Identification of functional immune and neuronal tumor cells in glioma

Rachel Naomi Curry, B.A.[1,2], Malcolm F. McDonald, B.S. [3,4,+], Qianqian Ma, Ph.D. [5+], Jochen Meyer, Ph.D.[6], Isamu Aiba, Ph.D.[6], Brittney Lozzi [2], Alexis Cervantes [2], Yeunjung Ko, M.S. [2], Estefania Luna-Figuera [2], Dong-Joo Choi, Ph.D. [2], Zhung-Fu Lee [2], Junzhan Jing, Ph.D. [2], Arif O. Harmanci, Ph.D. [7], Anna Rosenbaum [2], Peihao He [2], Carrie Mohila, M.D. Ph.D. [8], Ali Jalali, M.D. Ph.D. [9], Jeffrey Noebels, M.D., Ph.D. [5,6,11], Xiaolong Jiang, Ph.D. [5,10], Benjamin Deneen, Ph.D.[2,5,9,*,‡], Ganesh Rao, M.D.[9,*,‡], Akdes Serin Harmanci, Ph.D.[9,*,‡]

[1] Graduate School of Biomedical Sciences, Baylor College of Medicine, Houston, TX
[2] Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX
[3] Medical Scientist Training Program, Baylor College of Medicine, Houston, TX
[4] Development, Disease Models, and Therapeutics, Baylor College of Medicine, Houston, TX
[5] Department of Neuroscience, Baylor College of Medicine, Houston, TX
[6] Department of Neurology, Baylor College of Medicine, Houston, TX
[7] School of Biomedical Informatics, Center for Precision Health, University of Texas Health Science Center, Houston, TX
[8] Department of Pathology, Texas Children's Hospital, Houston, TX
[9] Department of Neurosurgery, Baylor College of Medicine, Houston, TX
[10] Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX
[11] Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX
[+] These authors contributed equally to this work
* These authors contributed equally to this work
[‡] Corresponding Author

Abstract

Despite advances in cancer molecular profiling, successful therapeutic development has been hindered by challenges in identifying tumor-specific mechanisms that can be targeted without consequence to healthy tissue. Discrimination between tumor and host cells that comprise the tumor microenvironment remains a difficult yet important task for defining tumor cell signatures. Correspondingly, a computational framework capable of accurately distinguishing tumor from non-tumor cells has yet to be developed. Cell annotation algorithms are largely unable to assign integrated genomic and transcriptional profiles to single cells on a cell-by-cell basis. To address this, we developed the Single Cell Rule Association Mining (SCRAM) tool that integrates RNA-inferred genomic alterations with co-occurring cell type transcriptional signatures for individual cells. Applying our pipeline to human and mouse glioma, we identified tumor cell trajectories that recapitulate temporally-restricted developmental paradigms and feature unique co-occurring genomic and transcriptomic identities. Specifically, we describe and validate two previously unreported tumor cell populations with immune and neuronal signatures as hallmarks of human glioma subtypes. *In vivo* modeling revealed an immune-like tumor cell population can direct CD8+ T cell responses and survival outcomes. In parallel, electrophysiology and Patch-seq studies in human tumors confirmed a frequent subset of neuronal-like glioma cells that fire action potentials but retain the morphology of glia. These collective studies report the existence of new glioma cell types with functional properties akin to their non-tumor analogs and demonstrate the ability of SCRAM to identify and characterize these cell types in unprecedented detail.

Introduction

In the era of cancer genomics, the advent of high-throughput single cell sequencing technologies has cleared the way for examination of cellular heterogeneity in genomic, transcriptomic and epigenomic detail[1]. Studies employing these technologies have demonstrated dynamic cellular and subcellular hierarchies that are spatiotemporally distinct and have revealed transcriptional profiles that can be utilized to classify tumors into clinically-relevant molecular subtypes[2–5]. While these technologies have elucidated the diverse heterogeneity that exists within glioma tumors, the ability of sequencing pipelines to reliably identify tumor cells and the co-occurring genomic and transcriptomic cell states that define them has yet to be attained. Within malignant glioma, resolution of this complexity represents a significant impediment to therapeutic development that is further compounded by a diverse cellular constituency and a complex array of tumor-specific genetic variants. Furthermore, diffusely infiltrating tumor cells resembling glia and associated neural progenitors pose a considerable challenge to distinguishing between glioma and host cells, which if overcome could enable more precise characterization of tumors and important clinical correlates. To this end, we have developed a novel single cell computational tool, Single Cell Rule Association Mining (SCRAM) capable of accurately identifying tumor cells and defining co-occurring cellular states. Within our pipeline, we integrate three orthogonal tumor features to identify tumor cells in single cell resolution: (1) cell type transcriptional profiling; (2) RNA-inferred copy number variant (CNV) calling; and (3) RNA-inferred mutational analysis for single nucleotide variants (SNVs). Our studies reveal that more than half of tumor cells feature transcriptional profiles matching more than one cell type, exposing the extraordinary cellular complexity of glioma and the ability of our SCRAM pipeline to define it with both accuracy and precision. Furthermore, we illustrate the unique potential for tumor cells to acquire functional immune-like and neuronal-like cellular states and highlight the utility of using our computational framework for characterization of pseudo cell types in cancer, development and disease.

Results

SCRAM accurately identifies transcriptional states and reliably discriminates between tumor and non-tumor cell types

To validate SCRAM as a reliable computational framework (**Fig. 1A**), we utilized previously published human and mouse scRNA-seq datasets to examine whether SCRAM annotation is consistent with established transcriptional profiles. We found that SCRAM was successful at annotating >75% of cells across eight scRNA-seq datasets (**Fig. 1B**) and uncovered that >83% of cells possess transcriptional profiles meeting the criteria for more than one cell type (**Fig. 1C**). Employing established glioma transcriptional profiles[5,6], we assigned our manually-curated cell type list to these transcriptional classes and found that SCRAM-assigned identities were concordant with previous reports. Surprisingly, our analysis found that more than two thirds of sequenced cells exhibited co-occurring cell type annotations (**Figs. 1D-G**), suggesting that glioma cells possess differentially expressed gene (DEG) transcriptional signatures consistent with multiple cell states. Further validation using the 10X Genomics Spatial Transcriptomics dataset of *IDH1* wildtype ($IDH1^{WT}$) glioma unveiled that various permutations of these co-occurring cell states exist within single cell clusters, demonstrating that clusters encompass a heterogenous group of cell types with a variety of genomic and transcriptomic profiles (**Figs. 1H-J**). Exploiting the ability of SCRAM to define both transcriptional signatures and RNA-inferred genomic CNVs, we were able to specify which cells were tumor and further characterize

their respective cell lineages. Our results confirmed that tumor cells are spatially dispersed throughout the non-tumor landscape and uncover diverse transcriptomic cellular profiles within a small anatomical region. These analyses validate that SCRAM is capable of recapitulating previously defined cell annotations and can be used to further resolve intratumoral heterogeneity by defining both co-occurring cell identities and cellular states with higher resolution.

Having confirmed SCRAM's capacity to accurately assign cell identities, we next assessed its ability to distinguish tumor cells from non-tumor cells. Accordingly, we analyzed non-tumor human and mouse cortex scRNA-seq datasets from the Allen Brain Atlas, finding that SCRAM could reliably decipher between tumor and non-tumor constituents with >99% accuracy across species. SCRAM was able to annotate >88% of cells in these datasets, replicating 60% and 86% of the same cell annotations provided by Allen Brain Atlas for human and mouse datasets, respectively (**Figs. 1K-N**). Our analyses unexpectedly showed that >60% of cells in the mammalian non-tumor brain have expression profiles that qualify as more than one cell type and alluded to the existence of cell populations in both species that do not meet the transcriptional requirements of known cell types. These data demonstrate that SCRAM reliably distinguishes between tumor and non-tumor cell types and accurately assigns cell identities in both human and mouse. Furthermore, these analyses expose previously unreported cellular heterogeneity in the mammalian brain by defining unique co-occurring transcriptional cellular states in which cells possess co-occurring cell signatures.

SCRAM reveals unique integrated genomic and transcriptomic signatures in human glioma

Having validated SCRAM, we next aimed to define the genomic and transcriptomic landscape of tumor cells in our integrated scRNA-seq dataset consisting of 195,063 cells from seven IDH1^{WT} and *IDH1* mutant (IDH1^{mut}) glioma patients. SCRAM analyses employing the CaSpER[7] CNV calling algorithm confirmed that IDH1^{WT} tumors had increased CNV incidence (**Figs. 2A-C,2F-H**) whereas chromosome 1p19q *co*deletions were the predominant alterations for IDH1^{mut} oligodendroglioma. Parallel analyses using XCVATR[8] to call SNVs validated that IDH1^{mut} tumors are marked by *IDH1* mutations and that IDH1^{WT} tumors feature *EGFR*, *TP53* and *PTEN* mutations (**Figs. 2D,E,I,J**). While DNA-based sequencing studies have demonstrated a higher mutational burden in IDH1^{WT} tumors, our RNA-inferred analyses exposed more frequent transcriptional penetrance of SNVs in IDH1^{mut} tumors in our cohort, signifying that transcribed mutations at the mRNA level are more frequently encountered in IDH1^{mut} glioma. Of particular interest was high incidence of mRNA mutations in a transcript mapping to the polycomb repressive complex 2 (PRC2) *SUZ12*. Further examination of the mutation-bearing region revealed that the mutated *SUZ12* sequence is identical to the *SUZ12* pseudogene sequence, *SUZ12P1*. This transcript is detected at high levels in all glioma patients, suggesting that either *SUZ12* mutations are the most transcribed SNVs in our cohort, or that the *SUZ12P1* pseudogene is being actively transcribed in glioma patients (**Figs. 2K,L**). Prior reports have observed hypermethylation of PRC2 regions in malignant tumors, which is pronounced in IDH1^{mut} subtypes[9,10]. Whether high transcriptional penetrance of *SUZ12P1* or mutated *SUZ12* mechanistically contributes to these hypermethylation phenotypes in glioma warrants future mechanistic inquiries.

Using CNV and SNV annotation in conjunction with high expression of established tumor markers (**Fig. 2M**), cells were assigned tumor or non-tumor cell status (**Fig. 2N**). Tumor cells in both IDH1^{WT} and IDH1^{mut} samples primarily mapped to one of three SCRAM-assigned cell classes (embryonic-neural, embryonic or neural) and corresponded to one of four co-occurring cell lineages (glial, glial-neuronal, myeloid and neurodevelopmental), emphasizing convergent

transcriptomic features shared between glioma subtypes (**Figs. 2O,P**). An examination of cell types by patient sample uncovered divergent cellular heterogeneity both within and between patients and illuminated prodigious myeloid-derived immune cell signatures within IDH1^{mut} tumor core samples (**Fig. 2Q**). Importantly, 17% of tumor cells were marked only by CNVs, demonstrating that a population of tumor cells exists for which no corresponding cell type signature can be assigned. Using the SCRAM *de novo* marker function, we were able to identify differentially expressed markers of these tumor cells (**Table S2**), the function of which may serve as the basis for future scientific endeavors. Collectively, these data demonstrate that IDH1^{WT} and IDH1^{mut} tumors are defined by unique co-occurring genomic and transcriptomic profiles that exist in varying proportions between glioma subtypes and highlight the utility of SCRAM in defining transcribed mutations and integrated cell states in human glioma.

SCRAM elucidates novel developmental tumor cell trajectories and defines new intermediate cell states in glioma

Given the former observations and prior work demonstrating that spatially-distinct tumor regions possess unique transcriptomic and genomic signatures within individual tumors[11], we sought to identify which co-occurring cell types occur in our glioma dataset and to resolve these cell types spatially. An analysis of each Seurat-generated cluster revealed that clusters are exceedingly diverse and frequently feature a combination of tumor and non-tumor cells that share similar transcriptional expression patterns (e.g., astrocyte and astrocyte-like tumor). The largest cluster of tumor cells in our human dataset (Cluster 1) contained cells that were uniformly assigned the cell type annotation of astrocyte and were marked by a variety of co-occurring chromosome 5p, 5q, 7p, 7q and 17q amplifications and chromosome 10p, 10q and 22q losses (**Figs. 3A-D**). Visualization of co-occurring CNVs by patient showed that IDH1^{WT} tumor cells in the leading edge feature less co-occurring chromosomal rearrangements than tumor cells in the core (**Fig. 3E**). This finding suggests that clonal tumor cells in the leading edge emerge from specific glioma cell subclones residing in the tumor core and is further supported by lineage tracing studies that show leading edge cells emanate from more primitive cell types in the tumor core (**Figs. 3F-H**). Employing RNA velocity pseudotime analysis[12], we found that these astrocytic tumor cells emanate from stem cell-like tumor cells found in Cluster 23 and transition through an intermediate cell type found mainly in Cluster 3. Notably, these intermediates possess neurodevelopmentally-restricted transcriptional profiles resembling those of embryonic astrocytes and/or RG but retain the frequently co-occurring CNV annotations observed in Cluster 1. In contrast, the progenitor and/or stem cell-like cells found in Cluster 23 are characterized by the addition of more primitive co-occurring cell identities including neural stem cell (NSC) and GSC as well as oligodendrocyte (**Figs. 3I-L)** but are largely marked by a single CNV occurrence in each cell. These observations suggest that astrocyte-like tumor cells with widespread chromosomal anomalies derive from a chromosomally-intact stem cell-like population of cells that have transcriptional profiles resembling oligodendrocytes. These cells transition through an intermediate cell state in which stem cell signatures are lost and pervasive co-occurring CNVs are gained. Synchronously, SCRAM was able to identify analogous populations of glioma cells in our scRNA-seq dataset from our piggyBac transposase-based *in utero* electroporation (pB-IUE) model of glioma, which is yields GFP-labeled *de novo* tumors in immunocompetent mice (**Figs. 3M-P**). These combined analyses elucidate a compendium of individual cell states that parallel neurodevelopmental trajectories and define co-occurring spatiotemporal hierarchies that are elucidated by SCRAM.

SCRAM exposes an immune-like cell population in *IDH1* wildtype and *IDH1* mutant astrocytoma

Our preceding observations uncovered that co-occurring neurodevelopmental cellular states are found within glioma cell transcriptomes, leading us to investigate whether cell signatures deriving from non-neural lineages also exist in these tumors. Unexpectedly, we found 229 cells in Cluster 19 for which tumor cell, antigen presenting cell (APC) and tumor cell with co-occurring APC annotations were detected. These cells were unique to astrocytoma tumors in both IDH^{WT} and $IDH1^{mut}$ patients and featured CNVs, cycling progenitor cell annotation and high expression of the activated APC markers, *CD83*[13] and major histocompatibility class I and II genes[14] (**Figs. 4A-D**). A number of immune cell subtypes are known to actively surveil the tumor microenvironment where they engulf and display tumor antigens through processes of antigen presentation[15]; to confirm that these cells were not an artifact of immune cells phagocytosing tumor cells, we examined expression of the immune cell lineage marker, *PTPRC* (CD45) finding that 69% of APC-like tumor cells did not express *PTPRC*. SCRAM identified an analogous population of *Ptprc* negative APC-like tumor cells in mouse glioma, which featured high *GFP* expression in addition to *Cd83* (**Figs. 4E-H**). These data suggest that a rare population of CNV-altered tumor cells exist in high-grade astrocytoma that are endowed with APC transcriptional signatures and confirm the existence of a parallel cell population in our pB-IUE model.

Modulation of APC-like tumor cell types directs disease progression and CD8+ T cell responses in an immunocompetent *de novo* model of glioma

The former observations uncover a rare population of APC-like tumor cells in human glioma that are similarly detected in our pB-IUE model of murine glioma (**Fig. 5A**). To experimentally discern if APC-like tumor cells can be confirmed *in vivo*, we exploited GFP-labeling of our established pB-IUE model to perform multispectral imaging flow cytometry. Examination of endogenous GFP expression alongside CD83 revealed two distinct populations of CD83+GFP+ cells. The first population consisted of cells featuring dendrite-like projections and punctate expression of GFP that was either present intracellularly or localized adjacent to CD83 on the cell surface (**Fig. 5B**). As per prior reports employing this imaging modality to examine phagocytosis in live cells[16], this expression pattern is indicative of CD83+ immune cells actively phagocytosing or displaying GFP+ tumor antigen on their surface. In contrast, a second population of CD83+GFP+ cells was characterized by high, diffuse intracellular expression of GFP, which is a hallmark feature of targeted cells in the pB-IUE model (**Fig. 5C)**[17]**.** Subsequent analysis revealed a rare number of events wherein CD83-GFP- cells were bound to GFP+ tumor cells at a CD83+ interface (**Fig. 5D**), further implicating CD83 as a mediator of tumor-host cell interactions in glioma. These imaging studies confirm the existence of an endogenous CD83+GFP+ APC-like tumor cell population in the pB-IUE glioma model that was identified by SCRAM in both human and mice.

Insofar as CD83 is required for antigen presentation in non-tumor immune cells, we sought to define how CD83 expression alters tumor progression *in vivo*. To this end, we generated CD83 gain-of-function (GOF) tumors using overexpression of murine *Cd83* and CD83 loss-of-function (LOF) tumors via CRISPR/Cas9 technology. Kaplan-Meier survival analysis revealed that tumor loss of CD83 reduced survival times while tumor overexpression of CD83 extended survival when compared to control tumors (**Fig. 5E**). Correspondingly, end-stage tumors were smaller in CD83-GOF mice, whereas CD83-LOF tumors appeared significantly larger (**Figs. 5F,G**). Cell proliferation assays using Ki67 showed that CD83-GOF end-stage tumors were less proliferative than those of CD83-LOF and control, however, no difference was observed between control and CD83-LOF cohorts (**Figs. 5H,I**). Considering worse survival outcomes in CD83-LOF tumors, we analyzed the number of nuclei per tumor field of view finding that CD83-LOF tumors had nearly 50% more cells than controls (**Fig. 5J**), raising the possibility that impaired immune clearance of tumor cells may account for CD83-LOF phenotypes. These collective observations demonstrate that intratumoral expression of CD83 modulates disease

progression in mouse glioma and suggest that tumor extrinsic mechanisms may contribute to poor survival outcomes when CD83 expression is lost.

Given that CD83 is known to direct T cell responses by enabling antigen presentation by APCs[18], we sought to characterize how changes to tumor cell CD83 expression alters T cell responses in glioma. To this end, we derived glioma cell lines from our pB-IUE tumors and performed coculture experiments with naïve CD8+ T cells (**Figs. 5K**). Following one week of coculture, T cells were harvested for imaging flow cytometry. Analysis revealed that CD83-GOF cocultures had the highest population of CD3+CD8+ cells remaining (**Figs. 5L-O**). Closer examination uncovered a subset of CD3-CD8+ cells from CD83-LOF cocultures, >90% of which expressed the naïve T cell marker CCR7 (**Figs. 5L,M**). Importantly, prior studies have demonstrated that loss of CD3 can reduce T cell expansion[19,20] and is required for acquisition of cytotoxic functions in CD8+ effector memory T cells[21,22]. Additionally, a subset of CD8+CCR7+ T cells have been implicated as potent immunosuppressive mediators[23], suggesting that CD3-CD8+CCR7+ T cells may contribute to subpar T cell responses in CD83-LOF tumors. Consistent with these data, ELISA analyses of cell media from cocultures revealed that while both CD83-GOF and CD83-LOF cocultures had increased production of the T cell activating cytokines IFNγ and TNFα[24,25], cell media taken from CD83-GOF tumor cells alone had increased IL-2 secretion, which is required for memory and effector T cell subsets (**Figs. 5P,Q**). Our scRNA-seq analyses of CD83-altered tumors confirmed that CD83-GOF tumors are enriched for T cell activating cytokines (**Fig. 5R**). Notably, T cells from CD83-LOF tumors displayed increased expression of anergic and regulatory T cell markers, whereas CD83-GOF T cells were marked by high expression of *CD44*, a prominent activation marker of memory and effector T cells[26] (**Fig. 5S**). *In vitro* EdU assays revealed that CD83-GOF cells are intrinsically more proliferative than controls, having similar proliferation rates to CD83-LOF tumor cells (**Figs. 5T,U**). We confirmed this through scRNA-seq cell cycle scoring of CD83-altered tumors, which showed that CD83-GOF and CD83-LOF tumor cells have increased expression of proliferation markers as compared to controls (**Fig. 5V**). When considered alongside results from our *in vivo* proliferation and survival studies, these findings suggest that T cell-CD83 interactions may promote tumor cell clearance and counteract glioma cell proliferation, which may ameliorate poor survival outcomes. These results demonstrate that modulation of CD83 expression in tumors can direct glioma progression through both cell autonomous effects on cell proliferation and cell non-autonomous responses mediated by T cell-dependent changes to the cytokine milieu. Moreover, these studies emphasize the ability of SCRAM to report rare, co-occurring tumor cell populations with important and significant contributions to tumor biology in glioma.

Functional neuronal-like tumor cells are a prominent transcriptional subset of *IDH1* mutant glioma

Parallel analyses using SCRAM revealed that most tumor cells from IDH1^{mut} samples mapped to Cluster 4, which displayed a high density of *IDH1* mutations and featured chromosome 1p19q codeletions in oligodendroglioma patients (**Figs. 4I-K**). Surprisingly, we found that 40% of cells with CNV or GSC annotation had co-occurring cell type annotations of mature neurons, GABAergic neurons and/or synaptic neurons as well as oligodendrocytes and OPCs (**Fig. 4L**); in contrast, only 18 tumor cells had co-occurring neuronal signatures in IDH1^{WT} samples. SCRAM identified an similar population of neuronal-like tumor cells in our mouse glioma dataset (**Figs. 4M,N**), which prompted us to further investigate these cell types.

To validate the existence of neuronal-like tumor cells experimentally, we performed whole-cell patch clamp electrophysiology recordings and RNA patch-sequencing (Patch-seq) on two IDH1^{mut} glioma (Grades II-III) specimens that were surgically resected from the tumor core

(**Figs. 6A,B**). In total, we recorded from 53 cells, 31 of which were used for Patch-seq and 20 for which morphology was preserved using biocytin (**Figs. 6C-H**). Of the 50 cells for which recordings were obtained, 27 had electrophysiological and morphological properties consistent with glia or neurons found in the non-tumor brain[27,28]. Surprisingly, 26 cells displayed electrophysiological properties of neurons but were morphologically consistent with glia. Recorded cells had higher input resistances than neurons or glia (**Fig. 6I**) and had the capacity to fire single APs (**Figs. 6J,K**), signifying that these cells have acquired transcriptional signatures that endow them with functional characteristics of neurons. Prior work has described similar electrophysiological properties in neurogenesis and has implicated hyperexcitability, high input resistance and firing of single APs as a defining feature of young post-mitotic neurons[29]. These newborn neurons feature high levels of GABA receptors[30,31], which are similarly expressed by neuron-like tumor cells and frequently reported by SCRAM as co-occurring GABAergic neuron annotation in IDH1^{mut} tumor cells. Our Patch-seq results confirmed the presence of the *IDH1* mutation in five of the six neuron-like tumor cells sequenced, which confirms these cells are tumor in origin (**Fig. 6L**); transcripts mapping to either mutated *SUZ12* or the *SUZ12P1* pseudogene were also detected in a subset of these cells. Subsequent PCA and cell type enrichment of Patch-seq data using the enrichR tool[32,33] illuminated that neuronal-like tumor cells are transcriptionally akin to glia and are enriched for embryonic astrocyte and OPC gene sets (**Figs. 6M,N**). For comparison, we recorded from an IDH1^{WT} GBM (Grade IV) core sample, finding that only a few cells were capable of being patched, due in part to extensive gliosis and necrosis, but that these cells were electrophysiologically consistent with glia. Whether neuronal-like tumor cells exist in IDH1^{WT} glioma remains unanswered and may be the subject of future scientific investigations.

Using *in vivo* modeling, whole-cell patch clamp recordings were obtained for GFP+ tumor cells from control pB-IUE tumor mice, which revealed an analogous population of neuron-like tumor cells, some of which also featured high input resistances (**Fig. 6O**). To confirm these cells were of tumor origin, we performed complementary experiments employing live *in vivo* 2-photon imaging of mouse tumors to examine Synapsin-driven GCaMP-labeled calcium activity in tumor-bearing mice. Here, the fluorescent ubiquitination-based cell cycle indicator (FUCCI) was used to visualize actively proliferating cells and revealed a small population of FUCCI-labeled cells exhibiting calcium transients consistent with neuronal AP firing (**Figs. 6P**). Taken together, these human and mouse studies confirm the existence of neuronal-like tumor cells endowed with transcriptional and electrophysiological properties similar to but distinct from neurons and characterize these cells as a defining feature of human IDH1^{mut} gliomas.

Discussion

A new computational framework for defining genomic and transcriptional states in cancer, development and disease

By 2040 the number of new cancer cases per year will stand at 29.5 million worldwide, more than half of which will end in cancer-related deaths[34], highlighting a need for molecular insights that can guide therapeutic development. While advances in single-cell sequencing technologies have given scientists the opportunity to examine individual cells in unprecedented detail, the power of these data has yet to be harnessed by computational platforms capable of resolving these profiles on a cell-by-cell basis. To address this unmet need, we developed SCRAM, which we report here as the first computational algorithm of its kind capable of defining integrated co-occurring genomic and transcriptomic features from single cell datasets. In this report, we employ SCRAM using a curated brain tumor-specific cell type marker list to elucidate previously unreported cellular states in glioma and validate these cell types experimentally. However, our

pipeline can be easily modified for use in any cancer type, non-oncologic disease or developmental context. To this end, we believe SCRAM will offer researchers a reliable and potent tool that can be used to uncover molecular complexities in high resolution across a multitude of disease and non-disease states and will enable new lines of scientific inquiry with clinically relevant implications for cancer, development and disease.

A key finding of our studies is the identification of two previously unreported glioma cell types, which we define here as APC-like and neuronal-like tumor cells. Attempts to identify ways in which brain tumors coopt the surrounding microenvironment for continued growth and invasion have led to the development of two highly specialized fields of cancer biology, tumor immunology and cancer neuroscience, each of which seek to characterize tumor-host cell interactions in exquisite detail. Using SCRAM to uncover both rare and widespread co-occurring cell types in human glioma, we characterize two new glioma cell populations with APC-like and neuronal-like functional properties that contribute to the immune and electrophysiological landscapes of these tumors.

APC-like glioma cells alter CD8+ T cell responses and tumor progression

Owing in part to the existence of a highly immunosuppressive tumor microenvironment, immunotherapy trials for the treatment of glioma has proven largely unsuccessful, conferring little to no effect on overall survival[35,36]. Prior research in the field of tumor immunology has sought to characterize the molecular mechanisms employed by glioma cells that mediate this immunosuppressive constituency and has implicated anergic and regulatory T cells as important mediators of these processes[37,38]. To this end, identification of a tumor cell subset with immune-like transcriptomic signatures and functional effects on CD8+ T cell responses offers novel insights into the ways in which glioma cells can alter the immune landscape. Accordingly, CD8+ cytotoxic T cells have been recognized as potent antitumor mediators[39,40], raising the question of whether changes to the proportions of CD8+ T cells in this study alter the ability of these cells to recognize and destroy glioma cells. Indeed, our imaging flow cytometry studies demonstrate that tumor cells are actively surveilled and phagocytosed by the surrounding immune constituency, implicating downstream immunosuppressive phenotypes as a likely culprit for poor tumor clearance that should be further explored through future scientific investigations.

Neuronal-like tumor cells can generate action potentials and are a defining feature of *IDH1* mutant glioma

Over the past decade, the emerging field of cancer neuroscience has elucidated the ways in which glioma cells interact with surrounding neural networks to direct disease progression. Previous work has shown that brain tumors alter peritumoral networks by engendering synaptic dysregulation that promotes hyperexcitability and seizures, which in turn serve to enhance tumor dissemination and growth[41,42]. Prior studies have reported the ability of human glioma cells to receive synaptic inputs from the surrounding neuronal circuitry in the form of excitatory postsynaptic currents[43,44] and that glutamatergic inputs to glioma cells can direct disease outcomes[44]. Our studies build upon these existing reports to reveal that neuronal-like tumor cells may communicate with peritumoral neuronal networks by directing outgoing electrochemical information in the form of APs. Our observations suggest that dysregulated electrophysiological activity in the glioma-bearing brain may originate from tumor cells endowed with the ability to initiate and propagate outgoing neurochemical signals. Importantly, SCRAM predicted the existence of these neuronal-like tumor cells as a defining feature of $IDH1^{mut}$ tumors representing nearly half of tumor cells in our human datasets, which we confirmed experimentally. Notably, $IDH1^{mut}$ glioma frequently presents with seizures[45] despite a slower disease course, raising the possibility that these neuronal-like cell types may contribute to hyperexcitable phenotypes that present clinically. Future investigations may aim to better define

the electrophysiological contributions of these cells to tumor progression and may seek to characterize how modulation of electrophysiological activity originating from tumor cells alters disease outcomes in glioma.

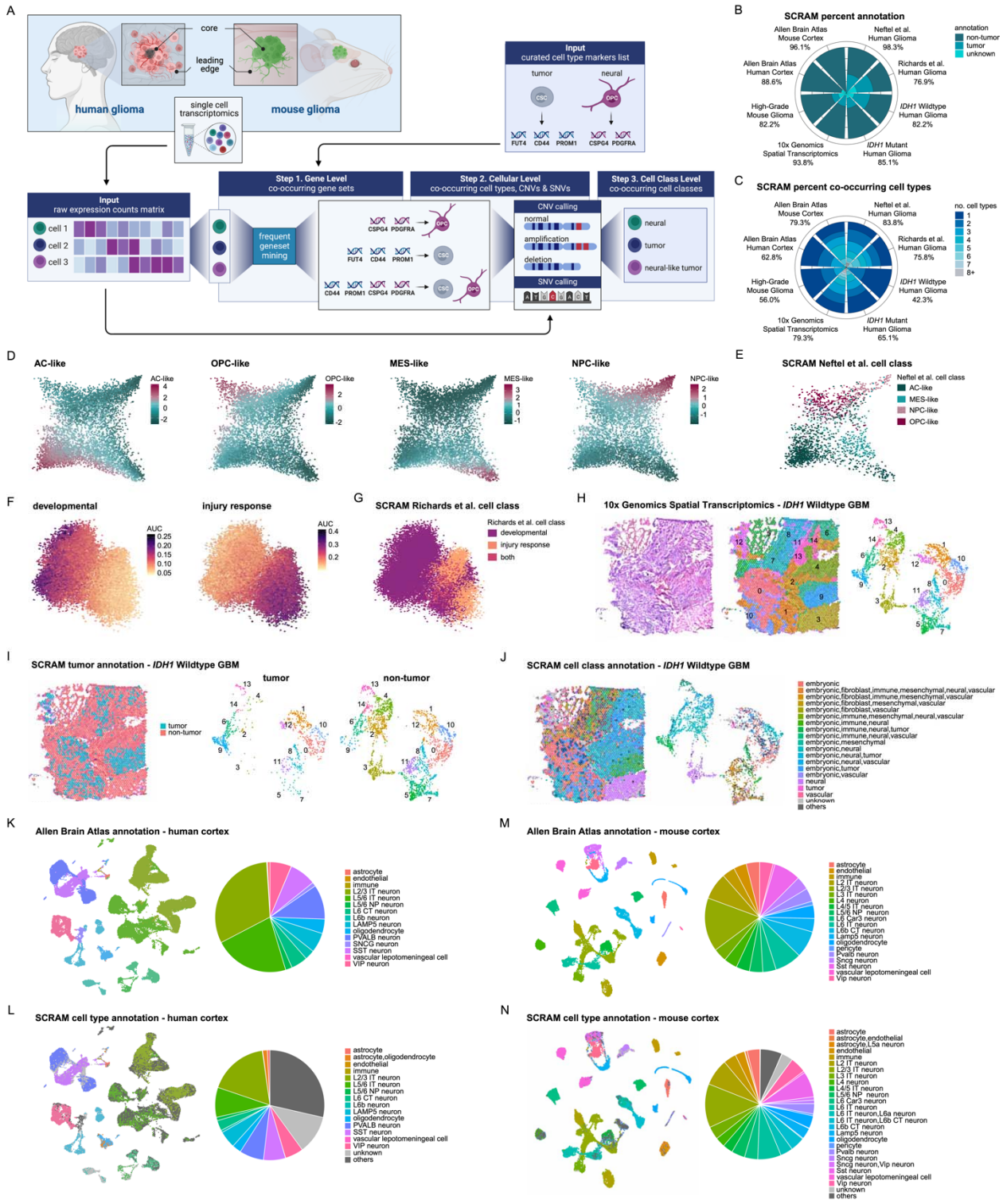

Figure 1. SCRAM reliably assigns cell type annotations to scRNA-seq datasets on a cell-by-cell basis. (A) Schematic representation of the SCRAM pipeline. (B) Percentages of SCRAM-annotated cells assigned non-tumor, tumor or unknown cell types across eight datasets. Percentages shown below dataset names indicate the total percentages of cells

assigned cell type annotation using SCRAM. (C) Percentages of SCRAM-annotated cells by number of co-occurring cell types. Percentages shown below dataset names indicate the percentages of cells having more than one cell type annotation. (D) Reproduction of meta-modules published by Neftel et al.[5] showing astrocyte-like (AC-like), oligodendrocyte precursor cell-like (OPC-like), mesenchymal-like (MES-like) or neural precursor cell-like (NPC-like) cellular states. (E) Meta-module of SCRAM-identified tumor cells from Neftel et al. dataset. (F) Reproduction of cell state modules published by Richards et al.[6] showing developmental and injury response programs. (G) SCRAM-assigned developmental and injury response cell classes for the Richards et al. dataset. (H) H&E and corresponding Seurat clusters are shown for the 10x Genomics Spatial Transcriptomics *IDH1* wildtype GBM dataset. (I) SCRAM-assigned tumor and non-tumor annotation for the 10x Genomics Spatial Transcriptomics *IDH1* wildtype GBM dataset shows tumor cells are embedded throughout non-tumor tissue. Seurat clusters split by tumor and non-tumor annotation shows clusters are comprised of tumor and non-tumor cells. (J) SCRAM cell class annotation for the 10x Genomics Spatial Transcriptomics *IDH1* wildtype GBM dataset. Cells harboring RNA-inferred CNVs are outlined in black. (K) Cell type annotation for human cortex dataset provided by Allen Brain Atlas. (L) SCRAM-assigned cell type annotation for the Allen Brain Atlas human cortex dataset reveals co-occurring cell identities (dark grey) and unknown transcriptional cell states (light grey). (M) Cell type annotation for mouse cortex dataset provided by Allen Brain Atlas. (N) SCRAM-assigned cell type annotation for the Allen Brain Atlas mouse cortex dataset reveals co-occurring cell identities (dark grey) and unknown transcriptional cell states (light grey).

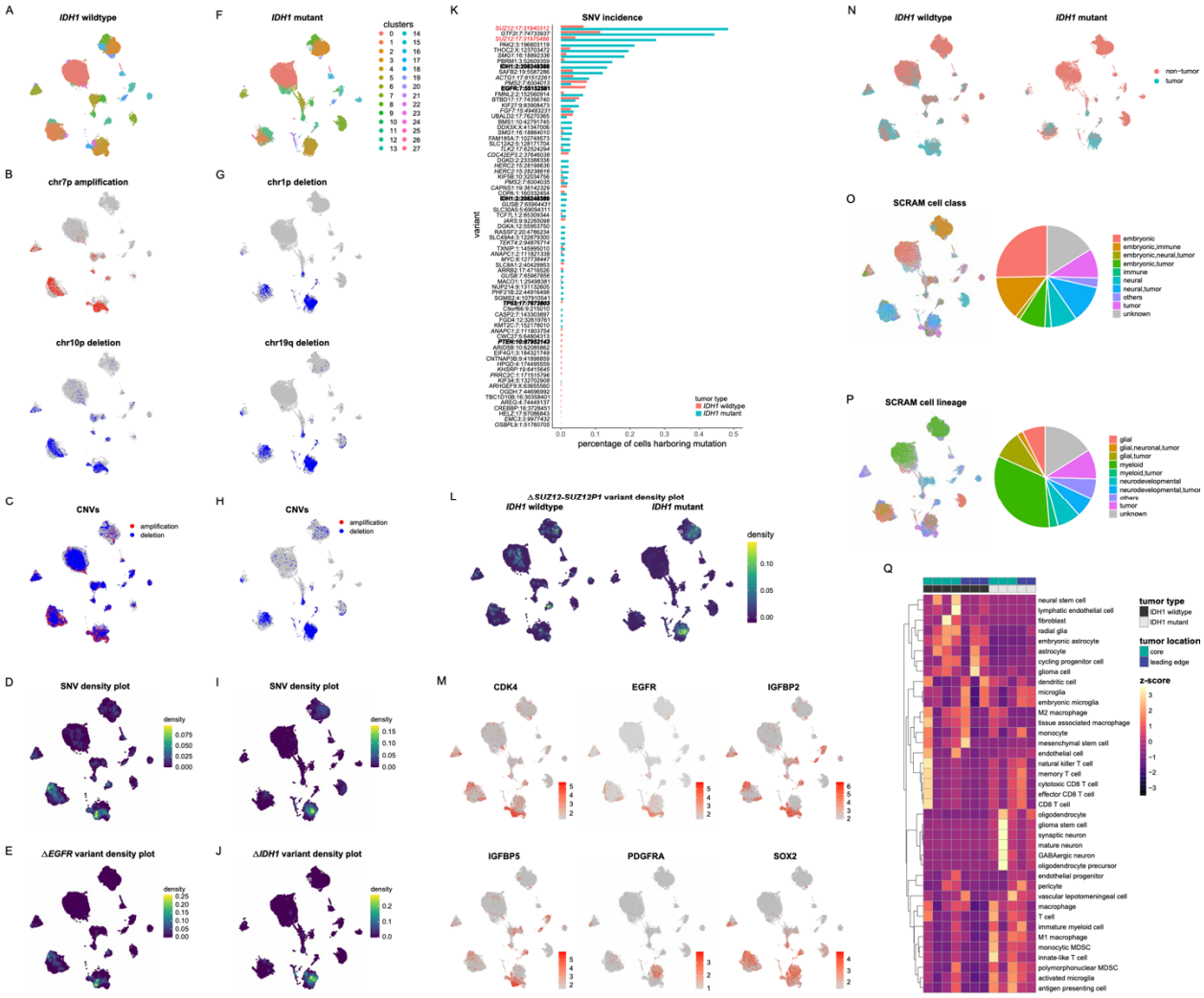

Figure 2. Integrated RNA-inferred genomic and transcriptomic cell states reveal unique molecular signatures in *IDH1* wildtype and *IDH1* mutant glioma. (A) Seurat clusters are shown for 157,316 cells from IDH1^{WT} glioma patients (n=4). (B) Feature plots show SCRAM detects large-scale chromosome 7p amplifications and chromosome 10p deletions in IDH1^{WT} glioma using CaSpER. (C) Feature plots showing total CNV calls by SCRAM for and IDH1^{WT} tumors. (D) Density plot showing SNV allele frequency for rare and COSMIC variants in and IDH1^{WT} glioma. (E) Density plot showing *EGFR* mutations are detected in IDH1^{WT} samples. (F) Seurat clusters are shown for 37,747 cells from IDH1^{mut} glioma patients (n=3). (G) Feature plots show SCRAM detects large-scale chromosome 1p19q codeletions in IDH1^{mut} oligodendroglioma using CaSpER. (H) Feature plots show total CNV calls by SCRAM for and IDH1^{mut} tumors. (I) Density plot showing SNV allele frequency for rare and COSMIC variants in IDH1^{mut} glioma. (J) Density plot showing *IDH1* mutations are detected in IDH1^{mut} samples. (K) Bar plot showing the percentage of cells harboring rare and COSMIC SNVs in IDH1WT and IDH1mut glioma; italics denotes genes with known pseudogenes; bold denotes known driver genes in glioma. (L) Density plot showing either *SUZ12* mutations or the *SUZ12P1* pseudogene are detected in IDH1^{WT} and IDH1^{mut} samples. (M) Feature plots for tumor cell markers used to assign tumor annotation. (N) SCRAM-assigned tumor annotation is shown for IDH1^{WT} and IDH1^{mut} glioma. (O) SCRAM-assigned cell classes are shown for all glioma samples. (P) SCRAM-assigned cell

lineages are shown for all glioma samples. (Q) Heatmap showing enrichment for SCRAM-assigned cell types by tumor type and sample location; z-score for expression of cell types is shown.

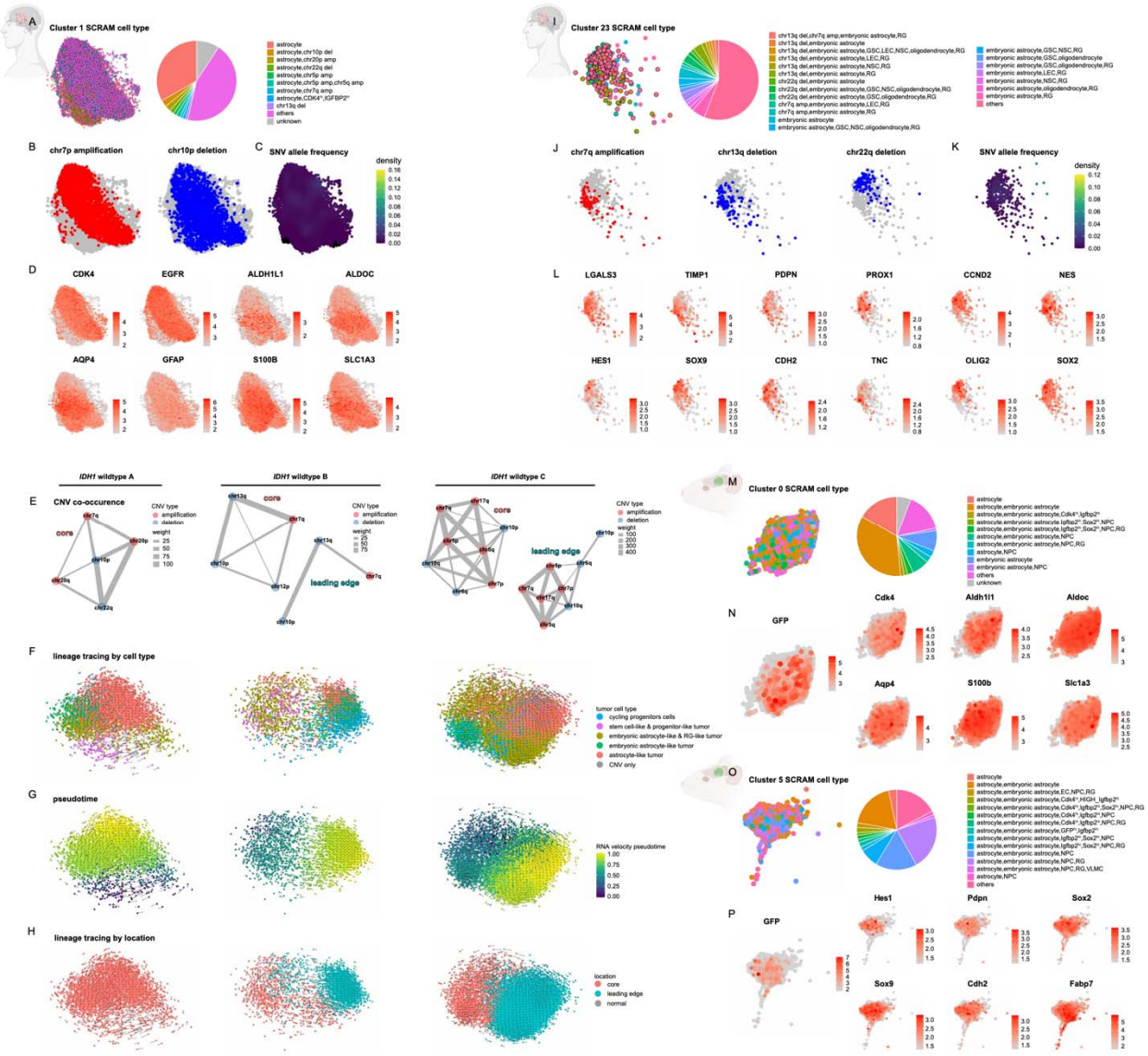

Figure 3. SCRAM uncovers novel tumor cell states and developmental hierarchies in glioma. (A) Dimplot and corresponding pie chart showing SCRAM-identified astrocyte-like tumor cells for Cluster 1 in human glioma samples. CNV-harboring cells are outlined in black. (B) Feature plots showing exemplary large-scale CNVs. (C) Density plot showing SNV allele frequency. (D) Feature plots for select tumor cell (*CDK4*, *EGFR*) and astrocyte (*ALDH1L1*, *ALDOC*, *AQP4*, *GFAP*, *S100B*, *SLC1A3*) markers. (E) Co-occurring CNVs for five patient-matched IDH1^{WT} glioma samples shows leading edge cells emanate from specific co-occurring CNV lineages. Increased line thickness (weight) indicates more frequent co-occurrence. (F) RNA velocity lineage tracing by SCRAM-assigned cell type shows astrocyte-like tumor cells derive from stem-like and progenitor-like tumor cells and transition through an intermediate cell state marked by embryonic astrocyte and RG signatures. (G) RNA velocity pseudotime plots.

(H) RNA velocity lineage tracing by tumor sample location shows leading edge cells derive from glioma cells in the tumor core. (I) Dimplot and corresponding pie chart showing SCRAM-identified stem-like and progenitor-like cell type annotations for Cluster 23 cells in human glioma samples. CNV-harboring cells are outlined in black. (J) Feature plots showing exemplary large-scale CNVs. (K) Density plot showing SNV allele frequency. (L) Feature plots for select embryonic astrocyte (*LGALS3, TIMP1*), GSC (*CCND2, OLIG2, SOX2*), LEC (*PDPN, PROX1*), NSC (*HES1, NES, SOX2, SOX9*), RG *(CDH2, HES1, NES, SOX2, TNC*) and oligodendrocyte (*OLIG2*) markers. (M) Dimplot and corresponding pie chart showing analogous cell types are detected by SCRAM in mouse glioma Cluster 0. (N) Feature plots showing GFP expression and select tumor (*Cdk4*) and astrocyte (*Aldh1l1, Aldoc, Aqp4, S100b, Slc1a3*) markers. (O) Dimplot and corresponding pie chart showing analogous cell types are detected by SCRAM in mouse glioma Cluster 5. (N) Feature plots showing GFP expression and select RG (*Cdh2, Hes1, Fabp7*), NPC (*Fabp7, Sox2, Sox9*) and EC (*Pdpn*) markers. EC: endothelial cell; GSC: glioma stem cell; LEC: lymphatic endothelial cell; NPC: neural precursor cell; NSC: neural stem cell; RG: radial glia.

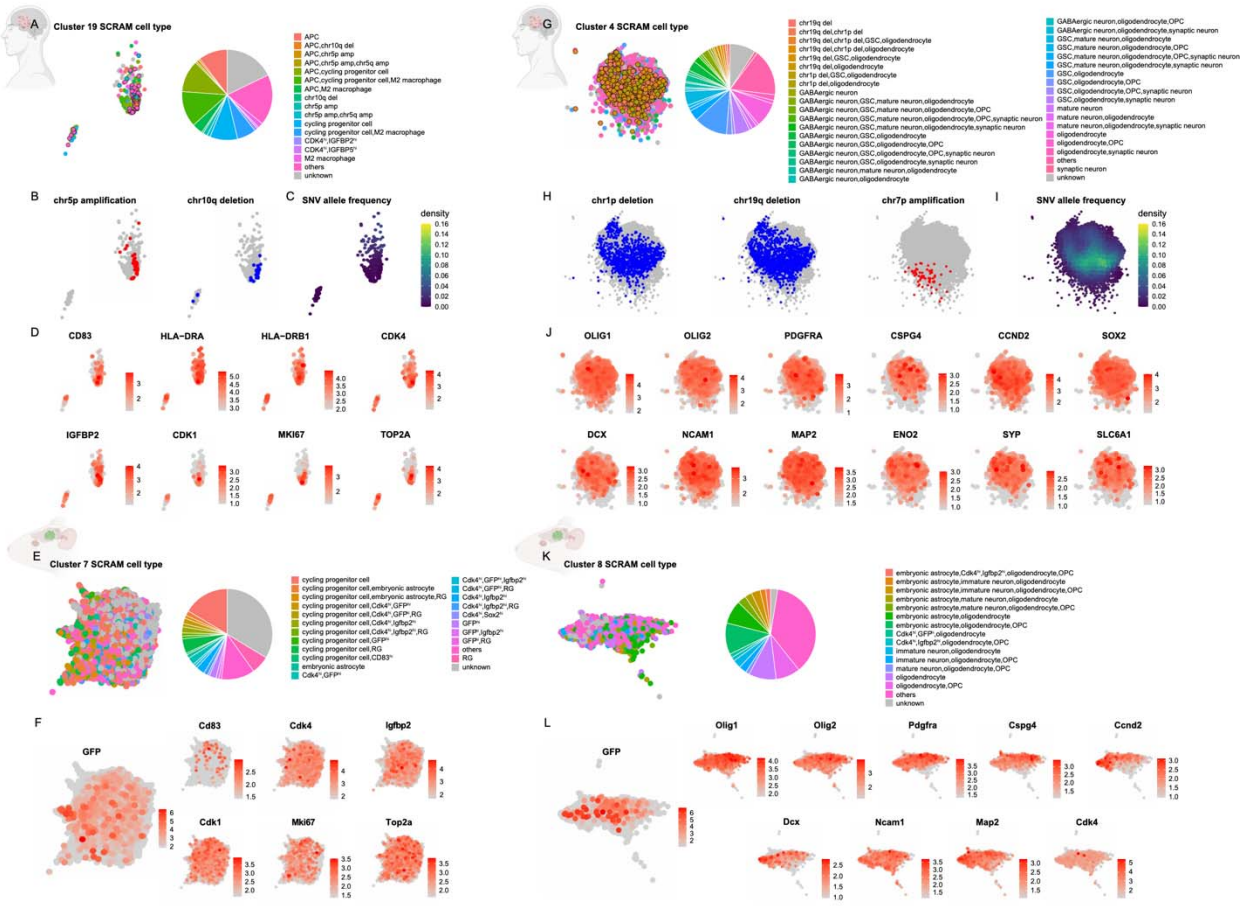

Figure 4. Identification of novel immune-like and neuronal-like tumor cells in glioma using SCRAM. (A) Dimplot and corresponding pie chart showing SCRAM-identified APC-like tumor cells for Cluster 19 in human glioma samples. CNV-harboring cells are outlined in black.

(B) Feature plots showing exemplary large-scale CNVs. (C) Density plot showing SNV allele frequency. (D) Feature plots for select tumor cell (*CDK4, IGFBP2*), cycling progenitor cells (*CDK1, CENPF, MKI67, TOP2A*) and APC (*CD83, HLA-DRA, HLA-DRB1*) markers. (E) Dimplot and corresponding pie chart showing SCRAM-identified APC-like tumor cells for Cluster 7 in mouse glioma samples. (F) Feature plots showing GFP expression and select tumor cell (*Cdk4, Igfbp2*), cycling progenitor cells (*Cdk1, Mki67, Top2a*) and APC (*Cd83*) markers. (G) Dimplot and corresponding pie chart showing SCRAM-identified neuronal-like tumor cells for Cluster 4 in human glioma samples. CNV-harboring cells are outlined in black. (H) Feature plots showing exemplary large-scale CNVs. (I) Density plot showing SNV allele frequency. (J) Feature plots for select GSC (*CCND2, OLIG2, SOX2*), oligodendrocyte (*OLIG1, OLIG2*), OPC (*PDGFRA, CSPG4*), mature neuron (*MAP2, ENO2*), synaptic neuron *(MAP2, ENO2, SYP*) and GABAergic neuron (*DCX, NCAM1, SLC6A1*) markers. (K) Dimplot and corresponding pie chart showing SCRAM-identified neuronal-like tumor cells for Cluster 8 in mouse glioma samples. (L) Feature plots showing GFP expression and select tumor cell (*Cdk4*), GSC (*Ccnd2, Olig2*), oligodendrocyte (*Olig1, Olig2*), OPC (*Pdgfra, Cspg4*), mature neuron (*Map2, Eno2*), immature neuron (*Dcx, Ncam1*) markers. APC: antigen presenting cell; GSC: glioma stem cell; OPC: oligodendrocyte precursor cell.

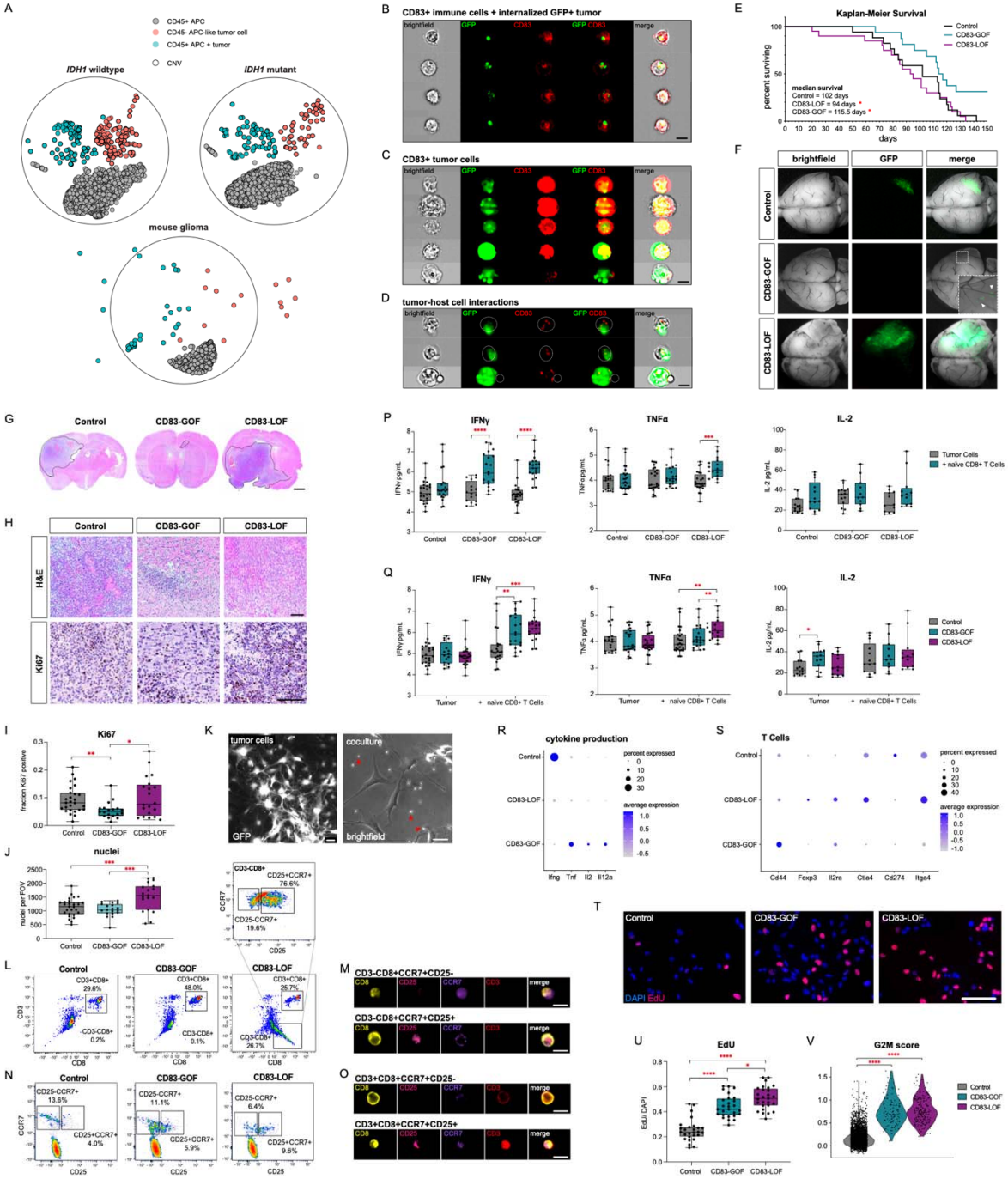

Figure 5. APC-like tumor cells are a potent mediator of disease progression and the immune landscape in glioma. (A). Feature plots of human and mouse scRNA-seq datasets showing APC-like tumor cells. (B) Representative images from imaging flow cytometry of CD83+ APCs with punctate GFP; scale bar = 7 µm. (C) Representative images from imaging flow cytometry of CD83+GFP+ APC-like tumor cells demonstrating diffuse GFP expression throughout the cell; scale bar = 7 µm. (D) Representative images from imaging flow cytometry showing GFP+ tumor cells interfacing with GFP- host cells at a CD83+ interface; scale bar = 7 µm. (E) Kaplan-Meier survival analysis of control (n=17), CD83-GOF (n=16) and CD83-LOF

(n=20) tumor mice; *p<0.05. (F) Exemplary brightfield and GFP images from end-stage control, CD83-GOF and CD83-LOF tumors. (G) Representative H&E whole brain coronal sections from end-stage control, CD83-GOF and CD83-LOF tumors; black dotted lines denote tumor regions, scale bar = 1 mm. (H) Representative H&E and Ki67 images from end-stage control, CD83-GOF and CD83-LOF tumors; scale bar = 100 μm. (I) Quantification of Ki67 for end-stage tumors shows CD83-GOF tumors are less proliferative than controls (**p<0.01) and CD83-LOF (*p<0.05) tumors. (J) Quantification of nuclei per tumor field of view (FOV) shows end-stage CD83-LOF tumors have more cells than controls (**p<0.01) and CD83-GOF (***p<0.001) tumors. (K) Representative images from established tumor cell lines age-matched control, CD83-GOF and CD83-LOF tumors; scale bar = 20 μm. (L) Flow cytometry plots from imaging flow cytometry experiments showing an increased percentage of CD3+CD8+ T cells from CD83-GOF (48.0%) cocultures as compared to controls (29.6%) and CD83-LOF (25.7%). CD3-CD8+ T cells were detected at 26.7% in CD83-LOF T cells compared to 0.2% and 0.1% in controls and CD83-GOF, respectively. Gating on CD3-CD8+ T cells from CD83-LOF cocultures shows two populations of T cells: CD3-CD8+CCR7+CD25- T cells and CD3-CD8+CCR7+CD25+ T cells. (M) Representative images of CD3-CD8+CCR7+CD25- and CD3-CD8+CCR7+CD25+ T cells from imaging flow cytometry experiments from cocultures; scale bar = 7 μm. (N) Flow cytometry plots from imaging flow cytometry experiments showing CCR7 and CD25 expression for CD3+CD8+ gated cells. T cells from CD83-LOF cocultures show an increased percentage of CD3+CD8+CCR7+CD25+ T cells. (O) Representative images of CD3+CD8+CCR7+CD25- and CD3+CD8+CCR7+CD25+ T cells from imaging flow cytometry experiments; scale bar = 7 μm. (P) Box and whisker plots for ELISA assays shows IFNγ production is increased in CD83-GOF (****p<0.0001) and CD83-LOF (****p<0.0001) cocultures when tumor cells cultured with naïve CD8+ T cells. TNFα production was only increased in CD83-LOF (***p<0.001); IL-2 production was unchanged. (Q) Box and whisker plots for ELISA assays shows IFNγ production is increased in CD83-GOF (**p<0.01) and CD83-LOF (****p<0.001) T cell cocultures when compared to controls (***p<0.001). TNFα production was increased in CD83-LOF T cell cocultures when compared to CD83-GOF (**p<0.01) and controls (**p<0.01). IL-2 production was increased in CD83-GOF tumor cells cultured in the absence of T cells when compared to controls (*p<0.05). (R) Dot plots from control, CD83-GOF and CD83-LOF tumor scRNA-seq datasets shows T cells are enriched for *Ifng* (IFNγ) expression in control tumors and enriched for *Tnf* (TNFα), *Il2* (IL-2) and *Il12a* (IL-12) in CD83-GOF tumors, *in vivo*. (S) Dot plots from control, CD83-GOF and CD83-LOF tumor scRNA-seq datasets shows T cells from CD83-GOF tumors are enriched for effector and memory T cell markers (*Cd44, Il2ra*) and T cells from CD83-LOF tumors are enriched for anergic and regulatory T cell markers (*Foxp3, Il2ra, Ctla4, Itga4*). (T) Representative images from EdU proliferation assays on control, CD83-GOF and CD83-LOF tumor cell lines *in vitro*. (U) Quantification of fraction of EdU+ cells shows CD83-GOF (****p<0.0001) and CD83-LOF (****p<0.0001) tumor cells are more proliferative than controls and that CD83-LOF tumor cells are more proliferative than CD83-GOF tumor cells (*p<0.05). (V) Cell cycle scoring analysis for control, CD83-GOF and CD83-LOF proliferating tumor cells from scRNA-seq datasets shows CD83-GOF (****p<0.0001) and CD83-LOF (****p<0.0001) tumor cells are more proliferative than controls.

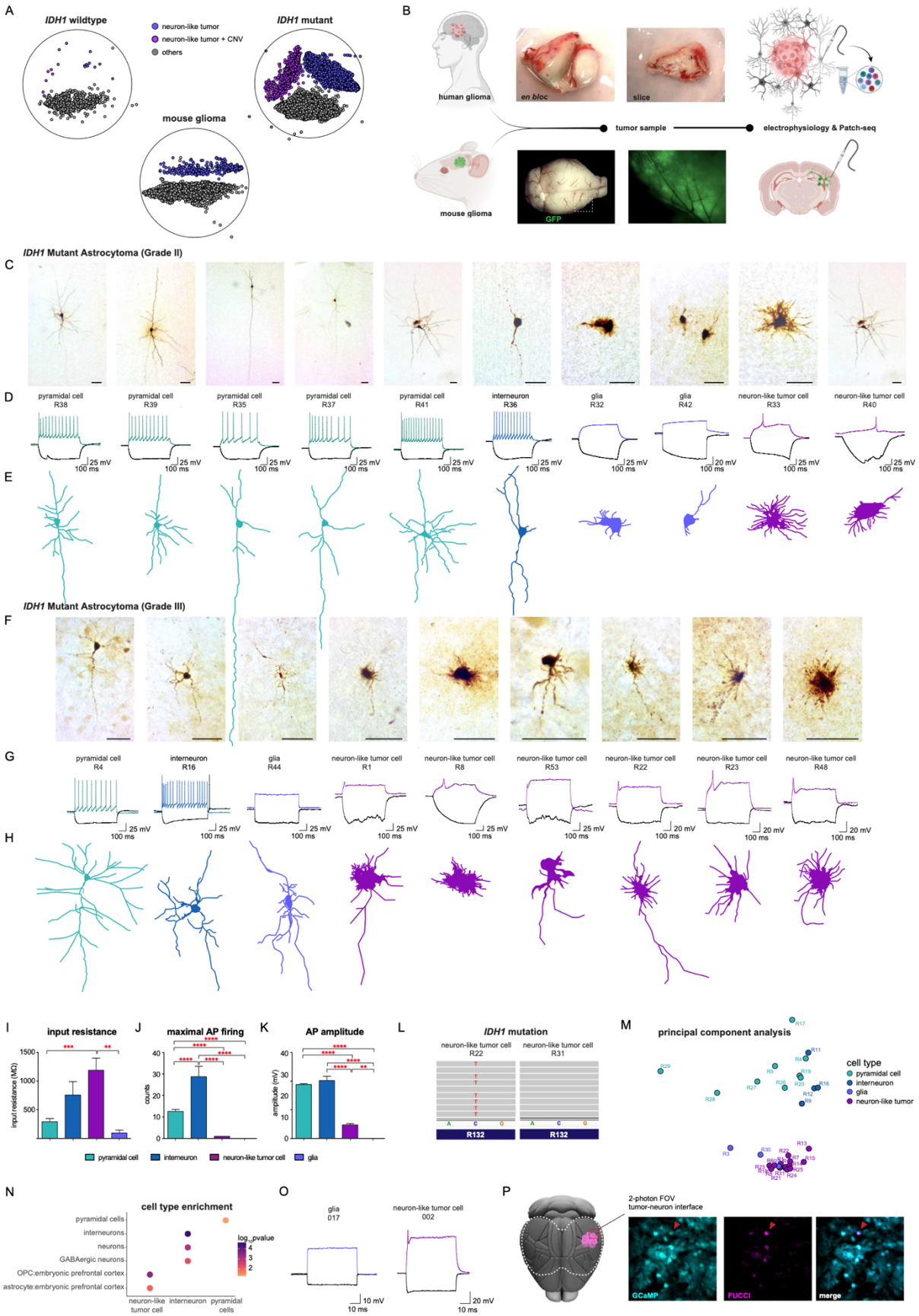
A
IDH1 wildtype
neuron-like tumor
neuron-like tumor + CNV
others
IDH1 mutant
mouse glioma
B
human glioma
en bloc
slice
tumor sample
electrophysiology & Patch-seq
mouse glioma
GFP
IDH1 Mutant Astrocytoma (Grade II)
C
D
pyramidal cell R38
pyramidal cell R39
pyramidal cell R35
pyramidal cell R37
pyramidal cell R41
interneuron R36
glia R32
glia R42
neuron-like tumor cell R33
neuron-like tumor cell R40
E
IDH1 Mutant Astrocytoma (Grade III)
F
G
pyramidal cell R4
interneuron R16
glia R44
neuron-like tumor cell R1
neuron-like tumor cell R8
neuron-like tumor cell R53
neuron-like tumor cell R22
neuron-like tumor cell R23
neuron-like tumor cell R48
H
I
input resistance
input resistance (MΩ)
J
maximal AP firing
counts
K
AP amplitude
amplitude (mV)
pyramidal cell
interneuron
neuron-like tumor cell
glia
L
IDH1 mutation
neuron-like tumor cell R22
neuron-like tumor cell R31
R132
M
principal component analysis
cell type
pyramidal cell
interneuron
glia
neuron-like tumor
N
cell type enrichment
pyramidal cells
interneurons
neurons
GABAergic neurons
OPC:embryonic prefrontal cortex
astrocyte:embryonic prefrontal cortex
neuron-like tumor cell
interneuron
pyramidal cells
O
glia 017
neuron-like tumor cell 002
P
2-photon FOV
tumor-neuron interface
GCaMP
FUCCI
merge

Figure 6. Neuronal-like tumor cells are functionally akin to neurons and are a hallmark of *IDH1* mutant glioma. (A). Feature plots of human and mouse scRNA-seq datasets showing neuronal-like tumor cells. (B) Experimental workflow for electrophysiology and Patch-seq assays. (C) Images of biocytin-reconstructed cell morphologies for patched cells from an $IDH1^{mut}$ diffuse astrocytoma (Grade II); scale bar = 50 µm. (D) Matched whole-cell electrophysiology traces for patched cells. Two voltage traces are shown: the hyperpolarization trace obtained with the smallest current stimulation (black) and the depolarization trace showing maximal AP firing rate. Stimulation length = 600 ms. (E) Traced cell morphologies are shown for recorded cells. (F) Images of biocytin-reconstructed cell morphologies for patched cells from an $IDH1^{mut}$ diffuse astrocytoma (Grade III); scale bar = 50 µm. (G) Matched whole-cell electrophysiology traces for patched cells. Two voltage traces are shown: the hyperpolarization trace obtained with the smallest current stimulation (black) and the depolarization trace showing maximal AP firing rate. Stimulation length = 600 ms. (H) Traced cell morphologies are shown for recorded cells. (I) Bar graph showing neuron-like tumor cells have the highest input resistance amongst recorded cells; *p*-values for pairwise comparisons are noted in the figure. (J) Bar graph showing neuron-like tumor cells have reduced maximal firing rates compared to neurons; *p*-values for pairwise comparisons are noted in the figure. (K) Bar graph showing neuron-like tumor cells have reduced AP amplitude compared to neurons; *p*-values for pairwise comparisons are noted in the figure. (L) Examples of Patch-seq results showing the *IDH1R132H* mutation is detected in 5/6 sequenced neuron-like tumor cells and is absent in 1/6 cells. (M) PCA plot of 31 Patch-seq cells shows neuron-like tumor cells are transcriptionally similar to each other and glia. (N) EnrichR analysis of 31 Patch-seq cells confirms neuron-like tumor cells are enriched for gene sets corresponding to OPCs and embryonic astrocytes; $p<0.05$. (O) Example traces of GFP+ tumor cells shows neuron-like tumor cells are detected in mouse glioma. (P) 2-photon images reveal Synapsin-driven GCaMP-labeled calcium activity is detected in select FUCCI-labeled tumor cells in mouse glioma.

Acknowledgements

This work was supported by grants from the NIH (R01NS071153 to B.D., R01NS124093 to B.D. and G.R., R01CA223388 to B.D, R01NS094615 to G.R.). R.N.C. is supported by grants from the NIH (5T32HL92332-15, F31CA265156, and F99CA274700). The Baylor College of Medicine Single Cell Genomics Core is supported by NIH Shared Instrument Grants (S10OD023469, S10OD025240), P30EY002520 and CPRIT grant RP200504. Schematics were created using Biorender.com. The Cytometry and Cell Sorting Core at Baylor College of Medicine is supported by funding from the CPRIT Core Facility Support Award (CPRIT-RP180672), the NIH (CA125123 and RR024574) and the assistance of Joel M. Sederstrom." If your publication used data generated with the BioSorter, please also acknowledge NIH grant S10 OD025251.

Data Availability

The RNA-Seq and scRNA-seq datasets generated during this study will be made available through the NCBI Gene Expression Omnibus (GEO) website. All other study data are included in the article and/or supporting information.

Code Availability

Source code of SCRAM will be available after publication of the manuscript.

Author Contributions

R.N.C. and A.S.H. are responsible for conception of this project, study and pipeline design and interpretation of results. R.N.C. performed all pB-IUE surgeries, *in vitro* and FACS experiments, immunostaining, statistical analyses and prepared all scRNA-seq samples from human and mouse with assistance from M.F.M. and A.S.H. wrote the code for the SCRAM pipeline. Q.M. and J.J. performed all human electrophysiology and Patch-seq experiments. J.M. performed 2-photon and widefield imaging studies in mice and subsequent analyses. I.A. performed all mouse electrophysiology and respective analyses. R.N.C., A.O.H. and A.S.H. performed bioinformatics analyses on human and mouse scRNA-seq datasets. A.C. maintained all mouse colonies and assisted with tissue harvesting and fixation. B.L., Z.L., D.C, E.L.F., Y.K., P.H., and A.R. provided experimental support for *in vitro* cell culture, IHC, IF, ICC and surveyor experiments. C.M. performed histopathological diagnoses of mouse tumors. A.J. and G.R. identified and consented patients for the study. R.N.C. and A.S.H prepared the manuscript. J.N., X.J., B.D. and G.R. contributed to the manuscript with feedback from all authors.

Declaration of Interests

The authors declare no competing interests.

Contact for reagent and resource sharing

Dr. Akdes Serin Harmanci (akdes.serinharmanci@bcm.edu) is the lead contact for reagent and resource sharing. All published reagents will be shared on an unrestricted basis; reagent requests should be directed to the corresponding author.

Experimental model and subject details

All experimental animals were treated in compliance with the US Department of Health and Human Services, and Baylor College of Medicine IACUC guidelines. All mice were housed with food and water available *ad libitum* in a 12-hour light/dark environment. Both male and female mice were used for experiments and mice were randomly allocated to experimental groups. All scRNA-seq studies were performed on mice of the same gender. For *ex vivo* and *in vivo* experiments, adult mice aged >3 months were used unless otherwise described. Adult patients at St. Luke's Medical Center and Ben Taub General Hospital provided preoperative informed consent to participate in these studies and were consented under Institutional Review Board Protocol H35355. Patients were males and females. Clinical characteristics were summarized and maintained in a deidentified patient database (**Table S1**).

Experimental Methods

SCRAM pipeline and methodology

SCRAM input consists of aligned scRNA-seq reads, our manually curated host cell type marker database named *BrainTumorHostCellDB* and scRNA-seq clustering with cluster markers (i.e. DEGs). Tumor and host cells are annotated independently for two reasons: (1) there is significant overlap in expression markers in tumor and non-tumor cells. For example, *EGFR* and *PDGFRA* genes are often used to denote tumor cells[5], however, *PDGFRA* and *EGFR* are also DEG markers of OPC and ependymal host cells, respectively. (2) We hypothesize that a separation of tumor and host-specific features can provide more robust results. This is based on our observation that cell-type assignment methods that jointly classify tumor and host cells fail at assigning cell types due to overlaps between markers. Therefore, the SCRAM pipeline employs orthogonal tumor features to identify tumor cells in single cell resolution: (1) marker-expression modeling; (2) RNA-inferred genotyping of large-scale chromosome alterations using a modified version of our CNV calling algorithm, CaSpER[7]; and (3) RNA-inferred mutational profiling of recently developed XCVATR[8] tool to deduce rare deleterious SNV (COSMIC[46]-

reported and dbSNP[47] <0.1% frequency). The following sections explain how these three features are employed systematically in our pipeline:

Step 1. Annotation of the host cells. We annotate each single cell within each Seurat cluster with host cell types using our manually curated host cell type marker database, *BrainTumorHostCellDB*. We only consider the genes in BrainTumorHostCellDB that are in our Seurat cluster markers. Seurat cluster markers are calculated using FindAllMarkers function with logfc.threshold = 0.5 for human and logfc.threshold = 0.25 for mouse. Our *BrainTumorHostCellDB* consists of a list of 87 host cell types and their respective markers for gliomas that represent a diverse range of cell lineages including myeloid, lymphoid, neuronal, embryonic, endothelial, fibroblast, tumor, hematopoietic stem cell, neurodevelopmental, lymphatic, glial, mesenchymal, and neurovascular. Most cell types require a minimum of 2 markers to qualify for annotation. Cell types are stratified into different hierarchies including cell lineage (e.g. neuronal, glial, lymphoid, myeloid) and cell class (e.g. neural, immune) levels. To annotate each cell using marker list, we analyzed the enrichment of gene marker sets as "rules" on cells. For example, to be called "CD4+ regulator T cell", a cell must express a reliable fraction of the following markers concurrently, {*CD3E, CD3D, CD3G, CD4, IL2RA, FOXP3, TIGIT, etc.*}. Results are tabulated in brute force annotation where each cell is evaluated in terms of a simple rule-based assignment by cell type.

Step 2. Annotation of tumor cells. Because tumor cells exhibit a wide range of transcriptional states, we employ redundant and stringent approaches to annotate tumor cells using 3 modular components: (1) marker-expression modeling, (2) genotyping of CNVs on all cells (3) RNA-inferred mutational profiling of known glioma mutations (i.e. *IDH1*, *EGFR*).

a. Module 1. Marker expression-based tumor classification model. Given the expressional heterogeneity of tumor markers in host cells, we used previously published datasets of tumor and non-tumor cells to establish a marker expression-based tumor classification model (i.e. thresholding requirements for "high expression" annotation) for the tumor markers *PDGFRA, EGFR, CDK4, IGFBP2, IGFBP5* and *SOX2*. For each tumor marker gene, an independent classifier model is built using: (1) Allen Brain mouse and human scRNA-seq data, which is the largest compendium of healthy brain data, as a training set for host cells; and (2) a compendium of publicly available brain-tumor scRNA-seq datasets as a training set for tumor cells[5]. Finally, the following statistical models are used to infer the class (host vs. tumor) of our in-house tumor scRNA-seq data:

We model the expression as a mixture of Gaussian distributions for identification and classification of host and tumor cells:

Let $X_j = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$ be the training expression vector of normal and tumor cells for gene j, where x_i is the expression value at cell i. The distribution of every expression value is specified by a probability density function through a finite mixture model of *G*=2 classes (host vs tumor):

$$f(x_i; z) = \sum_{k=1}^{G} \pi_k f_k(x_i;\ \theta_k)$$

where $z = \{\pi_{1,\ldots,}\pi_G, \theta_1, \ldots \theta_G\}$ is the parameters of the mixture model and $f_k(x_i;\ \theta_k)$ is the kth component density, which assumes to follow Gaussian distribution $f_k(x_i;\ \theta_k) \sim N(\mu_k, \sigma_k)$. $\{\pi_{1,\ldots,}\pi_G\}$ is the vector of probabilities, non-negative values which sum to 1, known as the mixing proportions. Mixing proportions, π, follows a multinomial distribution.

We used the above model to predict host vs. tumor class in our inhouse glioma cells. For each gene j, z parameters are estimated by maximizing log-likelihood function via the EM algorithm. The log-likelihood function is formulated as:

$$l(z; x) = \sum_{i=1}^{n} \log f_k(x_i;\ z)$$

For each tumor marker gene, we generate a matrix with genes in the rows and cells in the columns and cell value indices of 1 if that cell has high "tumor class" probability for the corresponding gene. A cell is marked as "tumor" if there are at least two marker genes with high "tumor class" probability. We used mclust R package for Gaussian mixture model (GMM) implementation[48].

b. Module 2. Identifying CNV events from scRNA-seq data. CNVs are a hallmark feature of tumor cells that can be used to classify tumor vs. non-tumor cells alongside or in the absence of expression markers. However, detection of CNVs from scRNA-seq data is inherently noisy due to a multitude of factors, including drop-outs and unmatched control sets and requires a set of cells that are known to be tumor cells. To estimate a "clean" set of CNV calls that can provide reliable CNV-based tumor scores, we used a *pure tumor pseudobulk* sample.

Estimation of CNV profiles using patient-specific pure tumor pseudobulk samples. We first use our expression-based marker model from **Module 1** to identify tumor cells. The collection of cells that are assigned as "tumor" using **Module 1** is treated as a pure tumor cell cohort.

CNV calling on patient-specific pure pseudobulk samples. We hypothesize that the pseudobulk sample contains representative sets of CNVs with high probability and therefore should be useful to identify a clean CNV call-set. The CNV calling on the pseudobulk samples is performed using our CNV calling algorithm, CaSpER, for each patient. CaSpER CNV calls are used as the ground truth large-scale CNV calls for each patient.

Genotyping of CNVs on all cells. After CNVs are identified from the pseudobulk sample, we genotype the set of CNVs on all cells and generate a binary matrix that represents the existence of CNVs on the cells, i.e., $CNV_{i,j}$.

c. Module 3. We perform RNA-inferred rare deleterious (COSMIC[46]-reported and dbSNP[47], <0.1% frequency) mutational profiling via our recently developed XCVATR[8] tool. We detected mutations in *IDH1*, *EGFR*, *TP53* and *PTEN* and annotated cells with those mutation harboring cells as tumor.

Visualization of tumor cells. We calculate the accumulation of expressed rare deleterious COSMIC reported variants in our clusters to visualize the tumor cells.

Step 3. Summarizing co-occurring cell types using maximum frequent gene set identification.

We summarized co-occurring cell types using a frequent itemset rule mining approach. CNV and SNV calls are added to provide an integrated transcriptomic and genomic summary for each cell. An example SCRAM output for a single cell is given as "glioma stem cell, mature neuron, synaptic neuron, oligodendrocyte precursor cell, chr1p_deletion, chr19q_deletion + *IDH1:2:208248389* mutation". We use the tumor and host cell assignments of **Step 1** and **Step 2** to integrate co-occurring tumor and host cell features.

The simplest method for detecting maximally frequent tumor and host feature sets is a brute force approach in which each possible subset of features is a candidate frequent set. The *apriori* algorithm is an efficient implementation for finding maximally frequent sets with support above a

given threshold. In *apriori* algorithm minimum *support* threshold is set to min(50, number_cells_in cluster*0.1) and maximum number of genes in a geneset is set to 50. Using the *apriori* algorithm we identified co-occurring gene sets expressed concurrently within each cell and provided annotation of high-resolution cellular identities using a three-step co-occurrence analysis. We performed our co-occurrence analysis in multiple levels: 1) gene level (an example output of this step: {*NES, SOX2, SLC1A3, CDH2*} is *radial glial cells*), 2) cell type level (an example output of this step: *tumor* AND *radial glial* AND *mature astrocyte*) 3) cell lineage level (an example output of this step: *neurodevelopmental* AND *tumor* AND *glial* is commonly upregulated) 4) cell class level (an example output of this step: *tumor* AND *neural cells* are commonly upregulated).

In maximum frequent gene set co-occurrence analysis:

Within each cluster *m* and cell type *t* we calculate the maximum frequent gene sets using *apriori* algorithm. Input is the binarized matrix E^{mt} where the cell type marker genes are on the row and the cells in cluster *m* are on the columns.

$$e_{ij}{}^{mt} = \begin{cases} 1, & if\ \ cell\ type\ \ marker\ gene\ i\ is\ a\ cluster\ marker\ AND\ expressed\ in\ cell\ j\ in\ cluster\ m \\ 0, & otherwise \end{cases}$$

In maximum frequent cell type (or cell lineage or cell class) co-occurrence analysis:

Within each cluster *m,* we calculate the maximum frequent cell types (or cell lineage or cell class) using *apriori* algorithm. Input is the binarized matrix E^{m} where the cell types (or cell lineage or cell class) are on the row and the cells in cluster *m* are on the columns.

$$e_{ij}{}^{m} = \begin{cases} 1, & if\ \ cell\ type(or\ cell\ lineage\ or\ cell\ class)\ i\ is\ annotated\ in\ cell\ j\ in\ cluster\ m \\ 0, & \text{otherwise} \end{cases}$$

<u>Gene set significance estimation:</u>

For a given frequent gene set G_i on n cells with k genes, we compute the probability of selecting more than n cells that can have the observed gene set with higher or same pattern:

$$P_{\{extreme\}}(G_i = \{g_1, g_2, \dots, g_k\} \mid H_0) = \sum_{n'>n} \binom{N}{n'} \underbrace{\left(\prod_{n'} P(G_i|H_0)\right)}_{\substack{n'\,cell\ with \\ pattern}} \times \underbrace{\left(\prod_{N-n'} P(\tilde{G}_i|H_0)\right)}_{\substack{N-n'\,cells\ w/o \\ pattern}}$$

$$P(G_i|H_0) = P(g_1, g_2, \dots, g_k \mid H_0) = \prod_{1\le i\le k} P(g_i > 0 \mid H_0)$$

<u>*Single cell processing.*</u> We run samples on the 10x Chromium platform to produce next-generation sequencing (NGS) libraries. We performed standard procedures for filtering, mitochondrial gene removal, variable gene selection using Seurat pipeline. Criteria for cell/gene inclusion is: include genes that are present in >3 of cells, include cells that express >300 genes, the number of genes detected in each cell should be > 200 and < 5000 and mitochondria ratio <10 (<20 in mouse). We integrated cells coming from different patients using Harmony algorithm[49]. Next, we visualized clusters using uniform manifold approximation and projection (UMAP) constructed from the Harmony corrected PCA. We also performed lineage tracing,

trajectory analysis, and RNA velocity assessments to create developmental hierarchies and lineage histories of glioma cells using scvelo R package[12].

Patch-seq data processing The Patch-seq reads are mapped using STAR[50] to hg38 assemblies for human. Read count matrices are generated using FeatureCounts[51] with the latest gene annotations from GENCODE[52] consortia. Differentially expressed genes and transcripts identified using DESeq2[53] and limma[54]. Cells are clustered and visualized using PCA methods. Cell type enrichment analysis are performed with enrichR[33] using the *PanglaoDB_Augmented_2021*, *CellMarker_Augmented_2021* cell type marker sets. *IDH* mutations are identified using our variant detection tool XCVATR[8] and visually confirmed using Integrative Genomics Viewer (IGV)[55].

Human data

Tumor samples were collected during surgery and immediately placed on ice. Tissue was divided for use in subsequent transcriptomic, histopathological, proteomic and/or biochemical studies. Patient samples were collected separately for pathology and molecular subtyping. Histopathology and molecular subtyping of *IDH1* and 1p19q deletion status were confirmed by board-certified pathologists. Samples for scRNA-seq and immunoprecipitation assays were fixed in LN_2 and kept at -80°C.

piggyBac *in utero* electroporation model

Tumor mice were generated according to previously published protocols[56]. Briefly, *in utero* electroporation and single-sided intraventricular injection of *Pten, Nf1* and *Trp53* CRISPR/Cas9 pX330 constructs targeting *Glast*-expressing mouse neural precursor cells via piggyBac transposase technology were performed on CD1 wildtype damns at E16.5. sgRNA guides employed for these experiments are as follows:

Δ*Pten* sgRNA: GAGATCGTTAGCAGAAACAAAAGG

Δ*Nfl* sgRNA: GCAGATGAGCCGCCACATCGAGGG

Δ*Trp53* sgRNA: CCTCGAGCTCCCTCTGAGCCAGG

Δ*Cd83-1* sgRNA: CTGCAGCCTGGCACCCGCGA

Δ*Cd83-2* sgRNA: CTTGGCCCAGGACACTGCAT

CD83-GOF tumor mice were generated using piggyBac constructs driving overexpression of mouse *Cd83*. All mice received co-electroporation of piggyBac-GFP constructs. Tumor brains were collected from mice either at matched time points or end-stage disease. Mice were monitored for symptoms indicative of tumor burden, including lethargy, hunched posture, decreased appetite, poor grooming maintenance, squinting of the eyes, partial limb paralysis and/or ataxia, denoting the IACUC permitted endpoint.

Single cell RNA sequencing

Human and GFP+ mouse tumors were prepared as single-cell suspensions. Briefly, samples were coarsely chopped with surgical scissors and enzymatically digested with Papain supplemented with DNase I (Worthington Biochemical Corporation, LK003150). Samples were incubated for 15 minutes at 37°C on a thermocycler kept at 1400 *g* and briefly pipetted every 5 minutes. Cells were pelleted at 13,000 *g* for 10 seconds and resuspended in PBS before processing for debris and dead cell removal. Cell suspensions were processed using the MACS Debris Removal Kit (Miltenyl, 130-109-398) and MACS Dead Cell Removal Kit (Miltenyl, 130-090-101), per the manufacturer's instructions. Live cells were collected through negative

selection using a MS Column in the magnetic field of a MiniMACS Separator (Miltenyl, 130-042-102). Eluted cells were spun at 300 *g* for 5 minutes and resuspended in Gibco Dulbecco's Modified Eagle Medium with GlutaMAX (DMEM; Thermofisher, 10566016) supplemented with 10% fetal bovine serum serum (FBS; Thermofisher, 16000044). Single cells were processed with the 10X Chromium 3′ Single Cell Platform using the Chromium Single Cell 3′ Library, Gel Bead and Chip Kits (10X Genomics) following the manufacturer's protocol. Briefly, approximately 5,000-15,000 cells were added to each channel of a chip to be partitioned into Gel Beads in Emulsion (GEMs) in the Chromium instrument, followed by cell lysis and barcoded reverse transcription of RNA in the droplets. GEMs were then broken and cDNA from each single cell is pooled together. Cleanup was performed using Dynabeads MyOne Silane Beads (Thermofisher, 37002D). Subsequent cDNA was then amplified and fragmented to optimal size before end-repair, A-tailing, and adaptor ligation. Libraries were run individually using a NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (Illumina, 20024907) and sequenced on an Illumina NextSeq550.

Histology

Mice were humanely euthanized and brain tissue was harvested for subsequent processing. Mouse brain samples were fixed through intracardial perfusion of 4% paraformaldehyde in PBS and kept in solution for 12 hours at 4°C before being transferred to 70% EtOH. Human samples were drop fixed in 4% paraformaldehyde in PBS for 12 hours at 4°C before being transferred to 70% EtOH. Paraffin embedding was performed by the Breast Cancer Pathology Core at Baylor College of Medicine.

Hematoxylin and eosin (H&E) staining was performed on 10 µm paraffin-embedded sections cut on a microtome. Slides were deparaffinized and processed using Harris hematoxylin (Poly Scientific R&D, S212A) and eosin (Poly Scientific R&D, S176) for 1 minute and 30 seconds, respectively. Slides were mounted with Permount Mounting Media (Electron Microscope Sciences, 17986-01) before cover slips were placed. Histological diagnoses of mouse tumors were validated across n≥3 tumors per experimental group by a board-certified neuropathologist.

For immunohistology, 10 µm paraffin-embedded sections were cut, deparaffinized and subject to heat-induced epitope retrieval (HIER) using antigen retrieval buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6.0). Sections were blocked for 1 hour at room temperature and kept in primary antibodies overnight at 4°C. The following primary antibodies were used in this report: goat anti-CD83 (1:200; Novus Biologicals, AF1437). Species-specific secondary antibodies tagged with Alexa Fluor corresponding to 568 nm emission spectra (1:1,000, Thermofisher) was used for immunofluorescence. Hoechst nuclear counter staining (1:50,000; Thermofisher, H3570), was performed before coverslips were mounted using Vectashield antifade mounting medium (Vector Laboratories, H-1000).

In vivo cell proliferation assay

Tumor proliferation was assessed through quantitative immunohistochemistry using rabbit anti-KI67 (1:200; Abcam, ab16667). In total, 27 brightfield images were acquired per experimental group (n=3) (n=3 images per coronal section × n=3 sections per animal × n=3 animals per experimental group). Quantification was performed using the Analyze Particles plugin in ImageJ.

Mouse tumor cell lines

Tumor cell lines from control, CD83-GOF and CD83-LOF tumors were established from P65 pB-IUE tumor mice. Mice were humanely euthanized and brains were dissected. GFP-guided

microdissection of tumor tissue was done under a dissection microscope. Tissue was processed into single-cell suspensions according to our scRNA-seq protocol above and seeded into T75 flasks with Gibco Dulbecco's Modified Eagle Medium with GlutaMAX (DMEM; Thermofisher, 10566016) supplemented with 10% fetal bovine serum serum (FBS; Thermofisher, 16000044). Cells were grown and passaged after reaching 85% confluency for two weeks before being used for subsequent assays. Tumor cell lines were validated using a Surveyor Assay for *Pten*, *Trp53*, *Nf1* and *Cd83* indels according to the manufacturer's instructions (IDT, 1075932). The following primers were used for PCR amplification of the indel-containing locus:

Trp53 forward	GCTTTCCCACCCTCGCATAA
Trp53 reverse	TCACACGAAAGACAACTCCCC
Nf1 forward	TCTGTACCTCTTGGACTATGCC
Nf1 reverse	TGAGCCTCAAAACTTGCTTGG
Pten forward	AGGATTATCCGTCTTCTCCCCA
Pten reverse	ACCCTCAAATGTGCACCGTC
Cd83 forward	CCAAGCGCGGGTACAAGA
Cd83 reverse	CTCTCTCAGAACCTCGCTGA

In vitro cell proliferation assay

In vitro cell proliferation was assessed for mouse tumor cell lines using the Click-iT EdU Assay (Thermofisher, C10340) according to the manufacturer's guidelines. Briefly, $6x10^4$ cells were seeded onto poly-D-Lysine coated 12 mm coverslips in a 12-well culture dish. After 48 hours, cells were pulsed with EdU for 2 hours before being fixed with 4% paraformaldehyde and processed for immunostaining. Quantification was performed using 27 images per experimental group (n=3) (n=3 images per coverslip × n=9 coverslip per experimental group). EdU positivity was assessed as a fraction of total cells using Hoechst counterstaining. Images were processed using the Analyze Particles plugin in ImageJ.

Coculture experiment

Mouse tumor cell lines were seeded into T25 flasks and grown to 50% confluency. Naïve CD8+ T cells were harvested from CD1 wildtype P25 mice and processed using the EasySep™ Mouse Naïve CD8+ T Cell Isolation Kit (Stemcell Technologies, 19858). Following isolation, $10x10^5$ naïve CD8+ T cells were seeded onto mouse tumor cell lines and maintained in RPMI 1640 media (Thermofisher, 11875093) supplemented with 10% fetal bovine serum (FBS; Thermofisher, 16000044) and 1% Penicillin-streptomycin solution. Cells were left in coculture for one week before T cell-containing media was harvested for use in imaging flow cytometry experiments.

Multispectral imaging flow cytometry

For CD83 experiments, GFP+ tumor tissue from control pB-IUE tumor mice at P65 was processed into single-cell suspension according to our above protocol. Briefly, cells were spun at 400 *g* for 3 minutes at 4°C and washed in MACS Buffer (Miltenyi Biotec, 130-091-222) supplemented with 4% FBS (FBS; Thermofisher, 16000044) 3× before being incubated with APC-conjugated rat anti-CD83 antibody (1.25 μL/$10x10^6$ cells; Biolegend, 121510) for 30 minutes on ice. Cells were spun at 400 *g* for 3 minutes at 4°C and washed in MACS Buffer 3×

again before being resuspended in 50 µL of MACS Buffer for imaging experiments. High expression of endogenous GFP was used to gate tumor cells.

Image cytometry assays were acquired on an Imagestream X MKII (Luminex) equipped with a 405nm, 488nm, 561nm, 633nm and 785nm scatter laser. Collection was performed on as many objects as feasible in two hours. The number of objects collected ranged from 100,000 - 300,000 to allow for analysis of rare CD83+ events. Objects were analyzed and gated using IDEAS software 6.3.23.0. Cells were gated using Aspect Ratio and Area parameters for the brightfield channel. Focused Cells were gated using the Gradient_RMS parameter for Channel 1 (Brightfield). CD83+GFP+ were gated by signal intensities for Channel 2 (GFP) and Channel 11 (CD83 APC).

For T cell experiments, T cell-containing media was harvested from cocultures, spun at 700 *g* for 3 minutes at 4°C and washed in MACS Buffer 3× before being incubated with the following antibodies: APC-conjugated rat anti-CD3 (5 µL/10x10^6 cells; Biolegend, 100408), PE-conjugated rat anti-CD8b (2.5 µL/10x10^6 cells; Biolegend, 126608), PE/Cyanine7-conjugated rat anti-CD25 (2.5 µL/10x10^6 cells; Biolegend, 101916), Brilliant Violet 421-conjugated rat anti-CCR7 (5 µL/10x10^6 cells; Biolegend, 120120) and Cells were incubated for 30 minutes on ice, spun at 700 *g* for 3 minutes at 4°C and washed in MACS Buffer 3×. Cells were resuspended in CytoFix/CytoPerm Solution (Thermofisher, 00-5523-00) and incubated for 30 minutes at 4°C in the dark and washed 3× before being resuspended in 50 µL of MACS Buffer for imaging experiments. 20,000 cells were acquired based on Aspect Ratio vs Area of Brightfield. Data were analyzed by gating on Focused, Single Cells, followed by Channel 2 (CD8 PE) and Channel 11 (CD3 APC) to determine CD3+/-CD8+/- cells. These were then analyzed using Channel 7 (CCR7 BV421) and Channel 6 (CD25 PE-Cy7) intensities.

Representative images used for figures were exported from the IDEAS image gallery as .tif files and inserted into manuscript. Raw data files are available upon request.

ELISA

Media was collected from mouse tumor cells either cocultured with naïve CD8+ T cells for one week or in the absence of T cells. Cells were maintained in T25 tissue culture flasks with RPMI 1640 media (Thermofisher, 11875093) supplemented with 10% fetal bovine serum (FBS; Thermofisher, 16000044) and 1% Penicillin-streptomycin solution. Harvested media was spun at 3000 *g* for 15 minutes to remove debris and cells. Supernatant was collected and used for subsequent ELISA experiments. The following ELISA kits were used per the manufacturer's instructions: mouse IFNγ Quantikine ELISA kit (R&D, MIF00), mouse TNFα Quantikine ELISA kit (R&D, M2000), and mouse IL-2 Quantikine ELISA kit (R&D, M2000).

Human tumor slice preparation

Fresh tumor samples were immediately placed into a cold (0−4 °C) oxygenated NMDG solution (93 mM NMDG, 93 mM HCl, 2.5 mM KCl, 1.2 mM NaH2PO4, 30 mM NaHCO3, 20 mM HEPES, 25 mM glucose, 5 mM sodium ascorbate, 2 mM Thiourea, 3 mM sodium pyruvate, 10 mM MgSO4 and 0.5 mM CaCl2, pH 7.35). Slices were cut at 300 µm thickness with a microslicer (Leica VT 1200) and kept at 37.0 ± 0.5 °C in oxygenated NMDG solution for 10-15 minutes before being transferred to ACSF (125 mM NaCl, 2.5 mM KCl, 1.25 nM NaH2PO4, 25 mM NaHCO3, 1 mM MgCl2, 25 mM glucose and 2 mM CaCl2, pH 7.4) for 1 hour before recording.

Patch-seq recording procedures

Electrophysiological, morphological and transcriptomic data from the same cell were obtained simultaneously using the Patch-seq protocol described previously[57,58] . Briefly, patch pipettes

(5−7 MΩ) were filled with RNase-free intracellular solution (111 potassium gluconate, 4 KCl, 10 HEPES, 0.2 EGTA, 4 MgATP, 0.3 Na3GTP, 5 sodium phosphocreatine, and 13.4 biocytin). Whole-cell recordings were performed using 1 Quadro EPC 10 amplifiers (HEKA Electronic). After 5-10 minutes of the whole-cell recording of firing patterns, the nucleus is extracted using gentle and continuous negative pressure. The contents in the pipette were ejected into a 0.2-mL PCR tube containing 4 mL lysis buffer[58]. RNA in the lysis buffer was denatured and reversed transcribed, amplified and purified following Smart-seq2-based protocol[59]. Only high-quality cDNA samples (yield ≥ 2 ng, average length ≥ 1500 bp) were sequenced.

Sequencing libraries were constructed from the cDNA using the Illumina Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1096). The cDNA library was sequenced on the NovaSeq 6000 instrument using 150-bp paired-end reads.

Biocytin staining and morphological reconstructions

Following slice recordings, slices were fixed by immersion in the fixation solution at 4° for at least 48 hours, and then processed with an avidin-biotin-peroxidase method to reveal cell morphology. The morphology of the cells was reconstructed and analyzed using a 100× oil-immersion objective lens and camera lucida system (Neurolucida, MicroBrightField).

Mouse tumor slice preparation

Mouse tumor brain slices were obtained following previously described protocols[60]. In brief, the animals were deeply anaesthetized using 3% isoflurane and decapitated. The brain was rapidly removed and collected into cold (0–4 °C) oxygenated NMDG (N-methyl-d-glucamine) solution containing 93 mM NMDG, 93 mM HCl, 2.5 mM KCl, 1.2 mM NaH2PO4, 30 mM NaHCO3, 20 mM HEPES, 25 mM glucose, 5 mM sodium ascorbate, 2 mM thiourea, 3 mM sodium pyruvate, 10 mM MgSO4 and 0.5 mM CaCl2, pH 7.35 (all from Sigma-Aldrich). We cut 300-μm-thick coronal slices using a Leica VT1200 microtome following coordinates provided in the Allen Brain Atlas for adult mouse (http://atlas.brain-map.org). The slices were subsequently incubated at 34.0 ± 0.5 °C in oxygenated NMDG solution for 10–15 min before being transferred to the artificial cerebrospinal fluid (ACSF) solution containing: 125 mM NaCl, 2.5 mM KCl, 1.25 mM NaH2PO4, 25 mM NaHCO3, 1 mM MgCl2, 11.1 mM glucose and 2 mM CaCl2, pH 7.4 (all from Sigma-Aldrich) for about 1 h. The slices were allowed to recover in ACSF equilibrated with CO2/O2 gas mixture (5% CO2, 95% O2), at room temperature (approximately 25 °C) for 1 h before experiments. During the recordings, slices were submerged in a customized chamber continuously perfused with oxygenated physiological solution. Recorded cells were generally located 15–60 μm deep under the slice surface.

Mouse surgery for *in vivo* imaging studies

Tumor mice were generated using the aforementioned pB-IUE system. A piggyBac construct containing mCherry-hGeminin (1–110), a fluorescently-labeled, ubiquitination-based, cell cycle indicator (FUCCI) was cloned from the pRetroX-SG2M-Red Vector (Takara Bio, 631465) and added to the pB-IUE cocktail. This FUCCI construct allows for identification of cell nuclei that are transitioning from S to G2/M phases. Mouse surgeries were performed according to our previously published protocols[61].

Intracranial viral injections

A nanoliter injector (Nanoject II, Drummond scientific, Broomall, USA), was used to inject AAV-FLEX-GCaMP8m virus. Glass pipettes were pulled using a Sutter P-87 horizontal pipette puller (Sutter Instruments) and tips were broken on the filament of a vertical puller (Narishige). The pipette was backfilled with corn oil, and 5 μL virus solution was aspirated from a sterile piece of

parafilm using the Nanoject II. A total of 8 or 6 (depending on the unobstructed skull area available) equidistant injection locations were selected throughout the sensory and posterior motor cortex. For each hemisphere, these were at [bregma –3.5 mm; 2.5 mm lateral], [bregma –2 mm; 1 mm lateral], [bregma –2 mm; 4 mm lateral], and [bregma –0.5 mm; 2.5 mm lateral]. At each location, a total of 400 nL AAV-solution at 2 depths, 300 µm, 600 µm, were injected into the cortex in 9.2 nL/pulse increments separated by 10 seconds. The Nanoject was mounted at a 25-degree angle relative to the skull surface at each location and actuated by a manual and 1-direction motorized micromanipulator (WPI) at speeds of 600 µm per min. After the last injection, the skull was covered with Vetbond and dental cement.

2-photon *in vivo* imaging

2-photon images of calcium reporter activity were assessed according to previously published protocols[61]. Briefly, calcium reporter activity was acquired using a Prairie Ultima IV 2-photon microscope through a 25× objective, 1.1 NA, or a 16× objective, 0.8 NA, at 920 nm (GCaMP) under spiral (10–20 Hz frame rate) or resonant scan mode (30-35 Hz). A 525/70 nm emission filter was used for GCaMP indicators, and a 620/60 nm filter for mCherry emission. Mice were imaged while awake, head-posted in a holding frame and allowed to run freely on a circular treadmill. FOVs for further analysis were chosen based on proximity to the primary tumor mass as identified by FUCCI-labeled cells.

2-photon images were processed using a MATLAB pipeline modified from suite2p. Images were acquired using the Bruker acquisition software (Prairieview 5.3), and converted into 4GB ".hdf5"-stacks in ImageJ. DF/F calcium traces were resampled to the original sampling rate, and detrended DF/F traces were deconvolved using the suite2p/OASIS approach[62] to compute activity traces equivalent to action potential firing rates, integrated over time scales of 30-100 ms. All further analysis were conducted using these deconvolved traces.

Statistical analysis

Sample sizes and statistical tests are provided in the figure legends. The following tests were used for statistical analysis, unless otherwise noted. For Kaplan–Meier survival analysis, the log-rank test was used to compare survival differences across groups. For quantified results, one-way ANOVA was used followed by Welch's t test to compare individual means. For electrophysiology analyses, a Kruskal-Wallis test or two-way ANOVA was used followed by unpaired t tests with a two-stage step-up (Benjamini, Krieger, and Yekutieli). For RT-qPCR, a two-tailed Student's t test was used to determine statistical significance. Significant differences are denoted by asterisks in associated the graphs. Data are presented as the mean ± SEM. Levels of statistical significance are indicated as follows: ns: not significant, *p<0.05, **p<0.01, ***p<0.001, and ****p<0.0001.

REFERENCES

1. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).

2. Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nature Communications* vol. 11 Preprint at https://doi.org/10.1038/s41467-020-17186-5 (2020).

3. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
4. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
5. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
6. Richards, L. M. *et al.* Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* **2**, 157–173 (2021).
7. Harmanci, A. S., Harmanci, A. O. & Zhou, X. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nature Communications* vol. 11 Preprint at https://doi.org/10.1038/s41467-019-13779-x (2020).
8. Harmanci, A. O., Harmanci, A. S., Klisch, T. & Patel, A. J. XCVATR: Characterization of Variant Impact on the Embeddings of Single -Cell and Bulk RNA-Sequencing Samples. Preprint at https://doi.org/10.1101/2021.06.01.446668.
9. Bai, H. *et al.* Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nat. Genet.* **48**, 59–66 (2016).
10. Harmancı, A. S. *et al.* Integrated genomic analyses of de novo pathways underlying atypical meningiomas. *Nat. Commun.* **9**, 16215 (2018).
11. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4009–4014 (2013).
12. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
13. Fujimoto, Y. *et al.* CD83 expression influences CD4+ T cell development in the thymus. *Cell* **108**, 755–767 (2002).

14. Neefjes, J., Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
15. Lecoultre, M., Dutoit, V. & Walker, P. R. Phagocytic function of tumor-associated macrophages as a key determinant of tumor progression control: a review. *J. Immunother. Cancer* **8**, (2020).
16. Park, Y. *et al.* Imaging flow cytometry protocols for examining phagocytosis of microplastics and bioparticles by immune cells of aquatic animals. *Front. Immunol.* **11**, 203 (2020).
17. Chen, F. & LoTurco, J. A method for stable transgenesis of radial glia lineage in rat neocortex by piggyBac mediated transposition. *J. Neurosci. Methods* **207**, 172–180 (2012).
18. Breloer, M. & Fleischer, B. CD83 regulates lymphocyte maturation, activation and homeostasis. *Trends Immunol.* **29**, 186–194 (2008).
19. Irving, B. A. & Weiss, A. The cytoplasmic domain of the T cell receptor zeta chain is sufficient to couple to receptor-associated signal transduction pathways. *Cell* **64**, 891–901 (1991).
20. Baniyash, M. TCR zeta-chain downregulation: curtailing an excessive inflammatory immune response. *Nat. Rev. Immunol.* **4**, 675–687 (2004).
21. Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.* **14**, 24–35 (2014).
22. Ziegler, E. *et al.* CCR7 signaling inhibits T cell proliferation. *J. Immunol.* **179**, 6485–6493 (2007).
23. Suzuki, M. *et al.* CD8+CD45RA+CCR7+FOXP3+ T cells with immunosuppressive properties: a novel subset of inducible human regulatory T cells. *J. Immunol.* **189**, 2118–2130 (2012).
24. Bradley, J. R. TNF-mediated inflammatory disease. *J. Pathol.* **214**, 149–160 (2008).

25. Siegel, J. P. Effects of interferon-gamma on the activation of human T lymphocytes. *Cell. Immunol.* **111**, 461–472 (1988).
26. Dutt, S. *et al.* CD8+CD44(hi) but not CD4+CD44(hi) memory T cells mediate potent graft antilymphoma activity without GVHD. *Blood* **117**, 3230–3239 (2011).
27. Jiang, X. *et al.* Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015).
28. Scala, F. *et al.* Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).
29. Drew, L. J., Fusi, S. & Hen, R. Adult neurogenesis in the mammalian hippocampus: why the dentate gyrus? *Learn. Mem.* **20**, 710–729 (2013).
30. Espósito, M. S. *et al.* Neuronal differentiation in the adult hippocampus recapitulates embryonic development. *J. Neurosci.* **25**, 10074–10086 (2005).
31. Overstreet Wadiche, L., Bromberg, D. A., Bensen, A. L. & Westbrook, G. L. GABAergic signaling to newborn neurons in dentate gyrus. *J. Neurophysiol.* **94**, 4528–4532 (2005).
32. Xie, Z. *et al.* Gene set knowledge discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).
33. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-7 (2016).
34. Saracci, R., IARC & Wild, C. P. *International Agency for Research on Cancer*. (IARC, 2015).
35. Habashy, K. J., Mansour, R., Moussalem, C., Sawaya, R. & Massaad, M. J. Challenges in glioblastoma immunotherapy: mechanisms of resistance and therapeutic approaches to overcome them. *Br. J. Cancer* **127**, 976–987 (2022).
36. Buerki, R. A., Chheda, Z. S. & Okada, H. Immunotherapy of primary brain tumors: Facts and hopes. *Clin. Cancer Res.* **24**, 5198–5205 (2018).
37. Grabowski, M. M. *et al.* Immune suppression in gliomas. *J. Neurooncol.* **151**, 3–12 (2021).

38. Woroniecka, K. I., Rhodin, K. E., Chongsathidkiet, P., Keith, K. A. & Fecci, P. E. T-cell dysfunction in glioblastoma: Applying a new framework. *Clin. Cancer Res.* **24**, 3792–3802 (2018).

39. Yu, J. S. *et al.* Vaccination with tumor lysate-pulsed dendritic cells elicits antigen-specific, cytotoxic T-cells in patients with malignant glioma. *Cancer Res.* **64**, 4973–4979 (2004).

40. Bu, N. *et al.* Exosome-loaded dendritic cells elicit tumor-specific CD8+ cytotoxic T cells in patients with glioma. *J. Neurooncol.* **104**, 659–667 (2011).

41. Yu, K. *et al.* PIK3CA variants selectively initiate brain hyperactivity during gliomagenesis. *Nature* **578**, 166–171 (2020).

42. John Lin, C.-C. *et al.* Identification of diverse astrocyte populations and their malignant analogs. *Nat. Neurosci.* **20**, 396–405 (2017).

43. Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539–545 (2019).

44. Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573**, 532–538 (2019).

45. Correia, C. E., Umemura, Y., Flynn, J. R., Reiner, A. S. & Avila, E. K. Pharmacoresistant seizures and IDH mutation in low-grade gliomas. *Neurooncol Adv* **3**, vdab146 (2021).

46. Tate, J. G. *et al.* COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).

47. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

48. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8**, 289–317 (2016).

49. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

50. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

52. Harrow, J. *et al.* GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

54. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

55. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

56. Zhu, W. *et al.* Daam2 driven degradation of VHL promotes gliomagenesis. *Elife* **6**, (2017).

57. Cadwell, C. R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).

58. Cadwell, C. R. *et al.* Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using Patch-seq. *Nat. Protoc.* **12**, 2531–2553 (2017).

59. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

60. Aiba, I., West, A. K., Sheline, C. T. & Shuttleworth, C. W. Intracellular dialysis disrupts Zn2+ dynamics and enables selective detection of Zn2+ influx in brain slice preparations. *J. Neurochem.* **125**, 822–831 (2013).

61. Meyer, J., Yu, K., Deneen, B. & Noebels, J. Glioblastoma disrupts cortical network activity at multiple spatial and temporal scales. *bioRxiv* 2022.08.31.505988 (2022) doi:10.1101/2022.08.31.505988.

62. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).

Table S1. Characteristics of patient samples included in this study.

Patient Identity	Tumor Type	*IDH1* Status	1p19 Codeletion	WHO Grade	Samples	Experiments
A	diffuse astrocytoma	mutant	no	IV	core leading edge	scRNA-seq
B	GBM	wildtype	no	IV	core leading edge	scRNA-seq
C	GBM	wildtype	no	IV	core leading edge	scRNA-seq
D	oligodendroglioma	mutant	yes	II	core	scRNA-seq
E	GBM	wildtype	no	IV	core leading edge	scRNA-seq
F	GBM	wildtype	no	IV	core	scRNA-seq
G	diffuse astrocytoma	mutant	no	IV	core leading edge	scRNA-seq
H	diffuse astrocytoma	mutant	no	III	core	whole-cell patch clamp
I	diffuse astrocytoma	mutant	no	IV	core	Patch-seq scRNA-seq
J	GBM	wildtype	no	IV	core	whole-cell patch clamp