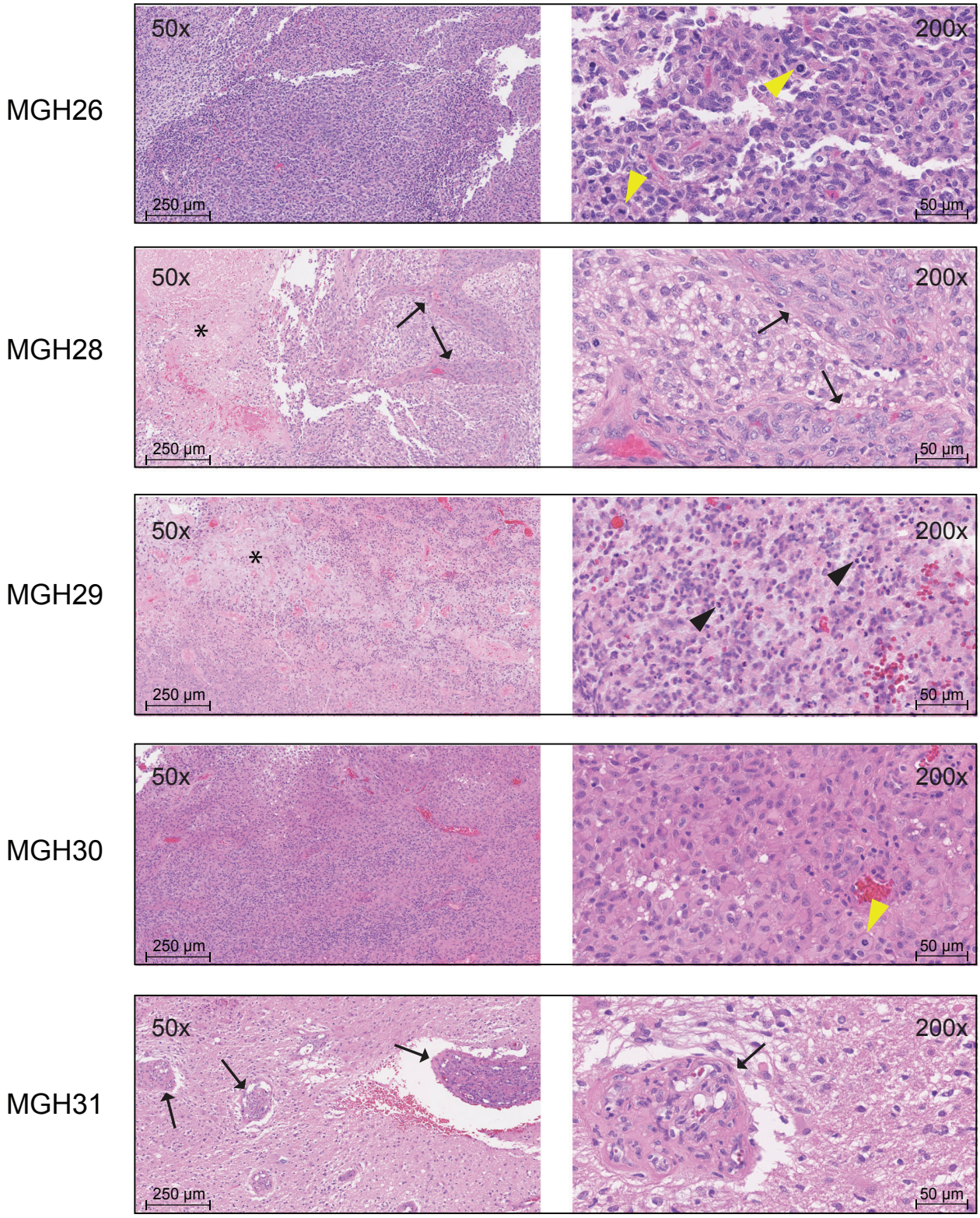
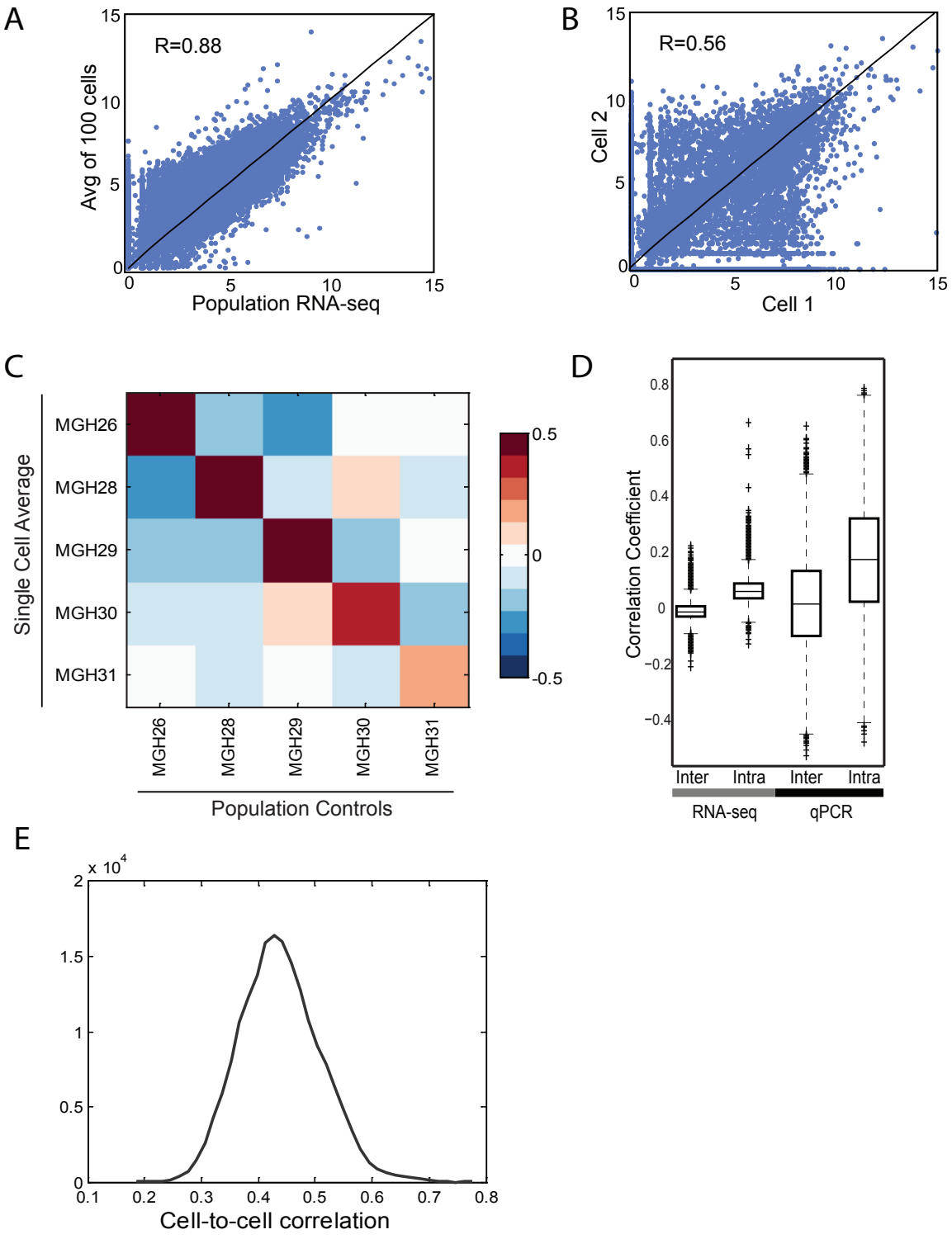
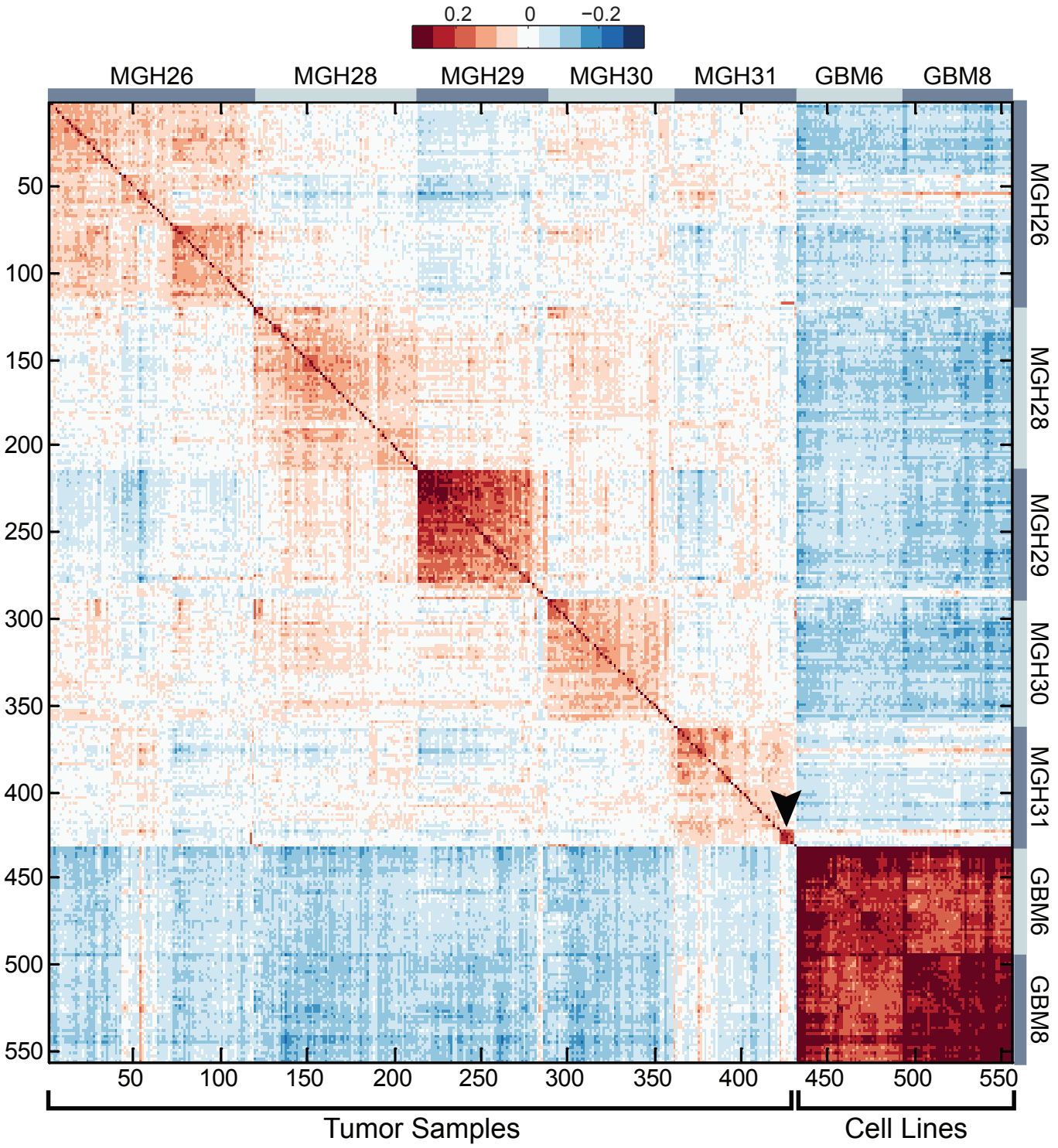
Supplemental Figures



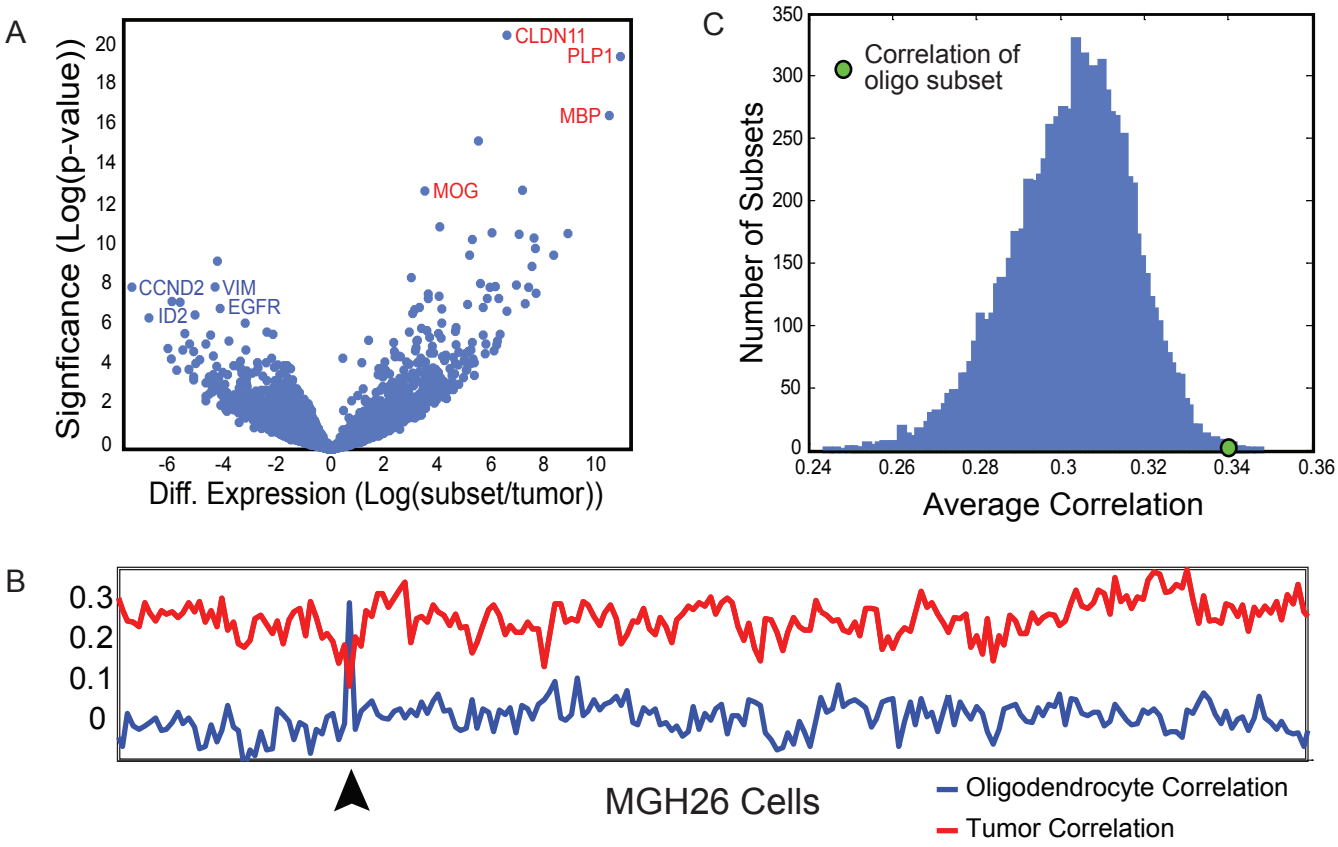
Supplemental Figure 1: Histology of primary glioblastoma cases used in this study. All tumors were confirmed to be primary glioblastoma, WHO grade IV/IV. Representative examples of morphology for each tumor at 50x (left) and 200x (right) magnification are shown. MGH26 and 30 show characteristic cytonuclear 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.



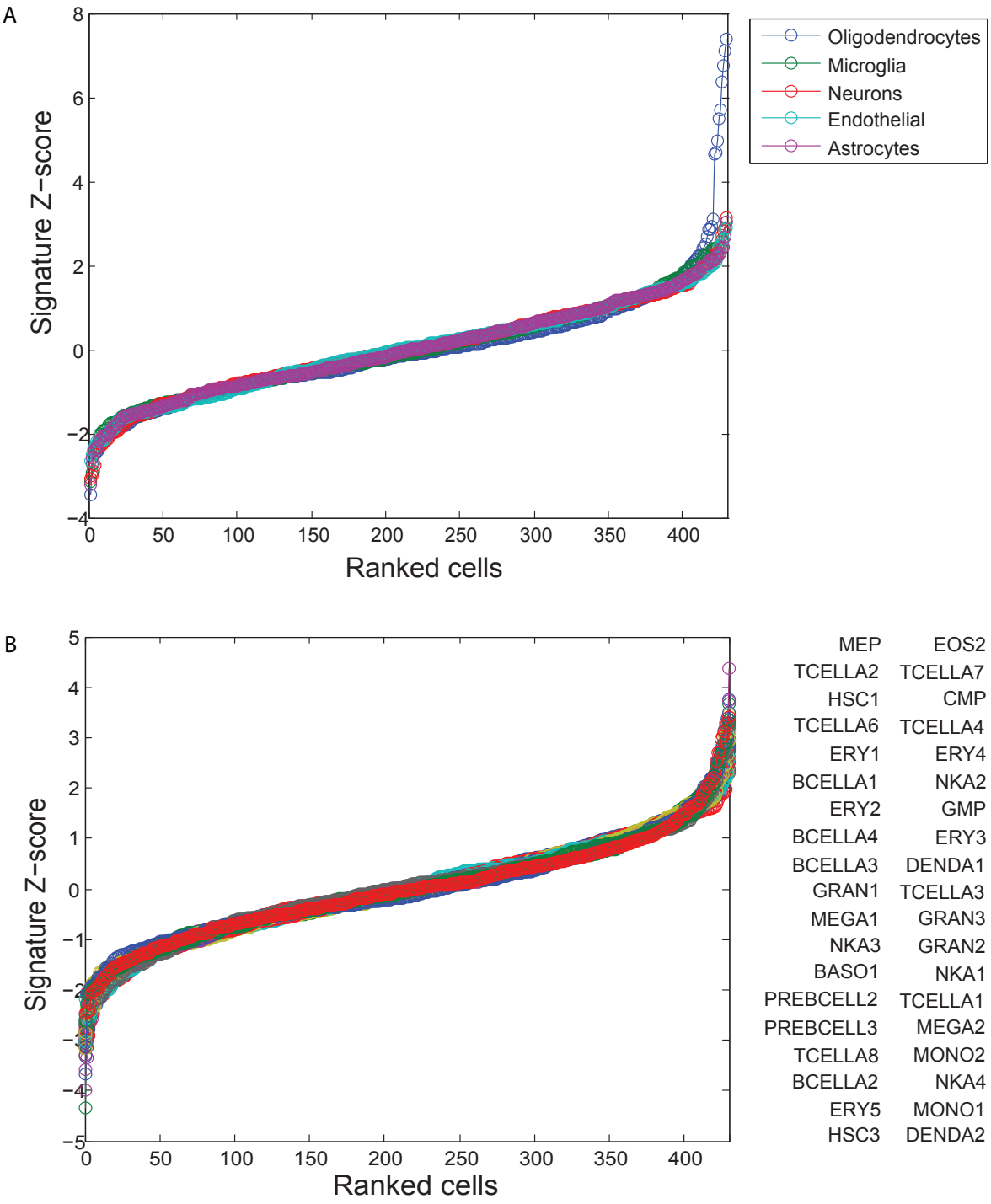
Supplemental Figure 2: Correlation between single cell profiles and (bulk) population profiles. (A) Scatter plot of gene expression values (log_2(TPM+1)) for the population (bulk) RNA-Seq (*x*- axis) and the average log_2(TPM+1) from 100 single cells for MGH26 (*y*-axis), showing high correlation (R=0.88). (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).



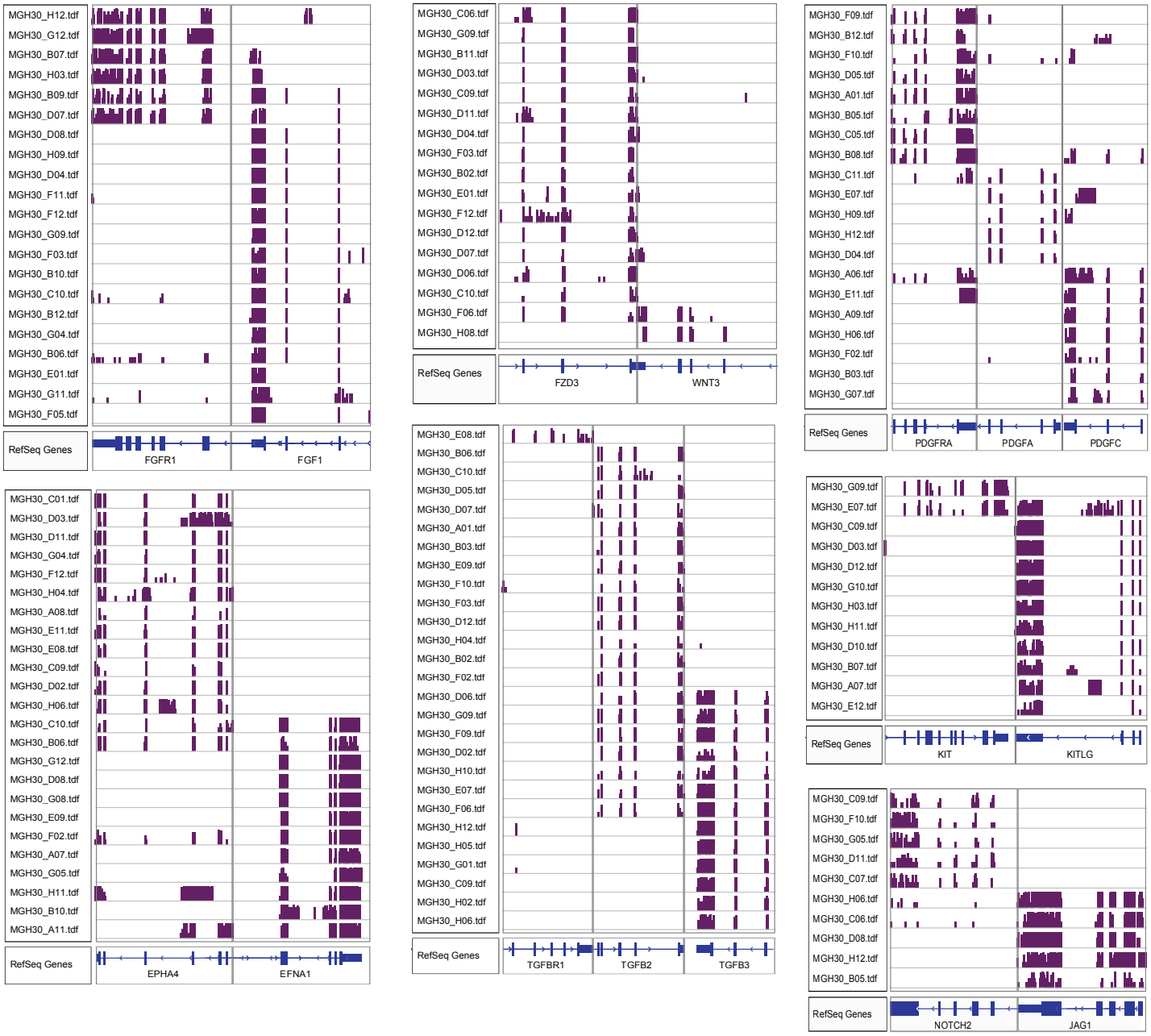
Supplemental Figure 3: Similarity and variation in single cell profiles. Correlation matrix for all 430 cells in the dataset (rows), ordered by tumor (MGH 26, 28, 29, 30, 31), and single cell profiles for the gliomasphere cell lines (GBM6, GBM8). While cells from each tumor demonstrate higher overall 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).



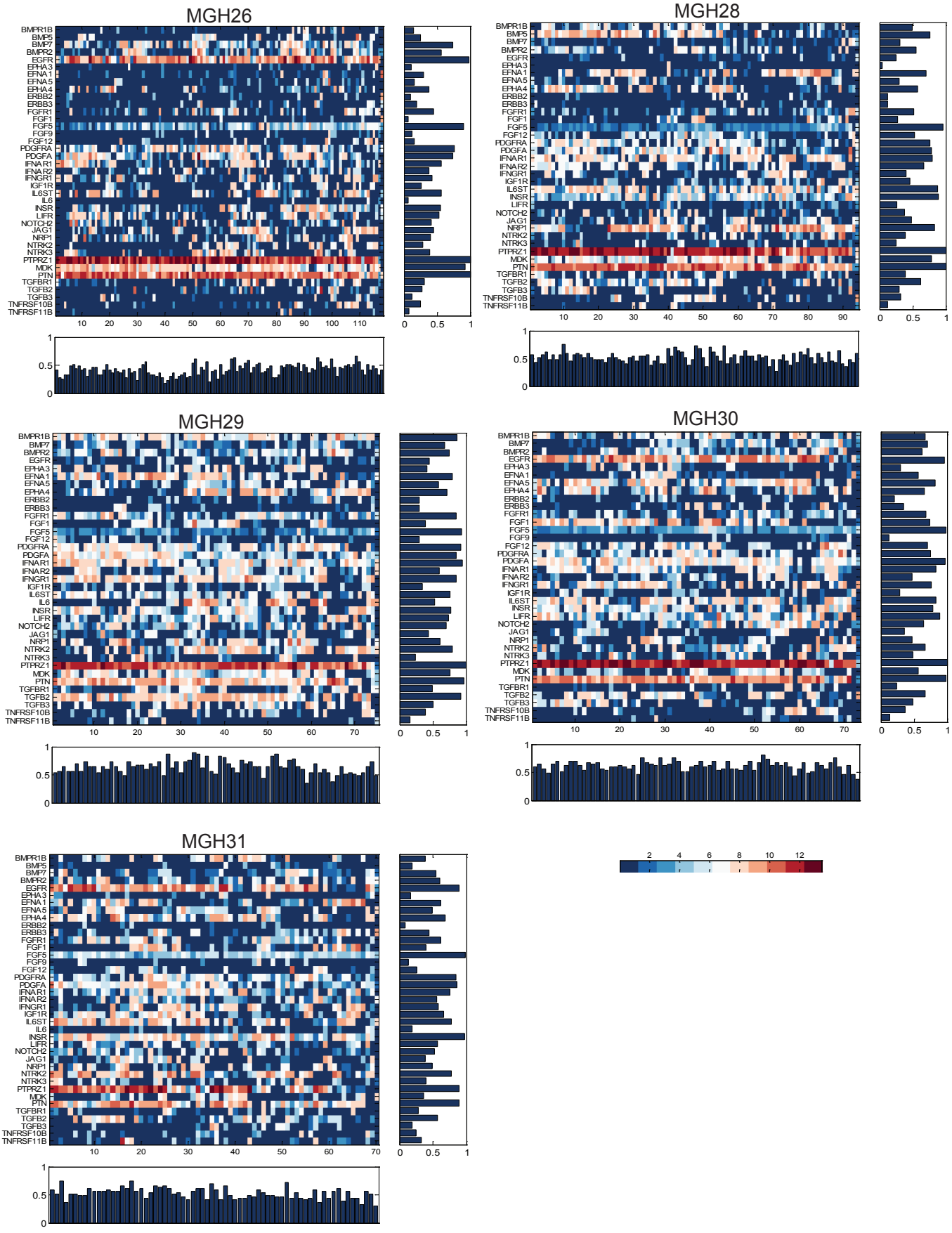
Supplemental Figure 4: Identification of oligodendrocytes. (A) Volcano plot shows for each gene the significance (*y*-axis) and ratio (*x*-axis) of its differential expression between the highly correlated subset of cells in **Fig. S3** and the remainder of cells in MGH31 (significance: $-\log_{10}$(P-value), 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).



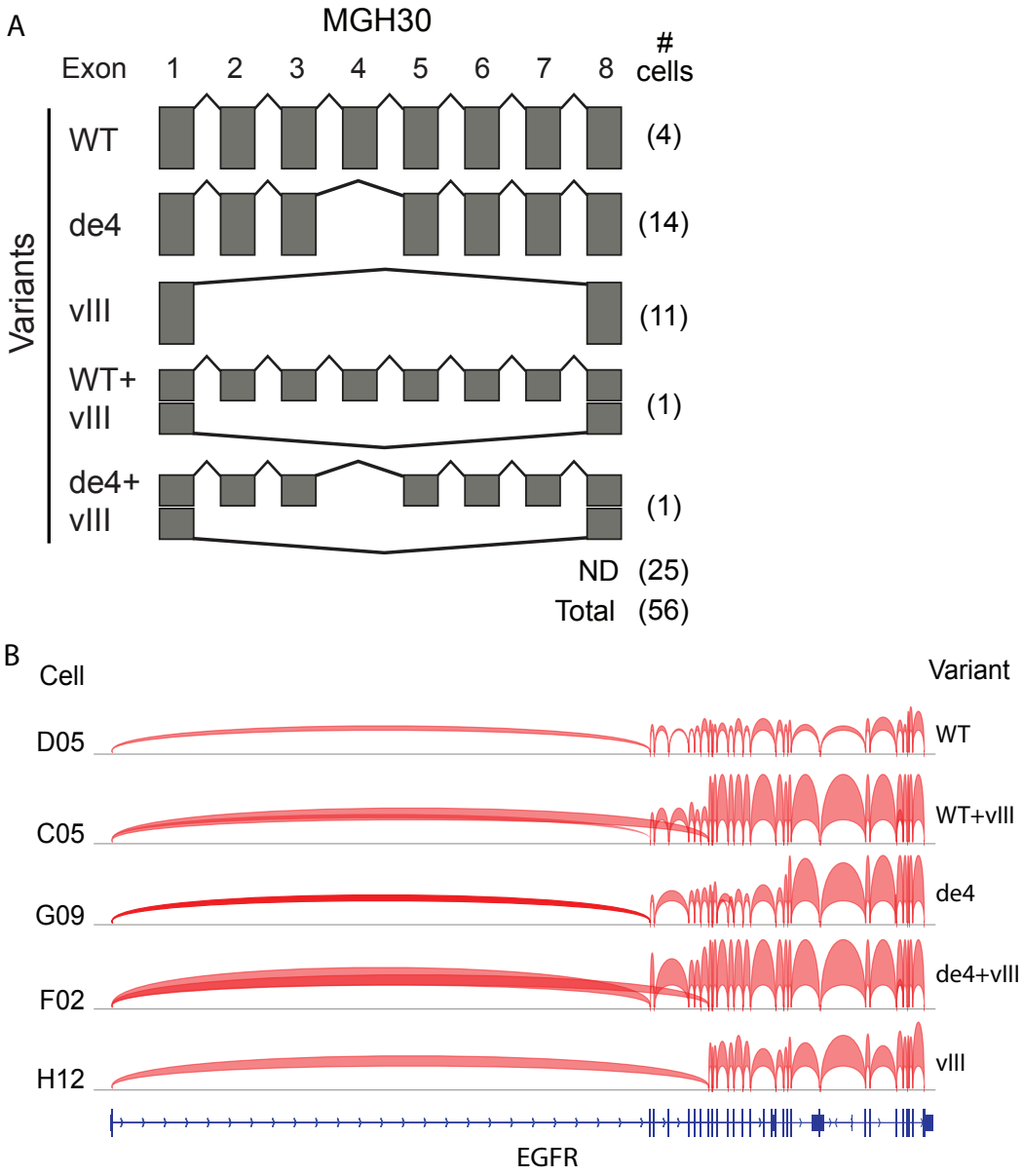
Supplemental Figure 5: (A) For each cell (*x*-axis), plot shows Z-scores (*y*- axis) reflecting similarity to each of five normal cell type signatures (color legends). Cells are ranked by their scores for each signature independently. Aside from a high correlation of several cells with normal oligodendrocyte 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.



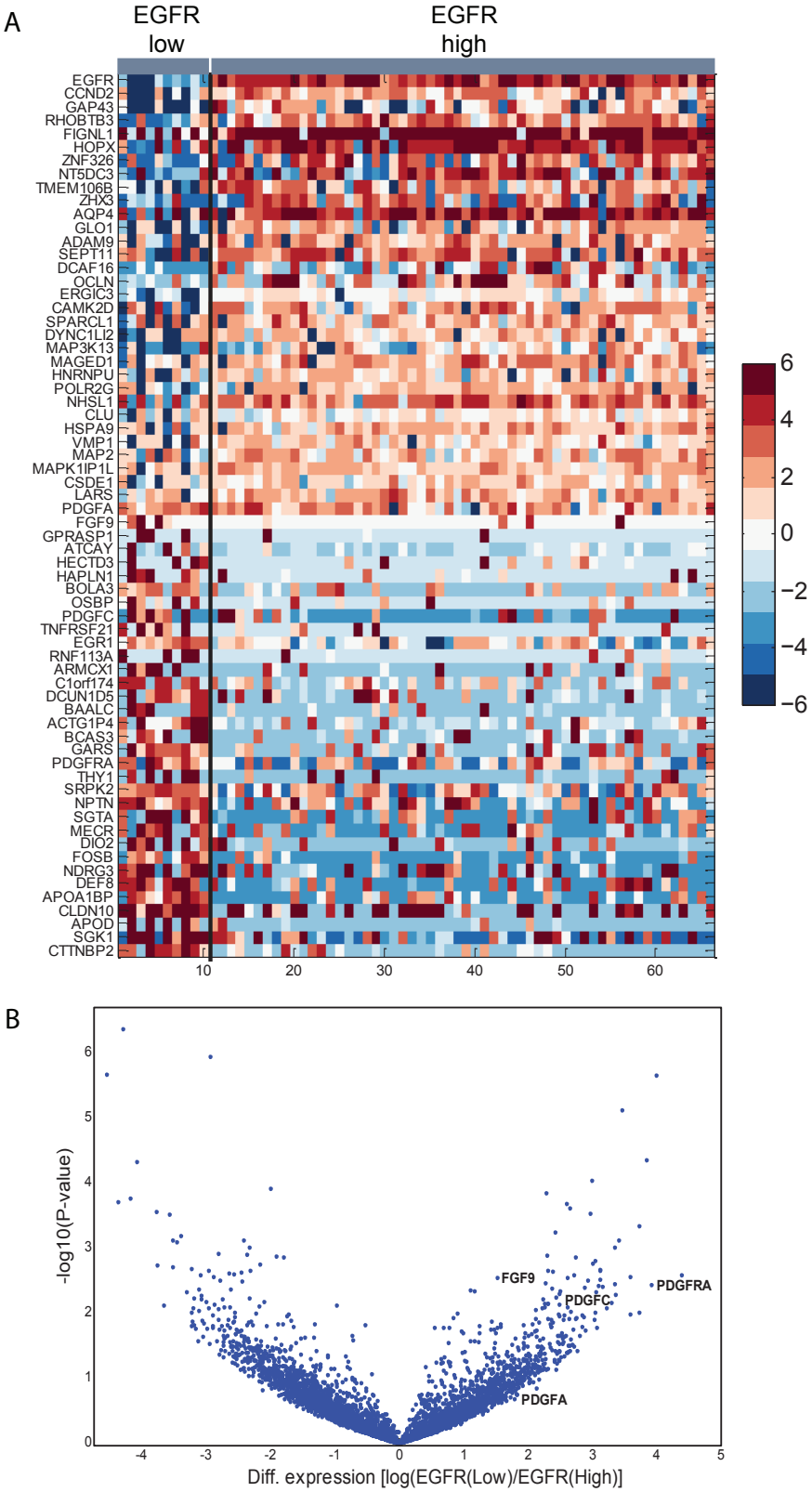
Supplemental Figure 6: Variation in expression of receptors and ligands in MGH30. Shown are RNA-seq read densities (vertical scale 10) for selected receptors and their ligands in MGH30, demonstrating variability in expression across individual cells (rows).



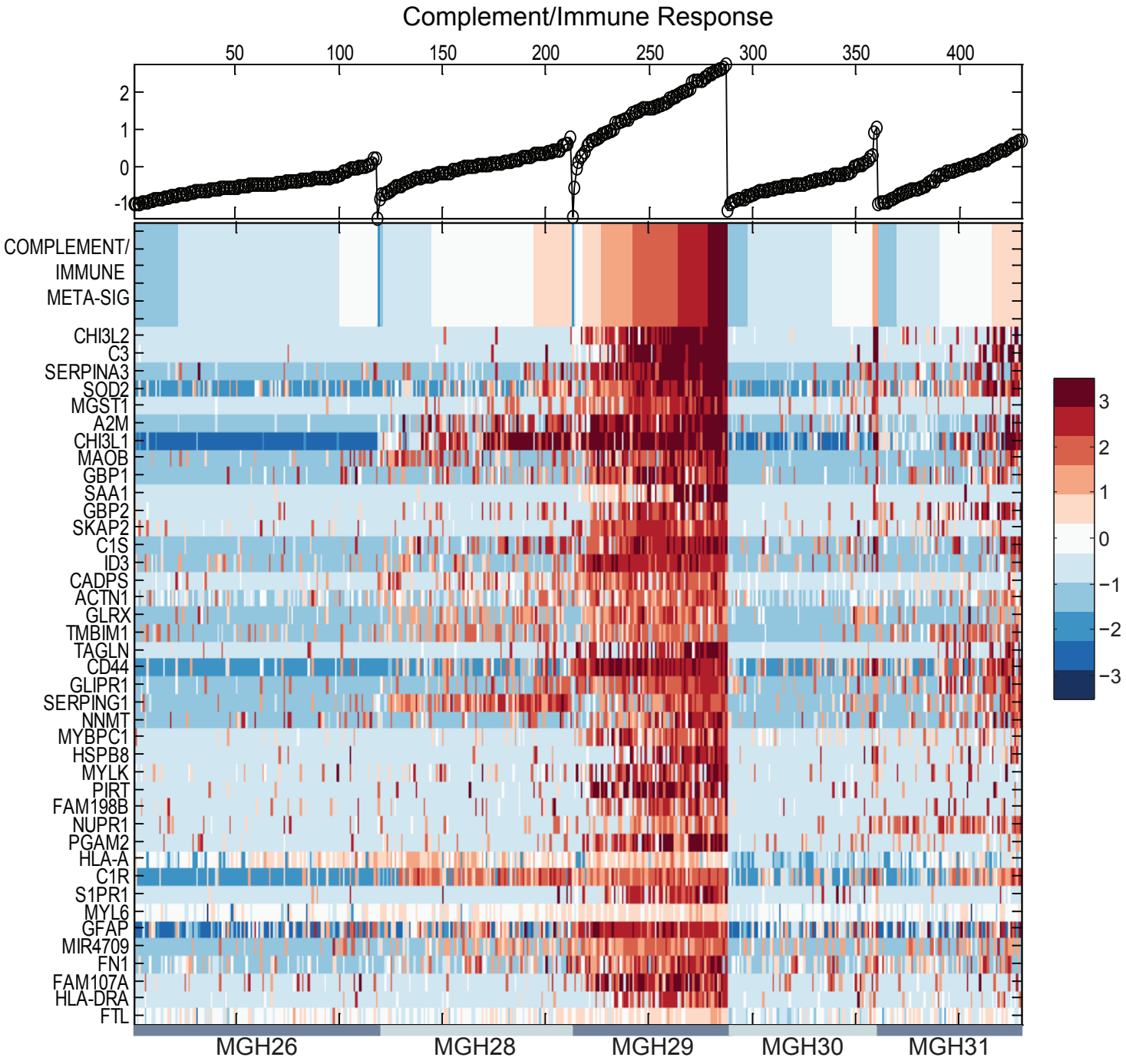
Supplemental Figure 7: Variation in expression of receptors and ligands in each tumor. Heatmaps for each tumor show expression values for receptors and ligands (rows, alphabetically ordered with receptors and ligands paired) in each cell (columns). All receptors and ligands represented among the 5,948 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).



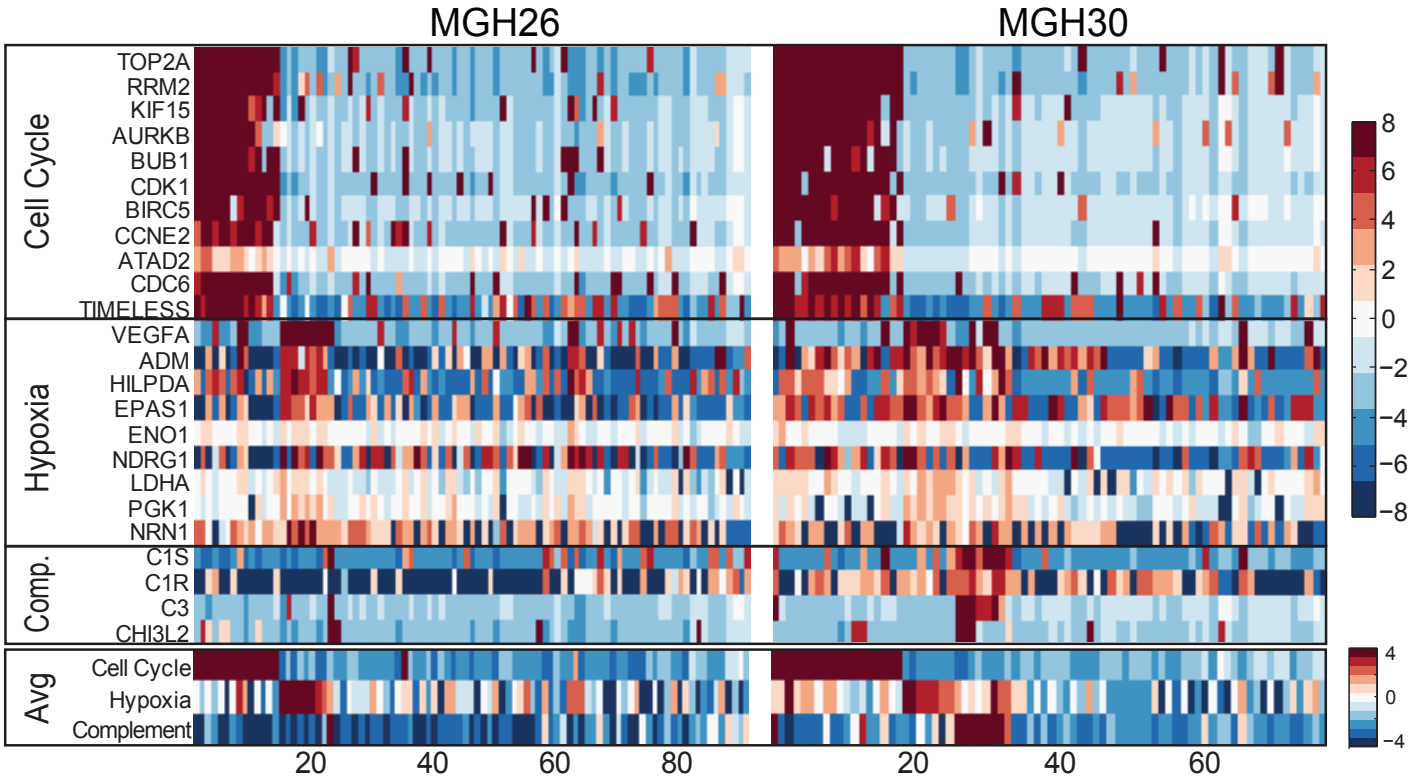
Supplemental Figure 8: Variation in EGFR splicing patterns in single cells. (A) Schematic depicts *EGFR* splice variants detected by mapping of junction-spanning 'spliced' reads. Three different variants were expressed in relatively mutually exclusive patterns across individual cells from MGH30, an *EGFR* 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).



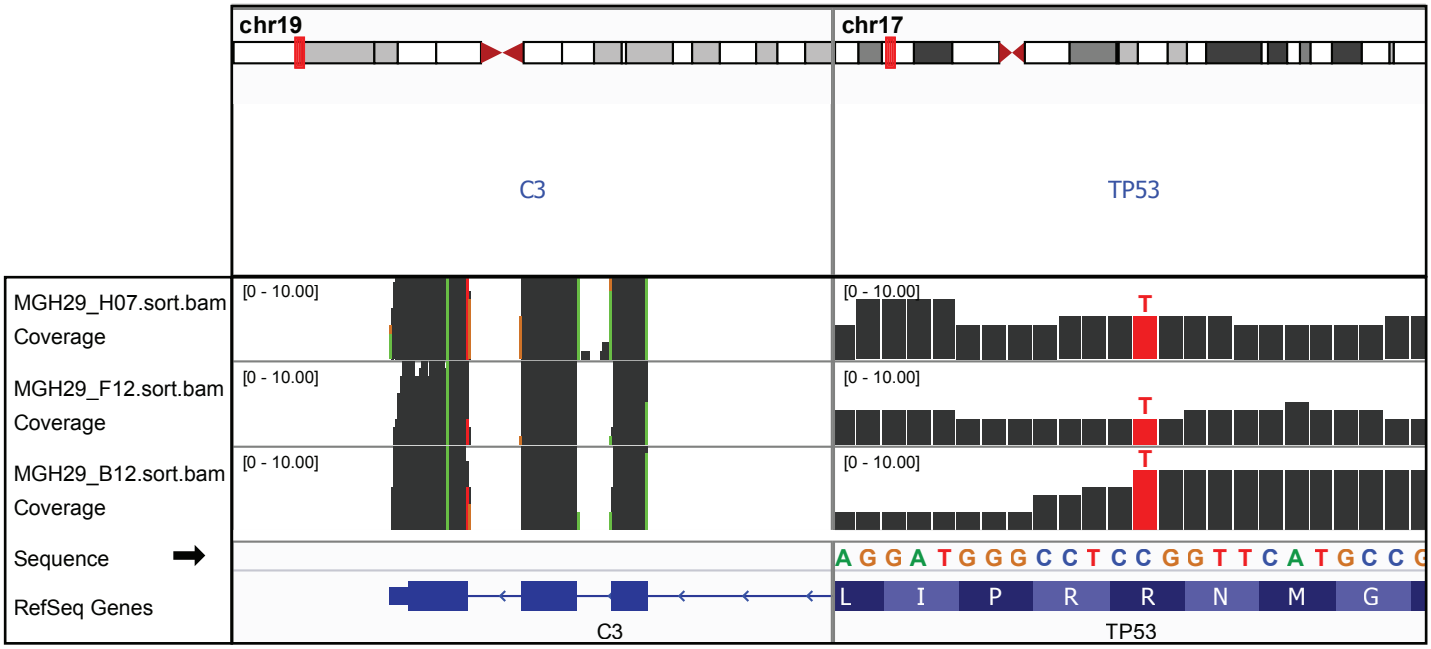
Supplemental Figure 9: Expression variation across cells associated with EGFR expression status. (A) Heatmap of genes (rows) that are differentially expressed between EGFR high and low cells (columns) in MGH30. (B) Volcano plot showing for each gene its differential expression in EGFR low vs. high single 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.



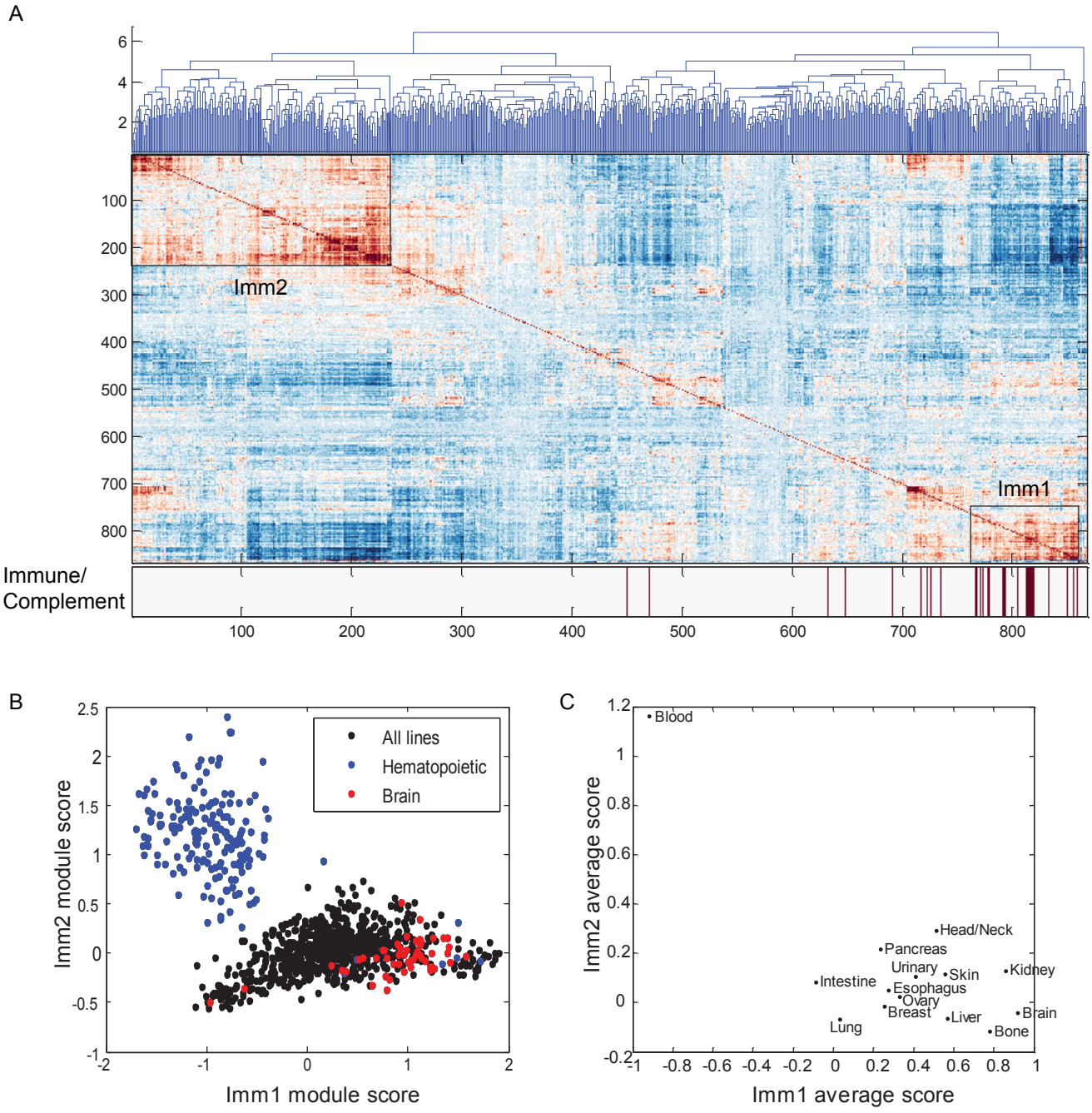
Supplemental Figure 10: Complement meta-signature. Heatmap shows the average expression of the meta-signature (top row) and individual genes from the signature (rows) in single glioblastoma cells (columns), grouped by tumor and ordered within each tumor from lowest to highest meta-signature score (*y*-axis,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



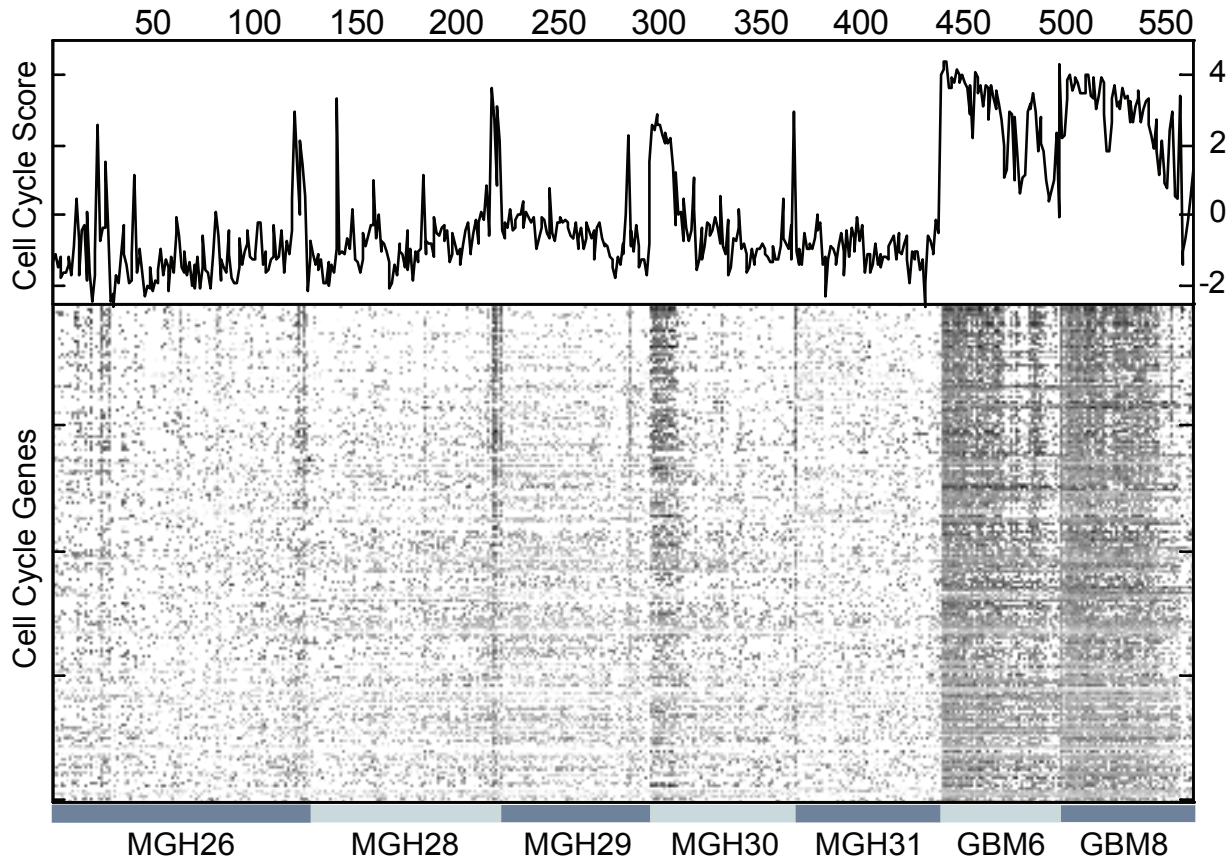
Supplemental Figure 11: Validation of meta-signatures by single cell qPCR. Top three panels show expression levels from single cell qPCR for 4-10 genes (rows) per meta-signature (labeled on far left) in 91 (MGH26, left panel) and 76 (MGH30, right panel) additional cells. Bottom panel shows average expression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



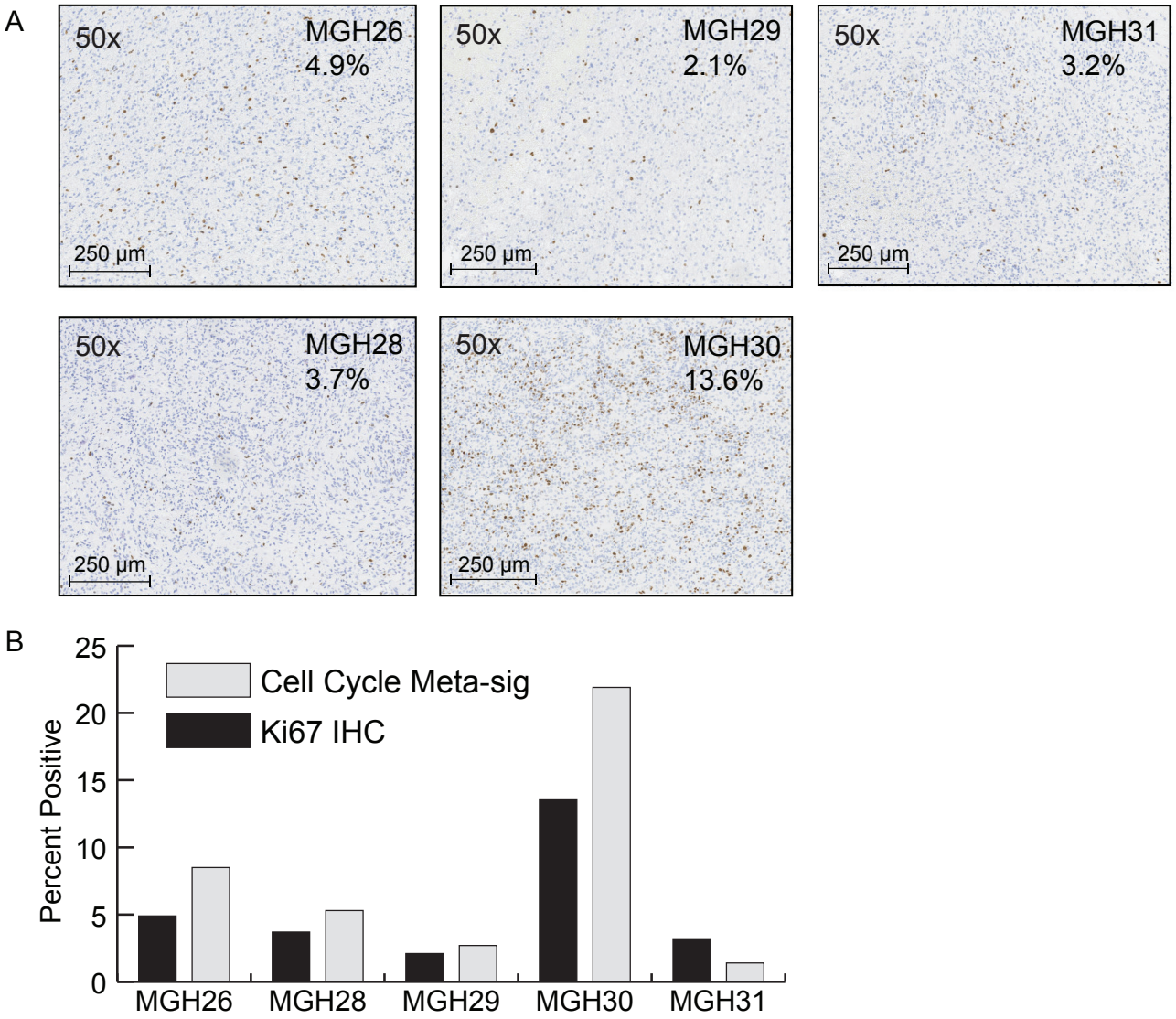
Supplemental Figure 12: Expression of complement C3 in tumor cells. Three cells were identified in MGH29 that definitively contain the R248L (C→T) mutation (right panel, highlighted in red). This mutation is highly oncogenic and occurs in the setting of malignancy. The same three cells also express complement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



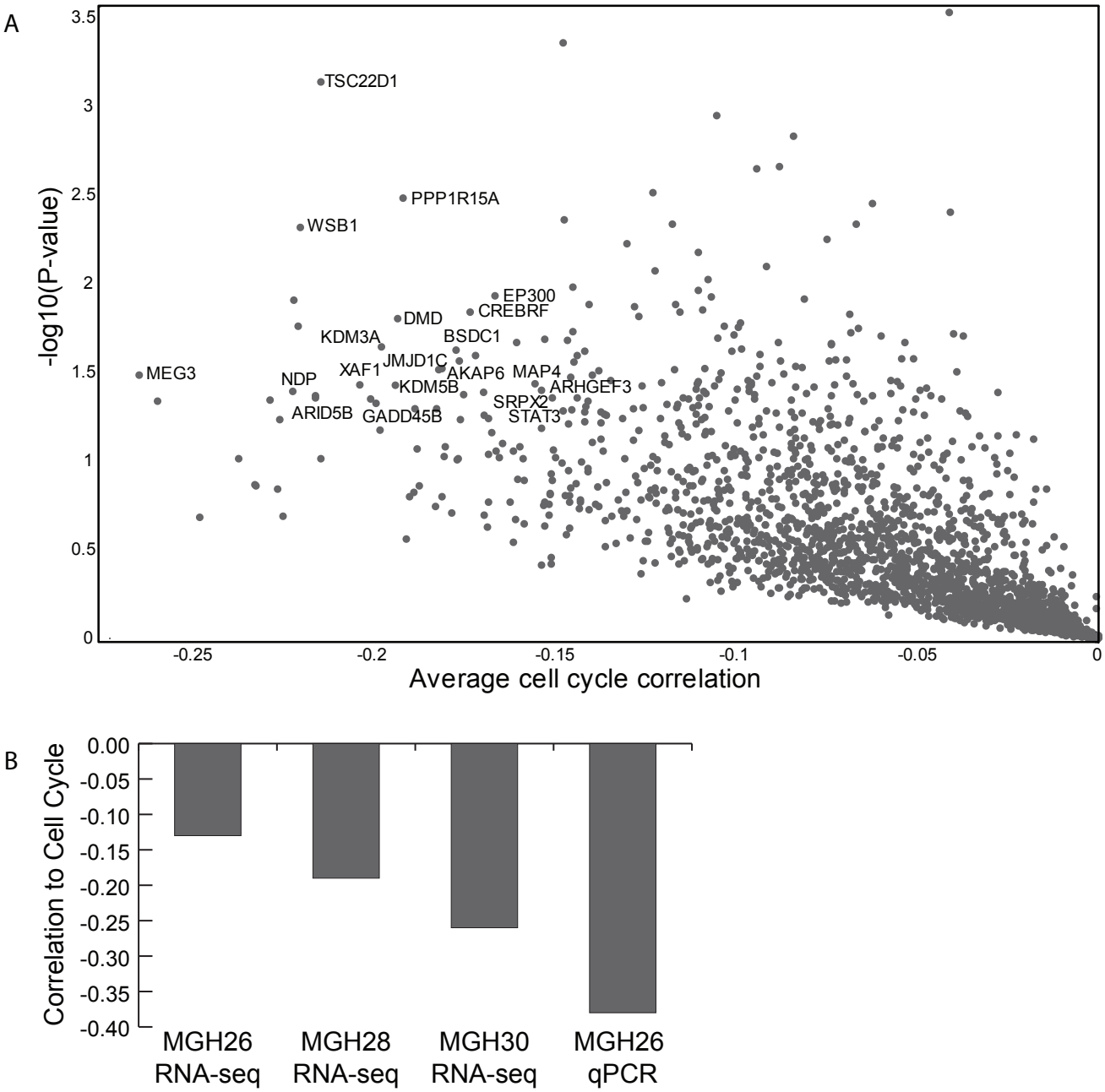
Supplemental Figure 13: Co-expression of the complement/immune module in cancer cell lines. (A) Unsupervised clustering of all annotated immune genes in ~1,000 cell lines from the Cancer Cell Line Encyclopedia identifies two clusters of co-expressed genes. Imm1 is highly enriched for genes from the complement/immune 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.



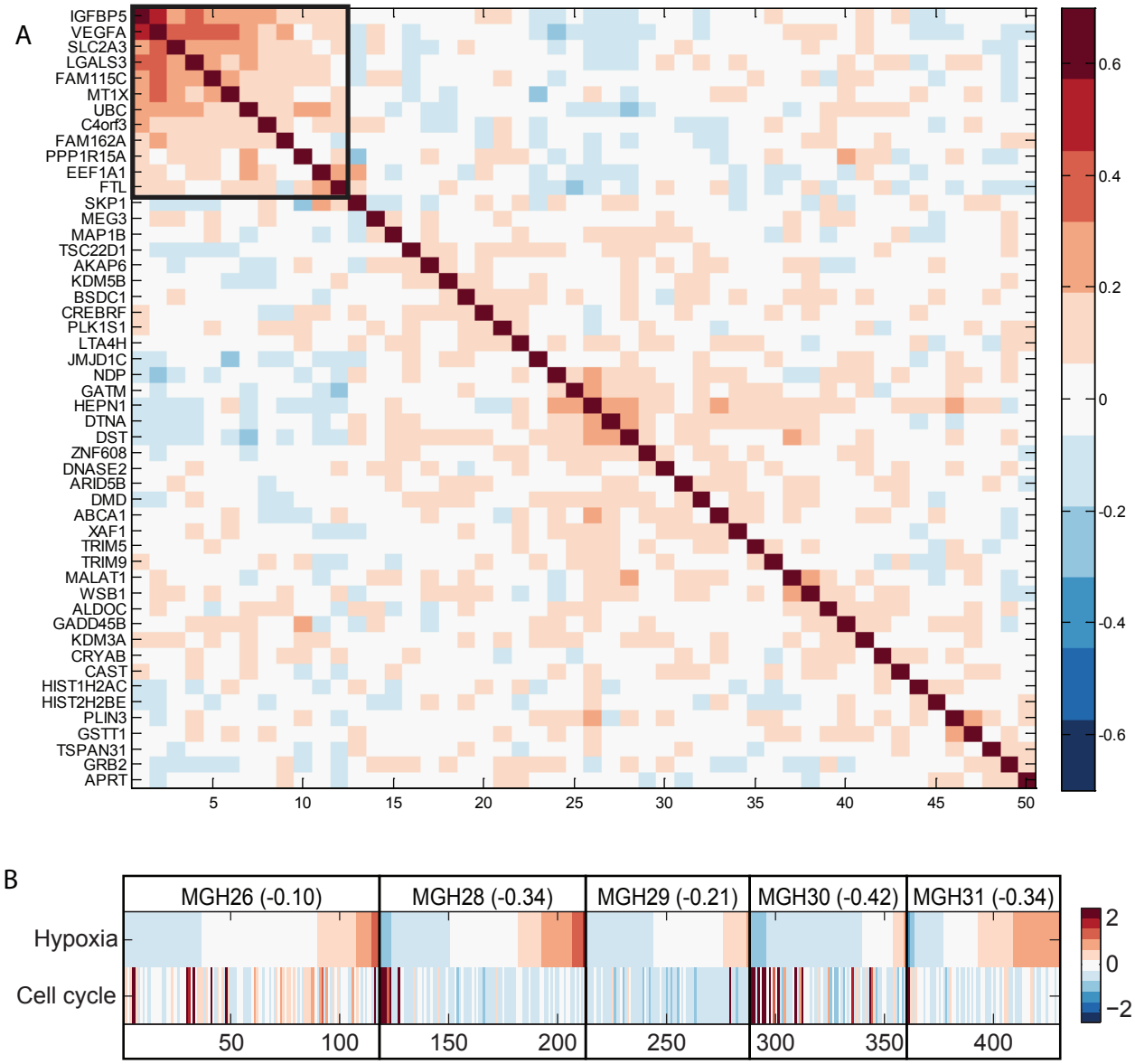
Supplemental Figure 14: Cell cycle meta-signature. Cell cycle meta-signature scores (*y*-axis, top plot) and the expression of each individual gene in the meta-signature (rows in grayscale heatmap) are shown for each individual cell (columns) in the five tumors and two gliomasphere cell lines (GBM6,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



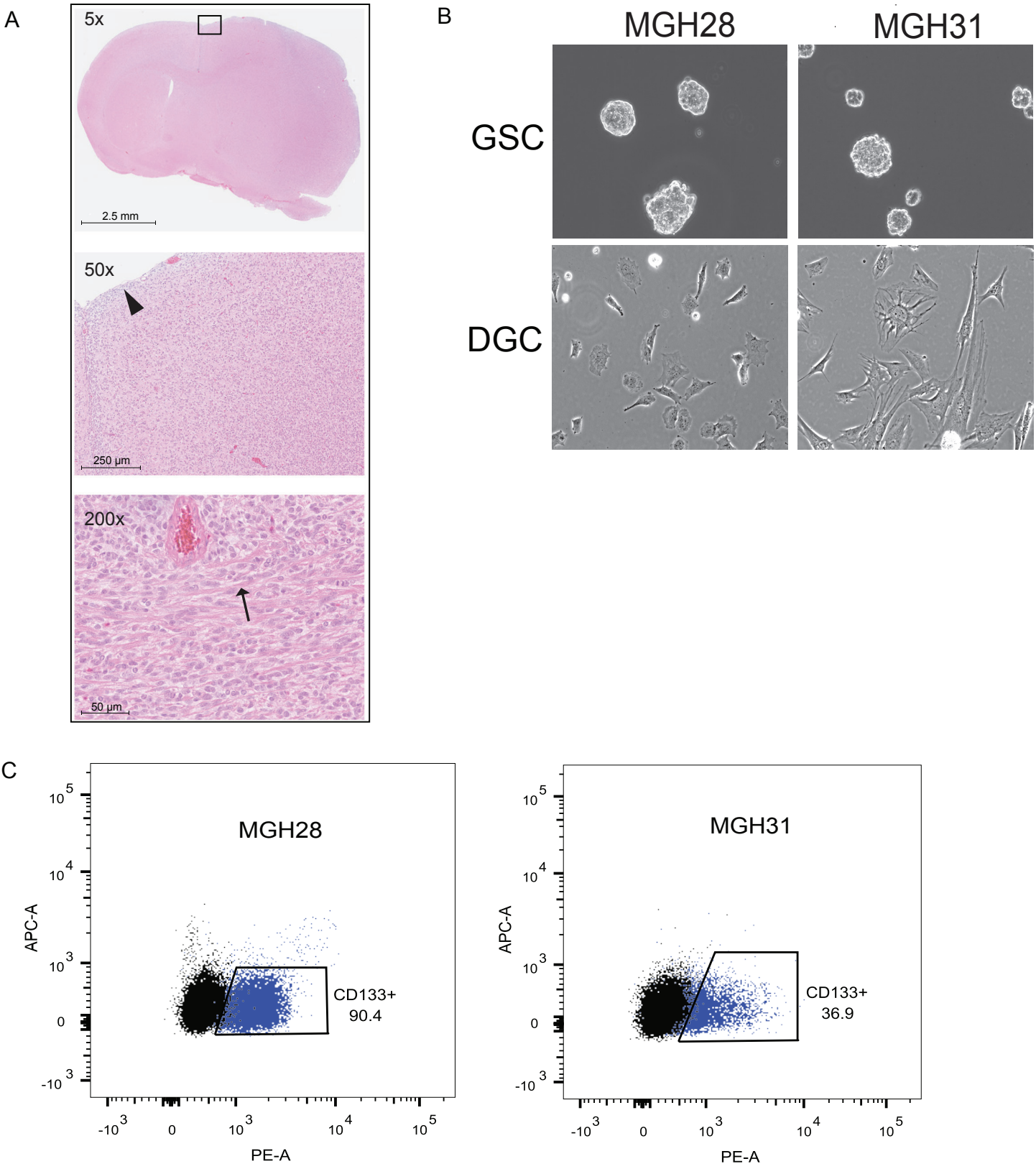
Supplemental Figure 15: Ki67 quantification for tumors. (A) Immunohistochemistry staining of Ki67 positive cells (brown) from representative fields of each tumor. Quantifications were performed by counting at least 1000 nuclei per tumor. **(B)** Plot compares percentage of cells predicted to be cycling,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



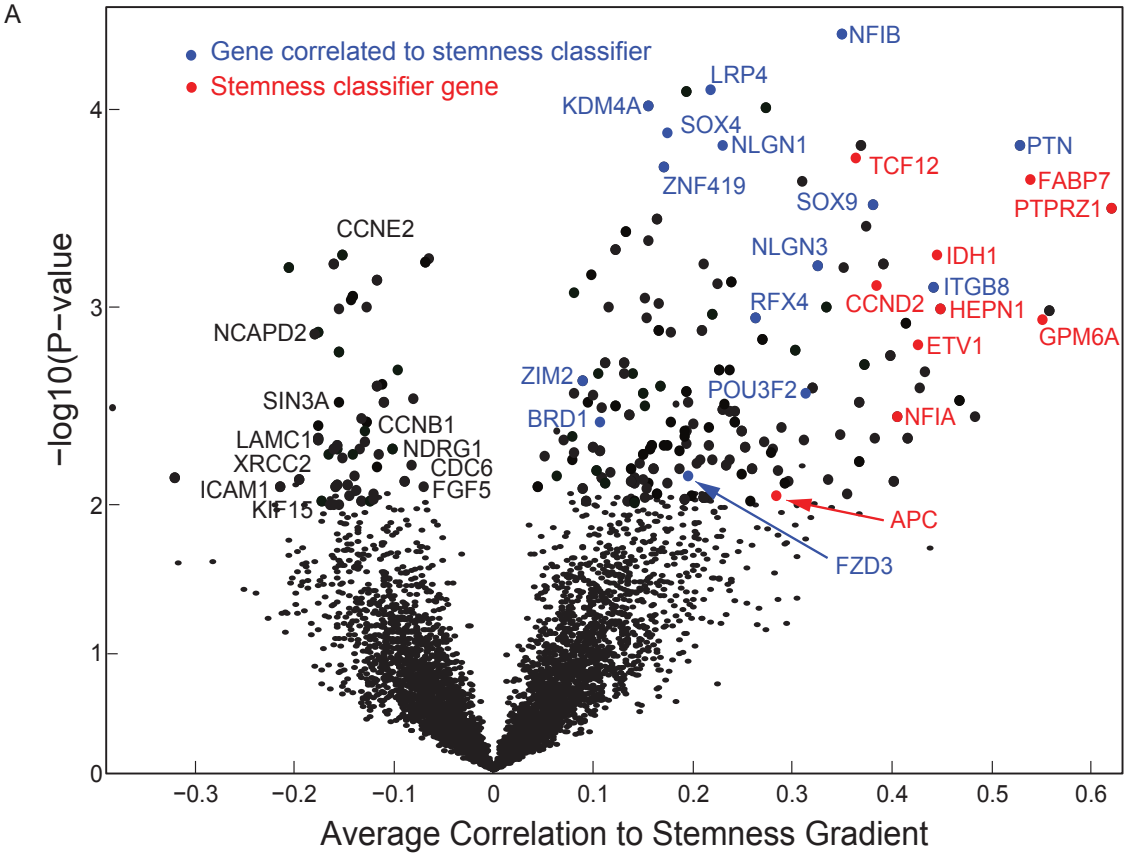
Supplemental Figure 16: Identification of genes anticorrelated to the cell cycle meta-signature. (A) Volcano plot of genes that are preferentially upregulated in non-cycling cells. Pearson correlations were calculated between the expression of each gene and the cell cycle meta-signature within each of 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.



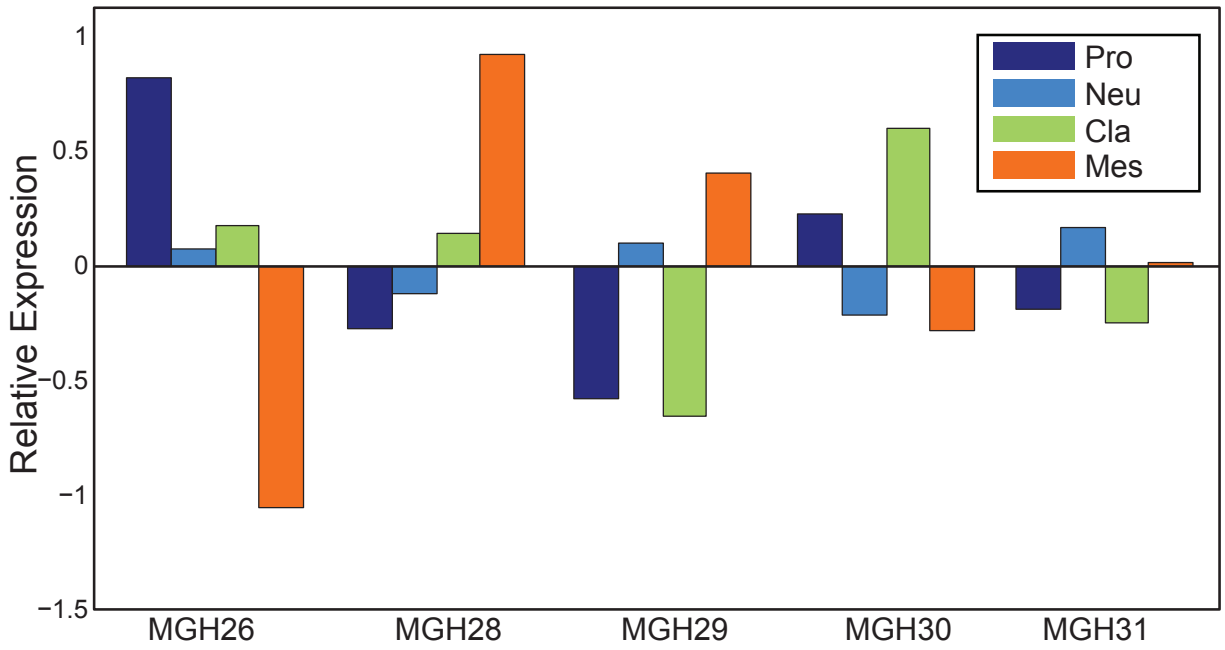
Supplemental Figure 17: Relationship between cell cycle and hypoxia meta-signatures. Clustering of genes anticorrelated to the cell cycle meta-signature (rows, columns: genes; color bar indicates Pearson correlation coefficient). A sub-cluster (black box) consists of genes from the hypoxia meta-signature. (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.



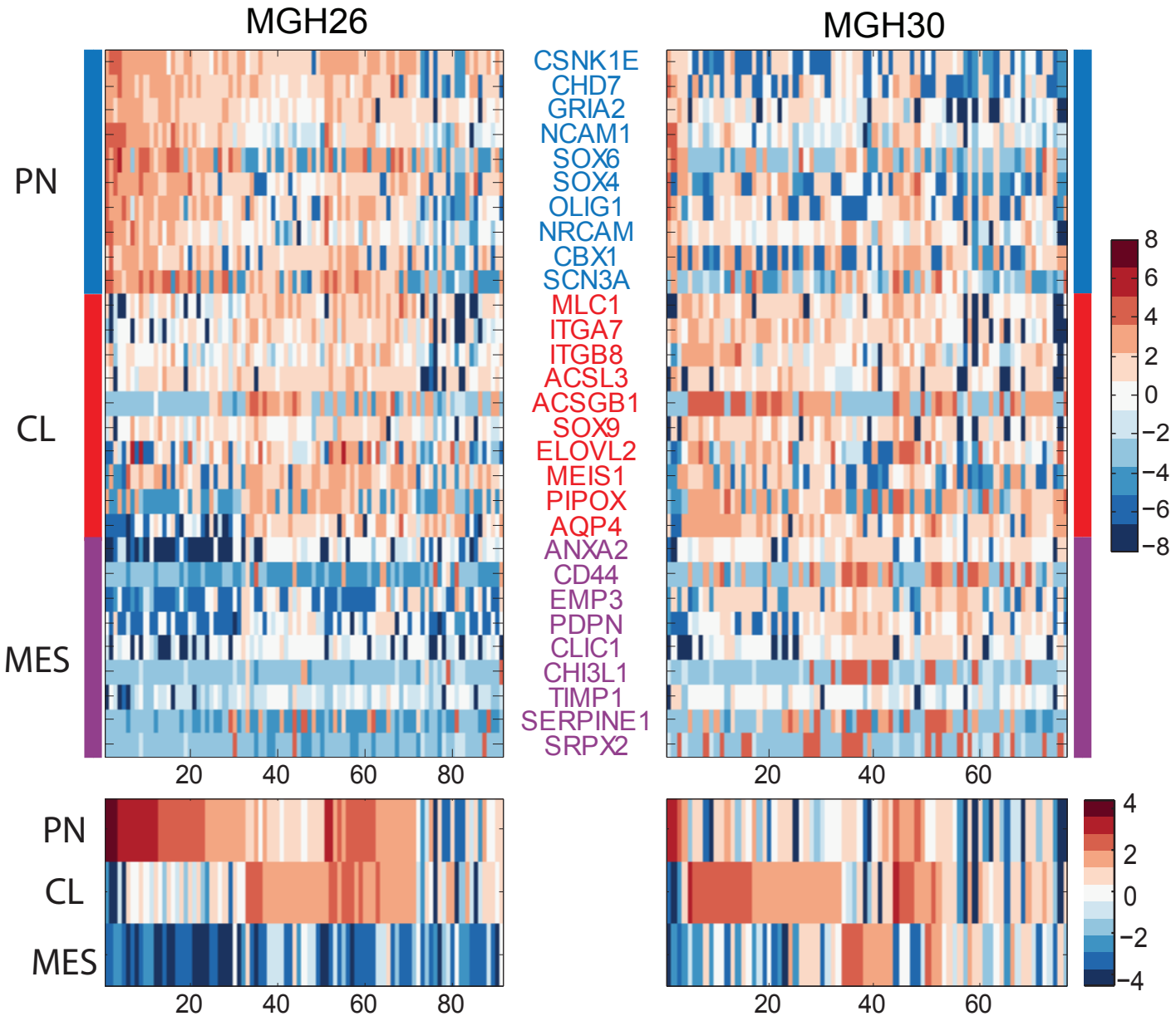
Supplemental Figure 18: *In vitro* tumor models. (A) Mouse xenografts of MGH26 GSCs demonstrate key features of diffuse infiltrative glioblastoma. Low magnification (5x) coronal section shows enlargement of the right hemisphere and compression of ipsilateral ventricle. Boxed area is magnified below at 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.



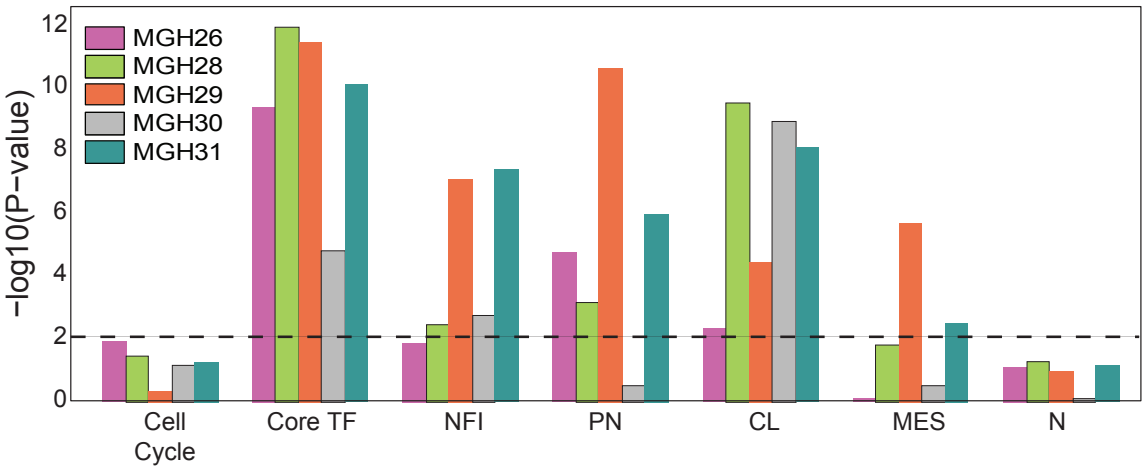
Supplemental Figure 19: Stemness associated genes *in vivo*. Volcano plot showing for each gene its average correlation to the stemness score among the five tumor (*x*-axis) and the –log10(P-value), reflecting the significance of the deviation of the average of these scores from zero, as defined by a 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*.



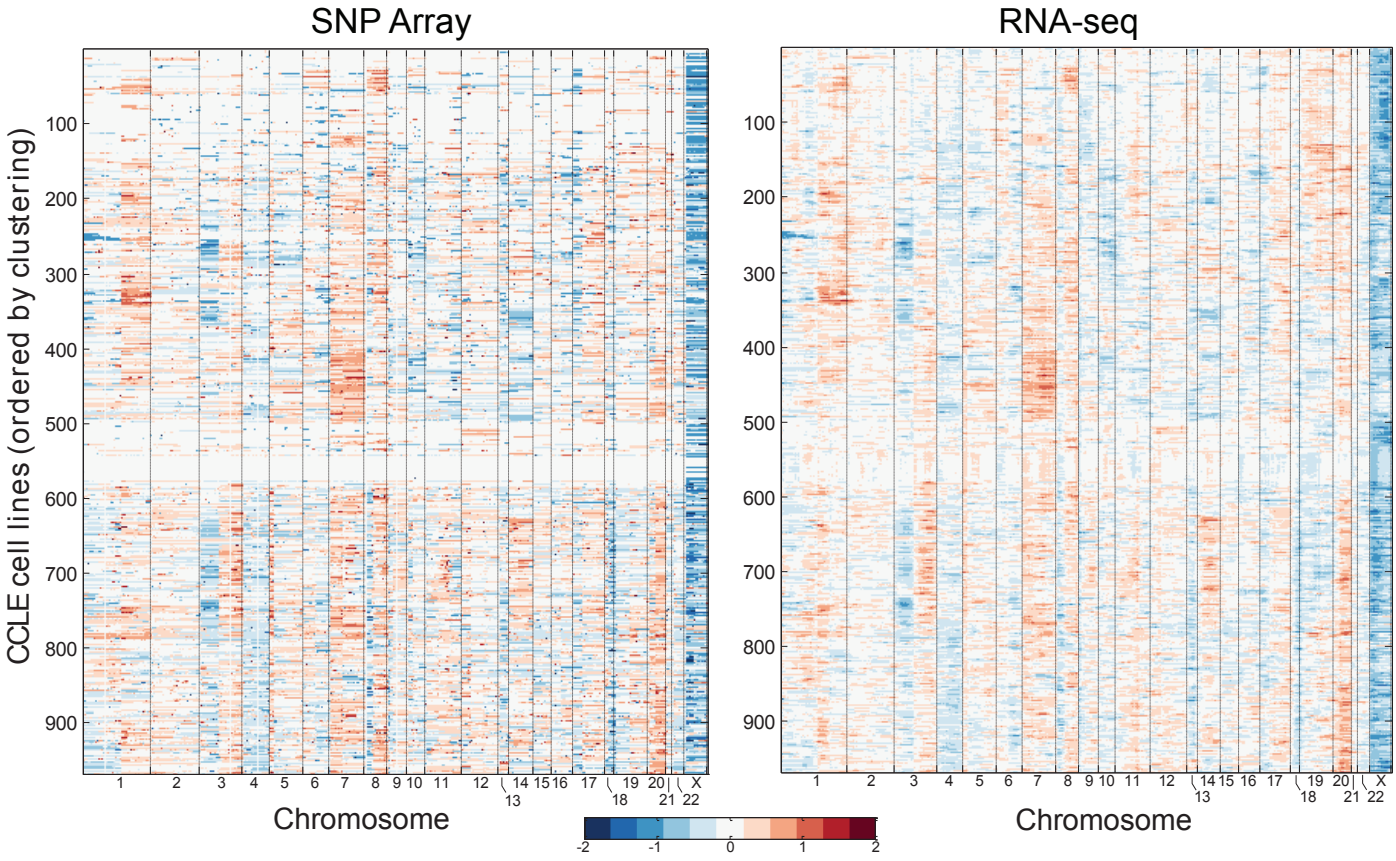
Supplemental Figure 20: Subtype classification of primary tumors based on population averaged data. Plot shows for each tumor (*x*-axis) the relative expression of each of four tumor sub-type signatures (bars; color legend). MGH26 is dominantly proneural, MGH28 and 29 are mesenchymal and MGH30 is classical. 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.