

Glioblastoma states are defined by cohabitating cellular populations with progression-, imaging- and sex-distinct patterns

Kamila M. Bond^{1,2}, Lee Curtin¹, Andrea Hawkins-Daarud¹, Javier C. Urcuyo¹, Gustavo De Leon¹, Christopher Sereduk³, Kyle W. Singleton¹, Jazlynn M. Langworthy¹, Pamela R. Jackson¹, Chandan Krishna⁴, Richard S. Zimmerman⁴, Devi Prasad Patra⁴, Bernard R. Bendok⁴, Kris Smith⁵, Peter Nakaji⁵, Kliment Donev⁶, Leslie Baxter^{7,8}, Maciej M. Mrugała⁹, Osama Al-Dalahmah^{10,11}, Leland S. Hu⁸, Nhan L. Tran³, Joshua B. Rubin^{12,13}, Peter Canoll^{10,11}, Kristin R. Swanson¹

1. Mathematical NeuroOncology Lab, Precision Neurotherapeutics Innovation Program, Mayo Clinic, Phoenix, AZ
2. Alix School of Medicine, Mayo Clinic, Rochester, MN
3. Department of Cancer Biology, Mayo Clinic, Phoenix, AZ
4. Department of Neurosurgery, Mayo Clinic, Phoenix, AZ
5. Department of Neurosurgery, Barrow Neurological Institute, Phoenix, AZ
6. Department of Pathology, Mayo Clinic, Phoenix, AZ
7. Department of Psychiatry & Psychology, Mayo Clinic, Phoenix, AZ
8. Department of Radiology, Mayo Clinic, Phoenix, AZ
9. Department of Neurology, Mayo Clinic, Phoenix, AZ
10. Department of Pathology and Cell Biology, Columbia University Vagelos College of Physicians and Surgeons, New York, NY
11. Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY
12. Department of Pediatrics, Washington University School of Medicine, St Louis, MO
13. Department of Neuroscience, Washington University School of Medicine, St Louis, MO

Abstract

Glioblastomas (GBMs) are biologically heterogeneous within and between patients. Many previous attempts to characterize this heterogeneity have classified tumors according to their omics similarities. These discrete classifications have predominantly focused on characterizing malignant cells, neglecting the immune and other cell populations that are known to be present. We leverage a manifold learning algorithm to define a low-dimensional transcriptional continuum along which heterogeneous GBM samples organize. This reveals three polarized states: invasive, immune/inflammatory, and proliferative. The location of each sample along this continuum correlates with the abundance of eighteen malignant, immune, and other cell populations. We connect these cell abundances with magnetic resonance imaging and find that the relationship between contrast enhancement and tumor composition varies with patient sex and treatment status. These findings suggest that GBM transcriptional biology is a predictably constrained continuum that contains a limited spectrum of viable cell cohabitation ecologies. Since the relationships between this ecological continuum and imaging vary with patient sex and tumor treatment status, studies that integrate imaging features with tumor biology should incorporate these variables in their design.

Introduction

Glioblastoma (GBM) is the most common primary brain tumor in adults and is known to be genetically, transcriptionally, and compositionally heterogeneous within and between patients (1–3). Treatment that effectively treats one tumor may be ineffective for another, and even cells within the same sample can have different therapeutic sensitivities (3). Since patients typically undergo a single surgery, at which time malignant cells are invariably left behind due to the diffuse nature of the disease, the ability to use routine magnetic resonance imaging (MRI) to non-invasively and longitudinally characterize tumor heterogeneity is invaluable.

Several studies have sought to characterize the heterogeneity of GBM using bulk and single-cell transcriptional profiling (4–9). In general, these studies have discretized GBM into biological groups and have focused on characterizing tumor cell populations rather than non-tumor cells. These simplified models of high-dimensional data, while easily interpretable, may not fully capture the biological reality of GBM. Mounting evidence suggests that GBM diffusely infiltrates the brain and cross-talks with normal cells (10–12). These pioneering tumor cells engage their environment, interact with the immune system (which can be co-opted by the tumor to facilitate its growth and invasion), and trigger other reactive responses (e.g., reactive astrocytes, expansion of progenitor cells). Under this conceptual framework, the biology of a tumor sample is a direct reflection of its cellular constituents and the environmental variables that influence them.

MRI is the mainstay of treatment assessment for GBM, but we do not fully understand how signal intensities from these images reflect the underlying biology. Under the prevailing gold standard for the radiographic evaluation of GBM, there are four categories of tumor responsiveness: complete response, partial response, stable disease, and progressive disease (13). Assignment to these categories is determined by the changes in the contrast-enhancing (CE) and non-enhancing (NE) tumor volumes in light of patients' steroid usage and clinical status. This paradigm is fundamentally limited by the non-specificity of contrast enhancement on T1-weighted MRI and T2/FLAIR MRI hyperintensity. This is evident with the use of anti-angiogenic agents (i.e., bevacizumab), where an acute reduction in

capillary permeability leads to a decrease in intratumoral contrast enhancement without a change in tumor burden (14). Conversely, about a third of patients demonstrate an increase in T1 contrast enhancement and progressive T2/FLAIR abnormality upon the initiation of temozolomide and radiation in a phenomenon called pseudoprogression (14, 15). It is assumed that T2/FLAIR hyperintensity represents infiltrative tumor, but it could represent several unrelated entities including vasogenic edema, microvascular ischemia, and leukoencephalopathies.

In this study, we leverage a manifold learning algorithm to discern the continuous nature of GBM's transcriptomes and the cellular compositions associated with them. We connect the resultant spectrum of tumor ecologies with MRI features in light of covariates like patient sex and tumor treatment status. In doing so, we assess the ability to infer tumor biology from routine imaging.

Methods

Patient populations and tumor biopsies

525 patients (205F, 320M) with GBM underwent tissue characterization as a part of The Cancer Genome Atlas (TCGA) effort. Microarray data from these samples were accessed using UCSC Xena (16). After obtaining informed consent, we enrolled 44 patients (18F, 26M) with primary or recurrent GBM in our IRB-approved study. Patients underwent surgical tumor resection with intraoperative navigation, and a total of 157 image-localized biopsies were harvested. The tissue was flash-frozen, and the Illumina TruSeq v2 RNAseq kit was used to prepare sequencing libraries. An Illumina HiSeq 4000 sequencer obtained paired-end reads, and FASTQ files were aligned to a reference genome (GRCh38.p37). Read counts were compiled using htseq-count, and batch effects were corrected with ComBat-Seq (17).

MRI acquisition and image localization

Pre- and post-contrast T1-weighted (T1W), pre-contrast T2-weighted (T2W), and fluid-attenuated inversion recovery (FLAIR) sequences were acquired at 3T field strength. Acquisition

parameters were as follows: T1W (TR/TE=600/12ms, matrix=320x240, FOV=24cm, thickness=2mm), T2W (TR/TE=4000/76ms, matrix=384x384; FOV=22cm; thickness=3mm), and FLAIR (TI/TR/TE=2600/10030/135ms, matrix=320x320, FOV=24cm, thickness=2mm). Post-contrast T1W images (T1Gd) were acquired following administration of a gadolinium-based contrast agent. Using previously-described methods, MRI coordinates for image-localized biopsies were captured and co-registered with multimodal preoperative imaging (18). Biopsy samples from regions that were T1Gd hyperintense were labeled contrast-enhancing (CE), and those from regions that were T1Gd isointense and T2/FLAIR hyperintense were considered non-enhancing (NE).

Deconvolving cell populations from RNA sequencing

While many have employed single-cell RNA sequencing as a means of characterizing intratumoral heterogeneity, this approach notoriously undersamples non-neoplastic populations (e.g., neurons) (19). Single nucleus RNAseq (snRNAseq) circumvents this limitation by more comprehensively sampling malignant, immune, and other cells. We leveraged a snRNAseq data set (20), derived from seven patients with GBM, which contained eighteen cell types: six malignant, five immune, and seven “other.” Malignant cell populations included mesenchymal/immunoreactive (gl_Mes1 and gl_Mes2), proliferative (gl_Pro1 and gl_Pro2), and neural/oligodendrocyte precursors (gl_PN1 and gl_PN2). Immune cells were characterized as T-cells, myeloid cells present at baseline (Myel1), proliferative tumor-associated macrophages (prTAM), monocyte-derived cells (moTAM), and microglia-derived cells (mgTAM). Other populations were characterized as neurons, endothelial cells, oligodendrocytes, oligodendrocyte precursor cells (OPCs), and three subtypes of astrocytes (Ast1, protoplasmic; Ast2 and Ast3, reactive).

A bulk RNAseq transcriptome reflects the sum of transcriptional contributions from individual cells. Therefore, the cellular composition of heterogeneous samples can be estimated using deconvolution methods like CIBERSORTx (21). This method relies on marker gene input and assumes

that expression signals are linearly additive, such that marker gene expression is related to the proportion of that cell type in the sample. Each cell in our snRNAseq matrix was labeled as one of eighteen cell populations (20). This served as the input to CIBERSORTx, where a signature matrix was derived to define the unique features of each cell type. This signature was applied to the bulk RNAseq samples to estimate the abundance of all eighteen cell types in our image-localized samples. The analysis was performed in absolute mode with 500 permutations and quantile mode disabled. S mode batch correction was applied to account for the application of a signature derived from UMI-based snRNAseq to bulk RNAseq data. We present CIBERSORTx output as absolute and relative (normalized to the total of its category - glioma, immune, or other) values.

Trajectory inference and pseudotime ordering of TCGA microarray and image-localized RNAseq samples

The objective of trajectory inference is to reduce high-complexity, asynchronous data into an ordered, one-dimensional path. While it was initially described in the context of ordering single-cell RNAseq data to infer the biological progression of cell lineages (22), it has been successfully repurposed to determine the relative ordering of other types of complex data (23). Monocle initially reduces data dimensionality using independent component analysis (ICA) (24). The data is represented by a series of nodes (cells) and edges (weighted by distance between samples after ICA). A manifold learning algorithm (DDRTree) uses reverse graph embedding to identify potential backbones of the trajectory and orders samples along the longest path. Each sample is assigned to a branch on the tree, and pseudotime is calculated using the geodesic distance to the path's starting point. Here we used TCGA microarray and our biopsy RNAseq as inputs to visualize GBM's natural transcriptional organization. States were defined as each of the three concurrent line segments. The "fgsea" package was used to perform gene set enrichment analysis, with the MSigDB database as the reference (25, 26).

Statistical analyses

The TCGA cohort contains one sample per patient and thus each sample can be considered independent. Chi-square tests were used to determine the significance of differences between groups. Since the image-localized biopsy cohort contains multiple samples per patient, linear mixed-effects models coupled with ANOVA tests (Type II Wald tests) were used to determine the significance of differences in this cohort, with cumulative plots and box plots employed for data visualization (27). For categorical outcome variables, groups were binarized. A dummy variable of patient number was included as a random effect in each test within the lmer and glmer functions in the “lme4” R package (28, 29) for continuous and categorical dependent variables, respectively. The “car” package in R was used for ANOVA tests (30). The “rmcorr” package in R was used to determine the significance of correlations between CIBERSORTx populations alongside the Bonferroni method to adjust for multiple comparisons (31).

Results

Trajectory inference of GBM samples reveals three polarized states

To discern GBM's transcriptional continuum, a manifold learning algorithm inferred the biological trajectory of TCGA GBM microarray samples (n=525). This approach reduces data dimensionality, computes the minimum spanning tree trajectory, and orders samples along the longest path (Figure 1A). Pseudotime, the geodesic distance of a point from the trajectory's origin (pseudotime=0), reveals an initiating trunk that bifurcates into two paths (Figure 1B). There is no sex difference in pseudotime distribution (p=0.19; Figure 1C). The three concurrent line segments of the trajectory are labeled as “states” A, B, and C (Figure 1D). Each state has a unique transcriptional profile of the top 1000 most highly variable genes (Figure 1E). Pre-ranked GSEA, using known transcriptional subtype signatures as a reference, reveals that state A is concordant with neural and proneural signatures, state B is more mesenchymal, and state C agrees with the classical signature (Figure 1F). TCGA sample classification overlaid on the trajectory confirms that each state is associated with a transcriptional subtype ($X^2=239$,

$p < 0.0001$; Figure 1G). Relative distribution of age, IDH1 mutation status, and MGMT methylated status are shown in Figure S1A-D.

Bulk transcriptomic sequencing of 134 image-localized biopsies (median=3 per patient) from 39 patients with primary ($n=30$) or recurrent ($n=9$) GBM underwent trajectory inference. Patient demographics and biopsy characteristics are summarized in Table S1. The inferred biological trajectory of these samples again reveals three concurrent segments that recapitulate those seen with TCGA (Figures 2A-B). In this cohort, there is a bias for female samples towards lower pseudotime values (55.2 vs. 71.7, $p=0.002$; Figure 2C). There is no difference in IDH status between the states ($X^2=1.19$, $p=0.55$; Figure S1E).

States reflect transitions in cellular population ecologies

To interpret how tumor composition contributes to this trajectory, each image-localized sample underwent RNA sequencing deconvolution to estimate the abundance of the individual cell populations (Figure S2A; Table S2). Eighteen cell populations, as defined by snRNAseq (20), were predicted: six malignant glioma, five immune (e.g., myeloid cells, T-cells), and seven “other” (e.g., astrocytes, neurons, endothelial cells). The sums of malignant cells, immune cells, and other cells are highest in states C, B, and A, respectively (Figure 2D). The trajectory as labeled by each cellular component can be found in Figure S2B.

When ordering sample composition heatmaps by pseudotime, we find that transitions between states correlate with changes in composition (Figure 2E). Ecological diversity measures (e.g. Shannon entropy and evenness) have lower values at the distal poles of each state’s segment compared to the transitional arms (Figure S2B). Taken together, these data suggest that GBM states are on a spectrum between three polarized states: invaded brain, immunoreactivity, and proliferation. Any sample can be a transitional admixture between these states by way of changes in their underlying cellular composition. As such, the abundance of certain cell populations tends to correlate with one another.

Sex-specific cell abundance correlation networks are quantified in Figure 2F and overall cohabitation trends are schematized in Figure 2G.

Contrast-enhancing samples contain more malignant cells than non-enhancing samples

Since measures of contrast-enhancement (CE; from the T1Gd+ core) and non-enhancement (NE; from the T1Gd-/FLAIR+ penumbra) are central to surgical planning and the radiological assessment of treatment responsiveness, we were interested in the relationship between enhancement and tumor biology (i.e., pseudotime and cellular composition). State A samples tend to be NE, while samples from states B and C tend to be CE ($p=0.00034$; Figure 3A). Tumor status (i.e., primary and recurrent) has no significant relationship to states ($X^2=0.39$, $p=0.82$; Figure 3A). Overall cellular composition varies widely within and between CE and NE samples but is more predictable within states (Figure 3B).

The absolute abundance of malignant cells is higher in samples harvested from CE than NE regions for each sex individually ($p=0.0053$ males, $p=0.021$ females; Figure 3C). This remains true when considering all data together ($p=0.00025$; Figure S3A) or separating biopsies collected from primary ($p=0.037$; Figure S3B) and recurrent tumors ($p<0.0001$; Figure S3C). This finding was also seen within males for primary biopsies alone ($p=0.027$; Figure S3D), with an analogous but insignificant difference in the recurrent setting ($p=0.081$). Within females, the result did not hold for primary samples only ($p=0.65$) but did for recurrent samples only ($p<0.0001$; Figure S3E). Amongst NE samples, there are fewer malignant cells in recurrent compared to primary biopsies ($p=0.0045$; Figure S3F).

To determine if specific populations are driving the relationship between contrast enhancement and malignant cell abundance, each subtype was considered individually. There was an overrepresentation of gl_Mes2 cells in CE compared to NE biopsies ($p=0.0002$; Figure S3G). Notably, an increased abundance of gl_Pro1 cells in all CE samples ($p=0.0028$; Figure S3G) was driven by female patients ($p=0.00059$; Figure 3D). Conversely, NE biopsies have more gl_PN1 cells than their CE counterparts ($p<0.0001$; Figure S3G), which was also the case within male ($p=0.026$; Figure 3D).

and female samples ($p=0.00010$; Figure 3D). There were no significant regional differences in gl_Pro2 across all samples or by sex.

Some immune cells have sex-specific imaging correlates

Although GBM is known as a “cold” tumor with little lymphocytic infiltration, tumor-associated macrophages and other immune cells are abundant and can account for up to half of a sample (32, 33). Thus, we were interested in the ability of contrast enhancement to discern the presence of immune cells. We observed a difference in the total abundance of immune cells between NE and CE imaging regions for females but not males (Figure 3E).

We next considered each population individually under the assumption that the imaging representations of immune activity could be sex-dependent. Contrast enhancement had no significant relationship to Myel1 cells, T-cells, or prTAMs. In females, mgTAMs were more likely to be present in NE samples ($p=0.0019$; Figure 3F), with significance retained only in the recurrent subcohort ($p=0.00058$; Figure S4A). In both sexes combined, moTAMs were significantly more abundant in CE than NE regions ($p=0.00063$; Figure S4B), a relationship that was retained only in primary samples ($p=0.0078$; Figure S4C).

MRI contrast enhancement corresponds with the abundance of some normal brain cell populations

NE tumor regions are expected to have fewer malignant cells and more normal and/or reactive CNS populations (34). We observe more neurons in NE samples ($p<0.0001$; Figure S5A), the significance of which is retained in both male and female samples ($p=0.00035$ and $p=0.0046$, respectively; Figure 3H), and primary and recurrent settings ($p=0.00092$, Figure S5B; $p=0.00024$, Figure S5C, respectively). Ast1 prevalence in NE ($p<0.0001$; Figure S5A) is not recapitulated in male primary GBM biopsies ($p=0.36$) but is driven by females in the primary setting and males in the recurrent setting ($p=0.019$, Figure S5D; $p=0.014$, Figure S5E, respectively). Endothelial cells, which line the neovasculature from which contrast extravasates into tissue (35), were more abundant in CE compared

to NE samples ($p < 0.0001$; Figure S5A). This significance was retained in primary ($p = 0.00011$; Figure S5B) and recurrent samples ($p = 0.018$; Figure S5B). There were no cohort-wide significant relationships between imaging and Ast3, oligodendrocytes, or OPCs.

Overall tumor composition has little bearing on MRI contrast enhancement

To reflect on the connections between contrast enhancement, pseudotime state, and cellular composition, we share a case example. A 59-year-old female with primary IDH wild-type, MGMT methylated GBM underwent seven image-localized biopsies (Figure 4A; 5CE, 2NE). The biopsies are distributed across the pseudotime trajectory and all three states (Figure 4B). Inter-sample correlations of cellular composition, as enumerated by CIBERSORTx, revealed a spectrum of biological heterogeneity between the samples (Figure 4C). Biopsies with similar imaging appearance (i.e., both NE or both CE) were associated with a spectrum of cellular composition similarity (Figure 4D). Samples 2 and 3 have nearly identical cell compositions ($R = 0.95$), but one is CE and one is NE. Conversely, samples 4 and 5 are compositionally dissimilar ($R = 0.07$) and belong to different pseudotime states (B and C), yet they are both CE.

Discussion

An unsupervised trajectory inference algorithm revealed that the GBM transcriptome is a continuum of three polarized states: invasive, immune/inflammatory, and proliferative. The states are conserved regardless of patient sex and tumor treatment status. The trajectory's constrained nature may have implications for treatment, as transcriptional dynamics should be predictably modifiable as a function of therapy and one should theoretically be able to artificially transmute the tumor into a more favorable (less aggressive) state. This emboldens the potential for adaptive therapy schemas (36) which are increasingly focused away from cytotoxic treatments and towards cytostatic population control approaches.

We next employed a cellular deconvolution strategy to predict the abundance of eighteen cell types amongst these states. Previous efforts to characterize GBM's cellular composition in totality have been impeded by the limitations of single-cell RNA sequencing (e.g., under-representation of cell populations) (19). Recent methodological advancements, (i.e., single nucleus RNA sequencing) circumvent these limitations and allow for a more complete characterization of all populations including malignant, immune, and other cells. We leveraged a unique cohort of GBM snRNAseq as training data for a cellular deconvolution algorithm (20). This paradigm allowed us to quantify all GBM cell subpopulations in a large number of bulk RNA sequencing samples, which has never been done before. In connecting cell abundances to the three tumor states, we observe that transitions between states directly correlate with changes in population ecologies. Many prior efforts to classify transcriptomics data have underappreciated the continuity of these low-dimensional spaces.

The widespread clinical utility of these results hinges on the ability to integrate them with routine clinical imaging. Several factors have historically hindered our ability to understand the relationships between tumor biology and MRI. First, since GBM has such profound intratumoral heterogeneity, biological data from tissue biopsies and corresponding imaging features cannot be aligned without knowing the exact location from which the sample was harvested. Since image-localized biopsy collection is a resource-intensive process, the vast majority of GBM samples are non-localized. Second, GBM has known sex differences in incidence, treatment responsiveness, and prognosis (37, 38). For this reason, its biological states, and the imaging representations associated with them, cannot be assumed to be the same for males and females. However, across all of biomedicine, few studies have considered sex differences in their design (39). Despite our lack of biological understanding of what imaging means due to these limitations, serial MRIs are paramount in the evaluation of GBM and many assumptions are made in their interpretation.

To address these limitations, we leveraged image-localized biopsies to directly connect MRI features with transcriptional states and cellular composition. A sample's biological state is generally associated with measures of contrast enhancement or non-enhancement. Samples from regions of

contrast enhancement had higher abundances of malignant cell populations, whereas non-enhancing samples generally contained more normal brain cells. These findings in isolation support the traditional notion that CE and NE represent proliferative tumor and invaded brain, respectively. However, an assessment of cellular composition as it relates to imaging and pseudotime revealed a more nuanced story.

Patterns of cellular cohabitation could most simply be described as a function of the pseudotime trajectory. However, these ecological patterns, as well as their relationships to CE and NE, varied based on patient sex and tumor treatment status. Notably, the polarization of the malignant gl_Pro1 population towards CE regions and gl_PN1 population towards NE regions was driven primarily by female samples. Further, enrichment for the Myel1 population was spatially skewed towards NE for males and CE for females. Taken together, these findings support the potential for sex-distinct imaging patterns of tumor ecologies.

Of particular clinical relevance, we presented a case example of a patient who underwent seven biopsies from both CE and NE regions (Figure 4). The cellular composition of these samples had very little bearing on contrast enhancement status. This result highlights the limitations of current methods of treatment assessment, where a tumor is considered to be responsive if the CE volume regresses. The inability of contrast enhancement alone to fully predict cellular composition motivates a role for advanced imaging strategies like radiomics. Radiomics is a quantitative approach to predict biological attributes from complex imaging features (40, 41). Early radiomics efforts focused on connecting an entire image to a biological prediction, thus failing to capture the spatial heterogeneity within individual tumors. More recently, our group and others have leveraged image-localized biopsies to directly connect spatially-resolved imaging features with tissue biology (42–48). The differences in bioimaging relationships seen in the primary and recurrent setting as well as between the sexes embolden the careful incorporation of these variables in the context of radiomics. Elucidating these bioimaging relationships will provide the opportunity to non-invasively characterize tumors in their entirety and personalize each patient's therapy to exploit the unique weaknesses of their tumor's dominant state.

Acknowledgments

We are grateful to all of those who have contributed to elements of this work, particularly current and past members of the image analysis teams and the glioma biopsy protocol teams, including: Barrett Anderies, Jessica Bauer, Spencer Bayless, Hend Bcharach, Regina Becker, Sameer Channer, Brenden Doyle, Lysette Elsner, Lily Esaleh, Ashlyn Gonzales, Crystal Harris, Morgan Hatlestead, Ryan Hess, Sandra Johnston, Yvette Lassiter-Morris, Julia Lorence, Ashley Napier, Ashley Nespodzany, Lisa Paulson, Cassandra Rickertsen, Sejal Shanbhag, Sarah Van Dijk, and Scott Whitmire.

References

1. Q. T. Ostrom, G. Cioffi, K. Waite, C. Kruchko, J. S. Barnholtz-Sloan, CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. *Neuro. Oncol.* **23**, iii1–iii105 (2021).
2. N. R. Parker, P. Khong, J. F. Parkinson, V. M. Howell, H. R. Wheeler, Molecular heterogeneity in glioblastoma: potential clinical implications. *Front. Oncol.* **5**, 55 (2015).
3. A. P. Patel, *et al.*, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
4. C. Neftel, *et al.*, An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835–849.e21 (2019).
5. M. Castellan, *et al.*, Single-cell analyses reveal YAP/TAZ as regulators of stemness and cell plasticity in Glioblastoma. *Nat Cancer* **2**, 174–188 (2021).
6. L. M. Richards, *et al.*, Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* **2**, 157–173 (2021).
7. L. Garofano, *et al.*, Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat Cancer* **2**, 141–156 (2021).
8. R. G. W. Verhaak, *et al.*, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
9. Q. Wang, *et al.*, Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* **32**, 42–56.e6 (2017).
10. H. S. Venkatesh, *et al.*, Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539–545 (2019).
11. H. S. Venkatesh, *et al.*, Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion. *Cell* **161**, 803–816 (2015).
12. F. Sahm, *et al.*, Addressing diffuse glioma as a systemic brain disease with single-cell analysis. *Arch. Neurol.* **69**, 523–526 (2012).
13. P. Y. Wen, *et al.*, Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J. Clin. Oncol.* **28**, 1963–1972 (2010).
14. L. C. Hygino da Cruz Jr, I. Rodriguez, R. C. Domingues, E. L. Gasparetto, A. G. Sorensen, Pseudoprogression and pseudoresponse: imaging challenges in the assessment of posttreatment glioma. *AJNR Am. J. Neuroradiol.* **32**, 1978–1985 (2011).
15. A. W. Abbasi, *et al.*, Incidence of Tumour Progression and Pseudoprogression in High-Grade Gliomas: a Systematic Review and Meta-Analysis. *Clin. Neuroradiol.* **28**, 401–411 (2018).
16. M. J. Goldman, *et al.*, Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).
17. Y. Zhang, G. Parmigiani, W. E. Johnson, : batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020).
18. P. D. Chang, *et al.*, A Multiparametric Model for Mapping Cellularity in Glioblastoma Using Radiographically Localized Biopsies. *AJNR Am. J. Neuroradiol.* **38**, 890–898 (2017).
19. P. Qiu, Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 1169 (2020).
20. O. Al-Dalahmah, *et al.*, Re-convolving the compositional landscape of primary and recurrent glioblastoma using single nucleus RNA sequencing. *bioRxiv* (2021) <https://doi.org/10.1101/2021.07.06.451295>.
21. C. B. Steen, C. L. Liu, A. A. Alizadeh, A. M. Newman, Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol. Biol.* **2117**, 135–157 (2020).
22. C. Trapnell, *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
23. S. Mukherjee, *et al.*, Molecular estimation of neurodegeneration pseudotime in older brains. *Nat. Commun.* **11**, 5781 (2020).

24. X. Qiu, *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
25. A. Liberzon, *et al.*, Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
26. G. Korotkevich, *et al.*, Fast gene set enrichment analysis. *bioRxiv* (2016) <https://doi.org/10.1101/060012>.
27. , ggplot2 (March 23, 2022).
28. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models Using lme4. *J. Stat. Softw.* **67** (2015).
29. , The R Project for Statistical Computing (March 23, 2022).
30. J. Fox, S. Weisberg, *An R Companion to Applied Regression* (SAGE Publications, 2018).
31. J. Z. Bakdash, L. R. Marusich, Repeated Measures Correlation. *Front. Psychol.* **8**, 456 (2017).
32. R. A. Morantz, G. W. Wood, M. Foster, M. Clark, K. Gollahon, Macrophages in experimental and human brain tumors. Part 2: studies of the macrophage content of human brain tumors. *J. Neurosurg.* **50**, 305–311 (1979).
33. M. Rahman, *et al.*, Adult immuno-oncology: using past failures to inform the future. *Neuro. Oncol.* **22**, 1249–1261 (2020).
34. B. J. Gill, *et al.*, MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12550–12555 (2014).
35. M. Oostendorp, M. J. Post, W. H. Backes, Vessel growth and function: depiction with contrast-enhanced MR imaging. *Radiology* **251**, 317–335 (2009).
36. R. A. Gatenby, A. S. Silva, R. J. Gillies, B. R. Frieden, Adaptive therapy. *Cancer Res.* **69**, 4894–4903 (2009).
37. W. Yang, *et al.*, Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data. *Sci. Transl. Med.* **11** (2019).
38. Q. T. Ostrom, J. B. Rubin, J. D. Lathia, M. E. Berens, J. S. Barnholtz-Sloan, Females have the survival advantage in glioblastoma. *Neuro. Oncol.* **20**, 576–577 (2018).
39. K. M. Bond, M. M. McCarthy, J. B. Rubin, K. R. Swanson, Molecular omics resources should require sex annotation: a call for action. *Nat. Methods* **18**, 585–588 (2021).
40. R. J. Gillies, P. E. Kinahan, H. Hricak, Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577 (2016).
41. G. Singh, *et al.*, Radiomics and radiogenomics in gliomas: a contemporary update. *Br. J. Cancer* **125**, 641–657 (2021).
42. L. S. Hu, *et al.*, Uncertainty quantification in the radiogenomics modeling of EGFR amplification in glioblastoma. *Sci. Rep.* **11**, 3932 (2021).
43. L. S. Hu, K. R. Swanson, Roadmap for the clinical integration of radiomics in neuro-oncology. *Neuro. Oncol.* **22**, 743–745 (2020).
44. L. S. Hu, A. Hawkins-Daarud, L. Wang, J. Li, K. R. Swanson, Imaging of intratumoral heterogeneity in high-grade glioma. *Cancer Lett.* **477**, 97–106 (2020).
45. N. Gaw, *et al.*, Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. *Sci. Rep.* **9**, 10063 (2019).
46. L. S. Hu, *et al.*, Accurate Patient-Specific Machine Learning Models of Glioblastoma Invasion Using Transfer Learning. *AJNR Am. J. Neuroradiol.* **40**, 418–425 (2019).
47. L. S. Hu, *et al.*, Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro. Oncol.* **19**, 128–137 (2017).
48. L. Wang, A. Hawkins-Daarud, K. R. Swanson, L. S. Hu, J. Li, Knowledge-infused global-local data fusion for spatial predictive modeling in precision medicine. *IEEE Trans. Autom. Sci. Eng.*, 1–13 (2021).

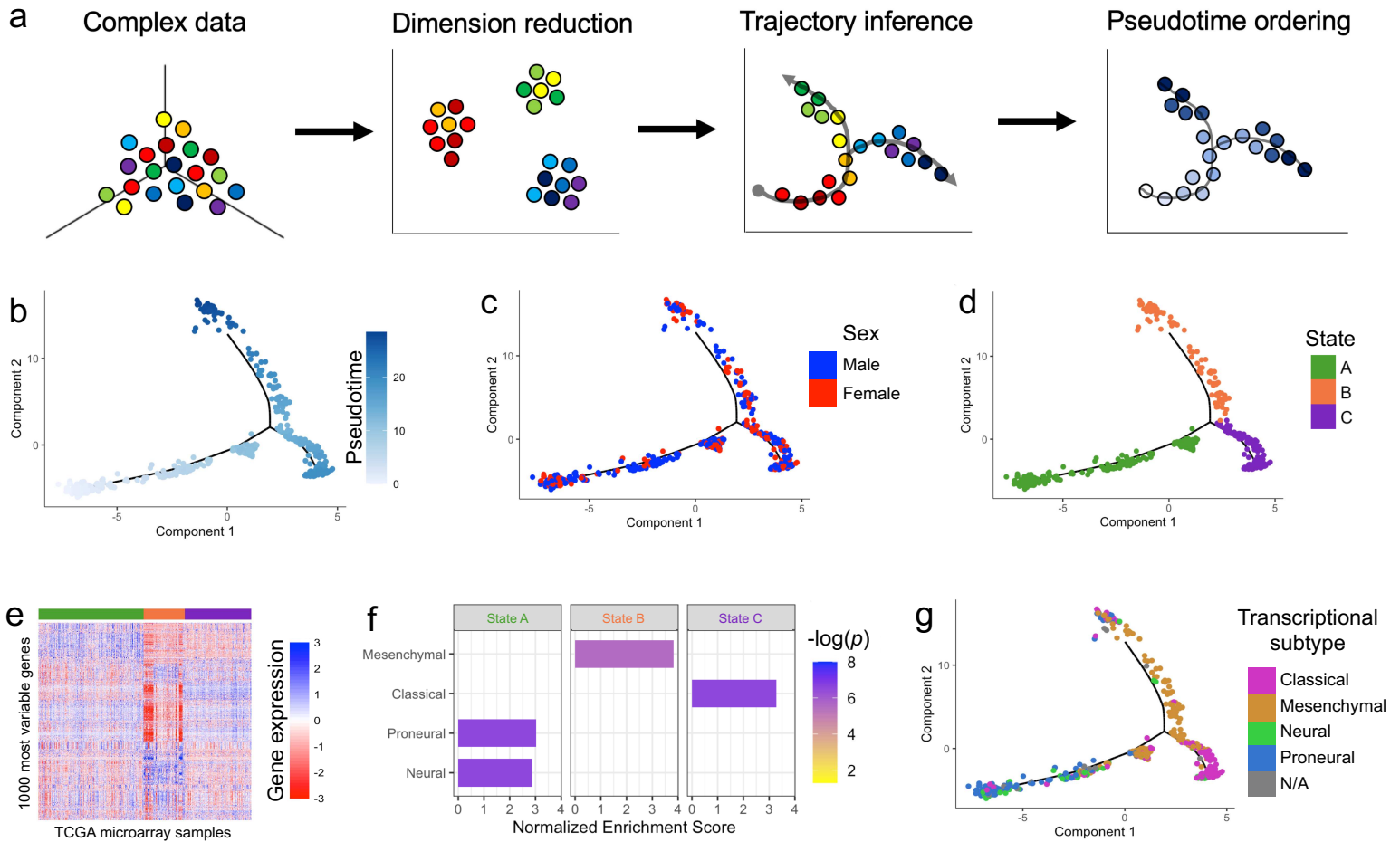


Figure 1. Trajectory inference and pseudotime ordering of TCGA GBM microarray samples reveal a continuum of three states. (A) An illustrative overview of the manifold learning approach to infer trajectories and order complex data along a progression. (B) When ordered by pseudotime, TCGA GBM microarray samples ($n=525$) fall along a continuum of three contiguous arms. (C) There is no sex bias in pseudotime distribution ($p=0.19$). (D) Each segment of the trajectory defines a “state.” (E) A heatmap of expression of the 1000 most highly variable genes reveals distinct transcriptomes for each state. (F) Pathway analysis quantifies the concordance of traditional transcriptional subtypes within each state. (G) Transcriptional subtype classification is related to state ($\chi^2=239$, $p<0.0001$).

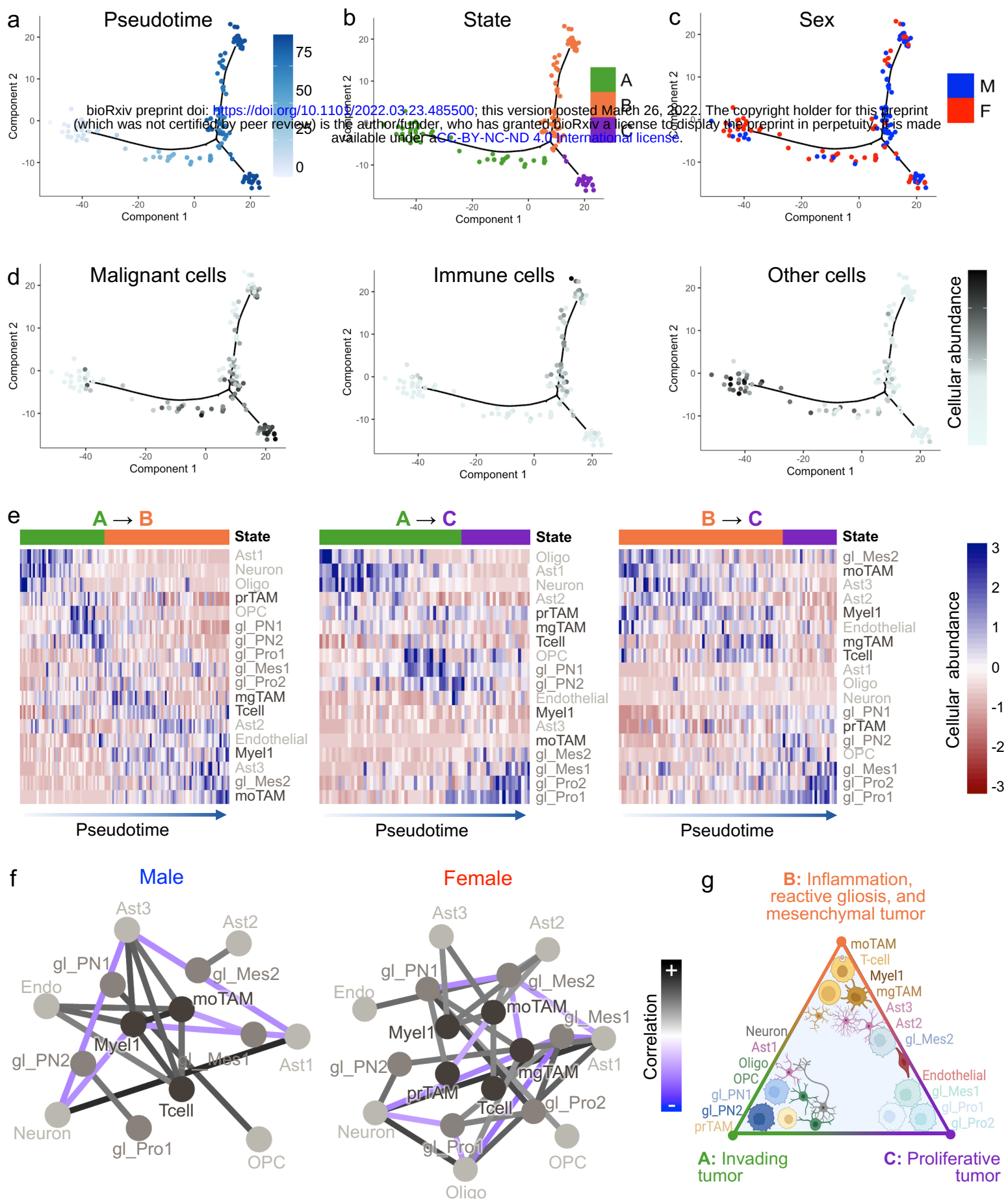


Figure 2. Trajectory inference of 134 image-localized GBM biopsies from 39 patients recapitulates the three states, each of which has distinct cellular cohabitation patterns. (A) Pseudotime trajectory and **(B)** states derived from biopsies mirror those that are derived from TCGA. **(C)** There is a sex bias in our samples such that F samples have lower pseudotime (55.2 vs. 71.7, $p=0.002$). **(D)** Deconvolution of bulk RNAseq allows for the estimation of cell population abundances. Malignant cells are present on all arms of the trajectory, with particular overrepresentation in state C. **(H)** Immune cells (i.e., moTAM, gmTAM, prTAM, Myel1, Tcell) are most abundant in state B. **(I)** Other cells (i.e., neurons, oligodendrocytes, OPCs, astrocytes, endothelial cells) are most populous in state A. **(E)** Estimates of all 18 malignant, immune, and other cell populations were z-scaled across all samples. Each state tends to be composed of distinct cellular constituents that transition over pseudotime. **(F)** Correlation networks between cell population abundances are sex-distinct. **(G)** Schematic of relative cohabitation tendencies of cell populations.

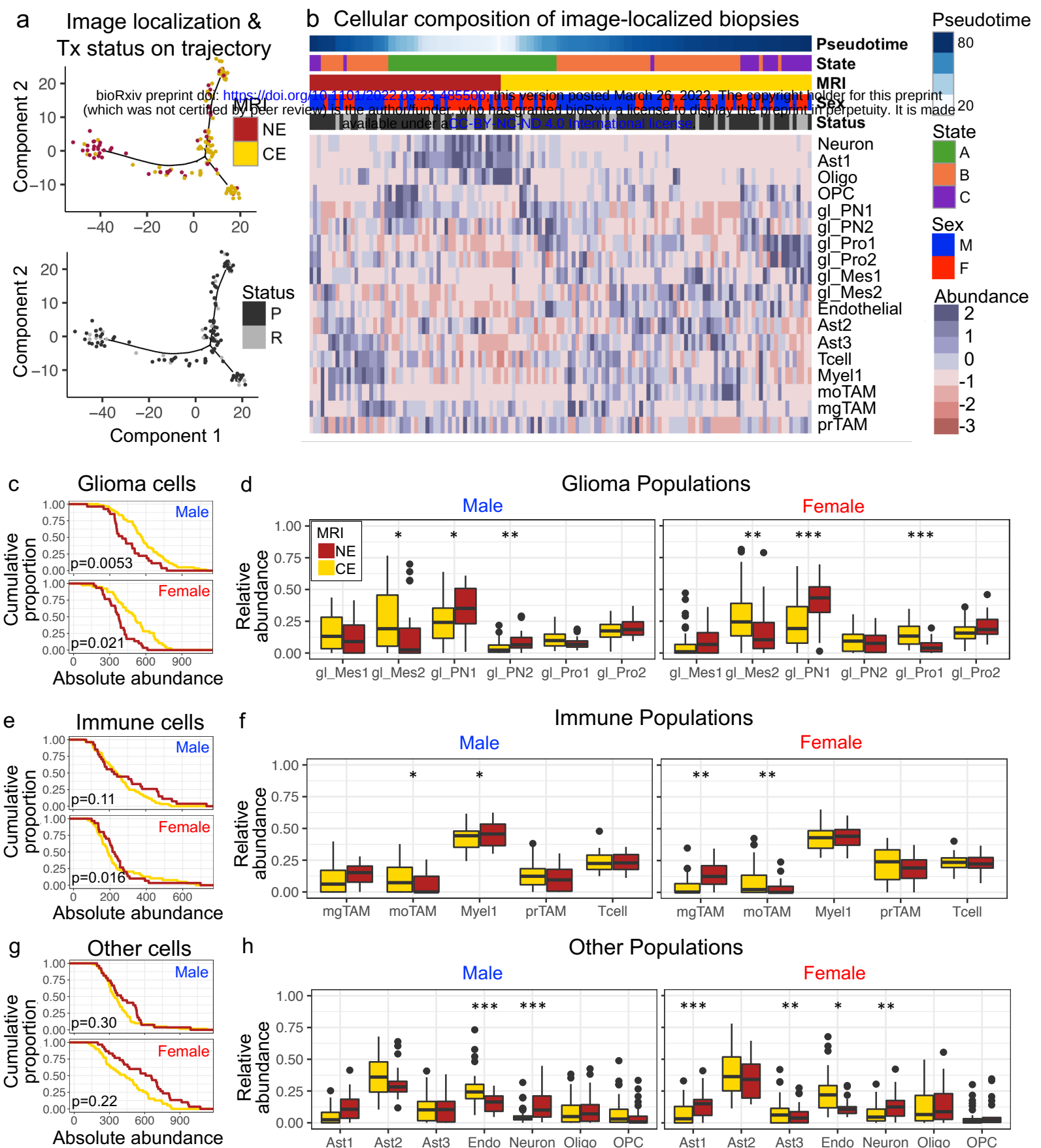


Figure 3. Pseudotime and imaging reveal complementary but distinct insights into tumor biology. (A) Biopsies from non-enhancing (NE) regions on MRI are more prevalent in state A, whereas contrast-enhancing (CE) biopsies are more common in states B and C ($p=0.00034$). Samples from primary (P) and recurrent (R) tumors are not significantly biased in their distribution ($p=0.75$). **(B)** Cellular composition of biopsy samples ordered by MRI localization (CE/NE) and pseudotime. **(C)** CE samples have more total glioma (malignant) cells than NE samples. **(D)** Glioma cell population abundance changes based on imaging localization and sex. **(E)** Immune cell abundance is enriched in CE for F with a trend toward significance in Ms. **(F)** Immune cell populations differ between imaging regions and the sexes. **(G)** Other (non-malignant, non-immune) cell abundance is not different between CE and NE samples. **(H)** The abundance of other cell populations varies based on sex and MRI localization. Differences in relative and absolute abundances are corrected for multiple observations within patients with significance levels denoted as $* < 0.05$, $** < 0.01$, $*** < 0.001$.

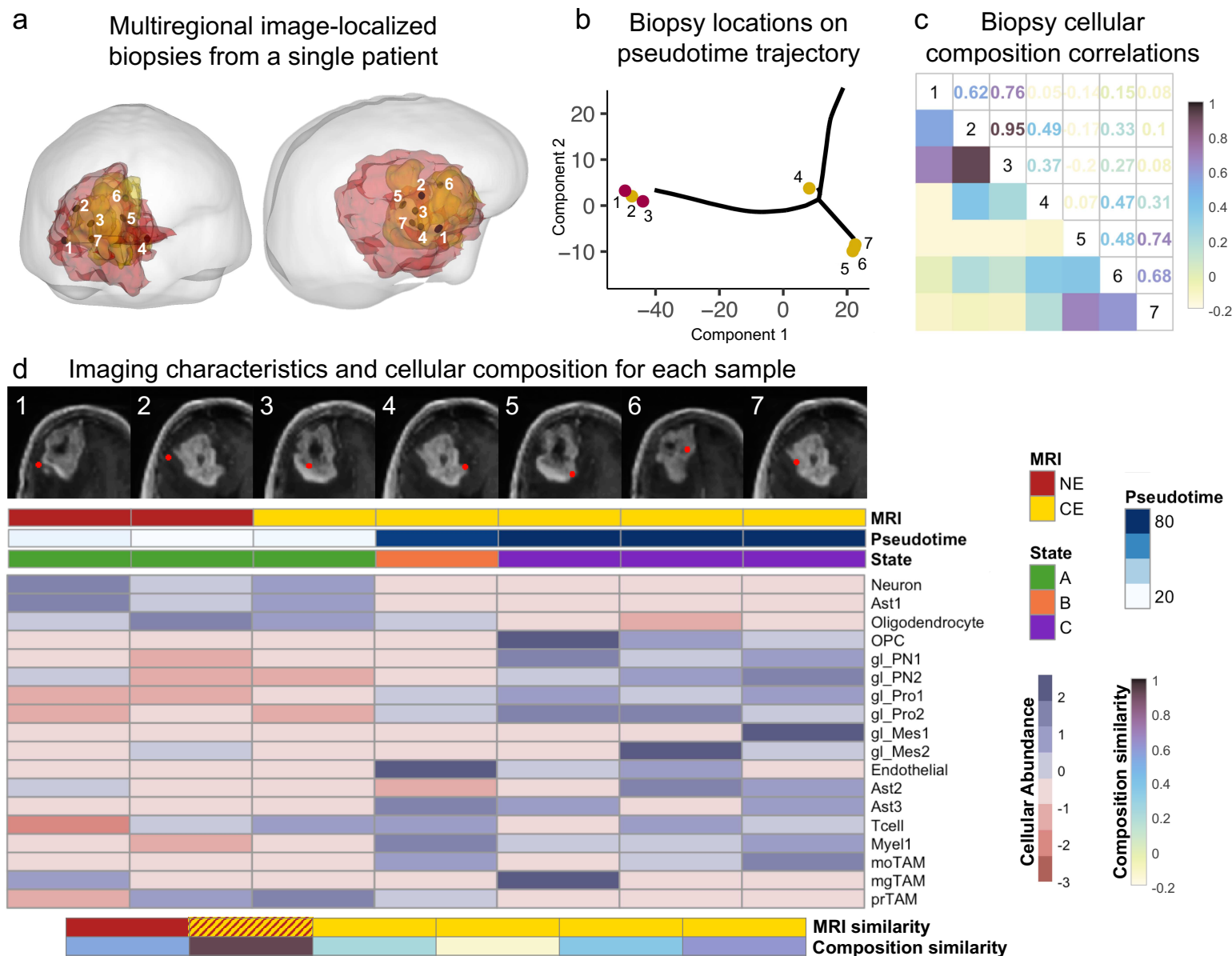


Figure 4. An illustration of biopsy locations and corresponding cell composition in a single patient. (A) A 59-year-old female with primary IDH wildtype, MGMT methylated glioblastoma underwent seven image-localized biopsies (5 CE, 2 NE). **(B)** Samples were distributed across the biological trajectory including States A and C. **(C)** A correlation matrix of cellular composition reveals a spectrum of biological heterogeneity and likeness between samples. **(D)** Though samples that are abundant in malignant cell populations were more commonly harvested from CE regions, imaging localization overall has little bearing on compositional similarity, whereas pseudotime similarity translated to compositional similarity.