Single-cell Bottlenecks and Dead-ends During Glioma Premalignancy

Matthew D. Sutcliffe[1], Rui P. Galvao[2], Lixin Wang[1], Jungeun Kim[2], Shambhavi Singh[1], Hui Zong[2], Kevin A. Janes[1,3]

[1]Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia.
[2]Department of Microbiology, Immunology & Cancer Biology, University of Virginia, Charlottesville, Virginia.
[3]Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, Virginia.

Corresponding Authors: Kevin A. Janes, University of Virginia, Box 800759 Health System, Charlottesville, VA 22908. Phone: 434-243-2126; Fax: 434-982-3870; E-mail: kjanes@virginia.edu; and Hui Zong, University of Virginia, Box 800577, Charlottesville, VA 22908. Phone: 434-982-1956; Email: hz9s@virginia.edu

The authors declare no potential conflicts of interest.

Running title: Cellular bottlenecks and dead-ends for gliomagenesis

Abstract

Cancer evolves from premalignant clones that accumulate mutations and adopt unusual cell states to achieve transformation. Tracking a cancer cell-of-origin through the cell-state alterations of premalignancy could provide clues for early-detection and cancer-prevention strategies. Previously we pinpointed the oligodendrocyte precursor cell (OPC) as a cell-of-origin for glioma. However, the early adaptations and cell-state changes of mutant OPCs during premalignancy are unknown. Using a genetically engineered mouse model (GEMM) of inducible *Nf1–Trp53* loss in OPCs, we acutely isolated labeled mutant OPCs by laser-capture microdissection and determined gene-expression changes in two ways: global changes in gene expression were measured by differential analysis of wild-type and mutant OPCs after bulk RNA sequencing; cell-to-cell state variations were identified by a fluctuation analysis, called stochastic profiling, which uses RNA-sequencing measurements from random pools of 10 mutant cells. We chose two time points for the analysis. At 12 days after *Nf1–Trp53* deletion, while bulk differences were mostly limited to increases in mitotic hallmarks and decreases in ribosome biosynthesis, stochastic profiling of mutant OPCs revealed a spectrum of stem-progenitor, proneural, and mesenchymal states as potential starting points for gliomagenesis. At 90 days after *Nf1–Trp53* deletion, while bulk sequencing detected very few differentially expressed transcripts, stochastic profiling revealed multiple cell states that are absent from glial tumors, suggesting that they marked dead-ends for gliomagenesis. In parallel, we identified cells without dead-end markers but abundantly expressing key effectors of nonsense-mediated decay and homology-dependent DNA repair. This suggests that resolution of replication stress may pose a considerable bottleneck for glioma initiation in premalignant mutant OPCs.

Statement of significance:

In situ heterogeneity profiling of cell states in a mouse model of glioma uncovers regulatory confusion in a glioma cell-of-origin and defines a state of replication stress that precedes tumor initiation.

Introduction

Glioblastoma multiforme (GBM) is the most common type of malignant brain tumor and remains incurable despite decades of basic and clinical research. There are multiple subtypes of GBM that differ in prevalence, mutation frequency, response to therapy, and cell-state composition (1). The latest conceptual frameworks support three subtypes (2), two of which encapsulate the proneural vs. mesenchymal regulatory state of GBM tumor cells. The third ("classical") describes mesenchymal tumor cells mixed with astrocytes instead of the microglial recruitment characteristic of mesenchymal GBMs (3). Malignant cells of different subtypes often reside in the same GBM (4), and advanced GBM clones can give rise to cells with OPC-, neural progenitor-, mesenchymal-, and astrocyte-like signatures (5), suggesting that subtypes could reflect dynamic cell states in glioma. Interestingly, malignant cell-state trajectories appear to be simpler in lower-grade gliomas, even though these tumors also show a mixture of cell-state signatures (6). For multiple glioma subtypes, tumor ecosystems are maintained by a progenitor-like subpopulation that yields terminally differentiated progeny with oligodendrocytic or astrocytic features (7-9). It is unclear whether cell state heterogeneity is even simpler or more chaotic during premalignancy, as glioma-predisposed cells test various states to reach a productive one for long-term tumor growth (10).

The Cancer Genome Atlas project identified three frequently mutated pathways in GBM: Arf/p53 signaling, receptor tyrosine kinase signaling, and the G1 checkpoint (11). Both patient sample analysis and genetically engineered mouse models (GEMMs) indicate oligodendrocyte precursor cells (OPCs) as a cell-of-origin for glioma (12-14). Toward modeling the very earliest steps of gliomagenesis, we previously developed a GEMM that incorporates OPC-selective conditional deletion of *Nf1–Trp53* together with lineage tracing to mark cells upon tumor-suppressor loss (15,16). Deletion in adult mice reactivates OPCs and restores proliferation to perinatal levels within 12 days. However, even though mutant OPCs show a net decrease in oligodendrocyte differentiation at 90 days after gene inactivation, proliferation rates decline to basal levels, right up until the onset of explosive gliomas. Besides these time-dependent changes in proliferation and differentiation, the cell-state changes during premalignant stages remain unknown.

Here, we deconstructed the glioma GEMM's premalignant phases by a combination of bulk transcriptomics and targeted heterogeneous cell-state analysis through 10-cell RNA sequencing (10cRNA-seq) (17-19). Both approaches used laser-capture microdissection (LCM) to isolate labeled cells in situ to avoid tissue dissociation-induced changes of gene expression

that would confound our analysis. Shortly after deletion–labeling, bulk transcriptomic differences largely indicated generic adaptations to *Nf1–Trp53* loss, whereas the heterogeneity analysis of transcriptional variation gave clues about the state of labeled OPCs. Intermediate times of premalignancy showed very limited bulk differences but substantial increases in heterogeneity of cell states. Importantly, we discovered a collection of cells that was not deconvolvable into any recognized cell types of the brain. These samples marked a cell state characterized by elevated DNA repair and nonsense-mediated decay, suggesting difficulty in resolving the replication stress caused by *Nf1–Trp53* deletion. Our results support a conceptual model of early gliomagenesis that is chaotic and largely unproductive. Even with two glioma-predisposing deletions, *p53*- and *NF1*-null mutant OPCs often end up in aberrant “dead-ends” for tumorigenesis while others freeze in a state of replication stress. Glioma precursors appear to struggle in bypassing proliferation–differentiation roadblocks, which might need to be surmounted concurrently through chromosomal gains–losses that enable progression into full-blown tumors.

Materials and Methods

Tissue sources

All mice were maintained according to practices prescribed by the National Institutes of Health in accordance with the IACUC protocol #3955. All animal procedures were approved by the Animal Care and Use Committee at the University of Virginia, accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care (AAALAC). The following mouse strains were crossed on a mixed background to obtain mutant and control mice used in the bulk analysis: Cspg4 (Ng2)-CreER (a kind gift from Dr. Akiko Nishiyama), Rosa-tdTomato (stock #007908, Jackson Laboratories), $NF1^{flox}$ (strain #01XM4, National Cancer Institute), $p53^{flox}$ (strain #01XC2, National Cancer Institute). The above mouse strains were crossed to obtain the mutant mice used in the 10cRNA-seq analysis: Cspg4-CreER; Rosa-tdTomato; $Nf1^{flox/flox}$; $p53^{flox/flox}$. Genotyping was performed by PCR as described (15). Tamoxifen citrate tablets (10 mg/tablet, Mylan) were crushed, resuspended in water, and delivered by oral gavage. All mutant mice were given daily doses of 200 mg tamoxifen/kg for five consecutive days starting at P45. Control mice were given tamoxifen at the end of each experiment instead of at P45 to ensure labeling of OPCs rather than mature oligodendrocytes.

Mouse brain tissue was isolated at either 12, 90, or 150 days post-induction (dpi) from the start of tamoxifen treatment. Animals were not randomized by time point but balanced to the extent possible by sex. For the bulk samples, isopentane was pre-chilled for ~30 min on dry ice to −50°C or lower. To dissect brains, mice were transcardially perfused with PBS supplemented with Procaine and Heparin (Sigma-Aldrich), then brains were dissected and slowly submerged in the pre-chilled isopentane. Brains were removed after one minute, wrapped in Parafilm and stored at −80°C until ready for sectioning and LCM. For the 10cRNA-seq samples, the freshly dissected brain hemispheres were separated and the olfactory bulbs were isolated. Each olfactory bulb was freshly cryoembedded in a dry ice-isopentane bath and stored at −80°C wrapped in aluminum foil until sectioning and LCM.

Before cryosectioning the bulk samples, untreated, plain glass slides were irradiated for 30 minutes with an ultraviolet lamp to eliminate RNases and the cryostat was cleaned with ethanol and RNase-Away (Sigma-Aldrich). Cryosections were 10 µm thick. Once a section was collected, the slide was kept inside the cryostat to avoid thawing and water condensation. For the 10cRNA-seq samples, samples were equilibrated to −24°C in a cryostat before sectioning. 8-µm sections were cut and wicked onto Superfrost Plus slides. To preserve tissue integrity, slides were pre-cooled on the cutting platform for 15–30 sec before wicking, and the section

was carefully placed atop the cooled slide. Then, the slide was gently warmed from underneath by tapping with a finger until the section was minimally wicked onto the slide. All wicked slides were stored in the cryostat before transfer to −80°C storage on dry ice. Frost build-up was minimized by storing cryosections in five-slide mailers.

Fluorescence-guided LCM

For the bulk samples, frozen slides were immediately placed in pre-chilled (−20°C) 100% ethanol for five minutes, then transferred to room temperature 100% xylene for five minutes twice. Finally, slides were removed from xylene and air-dried for five minutes. For the 10cRNA-seq samples, fluorescence-guided LCM was performed as described previously (18). Immediately from −80°C storage, slides were fixed in 70% ethanol for 15 seconds, 95% ethanol for 15 seconds, and then 100% ethanol for two minutes before clearing with xylene for two minutes. Slides were air-dried for 5–10 minutes before beginning LCM. Freshly fixed samples were microdissected on an Arcturus XT LCM instrument (Applied Biosystems) using Capsure HS caps (Arcturus). The smallest spot size on the instrument (10–12 μm) captured one OPC per laser shot.

RNA extraction and amplification

For bulk samples, total RNA from the Capsure HS cap was extracted and purified with the PicoPure RNA isolation kit (Arcturus) according to manufacturer instructions. Briefly, tissue was immediately incubated in 10 μl extraction buffer at 42°C and stored at −80°C. Samples were column purified and digested with DNase to remove genomic DNA. Total RNA was eluted in 11 μl RNase-free water and concentrated to 5 μl in a speedvac for five minutes at room temperature. Single-primer, isothermal linear amplification (Ribo-SPIA) of purified total RNA was performed with the Ovation RNAseq System V2 kit (NuGen) according to manufacturer instructions. Amplified cDNA molecules were purified before generating sequencing libraries.

For 10-cell samples, RNA extraction and amplification of microdissected samples was performed as described previously (18). Briefly, cells were eluted from Capsure HS caps (Arcturus) and digested. Biotinylated-cDNA was then synthesized from RNA eluted from captured cells and purified using streptavidin magnetic beads (Pierce) with a 96 S Super Magnet Plate (Alpaqua). Residual RNA was degraded by RNAse H (NEB) digestion and poly(A) tailed with terminal transferase (Roche). Poly(A)-cDNA was then amplified using an AL1 primer

(ATTGGATCCAGGCCGCTCTGGACAAAATATGAATTCTTTTTTTTTTTTTTTTTTTTTTTTT) and a blend of Taq polymerase (NEB) and Phusion (NEB).

Quantitative PCR (qPCR)-based quality assessment

Detection of transcripts by qPCR was performed on a CFX96 real-time PCR instrument (Bio-Rad) as described previously (20). Samples were filtered based on three loading controls: *Rplp1*, *Gapdh*, and *Apoe*. Samples were retained if the geometric mean of the *Rplp1*–*Gapdh*–*Apoe* quantification cycles was within 3.5x interquartile range of the median; samples outside that range were excluded because of over- or under-capture during LCM. We also sought to exclude samples that were overly contaminated with *Cspg4*-expressing pericytes. Therefore, samples with detectable *Pdgfrb*, *Des*, and *Anpep* were excluded from sequencing.

Library preparation and sequencing

For bulk samples, Illumina sequencing libraries were made by standard methods; briefly, double-stranded cDNA was mechanically sheared to 150–600 bp by sonication using a Bioruptor (Diagenode), and the fragments end-repaired, A-tailed and ligated to T-overhang adaptors. Samples were ligated to Illumina adapters each with a 5-nt index, amplified for 10 to 12 cycles, and size selected to 250 to 500 bp. Samples were sequenced on an Illumina HiSeq 2000 instrument.

10cRNA-seq libraries for sequencing were re-amplified, purified, and tagmented as described previously (18). Briefly, each poly(A) PCR cDNA sample was re-amplified for a number of cycles where the amplification remained in the exponential phase (typically 10 to 20). Re-amplified cDNA was then twice purified with Ampure Agencourt XP SPRI beads. After bead purification, samples were quantified on a CFX96 real-time PCR instrument (Bio-Rad) using a Qubit BR Assay Kit (Thermo Fisher). Samples were diluted to 0.2 ng/µl and tagmented with the Nextera XT DNA Library Preparation Kit (Illumina). Samples were sequenced as previously described (18). Samples were multiplexed at an equimolar ratio, and 1.3 pM of the multiplexed pool was sequenced on a NextSeq 500 instrument with NextSeq 500/550 Mid/High Output v1/v2/v2.5 kits (Illumina) to obtain 75-bp paired-end reads.

RNA sequencing alignments

Adapters were trimmed using fastq-mcf (version 1.1.2-779) in the EAutils package with the following options: -q 10 -t 0.01 -k 0 (quality threshold 10, 0.01% occurrence frequency, no nucleotide skew causing cycle removal). Quality checks were performed using FastQC (version

0.11.8) and MultiQC (version 1.7). Datasets were aligned to the mouse transcriptome (GRCm38.82) along with reference sequences for ERCC spike-ins, using RSEM (version 1.3.0) and Bowtie 2 (version 2.3.4.3). RSEM options for the 10cRNA-seq data also included the following options: --single-cell-prior --paired-end. RSEM read counts were converted to transcripts per million (TPM) by dividing each value by the total read count for each sample and multiplying by 10^6. Mitochondrial genes and ERCC spike-ins were not counted towards the total read count during TPM normalization.

Differential expression analysis

DESeq2 (version 1.26.0) was used to identify differentially expressed transcripts. For the bulk samples, outliers were first identified by principal components analysis and removed from subsequent analysis. The differential expression analysis was performed separately at each of the three timepoints. Mouse sex was included as a covariate with full model design "~ genotype + sex" and reduced model design "~ sex". Genes with an adjusted *p*-value less than 0.05 were considered significantly up- or downregulated between the control and mutant samples. For the 10cRNA-seq samples, DESeq2 was used to identify differentially expressed transcripts between the U group and the O–E–N union group with model design "~ group". Genes with an adjusted *p*-value less than 0.05 were considered significantly up- or downregulated between the two groups.

Overdispersion-based stochastic profiling

Stochastic profiling with 10cRNA-seq data was performed separately for the male and the female datasets exactly as described in the accompanying contribution (19).

Robust identification of transcriptional heterogeneities through crossvalidation

To identify heterogeneously expressed genes that were robust to sample outliers, we performed 100 leave-one-out simulations separately on the male and female datasets. For each simulation, one 10-cell sample and one pooled control were randomly chosen and removed from the dataset. To maintain a total of 28 10-cell samples and 20 pooled controls, one each of the remaining 10-cell samples and pooled controls was duplicated. The overdispersion analysis was performed as described above, returning a list of heterogeneously expressed genes for each simulation. Sex-specific heterogeneities were defined as those genes identified as heterogeneously expressed in at least 75% of the simulations.

Glioma subtyping

Bulk and 10cRNA-seq samples were projected onto the first two principal components of the GBM C1 principal components analysis (3). Mouse genes were first converted into their human orthologs using the "biomaRt" package in R. Of the 714 human genes contributing to the principal components, 629 genes had mouse orthologs in our dataset that were used for the projection. TPM values were scaled to TPM/100 + 1 and $\log_2$-transformed before applying the projections. Maximum-likelihood inference was performed on the 10cRNA-seq 12 dpi sample scores in the first principal component to determine constituent single cell contributions (21). Scores were right-shifted by 120 for parameterization of lognormal distributions; then, location parameters were left-shifted by 120 for display. The inference was performed with two lognormally-distributed cell populations and ten cells per sample. The probability density function for the mixed samples was determined by repeatedly sampling from the resulting single-cell distributions. To compare bulk samples and 10cRNA-seq samples along the proneural–mesenchymal and proliferation principal components at 12 dpi, 70–100 10-cell samples were randomly selected with replacement. TPM values were averaged before projecting onto the two principal components as described above.

Lineage and cell-state deconvolutions

CIBERSORT was used to unmix cellular samples into their composite cell types (22). Cell-type signatures were derived using cell-sorted RNA-seq data (23). The two biological replicates from each of five cell types (astrocyte, endothelial, myelinating oligodendrocyte, microglia, and neuron) were averaged. Because of ambiguity in the oligodendrocyte-lineage samples and because differentiated OPCs would continue to be fluorescently labeled, one OPC sample and one newly-formed oligodendrocyte sample from (23) were averaged to yield a hybrid "OPC" more characteristic of the expected population of labeled cells. Signature genes were defined as the transcripts with the highest fold enrichment of one cell type vs. the average of the remaining cell types. Only genes with FPKM > 20 were included in the analysis. We identified 40 signature genes per cell type, and FPKM values were converted to TPM before the CIBERSORT analysis. To account for pericyte labeling in the cell unmixing analysis, the endothelial signature was exchanged for a similarly derived pericyte signature based on mural cell RNA-seq data (24). The new signature matrix was optimized as described above. For the bulk 150 dpi control samples, CIBERSORT was run in relative mode because unclassified cell types were not expected. For the 90 dpi unmixing, absolute mode was used because of potentially unclassifiable cell states.

UMAP projections

All UMAP projections were generated using the R package "uwot" (version 0.1.5). For the 12 dpi UMAP projections, RSEM values were converted to TPM values before $\log_2$ scaling. Default parameters were used. For the combined bulk + 10cRNA-seq UMAP projections, RSEM values were converted to TPM values before $\log_2$ scaling. Bulk samples and 10cRNA-seq samples were then z-scored independently. The following UMAP parameter was used: n_neighbors = 20.

Statistics

Sample sizes for stochastic profiling were determined by Monte Carlo simulation (25). Animal sizes for the bulk analysis provided 66–93% power to detect an effect size of two standard deviations according to noncentral *t* statistics. Differences in genes detected per sample between 10cRNA-seq and scRNA-seq were assessed by Kolmogorov-Smirnov test using the "ks.test" function in R. Significance of overlap was assessed by Fisher's exact test using the "fisher.test" function in R. In the male and female overlaps, the total number of genes was defined by the intersection of filtered genes in the male and female datasets (10,585 genes for 12 dpi and 11,281 genes for 90 dpi). In the RHEG and DE overlaps, the total number of genes was defined by the intersection of the previously filtered gene list and the detected genes from the DE analysis (10,092 genes for 12 dpi and 10,503 genes for 90 dpi). Significance of the relative size of RHEG or DE gene lists was assessed by binomial test using the "binom.test" function in R. Hierarchical clustering was performed using "pheatmap" in R using Euclidean distance and "ward.D2" linkage. Venn diagrams were visualized using the "matplotlib_venn" package in Python. Gene set enrichment analyses were performed through the Molecular Signatures Database (26). Overlaps between gene lists and hallmark gene sets were computed using a hypergeometric test with false-discovery rate correction for multiple comparisons.

Data availability

Bulk and 10cRNA-seq data from this study is available through the NCBI Gene Expression Omnibus (GSE147360, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147360 Reviewer token: odgbkqqmlxgdrof).

Results

Comparing bulk-transcriptome and cell-state changes in a GEMM of glioma

We previously described a GEMM for low-grade glioma and secondary GBM in which the glioma-related tumor suppressors *Nf1* and *Trp53* are deleted with a tamoxifen-inducible Cre (CreER) (15) under the control of a *Cspg4* promoter (P_{Cspg4}) that specifically targets OPCs (12). *Nf1* deletion mimics receptor tyrosine kinase signaling and drives cells through the G1/S transition; *Trp53* deletion disrupts tumor suppression through the Arf/p53 pathway (11). A Cre reporter is incorporated into the GEMM so that mutant OPCs express tdTomato for fluorescence visualization. Serial administration of tamoxifen creates tdTomato+ OPCs with an $Nf1^{\Delta/\Delta}Trp53^{\Delta/\Delta}$ mutant genotype, which clonally expand and form malignant gliomas in ~150 days (**Fig. 1A**) (15).

To gain insight into the cell-state changes in mutant OPCs before gliomagenesis, we devised a study in which mutant OPCs were purified at two premalignant stages for gene-expression analysis: 12 dpi, when mutant OPCs show evidence of proliferative expansion (15); and 90 dpi, when mutant OPCs appear largely quiescent, presumably adapting to prolonged loss of *Nf1* and *Trp53*. Given sex differences in the prevalence and molecular subtypes of human glioblastoma (27), we sought a balanced design that split observations evenly between males and females (**Fig. 1B** and **1C**).

Accessing labeled cells required some form of OPC isolation or purification with single-cell resolution. However, an earlier attempt to enrich OPCs by flow sorting (28) yielded fewer transcript alignments compared to other methods (18). Therefore, we used fluorescence-guided LCM under conditions of cryoembedding, sectioning, and fixation that retained tdTomato fluorescence and RNA integrity in labeled cells (18,29). *Nf1–Trp53* deletion should cause not only population-wide changes in gene expression compared to control OPCs by bulk analysis, but also single-cell variations in transcript abundance requiring the stochastic-profiling method developed by our group (17,19). For the bulk analysis, we microdissected many hundreds of cells to minimize the impact of cell-to-cell variation (Supplementary Fig. 1) (25) and sequence without any exponential amplification (see Materials and Methods; **Fig. 1B**). The stochastic-profiling implementation, by contrast, involved repeated, random collection of 10-cell pools (17), which were preamplified, sequenced, and assessed for overdispersion as described in the accompanying contribution (19) (**Fig. 1C**). By applying different unbiased methods to the same microdissected subpopulation, we hoped to assess their relative merits for deciphering premalignancy.

LCM-based RNA sequencing validates differential gene expression and proneural subtype of the glioma model

According to bulk microarray profiles and an early glioma-classification system (1), the glioma model is most similar to the human proneural subtype (15). Subsequent work at the single-cell level suggested that intratumor heterogeneity of advanced gliomas is substantial (4), with the proportion of different cell types (recruited astrocytes vs. microglia) and cell states (proneural vs. mesenchymal) dictating the bulk classification. Glioma-propagating cells are now proposed to reside on a spectrum of proneural-to-mesenchymal cell fate, which is orthogonal to a cell's proliferative state (3). The question of subtyping is also separate from differential expression, as many of the markers delineating a subtype will reflect vestiges of the cell type in which the tumor appeared. Together, these considerations motivated a reassessment of 150 dpi glioma cells by using the latest experimental and analytical methods (**Fig. 1B**).

Although gliomas were visibly macrodissectable at 150 dpi (**Fig. 2A**), mutant OPCs at earlier time points remained sparse throughout the tissue; thus, LCM was used for all sample types throughout the study (**Fig. 2B**). Several hundred laser shots of ~one-cell diameter could be fired per cryosection to generate enough material for RNA sequencing after Ribo-SPIA (30) linear amplification (see Materials and Methods). We confirmed that transcriptomes of separate LCM bulk samples from the same tumor were highly correlated after linear amplification (Supplementary Fig. S2). Control OPCs were singular and much more entangled with unlabeled neighbors compared to prior applications of fluorescence-guided LCM by our group (29,31). To assess overall purity of normal controls, we used a collection of brain cell type-enriched marker transcripts (23) to construct a signature matrix for bulk sample deconvolution using CIBERSORT (22). The signature matrix accurately classified the composition of flow-sorted OPCs, astrocytes, neurons, myelinating oligodendrocytes, endothelial cells, and microglia (Supplementary Fig. S3A and S3B). When applied to 150 dpi control samples, the inferred cell composition was overwhelmingly OPC with minor contributions of astrocytes and neurons, fragments of which likely wrapped around cells targeted by LCM (**Fig. 2C**). We considered *Cspg4*-expressing pericytes as an alternative source of contamination and substituted a set of pericyte markers (24) in place of the endothelial signature (Supplementary Fig. S3C and S3D). The inferred pericyte fraction was negligible, and the other cell proportions were similar to the first deconvolution but more variable as a result of pooling marker sets from different studies. We obtained similar results with 12 dpi and 90 dpi control samples (Supplementary Fig. S3E–H), supporting that cell purity does not change with tissue age.

Together with the near-instantaneous cryopreservation (less than two minutes from sacrifice to embedding), the results bolster LCM as an effective means for labeled OPC isolation and transcriptomic profiling.

We sequenced seven samples from tumors and six samples of OPCs from age-matched control mice, detecting 11,748 ± 651 genes at greater than five transcripts per million (TPM). To revisit the molecular subtype of glioma cells in bulk, we used a principal components-based projection that stratifies cells along two axes: proneural-vs.-mesenchymal differentiation and quiescent-vs.-proliferative activity (3). Both control OPCs and mutant tumors clearly resided on the proneural side of the projection, but there was greater variability observed among the tumors (**Fig. 2D**). As expected, tumors were also consistently more proliferative according to their projections along the second axis. We examined quantitative differences in transcript abundance between the two groups with DESeq2 (32), incorporating sex as a covariate (see Materials and Methods). The analysis identified 6057 differentially abundant transcripts; there was a bias toward upregulation, consistent with the accelerated proliferation of the tumors (3453 genes vs. 2604 genes, $p < 10^{-15}$ by binomial test; **Fig. 2E** and Supplementary File S2). In search of coherent pathways, we queried the Molecular Signatures Database (26), finding hallmarks related to *Trp53* loss (DNA damage), *Nf1* loss (Ras signaling, proliferation), and immune responses consistent with local activation of microglia (2) (**Fig. 2E** and Supplementary File S1). The bulk-profiling results from 150 dpi samples validate the LCM procedure and reinforce that this GEMM gives rise to proneural gliomas with gene-expression patterns reflecting *Nf1*–*Trp53* loss.

Transcript abundance changes and cell-state heterogeneities are uncoupled in early mutant OPCs

We anticipated that the early responses to *Nf1*–*Trp53* loss would be more informative for tumorigenesis and more variable among labeled cells. Therefore, we coupled bulk transcriptomics with heterogeneity analysis by stochastic profiling (17,19) at 12 dpi to capture the immediate response of *Nf1*–*Trp53* deletion (**Fig. 1**). The two-pronged approach would enable a comparative assessment and provide complementary perspectives on the single-cell changes elicited by a two-hit tumor-suppressor loss in OPCs.

For the seven mutant OPC samples and five controls measured in bulk, we found that overall transcript content was very similar to that of the 150 dpi samples (11,674 ± 150 genes detected at greater than five TPM). However, when the 12 dpi bulk samples were projected onto the earlier principal-component axes (3), mutant and control OPCs overlapped as a tight

cluster, all with moderate proliferation and a proneural phenotype (**Fig. 3A**). Mutant OPCs incorporate BrdU at an elevated rate during the week after tamoxifen addition (15). The lack of a distinguishing RNA-seq projection here suggests that the proliferative burst of the GEMM occurs immediately after *Nf1–Trp53* deletion is induced and subsides by 12 dpi.

We performed a differential-expression analysis on the bulk samples and identified 469 transcripts elevated or reduced in mutant OPCs compared to control, 204 of which were identified at 150 dpi ($p < 10^{-15}$ by hypergeometric test; **Fig. 1E, 3B**, and Supplementary File S2). Notably, over half of the shared transcripts switched the direction of their abundance change between 12 dpi and 150 dpi. This reversal between the two time points illustrated the caveats of using late-stage comparisons between tumors and normal cells to deduce premalignant trajectory. The fewer differences and lower fold changes overall at 12 dpi compared to 150 dpi supported that secondary transcriptional effects of transformation at 150 dpi had been removed. Among individual transcripts, we noted large increases in the axon guidance receptor *Robo3* (33) ($\log_2$ fold change = 4.4) and *Clspn* ($\log_2$ fold change = 2.2), a scaffold coordinating the DNA replication checkpoint (34) (**Fig. 3B** and Supplementary File S2). Many ribosomal subunits were reciprocally decreased, likely due to the concomitant reduction of *Mycn* (35) ($\log_2$ fold change = –0.82). Unexpectedly, there was not an obvious coordination in expression changes stemming from recognized pathways. Hallmark gene-set enrichments were absent from the 12 dpi data except for the mitotic spindle, which among increased transcripts was enriched just beyond the false-discovery rate ($q < 0.01$; Supplementary File S1). Our observations at 12 dpi suggested that the bulk RNA-seq data were capturing generic adaptations to *Nf1* loss (36) unencumbered by Trp53 checkpoints.

Although bulk transcriptomes were very consistent at 12 dpi (**Fig. 3A**), we reasoned that the state of individual OPCs during and shortly after *Nf1–Trp53* deletion would be different. To evaluate regulatory heterogeneities in mutant OPCs, we carefully devised an experimental plan to apply stochastic profiling by 10cRNA-seq at 12 dpi. First, random 10-cell sampling was restricted to labeled OPCs in the olfactory bulb, where gliomas are most likely to arise in the GEMM (15). Samples were excluded if there was evidence of pericyte contamination by qPCR (Supplementary File S3). Second, we hedged against mouse-to-mouse variations by allocating 28 random samples and 20 pool-and-split controls equally between olfactory-bulb hemispheres of two separate animals. The samples and controls were merged for analysis (19) and inspected afterwards for mouse-dependent batch effects. Statistically, we built upon the resampling approach of the accompanying contribution (31) by evaluating overdispersion stability through leave-one-out cross-validation (see Materials and Methods). Consistent

regulatory heterogeneities were readily delineated as those appearing in >75% of cross-validation runs (Supplementary Fig. S4). Finally, given the sex differences documented in human GBM (27), the entire stochastic-profiling study was performed once in males and once in females. Intersecting the male and female candidates at the 12 dpi time point led to a set of recurrent, heterogeneously expressed genes (RHEGs) for gleaning cell states.

Before fluctuation analysis, we projected the 10cRNA-seq observations onto the principal-component axes (3) used previously with bulk data (**Fig. 2D** and **3A**). Proliferation scores were reduced and more variable (**Fig. 3C**), as expected from the ~100-fold reduction in cellular material. Projections on the proneural-to-mesenchymal spectrum were similarly displaced because 10-cell observations did not detect the strongest proneural markers as reliably as bulk RNA-seq (Supplementary Fig. S5). Notwithstanding these location differences, we found that the 10cRNA-seq samples were distinctly bimodal, half with more-proneural character and half with more-mesenchymal character (Supplementary Fig. S6A). The observations were consistent with single-cell profiles (21) that were more frequently mesenchymal than proneural (see Materials and Methods; Supplementary Fig. S6B). However, the mesenchymal bias accentuated by 10-cell data reverted to the mean and disappeared when samples were agglomerated computationally as bulk averages of 700–1000 cells (Supplementary Fig. S6C). This first-pass comparison of measurement types indicated that 10cRNA-seq data at the 12 dpi time point were fundamentally different from bulk observations.

We inspected 10-cell differences more globally by using uniform manifold approximation and projection (UMAP) (37) to reduce dimensionality of the 10cRNA-seq profiles. Across 56 10-cell samples of mutant OPCs (28 male + 28 female), we detected 4570 ± 1900 genes at greater than five TPM, suggesting that ~40% of the bulk transcriptome would be amenable for analysis. [By comparison, applying a similar threshold to scRNA-seq data from flow-sorted OPCs (28) yielded ~15% fewer transcripts (3852 ± 956 genes, $p < 0.01$ by K-S test).] Using two UMAP dimensions, there were 2–3 clusters of gene expression, each populated with 10-cell observations from different tissue batches at 12 dpi (**Fig. 3D**). The global projections illustrated that confounding batch effects were negligible and that cell-state heterogeneities might be embedded in the 10-cell transcript fluctuations from sample to sample.

Regulatory heterogeneities were assessed formally through abundance-dependent dispersion modeling (38,39) that used pool-and-split controls of 10-cell equivalents to exclude technical artifacts (19). High-ranking candidates showed 10-cell sampling fluctuations much greater than expected given their abundance and technical noise (**Fig. 3E**). At 12 dpi, we identified 707 candidate regulatory heterogeneities in males and 1057 in females

(Supplementary File S4). The significantly higher proportion of candidates in females ($p \sim 0$ by sign-rank test) coincided with a modest enrichment for human GBM transcripts specific to females ($p < 0.05$ by hypergeometric test), who have a reduced GBM incidence and longer survival compared to males (27). We observed no such enrichment for male-specific transcripts among male regulatory heterogeneities, supporting a recent report that OPC diversity may be less pronounced in males (40). Importantly, between sexes, there was a highly significant enrichment of 138 RHEGs shared at 12 dpi ($p < 10^{-14}$ by hypergeometric test; **Fig. 3F**, left, and Supplementary Fig. S7). Yet, only four of the RHEGs identified by stochastic profiling overlapped with the transcripts found to be differentially expressed in bulk 12 dpi samples ($p = 0.53$ by hypergeometric test; **Fig. 3F**, right). We interpreted 12 dpi RHEGs as sex-independent variations in mutant cells that are almost entirely obscured at the population level.

There were multiple transcripts of interest in the 12 dpi RHEG set (Supplementary File S4). The receptor tyrosine kinase *Axl*, for instance, is implicated in oligodendrocyte survival and myelination after injury (41) and promotes self-renewal of mesenchymal progenitors in GBM (42). The stem-cell marker *Aldh1a1* was also a RHEG with links to an OPC subpopulation found at roughly equal proportions in males and females (40). Relevant to the mixed proneural–mesenchymal states (**Fig. 3C**), we found extensive evidence for heterogeneous regulation of transforming growth factor β (TGFβ)-family signaling (43,44). The latent TGFβ-binding protein *Ltbp4*, the TGFβ-superfamily ligand *Bmp1*, and the Bmp-subfamily inhibitor *Nbl1* (45) were all RHEGs (Supplementary File S4). Using *Nbl1* as a reference, we searched the 10cRNA-seq transcriptomes for strong covariates that fell below the statistical thresholds of the dispersion test (46), identifying the TGFβ-induced transcripts, *Snai2* ($R = 0.58$) and *Vip* ($R = 0.52$) (47). Regulators of OPC state transitions were also prevalent RHEGs, including *Wnt7a* and *Ctnnb1* (48). Although some 12 dpi RHEGs reflect normal asynchronies in OPC maturation, they nonetheless give clues about the different cellular contexts in which *Nf1–Trp53* deletion occurs before premalignancy.

Premalignant glioma is punctuated by aberrant differentiation and a neomorphic state coupling DNA repair with nonsense-mediated decay

The 12 dpi studies documented the initial OPC cell states and early adaptations to *Nf1–Trp53* loss. We suspected a very different set of transcriptional changes and variation at an intermediate time, when unproductive mutants have stalled or been culled and the remaining expansion of mutant OPCs enters premalignancy (**Fig. 1A**). For transcriptomic profiling, we selected 90 dpi, a time point when mutant OPC proliferation has returned to basal levels in the

GEMM and the overall number of labeled cells has stabilized (15). Even though tdTomato+ cells were more numerous in the olfactory bulb at 90 dpi, there is no evidence for local expansion of mutant clones (**Fig. 1A**) (15). Therefore, the study design for bulk and 10-cell sample acquisition was identical to 12 dpi to enable comparisons between time points.

The bulk 90 dpi profiling of seven mutant OPC samples and four controls yielded 11,577 ± 106 genes detected at greater than five TPM, very similar to the other time points. Despite months of evolution, the projection of bulk 90 dpi samples onto proliferative and proneural–mesenchymal principal components was indistinguishable from their 12 dpi counterparts (**Fig. 3A** and **4A**). The subsequent divergence of 150 dpi gliomas on these principal components (**Fig. 2D**) suggests a late event drives malignant transformation in the GEMM. Contrary to expectation, there were significantly fewer differentially expressed genes at 90 dpi than at 12 dpi, with only 82 increased and 61 decreased ($p < 10^{-15}$ by binomial test; **Fig. 4B**). Although there was no hallmark gene-set enrichment among the upregulated transcripts, we noted a few proliferation markers—*Top2a*, *Ccna2*, *Hjurp*—that increased with $\log_2$ fold change = 1.2–1.9 (**Fig. 4B** and Supplementary File S2). Thus, despite undetectable increases in BrdU incorporation at 90 dpi (15) and lack of an overarching proliferative signature (**Fig. 4A**), more mutant OPCs appeared to be entering S phase than control. Reduced transcripts included many high-abundance markers for pericytes (*Pdgfrb*, *Cd248*, *Anpep*) (24), suggesting that their minor contamination had been further diluted out by the expanded mutant OPCs.

The projections of 10cRNA-seq data collected at 90 dpi (5649 ± 1197 genes at greater than five TPM) were likewise comparable to those at 12 dpi, except that the range of proliferation scores was 1.5-fold smaller (**Fig. 3C** and **4C**), indicating a narrowed breadth of cycling cells. Additionally, we found that cell-state scores became sexually dimorphic, with female samples residing on the proneural end of the spectrum and males on the mesenchymal end (Supplementary Fig. S8). Intersection of male and female candidate heterogeneities identified 231 RHEGs ($p < 10^{-15}$ by hypergeometric test), a significant increase compared to the 138 RHEGs at 12 dpi ($p < 10^{-12}$ by binomial test; **Fig. 3F** and **4D**, Supplementary Fig. S9, and Supplementary File S4) and a significant increase compared to the bulk analysis at 90 dpi ($p < 10^{-11}$ by binomial test) . As before, very few RHEGs were differentially abundant in bulk samples and vice versa (**Fig. 4D**), reiterating the stark distinction between the two approaches.

The larger number of RHEGs ($p < 10^{-10}$ by binomial test) suggested that a deeper investigation of cell-state heterogeneity at 90 dpi would be informative. The RHEG *Ly6a* (a murine stem-cell marker) covaried with the reprogramming factor *Klf4* ($\rho = 0.54$) and the mesenchymal transcript *Vim* ($\rho = 0.41$) that lay in the union set of candidates (**Fig. 4D**).

Further, there were various markers for oligodendrocytes (*Mog*, *Cldn11*, *Opalin*, *Mbp*) as reported in human gliomas (9), neurons (*Syt7*, *Resp18*, *Stmn2*) as reported in human GBMs (5), and pericyte-like cells (*Rgs5*, *Vtn*) (49) within the 90 dpi RHEG set (Supplementary File S4). In bulk samples, the 90 dpi time point showed no differences in overall purity compared to 12 dpi or 150 dpi (Supplementary Fig. S3C–H). We interpreted the mixed distribution of fate markers as a chaotic (de)differentiation of mutant OPCs into states partly reflecting other cells of the brain.

To quantify cell-state character, we returned to CIBERSORT, applying the algorithm in absolute mode for estimating how much of each 10-cell profile was attributable to markers in the signature matrix (**Fig. 4E** and Supplemental Fig. S3A). 34% of 90 dpi observations retained OPC as the most dominant contributing signature (O group, **Fig. 4E**). Observations with strong endothelial contributions suggested a confused mural-cell state (49) (E group, **Fig. 4E**). We also found 18% of samples with dominant (albeit weaker) contributions from neuronal signatures (N group, **Fig. 4E**). Neuron-like cells are absent in various subtypes of glioma (7-9), suggesting the N group may contain cellular dead-ends for gliomagenesis. Most importantly, we identified 14% of 10-cell samples lacking strong contributions from any conventional cell type in the brain, marking the group as undefined (U group, **Fig. 4E**). Consistent with prior immunocytochemistry (15), samples with properly differentiated myelinating oligodendrocytes were rare. The results together indicated that bulk tumor deconvolution is compatible with 10cRNA-seq pools and yields coherent sample categories for further analysis.

We hypothesized that the mysterious U group would contain information about a premalignant bottleneck. When the U group was separated from the O–E–N samples and evaluated for differential transcripts (see Materials and Methods), we found widespread downregulation of 163 genes. Among the 13 transcripts with increased abundance in the U group were *Upf3b*, an X-linked regulator of nonsense-mediated decay (50), and a set of genes directly involved in resolving Holliday junctions during homology-dependent repair: *Rad51c*, *Slx1b*, and *Ercc4* (51,52) (**Fig. 4F**). The U group also showed an increase in *Bicra*, a glioma tumor-suppressor candidate with a human polymorphism that predisposes for 19q deletions in oligodendroglioma (53). Without *Trp53* function, homology-dependent repair appears to create a regulatory choke point during premalignancy, which blurs cell identity and may lead to chromosomal instability.

Informative differences in bulk and 10-cell transcriptomes define self-consistent trajectories of gliomagenesis

Differential transcripts and RHEGs showed little overlap at 12 dpi and 90 dpi (**Fig. 3F** and **4D**), raising the possibility that the data types were too dissimilar for meaningful comparison. We asked whether the bulk and 10cRNA-seq data could be reconciled by fusing the observations according to only those transcripts that were differentially abundant or recurrently heterogeneous at least once (see Materials and Methods). The datasets were fused after standardization and reduced to two UMAP dimensions, with samples displayed as subsets for inspection of the median centroids pertaining to each group. The goal was to evaluate bulk and 10-cell sample trajectories on the UMAP for themes common to both data types.

With the bulk samples obtained at 150 dpi, we noted a diagonal trajectory from control OPC samples to tumor cells along the two UMAP dimensions (**Fig. 5A**). We interpreted this trajectory as a gliomagenesis "axis" within the UMAP. For 12 dpi samples, the trajectory from control to mutant was nearly perpendicular to that at 150 dpi (**Fig. 5B**) and delineated an adaptive axis, which was distinct from the late readouts of tumor-suppressor loss projected by the 150 dpi tumors. Most interesting were the 90 dpi trajectories, in which bulk mutant samples appeared to bifurcate in two directions from the controls (**Fig. 5C**). Half of the 90 dpi mutant samples were similar to 12 dpi bulk observations, indicating an averaged population that was still undergoing adaptation, while the other half mapped toward the 150 dpi tumors, suggesting a fraction of cells in those bulk samples were transitioning to a gliomagenic state. The confused (de)differentiation of mutant OPCs during premalignancy undoubtedly contributed intrinsic variance to the very small number of differential transcripts detected at 90 dpi (**Fig. 4B**).

Finally, we visualized the 10cRNA-seq data with the same UMAP cartography (**Fig. 5D**). The 12 dpi samples formed a cloud of observations that resided nearby controls and bulk mutant OPC observations at 12 dpi (**Fig. 5D** and **5E**), consistent with variable extents of progressive adaptation among mutant OPCs. In contrast, the 10-cell samples at 90 dpi were spanned by the bifurcating 90 dpi mutant OPC profiles measured in bulk (**Fig. 5D** and **5F**). We conclude that the 10cRNA-seq observations here reflect piecewise instances of premalignancy that contribute to the extreme variability in mutant OPC cell states before glioma onset. Thus, bulk and 10-cell data can be reconciled when viewed through their aggregate transcripts of interest.

Discussion

Modern studies of intratumor heterogeneity strive for robust assessments of differential expression at single-cell resolution (5,7,9). For bulk tumors, quantitative changes in transcript abundance are often explained by differences in cell-type composition (8,54). Our study goes to the next level of granularity by examining bulk vs. single-cell differences in one cell type marked with a genetically defined perturbation that predisposes for glioma. The heterogeneity analysis uncovered differences in stem–progenitor markers, proneural–mesenchymal states, and cell-fate assignments that were obscured in bulk averages. However, the 10-cell profiles also averaged over most cycling transcripts (19), many of which were detectably increased in bulk mutant OPCs compared to control. Other transcripts with bulk changes in expression did not fluctuate any more than expected given their measured abundance, implying that these alterations must occur somewhat uniformly in mutant OPCs. Amidst a relatively homogeneous adaptation to *Nf1*–*Trp53* deletion, stochastic events relating to OPC differentiation and proliferation stress accumulate until a productive mutant emerges.

For both time points analyzed by stochastic profiling, we identified substantially more regulatory heterogeneities in mutant OPCs from females compared to males. The difference is not easily explained by asynchronous hormonal cycles, because the observed variability was shared across samples. We also did not find an enrichment for female-specific candidates on the X chromosome, excluding random X inactivation as a possible explanation. The many female-specific heterogeneities at 12 dpi suggests a larger starting diversity of OPCs. Indeed, a recent study suggests that two OPC subpopulations dominate in males, whereas three subpopulations exist at roughly equal proportion in females (40). In addition to less regulatory diversity, the male 10-cell projections at 90 dpi scored as much more mesenchymal than females. This observation could relate to disease progression, as *Nf1*–*Trp53*-predisposed males are more likely to acquire further tumor-suppressor losses related to the mesenchymal subtype (55). Conceivably, reduced cellular diversity in males could increase the chance of gliomagenic events occurring in a mesenchymally susceptible subpopulation of OPCs.

The 90 dpi samples were unmixed into an assortment of cell states, but what ultimately enables a premalignant mutant OPC to become glioma? We surmise that mechanisms for gliomagenesis are embedded in the RHEG set at 90 dpi. For example, oligodendroglial tumors are tied to a loss of asymmetric cell divisions by OPCs and their differentiated progeny (56). The asymmetry regulator *Trim3* (57) is a 90 dpi RHEG, and Notch ligands appear separately as candidates in males (*Notch1*) and females (*Notch3*). Another probable contributor is cell-cycle

checkpoint resolution, as 90 dpi RHEGs contain many individual checkpoint regulators: *Bap1*, *Brcc3–Anapc7* (coexpressed), *Trp53bp1*, *Nek9*. Last, we leave open the possibility of entirely new regulatory states, which are specifically critical for glioma premalignancy. Around the stem-progenitor RHEG *Lgr6*, for instance, there is an intriguing group of co-clustering RHEGs that includes a survival kinase (*Pim2*), a neural-glial adhesion ligand (*Astn2*), a tumor-promoting lysine methyltransferase (*Smyd2*), and a proto-oncogene (*Mllt11*) (Supplementary Fig. S7). In the brain, murine *Lgr6* expression is restricted to the olfactory bulb (58), providing a hypothesis for why it is a "hotspot" of gliomagenesis in this GEMM.

Gliomas establish simpler, hierarchical ecosystems (7-9) compared to their more advanced GBM counterparts (5). However, our results with one GEMM here suggest that paths to glioma ecosystems are chaotic and wasteful. Despite premalignant expansion, mutant OPC clones remain sparsely distributed up until glioma onset, implying that multiple evolutionary dead-ends must be avoided, perhaps simultaneously, to progress. The bulk and 10-cell transcriptomic resources here give a roadmap for identifying the critical bottlenecks of proneural gliomagenesis. It will be of great interest to examine the generality of these themes in other GEMMs and in human gliomas (7-9). Understanding such regulatory choke points could give insight into strategies for early detection and cancer prevention.

Figures

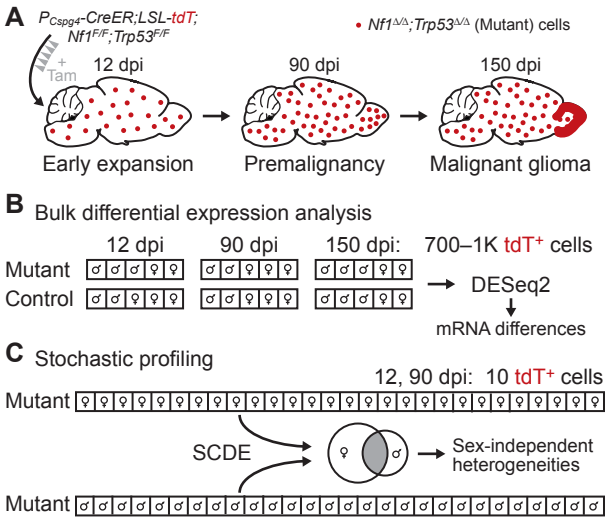


Figure 1. Transcriptomic analysis of bulk expression differences and regulatory heterogeneities in a GEMM of glioma. **A,** Summary of the glioma GEMM. Floxed alleles of *Nf1* and *Trp53* are inducibly deleted by tamoxifen (Tam) activation of CreER localized to oligodendrocyte precursor cells with a Cspg4 promoter (P_{Cspg4}). The resulting $Nf1^{\Delta/\Delta}Trp53^{\Delta/\Delta}$ (mutant) cells are labeled with tdTomato (tdT) and expand in the days post-induction (dpi) for form gliomas preferentially in the olfactory bulb (15). **B,** Experimental plan for bulk differential expression analysis. Large numbers of labeled cells from mutant and control animals were collected by LCM at the indicated dpi, linearly amplified, and sequenced for differential expression (n = 4–7 animals at roughly equal male-female proportion for each group). **C,** Experimental plan for heterogeneity analysis by stochastic profiling (17,19). Many 10-cell pools of mutant cells (n = 28 of each sex) were collected, measured by 10cRNA-seq (18), and assessed for regulatory heterogeneity with the abundance-dependent dispersion module of the SCDE package (38,39). Male and female candidates were intersected to arrive at the final set of regulatory heterogeneities.

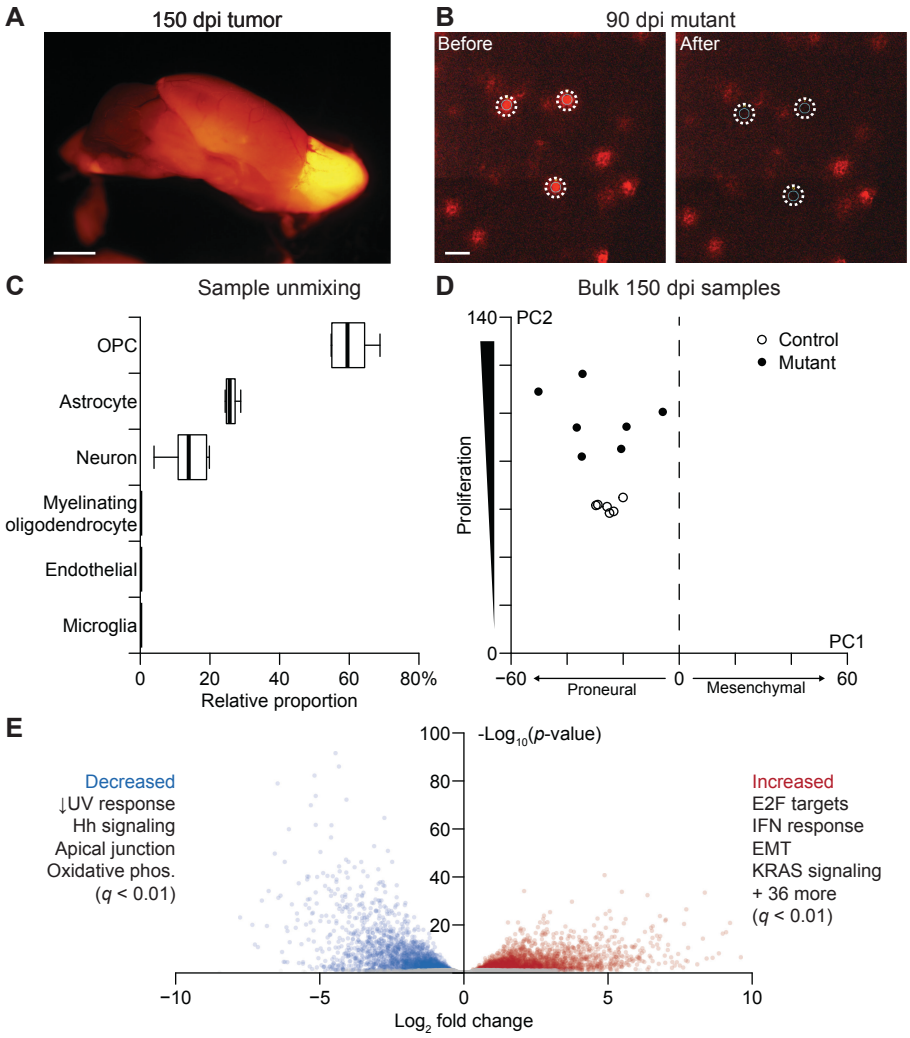


Figure 2. Bulk LCM transcriptomes of labeled gliomas are specific to OPC-derived cells and show gene-expression changes consistent with *Nf1*–*Trp53* deletion. **A,** Macroscopic image of a labeled glioma in the olfactory bulb of a 150 dpi animal. Scale bar is 2 mm. **B,** Microscopic image of 90 dpi mutant OPCs. Images were taken before and after LCM to show spatial fidelity of the isolation. Scale bar is 25 µm. **C,** Relative proportion of the indicated cell types in the control bulk RNA-seq samples estimated by CIBERSORT (22). Boxplots show the median and interquartile range with whiskers indicating the range from $n = 6$ independent samples of control 150 dpi OPCs. **D,** Projection of the bulk 150 dpi mutant and control samples onto principal components capturing proneural–mesenchymal state and relative proliferative status (3). **E,** Volcano plot of differentially expressed genes between tumors and controls. Enriched hallmark pathways among increased and decreased transcripts are summarized, and the complete list is available in Supplementary File S1.

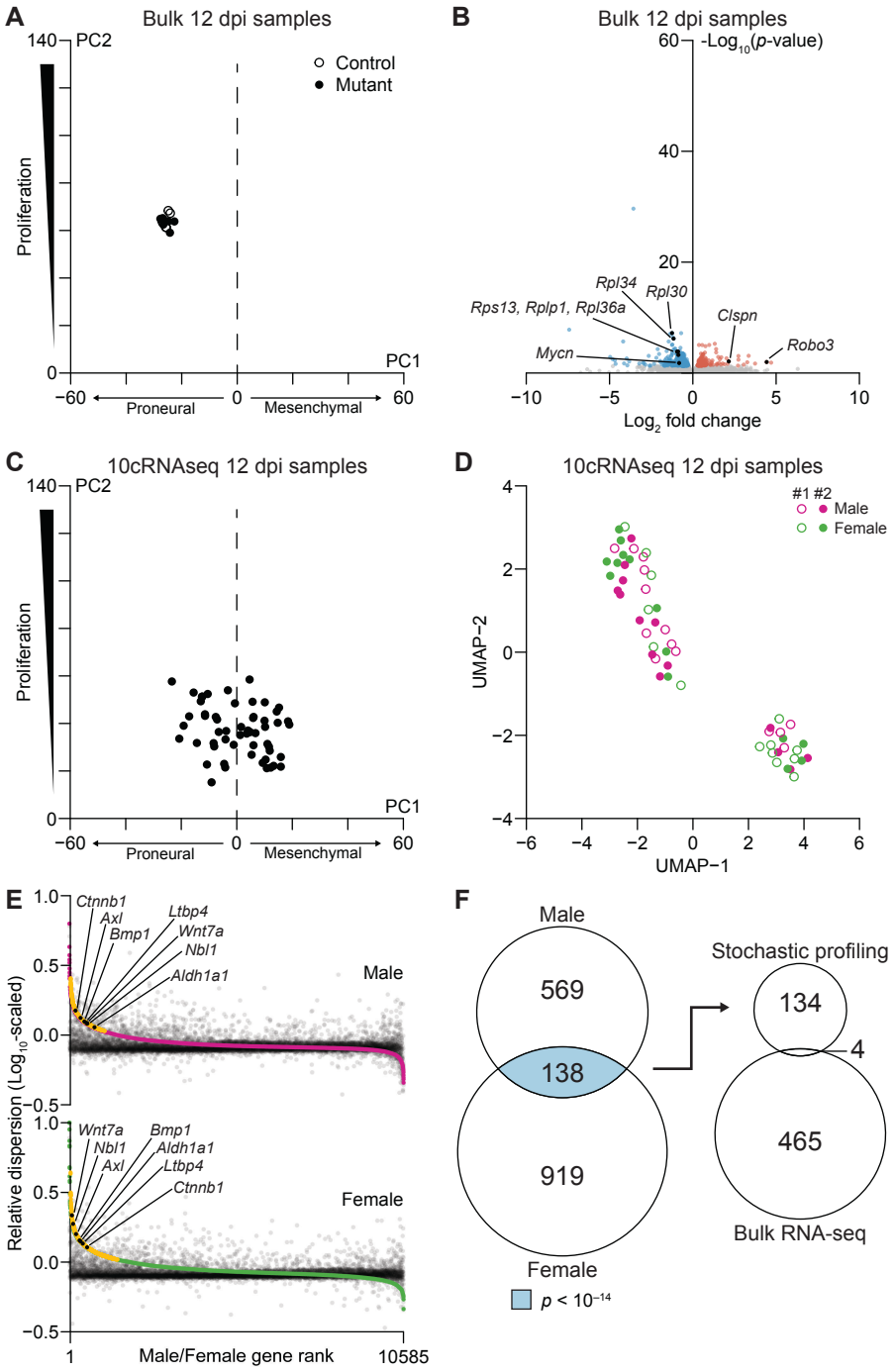

A
Bulk 12 dpi samples
PC2
Control
Mutant
Proliferation
PC1
Proneural
Mesenchymal
B
Bulk 12 dpi samples
-Log10(p-value)
Rpl34
Rpl30
Rps13, Rplp1, Rpl36a
Clspn
Robo3
Mycn
Log2 fold change
C
10cRNAseq 12 dpi samples
PC2
Proliferation
PC1
Proneural
Mesenchymal
D
10cRNAseq 12 dpi samples
#1 #2
Male
Female
UMAP−2
UMAP−1
E
Ctnnb1
Axl
Bmp1
Ltbp4
Wnt7a
Nbl1
Aldh1a1
Male
Female
Relative dispersion (Log10-scaled)
Male/Female gene rank
F
Male
569
138
919
Female
p < 10−14
Stochastic profiling
134
4
465
Bulk RNA-seq

Figure 3. Bulk and 10-cell transcriptomics give orthogonal perspectives on the early adaptations of mutant cells to *Nf1–Trp53* deletion. **A,** Projection of the bulk 12 dpi mutant and control samples onto principal components capturing proneural–mesenchymal state and relative proliferative status (3). **B,** Volcano plot of differentially expressed genes between 12 dpi mutant samples and controls. Enriched hallmark pathways among increased and decreased transcripts are shown along with specific genes discussed in the text. Complete lists are available in Supplementary Files S1 and S2. **C,** Projection of the 10cRNA-seq 12 dpi mutant samples onto principal components capturing proneural–mesenchymal state and relative proliferative status (3). **D,** UMAP projection of 10cRNA-seq observations labeled by sex and mouse sample number. **E,** Abundance-dependence dispersion of pool-and-split controls (black) and 10-cell samples from males (magenta) and females (green). Significantly overdisperse genes for each sex are indicated in yellow, and genes discussed in the text are indicated in black. **F,** Venn diagram (left) intersecting male and female candidates to yield RHEGs (light blue). The overlap with differentially abundant genes (right) was not statistically significant ($p = 0.53$).

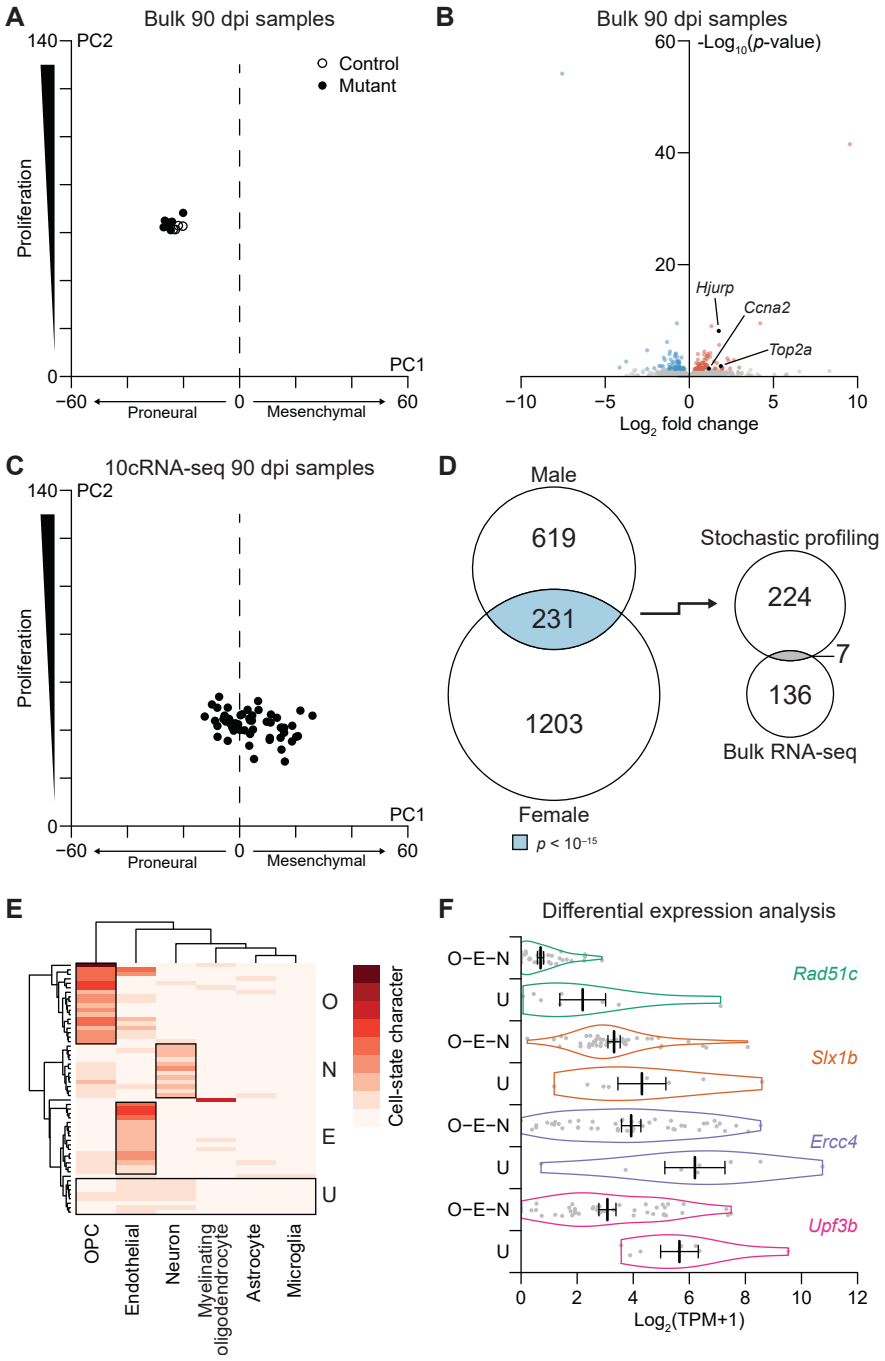

A
Bulk 90 dpi samples
PC2
140
0
Proliferation
Control
Mutant
PC1
−60
Proneural
0
Mesenchymal
60
B
Bulk 90 dpi samples
-Log10(p-value)
60
40
20
Hjurp
Ccna2
Top2a
−10
−5
0
5
10
Log2 fold change
C
10cRNA-seq 90 dpi samples
PC2
140
0
Proliferation
PC1
−60
Proneural
0
Mesenchymal
60
D
Male
619
231
1203
Female
p < 10−15
Stochastic profiling
224
7
136
Bulk RNA-seq
E
O
N
E
U
Cell-state character
OPC
Endothelial
Neuron
Myelinating oligodendrocyte
Astrocyte
Microglia
F
Differential expression analysis
O−E−N
U
Rad51c
Slx1b
Ercc4
Upf3b
0
2
4
6
8
10
12
Log2(TPM+1)

Figure 4. Premalignant mutants exhibit few population-level changes in transcript abundance despite gross alterations in single-cell fate profiles. **A,** Projection of the bulk 90 dpi mutant and control samples onto principal components capturing proneural–mesenchymal state and relative proliferative status (3). **B,** Volcano plot of differentially expressed genes between 90 dpi mutant samples and controls. Enriched hallmark pathways among increased and decreased transcripts are shown along with specific genes discussed in the text. Complete lists are available in Supplementary Files S1 and S2. **C,** Projection of the 10cRNA-seq 12 dpi mutant samples onto principal components capturing proneural–mesenchymal state and relative proliferative status (3). **D,** Venn diagram (left) intersecting male and female candidates to yield RHEGs (light blue). The overlap with differentially abundant genes (right) was marginally significant (gray, $p < 0.05$). **E,** Absolute cell-state character for the indicated marker sets in the 10cRNA-seq samples estimated by CIBERSORT (22). Sample groups with substantial cell-state character of OPCs (O), endothelial cells (E), or neurons (N) are marked along with an undefined group (U). **F,** Violin plots for a subset of transcripts significantly increased in the U group compared to the O–E–N groups. Data are shown as the mean ± s.e.m. of $n = 46$ (O–E–N) and 8 (U) 10-cell observations at 90 dpi.

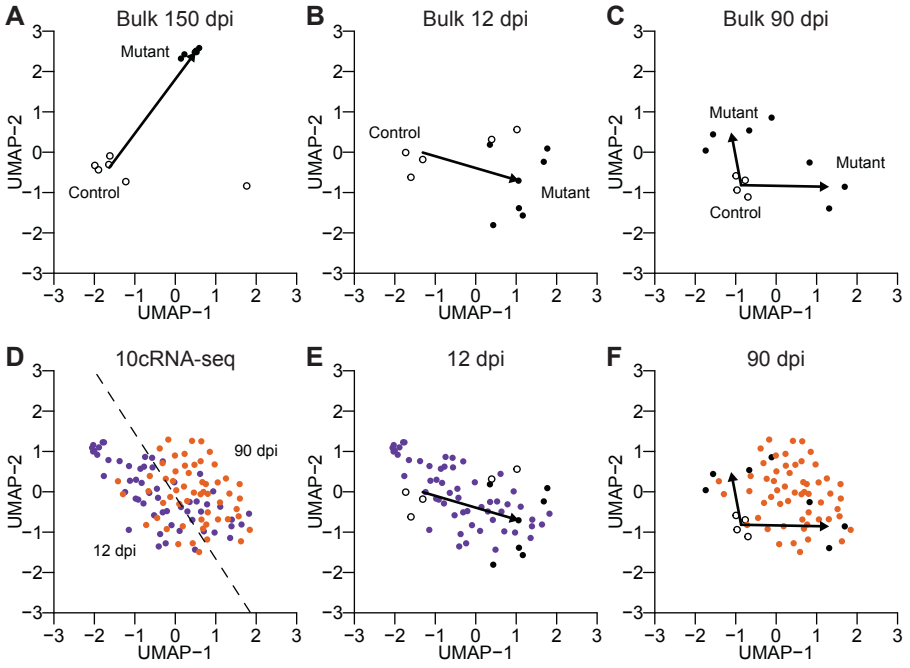
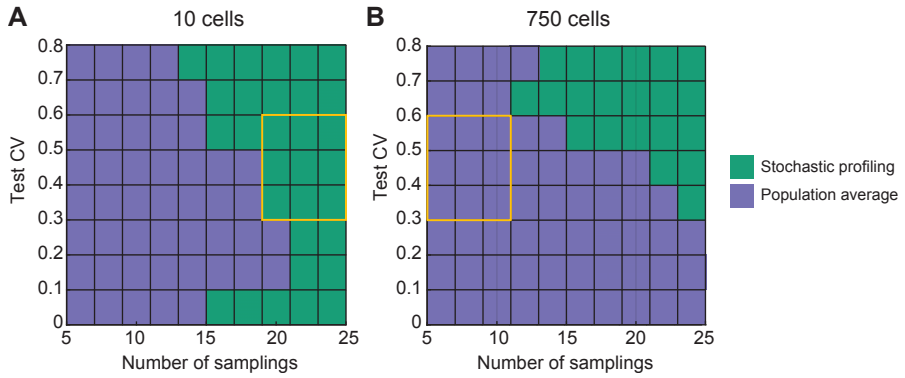
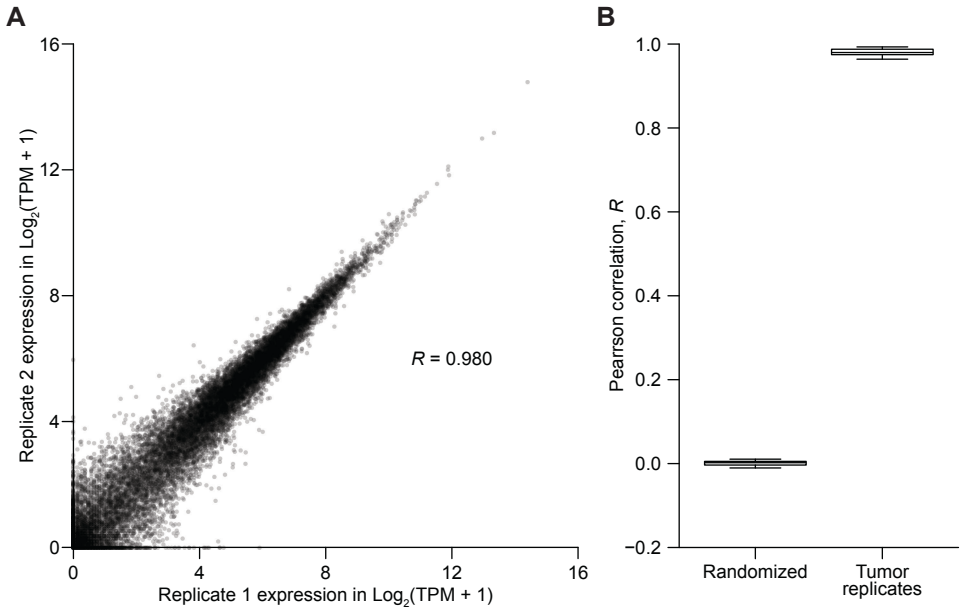


Figure 5. An integrated map of mutant gliomagenesis at the bulk and subpopulation levels. **A–C,** UMAP visualization of control and mutant bulk samples at 150 dpi (**A**), 12 dpi (**B**), and 90 dpi (**C**). Arrows indicate the vector of the control median centroid to the mutant median centroid(s). Two mutant centroids (identified by k-means clustering with two groups) were used at 90 dpi. **D–F,** UMAP visualization of 10cRNA-seq data at 12 dpi (purple) and 90 dpi (orange). The 12 dpi–90 dpi-separating hyperplane (identified by a support vector machine) is shown in **D**. Bulk and 10cRNA-seq samples are replotted together at 12 dpi (**E**) and 90 dpi (**F**).

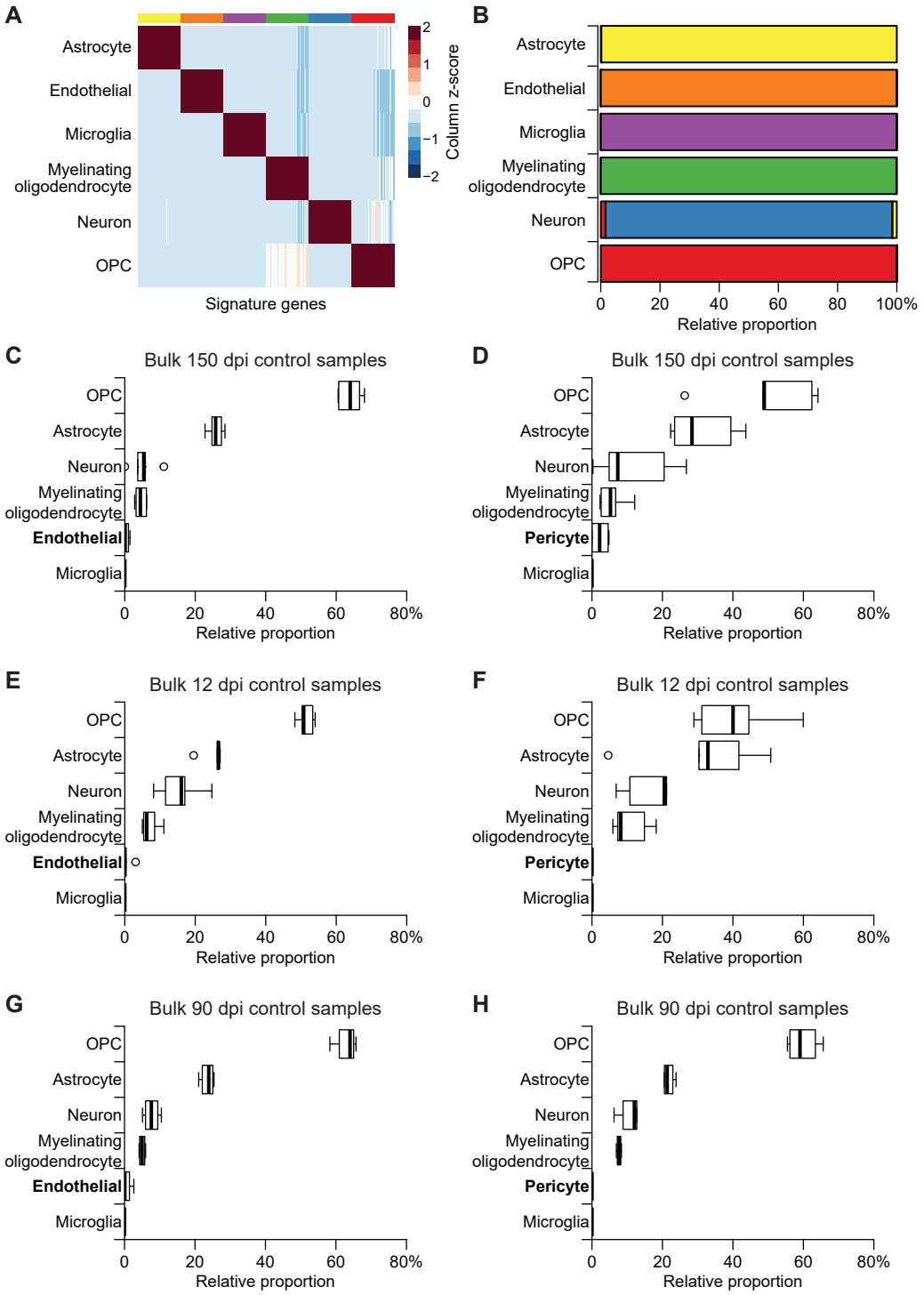
Supplementary Figures



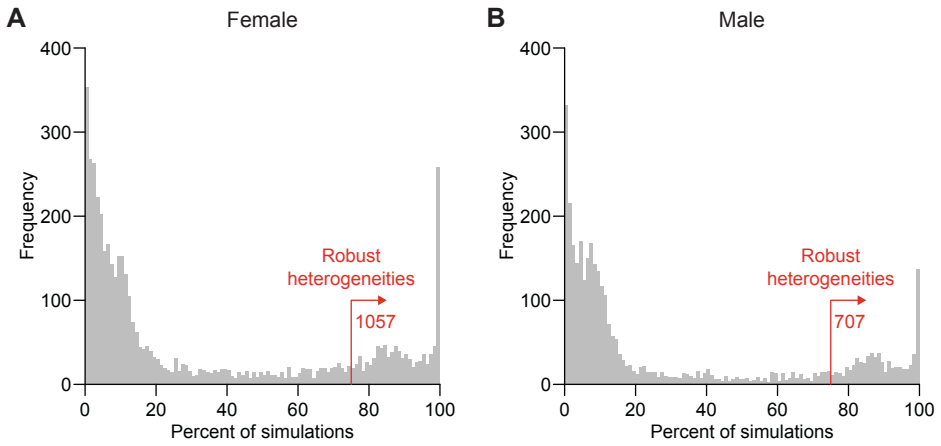
Supplementary Figure S1. Bulk analysis and stochastic profiling occupy different experimental regimes. **A** and **B,** Monte-Carlo simulations of stochastic-profiling experiments were simulated with 10-cell pools (**A**) or 750-cell pools (**B**) with the following parameters: $D = 5$, $F = 0.5$, $\sigma_{b1} = \sigma_{b2} = 0.3$. Yellow boxes highlight the range of experimental samples collected for each regime. See (25) for details.



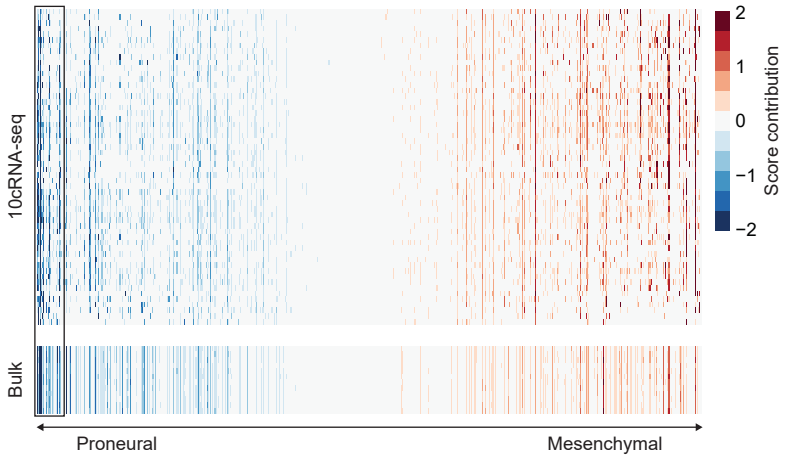
Supplementary Figure S2. RNA sequencing of mutant tumors after linear amplification is highly reproducible. **A,** Correlation plot of two 820-cell samples microdissected from the same 150 dpi mutant tumor and amplified independently. **B,** Boxplot of pairwise correlations within 150 dpi mutant tumors. Randomized transcripts are shown as a negative control. Data are from $n = 7$ pairs of independent samples.


A
Astrocyte
Endothelial
Microglia
Myelinating oligodendrocyte
Neuron
OPC
Signature genes
Column z-score
2
1
0
−1
−2
B
Astrocyte
Endothelial
Microglia
Myelinating oligodendrocyte
Neuron
OPC
0
20
40
60
80
100%
Relative proportion
C
Bulk 150 dpi control samples
OPC
Astrocyte
Neuron
Myelinating oligodendrocyte
Endothelial
Microglia
0
20
40
60
80%
Relative proportion
D
Bulk 150 dpi control samples
OPC
Astrocyte
Neuron
Myelinating oligodendrocyte
Pericyte
Microglia
0
20
40
60
80%
Relative proportion
E
Bulk 12 dpi control samples
OPC
Astrocyte
Neuron
Myelinating oligodendrocyte
Endothelial
Microglia
0
20
40
60
80%
Relative proportion
F
Bulk 12 dpi control samples
OPC
Astrocyte
Neuron
Myelinating oligodendrocyte
Pericyte
Microglia
0
20
40
60
80%
Relative proportion
G
Bulk 90 dpi control samples
OPC
Astrocyte
Neuron
Myelinating oligodendrocyte
Endothelial
Microglia
0
20
40
60
80%
Relative proportion
H
Bulk 90 dpi control samples
OPC
Astrocyte
Neuron
Myelinating oligodendrocyte
Pericyte
Microglia
0
20
40
60
80%
Relative proportion

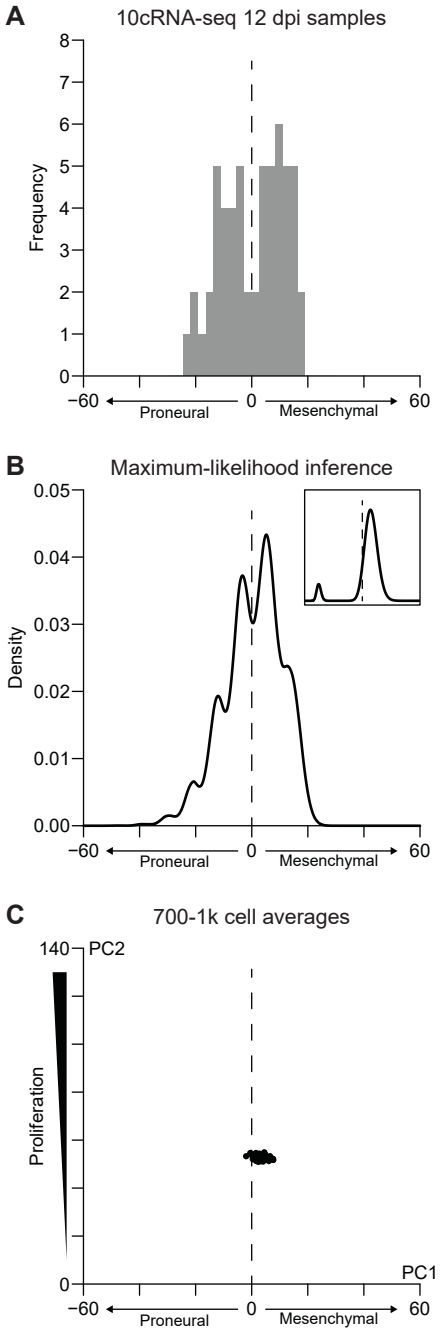
Supplementary Figure S3. OPC enrichment of tdTomato-directed LCM is confirmed by CIBERSORT. **A,** CIBERSORT signature matrix (22) derived from cell type-enriched markers obtained by RNA sequencing (23) and used in **Fig. 2C**. Signature genes were column standardized and hierarchically clustered. **B,** Validation of the derived signature matrix for cleanly deconvolving transcriptomes from sorted populations of the indicated cell types into astrocytes (yellow), endothelial cells (orange), microglia (purple), myelinating oligodendrocytes (green), neurons (blue), and OPCs (red). **C–H,** CIBERSORT deconvolution of normal 150 dpi (**C** and **D**), 12 dpi (**E** and **F**), or 90 dpi (**G** and **H**) control samples with a separate signature matrix using additional RNA sequencing-based markers of pericytes (24) and optimized to swap endothelial cells (**C**, **E**, **G**) for pericytes (**D**, **F**, **H**).



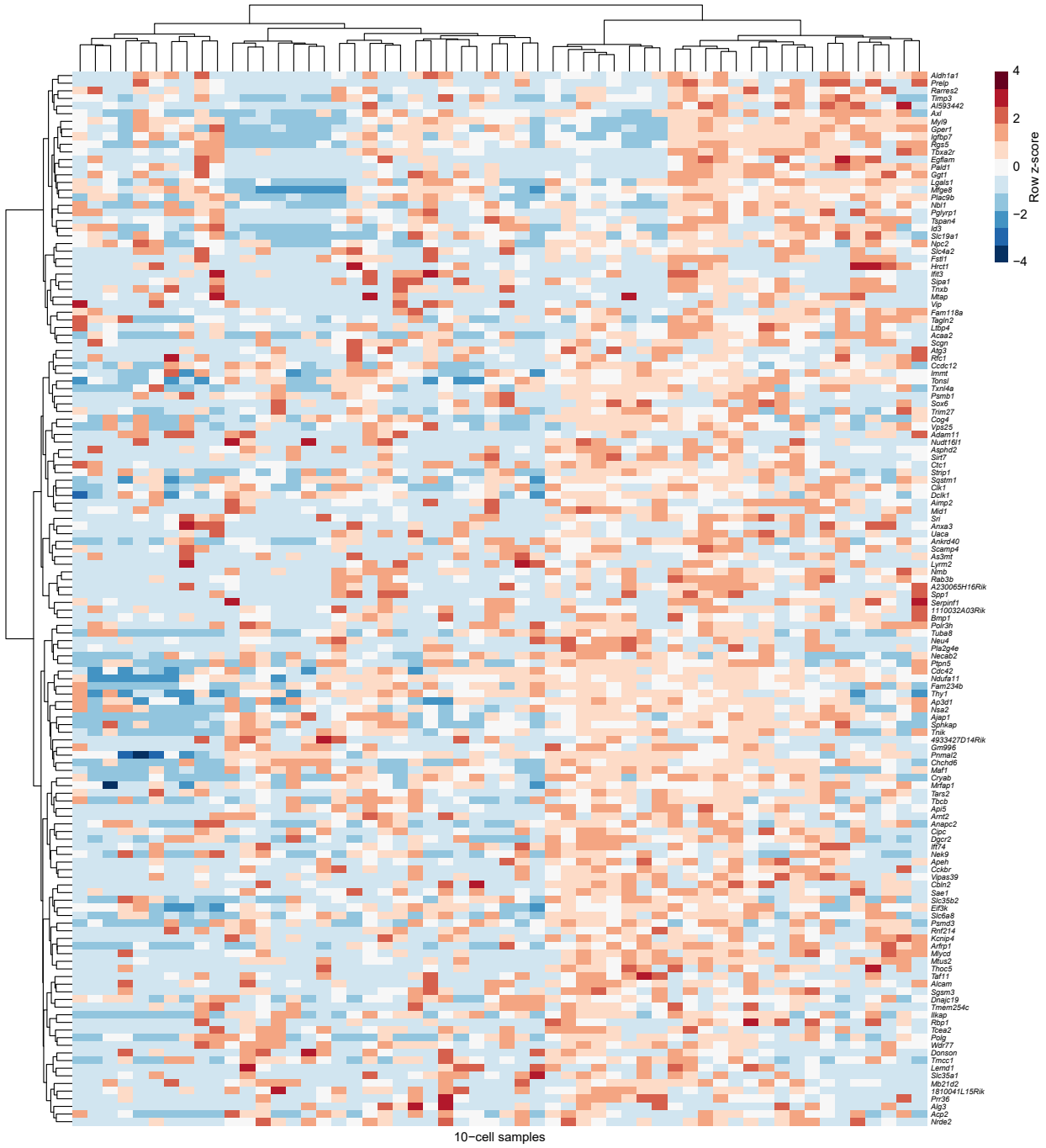
Supplementary Figure S4. Leave-one-out cross-validation identifies robust transcriptional heterogeneities in 12 dpi 10-cell samples. **A** and **B,** Cross-validated dispersion analysis of 24 10-cell observations in females (**A**) or males (**B**). Cross-validation runs removed one 10-cell observation and one pool-and-split control, substituting each with replacement from the remaining samples. Each run was analyzed for overdispersion as described (19), and candidates appearing in >75% of cross-validation runs were considered robust heterogeneities.



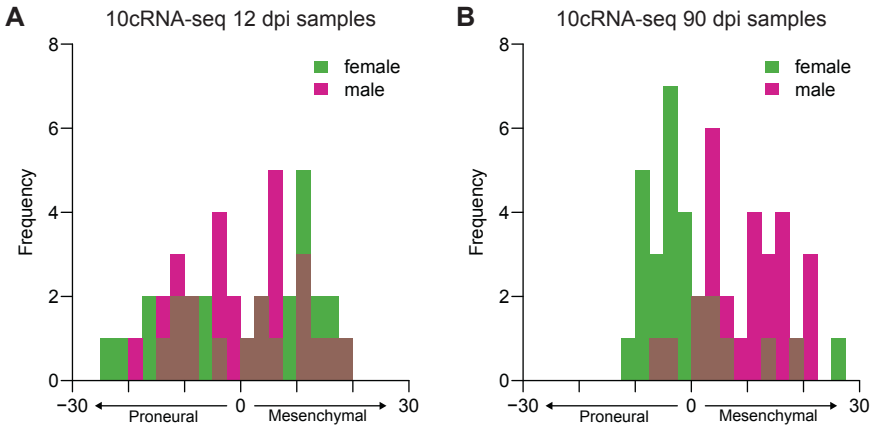
Supplementary Figure S5. Contribution of transcripts to the score along the proneural–mesenchymal principal component in 10cRNA-seq samples and bulk samples at 12 dpi. High-abundance transcripts strongly contributing to the proneural score of bulk samples are boxed.



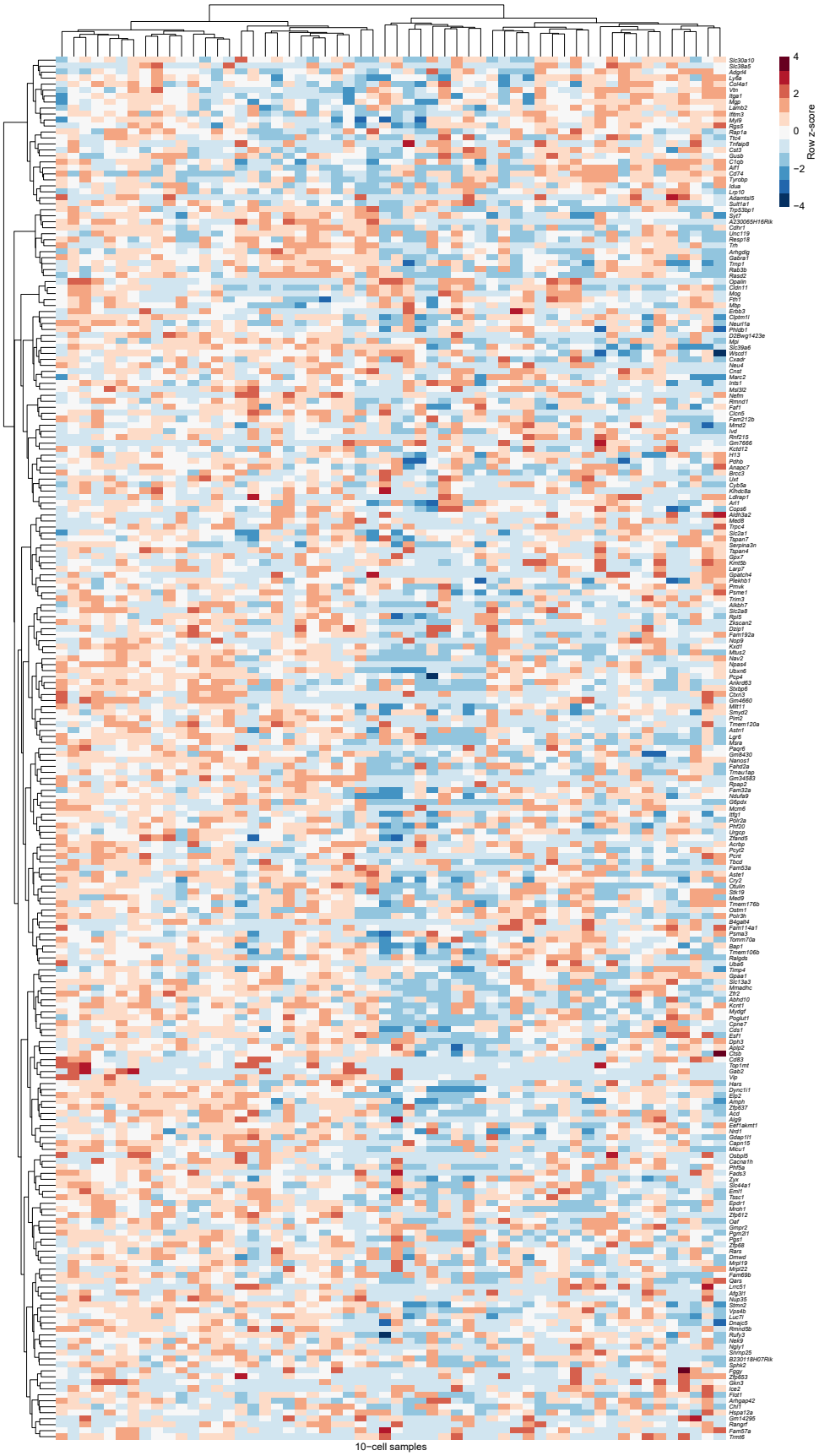
Supplementary Figure S6. Deconvolution of single-cell projections from 10-cell data. **A,** Histogram of 12 dpi 10cRNA-seq projections collapsed along the proneural-mesenchymal axis of **Fig. 3C**. **B,** Single-cell parameterization of 10-cell probability densities estimated by maximum-likelihood inference (21) based on the 10-cell averages in **A**. The inset shows the proportion and variance of the two underlying single-cell distributions. **C,** Simulated projection of 700–1000 cells averaged from the 10-cell projections observed in **Fig. 3C** (see Materials and Methods).



Supplementary Figure S7. Sample-by-sample clustergram of RHEGs in mutant OPCs at 12 dpi.



Supplementary Figure S8. Histogram of 10cRNA-seq projections along the proneural-mesenchymal axis separated by male and female samples. **A** and **B,** Data from **Fig. 3C** (**A**) and **Fig. 4C** (**B**) were collapsed along PC1 and separated by sex. Histograms summarize 10cRNA-seq projections from *n* = 28 10-cell samples from males (magenta) and 28 10-cell samples from females (green).


Row z-score
4
2
0
−2
−4
10−cell samples

Supplementary Figure S9. Sample-by-sample clustergram of RHEGs in mutant OPCs at 90 dpi.

Acknowledgements

We thank Emily Farber and Suna Onengut-Gumuscu at the UVA Center for Public Health Genomics for RNA-seq library preparation and sequencing, Craig Rumpel at the UVA Biorepository and Tissue Research Facility for LCM maintenance and histology services, UVA Research Computing for high-performance computing access and consulting, Henry Pritchard for assistance with NCBI GEO deposition, and Benjamin Purow, Roger Abounader, and Cheryl Borgman for critical evaluation of the manuscript. This work was supported by the National Institutes of Health #R01-CA194470 (K.A.J.), #U01-CA215794 (K.A.J.), and #R01-CA136495 (H.Z.), the David & Lucile Packard Foundation #2009-34710 (K.A.J.), the W. M. Keck Foundation (H.Z.), a UVA Cancer Center support grant #P30-CA044579, the UVA Medical Scientist Training Program (S.S.), and a Wagner Fellowship (S.S.).

References

1. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD*, et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell **2010**;17:98-110
2. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L*, et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell **2017**;32:42-56 e6
3. Wang L, Babikir H, Muller S, Yagnik G, Shamardani K, Catalan F*, et al.* The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. Cancer Discov **2019**;9:1708-19
4. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H*, et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science **2014**;344:1396-401
5. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ*, et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell **2019**;178:835-49 e21
6. Yuan J, Levitin HM, Frattini V, Bush EC, Boyett DM, Samanamud J*, et al.* Single-cell transcriptome analysis of lineage diversity in high-grade glioma. Genome Med **2018**;10:57
7. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K*, et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature **2016**;539:309-13
8. Venteicher AS, Tirosh I, Hebert C, Yizhak K, Neftel C, Filbin MG*, et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science **2017**;355
9. Filbin MG, Tirosh I, Hovestadt V, Shaw ML, Escalante LE, Mathewson ND*, et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. Science **2018**;360:331-5
10. Parsa CF, Hoyt CS, Lesser RL, Weinstein JM, Strother CM, Muci-Mendoza R*, et al.* Spontaneous regression of optic gliomas: thirteen cases documented by serial neuroimaging. Arch Ophthalmol **2001**;119:516-29
11. Network TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature **2008**;455:1061-8
12. Liu C, Sage JC, Miller MR, Verhaak RG, Hippenmeyer S, Vogel H*, et al.* Mosaic analysis with double markers reveals tumor cell of origin in glioma. Cell **2011**;146:209-21
13. Persson AI, Petritsch C, Swartling FJ, Itsara M, Sim FJ, Auvergne R*, et al.* Non-stem cell origin for oligodendroglioma. Cancer Cell **2010**;18:669-82
14. Ledur PF, Liu C, He H, Harris AR, Minussi DC, Zhou HY*, et al.* Culture conditions tailored to the cell of origin are critical for maintaining native properties and tumorigenicity of glioma cells. Neuro Oncol **2016**;18:1413-24
15. Galvao RP, Kasina A, McNeill RS, Harbin JE, Foreman O, Verhaak RG*, et al.* Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process. Proc Natl Acad Sci U S A **2014**;111:E4214-23
16. Gonzalez PP, Kim J, Galvao RP, Cruickshanks N, Abounader R, Zong H. p53 and NF 1 loss plays distinct but complementary roles in glioma initiation and progression. Glia **2018**;66:999-1015
17. Janes KA, Wang CC, Holmberg KJ, Cabral K, Brugge JS. Identifying single-cell molecular programs by stochastic profiling. Nat Methods **2010**;7:311-7
18. Singh S, Wang L, Schaff DL, Sutcliffe MD, Koeppel AF, Kim J*, et al.* In situ 10-cell RNA sequencing in tissue and tumor biopsy samples. Sci Rep **2019**;9:4836

19. Singh S, Sutcliffe MD, Repich K, Atkins KA, Harvey J, Janes KA. Pan-cancer drivers are recurrent transcriptional regulatory heterogeneities in early-stage luminal breast cancer. Cancer Res **2020**:co-submitted
20. Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA. Common effector processing mediates cell-specific responses to stimuli. Nature **2007**;448:604-8
21. Bajikar SS, Fuchs C, Roller A, Theis FJ, Janes KA. Parameterizing cell-to-cell regulatory heterogeneities via stochastic transcriptional profiles. Proc Natl Acad Sci U S A **2014**;111:E626-35
22. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y*, et al.* Robust enumeration of cell subsets from tissue expression profiles. Nat Methods **2015**;12:453-7
23. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S*, et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci **2014**;34:11929-47
24. He L, Vanlandewijck M, Raschperger E, Andaloussi Mae M, Jung B, Lebouvier T*, et al.* Analysis of the brain mural cell transcriptome. Sci Rep **2016**;6:35108
25. Wang L, Janes KA. Stochastic profiling of transcriptional regulatory heterogeneities in tissues, tumors and cultured cells. Nat Protoc **2013**;8:282-301
26. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst **2015**;1:417-25
27. Yang W, Warrington NM, Taylor SJ, Whitmire P, Carrasco E, Singleton KW*, et al.* Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data. Sci Transl Med **2019**;11
28. Marques S, Zeisel A, Codeluppi S, van Bruggen D, Mendanha Falcao A, Xiao L*, et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science **2016**;352:1326-9
29. Yao M, Ventura PB, Jiang Y, Rodriguez FJ, Wang L, Perry JSA*, et al.* Astrocytic trans-Differentiation Completes a Multicellular Paracrine Feedback Loop Required for Medulloblastoma Tumor Growth. Cell **2020**;180:502-20 e19
30. Kurn N, Chen P, Heath JD, Kopf-Sill A, Stephens KM, Wang S. Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications. Clin Chem **2005**;51:1973-81
31. Schaff DL, Singh S, Kim KB, Sutcliffe MD, Park KS, Janes KA. Fragmentation of small-cell lung cancer regulatory states in heterotypic microenvironments. Cancer Res **2020**:co-submitted
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **2014**;15:550
33. Zelina P, Blockus H, Zagar Y, Peres A, Friocourt F, Wu Z*, et al.* Signaling switch of the axon guidance receptor Robo3 during vertebrate evolution. Neuron **2014**;84:1258-72
34. Smits VAJ, Cabrera E, Freire R, Gillespie DA. Claspin - checkpoint adaptor and DNA replication factor. FEBS J **2019**;286:441-55
35. Boon K, Caron HN, van Asperen R, Valentijn L, Hermus MC, van Sluis P*, et al.* N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. EMBO J **2001**;20:1383-93
36. Fambrough D, McClure K, Kazlauskas A, Lander ES. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. Cell **1999**;97:727-41
37. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG*, et al.* Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol **2018**
38. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods **2014**;11:740-2

39. Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL*, et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat Methods **2016**;13:241-4
40. Beiter RM, Fernández-Castañeda A, Rivet-Noor C, Merchak A, Bai R, Slogar E*, et al.* Evidence for oligodendrocyte progenitor cell heterogeneity in the adult mouse brain. bioRxiv **2020**:2020.03.06.981373
41. Binder MD, Xiao J, Kemper D, Ma GZ, Murray SS, Kilpatrick TJ. Gas6 increases myelination by oligodendrocytes and its deficiency delays recovery following cuprizone-induced demyelination. PLoS One **2011**;6:e17727
42. Cheng P, Phillips E, Kim SH, Taylor D, Hielscher T, Puccio L*, et al.* Kinome-wide shRNA screen identifies the receptor tyrosine kinase AXL as a key regulator for mesenchymal glioblastoma stem-like cells. Stem Cell Reports **2015**;4:899-913
43. McKinnon RD, Piras G, Ida JA, Jr., Dubois-Dalcq M. A role for TGF-beta in oligodendrocyte differentiation. J Cell Biol **1993**;121:1397-407
44. Petersen MA, Ryu JK, Chang KJ, Etxeberria A, Bardehle S, Mendiola AS*, et al.* Fibrinogen Activates BMP Signaling in Oligodendrocyte Progenitor Cells and Inhibits Remyelination after Vascular Damage. Neuron **2017**;96:1003-12 e7
45. Nolan K, Thompson TB. The DAN family: modulators of TGF-beta signaling and beyond. Protein Sci **2014**;23:999-1012
46. Pereira EJ, Burns JS, Lee CY, Marohl T, Calderon D, Wang L*, et al.* Sporadic activation of an oxidative stress-dependent NRF2–p53 signaling network in breast epithelial spheroids and premalignancies. bioRxiv **2019**:862474
47. Pitts RL, Wang S, Jones EA, Symes AJ. Transforming growth factor-beta and ciliary neurotrophic factor synergistically induce vasoactive intestinal peptide gene expression through the cooperation of Smad, STAT, and AP-1 sites. J Biol Chem **2001**;276:19966-73
48. Dai ZM, Sun S, Wang C, Huang H, Hu X, Zhang Z*, et al.* Stage-specific regulation of oligodendrocyte development by Wnt/beta-catenin signaling. J Neurosci **2014**;34:8467-73
49. Marques S, van Bruggen D, Vanichkina DP, Floriddia EM, Munguba H, Varemo L*, et al.* Transcriptional Convergence of Oligodendrocyte Lineage Progenitors during Development. Dev Cell **2018**;46:504-17 e7
50. Lykke-Andersen J, Shu MD, Steitz JA. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell **2000**;103:1121-31
51. Liu Y, Masson JY, Shah R, O'Regan P, West SC. RAD51C is required for Holliday junction processing in mammalian cells. Science **2004**;303:243-6
52. Fekairi S, Scaglione S, Chahwan C, Taylor ER, Tissier A, Coulon S*, et al.* Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. Cell **2009**;138:78-89
53. Yang P, Kollmeyer TM, Buckner K, Bamlet W, Ballman KV, Jenkins RB. Polymorphisms in GLTSCR1 and ERCC2 are associated with the development of oligodendrogliomas. Cancer **2005**;103:2363-72
54. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ*, et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science **2016**;352:189-96
55. Sun T, Warrington NM, Luo J, Brooks MD, Dahiya S, Snyder SC*, et al.* Sexually dimorphic RB inactivation underlies mesenchymal glioblastoma prevalence in males. J Clin Invest **2014**;124:4123-33

56. Sugiarto S, Persson AI, Munoz EG, Waldhuber M, Lamagna C, Andor N*, et al.* Asymmetry-defective oligodendrocyte progenitors are glioma precursors. Cancer Cell **2011**;20:328-40
57. Chen G, Kong J, Tucker-Burden C, Anand M, Rong Y, Rahman F*, et al.* Human Brat ortholog TRIM3 is a tumor suppressor that regulates asymmetric cell division in glioblastoma. Cancer Res **2014**;74:4536-48
58. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A*, et al.* Genome-wide atlas of gene expression in the adult mouse brain. Nature **2007**;445:168-76