

UCSF

UC San Francisco Previously Published Works

Title

The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation

Permalink

<https://escholarship.org/uc/item/7h82g77j>

Journal

Cancer Discovery, 9(12)

ISSN

2159-8274

Authors

Wang, Lin

Babikir, Husam

Müller, Sören

et al.

Publication Date

2019-12-01

DOI

10.1158/2159-8290.cd-19-0329

Peer reviewed



Published in final edited form as:

Cancer Discov. 2019 December ; 9(12): 1708–1719. doi:10.1158/2159-8290.CD-19-0329.

The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation

Lin Wang^{#1}, Husam Babikir^{#1}, Sören Müller^{#1}, Garima Yagnik¹, Karin Shamardani¹, Francisca Catalan¹, Gary Kohanbash², Beatriz Alvarado^{1,3}, Elizabeth Di Lullo³, Arnold Kriegstein³, Sumedh Shah¹, Harsh Wadhwa¹, Susan M. Chang¹, Joanna J. Phillips¹, Manish K. Aghi¹, Aaron A. Diaz¹

¹Department of Neurological Surgery, University of California, San Francisco, San Francisco, California.

²Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, Pennsylvania.

³Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, California.

These authors contributed equally to this work.

Abstract

Although tumor-propagating cells can be derived from glioblastomas (GBM) of the proneural and mesenchymal subtypes, a glioma stem-like cell (GSC) of the classic subtype has not been identified. It is unclear whether mesenchymal GSCs (mGSC) and/or proneural GSCs (pGSC) alone are sufficient to generate the heterogeneity observed in GBM. We performed single-cell/single-nucleus RNA sequencing of 28 gliomas, and single-cell ATAC sequencing for 8 cases. We found that GBM GSCs reside on a single axis of variation, ranging from proneural to mesenchymal. *In silico* lineage tracing using both transcriptomics and genetics supports mGSCs as the progenitors of pGSCs. Dual inhibition of pGSC-enriched and mGSC-enriched growth and survival pathways provides a more complete treatment than combinations targeting one GSC phenotype alone. This study sheds light on a long-standing debate regarding lineage relationships

Corresponding Author: Aaron A. Diaz, University of California, San Francisco, 1450 3rd Street, San Francisco, CA 94158. Phone: 415-514-0408; aaron.diaz@ucsf.edu.

Authors' Contributions

Conception and design: H. Babikir, A. Kriegstein, A.A. Diaz

Development of methodology: L. Wang, H. Babikir, A. Kriegstein

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G. Yagnik, K. Shamardani, B. Alvarado, E. Di Lullo, A. Kriegstein, S. Shah, H. Wadhwa, S.M. Chang, J.J. Phillips, M.K. Aghi

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Wang, H. Babikir, S. Müller, K. Shamardani, F. Catalan, G. Kohanbash, A. Kriegstein, S. Shah, A.A. Diaz

Writing, review, and/or revision of the manuscript: L. Wang, H. Babikir, G. Kohanbash, A. Kriegstein, S. Shah, S.M. Chang, J.J. Phillips, A.A. Diaz

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K. Shamardani, F. Catalan, A. Kriegstein, M.K. Aghi

Study supervision: M.K. Aghi, A.A. Diaz

Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

Disclosure of Potential Conflicts of Interest

S. Müller is a scientist at and has ownership interest (including patents) in Genentech. No potential conflicts of interest were disclosed by the other authors.

among GSCs and presents a paradigm by which personalized combination therapies can be derived from single-cell RNA signatures, to overcome intratumor heterogeneity.

INTRODUCTION

Malignant gliomas are the most common primary tumors of the adult brain and are essentially incurable. Glioma genetics have been studied extensively, and yet targeted therapeutics have produced limited results. Glioblastomas (GBM) have been classified into subtypes based on gene expression (1). However, we and others have shown that GBMs contain heterogeneous mixtures of cells of distinct transcriptomic subtypes (2, 3). This intratumor heterogeneity is at least partially to blame for the failures of targeted therapies.

Tumor-propagating cells expressing markers of the mesenchymal and proneural transcriptomic subtypes can be derived from GBMs (4). However, no glioma stem-like cell (GSC) of the classic subtype has been convincingly identified. The lineage relationship between proneural GSCs (pGSC) and mesenchymal GSCs (mGSC) is unknown. Surprisingly little is known about the cellular progeny of GSCs *in vivo*. It is unclear whether pGSCs and/or mGSCs are sufficient to generate the heterogeneity observed in GBM.

We performed single-cell RNA sequencing (scRNA-seq), single-nucleus RNA sequencing (snRNA-seq), single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq), and whole-exome DNA sequencing (exome-seq) of specimens from untreated human gliomas. Via *in silico* transcriptomic and genetic lineage tracing of these data, we defined the lineage relationships between glioma cell types. Integrating this with meta-analysis of sequencing data from The Cancer Genome Atlas (TCGA; <https://cancergenome.nih.gov/>) and spatial data from The Ivy Foundation Glioblastoma Atlas Project (Ivy GAP; <http://glioblastoma.alleninstitute.org/>), we mapped glioma cells to analogous cell types in the developing brain and to specific tumor anatomic structures. From the scATAC-seq, we elucidated cell type-specific *cis*-regulatory grammars and associated transcription factors. Using IHC and automated image analysis of human GBM microarrays, we validated learned phenotypes at the protein level. Finally, we performed an *in vitro* screen of drug combinations that target genes identified from our single-cell analysis of GSCs.

We show that proliferating IDH-wild-type GBM cells can be described by a single axis of gene signature, which ranges from proneural to mesenchymal. At the extremes of this axis reside stem-like cells that express canonical markers of mGSCs and pGSCs. We identify an mGSC signature that correlates with significantly inferior survival in IDH-wild-type GBM. Our analysis shows that mGSCs, pGSCs, their progeny, and stromal/immune cells are sufficient to explain the heterogeneity observed in GBM. We show that combination therapies that address intratumor heterogeneity can be designed on the basis of single-cell RNA signatures.

RESULTS

Single-Cell mRNA and Bulk DNA Profiling of Untreated Human Gliomas

We applied scRNA-seq or snRNA-seq to biopsies from 22 primary untreated human IDH-wild-type GBMs and 6 primary untreated IDH-mutant gliomas (Supplementary Table S1). Our goal was to profile both for cellular coverage (to survey cellular phenotypes) and for transcript coverage (to compare genetics). Therefore, we performed sc/snRNA-seq for 19 samples via the 10× Genomics Chromium platform (10×) and obtained 3′-sequencing data for 31,281 cells after quality control. scRNA-seq for 3 cases was done using the Fluidigm C1 platform (C1), which yielded full-transcript coverage for 291 cells. We incorporated 4 more published cases from our C1 pipeline, adding 384 cells (3). For 6 of the 10× cases and 5 of the C1 cases, the biopsies were minced and split, and both scRNA-seq and exome-seq were performed (Supplementary Table S1).

We applied our pipeline for sc/snRNA-seq preprocessing (5), quantification of expressed mutations (3, 6), and cell-type identification (7, 8). We identified 10,816 tumor-infiltrating stromal and immune cells based on expressed mutations, clustering, and canonical marker genes (Fig. 1A; Supplementary Fig. S1A). We termed the remaining 20,465 cells neoplastic, as they expressed clonal malignant mutations. Only neoplastic cells were used for all subsequent analyses.

The Transcriptional Phenotypes of Proliferating IDH-Wild-Type GBM Neoplastic Cells Can Be Explained by a Single Axis That Varies from Proneural to Mesenchymal

Unbiased principal component analyses (PCA) of the IDH-wild-type GBM cases revealed two patient-independent clusters of neoplastic cells consistent across all three platforms (Fig. 1B–I; Supplementary Table S2). A differential expression test between clusters identified markers of the proneural (e.g., *ASCL1*, *OLIG2*) and mesenchymal (e.g., *CD44*, *CHI3L1*) subtypes as significant (Supplementary Table S3). Mesenchymal cells significantly overexpress markers of response to hypoxia (e.g., *HIF1A*) and cytokines that promote myeloid-cell chemotaxis (e.g., *CSF1*, *CCL2*, *CXCL2*). However, mesenchymal cells do not express high levels of *MKI67* or other markers of cell-cycle progression. Conversely, proneural cells express high levels of *MKI67* as well as cyclin-dependent kinase genes. We estimated the fraction of actively cycling cells using the Seurat package (9). By this metric, 21% to 30% of proneural cells are cycling compared with 0.3% to 10% of mesenchymal cells (Fig. 1D and G). Importantly, all our clinical specimens contain cells of both phenotypes: proliferating proneural cells and mesenchymal cells with a quiescent, cytokine-secretory phenotype (Supplementary Fig. S1B).

In all PCAs, cells at the left and right extremes of principal component 1 (PC1) express high levels pGSC or mGSC markers (4), respectively, as indicated by PC1 gene loadings (Fig. 1B and E). We therefore interpreted the top-loading genes from each direction as representing pGSC and mGSC gene signatures. We used those signatures to score all cells for stemness, controlling for technical variation as described previously (9, 10). In all PCAs, principal component 2 (PC2) is loaded by markers of cell-cycle progression (e.g., *MKI67*). Thus,

although stemness correlates with cell cycle for both proneural and mesenchymal cells, more pGSCs than mGSCs express markers of cell-cycle progression, and at higher levels.

IDH-Mutant Gliomas Are Composed of Cycling Stem-Like Cells, Noncycling Astrocyte-Like, and Oligodendrocyte-Like Cells

A PCA analysis of the IDH-mutant gliomas identified 3 patient-independent populations (Supplementary Fig. S1C–S1F). Two of the clusters fall on opposite ends of PC1. They differentially express markers (e.g., *OLIG2*, *GFAP*) of astrocyte-like (IDH-A) and oligodendrocyte-like (IDH-O) cell types, as has been previously described in scRNA-seq of IDH-mutant glioma (11). PC2 is positively loaded by cell-cycle genes, such as *MKI67*. A third cluster of cells has high PC2 sample scores and specifically expresses markers of IDH-mutant GSCs (10, 11), as well as genes associated with cancer stem cells not previously described in IDH-mutant glioma (Supplementary Table S2). We term this cluster IDH-S following the convention of Venteicher and colleagues (11). Most cells in IDH-S have low PC1 sample scores, and do not express IDH-A or IDH-O genes. Almost all cycling cells are found in IDH-S (Supplementary Fig. S1G).

IDH-Wild-Type GBM Cells Are Stratified by Differentiation Gradients Observed in Gliogenesis

In our differential expression test, we observed that proneural cells specifically express markers of the oligodendrocyte lineage (e.g., *OLIG2*, *SOX10*), whereas mesenchymal cells instead express markers of astrocytes (e.g., *GFAP*, *AQP4*). We evaluated the mGSC and pGSC gene signatures in scRNA-seq of glia from fetal and adult human brain (12, 13). We found that pGSC signature genes are enriched in oligodendrocyte progenitor cells. The mGSC signature, however, is comprised of genes expressed by neural stem cells as well as markers of astrocytes (Fig. 2A).

In addition to mGSCs and pGSCs, we find neoplastic cells (possessing clonal malignant mutations) that do not express stemness or cell-cycle signatures above background levels. Instead they express either markers of differentiated astrocytes (e.g., *ALDOC*) or differentiated oligodendrocytes (e.g., *MAG*, *MOG*). Although the GSCs express high levels of positive WNT pathway regulators, these more differentiated cells express high levels of WNT pathway agonists (Supplementary Fig. S2A). Thus, GBMs contain mesenchymal and proneural populations that align with astrocyte and oligodendrocyte differentiation gradients, respectively.

PGSCs, mGSCs, Their Differentiated Progeny, and Stromal/Immune Cells Explain the Phenotypic Heterogeneity Observed in GBM

We found that our mGSC and pGSC gene signatures are coexpressed across TCGA datasets (Fig. 2B; Supplementary Fig. S2B). Although mGSC and pGSC signature genes are correlated among themselves, the mGSC and pGSC signatures are anticorrelated with each other. Moreover, the cell-cycle signature more strongly correlates in TCGA data with the pGSC signature than the mGSC signature, consistent with our scRNA-seq data (Supplementary Fig. S2C).

Using our scRNA-seq data and published scRNA-seq data from human brain tissue (12, 13) as a basis, we pooled reads across cells of the same type. This yielded data-driven profiles for mGSCs, pGSCs, astrocytes, oligodendrocytes, neurons, endothelial cells, myeloid cells, and T cells. We then used these cell-type signatures as predictors in a linear regression model. We fit our model to each TCGA GBM RNA-seq dataset individually. A heat map of cell type-specific genes (Fig. 2C) agreed with our regression analysis (Fig. 2D). We found that samples of both the mesenchymal and classic Verhaak subtypes are enriched for mGSCs and depleted of pGSCs. Whereas classic samples are distinguished by higher infiltration of astrocytes, mesenchymal samples contain high levels of infiltrating immune cells. Proneural samples are characterized by the highest levels of pGSCs, oligodendrocytes, and neurons. In summary, the full spectrum of heterogeneity observed in TCGA GBM data can be explained by varying proportions of pGSCs and mGSCs, their differentiated progeny, and infiltrating stromal/immune content. Cox regression analysis identifies the mGSC signature as correlating with significantly inferior survival in TCGA data. However, pGSC content is not prognostic (Fig. 2E).

GSC Niche Localization and Microenvironment Interactions Elucidated Using Reference Atlases and Quantitative IHC

We compared our mGSC, pGSC, and cell-cycle signatures to RNA sequencing from the Ivy GAP (Fig. 2F and G). The Ivy GAP has annotated, microdissected, and RNA-sequenced GBM anatomic structures from human specimens. We found that mGSCs are enriched in hypoxic regions. pGSCs are enriched in the tumor's leading edge and in regions of diffuse infiltration of tumor-adjacent white matter.

To visualize and quantify the associations between GSCs, the cell cycle, and hypoxia, we performed IHC for CD44 (mGSCs), DLL3 (pGSCs), CA9 (hypoxia), and Ki-67 (cell cycle) on GBM microarrays (Fig. 2H; Supplementary Table S4) and nonmalignant controls (Supplementary Fig. S2D). We did not find any cycling CD44⁺ cells in our samples, whereas approximately 5% of DLL3⁺ cells expressed Ki-67. On the other hand, CD44⁺ cells colocalized with CA9 at 2-fold greater frequency than DLL3⁺ cells. This dovetails with our findings in scRNA-seq, TCGA, and GAP data, which show that pGSCs are more proliferative whereas mGSCs are enriched in hypoxic regions.

Genetic and Transcriptomic *In Silico* Lineage Tracing of GBM GSCs Supports a Mesenchymal-to-Proneural Hierarchy

We identified all cycling cells from our IDH-wild-type GBM datasets (see Methods). These cells were then used to perform RNA velocity analysis via *velocyto* (14). This approach compares the fractions of spliced to unspliced transcripts per gene to estimate the time derivative of gene abundance in single cells. When projected back onto a PCA axis (Fig. 3A), we found that RNA velocity supports stable mGSC and pGSC populations as well as an intermediate population that gradually transitions from a mesenchymal to proneural phenotype. For example, in this intermediate population, mGSC markers (e.g., *CD44*, *CHI3L1*) are highly expressed, but have a negative time derivative; conversely, pGSC markers (e.g., *DLL3*, *PDGFRA*) are lowly expressed but with a high, positive velocity.

Mitochondrially expressed mutations have recently been identified as a reliable way to perform genetic lineage tracing due to the robust coverage of mitochondrial genes, even in scRNA-seq (15). Following this approach, for each sample we pooled cells and identified mitochondrial single-nucleotide variants (SNV; see Methods). For each cell, we then annotated the percentage of its sample's mitochondrial mutations it harbored. We found that the percentage of expressed mitochondrial mutations correlated with progression from the mesenchymal phenotype (Fig. 3B–D). Thus, both genetic and transcriptomic lineage tracing supports a mesenchymal-to-proneural hierarchy.

We also compared chromosomal SNVs and CNVs in our scRNA-seq data that were validated by exome-seq data via our previously described approach (3, 6). We restricted ourselves to mutations that occurred at a minimum of 10% variant allele frequency in exome-seq and identified cells in our scRNA-seq that expressed these mutations. For all patients, we found that all expressed, validated mutations are present in the GSCs (Supplementary Fig. S3A). This shows that the GSCs are sufficient to explain the genetic heterogeneity observed in our tumor specimens.

IDH-Mutant and Wild-Type GSCs Share a Core Signature That Is Prognostic

RNA-velocity analysis of IDH-mutant glioma datasets revealed a hierarchy with a single root focused on the IDH-S population and two terminal points corresponding to differentiated IDH-A and IDH-O populations (Supplementary Fig. S3B and S3C). We compared gene loadings of PC2 between the original PCAs of our IDH-wild-type and IDH-mutant gliomas. This identified a core signature of stemness enriched in the cycling cells of both diseases (Supplementary Fig. S3D). This signature is prognostic in a combined cohort of IDH-mutant grade II/III oligodendrogliomas and astrocytomas (Supplementary Fig. S3E). An analysis of expressed mitochondrial mutations showed that mitochondrial mutational load correlated with progression along the IDH-O and IDH-A lineages (Supplementary Fig. S3F and S3G), further supporting a cellular hierarchy with IDH-S at the apex.

scATAC-seq of Human Gliomas Identifies Cell Type-Specific Regulatory Grammars and Supports a Single-Axis Hypothesis

We performed scATAC-seq on 4 IDH-wild-type GBMs, 2 IDH-mutant grade II astrocytomas, and 2 IDH-mutant oligodendrogliomas. We considered the IDH-wild-type and IDH-mutant samples separately. IDH-wild-type GBM cells subjected to scATAC-seq were triaged according to the presence of observed genomic alterations (Fig. 4A; see Methods). For each cell and each gene, we computed a gene-body activity score (i.e., the gene-wise sum of scATAC-seq read-counts) using Seurat v3. We then clustered gene-body activity scores separately for neoplastic and nonneoplastic cells using Seurat. The gene-body activities of nonneoplastic cells were readily separated by canonical markers of myeloid, glial, and endothelial cells (Fig. 4B).

When we clustered the gene activity scores of neoplastic cells, we found 3 clusters of cells. We then performed differential motif enrichment tests between these groups of cells to identify enriched regulatory motifs and associated transcription factors (Fig. 4C–G; Supplementary Table S5; see Methods).

In the first cluster, the proneural bHLH genes *OLIG2*, *ASCL1*, and *NEUROG2* were enriched and both their gene-activity scores and representative motif frequencies were significant. The second cluster showed enrichment for mesenchymal makers (e.g., *CD44*, *STAT3*). When we compared the third cluster to the mesenchymal cluster, proneural markers were differentially enriched (Fig. 4E). Likewise, when we compared the third cluster to the proneural cluster, we saw differential enrichment of mesenchymal markers (Fig. 4F). Moreover, the most cluster-specific gene activities identified for this third cluster showed partial overlap with both the mesenchymal and proneural clusters, for example, *CD44* and *CDH9* (Fig. 4C; Supplementary Table S4). We plotted the expression of the cluster-specific genes from the third cluster on top of a PCA plot of cycling cells. We found those genes expressed a population that was intermediate to the mesenchymal and proneural populations in our RNA-velocity and mitochondrial-mutation analysis (e.g., *MET*; Supplementary Fig. S4A). Thus, we interpret this third cluster as an intermediate population between mesenchymal and proneural clusters.

IDH-mutant glioma cells were likewise separated by the presence of clonal mutations into neoplastic cells and nonmalignant glial/immune cells (Supplementary Fig. S4B and S4C). Clustering by gene-activity scores identified three clusters corresponding to the IDH-S, IDH-O, and IDH-A cell types (Supplementary Fig. S4D–S4G). Motif enrichment in each cluster was assessed to determine cell type-specific regulatory grammars (Supplementary Fig. S4H).

A Drug Combination Screen Based on Single-Cell Signatures

We identified FDA-approved drugs that target genes found to be specific to pGSCs and mGSCs in our single-cell analysis (Fig. 5A; Supplementary Fig. S5A). Drugs were screened in combination to assess synergy *in vitro* (see Methods). Only combinations that targeted both the pGSC and mGSC phenotypes were found to be synergistic (Fig. 5B; Supplementary Table S6). In particular, inhibition of EGFR and FGFR3 growth factor receptors did not synergize with WNT inhibition (all mGSC-specific targets), but targeting either EGFR, FGFR3, or WNT synergized with inhibition of pGSC-specific Survivin.

DISCUSSION

It is known that an individual tumor may contain multiple GSC clones (16). We applied exome-seq, sc/snRNA-seq, and scATAC-seq to human tumor specimens and found that all GBMs contain hierarchies of mesenchymal and proneural GSCs and their more differentiated progeny. We also observe these same cellular hierarchies in scRNA-seq profiling of recurrent GBM following treatment (Supplementary Fig. S5B). However, our study was limited to the analysis of primary tumors. Additional studies will be required to determine whether standard therapy exerts a selection pressure on this hierarchy.

While our manuscript was in revisions, Neftel and colleagues published a model of GBM heterogeneity based on scRNA-seq analysis of adult and pediatric human specimens (17). Neftel and colleagues describe a model with four cellular states: astrocyte and mesenchymal cell types enriched in Verhaak classic and mesenchymal TCGA samples and NPC- and OPC-like cell types enriched in proneural TCGA samples. Their data corroborate a

mesenchymal-to-proneural axis, and anticorrelation of mesenchymal (e.g., *CD44*) and proneural (e.g., *DLL3*) genes in neoplastic cells (see ref. 17; Fig. 2C). Both studies contribute to recent advances that use transcriptomics to identify clinically relevant phenotypes (18, 19).

Murine GBMs can be separated into two cell populations that have different capacities for tumorigenicity and self-renewal (20). The Id1^{hi}/Olig2⁻ and Id1^{lo}/Olig2⁺ populations found in the model of Barrett and colleagues (20) match the expression signatures of mGSCs and pGSCs, respectively. Moreover, the finding of Barrett and colleagues (20) that Id1^{hi} cells initiate tumors containing both Id1^{hi} and Olig2⁺ cells, whereas tumors from Id1^{lo}/Olig2⁻ cells do not generate Id1^{hi} progeny, is consistent with our transcriptomic and genetic lineage tracing in human specimens. More recently, Narayanan and colleagues showed that ASCL1 is a master regulator of the proneural phenotype of GBM, and that ASCL1 directly represses mesenchymal-phenotype genes such as *CD44* and *GFAP* (21). This is consistent with our data; in particular, our scATAC-seq analysis shows anticorrelation of *CD44* and other mesenchymal genes with *ASCL1* gene-body activity and motif enrichment (Fig. 4C and G). Knowledge of GBM cell types, their lineage relationships, and functional differences is needed to develop combination therapies that address intratumor heterogeneity.

METHODS

Tumor Tissue Acquisition

We acquired fresh tumor tissue and peripheral blood from patients undergoing surgical resection for glioma at University of California, San Francisco (UCSF). Deidentified samples were provided by the UCSF Neurosurgery Tissue Bank. Sample use was approved by the Institutional Review Board at UCSF. The experiments performed here conform to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report. All patients provided informed written consent.

Tissue Processing for scRNA-seq

Fresh tissues were minced in collection medium (Leibovitz L-15 medium, 4 mg/mL glucose, 100 U/mL penicillin, 100 µg/mL streptomycin) with a scalpel. Sample dissociation was carried out in a mixture of papain (Worthington Biochem. Corp) and 2,000 U/mL of DNase I freshly diluted in Earle's Balanced Salt Solution and incubated at 37°C for 30 minutes. After centrifugation (5 minutes at 300 × *g*), the suspension was resuspended in PBS.

Subsequently, suspensions were triturated by pipetting up and down ten times and then passed through a 70-µm strainer cap (BD Falcon). Finally, centrifugation was performed for 5 minutes at 300 × *g*. After resuspension in PBS, pellets were passed through a 40-µm strainer cap (BD Falcon), followed by centrifugation for 5 minutes at 300 × *g*. The dissociated, single cells were then resuspended in GNS [Neurocult NS-A (Stem Cell Technology), 2 mmol/L L-glutamine, 100 U/mL penicillin, 100 µg/mL streptomycin, N-2/B-27 supplement (Invitrogen), and sodium pyruvate]. Nuclei were extracted from frozen tissues following the "Frankenstein" protocol developed by Luciano Martelotto, PhD (Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Melbourne, Australia), and

available from 10X Genomics (<https://community.10xgenomics.com/t5/Customer-Developed-Protocols/ct-p/customer-protocols>).

Fluidigm C1-Based scRNA-seq

Fluidigm C1 Single-Cell Integrated Fluidic Circuit and SMARTer Ultra Low RNA Kit were used for single-cell capture and complementary DNA (cDNA) generation. cDNA quantification was performed using Agilent High Sensitivity DNA Kits and diluted to 0.15 to 0.30 ng/μL. The Nextera XT DNA Library Prep Kit (Illumina) was used for dual indexing and amplification with the Fluidigm C1 protocol. cDNA was purified and the size selection was carried out twice using 0.9× volume of Agencourt AMPure XP beads (Beckman Coulter).

10X Genomics-Based scRNA-seq/snRNA-seq

For fresh tissues, 10.2 μL of live cells, at a concentration of 1,700 live cells/μL, were loaded into the 10X Chromium Single Cell Capture Chip. For frozen tissues, approximately 15,000 nuclei were loaded per capture. Single-cell/nucleus capture, reverse transcription, cell lysis, and library preparation were performed per the manufacturer's protocol. Sequencing was performed on an Illumina NovaSeq using a paired-end 100 bp protocol.

Public Data Acquisition

Normalized counts from TCGA RNA-seq data were obtained from the Genomics Data Commons portal (<https://gdc.cancer.gov/>). Patients diagnosed as having GBM with wild-type IDH1 expression ($n = 144$) were normalized to $\log_2(\text{CPM} + 1)$ and used for analysis. TCGA glioma microarray and associated survival data were obtained from the GlioVis portal (22). Z-score normalized counts from regional RNA-seq of 122 samples from 10 patients were obtained via the web interface of the Ivy GAP (<http://glioblastoma.alleninstitute.org/>) database. RNA-seq of nonmalignant human brain tissues was obtained from the GTEx portal (<https://www.gtexportal.org/home/datasets>).

Exome Sequencing and Genomic Mutation Identification

The NimbleGen SeqCap EZ Human Exome Kit v3.0 (Roche) was used for exome capture on a tumor sample and a blood control sample from each patient. Samples were sequenced with an Illumina HiSeq 2500 machine (100-bp paired-end reads). Reads were mapped to the human genome build GRCh37 with Burrows-Wheeler Aligner (23), and only uniquely matched paired reads were used for analysis. PicardTools (<http://broadinstitute.github.io/picard/>) and the Genome Analysis Toolkit (GATK; ref. 24) carried out quality score recalibration, duplicate removal, and realignment around indels. Large-scale (>100 exons) somatic copy-number variants (CNV) were inferred with ADTex (25). To increase CNV size, proximal (<1 Mbp) CNVs were merged. Somatic SNVs were inferred with MuTect (<https://www.broadinstitute.org/cancer/cga/mutect>) for each tumor/control pair and annotated with the Annovar software package (26).

sc/snRNA-seq Data Preprocessing

Data processing of the C1 data was performed as described previously (3). Briefly, reads were quality trimmed and TrimGalore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to clip Nextera adapters. HISAT2 (27) was used to perform alignments to the human genome build GRCh37. Gene expression was quantified using the ENSEMBL reference with featureCounts (28). Only correctly paired, uniquely mapped reads were kept. In each cell, expression values were scaled to counts per million (CPM)/100+1 and log-transformed. Low-quality cells were filtered by thresholding number of genes detected at 1,000 and at least 100,000 uniquely aligned reads. For 10× scRNA-seq and snRNA-seq data, we utilized CellRanger (version 3.0.2) for data preprocessing and gene-expression quantification, following the guidelines from the CellRanger web-site (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references#premrna>).

Classification of Somatic Mutations

The presence/absence of somatic CNVs in 10× scRNA-seq and snRNA-seq data was assessed with CONICSmatrix (6). We retained CNVs with a CONICSmatrix likelihood ratio test <0.001 and a difference in Bayesian Criterion >300. For each CNV, we used a cutoff of posterior probability >0.5 in the CONICSmatrix mixture model to infer the presence/absence of that CNV in a given cell. For the Fluidigm C1 data, we utilized our previous CNV and SNV classifications (3, 6). We retained only SNVs detected in exome-seq at >10% variant frequency in the tumor and <10% variant frequency in patient-matched normal blood. Cells expressing those tumor-restricted variant alleles were considered positive for the respective SNVs. The presence/absence of somatic CNVs in 10× snATAC-seq data was estimated with CONICSmatrix (version 1.0; ref. 6). Here, the gene-body activity generated by snapATAC (29) was used to perform the CNV analysis with a CONICSmatrix likelihood-ratio test <0.001 and a difference in Bayesian Criterion >300. SNV calls in 10× sc/snRNA-seq data were obtained by pooling reads by patient and running the GATK RNA-seq best practices pipeline (<https://software.broadinstitute.org/gatk/best-practices/workflow?id=11164>). Variant assignments in single cells were then assessed via VarTrix (version 1.0) tool <https://github.com/10xgenomics/vartrix>).

Dimensionality Reduction and Calculation of Stemness Scores

First, we filtered cells that had >5% mitochondrial read counts. A gene with nonzero read counts in more than 10 cells was considered as expressed, and each cell was required to have at least 200 expressed genes. Seurat v3 was used for *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) plots based on the first 10 principal components (9). PCA was done using R 3.4.2. The top 1,000 genes with the highest regularized variances were identified via Seurat v3 for each case. PCA was then performed using those genes that were among the 1,000 most variable genes for at least 3 patients. We defined the mGSC and pGSC gene sets as the top 15% of genes most strongly loading PC1, positively for mGSCs and negatively for pGSCs. Stemness scores were calculated using these gene sets as input to the *AddModuleScore* function from the Seurat package. Cluster-specific genes were identified via the FindMarkers/FindAllMarkers function from Seurat package, using a likelihood-ratio

test. Only genes enriched and expressed in at least 25% of the cells of at least one population and with a log-fold change bigger than 0.25 were considered.

Deconvolution of TCGA RNA-seq Data via Linear Models

To deconvolve GBM RNA-seq data from TCGA according to the cell types learned from scRNA-seq, we first pooled scRNA-seq read counts by cell type across mGSCs, pGSCs, nonmalignant oligodendrocytes, astrocytes, neurons, tumor-associated macrophages, T cells, and endothelial cells. The data used for this were our GBM scRNA-seq, as well as scRNA-seq of human fetal and adult nonmalignant brain tissues (12, 13). The resulting 8 count vectors were independently normalized to $\log_2(\text{CPM}/10+1)$. We then fit a linear model to each TCGA RNA-seq dataset (also scaled to $\log_2(\text{CPM}/10+1)$) using these vectors as predictors. We assessed the relative contribution of each predictor to the overall variance explained using the Lindeman, Merenda, and Gold (lmg) method as implemented in the relaimpo R package (30).

Lineage Reconstruction via RNA Velocity

RNA velocities were computed via velocityto (14). For GBM cases, cells with PC2 sample score > 0 were deemed cycling cells and used for velocityto analysis under default parameters. Root and terminal points in lineage reconstruction were identified with the functions `prepare_markov` and `run_markov` included in velocityto. These use backward and forward Markov processes on the transition probability matrix of the cells to determine high-density regions for the start and end points of trajectories, respectively.

snATAC-seq Data Processing

The CellRanger ATAC software (version 1.1.0) was used for read alignment, deduplication, and identifying transposase cut sites (<https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/algorithms/overview>). The output matrix of CellRanger was further processed by using the snapATAC package (<https://github.com/r3fang/SnapATAC>; ref. 29). We selected the highest-quality barcodes for each case based on two criteria: (i) number of filtered fragments at least 1,000; (ii) fragments in promoter ratio (FRiP) at least 0.2 for the case. Clustering was performed using Seurat v3 SNN graph clustering “FindClusters,” with the gene-body accessibility scores generated by the snapATAC package as input. Differentially accessible regions, peaks, and motif enrichments were computed using snapATAC “findDAR,” “runMACSForAll,” and “runHomer.”

IHC

IHC optimization was performed on a Leica Bond automated immunostainer using conditions optimized for each antibody (Supplementary Table S4). Heat-induced antigen retrieval was performed using Leica Bond Epitope Retrieval Buffer 1 (citrate buffer, pH 6.0) and Leica Bond Epitope Retrieval Buffer 2 (EDTA solution, pH 9.0) for 20 minutes [ER2(20)]. Nonspecific antibody binding was blocked using 5% milk in PBST or Novocastra Protein Block (Novolink #RE7158). Positivity was detected using Novocastra Bond Refine Polymer Detection and visualized with 3,3'-diaminobenzidine (brown) and alkaline phosphatase (red). A hematoxylin nuclear counterstain (blue) was applied. Duplex

controls were performed to provide a reference of specificity of the selected primary antibody and secondary detection system. For the DLL3 and CA9 duplex, two sets of single-stained control tissues were used because the antigens of interest are not commonly coexpressed in normal tissue types. Image analysis of whole-slide images (8 slides from patients with GBM; 5 control slides) was performed using the Aperio software (Leica) and the ImageDx slide-management pipeline (Reveal Bio). All tissue and staining artifacts were digitally excluded from the reported quantification.

Drug Screen

The U87MG cell line was obtained from ATCC and cultured in DMEM, supplemented with 10% FBS and 1% penicillin-streptomycin. Cells were seeded in 96-well plates at a density of 5,000 cells/well with DMEM and held overnight at 37°C, 5% CO₂. The media were then aspirated, and test compounds were administered through serial dilution in DMEM. After 48 hours, the cells were washed with PBS. DMEM containing AlamarBlue (Invitrogen) supplemented with FBS, penicillin, and streptomycin was added and the plate was then incubated for 4 hours. Cell viability was assessed via absorbance using a microplate reader and compared with untreated control wells.

Ethics Approval and Consent to Participate

Study protocols were approved by the UCSF Institutional Review Board. All clinical samples were analyzed in a deidentified fashion. All experiments were carried out in conformity to the principles set out in the WMA Declaration of Helsinki as well as the Department of Health and Human Services Belmont Report. Informed written consent was provided by all patients.

Availability of Data and Material

The study data are available from the European Genome-phenome Archive repository, under EGAS00001002185, EGAS00001001900, and EGAS00001003845. Third-party data that were used in the study are available from the GlioVis portal (<http://gliovis.bioinfo.cnio.es/>), the gbmseq portal (<http://gbmseq.org/>), and The Glioblastoma Atlas Project (<http://glioblastoma.alleninstitute.org/>).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work has been supported by research awards from the UCSF Glioma Precision Medicine Program (to A.A. Diaz, J.J. Philips, and S.M. Chang) and the UC Cancer Research Coordinating Committee (CRN-19-586041; to A.A. Diaz).

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. [PubMed: 20129251]

2. 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. [PubMed: 24925914]
3. Müller S, Liu SJ, Di Lullo E, Malatesta M, Pollen AA, Nowakowski TJ, et al. Single-cell sequencing maps gene expression to mutational phylogenies in PDGF and EGF driven gliomas. *Mol Syst Biol* 2016;12:889. [PubMed: 27888226]
4. Jin X, Kim LJY, Wu Q, Wallace LC, Prager BC, Sanvoranart T, et al. Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat Med* 2017;23:1352–61. [PubMed: 29035367]
5. Diaz A, Liu SJ, Sandoval C, Pollen A, Nowakowski TJ, Lim DA, et al. SCell: integrated analysis of single-cell RNA-seq data. *Bioinformatics* 2016;32:2219–20. [PubMed: 27153637]
6. Müller S, Cho A, Liu SJ, Lim DA, Diaz A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics* 2018;34:3217–9. [PubMed: 29897414]
7. Bartlett TE, Müller S, Diaz A. Single-cell co-expression subnetwork analysis. *Sci Rep* 2017;7:15066. [PubMed: 29118406]
8. Müller S, Kohanbash G, Liu SJ, Alvarado B, Carrera D, Bhaduri A, et al. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol* 2017;18:234. [PubMed: 29262845]
9. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20. [PubMed: 29608179]
10. 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. [PubMed: 27806376]
11. 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: 1391–402.
12. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A* 2015;112:7285–90. [PubMed: 26060301]
13. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* 2017;358:1318–23. [PubMed: 29217575]
14. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. *Nature* 2018;560:494–8. [PubMed: 30089906]
15. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* 2019;176:1325–39. [PubMed: 30827679]
16. Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, et al. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc Natl Acad Sci U S A* 2015;112: 851–6. [PubMed: 25561528]
17. 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. [PubMed: 31327527]
18. Decarvalho AC, Kim H, Poisson LM, Winn ME, Mueller C, Cherba D, et al. Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. *Nat Genet* 2018;50:708–17. [PubMed: 29686388]
19. Hu X, Wang Q, Tang M, Barthel F, Amin S, Yoshihara K, et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res* 2018;46:D1144–9. [PubMed: 29099951]
20. Barrett LE, Granot Z, Coker C, Iavarone A, Hambardzumyan D, Holland EC, et al. Self-renewal does not predict tumor growth potential in mouse models of high-grade glioma. *Cancer Cell* 2012;21: 11–24. [PubMed: 22264785]
21. Narayanan A, Gagliardi F, Gallotti AL, Mazzoleni S, Cominelli M, Fagnocchi L, et al. The proneural gene ASCL1 governs the transcriptional subgroup affiliation in glioblastoma stem cells

- by directly repressing the mesenchymal gene NDRG1. *Cell Death Differ* 2019;26: 1813–31. [PubMed: 30538287]
22. Bowman RL, Wang Q, Carro A, Verhaak RGW, Squatrito M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro Oncol* 2017;19:139–41. [PubMed: 28031383]
 23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60. [PubMed: 19451168]
 24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303. [PubMed: 20644199]
 25. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, et al. Inferring copy number and genotype in tumour exome data. *BMC Genomics* 2014;15:732. [PubMed: 25167919]
 26. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:1–7. [PubMed: 19843612]
 27. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 2016;11:1650–67. [PubMed: 27560171]
 28. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30. [PubMed: 24227677]
 29. Fang R, et al. Fast and accurate clustering of single cell epigenomes reveals cis-regulatory elements in rare cell types. *bioRxiv* 615179.
 30. Groemping U Relative importance for linear regression in R: the package relaimpo. *J Stat Softw* 2006;17:139–47.

SIGNIFICANCE:

Tumor-propagating cells can be derived from mesenchymal and proneural glioblastomas. However, a stem cell of the classic subtype has yet to be demonstrated. We show that classic-subtype gliomas are comprised of proneural and mesenchymal cells. This study sheds light on a long-standing debate regarding lineage relationships between glioma cell types.

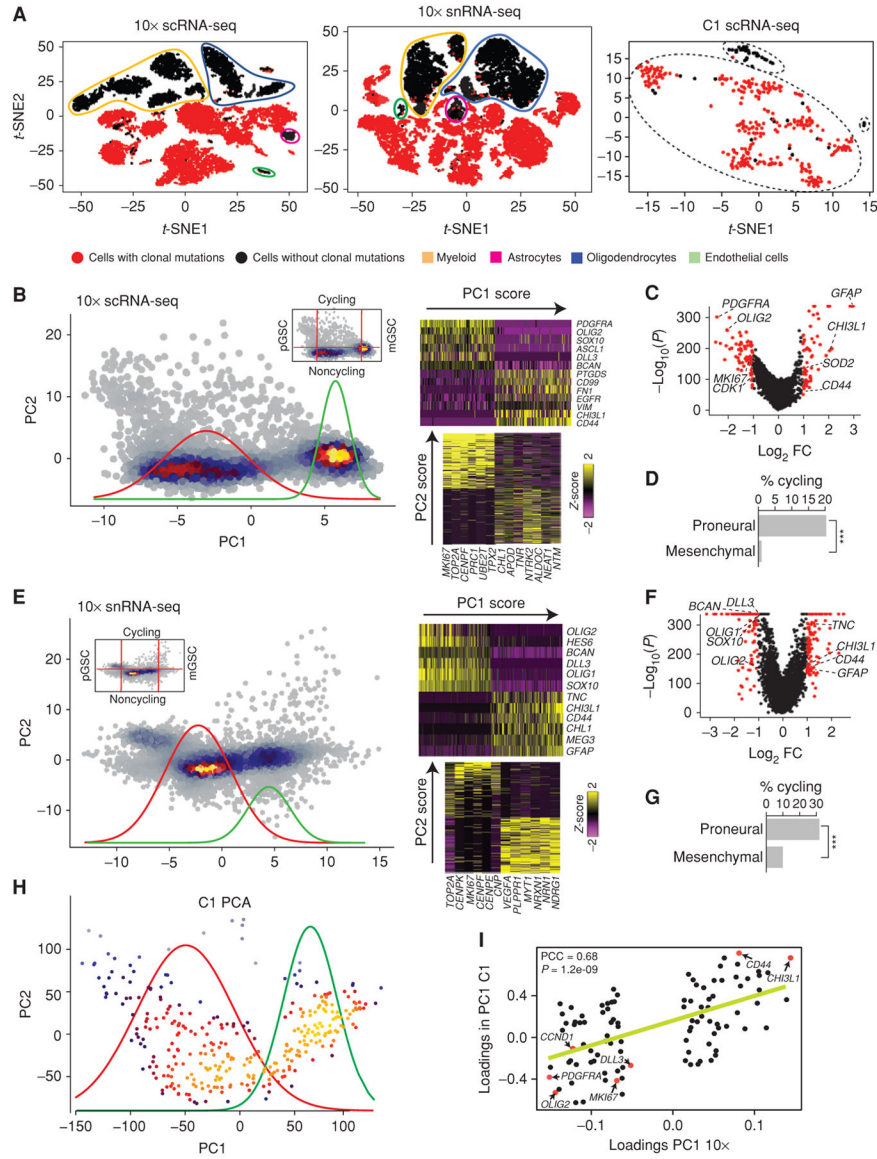


Figure 1. Single-cell sequencing reveals a single axis of variation in proliferating GBM cells. **A**, Left, *t*-SNE plot of 8,992 10× scRNA-seq cells from 6 patients; center, *t*-SNE plot of 15,975 nuclei from 10 patients; right, *t*-SNE plot of 568 cells from 7 patients. Cells/nuclei are colored by the presence (red) or absence (black) of clonal CNVs. **B**, Top, PCA of neoplastic cells from 10× scRNA-seq of IDH-wild-type GBMs. Curves represent the density of a Gaussian mixture model fit to PC1 sample scores. Heat maps display the expression of top-loading PC1 genes across cells, sorted by PC1 sample score. **C**, Differentially expressed genes between PCA clusters (abs. log₂ fold change > 1 and *P*_{adj} < 0.001 in red). **D**, Fractions of cycling cells. ***, Fisher *P* < 0.001. **E-G**, PCA and analysis as in **B-D** for 10× snRNA-seq of IDH-wild-type GBMs. ***, Fisher *P* < 0.001. **H**, PCA of neoplastic cells from C1 scRNA-seq of IDH-wild-type GBMs. **I**, Correlations between the C1-scRNA-seq and 1×-scRNA-seq loadings. PCC, Pearson correlation coefficient.

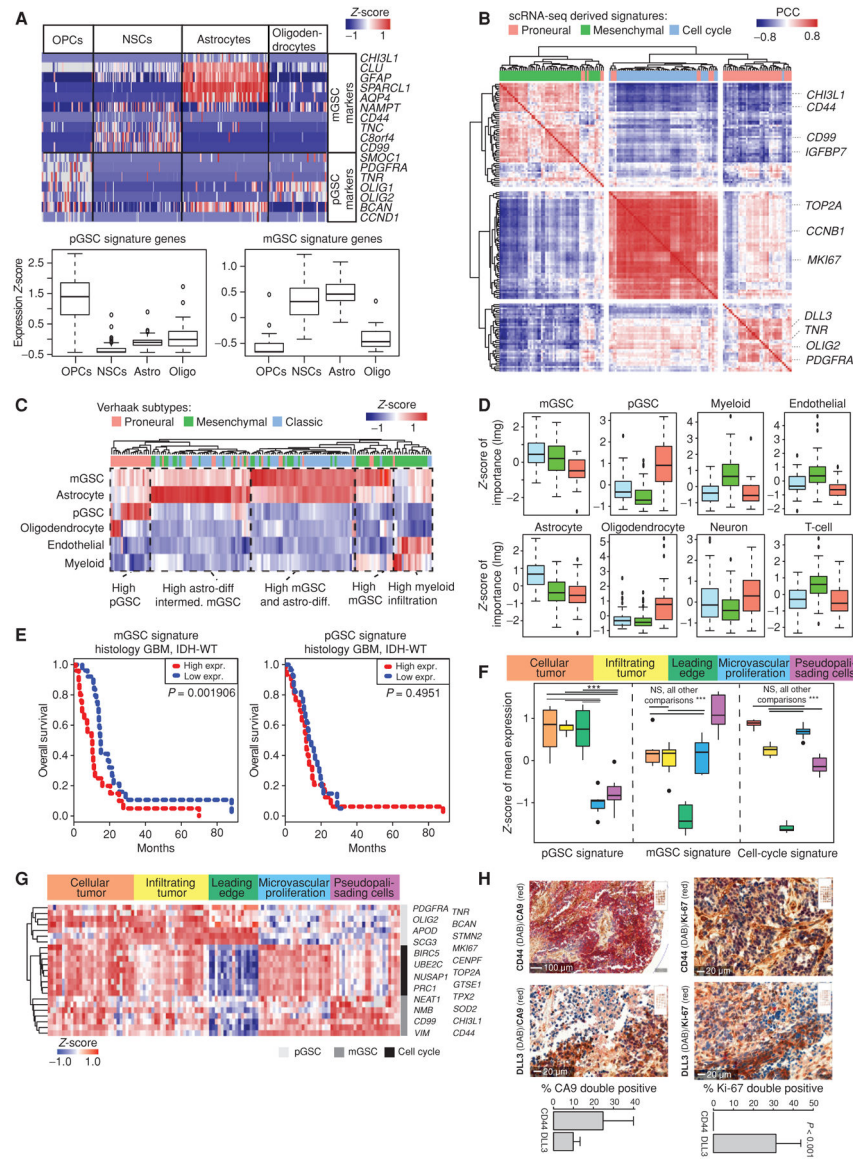


Figure 2. mGSCs and pGSCs explain the genetic and phenotypic heterogeneity of GBM. **A**, Expression of mGSC and pGSC markers in single cells from nonmalignant human brain. **B**, Hierarchical clustering of Pearson correlations between mGSC, pGSC, and cell-cycle genes in IDH-wild-type (IDH-WT) GBM RNA-seq samples from TCGA ($n = 144$). Heat map (**C**) and box plots (**D**) of the relative contributions of predictor cell types to the overall variance explained by a linear model fit to each TCGA sample. **E**, Kaplan-Meier analysis comparing survival of IDH-wild-type GBMs from TCGA to average expression of the mGSC and pGSC gene signatures in patient-matched RNA sequencing. **F** and **G**, MGSC, pGSC, and cell-cycle signatures in Ivy GAP RNA-seq of GBM-anatomic structures. **H**, Percentages of CD44⁺ DLL3⁺ cells also positive for CA9/Ki-67. ***, Wilcoxon $P < 0.001$. DAB, 3,3'-diaminobenzidine.

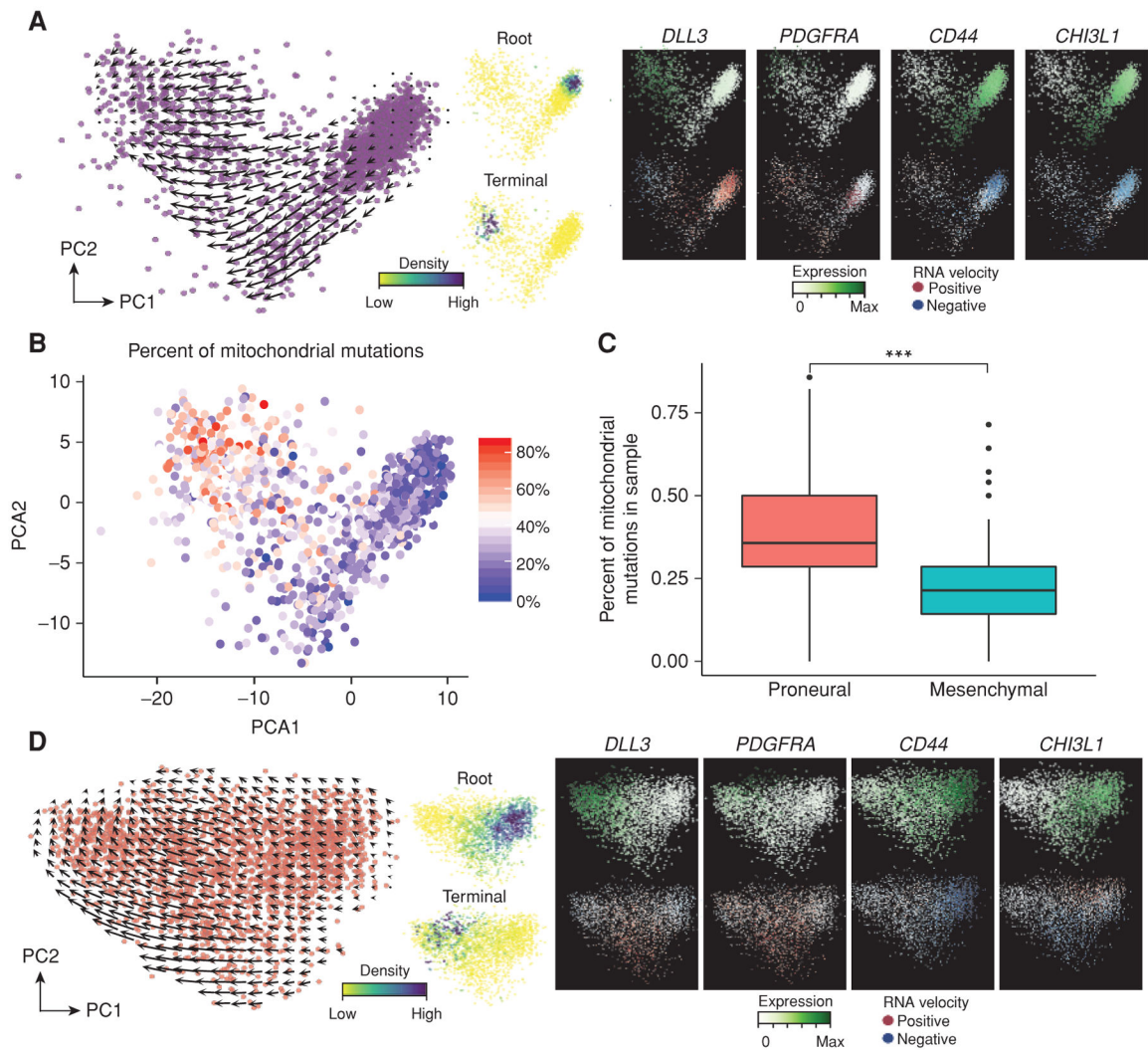


Figure 3. *In silico* genetic and transcriptomic lineage tracing supports a mGSC to pGSC hierarchy. **A**, Left, RNA velocities of cycling neoplastic 10× scRNA-seq IDH-wild-type GBM cells are projected onto a PCA axis. Center, velocyto-based lineage reconstruction identifies a stable mGSC root population and pGSC terminal population. Right, gene expression and velocity for mGSC and pGSC marker genes. **B**, The percent of expressed mitochondrial mutations found in a given cell, out of all mitochondrial mutations in a patient’s sample. **C**, Percent expressed mitochondrial mutations are compared between pGSCs and mGSCs. **D**, PCA analysis as in **A**, but for 10× snRNA-seq IDH-wild-type GBM data.

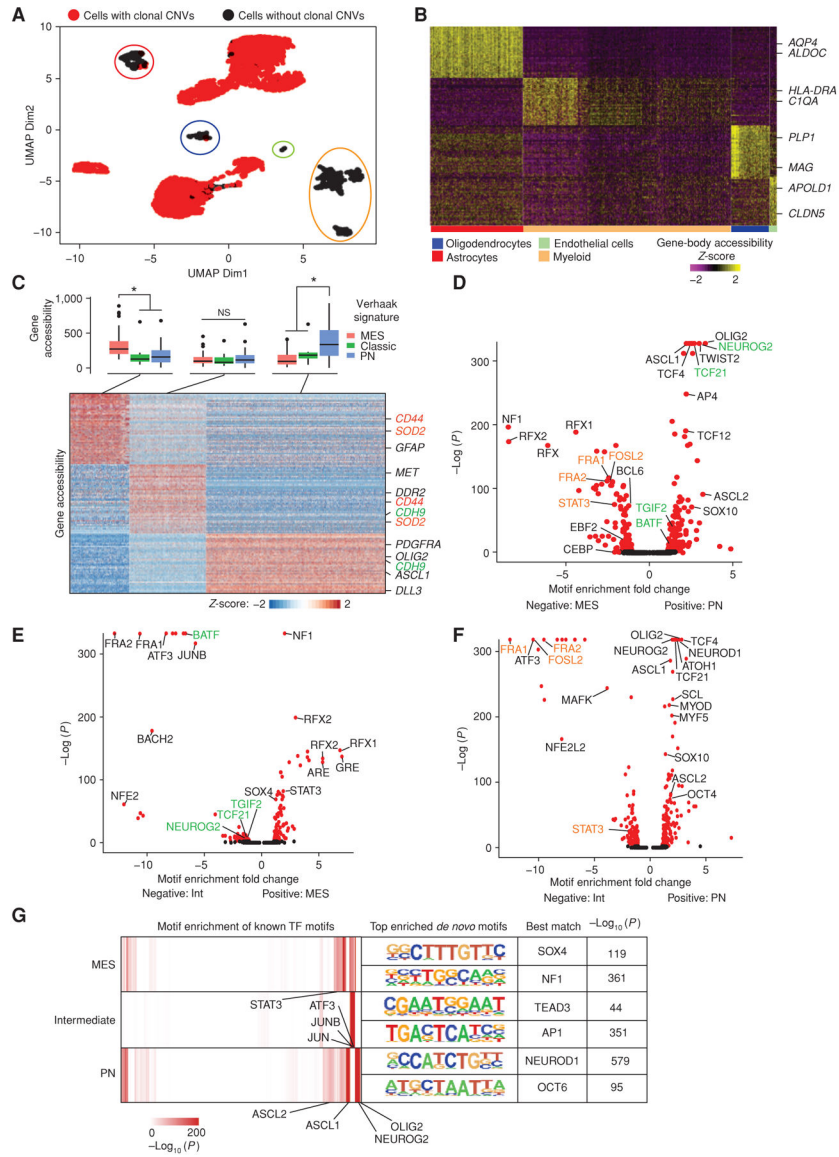


Figure 4.
A, t-SNE plot of IDH-wild-type GBM scATAC-seq annotated by the presence of clonal mutations. **B**, Clustering of scATAC-seq gene-body activity scores of nonneoplastic cells. **C**, Clustering of scATAC-seq gene-body activity scores of neoplastic cells, with box plots of activity scores for Verhaak-subtype gene sets annotated above. *, Wilcoxon $P < 0.05$. **D-F**, Differential motif enrichment tests between neoplastic cell clusters. **G**, Neoplastic cluster-specific motifs and associated transcription factors. Int, intermediate; MES, mesenchymal; PN, proneural; TF, transcription factor.

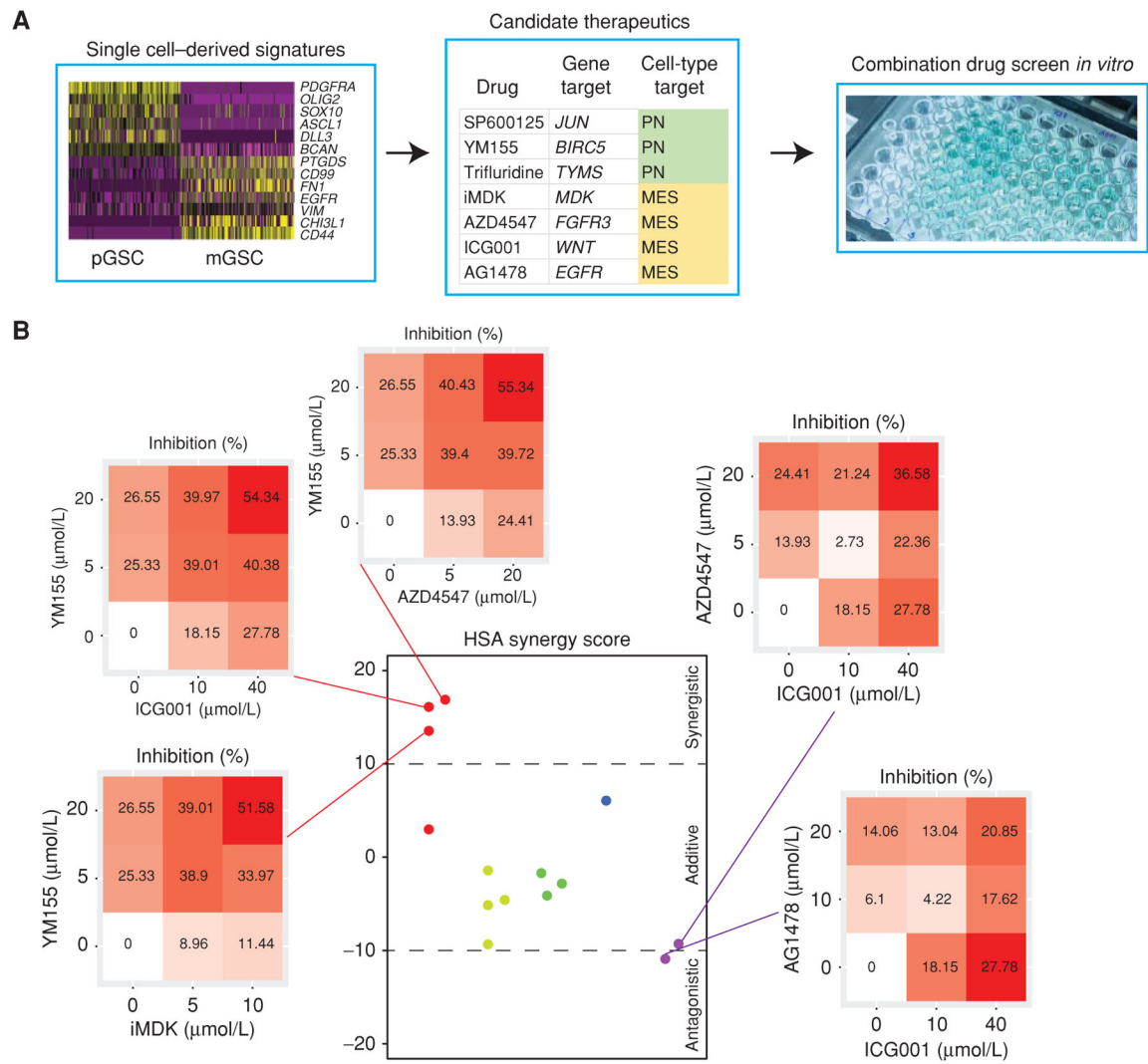


Figure 5. **A**, A schematic overview of the drug screen. **B**, HSA synergy scores and dose responses for drug combinations screened in U87 cells *in vitro*.