Spiking GABAergic OPC tumor cells prolong survival in *IDH1* mutant glioma

Rachel Naomi Curry, B.A. [1,2,+], Qianqian Ma, Ph.D. [3,4,+], Malcolm F. McDonald, B.S. [2,5,6], Yeunjung Ko[2,7,8], Snigdha Srivastava[3,5,9], Pey-Shyuan Chin[4], Peihao He[2,7,10], Brittney Lozzi[11], Junzhan Jing, Ph.D.[3,4], Prazwal Athukuri, B.S.[8], Su Wang, Ph.D. [8], Arif O. Harmanci, Ph.D.[12], Benjamin Arenkiel, Ph.D.[4,9], Xiaolong Jiang, Ph.D. [3,4,13,*,‡], Benjamin Deneen, Ph.D. [1,2,6,7,8,10,*,‡], Ganesh Rao, M.D. [7,8,*,‡], Akdes Serin Harmanci, Ph.D. [7,8,*,‡]

[1] The Integrative Molecular and Biomedical Sciences Graduate Program, Baylor College of Medicine, Houston, TX, USA

[2] Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX

[3] Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX

[4] Department of Neuroscience, Baylor College of Medicine, Houston, TX

[5] Medical Scientist Training Program, Baylor College of Medicine, Houston, TX

[6] Program in Development, Disease Models, and Therapeutics, Baylor College of Medicine, Houston, TX

[7] Center for Cancer Neuroscience, Baylor College of Medicine, Houston, TX

[8] Department of Neurosurgery, Baylor College of Medicine, Houston, TX

[9] Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX

[10] Program in Cancer Cell Biology, Baylor College of Medicine, Houston, TX, USA

[11] Program in Genetics and Genomics, Baylor College of Medicine, Houston, TX, USA

[12] McWilliams School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX

[13] Department of Ophthalmology, Baylor College of Medicine, Houston, TX, USA

[+] These authors contributed equally to this work

* These authors contributed equally to this work

[‡] Corresponding Author

Abstract

Prior studies have shown that glioma cells form synapses with neurons to receive synaptic inputs. To discern if glioma cells can send outgoing electrochemical signals in the form of action potentials (APs), we employed Patch-sequencing on surgically-resected human glioma slices. Results showed that half of patched cells in *IDH1* mutant ($IDH1^{mut}$) tumors demonstrate select properties of both neurons and glia and fire single, short APs. To define the transcriptional profiles of these hybrid cells (HCs) and discern if they are tumor derived, we developed a computational tool, Single Cell Rule Association Mining (SCRAM), to annotate features in each cell individually. SCRAM revealed that HCs represent a heterogenous group of tumor and non-tumor cells that are uniformly defined by both GABAergic neuron and oligodendrocyte precursor cell (GABA-OPC) transcriptional signatures. We found that GABA-OPC tumor cells express requisite voltage-gated ion channels needed to fire APs. We validated our findings in human single cell and bulk RNA-seq datasets, confirming that GABA-OPCs represent 40% of $IDH1^{mut}$ tumor cells, correlate with survival outcomes in $IDH1^{mut}$ human patients and are also found in select molecular subtypes of *IDH1* wild-type tumors. These studies describe a new cell type in $IDH1^{mut}$ glioma with unique electrophysiological and transcriptomic properties.

Introduction

Glioma is the most common central nervous system tumor with an estimated 20,000 cases diagnosed each year. These diffuse tumors include *IDH1* mutant ($IDH1^{mut}$) and *IDH1* wildtype ($IDH1^{WT}$) subtypes, each of which presents with unique clinical and histopathological correlates. Prognostic outcomes for $IDH1^{WT}$ tumors are poor, conferring a median survival of less than 14 months[1]. In contrast, $IDH1^{mut}$ tumors confer significantly better prognoses, with a median survival of 31 months after diagnosis[2]. Robust molecular and genomic studies conducted over the last two decades have revealed that the disparity in survival outcomes between glioma subtypes is primarily attributed to differences in tumor cell proliferative capacity and invasiveness[3], which are mediated by an intricate compendium of tumor intrinsic and extrinsic factors. Communication between tumor cells and their microenvironmental constituents has proven to be a critical mediator of glioma progression and is largely mediated by immunological and neural cellular components[4,5]. With regards to the latter, recent advances in the field of cancer neuroscience have revealed that glioma cells form functional synapses with peritumoral neurons and that heightened neuronal activity promotes glioma progression via increased proliferation and invasiveness[5–7]. While glioma cells have been shown to receive excitatory post synaptic potentials (EPSPs) from surrounding neurons[6], whether tumor cells can generate outgoing information in the form of propagating action potentials (APs) remains unclear.

One of the greatest impediments to successful treatment of glioma is the challenge of identifying tumor-specific attributes that can be exploited for therapeutic gain. Over the past decade, technological advances in single cell genomics have generated an

abundance of sequencing data, elucidating the robust transcriptional and genomic heterogeneity that exists in glioma [8–10]. Moreover, Patch-sequencing (Patch-seq), which couples whole cell recordings, morphological analysis and single cell RNA-sequencing (scRNA-seq) permits characterization of both electrophysiological and transcriptomic features from individual cells[11], and has been used to expound the extraordinary array of neuronal subtypes present in the mammalian brain[12]. Despite these technological advances, however, a computational framework capable of accurately distinguishing glioma cells 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. Because glioma cells frequently share molecular profiles with their non-tumor glial analogs, reliable identification of tumor cells within the brain is a difficult but necessary task.

To address the aforementioned knowledge gaps, we performed *in situ* Patch-seq studies in human glioma and developed a new single cell computational tool to characterize the genomic and transcriptomic features of recorded cells. We found that nearly half of patched cells in IDH1^{mut} tumors uniformly fire single, short APs and that this unique electrophysiological phenotype is indicative of hybrid cells (HCs) possessing select properties of both neurons and glia. Using our novel computational tool, Single Cell Rule Association Mining (SCRAM) that is capable of accurately annotating tumor cells and defining hybrid cellular states for individual cells, we found that HCs possess transcriptomic features of both GABAergic neuron and oligodendrocyte precursor cells (OPCs), which we term GABA-OPC cells. Prior work has documented shared features between GABAergic neurons and OPCs, including the expression of GABA receptors

(GABARs)[13,14] and the ability to synthesize and breakdown GABA[15]. Using large scale scRNA-seq datasets and immunostaining of human tumor samples, we confirmed that GABA-OPC cells account for 40% of IDH1^{mut} tumor cells and are less proliferative than other tumor cell types. Importantly, IDH1^{mut} tumor patients with more robust GABA-OPC molecular signatures show significantly better survival outcomes as compared to those without, and lower grade tumors were enriched for GABA-OPC profiles as compared to higher grade tumors. These data suggest that the ability of tumor cells to generate functional electrophysiological outputs is more prevalent in lower grade tumors and is lost as tumor cells become more proliferative and progress to higher grades. Collectively, our studies reveal a new tumor cell type in human glioma with important clinically-relevant outcomes and highlight the utility of using our computational framework for characterizing hybrid tumor cell types in glioma.

Results

IDH1^{mut} glioma cells fire single action potentials and are distinct from neurons and glia

To determine if glioma cells can fire APs, we performed whole-cell patch clamp recordings followed by scRNA-seq (Patch-seq) on brain slices resected from the tumor cores of two IDH1^{mut} glioma specimens (**Fig. 1a**). We recorded from 53 total cells; 31 were used to recover high-quality RNA for scRNA-seq, and 20 were preserved with biocytin for morphological analysis. For the 31 cells used for sequencing, an average of 4087 genes were identified per cell. Of the 53 total cells, 27 displayed electrophysiological and morphological properties consistent with either glia or neurons

in non-tumor brain[16,17] (**Figs. 1b-d**). In contrast, 26 cells displayed select neuronal electrophysiological properties but were morphologically consistent with glia (**Figs. 1e-g**). These cells, which we assigned as hybrid cells (HCs) electrophysiologically represented 49% of patched cells (**Fig. 1h**), had higher input resistances than neurons or glia (**Fig. 1i**), and were all capable of firing single, small action potentials (**Figs. 1j-k**). When we attempted to record from an IDH1^{WT} glioblastoma core sample, very few cells were patchable due to gliosis and necrosis; however, all three cells recorded were electrophysiologically consistent with glia and unable to generate APs. PCA analysis of scRNA-seq data showed that all HCs clustered together and most closely with glia rather than neurons, suggesting that HCs are transcriptionally more similar to glia than they are to neurons (**Fig. 1l**). Of the 31 cells used for Patch-seq, 15 had HC electrophysiology profiles (**Fig. 1m**). The canonical *IDH1R132H* mutation was detected in 5 of 31 cells, all 5 of which were HCs, confirming that IDH1^{mut} tumor cells are excitable and can fire *bona fide* APs. Interestingly, of the cells that had coverage of the *IDH1* locus, one HC cell was not *IDH1R132H* mutant, suggesting that both tumor and non-tumor cells can have HC electrophysiologies (**Fig. 1m**). These studies identified HCs as an electrophysiologically unique group of cells in the human brain and confirm that tumor and non-tumor cells in IDH1^{mut} gliomas possess HC electrophysiology signatures.

SCRAM is a reliable annotation tool for human scRNA-seq datasets

To better define the HCs identified in our Patch-seq studies, we sought to create a new computational platform that could annotate each cell from Patch-seq individually.

Because of the low cell numbers obtained using Patch-seq and the rarity of human glioma samples for use in these experiments, our annotation tool needed to be able to analyze each cell without a dependency on clustering methodologies, which requires hundreds of cells for optimal analysis. Accordingly, we developed the Single Cell Rule Association Mining (SCRAM) tool that can annotate each cell on a cell-by-cell basis, independently of clusters. SCRAM uses a three-step orthogonal process to provide detailed transcriptional and RNA-inferred genomic profiles for each cell: (1) cell type transcriptional annotation using machine-learned neural network models (NNMs); (2) single nucleotide variant (SNV) profiling using the XCVATR[18] tool; and (3) copy number variant (CNV) calling using the CaSpER[19] and NUMBAT[20] tools (**Fig. 2a-d**). SCRAM NNMs are trained from 11 previously published scRNA-seq (totaling ~1M cells) human tumor and non-tumor datasets, including developing brain, immune, and glioma cell atlases[9,10,21–28]. Each cell from Patch-seq is assigned a probability score for each cell type in the training datasets **(Fig. 2e)**. Using this methodology, individual cells may be annotated as more than one cell type, which permits for the characterization of hybrid cell states like those of HCs from our Patch-seq experiments. SNVs and CNVs are also considered for each cell and added to the cell annotation. Cells are assigned as “tumor” if they have ≥ 2 tumor features (see *Methods*). We validated that SCRAM can reliably discriminate between tumor and non-tumor tissue, which it does with >99% sensitivity (true positive rate: TPR) in Allen Brain Atlas[21], Bhaduri et al.[23] , Aldinger et al. [27], CoDEx[24] and >96% sensitivity in Human Protein Atlas (HPA)-Brain[28] (Allen Brain Atlas: 559 tumor cells out of 76,533 cells; HPA-Brain: 1333 tumor cells out of 76,533 cells; Bhaduri et al. Developing Brain: 4237 tumor cells out of 404,218 cells, Aldinger et al.

Developing Brain: 586 tumor cells out of 69,174 cells, CoDEx Developing Brain: 559 tumor cells out of 75,974 cells). Additionally, we ran SCRAM on glioma scRNA-seq datasets[8,10,22], in which tumor non-tumor annotations was reported in cluster resolution, SCRAM achieved 100% specificity (true negative rate: TNR) for tumor cells in $IDH1^{mut}$ and $IDH1^{WT}$ glioma datasets.

HCs express hybrid GABAergic neuron-OPC transcriptomes

We ran SCRAM on the 31 cells from Patch-seq experiments and found that cells with pyramidal cell electrophysiologies were correctly annotated as glutamatergic neurons, whereas those with interneuron electrophysiologies were appropriately annotated as GABAergic neurons (**Fig. 2e**). Consistent with their electrophysiological profiles, HCs showed concurrent annotation as tumor cells, OPCs, oligodendrocytes and GABAergic neurons (**Fig. 2e**), suggesting that HCs are endowed with a hybrid of neuronal and glial transcriptional features that may be responsible for the functional electrophysiological properties of these cells. To validate these findings in larger datasets, we used SCRAM to analyze our in-house scRNA-seq dataset consisting of 234,880 cells from 12 $IDH1^{mut}$ and $IDH1^{WT}$ glioma patients[29] (**Fig. 2b-d**). SCRAM-assigned probability scores >0.9 were used for final cell annotations and probability scores used to generate a SCRAM UMAP of our scRNA-seq glioma dataset, which clusters cells based on cell identity rather than similarity of transcriptional features (**Fig. 3a**). Visualization of SNVs, CNVs and tumor expression markers specific to glioma subtypes revealed that SCRAM segregates the majority of $IDH1^{mut}$ from $IDH1^{WT}$ tumor cells (**Fig. 3b**). Given that the HCs we observed in our Patch-seq experiment were found in $IDH1^{mut}$ patients, we focused on SCRAM clusters encompassing the majority of $IDH1^{mut}$ tumor cells. These

cells were predominantly distributed amongst three discrete SCRAM clusters (Clusters 13, 18, and 19) that collectively contained less than 5% of IDH1^{WT} tumor cells (**Figs. 3c-e**). Importantly, all IDH1^{mut} tumor patients had robust GABA-OPC expression profiles, whereas only one IDH1^{WT} patient showed the same expression profile (**Figs. 3f-g**). An analysis of tumor cells by patient revealed that on average 41.3% of IDH1^{mut} tumor cells were GABA-OPCs, whereas only 2.2% of IDH1^{WT} tumor cells received the same annotation (**Fig. 3h**). These SCRAM analyses revealed that HCs are defined by a hybrid GABA-OPC transcriptome, which is found in nearly half of IDH1^{mut} tumor cells.

Having identified that HCs are defined by GABA-OPC transcriptomes and that these cells are robustly present in most IDH1^{mut} patients, we sought to extract a GABA-OPC molecular signature by identifying transcriptional markers. To this end, we used SHapley Additive exPlanations (SHAP) analysis, which extracts essential features from machine-learning NNMs. We identified SHAP markers from each cell type in our training dataset (e.g. GABAergic PVALB neuron from Allen Brain Atlas; and compared these SHAP genes to the differentially expressed genes (DEGs) obtained for GABA-OPC tumor cells in our scRNA-seq dataset as compared to other non-GABA-OPC tumor cells. The intersection of SHAP genes and DEGs produced a list of 61 genes ($\log_2$FC > 1), which we termed the GABA-OPC gene set. The GABA-OPC gene set is comprised of SHAP genes that are critical transcriptional features of cell types from tumor and non-tumor training datasets (**Fig. 4a**). The training cell types that were most highly represented by GABA-OPC SHAP features were GABAergic neuron and OPC, demonstrating that the majority of GABA-OPC transcriptional features derive from

GABAergic neurons and OPCs. These data define a unique transcriptomic signature of GABA-OPC tumor cells that can be used to identify these cells in other datasets.

Strong GABA-OPC signatures are correlated with better survival in IDH1^{mut} patients

To validate our observation that GABA-OPC signatures were more prominent in our IDH1^{mut} cohort as compared to IDH1^{WT}, we used the GABA-OPC-like glioma geneset to score 225 IDH1^{WT} and 441 IDH1^{mut} bulk RNA-seq samples from TCGA (**Fig. 4b**). Consistent with our internal scRNA-seq dataset, we found that nearly all IDH1^{mut} samples had high GABA-OPC scores, whereas IDH1^{WT} samples were distributed more broadly across low and high scores. An analysis of scores by IDH1^{WT} molecular subtypes revealed the highest GABA-OPC scores belonged to OPC-like and NPC-like molecular subtypes (**Fig. 4c**). Consistent with these observations, examining the percentage of GABA-OPC tumor cells in Neftel et al.'s IDH1^{WT} scRNA-seq dataset[9] revealed that OPC-like samples had significantly higher GABA-OPC tumor percentages than the other molecular subtypes (**Fig. 4d**). Reexamining our internal scRNA-seq dataset revealed that the only IDH1^{WT} patient with a strong GABA-OPC signature was an OPC-like subtype. Interestingly, analysis of the TCGA IDH1^{mut} samples by histopathological subtype showed that GABA-OPC scores were higher in low-grade glioma (LGG) than high-grade glioma (HGG), suggesting that stronger GABA-OPC phenotypes are associated with lower grade tumors (**Fig. 4e**). RNAvelocity pseudotime analyses of three IDH1^{mut} and one OPC-like IDH1^{WT} samples revealed that GABA-OPC tumor cells emerge from more primitive tumor cells to become the largest population of tumor cells in tumors that bear them, demonstrating that select tumor cells from

divergent genetic backgrounds of IDH1^{mut} and IDH1^{WT} converge at a shared transcriptional phenotype (**Figs. 4f-g**). We assigned GABA-OPC scores to bulk RNA-seq IDH1^{mut} glioma samples for which long-term survival follow up was collected[30]. Samples were split into low (n=41) and high (n=41) GABA-OPC groups based on median expression. Kaplan-Meier survival analysis showed that low GABA-OPC tumor patients had a median survival of 84.5 months whereas high GABA-OPC tumor patients showed a median survival of 156.9 months (**Fig. 4h**). Survival outcomes for HGG IDH1^{mut} patients alone were also significantly different, with a median survival of 80.5 and 182.5 for low and high GABA-OPC groups in, respectively (**Fig. 4i**). Collectively, these analyses confirm that GABA-OPC tumor cells are a defining feature of IDH1^{mut} glioma and the OPC-like subtype of IDH1^{WT} glioma and demonstrate that a decreased GABA-OPC transcriptional signature confers significantly worse survival outcomes in IDH1^{mut} glioma patients **(Fig. 4h-i,)**.

GABAergic neurons and OPCs share transcriptional profiles in non-tumor human brain

To discern whether AP-firing GABA-OPC tumor cells represent a population of hybrid neuronal-glial cells unique to tumor, we examined expression data of the non-tumor adult human brain scRNA-seq cell datasets from Allen Brain Atlas and the Human Protein Atlas (HPA). We examined pan- and subtype-specific GABAergic neuron markers as well as established OPC markers, finding that 6 of 30 GABAergic neuron markers were similarly expressed by OPCs (**Fig. 5a**). Conversely, and unexpectedly, *PDGFRA*, a widely-used OPC-specific marker, is highly expressed by GABAergic LAMP5+ neurons both in Allen Brain Atlas and the HPA, demonstrating a transcriptional

overlap that is uniquely shared by OPCs and GABAergic neurons. To determine whether GABA-OPC glioma features are also present in non-tumor cells, we examined the percentage of each cell type that express the GABA-OPC transcriptional signature. We found that on average 48% percent of OPCs from non-tumor brain atlases possessed GABA-OPC molecular profiles (**Fig. 5b**), suggesting that GABA-OPC tumor cells are not hybrid neuronal-glial tumor cells, but are malignant manifestations of an OPC subclass that is found in non-tumor brain.

GABA-OPC tumor cells express voltage-gated ion channels and GABA-related genes

Given prior reports documenting the existence of GABARs and GABA synthesis genes in non-tumor OPCs, we investigated the prevalence of these gene sets in non-tumor developing and adult scRNA-seq datasets from the human brain[31,32], finding that they were more enriched in non-tumor adult OPCs than in development (**Fig. 5c**). We found that GABA metabolism genes are also widely expressed by GABA-OPC HCs from our Patch-seq studies, including high expression of the GABAergic neuron marker and GABA synthesis gene *glutamate decarboxylase 1 (GAD1)*, which was also found in GABAergic interneurons but not detected in glutamatergic pyramidal cells (**Fig. 5d**). Notably, *GABA transaminase (ABAT)*, which is a crucial enzyme in GABA catabolism, was also detected at high levels in GABA-OPC HCs. Analysis of these genes in our larger scRNA-seq cohorts confirmed that *ABAT* and *GAD1* are highly expressed by GABA-OPC tumor cells (**Fig. 5c**), the latter of which we confirmed through immunostaining of human IDH1^{mut} tumor samples (**Fig. 5e**). GABARs were also robustly

expressed in GABA-OPC tumor cells as compared to other non-GABA-OPC tumor cells, suggesting that GABA-OPC tumor cells are transcriptionally equipped to receive electrochemical signals from GABAergic neurons (**Figs. 5c-d**). Importantly, previous work has shown that quantal GABA release from interneurons to OPCs at interneuron-OPC synapses can elicit miniature depolarizations in OPCs[33]. Given this precedent, our data suggest that the electrophysiological activity of GABA-OPC tumor cells may be in part conferred through the expression of GABARs and GABA metabolism genes that enables these cells to respond to signals from GABAergic neurons of other GABA-OPC tumor cells.

Having confirmed that GABARs and GABA metabolism genes are expressed by GABA-OPC tumor cells, we next sought to elucidate the molecular constituents that may be responsible for AP firing in HCs. Importantly, previous work demonstrated that a subgroup of white matter OPCs in the rat brain fire single[34], short APs that are similar to those observed in GABA-OPC HC cells from our Patch-seq studies. The AP capacity of these spiking OPCs is dependent on voltage-gated ion channels, particularly voltage-gated sodium channels (Na_vs) and voltage-gated potassium channels (K_vs), which we confirmed were highly expressed by human non-tumor OPCs (**Fig. 5f**). We found that the majority of GABA-OPC tumor cells have high expression of Na_vs and but that high expression of K_vs is restricted to a smaller population of cells. These findings were confirmed in our Patch-seq dataset, which revealed that Na_vs were broadly expressed by GABA-OPC HC cells, whereas high expression of K_vs was detected in fewer of these cells (**Fig. 5g**). Immunostaining for the voltage-gated sodium channel $Na_v1.1$, encoded by the *SCN1A* gene, in a human IDH1^{mut} tumor confirmed GABA-OPC tumor cells

express Na_vs, which are essential for the rising phase of APs[34,35]. These data suggest that GABA-OPC tumor cells are endowed with the requisite machinery that can mechanistically elicit and sustain the APs observed in these cells (**Fig. 5h**).

Discussion

In recent years, studies in the field cancer neuroscience have elucidated how glioma cells interact with surrounding neural networks to direct disease progression[36,37]. These reports demonstrate that human glioma cells receive synaptic inputs from the surrounding excitatory glutamatergic neuronal circuitry that can evoke EPSPs in glioma cells[6,7,38]. Whole cell patch clamp experiments from the 1990s documented that cells firing short, single APs were the majority in human glioma slices; however, conclusively showing that these cells were tumor in origin was not possible without the advent of single cell transcriptomics[39–41]. These early electrophysiology studies described cells that fire single, short APs that were dependent on voltage-gated sodium currents. Separately, the identification of a spiking class of OPCs in rats revealed that neurons are not the only cells capable of firing APs and suggested that an analogous population of spiking OPCs exists in human[34]. Given these precedents, their electrophysiological similarities to our Patch-seq HCs and the enrichment of voltage-gated ion channels that are expressed transcriptionally by both adult OPCs and GABA-OPC tumor cells, we believe that GABA-OPC tumor cells are likely the malignant analogs of spiking OPCs and constitute the major tumor cell population in $IDH1^{mut}$ patients. To this point, while five of the HCs from Patch-seq had the *IDH1R132H* canonical mutation, a single HC that was confirmed *IDH1* wild-type showed the same electrophysiological profile as

those *IDH1R132H* mutant HCs and likely represents a non-tumor spiking OPC. There is mounting evidence to support the OPC as a cell of origin in glioma[42–44], which represents the largest proliferative neural cell population in the adult brain[45]. Our studies further support the notion that OPCs undergo malignant transformation to become glioma cells and adds to our understanding of cancer neuroscience by implicating GABAergic metabolism and signaling as an important mediator of tumor cell neurophysiology. Interestingly, GABARs, which are highly expressed by GABA-OPC tumor cells, function to limit neuronal excitability in neurons by triggering an influx of chloride anions into the cell. The accumulating negative charge decreases the resting membrane potential, making subsequent depolarizations more difficult to achieve[46], raising the question of whether GABARs restrict tumor cells from firing more than one AP. Our analyses revealed that the GABA-OPC molecular profile that supports AP activity in tumor cells are lost in the process of malignant transformation from LGG to HGG and concordantly, that diminished GABA-OPC scores correlate with significantly worse survival outcomes in IDH1^{mut} patients. These findings suggest that GABA-OPC tumor cells can serve as a biomarker of premalignant brain tumors and may serve as the target of tumor-specific treatments in future endeavors.

Acknowledgements

This work was supported by grants from the NIH (R35-NS132230 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.

Data availability

In our previous study, we published the scRNA-seq datasets of 12 samples under the accession number GSE221534[29]. The Patch-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.

Author contributions

R.N.C. and A.S.H. are responsible for conception of this project, the study and pipeline design, and interpretation of the results. R.N.C. devised all experimental set ups, performed immunostaining and statistical analyses, and prepared all scRNA-seq samples from with assistance from M.F.M. A.S.H. wrote the code for the SCRAM pipeline. Q.M. and J.J. performed all human electrophysiology and Patch-seq experiments. R.N.C., A.O.H., S.W. and A.S.H. performed bioinformatics analyses on human scRNA-seq datasets. B.L., E.L.F., and A.R. provided experimental support for immunohistochemistry. G.R. identified and obtained consent from patients for the study. R.N.C. and A.S.H prepared the manuscript. 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.

Figures & Figure Legends

Figure 1

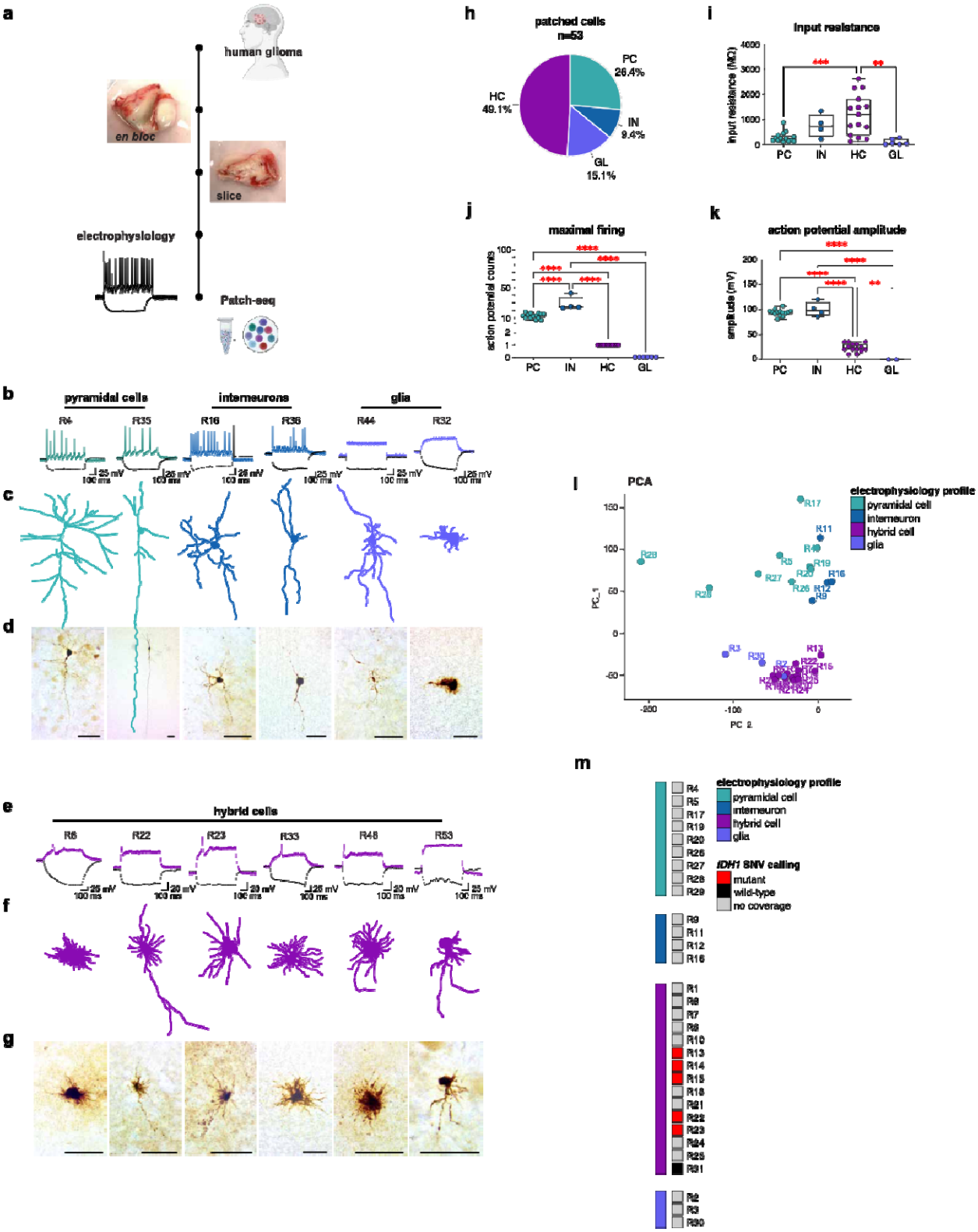

a
human glioma
en bloc
slice
electrophysiology
Patch-seq
b
pyramidal cells
interneurons
glia
R4
R35
R16
R36
R44
R32
c
d
e
hybrid cells
R6
R22
R23
R33
R48
R53
f
g
h
patched cells
n=53
PC
26.4%
IN
9.4%
GL
15.1%
HC
49.1%
i
input resistance
input resistance (MΩ)
PC
IN
HC
GL
j
maximal firing
action potential counts
k
action potential amplitude
amplitude (mV)
l
PCA
electrophysiology profile
pyramidal cell
interneuron
hybrid cell
glia
PC_1
PC_2
m
electrophysiology profile
pyramidal cell
interneuron
hybrid cell
glia
IDH1 SNV calling
mutant
wild-type
no coverage

Figure 1. Patch-seq of human glioma samples reveals tumor cells fire single action potentials. (a) Experimental workflow for whole-cell recordings and whole-cell patch clamp recordings, followed by singe-cell RNA-sequencing (Patch-seq) assays. (b) Exemplary membrane responses from patched *IDH1*-mutant ($IDH1^{mut}$) neuronal and glial cells to a 600-ms hyperpolarizing current step (black) and suprathreshold depolarizing current step (coloured). (c) Matched traced cell morphologies are shown for recorded neuronal and glial cells. (d) Matched images of biocytin-filled cell morphologies for patched neuronal and glial cells; scale bar: 50 µm. (e) Exemplary membrane responses from $IDH1^{mut}$ HCs to a 600-ms hyperpolarizing current step (black) and suprathreshold depolarizing current step (coloured). (f) Matched traced cell morphologies are shown for recorded HC cells. (g) Matched images of biocytin-filled cell morphologies for patched HC cells; scale bar: 50 µm**.** (h) Pie chart showing percentage of PC, IN, GL and HC patched cells. (i) Bar graph showing HC have higher input resistance compared with non-tumor neurons; *p*-values for pairwise comparisons are noted in the figure. (j) Bar graph showing HC cells fire fewer spikes compared to neurons; *p*-values for pairwise comparisons are noted in the figure. (k) Bar graph showing HC cells have AP amplitudes compared to neurons; *p*-values for pairwise comparisons are noted in the figure. (l) Principal component analysis (PCA) plot of 31 Patch-seq cells shows clustering of cells based on electrophysiological properties. HC cells are transcriptionally like each other and glia. Cells are coloured according to electrophysiological properties. (m) SNV profiling for the *IDH1R132H* mutation is shown for 31 Patch-seq cells. For Patch-seq, two voltage traces are shown: the hyperpolarization trace obtained with injected currents (black) and the depolarization

trace showing maximal AP firing rate; injected current: −100 pA. AP: action potential; GABA: γ-aminobutyric acid; GL: glia; GSC: glioma stem cell; IN: interneuron; OPC: oligodendrocyte precursor cell. PC: pyramidal cell.

Figure 2

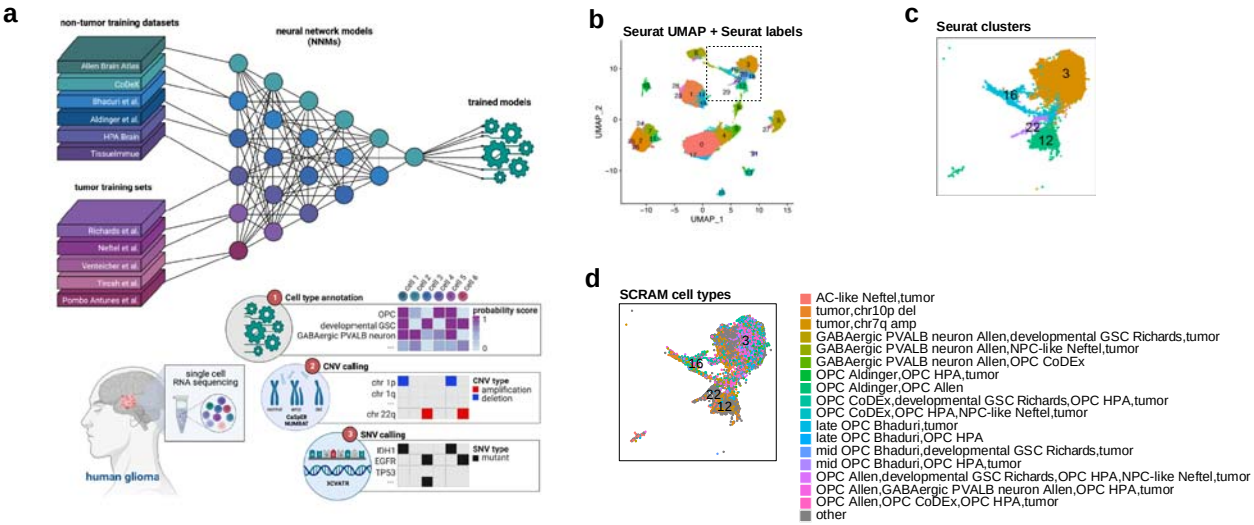
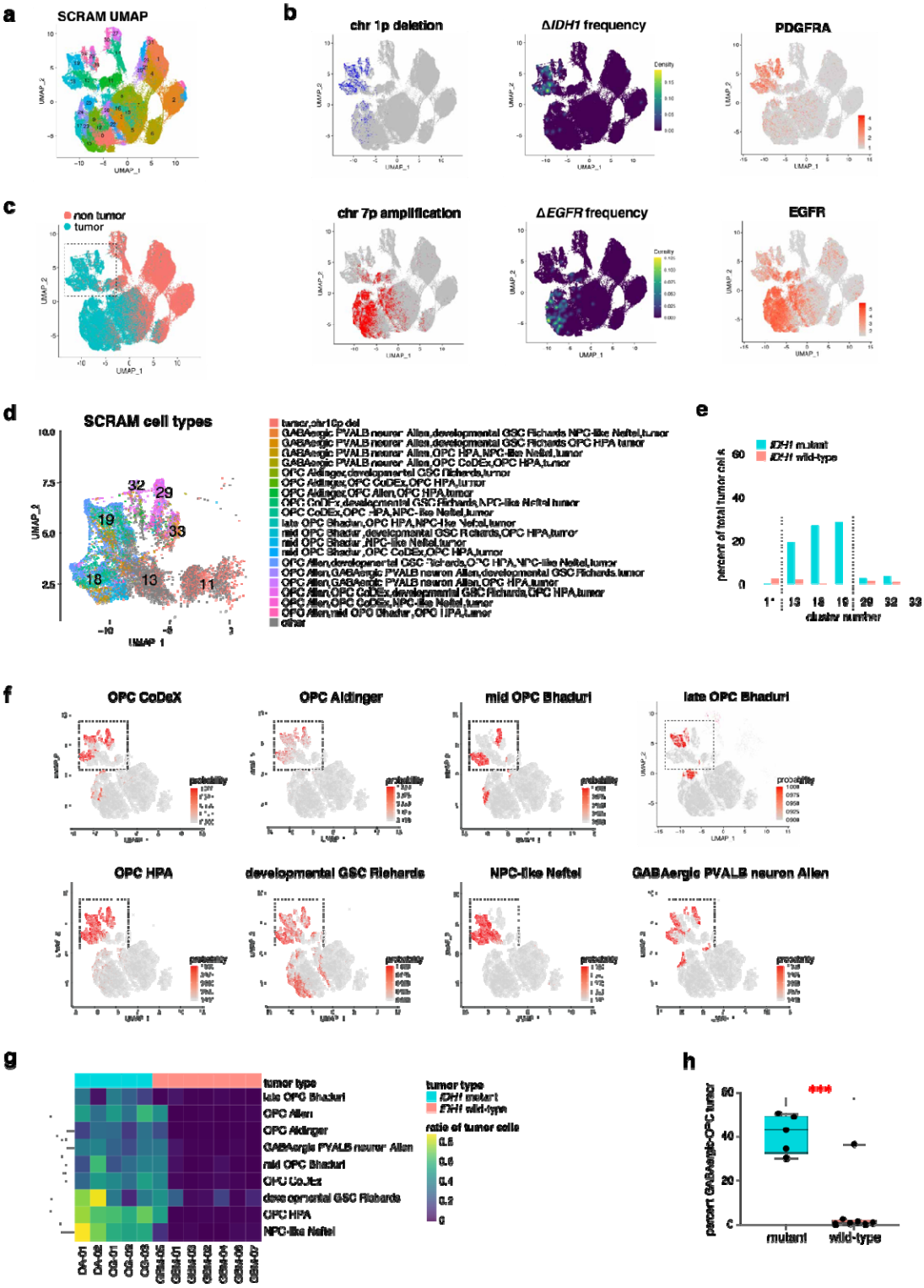

a
non-tumor training datasets
Allen Brain Atlas
CoDeX
Bhaduri et al.
Aldinger et al.
HPA Brain
Tissueimmune
neural network models (NNMs)
trained models
tumor training sets
Richards et al.
Neftel et al.
Venteicher et al.
Tirosh et al.
Pombo Antunes et al.
1 Cell type annotation
OPC
developmental GSC
GABAergic PVALB neuron
probability score
single cell RNA sequencing
2 CNV calling
chr 1p
chr 1q
chr 22q
CNV type
amplification
deletion
3 SNV calling
IDH1
EGFR
TP53
SNV type
mutant
human glioma
b
Seurat UMAP + Seurat labels
UMAP_1
UMAP_2
c
Seurat clusters
3
16
22
12
d
SCRAM cell types
AC-like Neftel,tumor
tumor,chr10p del
tumor,chr7q amp
GABAergic PVALB neuron Allen,developmental GSC Richards,tumor
GABAergic PVALB neuron Allen,NPC-like Neftel,tumor
GABAergic PVALB neuron Allen,OPC CoDEx
OPC Aldinger,OPC HPA,tumor
OPC Aldinger,OPC Allen
OPC CoDEx,developmental GSC Richards,OPC HPA,tumor
OPC CoDEx,OPC HPA,NPC-like Neftel,tumor
late OPC Bhaduri,tumor
late OPC Bhaduri,OPC HPA
mid OPC Bhaduri,developmental GSC Richards,tumor
mid OPC Bhaduri,OPC HPA,tumor
OPC Allen,developmental GSC Richards,OPC HPA,NPC-like Neftel,tumor
OPC Allen,GABAergic PVALB neuron Allen,OPC HPA,tumor
OPC Allen,OPC CoDEx,OPC HPA,tumor
other
e
IDH1R132H status
electrophysiology profile
tumor SCRAM
developmental GSC Richards
OPC-like Neftel
AC-like Neftel
NPC-like Neftel
tumor
oligodendrocyte HPA
oligodendrocyte Allen
OPC CoDEx
OPC Aldinger
mid OPC Bhaduri
late OPC Bhaduri
oligodendrocyte/OPC
mid dividing Bhaduri
late dividing Bhaduri
early neuron Bhaduri
cycling progenitor G2M CoDEx
cycling progenitor S CoDEx
intermediate progenitor CoDEx
Purkinje cell Aldinger
granule cell precursor Aldinger
neurodevelopmental progenitor
GABAergic neuron HPA
GABAergic SST neuron Allen
GABAergic PVALB neuron Allen
GABAergic neuron Allen
GABAergic LAMP5 neuron Allen
GABAergic VIP neuron Allen
GABAergic SNCG neuron Allen
late interneuron Bhaduri
interneuron CGE CoDEx
GABAergic neuron
glutamatergic neuron HPA
glutamatergic neuron Allen
glutamatergic L2/3 IT neuron Allen
glutamatergic L5 IT neuron Allen
mid neuron Bhaduri
late neuron Bhaduri
glutamatergic neuron
R13 R14 R15 R22 R23 R31 R10 R18 R1 R21 R24 R25 R6 R7 R8 R2 R30 R3 R11 R12 R16 R9 R17 R19 R20 R26 R27 R28 R29 R4 R5
electrophysiology profile
pyramidal cell
interneuron
hybrid cell
glia
IDH1R132H Status
mutant
wild-type
no coverage
probability score
1
0.5
0

Figure 2. SCRAM reveals HC cells are GABA-OPC tumor cells. (a) Schematic of the SCRAM pipeline. Briefly, 11 scRNA-seq datasets are used to train cell type neural network models (NNMs). Each cell from scRNA-seq is then assigned cell type annotation independently of all other cells using NNM trained models. CNVs are added for each cell using CaSpER and NUMBAT. SNVs are added for each cell using XCAVTR. (b) Seurat clusters are shown for 234,880 cells from our in-house glioma scRNA-seq dataset. (c) Zoom-in of Seurat clusters 3, 12, 16 and 22 colored by Seurat clusters. (d) Zoom-in of Seurat clusters 3, 12, 16 and 22 coloured by SCRAM cell-by-cell annotations. (e) Heatmap showing SCRAM cell type probability scores for 31 Patch-seq cells.

Figure 3


a
SCRAM UMAP
b
chr 1p deletion
ΔIDH1 frequency
PDGFRA
c
non tumor
tumor
chr 7p amplification
ΔEGFR frequency
EGFR
d
SCRAM cell types
tumor,chr1Cp del
GABAergic PVALB neuron Allen,developmental GSC Richards NPC-like Neftel,tumor
GABAergic PVALB neuron Allen,developmental GSC Richards OPC HPA tumor
GABAergic PVALB neuron Allen,OPC HPA,NPC-like Neftel,tumor
GABAergic PVALB neuron Allen,OPC CoDEx,OPC HPA,tumor
OPC Aldinger,developmental GSC Richards,tumor
OPC Aldinger,OPC CoDEx,OPC HPA,tumor
OPC Aldinger,OPC Allen,OPC HPA,tumor
OPC CoDEx,developmental GSC Richards,NPC-like Neftel tumor
OPC CoDEx,OPC HPA,NPC-like Neftel,tumor
late OPC Bhaduri,OPC HPA,NPC-like Neftel,tumor
mid OPC Bhaduri,developmental GSC Richards,OPC HPA,tumor
mid OPC Bhaduri,NPC-like Neftel,tumor
mid OPC Bhaduri,OPC CoDEx,OPC HPA,tumor
OPC Allen,developmental GSC Richards,OPC HPA,NPC-like Neftel,tumor
OPC Allen,GABAergic PVALB neuron Allen,developmental GSC Richards,tumor
OPC Allen,GABAergic PVALB neuron Allen,OPC HPA,tumor
OPC Allen,OPC CoDEx,developmental GSC Richards,OPC HPA,tumor
OPC Allen,OPC CoDEx,NPC-like Neftel,tumor
OPC Allen,mid OPC Bhaduri,OPC HPA,tumor
other
UMAP 1
UMAP 2
e
IDH1 mutant
IDH1 wild-type
percent of total tumor cell #
cluster number
f
OPC CoDeX
OPC Aldinger
mid OPC Bhaduri
late OPC Bhaduri
OPC HPA
developmental GSC Richards
NPC-like Neftel
GABAergic PVALB neuron Allen
probability
g
tumor type
late OPC Bhaduri
OPC Allen
OPC Aldinger
GABAergic PVALB neuron Allen
mid OPC Bhaduri
OPC CoDEx
developmental GSC Richards
OPC HPA
NPC-like Neftel
ratio of tumor cells
h
percent GABAergic-OPC tumor
mutant
wild-type

425

Figure 3. GABA-OPC tumor cells in human glioma. (a) SCRAM UMAP of 234,880 scRNA-seq cells. (b) Commonly used IDH1^{mut} (top row) and IDH1^{WT} (bottom row) tumor features are shown. Top (from left to right): chromosome 1p deletion feature plot, *IDH1R132* mutation density plot, *PDGFRA* expression feature plot. Bottom (from left to right): chromosome 7p amplification feature plot, *EGFR* mutation density plot, *EGFR* expression feature plot. (c) SCRAM tumor and non-tumor cell annotation. (d) Zoom-in of inset from *(c)* showing SCRAM cell type annotations for each cell. (e) Bar graph showing the majority of cells in *(d)* are from IDH1^{mut} tumor patients. (f) SCRAM probability scores are shown for cell types of interest. (g) Heatmap of SCRAM cell type annotations for IDH1^{mut} (n=5) and IDH1^{WT} (n=7) glioma patients. (h) Bar graph showing the percentage of tumor cells with GABA-OPC annotations; *p*-value is noted in the figure.

Figure 4

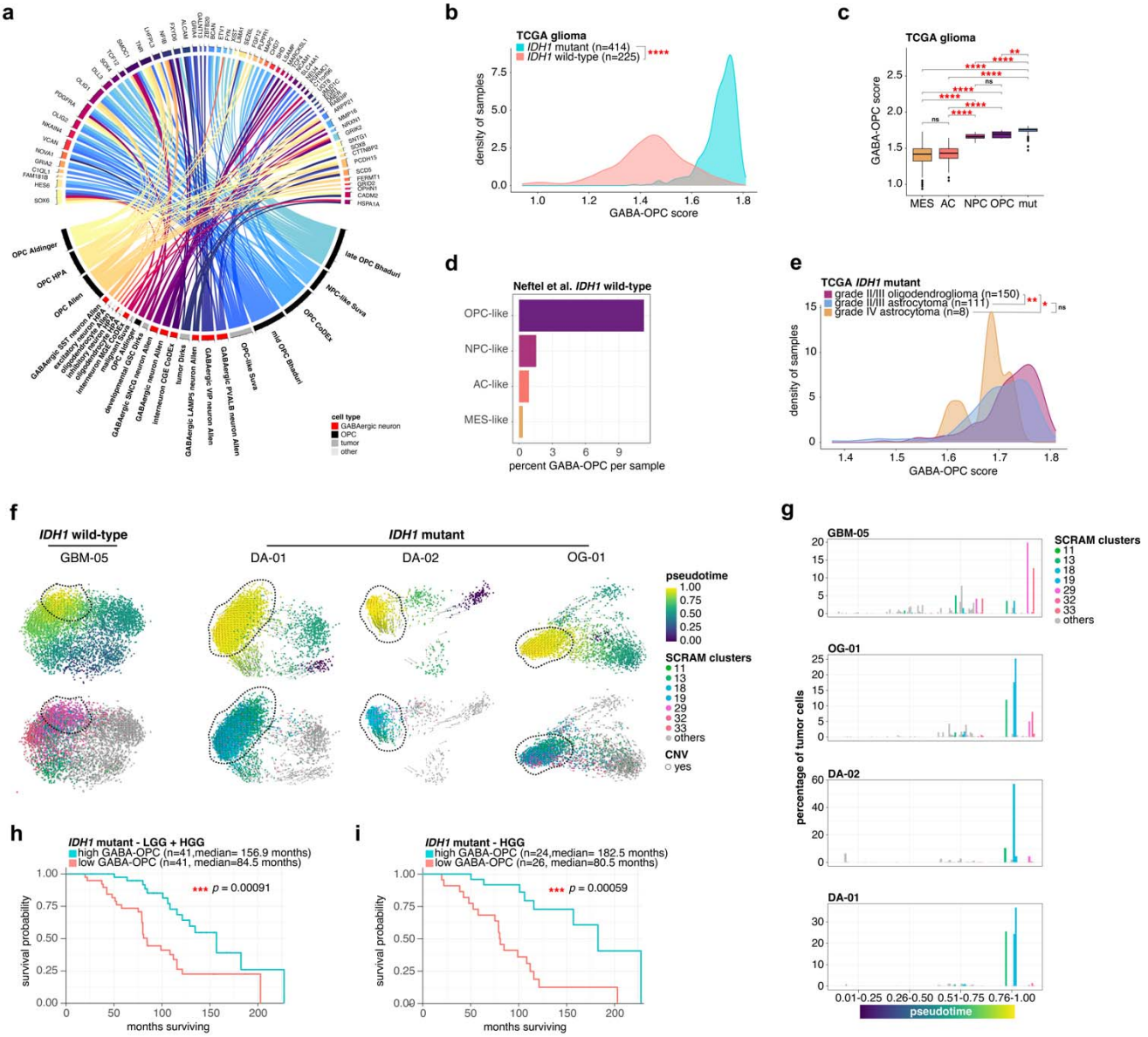

a
OPC Aldinger
OPC HPA
OPC Allen
late OPC Bhaduri
NPC-like Suva
OPC CoDEx
mid OPC Bhaduri
OPC-like Suva
cell type
GABAergic neuron
OPC
tumor
other
b
TCGA glioma
IDH1 mutant (n=414)
IDH1 wild-type (n=225)
density of samples
GABA-OPC score
c
TCGA glioma
GABA-OPC score
MES
AC
NPC
OPC
mut
d
Neftel et al. IDH1 wild-type
OPC-like
NPC-like
AC-like
MES-like
percent GABA-OPC per sample
e
TCGA IDH1 mutant
grade II/III oligodendroglioma (n=150)
grade II/III astrocytoma (n=111)
grade IV astrocytoma (n=8)
density of samples
GABA-OPC score
f
IDH1 wild-type
IDH1 mutant
GBM-05
DA-01
DA-02
OG-01
pseudotime
SCRAM clusters
CNV
yes
g
GBM-05
OG-01
DA-02
DA-01
SCRAM clusters
percentage of tumor cells
pseudotime
h
IDH1 mutant - LGG + HGG
high GABA-OPC (n=41,median= 156.9 months)
low GABA-OPC (n=41, median=84.5 months)
p = 0.00091
survival probability
months surviving
i
IDH1 mutant - HGG
high GABA-OPC (n=24,median= 182.5 months)
low GABA-OPC (n=26, median=80.5 months)
p = 0.00059
survival probability
months surviving

Figure 4. GABA-OPC tumor cell signatures are most highly expressed in IDH1^{mut} LGG and are reduced in both IDH1^{mut} and IDH1^{WT} HGG. (a) Circos plot showing the 61 genes ($log_2FC > 1$) comprising the GABA-OPC tumor cell signature. DEGs from GABA-OPC tumor cells versus all other tumor cells were extracted from our scRNA-seq dataset and crossmatched with SHAP genes from our trained NNMs. (b) Density plot of GABA-OPC tumor scores is shown for IDH1^{mut} (n=414) and IDH1^{WT} (n=225) TCGA bulk RNA-seq glioma samples; *p*-value is noted in the figure. (c) Box and whiskers plot showing GABA-OPC scores for TCGA samples by IDH1^{WT} tumor subtype: MES-like (n=149); AC-like (n=157); NPC-like (n=12); OPC-like (n=9); *p*-values for pairwise comparisons are noted in the figure. (d) Bar graph showing the percentage of tumor cells that are GABA-OPC tumor from Suva's IDH1^{WT} SMART-seq dataset. (e) Density plot of GABA-OPC tumor scores is shown for IDH1^{mut} TCGA bulk RNA-seq glioma samples by histology subtype: oligodendroglioma (n=150); astrocytoma (n=111); diffuse astrocytoma (n=8); *p*-values for pairwise comparisons are noted in the figure (f) RNAvelocity pseudotime analyses are shown for one IDH1^{WT} and three IDH1^{mut} samples. (g) Bar graphs showing percentage of tumor cells by cluster over pseudotime. (h) Kaplan-Meier survival analysis in IDH1^{mut} patient cohort. (i) Kaplan-Meier survival analysis in HGG IDH1^{mut} patient cohort.

Figure 5

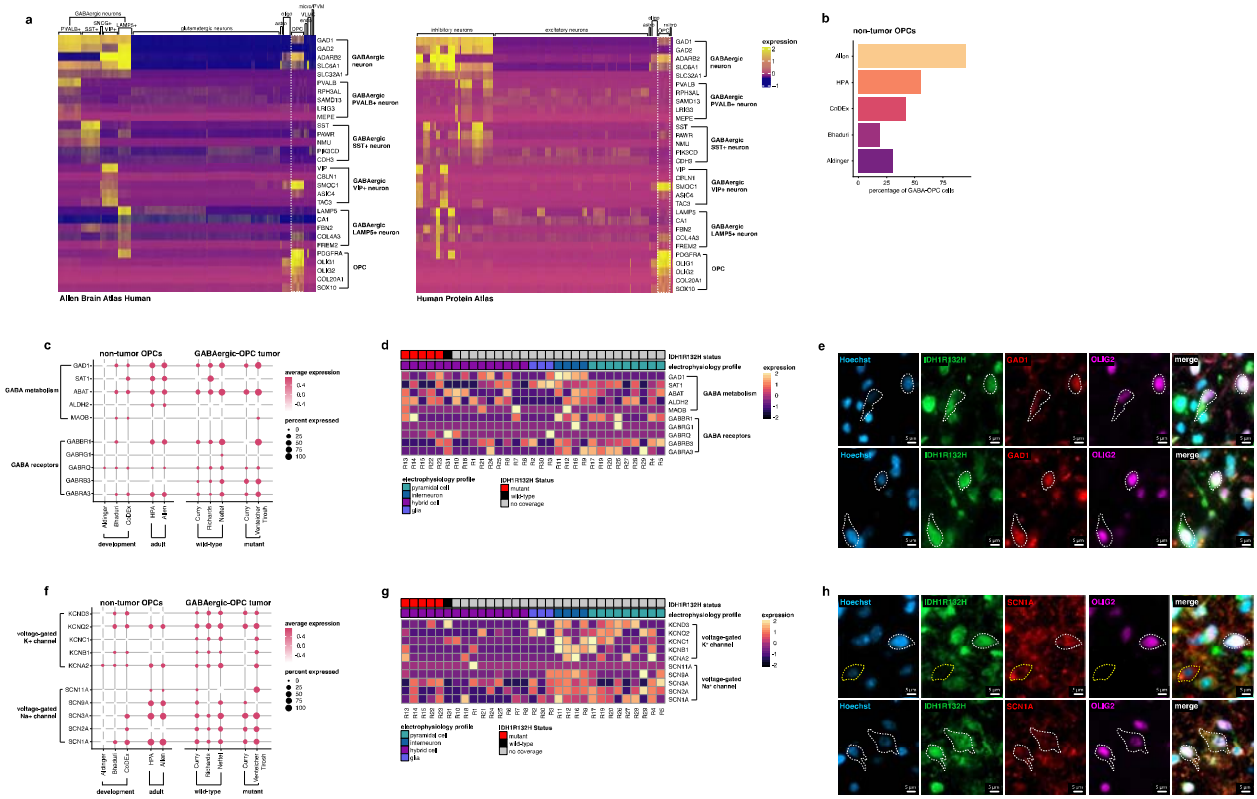

a
Allen Brain Atlas Human
Human Protein Atlas
b
non-tumor OPCs
percentage of GABA+OPC cells
c
non-tumor OPCs
GABAergic-OPC tumor
GABA metabolism
GABA receptors
development
adult
wild-type
mutant
d
IDH1R132H status
electrophysiology profile
GABA metabolism
GABA receptors
e
Hoechst
IDH1R132H
GAD1
OLIG2
merge
f
non-tumor OPCs
GABAergic-OPC tumor
voltage-gated K+ channel
voltage-gated Na+ channel
development
adult
wild-type
mutant
g
IDH1R132H status
electrophysiology profile
voltage-gated K+ channel
voltage-gated Na+ channel
h
Hoechst
IDH1R132H
SCN1A
OLIG2
merge

Figure 5. GABAergic neurons and OPCs share transcriptional profiles in non-tumor human brain (a) Heatmap showing subtype-specific GABAergic neuron markers as well as established OPC markers in Allen Brain Atlas and the Human Protein Atlas (HPA). White dotted box outlines OPCs. (b) Percentage of OPCs that express the GABA-OPC transcriptional signature in non-tumor brain atlases. (c) DotPlot showing the average expression of GABARs and GABA metabolism genes in human non-tumor OPCs and in GABA-OPC tumor cells. (d) Heatmap showing the expression GABARs and GABA metabolism genes in Patch-seq data. (e) IHC staining for GAD1 (red), IDH1R132H (green), OLIG2 (pink) and Hoechst (blue) in a human IDH1^{mut} tumor sample. (f) DotPlot showing the average expression of voltage-gated sodium channels (Na_vs) and voltage-gated potassium channels (K_vs) genes in human non-tumor OPCs and also in GABA-OPC tumor cells (g) Heatmap showing the Na_vs and K_vs genes in Patch-seq data. (h) IHC staining for Na_V1.1 (SCN1A; red), IDH1R132H (green), OLIG2 (pink) and Hoechst (blue) in a human IDH1^{mut} tumor sample.

Experimental Methods

Human data

Adult patients at St. Luke's Medical Center and Ben Taub General Hospital provided preoperative informed consent to participate in the study and gave consent under Institutional Review Board Protocol H35355. Patients included males and females. Clinical characteristics were maintained in a deidentified patient database and are summarized in **Table S4**.

Tumor samples were collected during surgery and immediately placed on ice. Tissue was divided for use in subsequent transcriptomic, histopathological, proteomic, 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.

Single-cell RNA-sequencing

Human 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 phosphate-buffered saline (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), according to the manufacturer's instructions. Live cells were collected through negative

selection using an 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% foetal bovine 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 droplets. GEMs were broken, and cDNA from each single cell was pooled. Clean-up was performed using Dynabeads MyOne Silane Beads (ThermoFisher, 37002D). Subsequently, the cDNA was 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 instrument.

Human tumor slice preparation

Fresh tumor samples were immediately placed into a cold (0−4°C) oxygenated N-methyl-d-glucamine (NMDG) solution (93 mM NMDG, 93 mM HCl, 2.5 mM KCl, 1.2 mM NaH_2PO_4, 30 mM $NaHCO_3$, 20 mM HEPES, 25 mM glucose, 5 mM sodium ascorbate, 2 mM thiourea, 3 mM sodium pyruvate, 10 mM $MgSO_4$, and 0.5 mM $CaCl_2$, 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 artificial cerebrospinal fluid (ACSF, 125 mM NaCl, 2.5 mM KCl, 1.25 nM NaH_2PO_4, 25

mM $NaHCO_3$, 1 mM $MgCl_2$, 25 mM glucose, and 2 mM $CaCl_2$, pH 7.4) for 1 hour before recording.

Single-cell processing

We ran samples on the 10X Chromium platform to produce next-generation sequencing libraries. We performed standard procedures for filtering, mitochondrial gene removal, and variable gene selection using the Seurat pipeline. The criteria for cell/gene inclusion were as follows: genes present in more than three cells were included, cells that expressed >300 genes were included, the number of genes detected in each cell was >200 and <5000, and the mitochondria ratio was 10. We integrated cells from different patients using the Harmony algorithm[47]. Next, we visualized clusters using a uniform manifold approximation and projection 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 the scvelo R package[48] and IntrExtract[49].

SCRAM pipeline and methodology

SCRAM input consisted of aligned scRNA-seq reads and our neural network model trained on 11 diverse single-cell RNA-Seq datasets encompassing 1 million cells of publicly available data from healthy adult and developing brain samples, as well as brain tumor samples. Tumor and normal cells were annotated independently for two reasons. (1) Significant overlap exists between tumor and non-tumor expression markers. For example, *EGFR* and *PDGFRA* are often used to denote tumor cells[9] but are also cell type markers for OPCs and ependymal host cells, respectively (2) We hypothesize that by separating tumor-specific and normal-specific features, we can

achieve more robust identification of hybrid cells (HCs). This hypothesis is supported by our observation that existing cell type assignment methods, which typically classify both tumor and normal cells together, fail to accurately characterize HCs. These tumor/normal features were systematically employed in our pipeline as follows:

Step 1. Annotation of non- tumor cells:

Training Neural Network Models (NNMs). We trained our neural network model (NNMs) on 11 diverse single-cell RNA-Seq datasets, which collectively contain 1 million cells. These datasets comprise publicly available data sourced from various datasets, including healthy adult and developing brain samples, as well as brain tumor samples[8–10,21–28]. We trained our model using a deep neural network (DNN), with an input layer of around 20K genes, three intermediate layers (256, 64, 32), and an output layer of size 16 or 21, depending on the number of referenced cell types. Following the dense connection within each hidden layer, there are batch normalization, activation, and dropout functions. We use the popular Rectified Linear Unit (ReLU) for hidden layer activation and set dropout rate to be 0.1. The output layer uses Softmax activation function so that each node outputs a non-negative value smaller than 1 and all the values sum up to 1. Therefore, each output corresponds to the probability of one cell type. We compile the model using categorical crossentropy as loss function, Adam as optimizer, and accuracy as metrics. In order to achieve a more balanced class distribution, we opted to subsample cell types within our training model. We train one neural network-based classifier on each developmental-like, normal, tumor cells datasets and save the model in repository. We predicted the developmental-like, normal and tumor cells, in our glioma scRNA-Seq data using our trained NNMs. Model probability scores >0.9 were used for final cell annotations. In building our NNMs, we

utilized the Python packages TensorFlow and Keras. Additionally, we used the Python Scipy package for processing the scRNA-Seq data. Prerequisite packages for data preprocessing and model training include Numpy 1.19.5, Pandas 1.1.5, Scanpy 1.7.2, Anndata 0.7.8, Scipy 1.5.4, and Scikit-learn 0.24.2.

In order for a single cell to be classified as "immune" within our framework, it needs be annotated as immune with a probability score above 0.9 in three or more trained datasets.

SHapley Additive exPlanations (SHAP) analysis.

SHAP analysis was employed to gain insights into the main features that affect the output of the NNMs, we used SHAP method, to explain how each feature affects the NNMs in inhouse glioma single-cell dataset. To perform the SHAP analysis, the model predictions were decomposed into contributions from individual features, allowing us to assess their impact on the final outcome using the *shap* python package.

Step 2. Annotation of tumor cells:

The SCRAM pipeline integrates multiple orthogonal tumor features to identify tumor cells at a single-cell resolution. These features include:

Module 1. Neural Network Model-Based Tumor Cell Prediction: Above explained NNMs is used to predict tumor cells based on the probability score. Cells with a probability score above 0.9 are classified as tumor cells.

Module 2. Marker based Expression Modeling: SCRAM employs finite Gaussian mixture modeling to model marker expression of three tumor marker genes: *SOX2, EGFR PDGFRA*. This approach helps to distinguish tumor cells based on their specific marker gene expression profiles (details explained below section "*Marker Expression Modeling for tumor annotation*").

Module 3. RNA-Inferred Genotyping of Chromosome Alterations: A modified version of our CNV-calling algorithm, CaSpER[19], another state of art CNV calling method numbat[20] is used in SCRAM to perform RNA-inferred genotyping of large-scale chromosome alterations.

Module 4. RNA-Inferred Mutational Profiling: SCRAM utilizes our XCVATR[18] tool, a recently developed tool, to deduce rare deleterious single-nucleotide variants (SNVs) present in the tumor cells. This analysis involves considering SNVs that are reported in the COSMIC[50]- database and have a frequency of less than 0.1% in the dbSNP[51]database.

These orthogonal tumor features are called separately in the SCRAM pipeline. By combining these different approaches, SCRAM aims to accurately identify tumor cells at a single-cell resolution.

In order for a single cell to be classified as "tumor" within our framework, it needs to meet <u>two or more</u> of the following criteria:

1. Neural Network Model-Based Tumor Cell Prediction (Module 1): The cell is annotated as "tumor" when it receives a probability score greater than 0.9 from the trained Neural Network Models (NNMs) in Richards et al.[10], or Venteicher et al.[22] or Neftel et al.[9] or Tirosh et al.[8] datasets.
2. Marker Expression Modeling (Module 2): The expression levels of at least two tumor cell markers, (*SOX2*, *EGFR*, or *PDGFRA*) should surpass a predetermined threshold. This threshold is established using finite Gaussian mixture modeling (details explained below section "*Marker Expression Modeling for tumor annotation*").

3. RNA-Inferred Genotyping of Chromosome Alterations (Module 3): The presence of large-scale copy number variations (CNVs) is considered a tumor cell.
4. RNA-Inferred Mutational Profiling (Module 4): Tumor cells that have SNVs in genes *IDH1* (R132H/R132C) or EGFR.

<u>*Module 2. Marker Expression Modeling for tumor annotation:*</u> Given the expressional heterogeneity of tumor markers in normal cells, we used previously published tumor and non-tumor cell datasets to establish a marker expression–based tumor classification model (i.e., thresholding requirements for "high expression" annotation) for the tumor markers *PDGFRA, EGFR,* and *SOX2*. For each tumor marker, an independent classifier model was built using (1) Allen Brain Atlas human scRNA-seq data, which represent the largest compendium of healthy brain data as a training set for normal cells; and (2) a compendium of publicly available brain tumor scRNA-seq datasets as a training set for tumor cells[9]. The following statistical models were used to infer the class (normal vs tumor) of our in-house tumor scRNA-seq data.

We modelled expression as a mixture of Gaussian distributions to identify and classify normal and tumor cells:

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

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

where $z = \{\pi_{1,\ldots,}\pi_G, \theta_1, \ldots \theta_G\}$ represents the parameters of the mixture model and $f_k\,(x_i;\ \theta_k)$ is the kth component density, which is assumed to follow a 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 that sum to 1, known as the mixing proportion. The mixing proportion, π, follows a multinomial distribution.

We used the above model to predict normal vs tumor class in our in-house glioma cells. For each gene, j, z parameters were estimated by maximizing the log-likelihood function via the expectation-maximization 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, we generated a matrix, with genes indicated by rows and cells indicated by columns, and the cell value index was 1 if the cell had a high "tumor class" probability for the corresponding gene. A cell was classified as "tumor" if at least two markers had high "tumor class" probabilities. We used the mclust R package for Gaussian mixture model implementation[52].

<u>Module 3. RNA-Inferred Genotyping of Chromosome Alterations:</u> CNVs are hallmark features of tumor cells that can be used to classify tumor vs non-tumor cells with or without expression markers. However, CNV detection from scRNA-seq data is inherently noisy due to dropouts and unmatched control sets, among other factors, requiring a set of known tumor cells. To estimate a "clean" set of CNV calls to 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 used our marker expression–based and NNMs models from Module 1 and Module

2 to identify tumor cells. Cells assigned as “tumor” cells using Module 1 and 2 were treated as a pure tumor cell cohort. Cells assigned as “immune” cells using our NNMs are considered control cells.

CNV calling of patient-specific pure pseudobulk samples: We hypothesized that the pseudobulk sample contained representative sets of CNVs with high probability and, therefore, should be useful to identify a clean CNV call set. CNV calling of the pseudobulk samples from each patient was performed using our CNV-calling algorithm, CaSpER. CaSpER CNV calls were used as the ground truth, large-scale CNV calls for each patient.

Genotyping of CNVs of all cells: After CNVs were identified from the pseudobulk sample, we genotyped the CNVs in all cells and generated a binary matrix that represents the existence of CNVs in cells, i.e., $CNV_{i,j}$.

Module 4. RNA-Inferred Mutational Profiling: We performed RNA-inferred rare deleterious (COSMIC[50]-reported and dbSNP[51], <0.1% frequency) mutational profiling via our recently developed XCVATR[18] tool. We detected mutations in *IDH1 (R132H/R132C)*, *EGFR* and annotated cells harbouring these mutations as tumor cells.

Visualization and clustering of single-cells

We used the probability score output from NNMs instead of relying solely on the most variable genes for clustering and visualization of our in-house single-cell data. This, named as SCRAM UMAP, involved applying UMAP and clustering techniques to the model probability scores using the Seurat package's runUMAP, FindNeighbors, and FindClusters functions. Additionally, we employed the most variable genes for cell data clustering and visualization, referring to this UMAP representation as the original/Seurat UMAP.

Summarizing co-occurring cell types using maximally frequent gene set identification

We summarized co-occurring cell types using a frequent itemset rule mining approach. CNV and SNV calls were 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, oligodendrocyte precursor cell, chr1p_deletion, chr19q_deletion + *IDH1:2:208248389* mutation". We used the tumor and normal cell assignments of Step 1 and Step 2 to integrate co-occurring tumor and normal 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 *a priori* algorithm is an efficient implementation for finding maximally frequent sets with support above a given threshold. In the *a priori* algorithm, the minimum *support* threshold is set to min (50, number_cells_in cluster*0.1), and the maximum number of genes in a gene set is set to 50. Using the *a priori* 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 at multiple levels: 1) cell type level (example output of this step: *tumor* AND *radial glia* AND *astrocyte*); 2) cell class level (example output of this step: *tumor* AND *neural cells* are commonly upregulated).

<u>Maximally frequent cell type (or cell class) co-occurrence analysis:</u> Within each cluster, *m,* we calculated the maximally frequent cell types (or cell lineage or cell class) using the *a priori* algorithm. The input was the binarized matrix E^m, where the cell types (or cell lineage or cell class) were the rows, and the cells in cluster *m* were the columns.

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

Analyzing bulk expression data and survival analysis

TCGA-GBM (high grade glioma), TCGA-LGG (low grade glioma) raw read counts and accompanying clinical data are downloaded using TCGAbiolinks R package[53]. TCGA-GBM, TCGA-LGG and our bulk RNA-Seq data of the IDH Mutant cohort were both normalized and applied variance stabilizing transformation using the DESeq2 package[54]. Single sample gene set enrichment analysis (ssGSEA) was performed using GSVA R package. We used our GABA-OPC tumor gene sets and also the MES-like, AC-like, NPC-like and OPC-like gene sets reported in a previous study[9]. SsGSEA GABA-OPC scores were split by median to assign high-OPC-GABA and low-OPC-GABA scored samples. Those groups are compared against "overall survival" in a Cox Proportional Hazards (Cox) survival model. used in survival analysis and compared using a Log-rank test P-value. We used survminer and survival R package for the survival analysis.

Patch-seq recording procedures

Electrophysiological, morphological, and transcriptomic data from the same cell were obtained simultaneously using the Patch-seq protocol described previously[11,55]. Briefly, patch pipettes (5−7 MΩ) were filled with RNase-free intracellular solution (111 mM potassium gluconate, 4 mM KCl, 10 mM HEPES, 0.2 mM EGTA, 4 mM MgATP, 0.3 mM Na_3GTP, 5 mM sodium phosphocreatine, and 13.4 mM biocytin). Whole-cell recordings were performed using a Quadro EPC 10 amplifier (HEKA Electronic). After 5–10 minutes of whole-cell recording of firing patterns, the nucleus was 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[55]. RNA in the lysis buffer was denatured, reverse transcribed, amplified, and purified following the Smart-seq2-based protocol[56]. 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 a NovaSeq 6000 instrument using 150-bp paired-end reads.

Biocytin staining and morphological reconstruction

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

Histology

Human samples were retrieved from the operating room on ice and then 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.

Immunohistochemistry

For immunohistology, 10 μm paraffin-embedded human glioma 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) when needed. Sections were blocked for 1 hour at room temperature and kept in primary antibody

incubation overnight at 4°C. The following primary antibodies were used: mouse anti-IDH1R132H (1:50; Dianova, DIA-H09), rabbit anti-OLIG2 (1:100; R&D, AF4788), rabbit anti-GAD1 (1:200; Synaptic Systems, 198013), rabbit anti-SCN1A ($Na_V1.1$) Antibody (1:200; Alomone Labs, ASC-001). Species-specific secondary antibodies tagged with Alexa Fluor corresponding to emission spectra 488 nm, 568 nm, or 647 nm (1:1,000, ThermoFisher) were used for immunofluorescence and Hoechst nuclear counter staining (1:50,000; ThermoFisher, H3570) was performed before coverslipping with Vectashield antifade mounting medium (Vector Laboratories, H-1000).

Patch-seq data processing

The Patch-seq reads were mapped using STAR[57] to hg38 assemblies for humans. Read count matrices were generated using FeatureCounts[58] with the latest gene annotations from GENCODE[59] consortia. DEGs and transcripts were identified using DESeq2[54] and limma[60]. Cells were clustered and visualized using PCA methods. Cell type enrichment analyses were performed with enrichR[61] using the *PanglaoDB_Augmented_2021* and *CellMarker_Augmented_2021* cell type marker sets. *IDH1* mutations were identified using our variant detection tool, XCVATR[18], and visually confirmed using the Integrative Genomics Viewer[62].

Statistical analysis

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. Significant differences are denoted by asterisks in associated graphs. Data are presented as the mean±standard

error of the mean. 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. Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
2. Yan, H. *et al. IDH1*and*IDH2*mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
3. Verhaak, R. G. W. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
4. Charles, N. A., Holland, E. C., Gilbertson, R., Glass, R. & Kettenmann, H. The brain tumor microenvironment. *Glia* **59**, 1169–1180 (2011).
5. Venkatesh, H. S. *et al.* Neuronal activity promotes glioma growth through neuroligin-3 secretion. *Cell* **161**, 803–816 (2015).
6. Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539–545 (2019).
7. Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573**, 532–538 (2019).
8. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
9. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

10. 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).
11. Cadwell, C. R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).
12. Lipovsek, M. *et al.* Patch-seq: Past, present, and future. *J. Neurosci.* **41**, 937–946 (2021).
13. Bai, X., Kirchhoff, F. & Scheller, A. Oligodendroglial GABAergic signaling: More than inhibition! *Neurosci. Bull.* **37**, 1039–1050 (2021).
14. Luyt, K. *et al.* Developing oligodendrocytes express functional GABA(B) receptors that stimulate cell proliferation and migration. *J. Neurochem.* **100**, 822–840 (2007).
15. Zhang, X. *et al.* NG2 glia-derived GABA release tunes inhibitory synapses and contributes to stress-induced anxiety. *Nat. Commun.* **12**, 5740 (2021).
16. Jiang, X. *et al.* Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* **350**, aac9462 (2015).
17. Scala, F. *et al.* Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature* **598**, 144–150 (2021).
18. 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.
19. 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).

20. Gao, T. *et al.* Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* **41**, 417–426 (2023).
21. Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
22. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, (2017).
23. Bhaduri, A. *et al.* An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200–204 (2021).
24. Polioudakis, D. *et al.* A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron* **103**, 785-801.e8 (2019).
25. Pombo Antunes, A. R. *et al.* Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization. *Nat. Neurosci.* **24**, 595–610 (2021).
26. Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
27. Aldinger, K. A. *et al.* Spatial and cell type transcriptional landscape of human cerebellar development. *Nat. Neurosci.* **24**, 1163–1175 (2021).
28. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, (2020).
29. Curry, R. N. *et al.* Glioma epileptiform activity and progression are driven by IGSF3-mediated potassium dysregulation. *Neuron* **111**, 682-695.e9 (2023).

30. Lee, S. *et al.* Role of CX3CR1 signaling in malignant transformation of gliomas. *Neuro. Oncol.* **22**, 1463–1473 (2020).
31. Boshans, L. L. *et al.* Direct reprogramming of oligodendrocyte precursor cells into GABAergic inhibitory neurons by a single homeodomain transcription factor Dlx2. *Sci. Rep.* **11**, 3552 (2021).
32. Serrano-Regal, M. P. *et al.* Expression and function of GABA receptors in myelinating cells. *Front. Cell. Neurosci.* **14**, 256 (2020).
33. Lin, S.-C. & Bergles, D. E. Synaptic signaling between GABAergic interneurons and oligodendrocyte precursor cells in the hippocampus. *Nat. Neurosci.* **7**, 24–32 (2004).
34. Káradóttir, R., Hamilton, N. B., Bakiri, Y. & Attwell, D. Spiking and nonspiking classes of oligodendrocyte precursor glia in CNS white matter. *Nat. Neurosci.* **11**, 450–456 (2008).
35. Ge, W.-P. *et al.* Long-term potentiation of neuron-glia synapses mediated by Ca2+-permeable AMPA receptors. *Science* **312**, 1533–1537 (2006).
36. Yu, K. *et al.* PIK3CA variants selectively initiate brain hyperactivity during gliomagenesis. *Nature* **578**, 166–171 (2020).
37. John Lin, C.-C. *et al.* Identification of diverse astrocyte populations and their malignant analogs. *Nat. Neurosci.* **20**, 396–405 (2017).
38. Barron, T. *et al.* GABAergic neuron-to-glioma synapses in diffuse midline gliomas. (2022) doi:10.1101/2022.11.08.515720.
39. Labrakakis, C. *et al.* Action potential-generating cells in human glioblastomas. *J. Neuropathol. Exp. Neurol.* **56**, 243–254 (1997).

40. Patt, S. *et al.* Neuron-like physiological properties of cells from human oligodendroglial tumors. *Neuroscience* **71**, 601–611 (1996).

41. Labrakakis, C., Patt, S., Hartmann, J. & Kettenmann, H. Glutamate receptor activation can trigger electrical activity in human glioma cells. *Eur. J. Neurosci.* **10**, 2153–2162 (1998).

42. Galvao, R. P. *et al.* Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4214-23 (2014).

43. Liu, C. *et al.* Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell* **146**, 209–221 (2011).

44. Monje, M. *et al.* Hedgehog-responsive candidate cell of origin for diffuse intrinsic pontine glioma. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4453–4458 (2011).

45. Geha, S. *et al.* NG2+/Olig2+ cells are the major cycle-related cell population of the adult human normal brain. *Brain Pathol.* **20**, 399–411 (2010).

46. Sigel, E. & Steinmann, M. E. Structure, function, and modulation of GABA(A) receptors. *J. Biol. Chem.* **287**, 40224–40231 (2012).

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

48. 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).

49. Harmanci, A. S. *et al.* scRegulocity: Detection of local RNA velocity patterns in embeddings of single cell RNA-Seq data. *bioRxiv* (2021) doi:10.1101/2021.06.01.446674.

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

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

52. 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).

53. Mounir, M. *et al.* Analyses of cancer data in the Genomic Data Commons Data Portal with new functionalities in the TCGAbiolinks R/Bioconductor package. *bioRxiv* (2018) doi:10.1101/350439.

54. 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).

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

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

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

58. 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).

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

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

61. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-7 (2016).

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