Title

Integrated analysis of single-cell chromatin state and transcriptome identified common vulnerability despite glioblastoma heterogeneity

Journal

Proceedings of the National Academy of Sciences of the United States of America, 120(20)

ISSN

0027-8424

Authors

Raviram, Ramya

Raman, Anugraha

Preissl, Sebastian

et al.

Publication Date

2023-05-16

DOI

10.1073/pnas.2210991120

Peer reviewed

PNAS RESEARCH ARTICLE | MEDICAL SCIENCES

Integrated analysis of single-cell chromatin state and transcriptome identified common vulnerability despite glioblastoma heterogeneity

Ramya Raviram[a], Anugraha Raman[a], Sebastian Preissl[b,c], Jiangfang Ning[d], Shaoping Wu[d], Tomoyuki Koga[d], Kai Zhang[a], Cameron W. Brennan[e], Chenxu Zhu[a], Jens Luebeck[f], Kinsey Van Deynze[a], Jee Yun Han[b], Xiaomeng Hou[b], Zhen Ye[a], Anna K. Mischel[a], Yang Eric Li[a], Rongxin Fang[a], Tomas Babak[g], Joshua Mugford[g], Claudia Z. Han[h], Christopher K. Glass[h,i], Cathy L. Barr[j,k,l], Paul S. Mischel[m], Vineet Bafna[f], Laure Escoubet[g], Bing Ren[a,b,h,1,2], and Clark C. Chen[d,1,2]

Edited by Monte M. Winslow, Stanford University School of Medicine, Palo Alto, CA; received June 27, 2022; accepted March 9, 2023 by Editorial Board Member Anton Berns

In 2021, the World Health Organization reclassified glioblastoma, the most common form of adult brain cancer, into isocitrate dehydrogenase (IDH)-wild-type glioblastomas and grade IV IDH mutant (G4 IDHm) astrocytomas. For both tumor types, intratumoral heterogeneity is a key contributor to therapeutic failure. To better define this heterogeneity, genome-wide chromatin accessibility and transcription profiles of clinical samples of glioblastomas and G4 IDHm astrocytomas were analyzed at single-cell resolution. These profiles afforded resolution of intratumoral genetic heterogeneity , including delineation of cell-to-cell variations in distinct cell states, focal gene amplifications, as well as extrachromosomal circular DNAs. Despite differences in IDH mutation status and significant intratumoral heterogeneity, the profiled tumor cells shared a common chromatin structure defined by open regions enriched for nuclear factor 1 transcription factors (NFIA and NFIB). Silencing of *NFIA* or *NFIB* suppressed in vitro and in vivo growths of patient-derived glioblastomas and G4 IDHm astrocytoma models. These findings suggest that despite distinct genotypes and cell states, glioblastoma/G4 astrocytoma cells share dependency on core transcriptional programs, yielding an attractive platform for addressing therapeutic challenges associated with intratumoral heterogeneity.

single cell | glioblastoma | extrachromosomal DNA | amplicons

Glioblastoma, also known as World Health Organization (WHO) grade 4 (G4) astrocytoma, is the most common form of primary brain cancer in adults and remains one of the deadliest of human cancers. Histologically, glioblastoma is defined by marked hypercellularity, nuclear atypia, pseudopalisading necrosis, and microvascular proliferation (1). Despite commonality in histologic appearance, molecular profiling of glioblastomas revealed significant intertumoral and intratumoral heterogeneity (2, 3). A key feature that defines intertumoral heterogeneity involves the isocitrate dehydrogenase genes (IDH) (4, 5). While histologically indistinguishable from wild-type IDH (IDHwt) glioblastomas, IDH mutant (IDHm) tumors affect younger patients, are associated with more favorable prognosis, and exhibit distinct responses to therapy (1, 6). In this context, the 2021 WHO classification recognized grade IV IDHm (G4 IDHm) astrocytomas and IDHwt glioblastomas as two distinct entities (7).

Single-cell RNA-seq (scRNA-seq) analysis of both IDHwt glioblastomas and G4 IDHm astrocytomas further revealed significant intratumoral heterogeneity, with copresence of tumor cells harboring distinct progenitor-like and differentiated-like cell states (3), including the neural progenitor cells (NPC-like), oligodendrocyte progenitor cells (OPC-like), astrocyte-like cells (AC-like), and mesenchymal-like (MES-like) cell states. This intratumoral heterogeneity poses significant challenges in therapeutic development, magnified by the dynamic transitions between cell states in response to therapy (8, 9). Meaningful therapeutic response requires the identification of effective agents that simultaneously target these distinct cell states or strategies that combine multiple therapeutic agents.

Single-cell sequencing technologies have been developed to study various features of chromatin such as open chromatin status (10, 11) and histone modifications (4, 12–15). Here, we analyzed clinical IDHwt glioblastoma and G4 IDHm astrocytoma specimens using single-nucleus ATAC-seq (snATAC-Seq), single-nucleus RNA-Seq (snRNA-seq), as well as a single-cell coassay of open chromatin and nuclear transcriptome (Paired-Seq) (16), with the goal of refining our understanding of intratumoral heterogeneity as well as to identify shared dependencies despite this heterogeneity.

Significance

Intratumoral heterogeneity forms the basis for cancer evolution and fundamentally dictates the therapeutic response. Here, we carried out genome-wide chromatin accessibility (snATAC-Seq) and transcription (snRNA-seq) profiling at single-cell resolution using clinical specimens derived from isocitrate dehydrogenase wild-type (IDHwt) glioblastomas and grade 4 IDH mutant (IDHm) astrocytomas, the most common forms of adult brain cancer. The integrated analysis achieved resolution of intratumoral heterogeneity not previously possible, including structural delineation of extrachromosomal circular DNA. IDHm grade 4 astrocytoma and IDHwt glioblastoma cells in distinct cell states share a dependency on a nuclear factor IA/B-mediated core transcriptional program. These findings demonstrate the potential for integrated single-nuclear platforms in addressing therapeutic challenges related to intratumoral heterogeneity.

This article is a PNAS Direct Submission. M.M.W. is a guest editor invited by the Editorial Board.

[1]B.R. and C.C.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: biren@health.ucsd.edu or ccchen@umn.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2210991120/-/DCSupplemental.

Published May 8, 2023.

Results

Clustering of snATAC-Seq Data Reveals Tumor Subpopulations Based on Focal Amplifications. In this proof-of-principle study, we first performed snATAC-Seq (combinatorial barcoding strategy) (17) and snRNA-seq (10X Genomics) on two G4 IDHm astrocytomas (RL1 and RL2) and three IDHwt glioblastomas (RL3, 4, and 5) (*SI Appendix*, Table S1 and Fig. S1 *A* and *B*). For two of the tumor specimens (RL1 and RL3), we examined multiple sections to characterize spatial heterogeneity within each tumor. We first performed clustering of snATAC-Seq profiles using SnapATAC (18) (binary counts in 5 kb bins) for each tumor individually and detected nine clusters in the RL1 tumor (Fig. 1*A*). Nonmalignant cell types (oligodendrocytes, myeloid, and neurons) were identified based on accessibility profiles on marker genes. We then performed CNV analysis (20 Mb bins) to confirm that the remaining six clusters are malignant evidenced by the amplification of a region on chromosome 11 in these cells (Fig. 1*B*).

Next, to classify the malignant cells into glioma cell states based on snATAC-Seq profiles, we analyzed the data with snRNA-seq profiles. Clustering of snRNA-seq profiles revealed nine clusters (Fig. 1*C*). Nonmalignant cells were identified based on the expression of marker genes (a small population of nonmalignant astrocytes was detected that was not found in the snATAC-Seq data). Due to data sparsity, malignant cells in snRNA-seq were determined based on cells that corresponded to malignant cells with CNVs upon integration of the two datasets, revealing five malignant clusters (Fig. 1*C*). We classified each cluster based on the enrichment of gene module scores from previously defined glioma cell state signatures (*SI Appendix*, Fig. S1 *C* and *D*). Some clusters were a mixture of closely related cell states, and the low number of genes captured per cell limits the ability to distinguish between these cell states. For instance, we observed a cluster with enrichment for the MES-like cell state from snRNA-seq data and cells in the AC-like cluster with enrichment for the MES-like signature in the snATAC-Seq data (Fig. 1 *C* and *D*).

Integration of snRNA-seq and snATAC-Seq revealed clustering that corresponded to the stem-like, OC-like, MES-like, or AC-like states (*SI Appendix*, Fig. S1*D*). We also calculated a module score based on ATAC-seq signal (±2 kb of the gene body) and observed similar patterns, though the clustering of MES-like state is less obvious (Fig. 1*D*). All the six clusters with malignant cells identified in snATAC-Seq contain cells that map to the stem/OC-like cell state from snRNA-seq data. A subset of malignant cells in snATAC-Seq defined cluster 2 additionally mapped to the snRNA-seq-defined AC/MES-like cell state (Fig. 1*E*). In aggregate, our data demonstrate that snATAC-Seq can be used to identify malignant cells based on CNV profiles and classify them into certain cell states defined from snRNA-seq data.

Additional snATAC-Seq Clusters Correspond to Tumor Subpopulations with Focal Amplifications. To identify the features that distinguish the snATACs-seq clusters, we performed differential analysis to identify cluster-specific peaks. First, we identified peaks across all malignant cells in RL1 and used the SnapATAC FindDAR function (18) to identify cluster-specific peaks. We observed that certain peaks specific to malignant clusters had a normalized count higher than that of most peaks (Fig. 2*A*). As this could represent copy number amplifications, we performed whole-genome sequencing (WGS) on section 3 of RL1 and indeed observed that these cluster-specific peaks corresponded to focal amplifications (Fig. 2*B*).

To detect these focal amplifications from snATAC-Seq data, we adapted a previous method (19) to detect CNVs from bulk ATAC-seq data (*SI Appendix*, Fig. S2). To improve the resolution of gene amplification detection, we used the aggregate counts in each 50 kb-sized genomic bin from cells in each cluster. For each bin, we calculated a $\log_2$-ratio using astrocytes and oligodendrocyte progenitor cells from the nontumor brain sample and then applied the circular binary segmentation (CBS) algorithm (20) to define continuously amplified regions in the genome (Fig. 2*B*). We detected 12 out of 14 events from WGS data (*SI Appendix*, Table S2), suggesting high accuracy and specificity of our approach

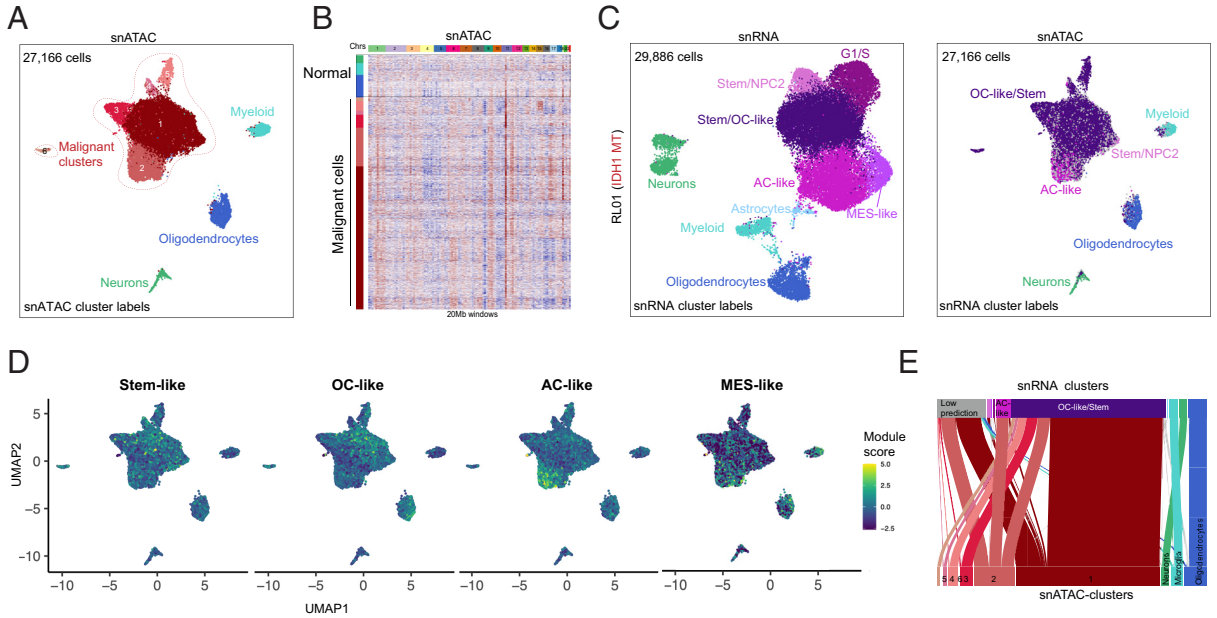

Fig. 1. Integration of snRNA-seq and snATAC-Seq data in IDH-mt grade 4 astrocytoma (G4 IDHm astrocytoma). (*A*) UMAP of snATAC-Seq data from five sections of RL1 G4 IDHm astrocytoma. Cells colored based on clustering identified from SnapATAC. (*B*) Heatmap of copy number variation (CNV) analysis from snATAC-Seq data in 20 MB windows. (*C*) UMAP of snRNA-seq data from five sections of RL1 and transferred labels from integration with snRNA-seq onto snATAC-Seq clustering (*Left*). (*D*) Cells colored based on module score for glioma cell states based on signal in peaks in ±2b around genes. (*E*) Sankey plot of cell label correspondence between snATAC-Seq and snRNA-seq.

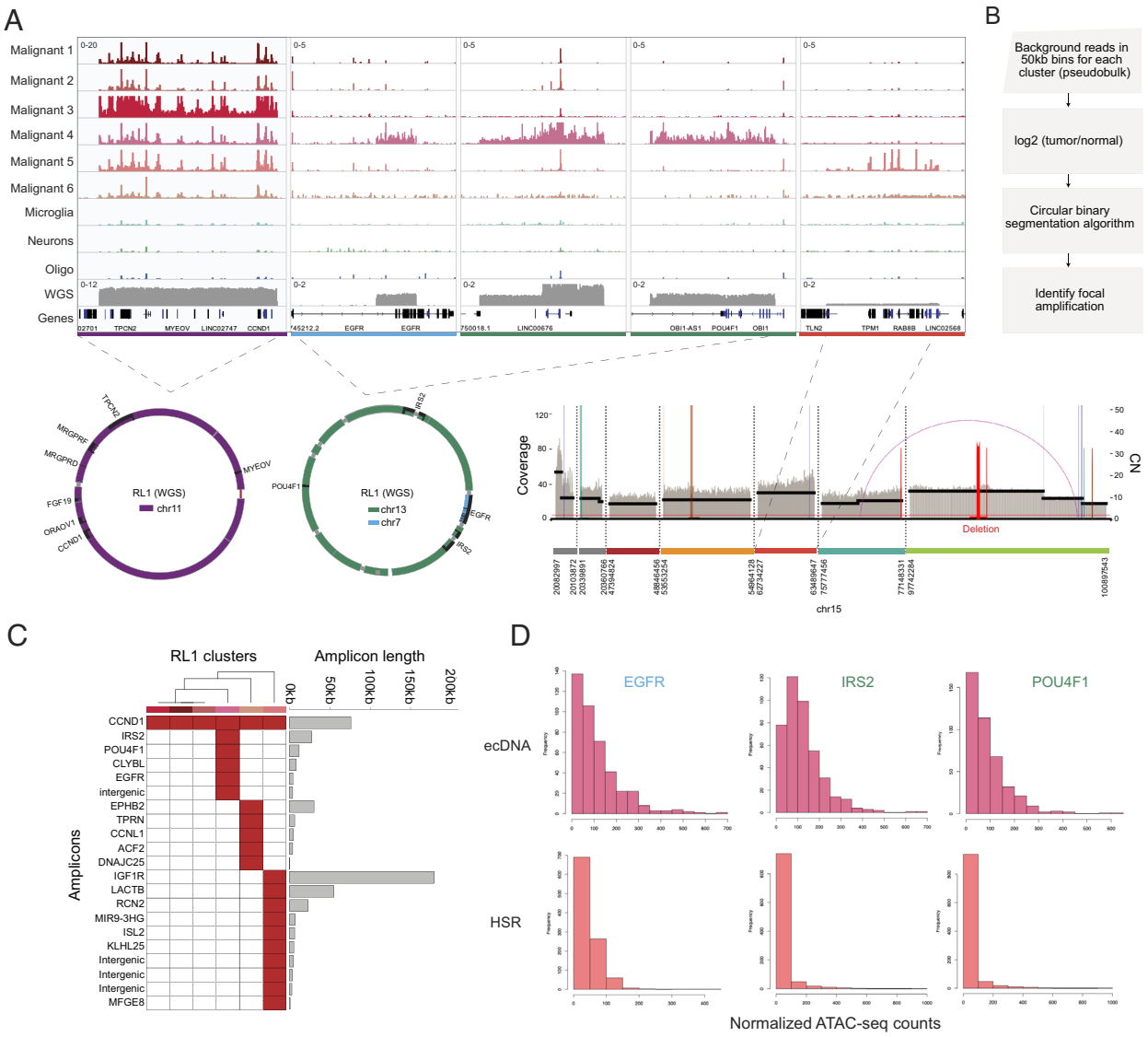

Fig. 2. Identification of focal amplifications from snATAC-Seq data. (*A*) snATAC-Seq signal and WGS signal for five focal amplifications detected across the genome and viewed using IGV (*Top*). Reconstruction of circular amplifications based on WGS data and noncircular amplifications (*Bottom*). (*B*) Workflow of focal amplification detection from aggregating cells from each snATAC-Seq cluster. (*C*) Length of amplicons detected from snATAC-Seq data and the clusters they were detected in (red squares). (*D*) Histogram of snATAC-Seq signal in cells where extrachromosomal circular DNA (ecDNA) (cluster4) and homogenous staining regions (HSR) (cluster6) were detected.

and nine additional events in tumor sections for which we did not perform WGS. One focal amplification (~700 kb region on chr11 containing the *CCND1* oncogene) was detected across all malignant clusters, while the remaining were highly cluster specific (Fig. 2*C*). This approach allowed us to characterize heterogeneity within the tumor based on cell-to-cell variations in focal amplifications and enhanced the resolution of subclonal genetic heterogeneity.

As ecDNA is an important mechanism of focal oncogene amplification, we tested whether these amplifications could be the result of ecDNA or double-minute formation (21). Using AmpliconArchitect (22, 23), which can analyze amplifications to identify ecDNA structures, we found that four out of the fourteen focal amplifications detected from WGS data showed evidence of ecDNA structures (Fig. 2*A*). The amplification around the *CCND1* oncogene region found across all malignant cells formed simple circular structures (Fig. 2*A*). We also detected a fusion amplification-based translocation of genomic segments involving the intragenic enhancer of *EGFR* (24), *IRS2*, and *POU5F1* oncogenes that was highly specific to malignant cluster 4 (Fig. 2*C*). Additionally, we found cluster-specific amplifications on chromosome 15 that were not ecDNA.

Previous studies using mathematical modeling and fluorescence in situ hybridization (FISH) have observed that there is a great deal of cell-to-cell heterogeneity in ecDNA copy numbers as random segregation occurs upon cell division. Here, using the snATAC-Seq signal within the amplified regions, we compared the cell-to-cell heterogeneity between regions that were identified as ecDNA versus chromosomal amplifications. Consistent with previous mathematical modeling, we observed a wider distribution in ecDNA compared to chromosomal amplifications (Fig. 2*D*).

To compare the ecDNA shown here to those previously reported, we compared the amplicons detected in our samples with those reported from a pan-cancer analysis (23). Based on snATAC-Seq and WGS data, we detected three ecDNAs and all three overlapped with ecDNAs reported across various cancers. In addition, we detected nine potential ecDNAs from snATAC-Seq (no corresponding WGS to confirm whether they are ecDNAs) and, of these, four overlapped with ecDNAs detected in different cancers (*SI Appendix*, Fig. S3).

To assess the expression of genes in the DNA exhibiting focal amplifications, we performed Paired-Seq (16), a single-cell multiomics

assay to jointly profile gene expression and chromatin accessibility from the same nucleus for section 3 of RL3 (Fig. 3*A*). We asked whether copy numbers correlated with gene expression at single-cell level. We classified the tumor cells with matched RNA and ATAC profiles (n = 1,963) into quartiles to group cells based on increasing copy number. Of the six genes on this amplified region, we observed that the expression of *TPCN2* and *CCND1* showed higher expression in cells with more copies based on the ATAC-seq inferred copy number (Fig. 3*B*).

We also examined the effect of focal amplifications without features of ecDNA, where we detected multiple events on chromosome 15 in RL3. Due to the small number of cells harboring these amplifications, we assessed the gene expression at the cluster level instead of at the single-cell level. We observed that 15/19 genes on these chromosome 15 amplifications had the highest expression in the amplified cluster compared to other clusters (Fig. 3 *C* and *D*). Several of the amplified genes such as *RAB8B*, *LRRC28*, *RCN2*, and *APH1B* are also highly expressed in bulk RNA-seq data from TCGA glioblastoma data compared to normal brain tissue (GTEx).

In aggregate, these results suggest that snATAC-Seq as a profiling platform afforded resolution of a high degree of intratumoral genetic heterogeneity .

Distinct Subclonal Populations with Either *EGFR* or *PDGFRA* Amplifications in IDH-Wild-Type Glioblastoma. We next characterized the intratumoral heterogeneity of *EGFR* and *PDGRA*, two receptor tyrosine kinases that play key roles in glioblastoma pathogenesis (2). To this end, we analyzed snATAC profile for the four sections of RL3, an *EGFR*-amplified IDH-wild-type glioblastoma that were taken from geographically distinct regions of the tumor (Fig. 4 *A* and *B*). In addition to *EGFR* amplification, we observed an amplification of the region surrounding the *PDGFRA* oncogene (Fig. 4*C*). AmpliconArchitect analysis of RL3 WGS identified *EGFR* amplification as an ecDNA (Fig. 4*D*), while *PDGFRA* amplification was chromosomal. *EGFR* and *PDGFRA* focal amplification was mutually exclusive (Fig. 4 *B*, *C*, and *E*).

Integrating the snATAC-Seq data with snRNA-seq, we next ask whether these cells with the amplifications are preferentially associated with a particular transcriptional cell state. Clustering of snRNA-seq for RL3 detected two MES-like, one AC/MES-like, and one OPC/NPC-like clusters (Fig. 4*F*). The *EGFR*-amplified cells corresponded to both the AC/MES-like and NPC-like tumor clusters defined by snRNA-seq. In contrast, the *PDGFRA*-amplified cells primarily corresponded to the NPC-like cluster (Fig. 4*G* and *SI Appendix*, Fig. S4 *A–C*). By integration of snATAC-Seq and snRNA-seq, we demonstrate a higher resolution of the heterogeneity observed in the tumor based on focal amplification and the effect genetic aberrations have on the overall transcriptional cell state of the malignant cell. With larger sample cohorts, this approach can allow determination of the amplification's downstream effects.

Shared Dependency of Glioblastoma Cell States on NFIA and NFIB. Since active regulatory elements are characterized by open chromatin regions, we characterized the open chromatin regions in each tumor cluster as a function of their cell states. For each tumor, we performed differential accessibility analysis between malignant cluster and then performed motif enrichment analysis on the cluster-specific peaks (Fig. 5*A*). This analysis identified motifs for the bHLH family of transcription factors, such as NEUROG2/D1 and OLIG2, enriched in the

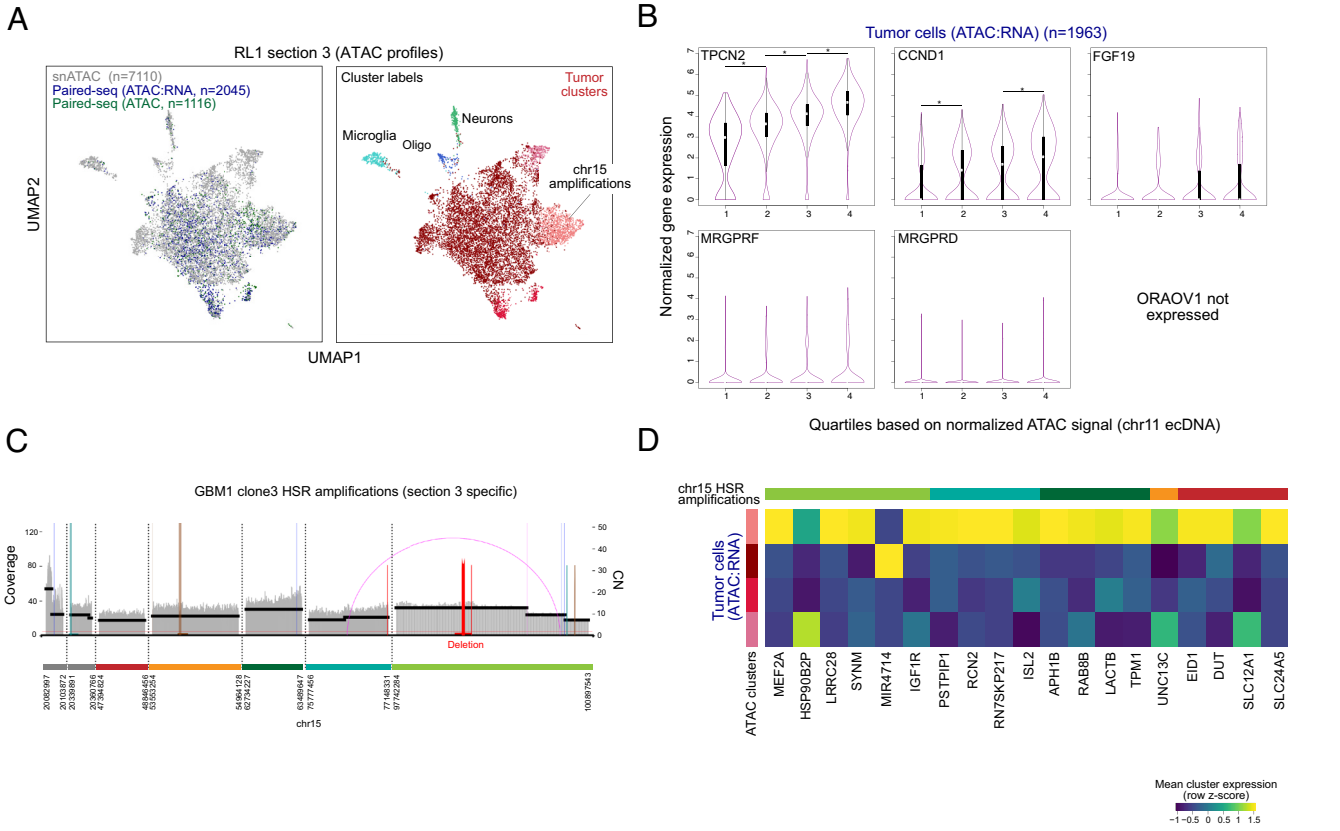

Fig. 3. Paired-Seq data for one tumor section of RL1. (*A*) Coclustering of snATAC-Seq profiles from Paired-Seq data and snATAC-Seq from the same tumor. (*B*) For genes on chr11 ecDNA in RL1, expression level of genes binned based on quartiles of normalized ATAC signal per cell. (*C*) Noncircular amplifications from multiple regions on chromosome 15. (*D*) Gene expression for genes on chr15 HSR-amplified regions.

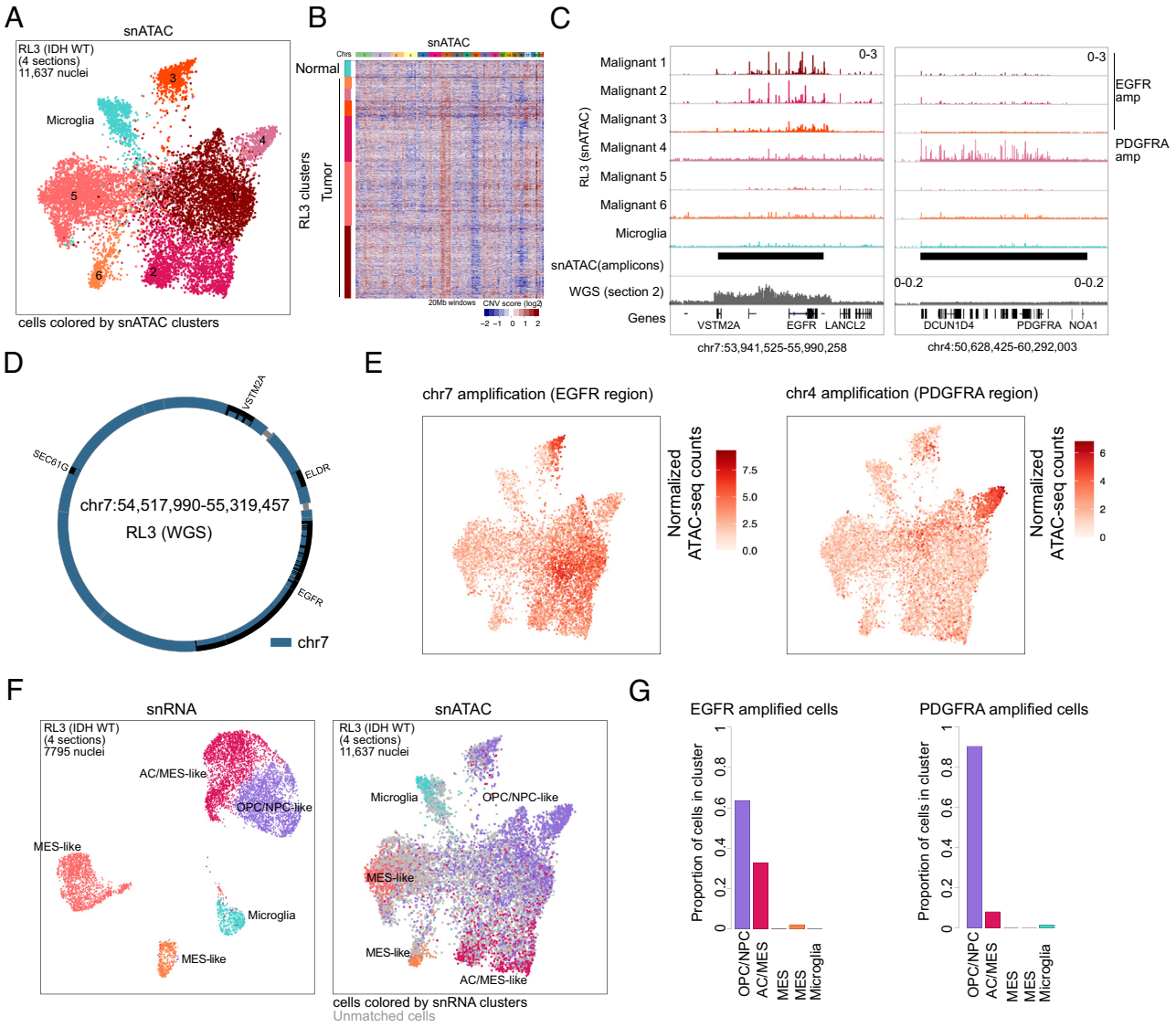

Fig. 4. Identification of focal amplification in IDH-wild-type glioblastoma. (*A*) UMAP of snATAC-Seq data from four sections of RL3 (IDH-wild type) glioblastoma. (*B*) Heatmap of CNV analysis from snATAC-Seq data in 20 MB windows. (*C*) snATAC-Seq signal and WGS signal focal amplifications detected on chr7 and chr4 and viewed using IGV. (*D*) Reconstruction of circular amplifications based on WGS data and noncircular amplifications. (*E*) Cells colored based on normalized ATAC-seq signal in *EGFR* and *PDGFRA*-amplified regions. (*F*) UMAP of snRNA-seq data from four sections of RL3 (IDH-wild type) glioblastoma (*Left*). UMAP of snATAC-Seq colored based on transferred labels from integration with snRNA-seq (*Right*). (*G*) Barplot showing proportion of cells in each cluster for each tumor section.

cells exhibiting OPC/NPC-like states. Enrichment analysis of peaks that were present in multiple cell states across each tumor revealed that NFI-halfsite motif representing the NFI transcription factor complex of NFIA, NFIB, NFIC, and NFIX is similarly enriched in both OPC/NPC-like and AC/MES-like cell states. This motif was also found to be enriched in all TCGA subtypes based on single-cell ATAC-seq data (25). Such enrichment was observed in both IDHwt glioblastomas and G4 IDHm astrocytomas.

NFIA has been shown to play an important role in the initiation of gliogenesis during early neural stem cell development (26, 27). Previous scRNA-seq studies revealed that NFIA and NFIB expression is associated with a stemness signature (28). Moreover, *NFIB* expression is associated with the maintenance of NPC-like cell state (29). We used Gepia2 (30) to assess the expression of NFI complex members by comparing bulk TCGA RNA-Seq data to normal brain tissue samples from GTEx data and found that only *NFIA* and *NFIB* were up-regulated in glioblastoma tumor samples (*SI Appendix*, Fig. S5).

In our snRNA dataset, we reliably detected *NFIA/IB* expression in both RL1 (IDHm) and RL3 (IDHwt, *SI Appendix*, Fig. S6 *A*, *B*, *D*, and *E*). Similar levels of *NFIA* and *NFIB* were observed in all glioblastoma cell states (*SI Appendix*, Figs. S6 *C* and *F* and S7). Finally, using methods described by Zhang et al. (31), we performed an integrative analysis of snATAC-Seq and snRNA-seq to identify a NFIA/NFIB-associated core-transcriptional program (Fig. 5*B* and *SI Appendix*, Table S3). Using network analysis and PageRank analysis, we show that this core-transcriptional program is highly active in all glioblastoma subtypes in RL3 (Fig. 5*C*) and RL1 (*SI Appendix*, Fig. S8 *A* and *B*). SPI1, a highly active transcription factor in myeloid cells (32) that plays unknown role in glioma pathogenesis, is included as a control for our analysis. To further characterize the mechanism of how NFIA and NFIB impact glioma pathogenesis, we extracted their target genes from the network and applied Gene Set Enrichment Analysis (GSEA) (33). This analysis revealed that NFIA/IB potentially contribute to the activation of Ras signaling as well as other cancer-related pathways (*SI Appendix*, Fig. S8*C*).

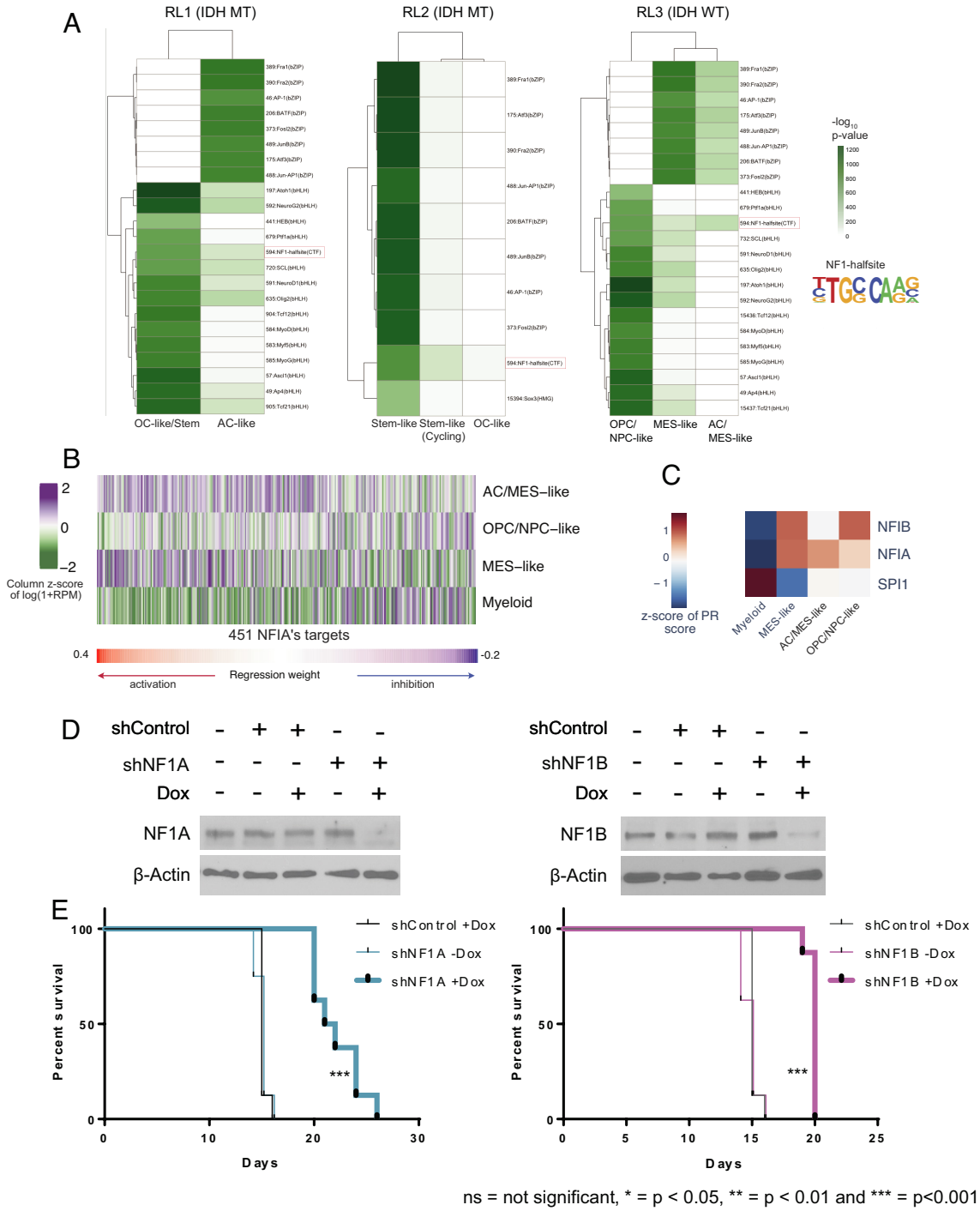

Fig. 5. NFI motif is enriched across glioma cell states. (*A*) Heatmap of p-values for enriched motifs in cell state-specific peaks across each tumor (only *P*-values <1e-100 are displayed). Motif is ordered based on probability value of enrichment. NFI-half-site (shown in red box) is enriched across multiple cell states in each sample. (*B*) Heatmap of NFIA's target genes' expression levels of RL3. The target genes are ordered by the weights from the regression model. Positive weights indicate positive regulations between NFIA and the targets, while negative weights indicate negative regulations. (*C*) Heatmap showing the PageRank scores of NFIA, NFIB, and SPI1 across different cell types/states in RL3. (*D*) Western blot analysis demonstrating silencing of *NFIA* (*Left*) and *NFIB* (*Right*) upon doxycycline induction. (*E*) Silencing of *NFIA* (*Left*) or *NFIB* (*Right*) prolonged survival of mice orthotopically implanted with a patient-derived xenograft (PDX) (GBM 83).

To assess whether *NFIA* and *NFIB* are essential in IDHwt glioblastoma, we performed RNA interference experiments to knockdown both genes in a previously described PDX model. First, doxycycline-inducible lentivector-containing shRNA sequences for *NFIA* and *NFIB* or scrambled sequence were introduced into the GBM83 PDX line (8), an IDHwt glioblastoma line enriched for gene expression signature observed in the AC/MES-like state. Induction of *NFIA* and *NFIB* silencing by siRNAs suppressed glioblastoma growth in vitro by more than 50%, at 7 d after doxycycline treatment (*SI Appendix*, Fig. S9 *A* and *B*). To validate these effects in vivo, the GBM83 line harboring doxycycline-inducible lentivector against *NFIA* and *NFIB* was orthotopically injected into the brain of nude mice. The mice were then fed with doxycycline-containing or control water in a randomized experimental design. For both *NFIA* and *NFIB* experiments, mice fed doxycycline-containing water (thereby silencing *NFIA* and *NFIB*) exhibited prolonged survival relative to those fed control water (Fig. 5*C*).

We next tested whether *NFIA* and *NFIB* are essential for the tumorigenicity of CMK3 (8), an IDHwt glioblastoma PDX line enriched for gene expression signature observed in the NPC/OPC-like cell state. Similar to that observed in GBM83, induction of *NFIA* and *NFIB* silencing in CMK3 inhibited growth in vivo (*SI Appendix*, Fig. S10). Finally, we tested the effects of *NFIA* and *NFIB* silencing in TS603, a G4 IDHm astrocytoma PDX model (34, 35). Induction of *NFIA* and *NFIB* silencing in TS603 inhibited in vivo tumorigenicity (*SI Appendix*, Fig. S10). Taken together, the above observations suggest that IDHwt and IDHm G4 astrocytomas share dependency on a NFIA/B-mediated transcriptional program.

Discussion

Here, we applied single-cell technologies to clinical samples to characterize the genetic and epigenetic heterogeneity within and between astrocytomas and glioblastomas. Previous studies using single-cell RNA-seq and single-cell ATAC-seq have identified tumor cells solely on the basis of gain or loss of whole chromosome or arm (28, 36). Deep profiling of the tumors using snATAC-Seq revealed the presence of focal amplifications and ecDNA in subpopulations of the tumor that are also specific to spatial regions of the tumor. This allowed us to gain a more detailed view of the heterogeneity within each tumor and help us better map the evolutionary trajectory of the tumor cell populations based on the genetic aberrations. We also leveraged the use of multiomics technologies to assess the expression of genes harbored on the amplicons (Fig. 6). As the multiomics field is expanding to include more genomic features that can be measured in parallel for each cell, we will attain a more refined map of heterogeneity and inferred tumor evolution. Using a combination of molecular modalities can help us understand the effect of genetic aberrations, including point mutations on the transcriptional and epigenetic states of the tumor cells.

Our appreciation of the glioblastoma heterogeneity escalates with each generation of technology development, including that presented in this study. The notable subclonal representation between distinct regions of glioblastoma in this study calls into question the appropriateness of the clinical practice where limited biopsies are performed targeting a single region of glioblastomas. Even for resected samples, genomic profiling is typically performed only on limited histological sections. In the context of our finding, inadequate tumor sampling and incomplete catalog of clonal and subclonal representation likely contributed to the failure of targeted therapy against glioblastomas (37–39). The complexity of glioblastoma heterogeneity revealed in this and other single-cell studies (25, 28, 29, 40–43) strongly suggests that a redesign of the paradigm for neurooncology surgical practices is warranted, including consideration for routine regional sampling and the need for clinical-grade, single-cell multiomic platforms.

The complex glioblastoma heterogeneity shown in our study and by others (2, 3) highlights challenges associated with molecularly targeted therapy. For instance, the functional redundancy between Receptor Tyrosine Kinases (RTKs) (44) and subpopulations with mutually exclusive RTK amplifications renders therapeutic success unlikely with single RTK targeting. However, our finding that IDHwt glioblastoma and G4 IDHm astrocytoma cells in distinct cell states not only share open chromatin sites enriched for NFIA and NFIB motifs but also are dependent on

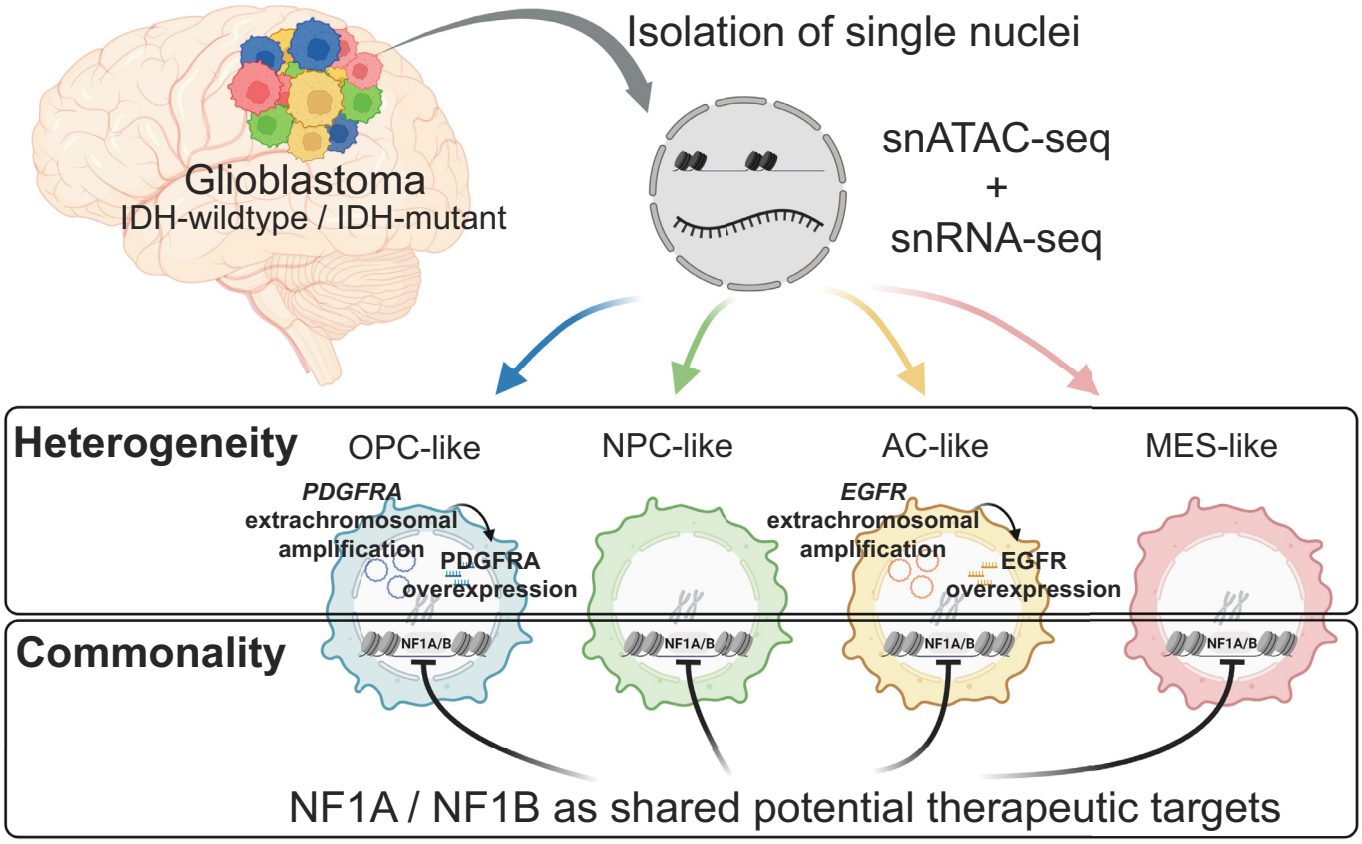

Fig. 6. Summary of the study. Integration of snATAC and snRNA afforded resolution of intratumoral genetic heterogeneity not previously possible, including delineation of cell-to-cell variations in distinct cell states as well as ecDNA. Despite inter- and intra-tumoral heterogeneity, IDHwt glioblastoma and IDHmt G4 astrocytoma cells share dependency on NFI-mediated core transcriptional programs.

NFIA and NFIB for in vitro and in vivo proliferation offers one potential therapeutic strategy in this regard. These findings suggest dependency on NFIA/IB-mediated transcriptional programs shared by tumor cells in distinct genotypic and cell state that may be exploited for therapeutic purposes (Fig. 5). This common dependency may reflect properties shared by the cell of origin and lend insights into glioblastoma pathogenesis. Building on this proof-of-principle study, refined algorithmic motif analysis coupled with snRNA-seq/snATAC-Seq profiling of survival-annotated glioblastoma specimens may unveil other shared vulnerabilities agnostic to intra- and inter-tumoral heterogeneity.

As a proof of principle, our results are limited in terms of mechanistic insight. While our analysis supports putative contribution of NFIA/IB chromatin association and transcriptional modulation in glioblastoma proliferation, we cannot exclude contributions from nontranscriptional roles of NFIA/IB (e.g., in replication). Moreover, multiple transcription factors in addition to NFIA/B likely contribute to tumorigenesis in different subclasses of glioblastomas. Finally, as is the case for many master regulatory transcription factors (8, 45), including STAT3 and MYC, the relationship between expression level and biologic effect appears pleiotropic and nonlinear. While our study suggests that a threshold level of NFIA/IB is necessary for tumorigenicity, others have reported that ectopic expression of NFIA/IB exerts the opposite effect of suppressing tumor proliferation (27, 46) or inducing differentiation (47, 48). These effects are likely influenced by the underlying molecular circuitry or cell state (46). Understanding the complexity of these interactions in astrocyte in distinct cell states is a prerequisite to clinical translation of NFIA/IB-targeting therapeutic strategies.

In summary, our proof-of-principle study demonstrated the potential for snATAC-Seq and multiomics approaches in profiling glioblastoma heterogeneity. Our results suggest that orthogonal intersection of these profiles yields insights into tumor evolution and offers potential for therapeutic discovery.

Materials and Methods

Glioblastoma Tumor Samples. Specimen collection was approved by the IRB at the University of California, San Diego Human Research Protections Program. Each patient was consented by a dedicated research coordinator prior to collection. Written consent was collected from each patient based on processes approved by the ethics committee. The specimens were collected under IRB 120345X. Pathological analysis of the specimens was performed to verify >80% tumor cell content by a board-certified neuropathologist. Samples were collected within 2 h of surgical resection, frozen in liquid nitrogen, and subsequently stored at −80 °C. Tissue was pulverized using a mortar and pestle on liquid nitrogen.

Brain Tissue. Human dorsolateral prefrontal cortex and hippocampus tissue of a healthy 31-y-old male were obtained from the National Institute of Child Health and Human Development Brain Bank for Developmental Disorders (GSE140493–UMB4540) (49). Ethical approval was obtained from the University Health Network and The Hospital for Sick Children for use of the tissues.

Single-Nucleus ATAC-Seq (snATAC-Seq). Combinatorial barcoding single-nucleus ATAC-seq was performed as described previously (17, 18) . Libraries were sequenced on a HiSeq2500 sequencer (Illumina) using custom sequencing primers and following read lengths: 50 + 43 + 37 + 50 (Read1 + Index1 + Index2 + Read2).

Analysis of snATAC-Seq Data.

Mapping, clustering, and visualization. Processing of snATAC-Seq data was performed using SnapATAC (18), and reads were aligned to the hg38 genome using bowtie2 with the following parameters: bowtie2 -X2000 –no-mixed –no-discordant. PCR duplicates were removed using the MarkDuplicate function from Picard tools (http://broadinstitute.github.io/picard) for each cell. Cells with less than 1,000 reads and less than 20% of reads at promoters (GENCODE v29 annotation) were filtered from downstream analysis. Clustering was performed using SnapATAC using 5 kb bins filtered for blacklist regions annotated by ENCODE. Clusters were determined using the Leiden algorithm and visualization using UMAP. For the merged clustering results, Harmony was used to correct for sample and batch effects (50). TDF files for visualizing the aggregate of cluster signal on IGV were generated using IGVtools (51).

Large-scale CNV analysis. To detect large CNVs, a CNV score was detected in 20 MB bins across the genome as previously described in the study by Satpathy et al. (36). For each cell, the CNV score is the log2 ratio of the binary counts in each 20 MB bin to the mean of the nearest 50 bins based on GC content. For heatmap visualization, hierarchical clustering was performed for cells in each snATAC-Seq-defined cluster separately using fastcluster and parallelDist R packages using default parameters.

Focal amplification detection. Amplifications were detected for each cluster in each tumor sample individually. First, background reads in 50 kb bins (filtered for blacklist regions) were normalized to the total number of reads across all cells in that cluster and the mean of the nearest bins based on GC content. The same procedure was applied to astrocyte and oligodendrocyte progenitor cells from a nontumor sample to serve as a control. A log-ratio score was calculated for each 50 kb bin between tumor cluster and the aggregate of the nontumor cells. This was used as the input to the circular binary segmentation algorithm (implemented in DNAcopy R package to determine contiguous segments in the genome). To determine amplified regions, we calculated a threshold cutoff based on the mean and three times the standard deviation of the segments.mean values. Segments with a mean value greater than the threshold were further filtered for regions of low mappability (score < 0.6) downloaded from ichorCNA (52) and having a score greater than the 50th percentile of normalized counts in the tumor cluster. Amplifications within 1 Mb were merged.

Clustering of CNV and focal amplification profiles. The top 10% of cells based on read count from each cluster were used for clustering analysis. For each cell, the CNV score for the chromosomes was determined based on the mean score across 20 MB windows in that chromosome. A z-score was used for the focal amplifications based on the normalized ATAC-seq in that region. Clustering was performed based on the chromosome and focal amplification profiles for each cell using the Manhattan distance and the complete-linkage clustering method.

Peak calling and differential analysis. Peaks were called for each individual cluster and then merged. MACS2 (53) was used for peak calling with the following parameters: macs2 callpeak -t sample -n sample -f BAM -g hs –nomodel –shift -100 –extsize 200 –slocal 1,000 –keep-dup all -B –SPMR –call-summits -q 1e-4. In order to account for the high signal in amplified regions, a local background was used for peak calling. Aggregate cluster peak counts in heatmaps based on counts in each peak normalized to the sum of reads in all 5kb bins for cells in each cluster. Differential analysis for the merged clustering was performed on peaks specific to the tumor clusters using Monocle2 (54) with clusterID as a variable. Significant peaks were determined based on q-value < 1e-20 and fold-change > 1.5 between each cluster and all other clusters (normalized counts in peaks). Motif analysis was performed on the cluster-specific peaks using Homer (55) (implemented in SnapATAC (18)).

Single-Nucleus RNA-Seq (snRNA-Seq). Frozen tissue was ground in liquid nitrogen using a mortar and pestle. Approximately 30 mg of ground tissue was incubated with 500 μL nuclei isolation buffer [2% BSA, 0.1% Triton-X cOmplete (Roche), 1 mM DTT, 0.2 U/μL RNAsin (Promega) in PBS] for 10 min at 4 °C. After centrifugation with 500 × g for 5 min, the nuclei were resuspended in sort buffer [2 % BSA, 1 mM EDTA, 0.2 U/μL RNAsin (Promega) in PBS]. Nuclei suspension was filtered and sorted using a SH800 sorter (Sony) after staining with the DNA-dye DRAQ7 (3 μM, Cell Signaling Technologies). The sorted nuclei were pelleted (15 min, 1,000 × g) and resuspended in PBS containing 1 % BSA and 0.2 U/μL RNAsin (Promega), at a concentration of 1,000 nuclei/μL. snRNA-Seq libraries were constructed using the Chromium™ Single Cell 3′ v2 Library kit (10× Genomics, PN-120237) according to manufacturer's descriptions. Reverse transcription and other amplification steps were carried out on a T100 thermal cycler (Bio-Rad). After reverse transcription, GEMs (gel beads in emulsion) were lysed and cDNA was cleaned up with MyOne Silane Beads (Thermo Fisher Scientific). Next, single stranded cDNA was PCR-amplified for 14 cycles and purified using SPRIselect Reagent Kit (Beckman Coulter). CDNA was enzymatically fragmented followed by double size selection with SPRIselect Reagent Kit (200 to 700 bp,

0.6× and 0.8×, Beckman Coulter). Subsequently, adaptors were ligated and libraries were constructed by PCR. Another round of double size selection was performed using SPRIselect Reagent Kit (200 to 700 bp, Beckman Coulter). Final libraries were quantified using Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific), and size distribution was measured using TapeStation (High Sensitivity D1000, Agilent). The average fragment size of successful libraries was 500 bp. The libraries were loaded at a concentration of 13 pM and sequenced on a HiSeq 2500 sequencer (Illumina) with the following parameters: Read 1 26 cycles; Index 1 8 cycles; Read 2 98 cycles.

Analysis of snRNA-Seq.

snRNA-Seq alignment, clustering, & marker gene identification. Samples were mapped using the STAR aligner (56) to the GENCODE version 27 (GRCh38.p10) genome. For downstream analysis, to identify cell types and merge samples, we used the Seurat version 3 package (57). We first filtered out cells that had less than 200 expressed genes and all genes that were expressed in ten or fewer cells. In order to filter out potential doublets, we used a filter on the ratio of unique molecular identifier (UMI)/genes in all cells. This ratio showed a bimodal distribution. Cells with a higher ratio of UMI/gene were filtered out. We scaled each gene UMI count by 1e4, normalized and log transformed the expression values in each cell (16), and then scaled to z scores for each gene across all cells. For cluster identification, we used the FindClusters function in the Seurat package, and then used FindMarkers function to identify genes that were significantly up-regulated in the cells in one cluster versus all other cells.

WGS. WGS was performed using the TruSeq DNA PCR-free low-throughput library prep kit (Cat: 20015962). Paired-end 100 bp reads were generated for RL3-section 3, and paired-end 50 bp reads were generated for RL1-section 2. Data were analyzed using the AmpliconArchitect (22) and PrepareAA (github.com/jluebeck/PrepareAA) using the hg19 genome and default settings. For RL3, additional filtering of reads was performed to remove everted read pairs. Amplicon intervals with copy number greater than 5 were converted to hg38 coordinates using the UCSC liftOver tool (58), and any amplicons with an average mappability score less than 0.6 (50 kb bins from ichorCNA) and duplicated regions in hg38 were removed.

WGS data were also mapped to the hg38 genome using bowtie2 (59) (options: –no-mixed –no-discordant), and reads were further filtered using SAMtools (60) (options: -F 1804 –q 30). TDF files were generated for visualization using IGVtools (51).

Paired-Seq. Paired-Seq data for RL3 were generated as previously described with minor modifications. Tumor dissections were ground and incubated in 1 mL NPB [5% BSA (Sigma, A7906), 0.2% IGEPAL-CA630 (Sigma, I8896), 1 mM DTT (Sigma, D9779) in PBS] supplemented with 1× cOmplete EDTA-free protease inhibitor (Roche, 05056489001), 0.4 U/μL RNase OUT (Invitrogen, 10777-019), and 0.4 U/μL SUPERase In (Invitrogen, AM2694) at 4 °C for 10 min. Single-nuclei suspensions were then filtered with 30 μM Cell-Tric (Sysmex) and counted with Bio-Rad TC20 cell counter. A total of 500,000 nuclei were split into five tubes for tagmentation, followed by reverse transcription with tube-specific T_{15} and N_6 barcoded primers. Two rounds of ligation-based combinatorial barcoding were then carried out to label the nuclei (16, 61). The nuclei were then lyzed, and DNA and cDNA were cleaned up with SPRI beads (Beckman coulter, B23319). Library construction was carried out as previously described (16). Briefly, preamplification, endonuclease digestion, and 2nd adaptor tagging were sequentially performed, followed by sequencing with Illumina HiSeq 4000 with the following read lengths: PE 100 + 7 + 100 (read1 + index 1 + read2). Cellular barcode sequences were first extracted from Read2 of both DNA and RNA libraries of Paired-Seq experiments and then mapped to all possible barcode combinations with no more than 1 mismatch. Barcode-collision rate was estimated as 4.3% according to the species-mixing experiment parallelly performed.

RNA reads were mapped to reference genome (GRCh38 with annotation GENCODE v29) using STAR (56) with default parameters. PCR duplicates were removed according to the genomic position of mapped reads, cellular barcodes, and UMI. Cell-to-gene count matrix was generated for the RNA data and processed using Seurat to obtain normalized gene counts per cell. DNA reads were mapped to reference genome (GRCh38) using bowtie2 (59) and processed similarly to snATAC-Seq data. The Paired-Seq:ATAC reads were combined with the snATAC-Seq data for RL1-section 3 sample using SnapATAC (18), and Harmony was used to merge the samples. Clustering was performed on the merged sample using Leiden clustering (50).

Construct a Transcription Factor Regulatory Network.

Link cCREs to genes. We analyzed single-cell ATAC-seq data to link cCREs to gene promoters. We first clustered and aggregated the cells into pseudocells using the Leiden algorithm (62) such that each pseudocell contains at least 25 cells. Following this, we generated two matrices: one illustrating the read counts for each peak in every pseudocell, and another representing the gene activity within each pseudocell. The gene activity is determined by the average chromatin accessibility of the gene promoter plus the gene body in the pseudocell. The initial cCRE–gene links were created by linking cCREs to nearby gene promoters (±5 Mb within gene's TSS). To refine these connections, an elastic net model was employed, which predicted gene activity based on the read counts of linked cCREs. Only the connections with nonzero coefficients were retained for further analysis.

Link transcription factor motifs to genes. We searched for the binding sites of 1,165 distinct TF motifs from the cisBP database in cCREs using the FIMO algorithm (63) with *P*-value cutoff of 1e-5. A motif is linked to a gene if a cCRE containing the motif is also linked to the same gene.

Refine the TF–gene links using single-cell RNA-seq data. We used single-cell RNA-seq data to refine the TF–gene links. We first clustered and aggregated the cells into pseudocells using the Leiden algorithm as described above. Then, we trained an elastic net model to predict the gene expression from the linked TFs. We kept the links with nonzero coefficients.

Perform GSEA of NFIA's target genes. We identified the target genes of NFIA from the TF regulatory network constructed above, weighted by their regression coefficients. The weights of target genes from multiple sample (RL1 and RL3) were averaged. The final gene list and weights were input into the GSEA preranked method (64) to identify the statistically significant enriched pathways using the "KEGG_2021_Human" gene set from the Enrichr website (65).

Perform PageRank analysis of the TF regulatory network. Using the approach outlined above, we constructed cell type-specific networks where edges connect transcription factors (TFs) to their target genes, with edge weights determined by the coefficients of the regression models. To calculate the relative contributions of TFs to the expression levels of target genes, we first determined the standardized expression levels of the parent nodes (TFs linked to the gene) for each gene in the network. Next, we multiplied the standardized TF expression level by the corresponding regression coefficient for each edge. This product represents the rate of change in gene expression, and we used these values as the updated edge weights in the network. Finally, we performed the PageRank analysis as described in ref. 66 on networks from each cell state to determine the ranking scores of TFs.

Inducible Knockdown of NFIA and NFIB in Glioblastoma and Grade 4 Astrocytoma Cells. ShRNA sequences against NFIA and NFIB or scrambled shRNA (control) listed below were synthesized (University of Minnesota Genomics Center) and cloned into inducible lentivector Tet-pLKO-puro (Addgene) according to the manufacturer's instructions.

NFIA-forward: 5′-CCGGTTCCCGACATCACCCATTATCCTCGAGGATAATGGGTGATG TCGGGAATTTTT-3′,

NFIA-reverse: 5′- AATTAAAAATTCCCGACATCACCCATTATCCTCGAGGATAATGGGT GATGTCGGGAA-3′,

NFIB-forward: 5′-CCGGGCAGCACTTACAATCACTAATCTCGAGATTAGTGATTGTAA GTGCTGCTTTTT-3′,

NFIB-reverse: 5′-AATTAAAAAGCAGCACTTACAATCACTAATCTCGAGATTAGTGATT GTAAGTGCTGC-3′,

Scrambled-forward: 5′-CCGGGCTTCGCGCCGTAGTCTTACTCGAGTAAGACT ACGGCGCGAAGCTTTTT-3′,

Scrambled-reverse: 5′- AATTAAAAAGCTTCGCGCCGTAGTCTTACTCGAGTAAGACT ACGGCGCGAAGC-3′.

293T cells were transfected with packaging plasmid (psPAX2), envelope plasmid (pCMV-VSV-G), and shRNA-cloned Tet-pLKO-puro using FuGENE HD Transfection Reagent (Promega). Lentivirus was harvested at 48h after transfection and used to infect glioblastoma PDX cells, followed by selection with puromycin (0.3 to 0.6 μg/mL, Sigma) for 7 to 10 d. Cells were induced with doxycycline (Sigma; 1 μg/mL) for 4 d, and silencing was assessed by western blot and/or qRT-PCR. Cells with NFIA or NFIB knockdown were further selected by single-cell cloning under puromycin and confirmed by western blot analysis. ShRNA sequences are listed above.

Western Blotting. Protein (10 μg/lane) was electrophoresed in 4 to 15% SDS-PAGE gel (Bio-Rad) and transferred to PVDF membranes (Bio-Rad) by electroblotting. Membranes were blocked with 5% nonfat dry milk or 5% bovine serum albumin (BSA, sigma) for 1 h at room temperature and then incubated with primary antibodies against NFIA (Active Motif) and NFIB (NSJ Bioreagents) at 4 °C overnight. Membranes were then washed in washing buffer TBST (20 mM Tris pH7.5, 150 mM NaCl, 0.1% Tween20) and incubated with peroxidase-conjugated secondary antibodies (Cell Signaling Technology) for 1h at room temperature. Signals were visualized with an ECL kit (BioRad).

Quantitative Reverse Transcription-PCR. Cells with or without shRNA against NFIA or NFIB, or control shRNA, were treated with Dox (1 μg/mL) for 4 d, and then harvested for RNA extraction with QIAGEN RNeasy Mini Kit according to the manufacturer's protocol. High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) was applied to perform reverse transcription reaction, followed by SYBR Green-Based Quantitative PCR (Bio-Rad). mRNA levels were normalized to GAPDH (dCt = Ct of the gene of interest – Ct of the GAPDH) and reported as relative mRNA expression [ddCt = $2^{\wedge -(\text{dCt sample} - \text{dCt control})}$] in fold change. The primers listed below were used.

NFIA-forward: 5′- AGCTCATGGAGCGGCAATAG-3′,
NFIA-reverse: 5′- ATTCATCCTGGGTGAGACAGAG-3′,
NFIB-forward: 5′- GGAAATCTAGCACGCCGAGT-3′,
NFIB-reverse: 5′- ATTCATCCTGAGTGAGACAGATGG-3′.

Animal Study. Glioblastoma cells were dissociated or trypsinized into single cells and stereotactically implanted intracerebrally (right striatum, 2.5-mm lateral from bregma and 2.5-mm deep) in 7 to 8-wk-old female athymic mice (Charles River Laboratories). To induce NFIA or NFIB knockdown in vivo, doxycycline (1 mg/mL) was administrated through drinking water at day 7 after implantation. All in vivo procedures were approved by the Institutional Animal Care and Use Committee at the University of Minnesota.

Cell Lines. The human astrocyte line was purchased from ScienCell Research Laboratories. Patient-derived glioblastoma cell lines, GBM83, CMK3 (8), and TS603 (35), were propagated under neurosphere conditions in NeuroCult medium (Stem Cell Technology) supplemented with B27, GlutaMAX, epidermal growth factor, fibroblast growth factor, and heparin according to the manufacturer's instructions.

Data, Materials, and Software Availability. Code for CNV analysis and focal amplification detection from snATAC-Seq data can be found on github: https://github.com/rr1859/AAmpD (67). Processed data have been deposited in GEO (GSE165037) (68). All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We thank the members of the Ren lab, Bojing Jia, Elise Vu, Sihan Wu, Shunichiro Miki, and Jeremy Rich, and members of his lab at UCSD for their input and discussion on this project. We also thank Federico Gaiti for comments on the manuscript. This work was supported by the Ludwig Institute for Cancer Research to B.R., NIH NS097649-01 and CA240953-01 to C.C.C, and the Ruth L. Kirschstein National Research Service Award NIH/NCI T32 CA009523 to R.R. Funding was also provided by Bristol Myers Squibb (Celgene). V.B. was supported in part by grants from the NIH (R01GM114362 and U24CA264379). P.S.M. and V.B. are part of the eDyNAmiC team supported by the Cancer Grand Challenges partnership funded by Cancer Research UK (CRUK) (P.S.M., CGCATF-2021/100012; V.B. CGCATF-2021/100025) and the National Cancer Institute (P.S.M. OT2CA278688; V.B. OT2CA278635).

Author affiliations: [a]Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, CA 92093; [b]Center for Epigenomics, University of California San Diego, La Jolla, CA 92093; [c]Institute of Experimental and Clinical Pharmacology and Toxicology, Faculty of Medicine, University of Freiburg, Freiburg, Germany; [d]Department of Neurosurgery, University of Minnesota, Minneapolis, MN 55455; [e]Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, NY 10065; [f]Department of Computer Science and Engineering, Halicioglu Data Science Institute, University of California San Diego, La Jolla, CA 92093; [g]Department of Computer Science and Engineering, Biomedical Sciences Graduate Program, San Diego, CA 92121; [h]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093; [i]Department of Medicine, University of California San Diego, La Jolla, CA 92093; [j]Program in Neurosciences and Mental Health, Hospital for Sick Children, Division of Experimental & Translational Neuroscience, Krembil Research Institute, University Health Network, Toronto, ON M5T 0S8, Canada; [k]Department of Psychiatry, University of Toronto, Toronto, ON M5T 0S8, Canada; [l]Department of Physiology, University of Toronto, Toronto, ON M5T 0S8, Canada; and [m]Department of Pathology, Stanford University, Stanford, CA 94305

Author contributions: B.R. and C.C.C. designed research; R.R., S.P., J.N., S.W., C.Z., J.Y.H., X.H., Z.Y., and A.K.M. performed research; R.R., A.R., K.V.D., Y.E.L., R.F., and B.R. analyzed data; and R.R., T.K., K.Z., C.W.B., J.L., T.B., J.M., C.Z.H., C.K.G., C.L.B., P.S.M., V.B., L.E., B.R. and C.C.C. wrote the paper.

Competing interest statement: B.R. is a cofounder of Arima Genomics Inc and Epigenome Technologies, Inc. P.S.M. is a cofounder of Boundless Bio, Inc. He has equity in the company and chairs the Scientific Advisory Board, for which he is compensated. V.B. is a cofounder, consultant, and SAB member and has equity interest in Boundless Bio, Inc. (BB) and Digital Proteomics, LLC (DP), and receives income from DP. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflicts of interest policies.

1. P. Y. Wen, S. Kesari, Malignant gliomas in adults. *N. Engl. J. Med.* **359**, 492–507 (2008).
2. C. W. Brennan *et al.*, The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
3. C. Neftel *et al.*, An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e21 (2019).
4. H. Noushmehr *et al.*, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
5. D. W. Parsons *et al.*, An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
6. R. G. Verhaak *et al.*, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
7. D. N. Louis *et al.*, The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro-Oncol.* **23**, 1231–1251 (2021).
8. D. Kozono *et al.*, Dynamic epigenetic regulation of glioblastoma tumorigenicity through LSD1 modulation of MYC expression. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4055–E4064 (2015).
9. M. L. Suvà *et al.*, Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014).
10. J. D. Buenrostro *et al.*, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
11. D. A. Cusanovich *et al.*, Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
12. D. Gomez, L. S. Shankman, A. T. Nguyen, G. K. Owens, Detection of histone modifications at specific gene loci in single cells in histological sections. *Nat. Methods* **10**, 171–177 (2013).
13. H. S. Kaya-Okur *et al.*, CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
14. A. Rotem *et al.*, Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
15. S. Turcan *et al.*, IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483 (2012).
16. C. Zhu *et al.*, An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 (2019).
17. S. Preissl *et al.*, Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).
18. R. Fang *et al.*, Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* **12**, 1–15 (2021).
19. M. R. Corces *et al.*, The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
20. A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
21. R. G. W. Verhaak, V. Bafna, P. S. Mischel, Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat. Rev. Cancer* **19**, 283–288 (2019).
22. V. Deshpande *et al.*, Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).
23. H. Kim *et al.*, Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).
24. N. M. Jameson *et al.*, Intron 1-mediated regulation of EGFR expression in EGFR-dependent malignancies is mediated by AP-1 and BET proteins. *Mol. Cancer Res.* **17**, 2208–2220 (2019).
25. L. Wang *et al.*, The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer Discov.* **9**, 1708–1719 (2019).
26. B. Deneen *et al.*, The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron* **52**, 953–968 (2006).
27. H.-R. Song *et al.*, Nuclear factor IA is expressed in astrocytomas and is associated with improved survival. *Neuro. Oncol.* **12**, 122–132 (2010).
28. A. P. Patel *et al.*, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
29. C. Neftel *et al.*, An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849 e821 (2019).
30. Z. Tang, B. Kang, C. Li, T. Chen, Z. Zhang, GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560 (2019).
31. K. Zhang, M. Wang, Y. Zhao, W. Wang, Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci. Adv.* **5**, eaav3262 (2019).
32. K. K. Resendes, A. G. Rosmarin, Sp1 control of gene expression in myeloid cells. *Crit Rev. Eukaryot Gene Expr.* **14**, 171–181 (2004).
33. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
34. S. Turcan *et al.*, Efficient induction of differentiation and growth inhibition in IDH1 mutant glioma cells by the DNMT Inhibitor Decitabine. *Oncotarget* **4**, 1729 (2013).
35. D. Rohle *et al.*, An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* **340**, 626–630 (2013).

36. A. T. Satpathy *et al.*, Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
37. A. Sottoriva *et al.*, Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4009–4014 (2013).
38. N. R. Parker, P. Khong, J. F. Parkinson, V. M. Howell, H. R. Wheeler, Molecular heterogeneity in glioblastoma: Potential clinical implications. *Front. Oncol.* **5**, 55 (2015).
39. H. Kim *et al.*, Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* **25**, 316–327 (2015).
40. S. Darmanis *et al.*, Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* **21**, 1399–1410 (2017).
41. P. Guilhamon *et al.*, Single-cell chromatin accessibility profiling of glioblastoma identifies an invasive cancer stem cell population associated with lower survival. *eLife* **10**, 364090 (2021).
42. A. Bhaduri *et al.*, Outer radial Glia-like cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell* **26**, 48–63.e6 (2020).
43. C. P. Couturier *et al.*, Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Commun.* **11**, 3406 (2020).
44. J. M. Stommel *et al.*, Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318**, 287–290 (2007).
45. L. Avalle *et al.*, MicroRNAs-143 and-145 induce epithelial to mesenchymal transition and modulate the expression of junction proteins. *Cell Death Differ.* **24**, 1750–1760 (2017).
46. B. W. Stringer *et al.*, Nuclear factor one B (NFIB) encodes a subtype-specific tumour suppressor in glioblastoma. *Oncotarget* **7**, 29306 (2016).
47. M. Piper *et al.*, NFIB-mediated repression of the epigenetic factor Ezh2 regulates cortical development. *J. Neurosci.* **34**, 2921–2930 (2014).
48. G. Steele-Perkins *et al.*, The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol. Cell Biol.* **25**, 685–698 (2005).
49. C. Luo *et al.*, Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genom.* **2**, 100107 (2022).
50. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289–1296 (2019).
51. J. T. Robinson *et al.*, Integrative genomics viewer. *Nature Biotechnology* **29**, 24–26 (2011).
52. V. A. Adalsteinsson *et al.*, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).
53. T. Liu, Use model-based analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. *Methods Mol. Biol.* **1150**, 81–95 (2014).
54. X. Qiu *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
55. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**, 576-589 (2010).
56. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. T. Stuart *et al.*, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
58. C. Tyner *et al.*, The UCSC Genome Browser database: 2017 update. *Nucleic Acids Research* **45**, D626–D634 (2017).
59. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
60. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
61. A. B. Rosenberg *et al.*, Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
62. V. A. Traag, L. Waltman, N. J. v. Eck, From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).
63. T. L. Bailey *et al.*, MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, 202–208 (2009).
64. A. Subramanian *et al.*, Gene set enrichment analysis. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
65. E. Y. Chen *et al.*, Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128–128 (2013).
66. K. Zhang, M. Wang, Y. Zhao, W. Wang, Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci. Adv.* **5**, eaav3262 (2019).
67. R. Raviram, Workflow to detect ecDNA from snATAC-Seq data. Github. https://github.com/rr1859/AAmpD. Deposited 30 January 2021.
68. R. Raviram *et al.*, snATAC-Seq and snRNA-Seq data of gliomas. NCBI *Gene Expression* Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165037. Deposited 19 January 2021.