability (R~0.2-0.7).



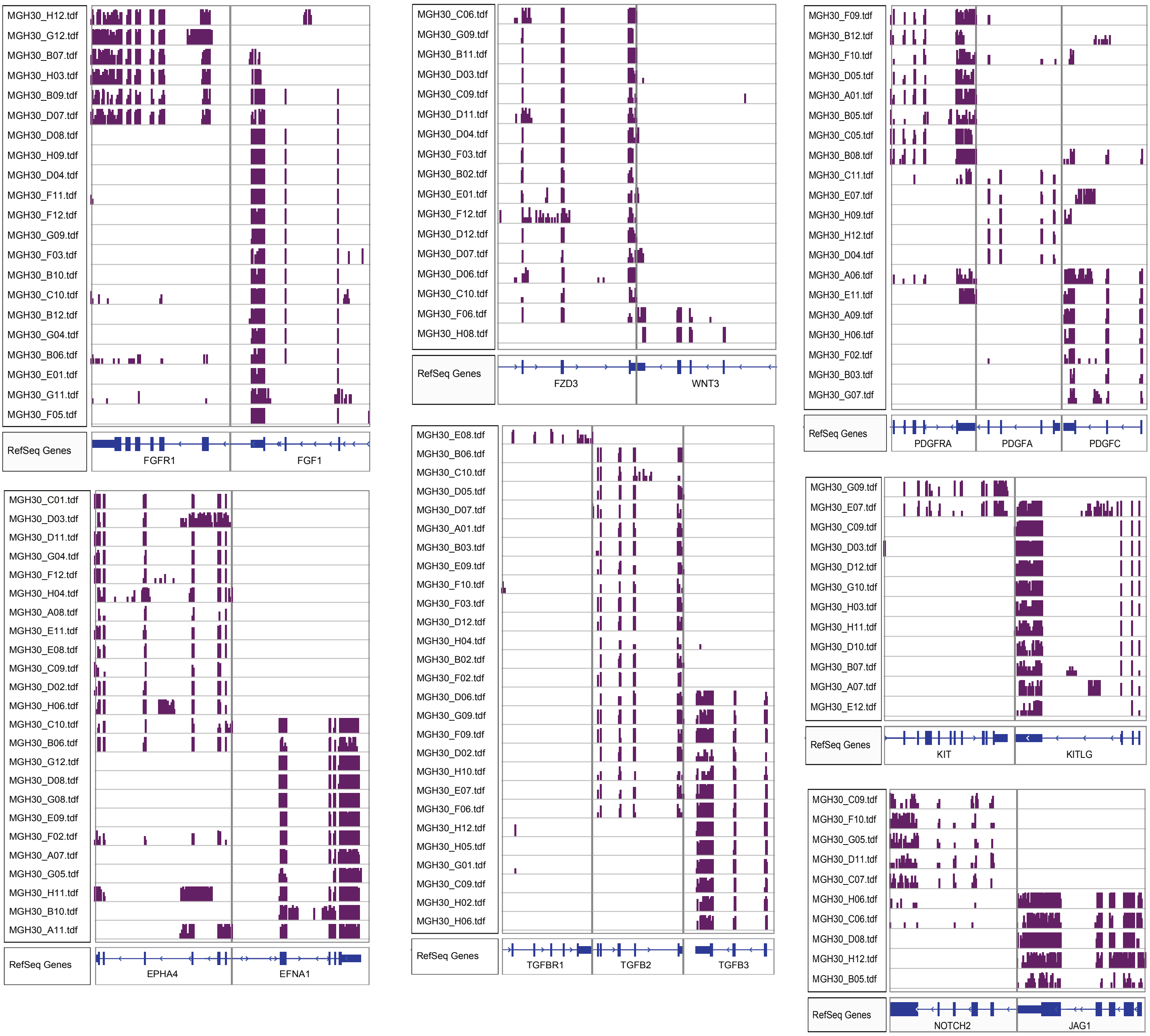
Supplemental Figure 3: Similarity and variation in single cell profiles. Correlation matrix for all 430 cells in the dataset (rows), ordered by tumor (MGH 26, 28, 29, 30, 31), and single cell profiles for the gliomasphere cell lines (GBM6, GBM8). While cells from each tumor demonstrate higher overall intratumoral coherence, there are several cells that have positive correlations with cells from other tumors. The two gliomasphere models show very high correlation with each other, but restricted similarity to only a subset of single cells from the primary tumors. Arrowhead highlights a group of cells in MGH31 with noticeably higher correlation to each other than other cells in that tumor, which likely correspond to oligodendrocytes (see Fig. 1,S4).



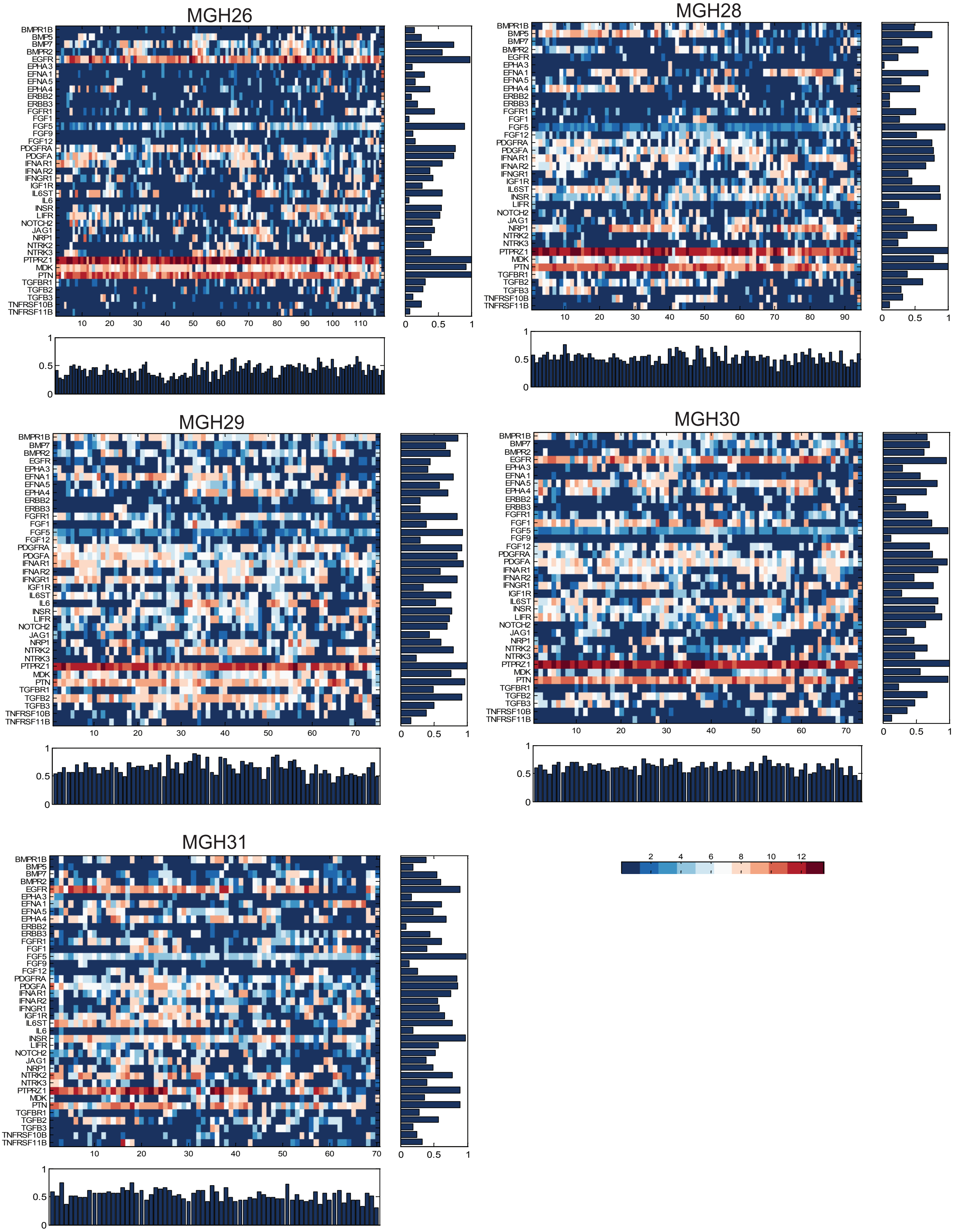
Supplemental Figure 4: Identification of oligodendrocytes. (A) Volcano plot shows for each gene the significance (*y*-axis) and ratio (*x*-axis) of its differential expression between the highly correlated subset of cells in **Fig. S3** and the remainder of cells in MGH31 (significance: -$\log_{10}$(P-value), two sample t-test). Genes upregulated in the subset include those associated with oligodendrocytes such as *MBP, MOG, CLDN11,* and *PLP1* (red), while those downregulated include genes characteristic of glioblastoma, such as *VIM, ID2, CCND2,* and *EGFR* (blue). (B) Each cell in MGH26 (*x*-axis) is shown with its score for correlation (*y*-axis) to the oligodendrocyte subset (red) and tumor subsets (blue). One additional oligodendrocyte is detected in MGH26 (arrowhead). This additional cell also clustered with normal samples in the inferred CNV analysis (Fig. 1). (C) Distribution of average cell-to-cell correlation coefficients for 1,000 randomly selected groups of eight cells from MGH31. The average correlation between the eight oligodendrocytes (green dot) is higher than that among tumor counterparts (p=0.006).



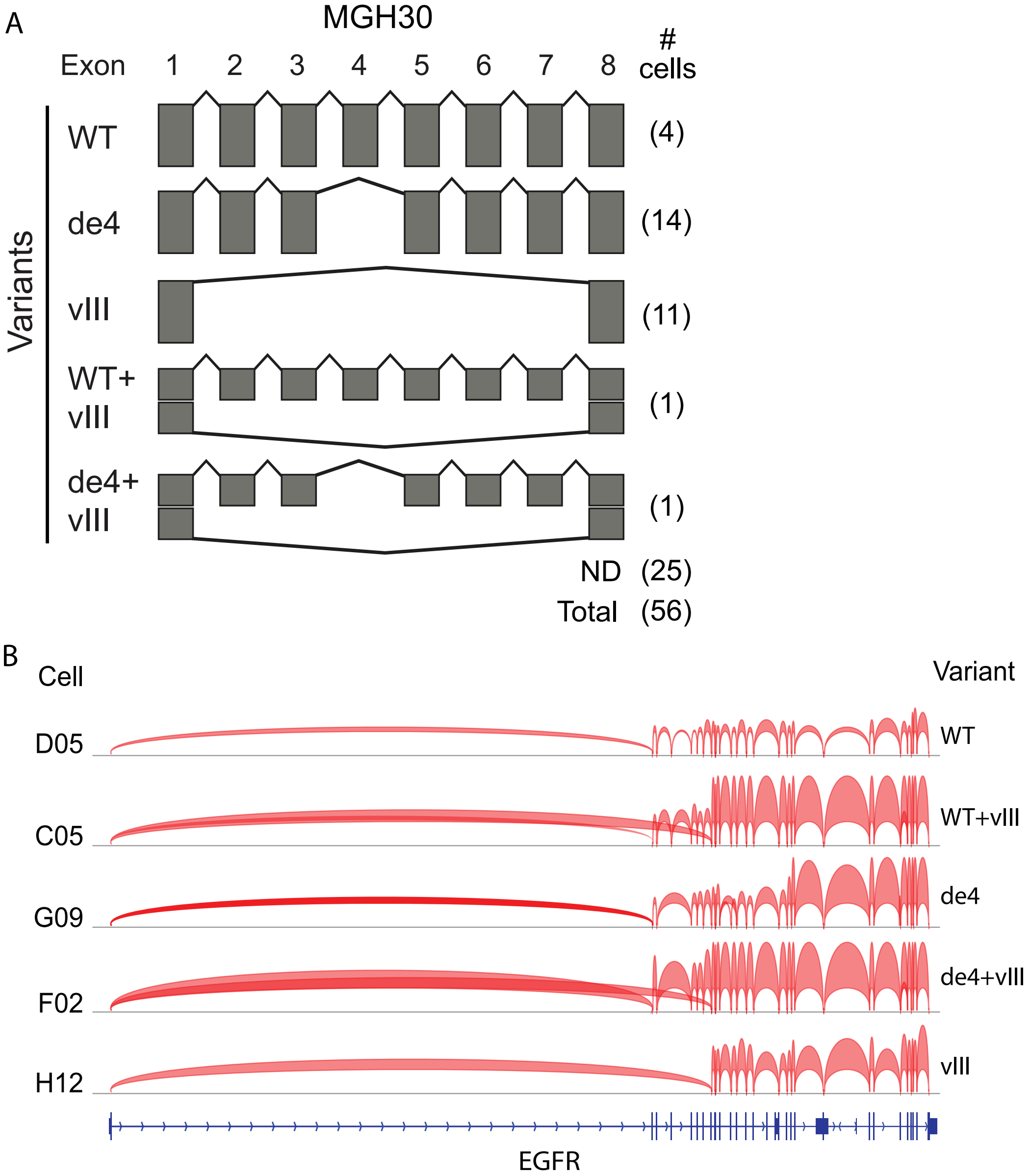
Supplemental Figure 5: (A) For each cell (*x*-axis), plot shows Z-scores (*y*- axis) reflecting similarity to each of five normal cell type signatures (color legends). Cells are ranked by their scores for each signature independently. Aside from a high correlation of several cells with normal oligodendrocyte transcriptional programs (blue line), we did not detect significant representation of other neural cell-related programs (including microglia, neurons, endothelial cells, and astrocytes). **(B)** Same analysis as (A) for 38 immune signatures did not identify a significant representation of immune cell-related programs.



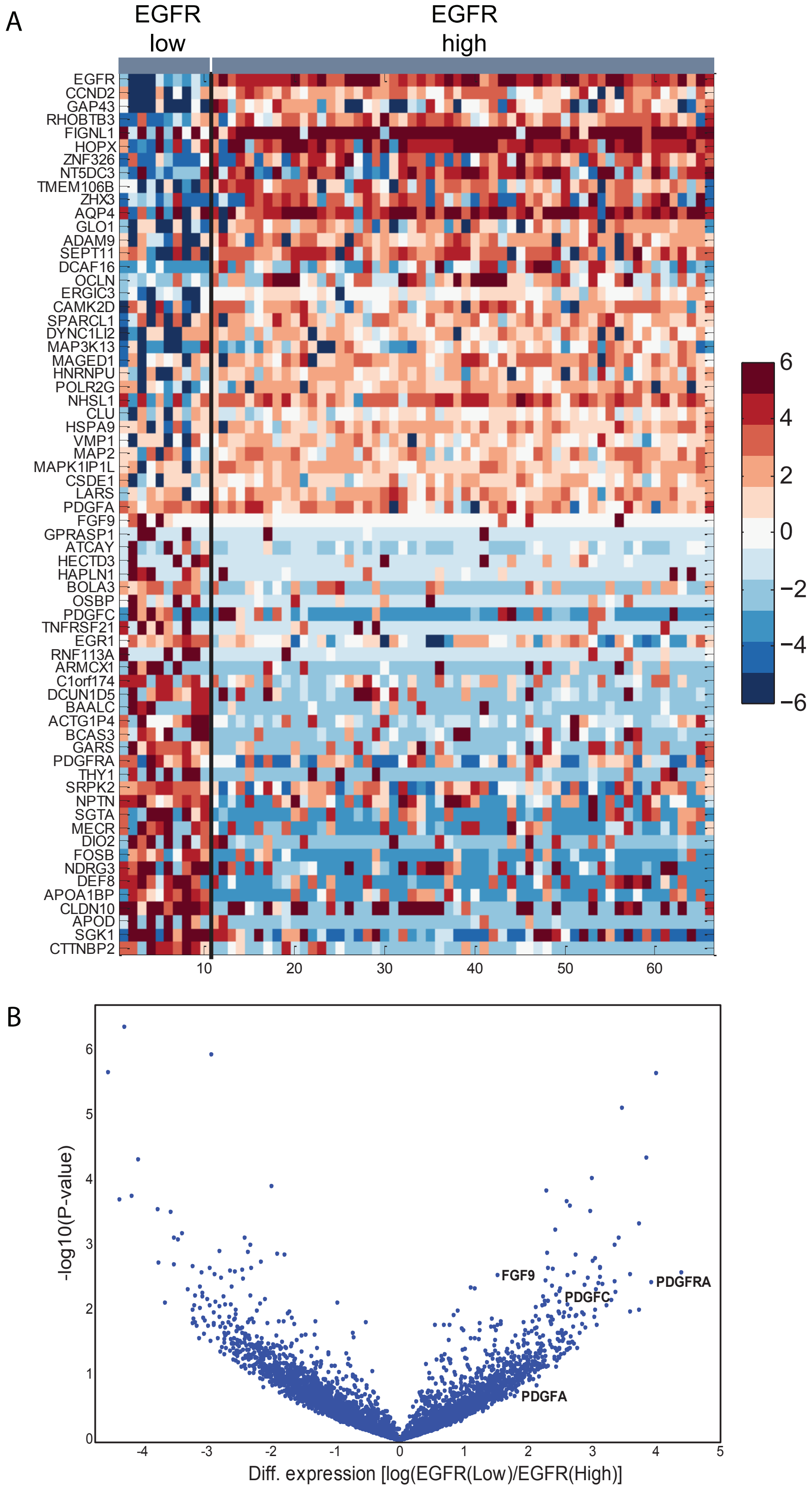
Supplemental Figure 6: Variation in expression of receptors and ligands in MGH30. Shown are RNA-seq read densities (vertical scale 10) for selected receptors and their ligands in MGH30, demonstrating variability in expression across individual cells (rows).



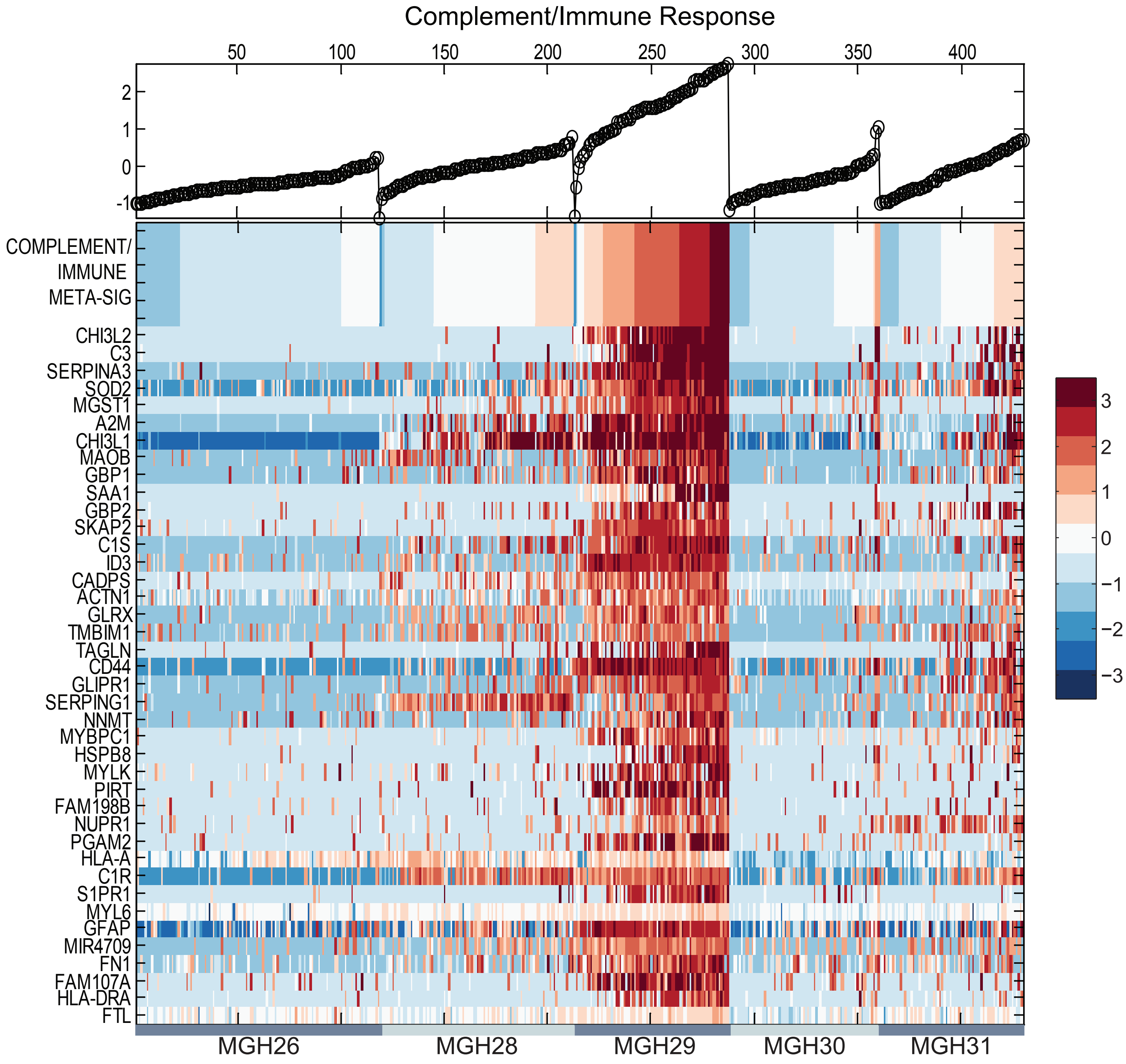
Supplemental Figure 7: Variation in expression of receptors and ligands in each tumor. Heatmaps for each tumor show expression values for receptors and ligands (rows, alphabetically ordered with receptors and ligands paired) in each cell (columns). All receptors and ligands represented among the 5,948 genes in the filtered dataset are included. Their expression varies considerably in each tumor. Bar plots represent the sum of the signal in rows (genes) and columns (cells).



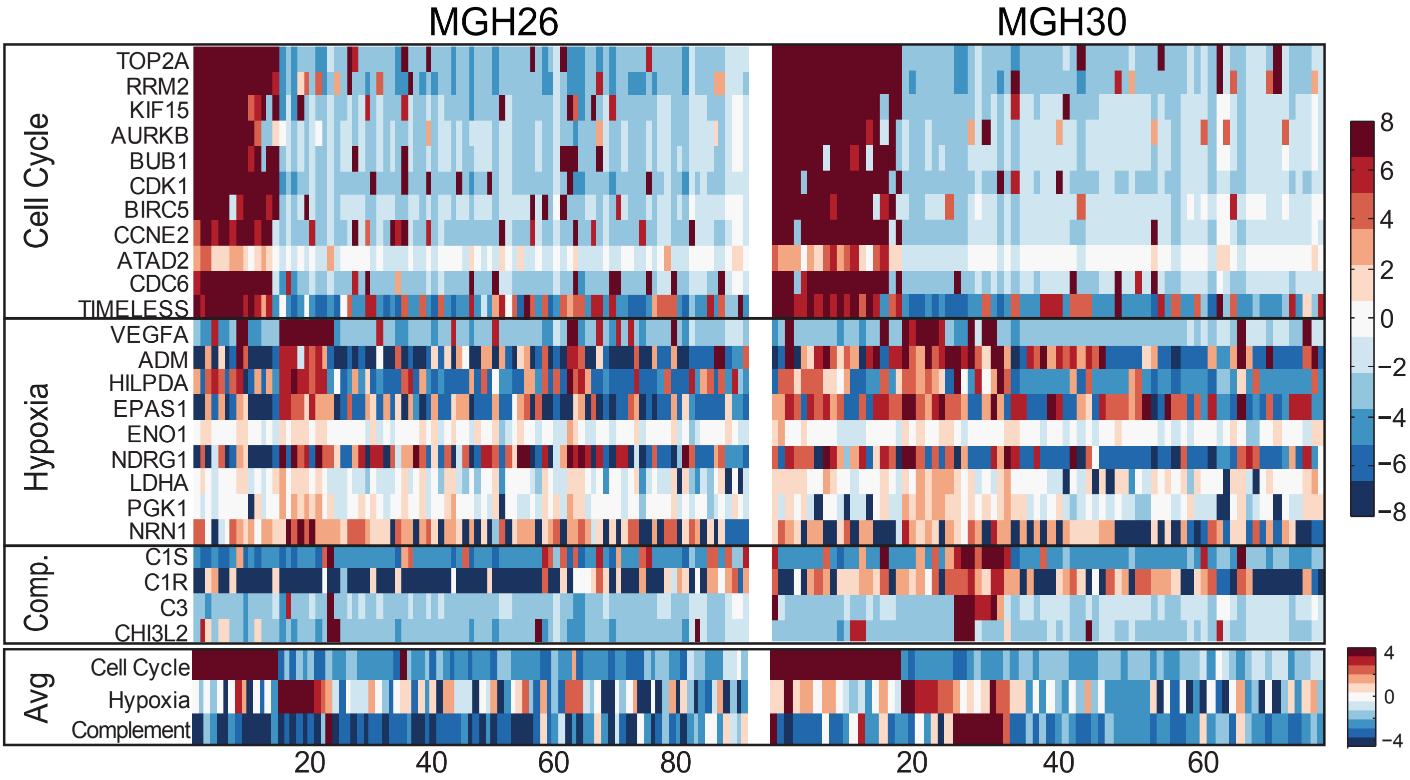
Supplemental Figure 8: Variation in EGFR splicing patterns in single cells. (A) Schematic depicts *EGFR* splice variants detected by mapping of junction-spanning 'spliced' reads. Three different variants were expressed in relatively mutually exclusive patterns across individual cells from MGH30, an *EGFR* amplified tumor. Coexpression of *EGFR* variants was observed in only a small fraction of cells. ND denotes cells that had only 3' coverage and were not assignable. (B) Shown are splice junction-spanning reads mapped in different individual cells (rows) in MGH30 representing the five combinations of EGFR variants (WT, de4, vIII, WT+vIII, and de4+vIII).



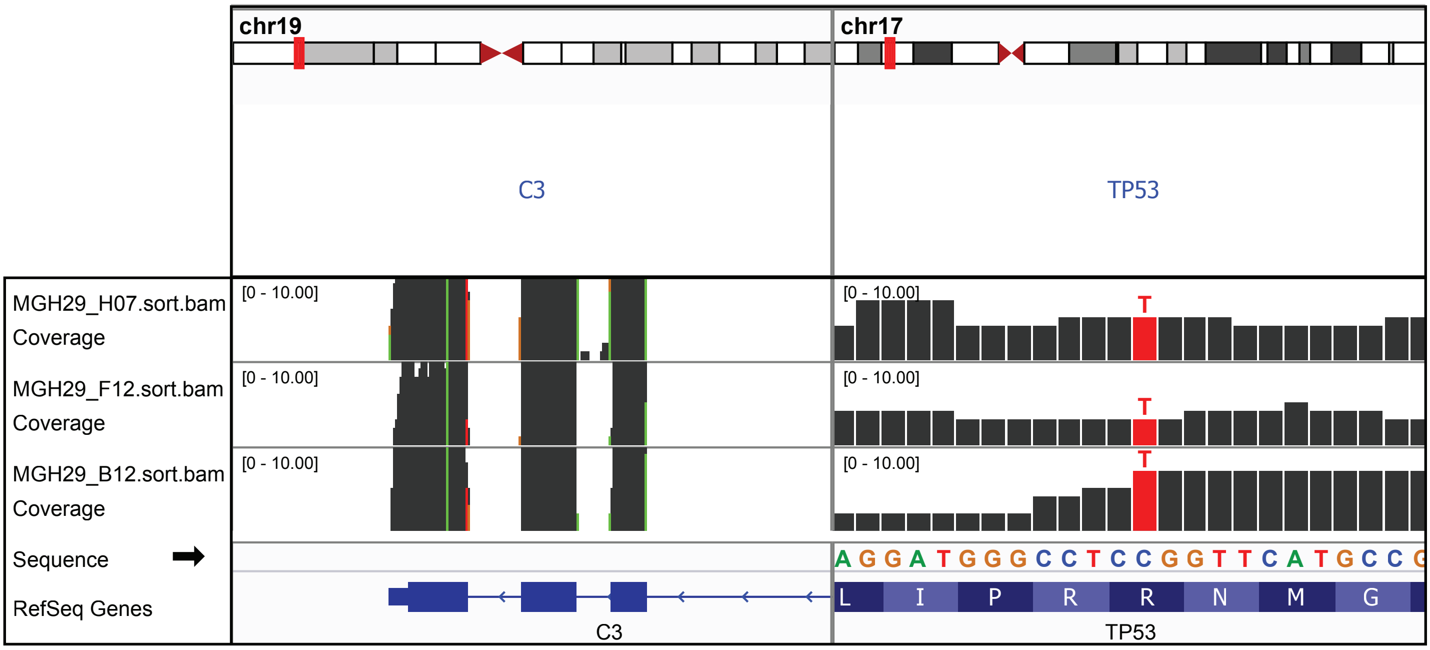
Supplemental Figure 9: Expression variation across cells associated with EGFR expression status. (A) Heatmap of genes (rows) that are differentially expressed between EGFR high and low cells (columns) in MGH30. (B) Volcano plot showing for each gene its differential expression in EGFR low vs. high single cells (*x*- axis) vs. the significance associated with this degree of differential expression (-$\log_{10}$(P-value), two-sample t-test). Key signaling genes expressed in EGFR low populations, including *PDGFRA, PDGFA, PDGFC,* and *FGF9*, are highlighted.



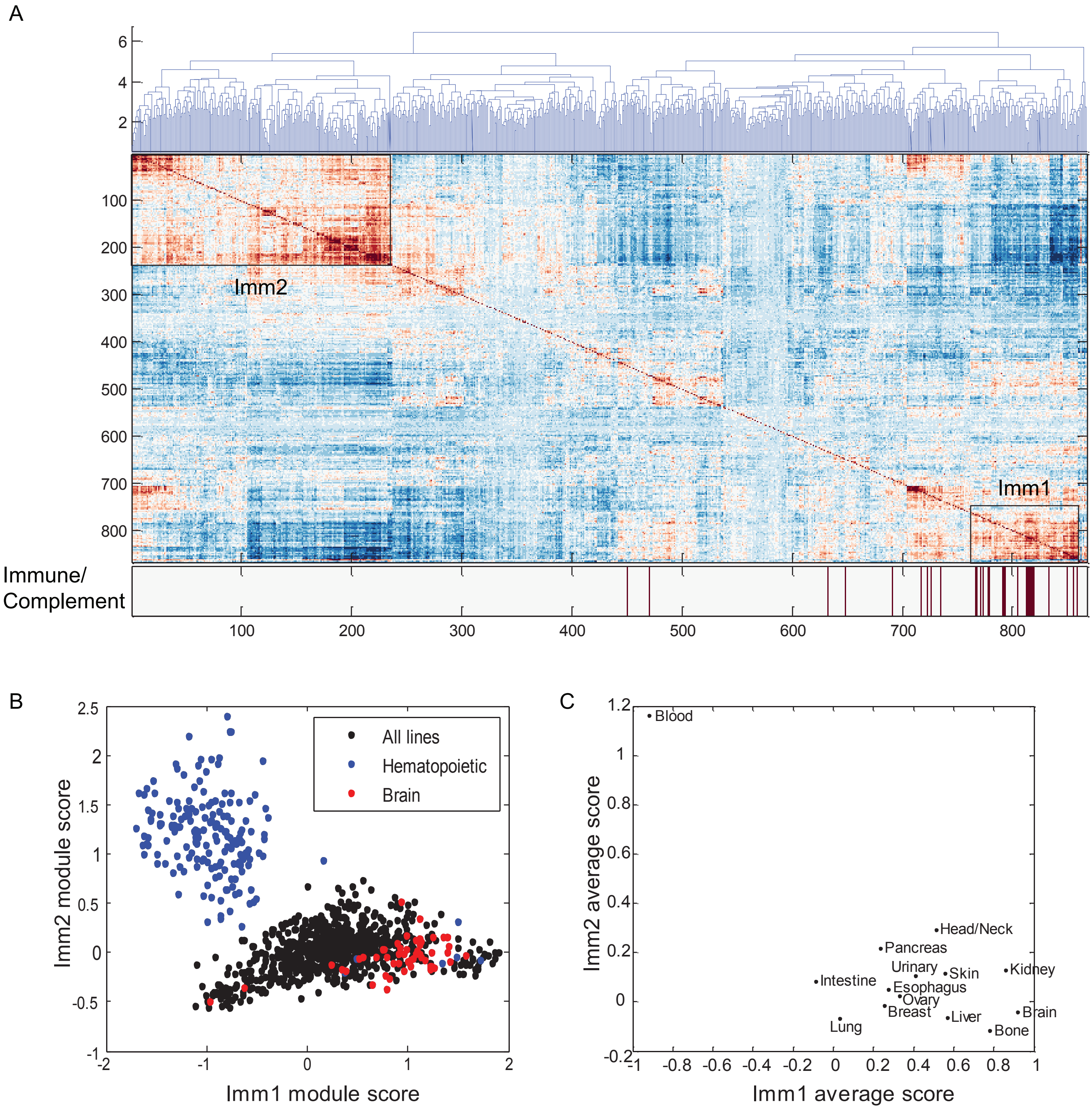
Supplemental Figure 10: Complement meta-signature. Heatmap shows the average expression of the meta-signature (top row) and individual genes from the signature (rows) in single glioblastoma cells (columns), grouped by tumor and ordered within each tumor from lowest to highest meta-signature score (*y*-axis, top panel). The meta-signature is expressed in most cells of MGH29 but is rare or absent from the other tumors.



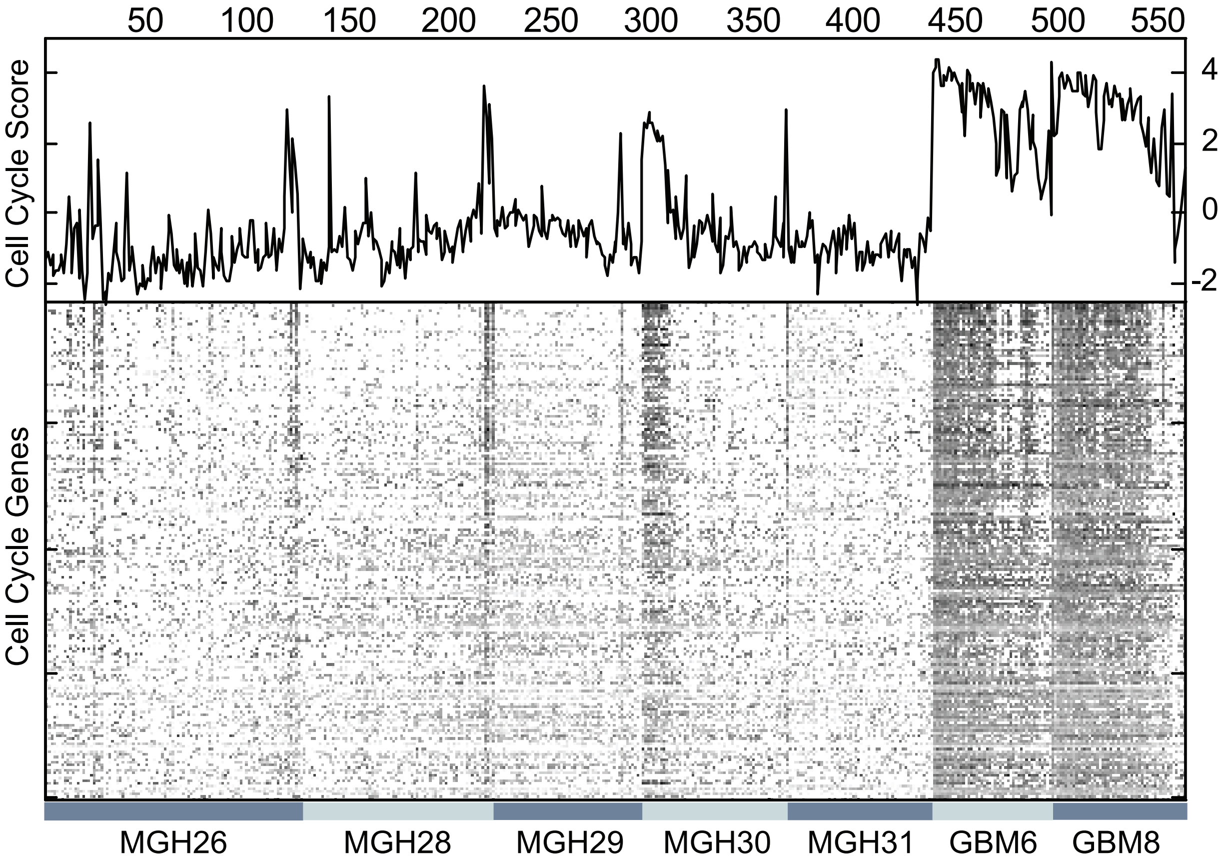
Supplemental Figure 11: Validation of meta-signatures by single cell qPCR. Top three panels show expression levels from single cell qPCR for 4-10 genes (rows) per meta-signature (labeled on far left) in 91 (MGH26, left panel) and 76 (MGH30, right panel) additional cells. Bottom panel shows average expression of genes in each meta-signature.



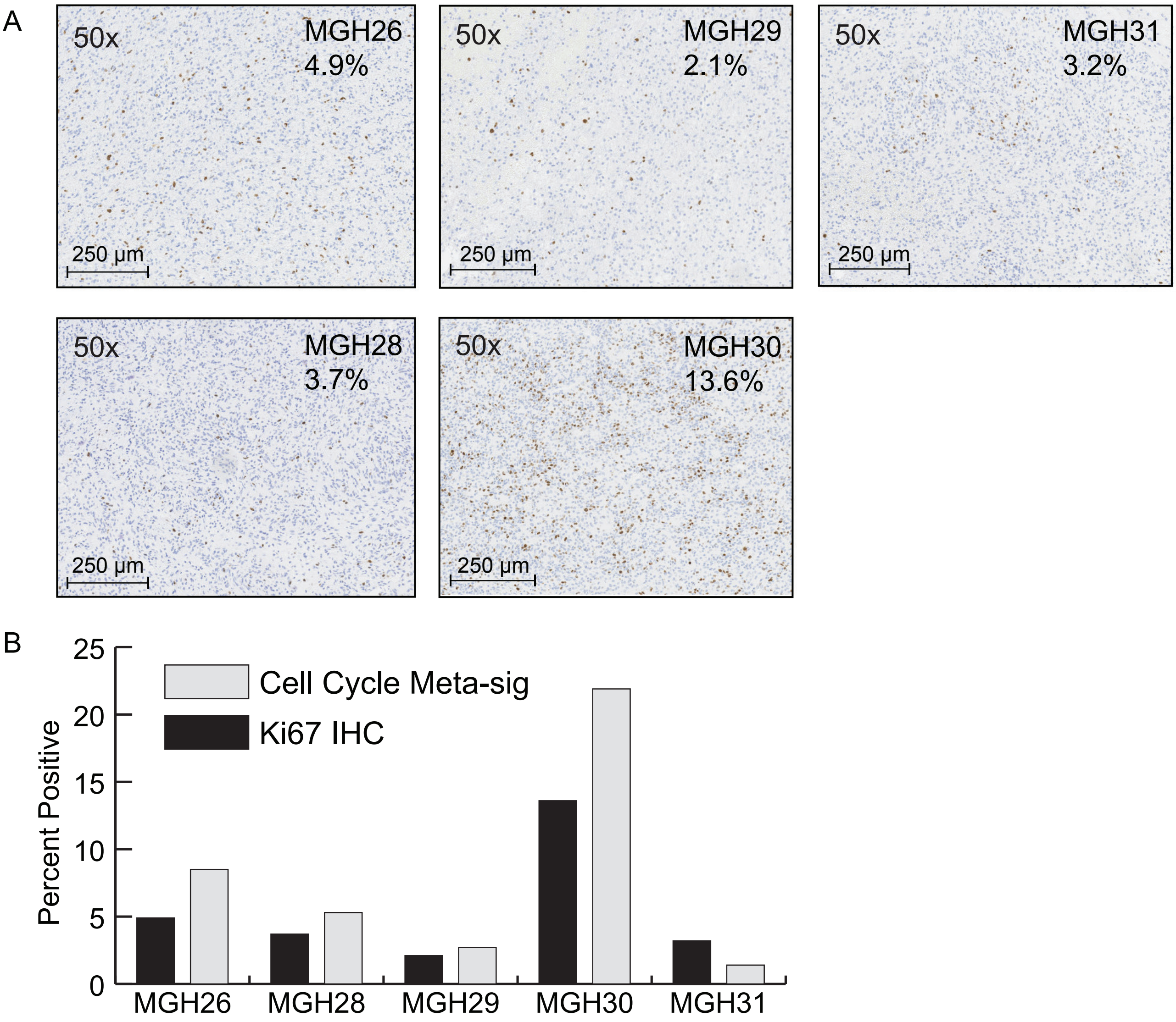
Supplemental Figure 12: Expression of complement C3 in tumor cells. Three cells were identified in MGH29 that definitively contain the R248L (C→T) mutation (right panel, highlighted in red). This mutation is highly oncogenic and occurs in the setting of malignancy. The same three cells also express complement C3 (left panel) demonstrating that glioblastoma cells are able to express complement pathway genes.



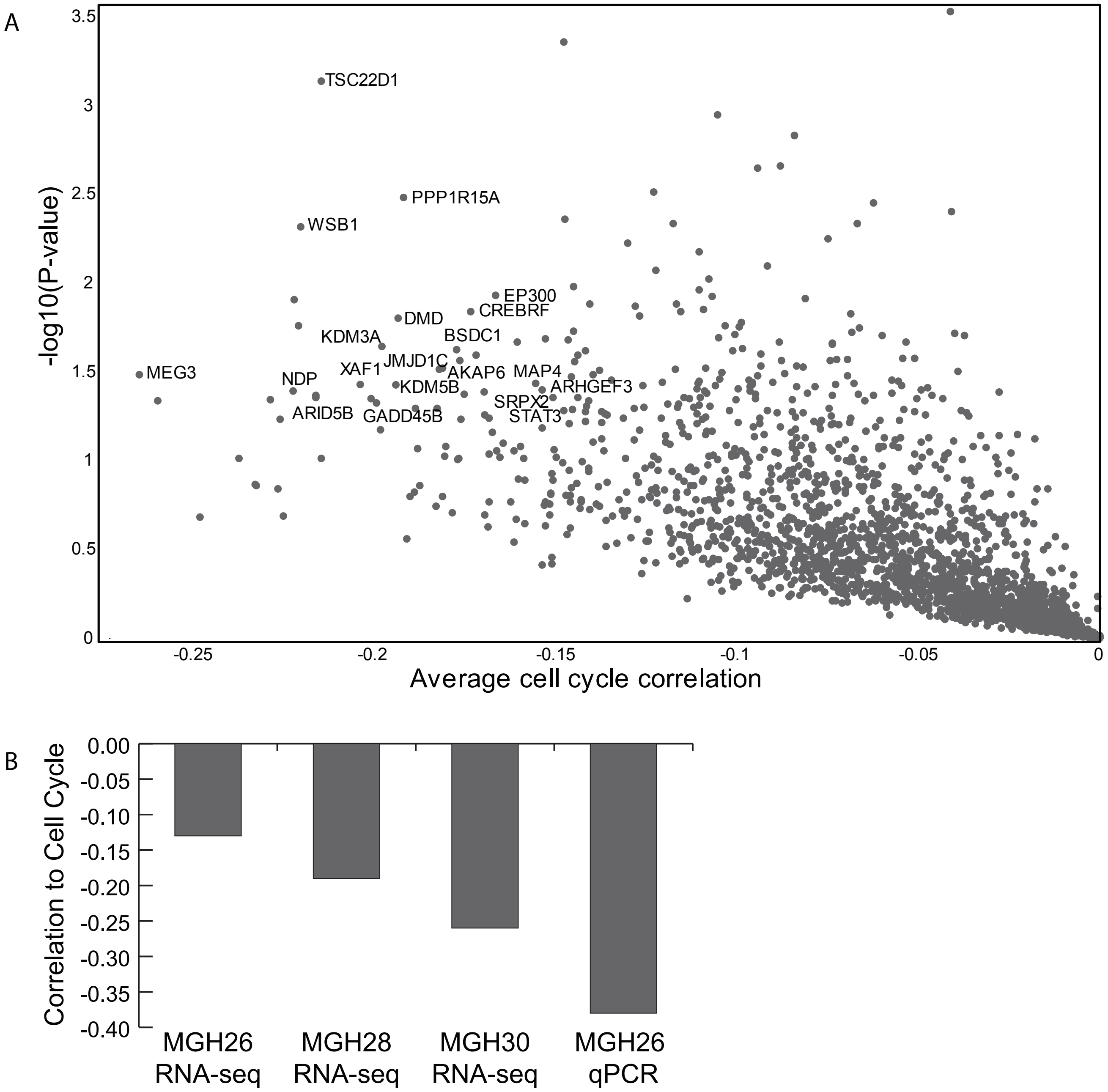
Supplemental Figure 13: Co-expression of the complement/immune module in cancer cell lines. (A) Unsupervised clustering of all annotated immune genes in ~1,000 cell lines from the Cancer Cell Line Encyclopedia identifies two clusters of co-expressed genes. Imm1 is highly enriched for genes from the complement/immune meta-signature (red bar). **(B)** Scatter plots depict Imm1 and Imm2 scores for these cancer cell lines. The Imm1 gene set, which contains complement/immune module genes, is co-expressed in cancer cell lines from many solid tumors (black), especially those derived from brain tumors (red). The Imm2 gene set, which is distinct from the module derived in this study, is coexpressed in hematopoetic cancer lines (blue). **(C)** Separation of the module scores based on tumor type demonstrates that the Imm1/complement module is on average most strongly expressed in lines derived from brain tumors.



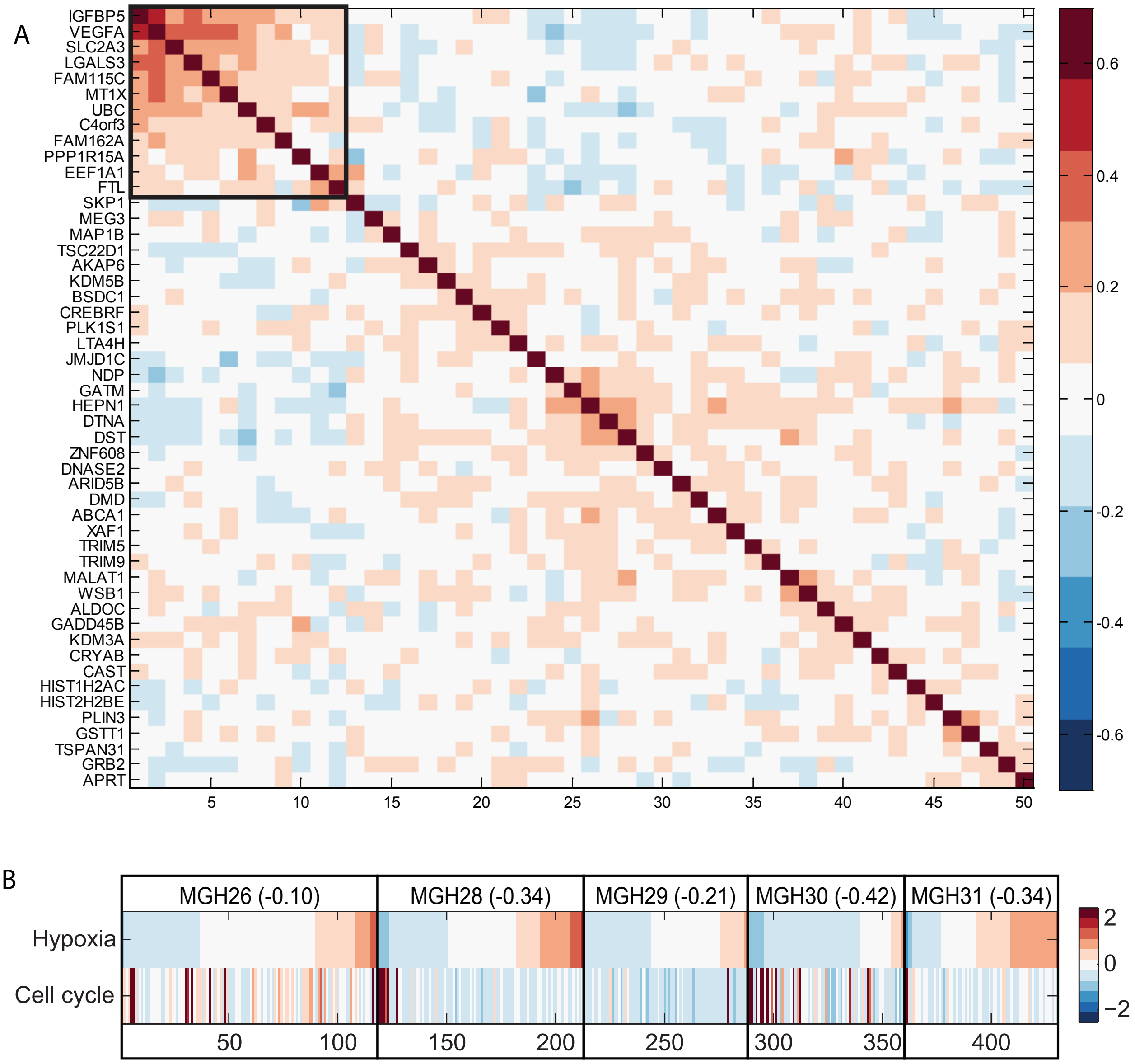
Supplemental Figure 14: Cell cycle meta-signature. Cell cycle meta-signature scores (*y*-axis, top plot) and the expression of each individual gene in the meta-signature (rows in grayscale heatmap) are shown for each individual cell (columns) in the five tumors and two gliomasphere cell lines (GBM6, GBM8). Almost all of the gliomasphere cells, but few tumor cells score highly for this meta-signature.



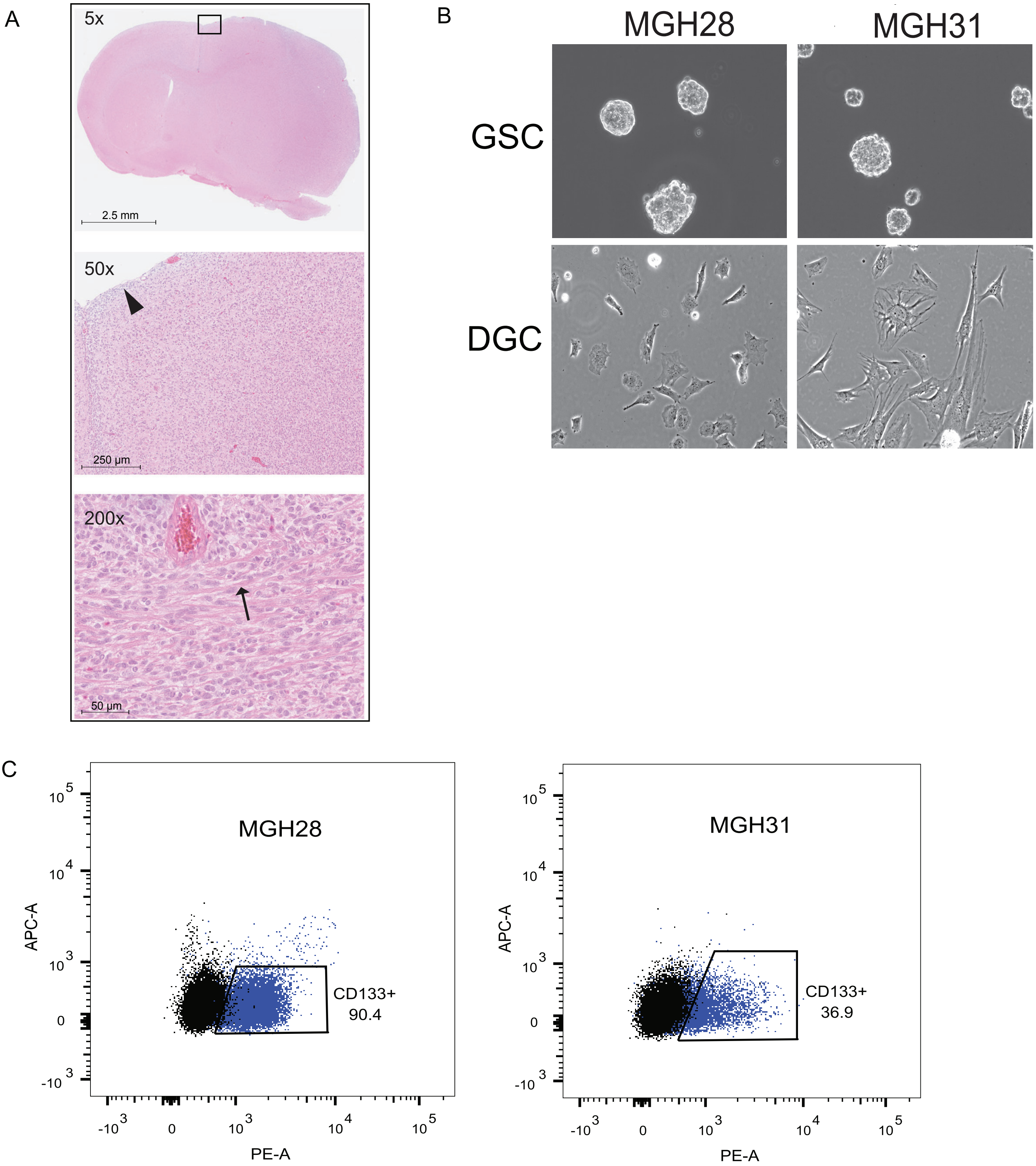
Supplemental Figure 15: Ki67 quantification for tumors. (A) Immunohistochemistry staining of Ki67 positive cells (brown) from representative fields of each tumor. Quantifications were performed by counting at least 1000 nuclei per tumor. **(B)** Plot compares percentage of cells predicted to be cycling, as determined by Ki67 IHC (black) or inferred from cell cycle meta-signature (gray). The two metrics show qualitative concordance.



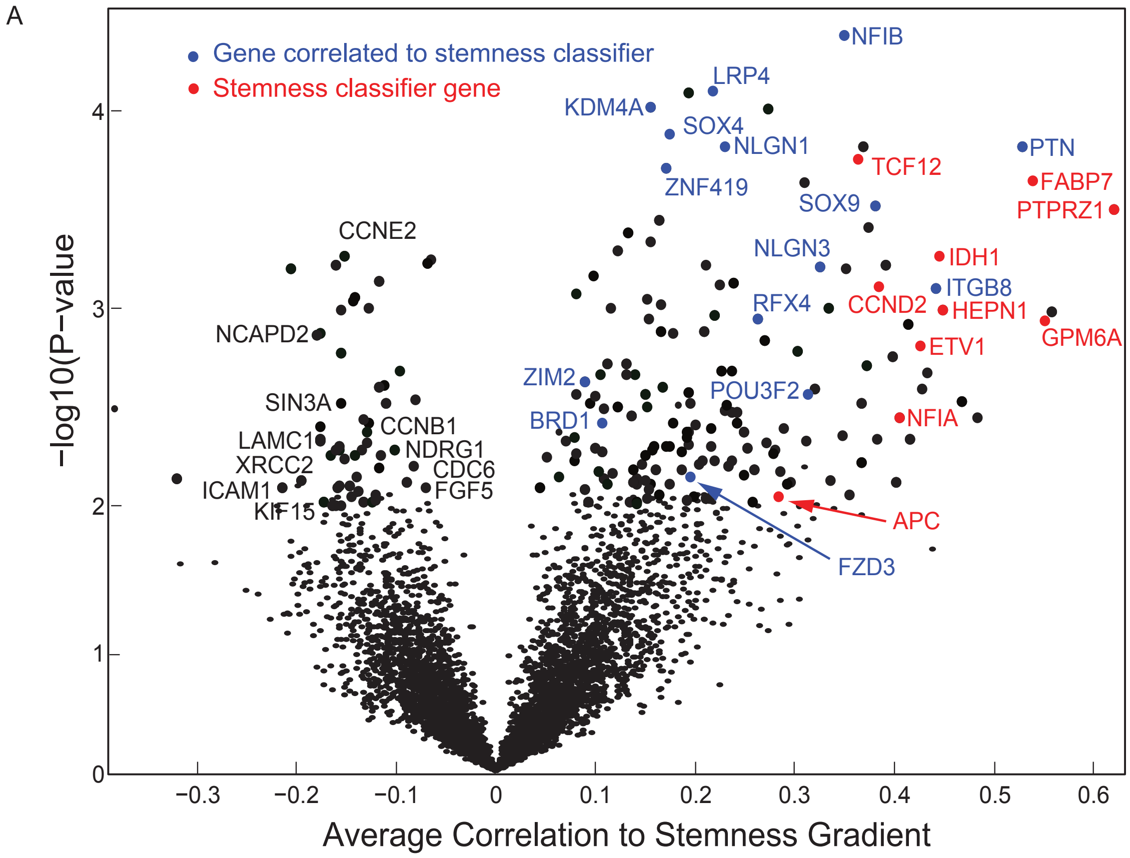
Supplemental Figure 16: Identification of genes anticorrelated to the cell cycle meta-signature. (A) Volcano plot of genes that are preferentially upregulated in non-cycling cells. Pearson correlations were calculated between the expression of each gene and the cell cycle meta-signature within each of the three tumors in which there were more than two cells that scored highly for the cell cycle meta-signature (MGH26, MGH28 and MGH30). The average of these three correlations (reflecting the three tumors) is shown as a measure of correlation of each gene with the cell cycle (*x*-axis), and the significance of the these correlations deviating from a mean of zero is shown on the y-axis as –log10(p), where p is the p-value from a one-sample t-test with these three correlations. Putative regulators of the non-cycling compartment are labeled, including *TSC22D1, GADD45B,* and *KDM5B.* (B) Plot depicts the (negative) correlation (*y*-axis) of *KDM5B* expression with the cell cycle meta-signature across cells in the indicated tumor. For RNA-seq, MGH26, 28, and 30 are shown, as they have significant expression of the cell cycle meta-signature. *KDM5B* expression was also tested by single cell qPCR in 91 cells MGH26, confirming the anti-correlation to cell cycle seen in RNA-seq data.



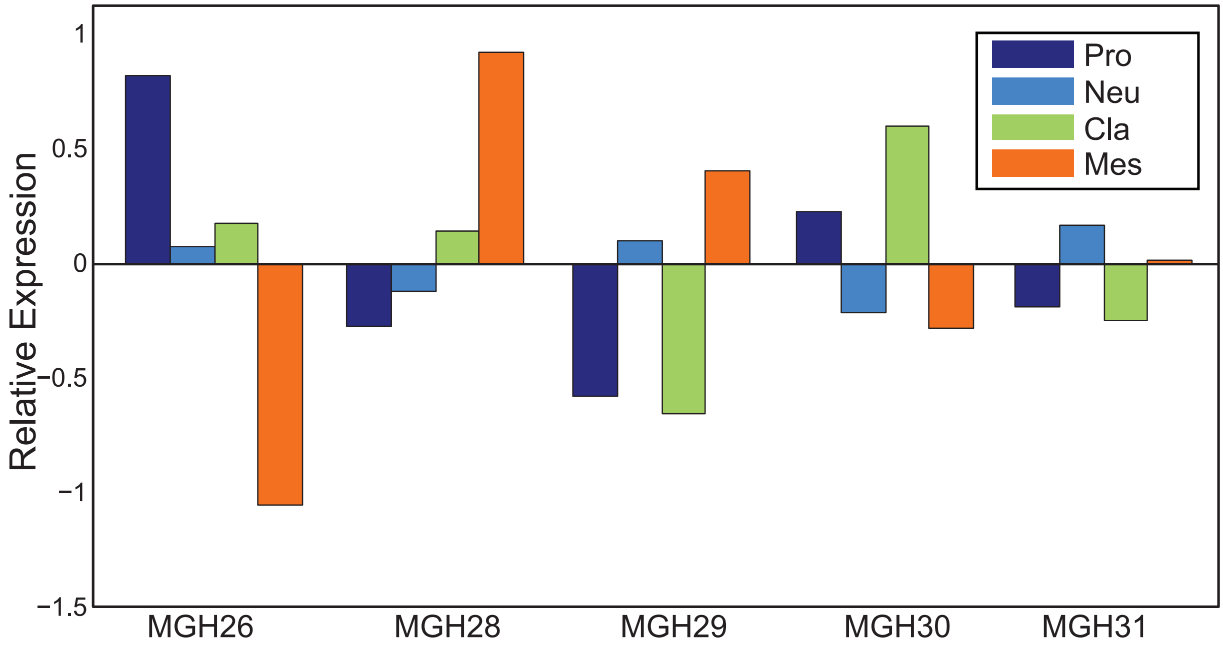
Supplemental Figure 17: Relationship between cell cycle and hypoxia meta-signatures. Clustering of genes anticorrelated to the cell cycle meta-signature (rows, columns: genes; color bar indicates Pearson correlation coefficient). A sub-cluster (black box) consists of genes from the hypoxia meta-signature. (B) Average relative expression values (red/blue) of the genes in the cell cycle and the hypoxia meta-signatures (rows) in each individual cell (columns; separated by tumors, and sorted in each by the strength of the hypoxia meta-signature). There is an overall anticorrelation (R value is represented in parenthesis next to tumor name) between these two meta-signatures across all tumors.



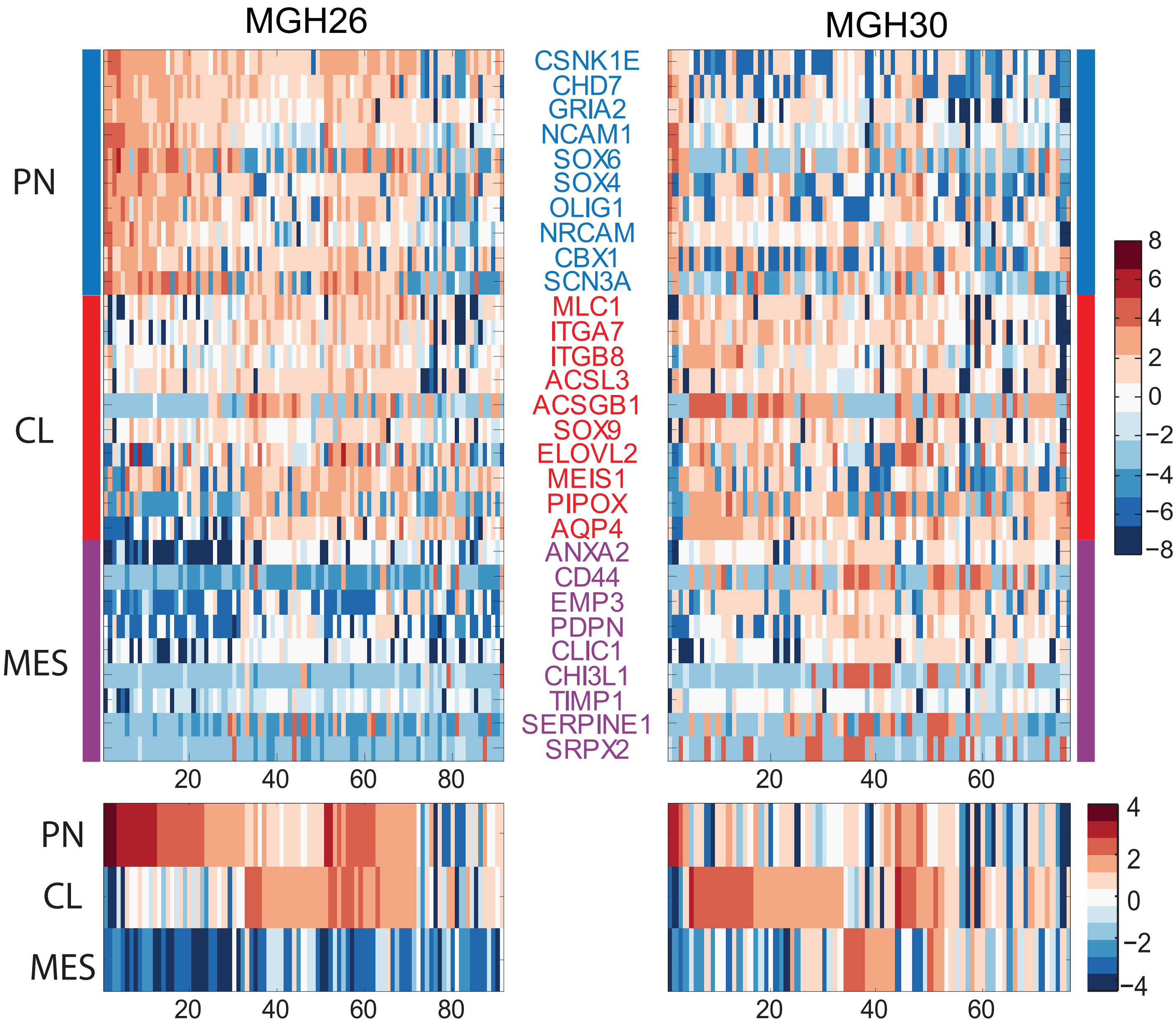
Supplemental Figure 18: *In vitro* tumor models. (A) Mouse xenografts of MGH26 GSCs demonstrate key features of diffuse infiltrative glioblastoma. Low magnification (5x) coronal section shows enlargement of the right hemisphere and compression of ipsilateral ventricle. Boxed area is magnified below at 50x, showing diffuse infiltration of brain parenchyma and subpial structures (arrowhead) by glioblastoma cells. High magnification (200x) shows cells with atypical features infiltrating along inter-hemispheric white matter tracks (corpus callosum). (B) Brightfield images of MGH28 and MGH31 gliomaspheres (GSC) and adherent (DGC) cultures. (C) GSCs express the stem cell surface marker CD133.



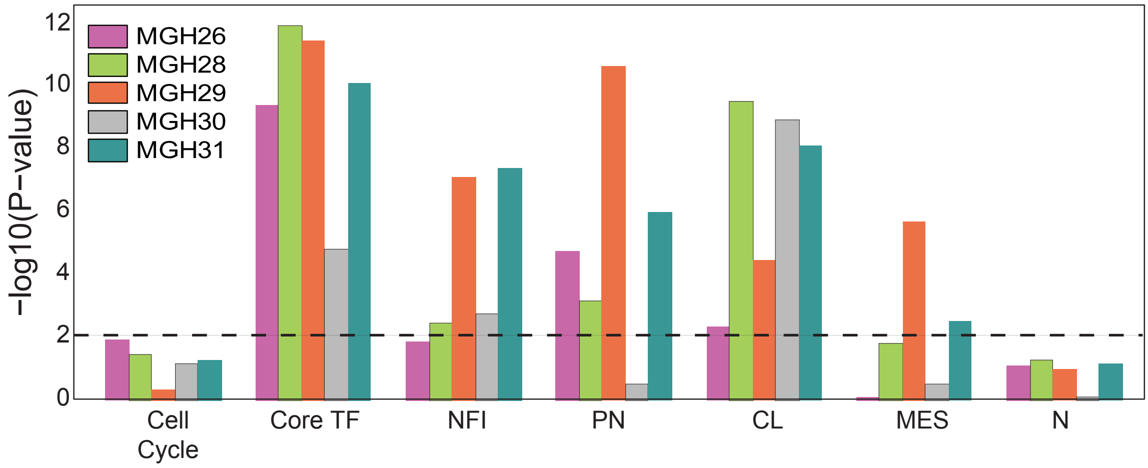
Supplemental Figure 19: Stemness associated genes *in vivo*. Volcano plot showing for each gene its average correlation to the stemness score among the five tumor (*x*-axis) and the –log10(P-value), reflecting the significance of the deviation of the average of these scores from zero, as defined by a one-sample t-test (*y*-axis). Genes differentially expressed in GSC high versus low cells are labeled. Red genes denote those that were in the original signature, and blue genes denote those that correlate with the signature *in vivo* but were not discovered *in vitro*, and suggest candidate regulators of the stem cell compartment, including *SOX4/9*, *IDH1*, *FZD3*, *PTN*, *NFIB, KDM4A*, and *RFX4*.



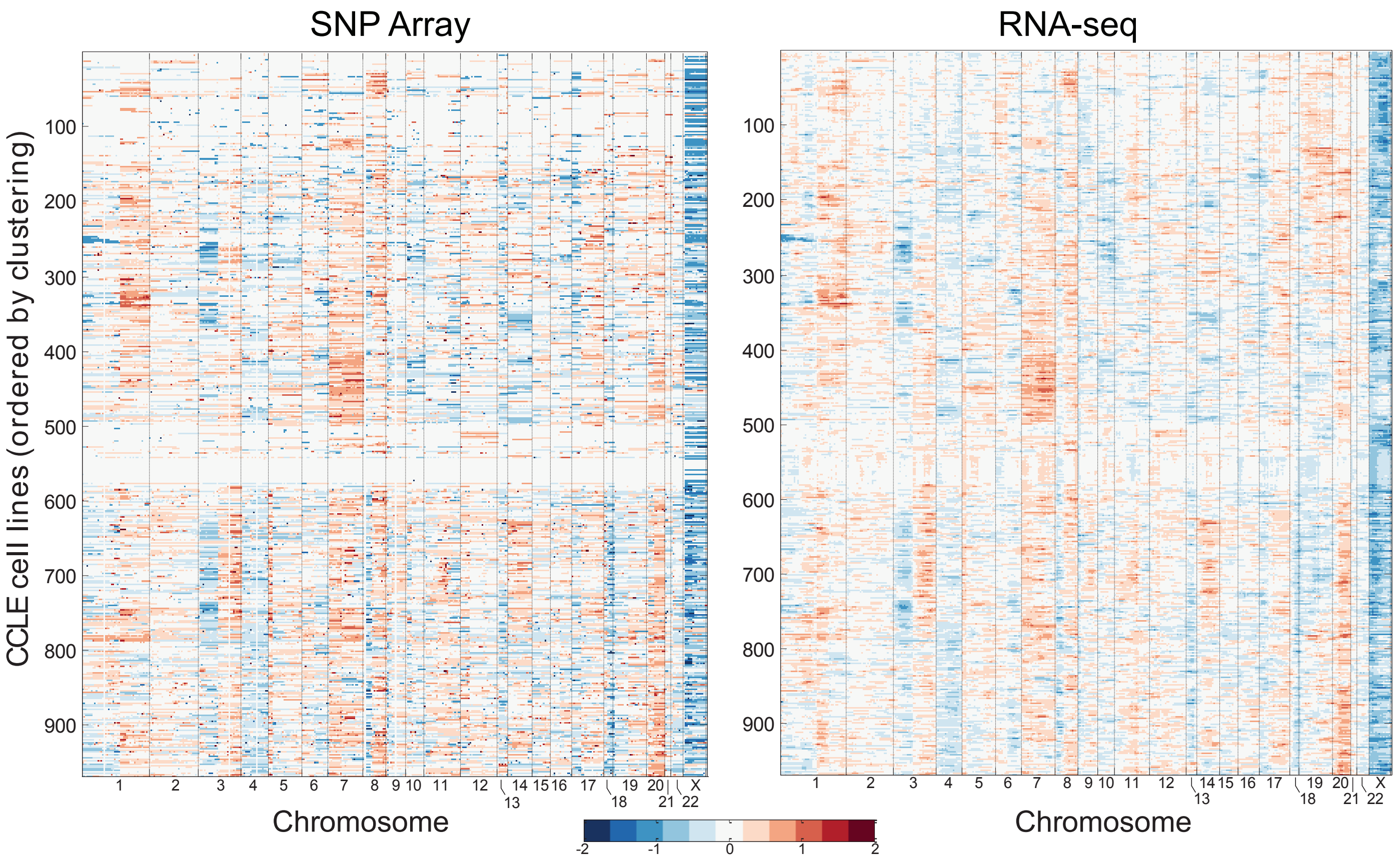
Supplemental Figure 20: Subtype classification of primary tumors based on population averaged data. Plot shows for each tumor (*x*-axis) the relative expression of each of four tumor sub-type signatures (bars; color legend). MGH26 is dominantly proneural, MGH28 and 29 are mesenchymal and MGH30 is classical. MGH31 was not clearly classifiable.



Supplemental Figure 21: Single cell qPCR validation of subtype heterogeneity. qPCR was performed on 91 (MGH26, left panel) and 76 (MGH30, right panel) additional cells using ten genes from the three most common subtype signatures (PN, CL, MES: blue, red and purple, respectively). Top: Heatmap of expression values for each gene (rows) in individual cells (columns), with genes sorted by signature. Bottom: average expression over the ten genes from each subtype. The relative proportions of cells from each subtype within MGH26 and MGH30 are consistent with those determined by single cell RNA-seq.



Supplemental Figure 22: Module/subtype correlations with stemness. Plot depicts –log10(p-values) from Student's t-test distributions for the correlations depicted in Fig. 3C. Dashed line represents a cutoff value of p<0.01. The proneural and classical subtype signatures and the Core TF and NFI modules all show significant correlations with stemness.



Supplemental Figure 23: Comparison of RNA-seq based copy number variation calls to SNP arrays. Plots depict CNV landscapes for 1,046 CCLE samples (row) for which there was both RNA-seq and SNP array data. CNV calling by SNP array (left) was highly correlated (r=0.72) with CNV calling by RNA-seq (right), supporting the use of this technique to infer CNV from RNA-seq profiles.

References and Notes

1. S. Yachida, S. Jones, I. Bozic, T. Antal, R. Leary, B. Fu, M. Kamiyama, R. H. Hruban, J. R. Eshleman, M. A. Nowak, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, C. A. Iacobuzio-Donahue, Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010). Medline doi:10.1038/nature09515

2. K. Eppert, K. Takenaka, E. R. Lechman, L. Waldron, B. Nilsson, P. van Galen, K. H. Metzeler, A. Poeppl, V. Ling, J. Beyene, A. J. Canty, J. S. Danska, S. K. Bohlander, C. Buske, M. D. Minden, T. R. Golub, I. Jurisica, B. L. Ebert, J. E. Dick, Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17**, 1086–1093 (2011). Medline doi:10.1038/nm.2415

3. D. W. Parsons, S. Jones, X. Zhang, J. C. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, I. M. Siu, G. L. Gallia, A. Olivi, R. McLendon, B. A. Rasheed, S. Keir, T. Nikolskaya, Y. Nikolsky, D. A. Busam, H. Tekleab, L. A. Diaz Jr., J. Hartigan, D. R. Smith, R. L. Strausberg, S. K. Marie, S. M. Shinjo, H. Yan, G. J. Riggins, D. D. Bigner, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu, K. W. Kinzler, An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008). Medline doi:10.1126/science.1164382

4. N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks, M. Wigler, Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011). Medline doi:10.1038/nature09807

5. M. Gerlinger, A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N. Q. McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C. R. Santos, M. Nohadani, A. C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P. A. Futreal, C. Swanton, Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012). Medline doi:10.1056/NEJMoa1113205

6. G. Driessens, B. Beck, A. Caauwe, B. D. Simons, C. Blanpain, Defining the mode of tumour growth by clonal analysis. *Nature* **488**, 527–530 (2012). Medline doi:10.1038/nature11344

7. A. G. Schepers, H. J. Snippert, D. E. Stange, M. van den Born, J. H. van Es, M. van de Wetering, H. Clevers, Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science* **337**, 730–735 (2012). Medline doi:10.1126/science.1224676

8. P. L. Bedard, A. R. Hansen, M. J. Ratain, L. L. Siu, Tumour heterogeneity in the clinic. *Nature* **501**, 355–364 (2013). Medline doi:10.1038/nature12627

9. R. Stupp, W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. C. Janzer, S. K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J. G. Cairncross, E. Eisenhauer, R. O. MirimanoffEuropean Organisation for Research and Treatment of Cancer Brain Tumor and Radiotherapy GroupsNational Cancer Institute of Canada Clinical Trials Group,

Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005). Medline doi:10.1056/NEJMoa043330

10. A. Sottoriva, I. Spiteri, S. G. Piccirillo, A. Touloumis, V. P. Collins, J. C. Marioni, C. Curtis, C. Watts, S. Tavaré, Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4009–4014 (2013). Medline doi:10.1073/pnas.1219747110

11. J. M. Stommel, A. C. Kimmelman, H. Ying, R. Nabioullin, A. H. Ponugoti, R. Wiedemeyer, A. H. Stegh, J. E. Bradner, K. L. Ligon, C. Brennan, L. Chin, R. A. DePinho, Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**, 287–290 (2007). Medline doi:10.1126/science.1142946

12. D. A. Nathanson, B. Gini, J. Mottahedeh, K. Visnyei, T. Koga, G. Gomez, A. Eskin, K. Hwang, J. Wang, K. Masui, A. Paucar, H. Yang, M. Ohashi, S. Zhu, J. Wykosky, R. Reed, S. F. Nelson, T. F. Cloughesy, C. D. James, P. N. Rao, H. I. Kornblum, J. R. Heath, W. K. Cavenee, F. B. Furnari, P. S. Mischel, Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014). 10.1126/science.1241328 Medline doi:10.1126/science.1241328

13. M. R. Gilbert, J. J. Dignam, T. S. Armstrong, J. S. Wefel, D. T. Blumenthal, M. A. Vogelbaum, H. Colman, A. Chakravarti, S. Pugh, M. Won, R. Jeraj, P. D. Brown, K. A. Jaeckle, D. Schiff, V. W. Stieber, D. G. Brachman, M. Werner-Wasik, I. W. Tremont-Lukats, E. P. Sulman, K. D. Aldape, W. J. Curran Jr., M. P. Mehta, A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N. Engl. J. Med.* **370**, 699–708 (2014). Medline doi:10.1056/NEJMoa1308573

14. C. E. Meacham, S. J. Morrison, Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013). Medline doi:10.1038/nature12624

15. S. K. Singh, C. Hawkins, I. D. Clarke, J. A. Squire, J. Bayani, T. Hide, R. M. Henkelman, M. D. Cusimano, P. B. Dirks, Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004). Medline doi:10.1038/nature03128

16. J. Chen, Y. Li, T. S. Yu, R. M. McKay, D. K. Burns, S. G. Kernie, L. F. Parada, A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522–526 (2012). Medline doi:10.1038/nature11287

17. Z. Li, S. Bao, Q. Wu, H. Wang, C. Eyler, S. Sathornsumetee, Q. Shi, Y. Cao, J. Lathia, R. E. McLendon, A. B. Hjelmeland, J. N. Rich, Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. *Cancer Cell* **15**, 501–513 (2009). Medline doi:10.1016/j.ccr.2009.03.018

18. S. Bao, Q. Wu, R. E. McLendon, Y. Hao, Q. Shi, A. B. Hjelmeland, M. W. Dewhirst, D. D. Bigner, J. N. Rich, Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756–760 (2006). Medline doi:10.1038/nature05236

19. K. P. Bhat, V. Balasubramaniyan, B. Vaillant, R. Ezhilarasan, K. Hummelink, F. Hollingsworth, K. Wani, L. Heathcock, J. D. James, L. D. Goodman, S. Conroy, L. Long, N. Lelic, S. Wang, J. Gumin, D. Raj, Y. Kodama, A. Raghunathan, A. Olar, K. Joshi, C. E. Pelloski, A. Heimberger, S. H. Kim, D. P. Cahill, G. Rao, W. F. Den Dunnen, H. W.

Boddeke, H. S. Phillips, I. Nakano, F. F. Lang, H. Colman, E. P. Sulman, K. Aldape, Mesenchymal differentiation mediated by NF-κB promotes radiation resistance in glioblastoma. *Cancer Cell* **24**, 331–346 (2013). Medline doi:10.1016/j.ccr.2013.08.001

20. C. W. Brennan, R. G. Verhaak, A. McKenna, B. Campos, H. Noushmehr, S. R. Salama, S. Zheng, D. Chakravarty, J. Z. Sanborn, S. H. Berman, R. Beroukhim, B. Bernard, C. J. Wu, G. Genovese, I. Shmulevich, J. Barnholtz-Sloan, L. Zou, R. Vegesna, S. A. Shukla, G. Ciriello, W. K. Yung, W. Zhang, C. Sougnez, T. Mikkelsen, K. Aldape, D. D. Bigner, E. G. Van Meir, M. Prados, A. Sloan, K. L. Black, J. Eschbacher, G. Finocchiaro, W. Friedman, D. W. Andrews, A. Guha, M. Iacocca, B. P. O'Neill, G. Foltz, J. Myers, D. J. Weisenberger, R. Penny, R. Kucherlapati, C. M. Perou, D. N. Hayes, R. Gibbs, M. Marra, G. B. Mills, E. Lander, P. Spellman, R. Wilson, C. Sander, J. Weinstein, M. Meyerson, S. Gabriel, P. W. Laird, D. Haussler, G. Getz, L. Chin; TCGA Research Network, The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013). Medline doi:10.1016/j.cell.2013.09.034

21. R. G. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B. A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H. S. Feiler, J. G. Hodgson, C. D. James, J. N. Sarkaria, C. Brennan, A. Kahn, P. T. Spellman, R. K. Wilson, T. P. Speed, J. W. Gray, M. Meyerson, G. Getz, C. M. Perou, D. N. Hayes; Cancer Genome Atlas Research Network, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010). Medline doi:10.1016/j.ccr.2009.12.020

22. D. Ramsköld, S. Luo, Y. C. Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtukova, J. F. Loring, L. C. Laurent, G. P. Schroth, R. Sandberg, Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012). Medline doi:10.1038/nbt.2282

23. A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park, A. Regev, Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013). Medline doi:10.1038/nature12172

24. Materials and methods are available as supplementary materials on *Science* Online.

25. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G. Korzeniewski, C. Shive, D. Tabor, L. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K. Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, D. MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong,

M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S. Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A. R. Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore; GTEx Consortium, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013). Medline doi:10.1038/ng.2653

26. N. Novershtern, A. Subramanian, L. N. Lawton, R. H. Mak, W. N. Haining, M. E. McConkey, N. Habib, N. Yosef, C. Y. Chang, T. Shay, G. M. Frampton, A. C. Drake, I. Leskov, B. Nilsson, F. Preffer, D. Dombkowski, J. W. Evans, T. Liefeld, J. S. Smutko, J. Chen, N. Friedman, R. A. Young, T. R. Golub, A. Regev, B. L. Ebert, Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011). Medline doi:10.1016/j.cell.2011.01.004

27. J. Lee, S. Kotliarova, Y. Kotliarov, A. Li, Q. Su, N. M. Donin, S. Pastorino, B. W. Purow, N. Christopher, W. Zhang, J. K. Park, H. A. Fine, 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* **9**, 391–403 (2006). Medline doi:10.1016/j.ccr.2006.03.030

28. H. Wakimoto, G. Mohapatra, R. Kanai, W. T. Curry Jr., S. Yip, M. Nitta, A. P. Patel, Z. R. Barnard, A. O. Stemmer-Rachamimov, D. N. Louis, R. L. Martuza, S. D. Rabkin, Maintenance of primary tumor phenotype and genotype in glioblastoma stem cells. *Neuro-oncol.* **14**, 132–144 (2012). Medline doi:10.1093/neuonc/nor195

29. T. F. Cloughesy, W. K. Cavenee, P. S. Mischel, Glioblastoma: From molecular pathology to targeted treatment. *Annu. Rev. Pathol.* **9**, 1–25 (2014). Medline doi:10.1146/annurev-pathol-011110-130324

30. M. Snuderl, L. Fazlollahi, L. P. Le, M. Nitta, B. H. Zhelyazkova, C. J. Davidson, S. Akhavanfard, D. P. Cahill, K. D. Aldape, R. A. Betensky, D. N. Louis, A. J. Iafrate, Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**, 810–817 (2011). Medline doi:10.1016/j.ccr.2011.11.005

31. B. D. Choi, C. T. Kuan, M. Cai, G. E. Archer, D. A. Mitchell, P. C. Gedeon, L. Sanchez-Perez, I. Pastan, D. D. Bigner, J. H. Sampson, Systemic administration of a bispecific antibody targeting EGFRvIII successfully treats intracerebral glioma. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 270–275 (2013). Medline doi:10.1073/pnas.1219817110

32. H. Wang, M. Zhou, B. Shi, Q. Zhang, H. Jiang, Y. Sun, J. Liu, K. Zhou, M. Yao, J. Gu, S. Yang, Y. Mao, Z. Li, Identification of an exon 4-deletion variant of epidermal growth factor receptor with increased metastasis-promoting capacity. *Neoplasia* **13**, 461–471 (2011). Medline

33. J. Xu, J. Qian, Y. Hu, J. Wang, X. Zhou, H. Chen, J. Y. Fang, Heterogeneity of Li-Fraumeni syndrome links to unequal gain-of-function effects of p53 mutations. *Sci. Rep.* **4**, 4223 (2014). Medline doi:10.1038/srep04223

34. M. S. Cho, H. G. Vasquez, R. Rupaimoole, S. Pradeep, S. Wu, B. Zand, H. D. Han, C. Rodriguez-Aguayo, J. Bottsford-Miller, J. Huang, T. Miyake, H. J. Choi, H. J. Dalton, C.

Ivan, K. Baggerly, G. Lopez-Berestein, A. K. Sood, V. Afshar-Kharghan, Autocrine effects of tumor-derived complement. *Cell Reports* **6**, 1085–1095 (2014). Medline doi:10.1016/j.celrep.2014.02.014

35. L. Sang, H. A. Coller, J. M. Roberts, Control of the reversibility of cellular quiescence by the transcriptional repressor HES1. *Science* **321**, 1095–1100 (2008). Medline doi:10.1126/science.1155998
36. K. O. Shostak, V. V. Dmitrenko, O. M. Garifulin, V. D. Rozumenko, O. V. Khomenko, Y. A. Zozulya, G. Zehetner, V. M. Kavsan, Downregulation of putative tumor suppressor gene TSC-22 in human brain tumors. *J. Surg. Oncol.* **82**, 57–64 (2003). Medline doi:10.1002/jso.10180
37. A. Roesch, M. Fukunaga-Kalabis, E. C. Schmidt, S. E. Zabierowski, P. A. Brafford, A. Vultur, D. Basu, P. Gimotty, T. Vogt, M. Herlyn, A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* **141**, 583–594 (2010). Medline doi:10.1016/j.cell.2010.04.020
38. L. Ricci-Vitiani, R. Pallini, M. Biffoni, M. Todaro, G. Invernici, T. Cenci, G. Maira, E. A. Parati, G. Stassi, L. M. Larocca, R. De Maria, Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature* **468**, 824–828 (2010). Medline doi:10.1038/nature09557
39. R. Wang, K. Chadalavada, J. Wilshire, U. Kowalik, K. E. Hovinga, A. Geber, B. Fligelman, M. Leversha, C. Brennan, V. Tabar, Glioblastoma stem-like cells give rise to tumour endothelium. *Nature* **468**, 829–833 (2010). Medline doi:10.1038/nature09624
40. Y. Soda, T. Marumoto, D. Friedmann-Morvinski, M. Soda, F. Liu, H. Michiue, S. Pastorino, M. Yang, R. M. Hoffman, S. Kesari, I. M. Verma, Transdifferentiation of glioblastoma cells into vascular endothelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4274–4280 (2011). Medline doi:10.1073/pnas.1016030108
41. B. Martynoga, J. L. Mateo, B. Zhou, J. Andersen, A. Achimastou, N. Urbán, D. van den Berg, D. Georgopoulou, S. Hadjur, J. Wittbrodt, L. Ettwiller, M. Piper, R. M. Gronostajski, F. Guillemot, Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev.* **27**, 1769–1786 (2013). Medline doi:10.1101/gad.216804.113
42. M. L. Suvà, E. Rheinbay, S. M. Gillespie, A. P. Patel, H. Wakimoto, S. D. Rabkin, N. Riggi, A. S. Chi, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, M. N. Rivera, N. Rossetti, S. Kasif, S. Beik, S. Kadri, I. Tirosh, I. Wortman, A. K. Shalek, O. Rozenblatt-Rosen, A. Regev, D. N. Louis, B. E. Bernstein, Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014). Medline doi:10.1016/j.cell.2014.02.030
43. H. Noushmehr, D. J. Weisenberger, K. Diefes, H. S. Phillips, K. Pujara, B. P. Berman, F. Pan, C. E. Pelloski, E. P. Sulman, K. P. Bhat, R. G. Verhaak, K. A. Hoadley, D. N. Hayes, C. M. Perou, H. K. Schmidt, L. Ding, R. K. Wilson, D. Van Den Berg, H. Shen, H. Bengtsson, P. Neuvial, L. M. Cope, J. Buckley, J. G. Herman, S. B. Baylin, P. W. Laird, K. Aldape; Cancer Genome Atlas Research Network, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010). Medline doi:10.1016/j.ccr.2010.03.017

44. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). Medline doi:10.1186/gb-2009-10-3-r25

45. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). Medline doi:10.1186/1471-2105-12-323

46. J. D. Cahoy, B. Emery, A. Kaushal, L. C. Foo, J. L. Zamanian, K. S. Christopherson, Y. Xing, J. L. Lubischer, P. A. Krieg, S. A. Krupenko, W. J. Thompson, B. A. Barres, A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008). Medline doi:10.1523/JNEUROSCI.4178-07.2008

47. O. Butovsky, M. P. Jedrychowski, C. S. Moore, R. Cialic, A. J. Lanser, G. Gabriely, T. Koeglsperger, B. Dake, P. M. Wu, C. E. Doykan, Z. Fanek, L. Liu, Z. Chen, J. D. Rothstein, R. M. Ransohoff, S. P. Gygi, J. P. Antel, H. L. Weiner, Identification of a unique TGF-β-dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014). Medline doi:10.1038/nn.3599

48. M. Ho, E. Yang, G. Matcuk, D. Deng, N. Sampas, A. Tsalenko, R. Tabibiazar, Y. Zhang, M. Chen, S. Talbi, Y. D. Ho, J. Wang, P. S. Tsao, A. Ben-Dor, Z. Yakhini, L. Bruhn, T. Quertermous, Identification of endothelial cell genes by combined database mining and microarray analysis. *Physiol. Genomics* **13**, 249–262 (2003). Medline