Title

Deconvolving cell type and pathway heterogeneity in disorders of the central nervous system: intertumoral and spatial heterogeneity in glioma

Author

Tessema, Kaleab

Publication Date

2021

UNIVERSITY OF CALIFORNIA

Los Angeles

Deconvolving cell type and pathway heterogeneity in disorders of the central nervous system: intertumoral and spatial heterogeneity in glioma

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy in Molecular Biology

by

Kaleab Tessema

2021

ABSTRACT OF THE DISSERTATION

Deconvolving cell type and pathway heterogeneity in disorders of the central nervous system: intertumoral and spatial heterogeneity in glioma

by

Kaleab Tessema
Doctor of Philosophy in Molecular Biology
University of California, Los Angeles, 2021
Professor Harley I. Kornblum, Chair

Glioma is the most common primary brain malignancy and remains a challenging and intractable disease. Two major barriers inhibiting progress in therapeutic development are the unclear clinical implications of molecular differences across patients and inevitable tumor recurrence likely due to the difficulty in removing the entire malignant compartment during the initial resection surgery. Here, I present a strategy to begin combatting these challenges by leveraging flexible and scalable sequencing analysis pipelines to deconvolve complex glioblastoma cellularity and discover functional drivers of disease progression and recurrence. I first describe an approach that aims to better characterize therapeutically relevant patterns of intertumoral glioblastoma heterogeneity by classifying tumors based on functional pathway enrichment rather than gene expression alone, which identified E2F1 as a key proliferation driver specifically in E2F1-activated tumors. We also found a biological parallel of one of our candidate gene lists, as it correlated significantly with markers of endothelial, mural, and mesenchymal-like glioma cells. These findings suggest that pathway enrichment is a more functionally relevant grouping metric and that we can use this approach to identify therapeutic targets in specific patient subsets. I then describe an approach aimed at addressing the challenge of recurrence by characterizing the cellular and molecular differences between the core and edge regions of recurrent glioma at single cell resolution. Using MRI-guided biopsy target selection and a multifaceted snRNA-seq analysis pipeline, we thoroughly profiled both malignant and non-malignant cell types in the edge region and identified a functional “state switching” event in mesenchymal-like malignant cells, with vascular remodeling consistently enriched in the tumor core and proliferation consistently upregulated in the tumor edge. Robust sequencing analysis pipelines thus represent an effective approach to identifying key functional drivers in multiple contexts and demonstrate the potential to identify targetable candidate cells and pathways in glioma.

The dissertation of Kaleab Tessema is approved.

Peyman Golshani

Janel Le Belle

Jeffrey A. Long

Alcino J. Silva

Harley I. Kornblum, Committee Chair

University of California, Los Angeles

2021

For my parents, whose unending love and sacrifice through
the hardest of times made me who I am today

Table of contents

List of figures

Acknowledgements

I would like to thank my advisor, Dr. Harley Kornblum, and all members of the Kornblum Lab, past and present, for their outstanding support and guidance throughout my graduate studies. In this wonderful training environment, I felt comfortable to freely explore and pursue my scholarly interests, which I don't take for granted. I also appreciate the mentorship and advice offered by my thesis committee and many of our collaborating scientists, especially Dr. Riki Kawaguchi and Dr. Vivek Swarup. Finally, I would like to thank my family, friends, colleagues, previous mentors, educators, and the wonderful UCLA administrative staff, all of whom played an important part in helping me grow as a scientist and student of life.

Chapter 2 is a version of: Alvarado, A. G., Tessema, K., Muthukrishnan, S. D., Sober, M., Kawaguchi, R., Bhaduri, A., Swarup, V., Nathanson, D. A., Goldman, S. A., & Kornblum, H. I. (2021). Pathway-based approach reveals differential sensitivity of glioblastoma to E2F1 inhibition. BioRxiv, 2021.06.19.449118. It is accessible with the DOI: 10.1101/2021.06.19.449118. A.G.A, S.M and M.S conducted in vitro and in vivo experiments. A.G.A, K.T, R.K, and V.S performed bioinformatics analysis. A.B and D.A.N contributed resources and reagents. A.B, V.S, D.A.N, and S.A.G provided valuable inputs in experimental design and data analysis. A.G.A and H.I.K conceptualized the study, designed the experiments, and wrote the manuscript. All authors read, revised, and approved the final manuscript.

Chapter 3 is unpublished work that is in preparation for publication and was performed by Kaleab Tessema, Kunal Patel, Riki Kawaguchi, Alvaro Alvarado, Deepthi Muthukrishnan, Akifumi Hagiwara, Vivek Swarup, Linda Liau, Anthony Wang, William Yong, Daniel H. Geschwind, Ichiro Nakano, Steven A. Goldman, Richard Everson, Benjamin Ellingson, and Harley I. Kornblum. KT and KP performed sample preparation for sequencing. KT, KP, RK, VS, and DHG performed bioinformatics analysis. KP, RE, and BE designed and performed the imaging-related procedures and biopsy collection. SAG and IN contributed to experimental planning. All authors provided valuable inputs regarding study design and/or data analysis.

The reported work was supported by the NIH Ruth L. Kirschstein National Research Service Award F31 MH122205 (KT), UCLA/Caltech MSTP T32 GM008042 (KT), UCLA Molecular Biology Institute Whitcome Fellowship (KT), UC President's Postdoctoral Fellowship (AGA), UCLA Broad Stem Cell Research Center Postdoctoral Fellowship (KP), Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (HIK), and UCLA SPORE in Brain Cancer P50 CA211015 (HIK).

Biographical Sketch

Education

Yale University (New Haven, CT) | 2010 – 2014

- Bachelor of Science in Biomedical Engineering with distinction

Duke Study in China (University of International Business & Economics; Beijing, China) | 2013

- Completed third-year level of Mandarin Chinese during 10-week summer program

University of Maine (Orono, ME) | 2012

- Summer research and coursework focused on sensor technologies

Richland College (Dallas, TX) | 2008 – 2009

- Dual-credit courses during high school

Honors and awards

2020: National Institutes of Health Ruch L. Kirschstein NRSA Individual Fellowship F31 (NIMH)

2019, 2020: Whitcome Pre-doctoral Fellowship in Molecular Biology (UCLA)

2015: National Institutes of Health MSTP Institutional Training Grant T32 (NIGMS)

2014: Davie Award for outstanding contributions to quality of life in Timothy Dwight College (Yale)

2013: Edward Allen Colby Scholarship (Yale)

2013: Light Fellowship/Greenberg Scholarship for East Asian Language Study (Yale, Duke)

2012: NSF Summer Research Fellowship in Sensor Science and Engineering (University of Maine)

2011: William J. Morden Scholarship (Yale)

2011: STARS Summer Research Fellowship (Yale)

2011, 2012, 2013: AACcolades Outstanding Academic Achievement Award (Yale)

2010: National Achievement Boule Foundation Scholarship

Publications

1) Alvarado, A. G., **Tessema, K.**, Muthukrishnan, S. D., Sober, M., Kawaguchi, R., Bhaduri, A., Swarup, V., Nathanson, D. A., Goldman, S. A., & Kornblum, H. I. (2021). Pathway-based approach reveals differential sensitivity of glioblastoma to E2F1 inhibition. BioRxiv, 2021.06.19.449118.

2) Hernandez C, Wang Z, Ramazanov B, Tang Y, Mehta S, Dambrot C, Lee YW, **Tessema K**, Kumar I, Astudillo M, Neubert TA, Guo S, Ivanova NB. Dppa2/4 Facilitate Epigenetic Remodeling during Reprogramming to Pluripotency. Cell Stem Cell. 2018 Sep 6;23(3):396-411.e8. PubMed PMID: 30146411; PMCID: PMC6128737.

Presentations

1) V. Gudapati, S. Meyer, A. Chen, **K. Tessema**, A. Abiri, Y. Ding, C. J. Kuo, M. Wang, T. K. Hsiai. Machine-Learning Enabled Interactive Virtual Reality Simulator for Preoperative Planning of Endoscopic Sinus and Skull-Based Surgery. Biomedical Engineering Society 2020 Annual Meeting; 2020 October 14-17; San Diego, CA. (Poster)

2) Roya Susan Moheimani, M.D., Sara Arastoo, B.S., Julian Landaw, B.S., **Kaleab Tessema, B.S.**, Rasheed Majzoub, & Agnes Wallbom, M.D. 8-Man football as an injury risk reduction strategy. Annual Assembly of the American Academy of Physical Medicine and Rehabilitation; 2019 November 14-17; San Antonio, TX. (Poster)

3) **K Tessema**, L Elahi, R Gau, JA Martinez-Agosto, JE LeBelle, H Kornblum. Examining a novel gene-environment interaction in autism spectrum disorder. Society for Neuroscience Annual Meeting; 2019 October 19-23; Chicago, IL. (Poster)

4) **K. Tessema**, L. Elahi, R. Gau, J. A. Martinez-Agosto, J. LeBelle, H. Kornblum. Understanding genetic and environmental regulation of mTOR signaling in autism spectrum disorder. MSTP 2019 Annual Research Conference; 2019 September 13; Los Angeles, CA. (Poster)

5) Muthukrishnan SD, Condro M, **Tessema K**, Johnson M, Kornblum H. TMIC-47. GLIOMA STEM CELL-VASCULAR INTERFACE IN GLIOBLASTOMA PROGRESSION, Neuro-Oncology, Volume 20, Issue suppl_6, November 2018, Page vi266. (Poster)

6) Hernandez C, Dambrot C, **Tessema K**, Zhang A, Ziller M, Meissner A, Ivanova N. Chromatin-associated factor Dppa2 regulates pluripotent stem cell differentiation and cellular reprogramming. Annual Genetics Department Minisymposium; 2014 Sep 2; New Haven, CT. (Poster)

Chapter 1: Introduction

“The most complex object in the universe” is a label to which the human brain has complete bragging rights. Along with the spinal cord, the brain is the primary component of the mammalian central nervous system (CNS), which manages information processing and controls the activities of the rest of the body. Inside the brain lies a complex cellular landscape composed of many cell types and subtypes exhibiting distinct functions and heterogeneous molecular pathway utilization profiles. Key players in this environment include neurons, astrocytes, oligodendrocytes, vascular cells, immune cells, and progenitor cells giving rise to neurons and glia. There is a fine balance among these brain cell types and their respective functions that ideally achieves an equilibrium state of normal CNS function. However, when there is dysregulation in one or more cell types or subtypes, disorders of the CNS may arise. To make matters worse, many of these CNS disorders are driven by complex pathophysiological patterns that involve compound cellular phenotypes, with dysregulation in some cells causing dysregulation in others. This presents a significant challenge for the research community, as it is difficult to pinpoint specific functional alterations in the cell types that are important for disease progression.

Recent advances in methods to analyze tissues at the level of single cells, such as single cell and single nucleus RNA-sequencing (sc/snRNA-seq), have drastically improved our ability to understand these complex cell-cell and cell-function relationships in the CNS and other tissues. Over the past decade, there has been a rapid increase in the number of sequencing studies performed, likely owing to the technology’s ever-increasing accessibility, both in terms of sample processing and data analysis. As a result, there are many publicly available databases containing sc/snRNA-seq datasets from different areas of the brain (**Fig. 1-1**) [1], which presents a unique opportunity to capitalize on this rapid technological evolution in ways that allow comprehensive molecular analyses of brain tissue at the level of single cells. While sc/snRNA-seq suffers from several limitations related to sequencing depth, gene drop-out, and feasibility, the ability to examine composite tissue at an unprecedented resolution is justified in many (though not all) contexts.

As an engineer by training and an aspiring physician-scientist, I was keen to harness this opportunity to leverage readily available big data in the hope of discovering important underlying biology in clinically relevant contexts that may help alleviate patient suffering. After initially focusing my studies on autism spectrum disorder and maternal inflammation, I was drawn to several ongoing investigations in our research group that focused on glioblastoma, as these studies presented an opportunity to directly examine tumor samples recently resected from patients. Glioblastoma is a highly intractable and aggressive cancer of the brain with a median overall survival of 12-15 months despite resection surgery followed by radiation and temozolomide therapy [2]. While decades of intensive research efforts have identified several key cellular and molecular features that characterize inter- and intra-tumoral heterogeneity, glioblastoma remains the deadliest CNS cancer with no clear avenue for clinical improvement. Two major barriers inhibiting progress in glioblastoma management are the unclear clinical implications of molecular differences across patients and the inevitability of tumor recurrence likely due to the difficulty in removing all of the malignant cells during the initial resection surgery. In terms of molecular intertumoral heterogeneity, glioblastoma molecular subtypes have been defined to classify individual tumors, but this has largely not improved our ability to develop targeted therapies for specific patient groups and improve survival. With respect to malignant cells left behind after resection, it remains unclear how these individual malignant cells at the periphery of the tumor might differ from those at the core, as current literature comprises bulk tissue studies or analyses with low numbers of cells from the periphery.

The complex cellular architecture within and surrounding glioblastomas is a strong contributor to both of these challenges and is thus critical to our research efforts that aim to address these issues. This presents an ideal opportunity to leverage sc/snRNA-seq to gain a deeper understanding of the cellular and molecular complexity of glioblastoma tumors and potentially disambiguate both the unclear role of molecular subtypes and the unknown functional profiles of malignant cells at the tumor periphery. However, current sc/snRNA-seq workflows in tumor biology research suffer from several technical challenges that limit robustness, as workflows often involve 1) manual intervention by users with inevitably varying skill sets and experience levels, 2) arbitrary decisions during parameter selection, and 3) inconsistent guidelines across expert groups.

Here, I present a strategy to address these biological and technical challenges by building unbiased, flexible, and scalable sequencing analysis pipelines to deconvolve complex glioblastoma cellularity and discover functional drivers of multiple facets of disease progression. In Chapter 2, I describe an approach that aims to better characterize functionally relevant patterns of intertumoral glioblastoma heterogeneity. We hypothesized that classifying tumors based on functional pathway enrichment would yield distinct groups with more functional implications than groups generated based on patterns of gene expression alone, which is implemented in the traditional subtyping method [3]. Our approach identified E2F1 as a key driver of tumor cell proliferation specifically in an E2F1-enriched cluster of tumors and a strong correlation between one of our candidate gene lists and markers of endothelial cells, mural cells, and mesenchymal-like glioma cells. These findings suggest that pathway enrichment is a more functionally relevant grouping metric and that we can use this approach to identify therapeutic targets in specific patient subsets. This work was performed primarily in collaboration with Dr. Alvaro Alvarado, a postdoctoral researcher in our lab. My contributions were the bulk gene list correlation analysis in **Supplementary Fig. 2-4**, the scRNA-seq analysis evaluating candidate gene list correlation with cell cycle and stemness signatures in **Supplementary Fig. 2-6**, the scRNA-seq analysis of pairwise correlation among all candidate gene lists, malignant cellular states, and adult/developing brain cell type markers in **Supplementary Fig. 2-7**, and revisions of the text and figures. In Chapter 3, I characterize the cellular and molecular differences between the contrast-enhancing core and non-enhancing edge region of recurrent glioma at single cell resolution using MRI-guided biopsy target selection followed by snRNA-seq analysis. We hypothesized that one or multiple cell types exhibit putative recurrence-related functional enrichment patterns in the edge region of glioma. Using a multidimensional analytical approach, I thoroughly characterized both malignant and non-malignant cell types in the edge region and identified a potential functional “state switching” event in MES-like cells, with vascular remodeling enriched in the tumor core and proliferation upregulated in the tumor edge, potentially contributing to tumor recurrence. This work was performed primarily in collaboration with Dr. Kunal Patel (UCLA Neurosurgery). My contributions were isolating single nuclei for sequencing from 3 of the tumors and performing all of the snRNA-seq data pre-processing and analysis reported in the figures (i.e., all figure panels except MRI in **Fig. 3-1a**, image modeling in **Fig. 3-4c-d**, IHC in **Fig. 3-4g-i** and **Supp. Fig. 3-4b**).

Through the strategies described in the following chapters, I hope to provide convincing evidence that approaches using unbiased, reproducible, creative, and efficient analytical workflows are significantly more robust in identifying key biological insights with high confidence, in this case revealing multiple critical drivers of several hallmark glioma features, most notably proliferation, vascular remodeling, and invasive potential. These findings present opportunities for future investigation to validate several novel pathways, cellular populations, and cell-cell interactions that serve as potential candidates for glioma therapeutic development.

Figures

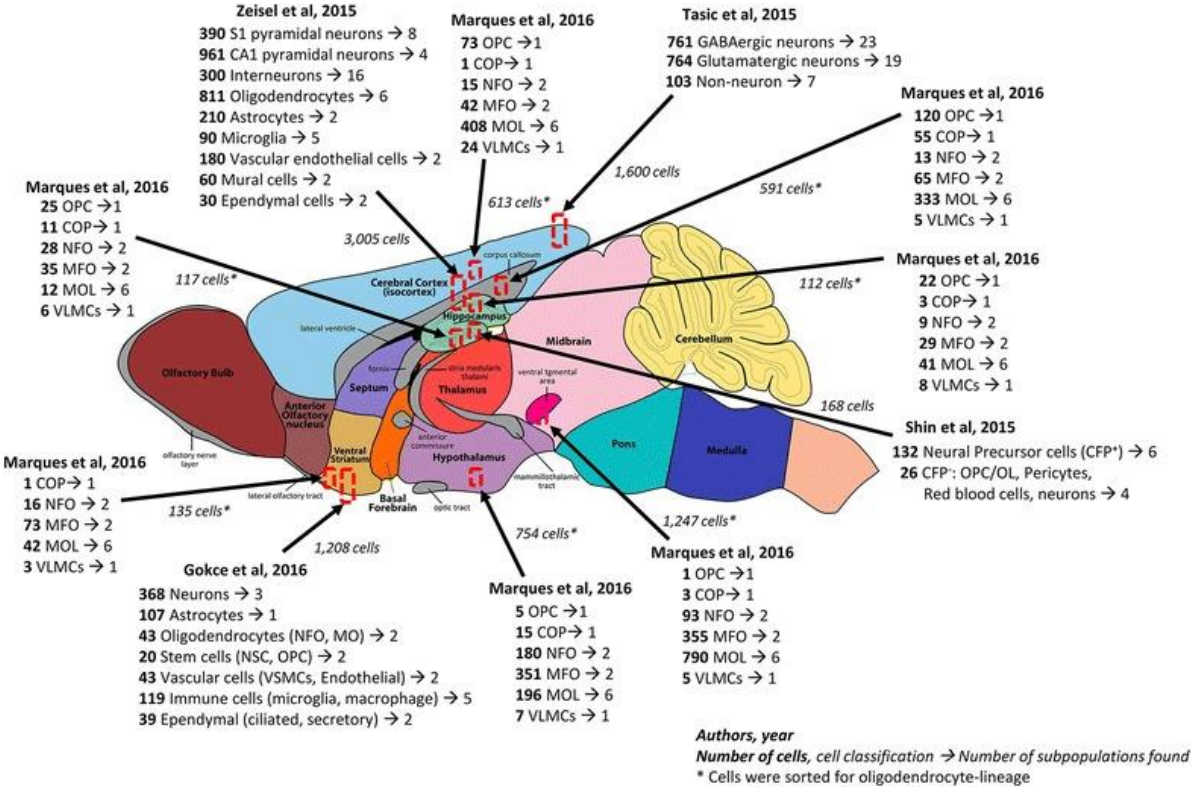


Figure 1-1. Mouse brain diversity surveyed by scRNA-seq studies [1]. Arrows correspond to collection region and number of isolated cells, numbers to the left are cell counts for coarse cell types, numbers to the right are number of subpopulations, asterisks mark oligodendrocyte-lineage enriched cells.

Creative Commons: “Selected relevant scRNA-seq studies revealing brain heterogeneity,” Figure 1 by Cuevas-Diaz Duran, R., Wei, H. & Wu, J. Q. in the article “Single-cell RNA-sequencing of the brain” on clintransmed.springeropen.com [1]. No changes were made. Licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

References

1. Cuevas-Diaz Duran, R., Wei, H. & Wu, J. Q. Single-cell RNA-sequencing of the brain. *Clinical and Translational Medicine* **6**, 20 (2017).

2. Wang, X., Lu, J., Guo, G. & Yu, J. Immunotherapy for recurrent glioblastoma: practical insights and challenging prospects. *Cell Death Dis* **12**, 1–15 (2021).

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).

Chapter 2:

A pathway-based approach reveals differential sensitivity of glioblastoma to E2F1 inhibition

A pathway-based approach reveals differential sensitivity of glioblastoma to E2F1 inhibition

Alvaro G. Alvarado[1], Kaleab Tessema[1], Sree Deepthi Muthukrishnan[1], Mackenzie Sober[1], Riki Kawaguchi[1], Aparna Bhaduri[2], Vivek Swarup[3], David A. Nathanson[4], Daniel H. Geschwind[1,5], Steven A. Goldman[6], Harley I. Kornblum[1,7,8, *]

Affiliations:

[1]Department of Psychiatry and Biobehavioral Sciences, and Semel Institute for Neuroscience & Human Behavior, David Geffen School of Medicine, UCLA, Los Angeles, CA

[2]Department of Biological Chemistry, David Geffen School of Medicine, UCLA, Los Angeles, CA

[3]Department of Neurobiology and Behavior, School of Biological Sciences, UCI, Irvine, CA

[4]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, UCLA, Los Angeles, CA

[5]Department of Neurology, David Geffen School of Medicine, UCLA, Los Angeles, CA

[6]Department of Neurology and the Center for Translational Neuromedicine, University of Rochester Medical Center, Rochester, NY, USA, and the University of Copenhagen, Denmark

[7]Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA

[8]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, UCLA, Los Angeles, CA

*** Corresponding author:**

Dr. Harley I. Kornblum
Email: hkornblum@mednet.ucla.edu

Abstract

Targeting glioblastoma (GBM) based on molecular subtyping has not yet translated into successful therapies. Here, we used gene set enrichment analysis (GSEA) to conduct an unsupervised clustering analysis to condense the gene expression data from bulk patient samples and patient-derived gliomasphere lines into new gene lists. We then identified key molecular pathways differentially regulated between tumors. These gene lists associated not only with cell cycle and stemness signatures, but also with cell-type specific markers and different cellular states of GBM. We identified the transcription factor E2F1 as a key regulator of tumor cell proliferation and self-renewal in only the subset of proliferating gliomasphere cultures predicted to be E2F1-activated and validated its functional significance in tumor formation capacity. E2F1 inhibition also sensitized E2F1-activated gliomasphere cultures to radiation treatment. Our findings indicate that a pathway-based approach can be leveraged to deconstruct inter-tumoral heterogeneity and uncover key therapeutic vulnerabilities for targeting GBM.

Introduction

Glioblastoma (GBM) is incurable, with an overall median survival of about 14 months(*1*) despite maximal surgical resection, radiation, and chemotherapy with temozolomide(*2*). The past decade has seen a revolution in the understanding of GBM, and studies of patient samples based on gene expression and oncogenic mutations have revealed that GBM can be parsed into distinct molecular categories--namely classical, mesenchymal, and proneural--and subsequently IDH mutated tumors(*3-8*). While these classification schemes have shown some relationship to prognosis, they have largely failed to provide new therapeutic approaches.

The driver mutations of GBM result in the activation of many well-known oncogenic pathways, including the PI3K/AKT, MAPK/ERK, Rb, and mTOR pathways(*9*). However, the use of pathway-specific inhibitors has not yet resulted in effective therapies. One potential explanation for this lack of efficacy is that tumors are comprised of multiple cell types with different pathway dependencies. Another is that inhibition of one pathway results in compensatory activation(*10*). Finally, it is possible that the identification of critical pathways driving GBM progression and recurrence is not yet complete. The prioritization of key pathways falls short of what would be required for tumor eradication because the combinatorial outcome of existing mutations, and resultant dominant pathways, cannot be conclusively inferred.

The advent of relatively easy and cost-effective sequencing methods and advances in bioinformatics open the door for rapid evaluation of gene expression, identification of key molecular pathways, and characterization of different cell types within individual tumors. For instance, we now appreciate that molecular subtypes previously used in the classification of GBM(*5*) can be linked to cell type-specific markers with the description of cellular states(*11*). Similarly, a subpopulation of tumor cells can express markers of outer radial glia and turn on developmental programs to promote invasion(*12*). Yet, a comprehensive and unbiased approach to group samples based on their pathway utilization has not yet been exploited to uncover therapeutic targets. We hypothesized that, rather than examining gene expression as a whole, analyzing targetable pathways could allow for the development of patient- or tumor class-specific therapeutics or combination therapies that go beyond the traditional inhibitors that have already been developed.

In this study, we developed a bioinformatics strategy that leverages Gene Set Enrichment Analysis (GSEA) to disambiguate tumor heterogeneity in GBM **(Figure 2-1).** We identified pathways and processes utilized by different tumors and individual cells within the tumors, without consideration of their driver mutations. We then implemented a downstream pipeline to ascertain key genes within enriched gene sets that were further evaluated for drug and/or molecular target selection. Using bulk RNA samples and existing single cell RNA-sequencing databases, we found that tumors can be clustered according to their enrichment of canonical and oncogenic pathway gene sets and that the new gene lists we derived from enriched gene sets reveal functionally significant differences between tumors and between cells within a tumor. We then utilized gliomasphere cultures, enriched or depleted in these gene lists, to demonstrate the functional significance of our approach using the pro-proliferative transcription factor E2F1 as an example.

Results

GBM clustering is clinically significant and modestly overlaps with prior molecular classification

To characterize intertumoral heterogeneity, we first studied presumptive pathway utilization by analyzing GBM patient samples in the Cancer Genome Atlas (TCGA(*5*)) using gene sets in the canonical (C2CP) and oncogenic (C6) pathway collections from the GSEA software(*13, 14*). The mRNA expression of each TCGA sample was compared to the average expression of all the samples (n = 538), and normalized enrichment scores were obtained for all the gene sets comprising the two pathway collections. A heatmap representing the normalized enrichment score profile for each sample (column) demonstrates the presence of heterogeneity when either canonical **(Figure 2-2A)** or oncogenic **(Figure 2-2B)** gene sets were analyzed. We determined that 3 clusters correctly represented the data when using oncogenic and canonical pathways via non-matrix factorization and consensus clustering; the robustness of the clusters was also tested and validated using the Random Forest approach **(Supplementary Figure 2-1)**. We then applied principal component analysis (PCA) for dimensionality reduction to better visualize the clustering of the samples using canonical and oncogenic pathways **(Figures 2-2C and 2-2D, respectively)**. Notably, when the clustered samples were colored by their known TCGA molecular subtype, we found there was one cluster that contained most of the mesenchymal samples, while the classical and proneural subtypes were present in all clusters **(Supplementary Figure 2-2)**. This indicates a lack of full correlation between molecular characterization and molecular pathway expression. These findings suggest that different tumors quantitatively utilize different molecular pathways, prompting further analysis.

To determine whether our pathway-based classification provided functionally significant information, we examined patient survival using the pathway-based clustering information. Prior studies using TCGA groupings have found only limited association with survival, with proneural tumors having longer survival--an observation largely driven by the subset of IDH mutant tumors. As shown in **Figure 2-3A**, our own analysis of the TCGA categories found significant differences only between the proneural and other two groups, as previously reported (*5*). However, when we utilized our new pathway-based clustering approach, we found statistically significant differences in median survival between the patients from each pair of clusters **(Figure 2-3B and 2-3C)**. For both collections of gene sets, canonical and oncogenic, the cluster with the lowest median survival was the one primarily composed of mesenchymal samples. However, approximately 30% of the samples within this cluster were characterized as classical and proneural. Similarly, using the canonical pathways collection, the cluster with the highest median survival included equal abundances of classical and proneural samples (47% each). These data challenge the idea that samples obtained from patients should be treated according to their TCGA-defined molecular phenotype and, in contrast, support the notion that tumors from different molecular backgrounds might have common signaling pathways that can be leveraged for therapeutic purposes.

GSEA gene signatures from patient-derived glioma database delineate actionable targets

Our findings thus far suggest that a pathway-based approach could be leveraged therapeutically in GBM. We recognized, however, that GSEA is an imperfect approach to assess functional pathway utilization and that individual genes or sets of genes would contribute to enrichment of multiple gene sets. Therefore, in order to more pragmatically develop potential interventions based on our analysis, we further distilled our pathway-based clustering of the whole TCGA dataset by extracting the top contributing gene sets to each principal component (PC) and direction and synthesizing their common elements into gene lists (Figure 2-1). We hypothesized that targeting the most highly determinative elements would allow us to target several key pathways simultaneously, even though we might be limiting our scope to shared targets and ignoring underrepresented pathways. In order to validate this approach, we performed a similar analysis on a microarray-derived database of patient-derived gliomasphere (GS) lines so that we could functionally test downstream targets. The enrichment patterns again showed heterogeneity between samples, yet they grouped into two major clusters when either canonical and oncogenic gene set collections were used **(Supplementary Figure 2-3)**. This is reminiscent of a number of gene expression studies, including our own, that classify cultured glioma cells into two major groups. Gene lists were also generated for the GS dataset based on the common elements shared among the top contributing gene sets for each PC and direction, as described above for the TCGA dataset (genes in each of the gene lists generated can be found in Supplementary Table 1, available online). The TCGA- and GS-based gene lists were then used to obtain enrichment scores in the gliomasphere lines **(Figure 2-4A)**. This dataset again separated into two main clusters in accordance with the pattern of enrichment scores generated when the oncogenic and canonical pathways were used **(Supplementary Figure 2-3)**. Interestingly, there was an overlap in enrichment between some of the gene lists generated from the TCGA dataset (TCGA_C2_PC1NEG) and the gliomasphere dataset (GS_C2_PC1NEG). This was confirmed by calculating correlation values for each pair of gene lists **(Supplementary Figure 2-4)**. In both datasets, the strongest positive correlation was between the C2_PC1NEG lists, with coefficients of 0.78 and 0.84 in the TCGA and GS datasets, respectively. Conversely, the strongest negative correlation was between the two C2_PC1NEG lists compared with GS_C2_PC1POS (-0.57, -0.66) in the GS dataset and with TCGA_C2_PC1POS (-0.39, -0.22) in the TCGA samples **(Supplementary Figure 2-4)**. These findings indicate that the identified gene lists represent independent groups of pathways for which tumor samples show differential enrichment that can potentially be exploited to uncover new therapeutic targets.

All gene lists were further analyzed using Ingenuity Pathways analysis (IPA) to identify critical pathways and targets that can be leveraged to identify molecular targets. Cluster 1 (right side of heatmap in **Figure 2-4A**) showed higher enrichment of the GS_C2_PC1NEG gene list and core analysis from IPA showed an increase in the expression of the E2F family of transcription factors and its downstream targets **(Figure 2-4B)**. Similarly, the contribution of each gene list can be appreciated in reference to both clusters **(Figure 2-4C)**. Activation of E2F1, together with inhibition of let7, reported to have a role in differentiation and tumor suppression (*15, 16*), had the most significant p-value for this particular gene list (data not shown). Since E2F1 has a known role in cell cycle progression, we examined the enrichment scores for cell cycle-related gene sets in the canonical pathways collection (C2CP). Indeed, samples that fell in the cluster with high enrichment of GS_C2_PC1NEG (and hence predicted to have activation of E2F1) had concomitant higher enrichment scores for cell cycle-related gene sets **(Figure 2-4D)**. Likewise, we examined the enrichment of E2F1-target gene sets from the C2CP collection and found great concordance with the samples predicted to have E2F1 activation **(Figure 2-4E)**. In addition, considering the correlation values between the signatures **(Supplementary Figure 2-4)**, we found that E2F1 activation seemed to oppose IFNγ and NFκB activation. Similarly, EGFR activation (TCGA_C2_PC2POS enrichment) appeared ubiquitously and showed correlation with most of the signatures in both datasets. These findings suggest that there are two clusters of gliomasphere samples based on their enrichment of gene lists that can be further analyzed to elucidate upstream regulators. Amongst the most meaningful differences between the samples was the fact that one cluster revealed an E2F1-activated signature exhibiting a high degree of enrichment for cell cycle and downstream target signatures.

We next re-clustered TCGA samples based on their enrichment for the gene lists described above and found 3 clusters, comparable to the original pathway analyses described in Figure 2-2 **(Supplementary Figure 2-3)**. The E2F1-activated signatures characterized one of the clusters, whereas the EGFR signature pointed in between two of the clusters. Additionally, we analyzed raw data from the available TCGA samples (n=160) from Broad Firehose. Differential expression analysis was performed on samples using the cluster identity from the gene signatures. These data were then used to highlight the most enriched gene ontology terms for each cluster. We found each of the 3 clusters had a defined set of GO terms: cell cycle-related, extracellular membrane and inflammation, and synapse and neurotransmitter signaling **(Supplementary Figure 2-3)**. These results suggest the existence of distinct clusters that can be parsed through their potential pathway utilization, highlighted by upstream regulator enrichment. Similarly, we found a greater complexity in the TCGA dataset compared with the GS dataset, as would be expected when analyzing primary and patient-derived lines, respectively.

Pathway analysis exposes limited intratumoral heterogeneity and underscores cell cycle signature

A key consideration in the evolution of thinking about glioma heterogeneity lies in the analysis of intratumoral heterogeneity. Several single cell RNA-sequencing studies have emphasized the finding of TCGA subtype heterogeneity within tumors—that is, tumor cells from the same tumor are often classified in different TCGA groups. To assess whether potential signaling pathway utilization is similar or different within individual tumors, we interrogated a single cell RNA-seq database derived from 6 primary GBM samples(*17*). Despite the reported molecular subtype diversity in each sample of this dataset, we found that, based on our GSEA approach, most cells from an individual sample clustered together based on their tumor of origin **(Supplementary Figure 2-5)** when either canonical or oncogenic gene sets were used. These findings suggest that, although clear intratumoral heterogeneity in gene expression exists, there may be some rationale in targeting dominant pathways as cells from each tumor have comparable enrichment profiles that contrast with cells from other tumors.

Although cells within any one tumor exhibited similarities in their enriched pathways, we recognized that our approach would not adequately identify those pathways that did indeed differ amongst cells within an individual tumor. Therefore, we analyzed each sample separately, in which each cell's gene expression was compared to its tumor bulk control. This analysis resulted in two clusters for each tumor. Principal component loadings were obtained in order to establish which gene sets in the collection contributed the most to the apparent spatial distribution of the gene sets in the 2-dimensional PC graph. We consistently observed a cluster of cells with a high enrichment of gene sets associated with cell cycle promotion and regulation (Supplementary Table 2, available online). In contrast, the second cluster was mainly described by gene sets associated with extracellular matrix processing and growth factor signaling pathways, suggesting that both cycling and non-cycling cells could be differentially targeted. In the cell cycle-enriched clusters, we identified genes such as cyclin dependent kinases and minichromosome maintenance complex subunits that have previously been associated with glioma progression and tumor growth (*18, 19*). Similarly, all 46 subunits of the proteasome complex were highly enriched in this cluster for all samples. It remains to be seen whether targeting the latter pathways will be therapeutically relevant for GBM.

In order to obtain some biological insight into the cluster not associated with cell cycle gene sets, we interrogated the gene sets associated with it using EnrichR and IPA. For each group, we selected the top common elements based on their frequency in order to generate new lists of genes that condense the information from all the gene sets. Lists were then uploaded to EnrichR (https://amp.pharm.mssm.edu/Enrichr/) and to the IPA software to obtain transcription factors, biological processes, and upstream regulators for each (Supplementary Table 3, available online). Although there are some differences between samples, there is a common theme of proliferative signals that are achieved through different mechanisms. For instance, CLOCK (circadian clock) is involved in the maintenance of pathways critical for tumor metabolism and that are upregulated during hypoxia in glioma (*20*). Other pathways of interest include those involving DNA repair, epithelial-mesenchymal transition, axon guidance, and axogenesis. There are also several transcription factors that are present in most samples, such as NUCKS1 and EGR1, both of which have been previously associated with glioma (NUCKS1 in pediatric (*21*) and EGR1 in adult (*22*)) and require further mechanistic analysis. Similarly, there is a report on a feed-forward loop between EGR1 and PDGFRA that promotes proliferation and self-renewal in GSCs (*23*). These findings in the scRNA-seq database indicate that there may be a potential combinatorial approach that can be tailored for each individual tumor to effectively target both cell cycle pathways and tumor-specific activated molecules.

Gene lists differentially associate with cell cycle and stemness programs, GBM cellular states, and cell type-specific signatures

One main focus of GBM research in the last decade has been the existence of a subset of cancer cells with activated stemness programs, namely glioma stem cells, that contribute to the malignancy of the tumor (*24-27*) and are refractory to therapy (*28, 29*). Together with the TCGA molecular subtypes, this paradigm has resulted in the development of specific therapies aimed at targeting molecules believed to be key regulators of tumor growth and invasion. The minimal clinical success of these efforts can be attributed to the limitations of the current *in vitro* and *in vivo* models used to validate these approaches. Importantly, we routinely work with patient-derived cell lines that behave differently than tumor cells in their intact tumor microenvironment in patients. It is important to establish whether critical pathways such as stem-like programs, differentiation pathways, or cell-cycle related signatures are predominantly active in these cells. To that end, we evaluated the association of our pathway-based gene signatures with cell-cycle and stemness scores in a GBM single-cell RNA-seq dataset (*17*). Consistent with our GS analyses, we found that both of our signatures predicted to have E2F1 activated (C2_PC1neg) strongly correlated with cell cycle scores **(Supplementary Figure 2-6)**, while other signatures showed weak or no association. Interestingly, we found two signatures (TCGA_C2_PC1neg and TCGA_C2_PC2pos) that had strong negative correlations and one with a positive correlation (TCGA_C2_PC1pos) with stemness score **(Supplementary Figure 2-6)**.

The negative correlation between activation of E2F1 and stemness score is not surprising given the fact that tumor cells are believed to be in either a proliferative or a stemness state. However, EGFR activation (TCGA_C2_PC2pos) has not been linked to a decrease in stemness and compels further investigation. The gene list with positive correlation to stemness (TCGA_C2_PC1pos) has three main targets per IPA analysis. First, phosphatidylinositol 3-kinase (PIK3R1) has been associated with GBM, and there are several inhibitors developed for molecules in this signaling pathway. GNA12 encodes for the G12 alpha subunit of G proteins and is of critical importance in regulating actin cytoskeletal remodeling in cells during migration, which is critical for tumor invasion. Finally, Early B-cell Factor 1 (EBF1) has been identified as a TET2 interaction partner in IDH-mutant cancers (*30*). These analyses establish a novel approach for uncovering new molecular targets based on a pathway-based approach that can be leveraged for the development of new therapies.

We next used the same scRNAseq dataset (*17*) and compared our gene lists to previously reported transcriptional signatures of cell types in adult cortex (*31*) and developing human brain (*32*), as well as to recent descriptions of cellular states in glioblastoma (*11*). As expected, our E2F1 activated gene lists both significantly correlated with G1/S and G2/M signatures in both datasets **(Supplementary Figure 2-7)**. Additionally, TCGA_C2_PC1POS (PIK3R1 and EBF1 activated) significantly associated with adult astrocytic markers and the AC-like molecular state, and TCGA_C2_PC2neg (ICMT activated) significantly associated with adult OPC markers and both OPC-like and NPC1-like molecular states. Notably, ICMT is a methyl transferase necessary for the targeting of CaaX proteins, which include the Ras family, to the cell membrane (*33, 34*). Ras/ERK signaling has been associated with the proneural subtype (*35*) and NRAS is expressed at higher levels in proneural subtype compared with mesenchymal and classical (https://gliovis.bioinfo.cnio.es/). TCGA samples classified as proneural also showed higher enrichment for genes in the NPC-like and OPC-like molecular states (*11*). These data suggest a targetable mechanism (trafficking to the cell membrane via ICMT) required for a specific protein (Ras) upregulated in a TCGA subtype (proneural) that has correlates in GBM molecular states (OPC-like, NPC1-like). Finally, we also found a significant correlation between GS_C2_PC1pos (IFNγ and NFκB activated) with both MES1-like and MES2-like states as well as adult endothelial and mural cell markers. The latter include blood vessel-associated cell types such as pericytes and vascular smooth muscle cells. This association relates to the extraordinary plasticity of glioma cells in response to their microenvironment. These data suggest IFNγ and NFκB pathways are activated in cells in the mesenchymal states that undergo vascular mimicry and express markers related to endothelial cells and pericytes that have been associated with tumor progression and recurrence (*36-38*). Of note, IL-10 is also predicted to be inhibited in this gene list; given IL-10 is an anti-inflammatory cytokine, this suggests an inflammatory microenvironment would promote this particular molecular state.

Weighted gene co-expression network analysis reveals distinct regulatory modules in each cluster

To further determine the biological relevance of the clusters identified in our GS dataset, we performed differential expression analysis and weighted gene co-expression network analysis (WGCNA). Gene ontology confirmed the enrichment of cell cycle signatures in the E2F1-enriched cluster **(Figure 2-5A)**. Likewise, WGCNA identified 26 modules **(Figure 2-5B)** that were distinctly associated with one of the two clusters **(Figure 2-5C)**. We took the top 3 enriched modules for each cluster and performed gene ontology enrichment analysis. Blue, brown, and light cyan modules (associated with E2F1 activation) showed enrichment of cell cycle, cell division, and DNA replication, in addition to processes associated with neurogenesis, neuron differentiation, and gliogenesis (Supplementary Table 4, available online). Moreover, two of the three modules also showed enrichment in their promoter region for E2F1 and other members of the E2F family. Conversely, black, green, and magenta modules (associated with the non-E2F1 enriched cluster) showed enrichment for inflammatory response, cell migration, and chemotaxis, as well as immune response, angiogenesis, and regulation of apoptotic processes. In these modules, we found genes with enrichment for transcription factors involved in inflammatory responses, such as CEBPB and the interferon regulatory transcription factor (IRF) family, as well as C2H2 zinc finger family members, including EGR1 and SP/KLF, which regulate proliferation, differentiation, and apoptosis cellular processes. We similarly identified the hub genes in each module, determined by how associated they are to the other members of their module(*39*), and colored them by module name in **(Figure 2-5B)**. Like the modules as a whole, the hub genes had different characteristics for modules associated with the E2F1 activated and non-activated clusters. Namely, the enriched E2F1-related modules show presence of known stem cell markers SOX2 and OLIG2, associated with self-renewal and persistent proliferation, as well as markers of cell division, like PLK4. Conversely, the hub genes of the other set of modules include IL-8, IL-6, and other inflammatory cytokines, together with CD44, which has been associated with a more invasive phenotype.

From a clinical perspective, we wanted to know if the hub genes would be viable as therapeutic targets. To this end, we took advantage of a prior study comparing gene expression in the cellular fraction containing tumor initiating cells, termed glioma-derived progenitors cells (GPCs) and normal, non-transformed glial progenitor cells (nGPCS) (*40*). For the E2F1-related modules, the blue module had several hub genes (DTL, CASC5, CDCA5, ASPM, CENPF, BUB1B) whose expression was at least 4-fold change higher in glial GPCs. Similarly, MYC in the light cyan module was 12-fold more highly expressed in glial GPCs relative to nGPCs. Evaluation of the non-E2F1 associated hub genes uncovered that CD44 and COL1A1, both associated with invasion, were highly expressed in glial GPCS (over 22-fold change higher related to nGPCs). Two other genes associated with migration and invasion, FN1 and SERPINE1, were also at least 4-fold higher in glial GPCs. These data confirm our pathway-based analyses that generated two clusters characterized by cell cycle enrichment contrasted with inflammatory and promotion of invasion pathways that are predicted to have lower chances of off-target effects based on their expression profiles on glioma-derived and normal progenitor cells.

Differential effects of E2F1 silencing and candidate therapeutics support the functional significance of pathway-based heterogeneity

Our findings of potentially differential dependence on E2F1 and its downstream targets in highly proliferative gliomasphere cultures was somewhat surprising, as this transcription factor is often thought to be primarily involved in proliferation and cell cycle regulation. To investigate this further and to validate our general approach, we used silencing technology to evaluate the cellular effects of E2F1 suppression in samples from E2F1-activated and non-activated clusters. HK217 and HK301, members of the E2F1-activated cluster, showed a marked decrease in stem (sphere-forming) cell frequency in a limiting dilution assay (LDA) in cells with E2F1 knockdown compared with control **(Figure 2-6A)**. These effects were not observed in HK357 or HK408, lines that were not enriched for an E2F1-activated signature **(Figure 2-6A)**. Likewise, knockdown of E2F1 resulted in compromised overall cell proliferation in E2F1-enriched samples when E2F1 expression was suppressed **(Figure 2-6B)**, compared with non-enriched cells where E2F1 knockdown did not significantly alter proliferation.

To determine the *in vivo* relevance of our findings, we assessed tumor formation capacity in HK301 (E2F1-activated) and HK408 (E2F1-non-activated) cell transduced with E2F1 KD (knockdown) and control (scrambled) lentivirus. *In vivo* bioluminescent images showed tumor growth in all mice intracranially transplanted with HK408 in both control and KD groups. Conversely, HK301 E2F1 KD cells showed limited tumor formation *in vivo* while animals injected with HK301 control cells showed tumors in four out of five mice **(Figure 2-6C),** and the quantification of the luminescence signal was significantly decreased only for this group **(Figure 2-6D)**. These data demonstrate that the targets uncovered by this pipeline have functional implications in patient-derived gliomaspheres.

The transcription factor E2F1 is a critical component of the cell cycle signaling machinery. However, having a cluster of samples that do not rely on this molecule for proliferative purposes, as demonstrated *in vivo* and *in vitro* by loss-of-function experiments in the non-E2F1 cluster, was surprising. We therefore decided to investigate whether E2F1 might be regulating other cellular processes in GBM. E2F1 has been demonstrated to have a role in the suppression of senescence in prostate cancer cells and proposed to be a key factor for the progression of tumors in the presence or absence of p53 or retinoblastoma (*41*). Similarly, it has been described in breast cancer cells that senescence sensitivity is regulated by an interaction between E2F1 and cellular inhibitor of PP2A (CIP2A) (*42*). Accordingly, we found increased expression of CIP2A in the E2F1-activated compared with the non-enriched cluster in the GS dataset **(Figure 2-7A)**. Correlation analyses also found highly significant correlations in both the TCGA and GS datasets between E2F1 and CIP2A expression **(Figure 2-7B)**. Of note, E2F1 knockdown affected CIP2A expression only in E2F1-enriched samples **(Figure 2-7C)**, which suggests a unique regulatory network primarily utilized in cells belonging to this cluster. We decided to test this functionally by treating cells with irradiation and measuring their capacity to resolve DNA damage as measured by γH2AX staining. HK217 and HK408 control and E2F1 KD cells were irradiated with a single dose of 8Gy, and cells were stained 12 hours later. HK408 showed comparable levels of H2AX-positive cells under both conditions (control = 80%, KD = 78%, n.s.), whereas E2F1 KD significantly impacted the capacity of HK217 cells to resolve DNA damage (control = 45%, KD = 68%, p-value = 0.01) **(Figure 2-7D)**. Additionally, we evaluated the enrichment of DNA repair- and senescence-related gene sets in samples from both clusters. As expected, samples in the E2F1-activated cluster showed higher enrichment for both groups of gene sets **(Supplementary Figure 2-8)**. These data further confirm that the pathway-based approach we have implemented in these studies has identified a specific molecular target for a cluster of samples that has both biological significance and therapeutic potential to advance treatment for patients with GBM.

Discussion

In this study, rather than focusing on driver mutations themselves, our goal was to focus on their impact on gene expression and to use the latter in an unbiased manner to assay molecular pathways that will influence the biology of the tumor. Our assumption is that while individual mutations may influence one of a number of different processes, ranging from protein phosphorylation to chromatin modifications, mutations will ultimately result in altered gene expression, which then results in modified cellular function. Although our strategy does result in a reclassification of tumors and tumor cells, the analyses described in this work present a pathway-based approach to uncover biologically relevant, actionable targets derived from the heterogenous biology inherent to glioblastoma.

GSEA is a powerful tool used to group sets of genes in a functionally relevant manner. However, the gene sets in GSEA, especially in the oncogenic collection, are often based on the responses of cells and certain tissues to genetic perturbations, and enrichment for a particular gene set does not prove that a specific process or pathway is involved. Furthermore, an individual gene or group of genes can be represented in multiple gene sets. In order to obtain more precise information from our analysis, we added an extra layer to our approach where we extracted genes that were common to multiple gene sets and that were highly associated with the directionality of the principal components. We then analyzed these “eigen-gene sets” for their enrichments in pathways and processes. This additional step allowed us to identify functionally important genes and processes.

Using our approach, gene lists were established form both bulk tumor samples and patient-derived gliomasphere datasets and associated with specific cell signatures in a single cell dataset. One of the clearest relationships we observed was the strong association between signatures associated with E2F1 activation and proliferative signatures. One potential explanation of such a finding would be that different tumors have different numbers of proliferating cells and thus differential gene expression based on their abundance. However, our findings in gliomaspheres suggest that there are more complex processes at play, as both E2F1-dependent and non-dependent cultures were highly proliferative at the time of study, indicating that the expression differences observed represented true differences in the biology of the cells. It is possible that another closely related member of the E2F family would serve the same function as E2F1 in the non-enriched population. However, one may surmise that such factors would result in similar downstream effectors and therefore would not have appeared to be enriched in our studies.

In addition to its role in proliferation, recent studies tie E2F1 function to other processes, including DNA repair. E2F1 is known to promote the expression of the DNA repair protein CIP2A, which was confirmed to be enriched in our E2F1-activated samples. Our studies confirmed inhibition of E2F1 reduced CIP2A expression and reduced the capacity to resolve irradiation-induced DNA damage in E2F1-activated gliomasphere cultures. Our analysis also identified small molecules that could selectively target this pathway and be considered for development of therapeutics in subclasses of cells. This study leveraged the availability of a large library of well-characterized gliomasphere samples. While these are thought to embody much of the complexity of GBMs, they are undoubtedly a simpler system and are likely to be enriched in actively proliferating cells, as opposed to quiescent stem or stem-like cells. Here we report differential sensitivity of tumor cells to E2F1 inhibition between distinct clusters of proliferating gliomasphere lines; yet we recognize the limitations of the gliomasphere culture system. Most studies based on expression have parsed gliomasphere cultures into two general categories, similar to our findings, rather than the multiple subtypes exhibited by tumors themselves. It is possible that other culture systems, such as organoids, would be better able to replicate the inter- and intratumoral heterogeneity observed in gliomas.

In addition to an E2F1-driven cluster of tumors and gliomaspheres, we observed clusters that were not E2F1-driven and that appeared to be more heavily reliant on other pathways. For example, the cluster on the left in Figure 2-4A showed diverse enrichment for gene lists whose main targets are more classical dysregulated pathways in glioblastoma, such as PIK3R1 and PDGF receptor (*9*). This cluster also had a strong enrichment in the majority of its samples for an IFNγ- and NFκB-activated signature. This inflammatory and/or damage response was also observed in the TCGA dataset as one of the main components for one the clusters described in our first analysis **(Supplementary Figure 2-3)**. Finally, samples predicted to have EGFR activation were equally distributed in both clusters, suggesting EGFR expression levels are not particularly informative in terms of functional diversity in glioblastoma samples.

In our single cell analysis, where we examined the pathway-based expression of all the cells within a tumor, we found that the most significantly enriched gene sets compared to the bulk tumor average were generally very similar. This suggests that the predominant pathways used by all or most of the cells within a single tumor might be targeted, but that these pathways would vary from tumor to tumor. It is important to note that individual sample analyses did separate cells from each tumor into a cell cycle-enriched cluster and non-cell cycle related cluster. In the latter, we identified EGR1 as a transcription factor upregulated in most samples (Supplementary Table 3, available online). This same molecule was underscored in the modules associated with the non-E2F1 cluster when we performed WGCNA (**Figure 2-5B** and Supplementary Table 4, available online). EGR1 has been associated with O-6-Methylguanine-DNA Methyltransferase (MGMT) methylation (*22*). Our findings link a potential prognostic marker to a subset of GBM samples that could have potential therapeutic implications.

Moreover, using cell signature scores rather than setting arbitrary values for cell type identities, we were able to determine some of the characteristics of individual cells within tumors. This analysis confirmed that an E2F1-driven signature correlated with genes that were related to mitosis, but inversely correlated with putative markers of stemness. It is unclear whether this is because “true” cancer stem cells are slowly dividing, or whether other factors are involved. We were able to further our correlational analysis to include cellular states described in glioblastoma (*11*) and cell types from normal brain development. The rationale to do these analyses was based on a recent report using scRNA-seq that uncovered a subset of glioblastoma cells with outer radial-glia signatures that were able to activate an embryonic pathway to promote invasion (*12*). Our studies link cellular states to cell type-specific signatures and potential targets from our gene lists. For example, predicted ICMT activation was negatively associated with stemness and positively associated with NPC-like and OPC-like states, as well as adult OPC signatures. Similarly, IFNγ and NFκB activation were positively correlated with MES-like1 and MES-like2, as well as endothelial and mural (vasculature, pericyte) signatures. This last interaction is particularly interesting since it suggests an inflammatory environment as a driver for the expression of tumor vasculature markers. This is consistent with the capacity of glioma cells to undergo transdifferentiation into endothelial cells and pericytes to promote invasion(*36-38*). These analyses provide new potential avenues for the development of innovative treatments.

Recent reports have appreciated the complexity of tumor cell expression signatures, and now the emphasis has been on providing a more holistic view of the cells within each tumor, such as cellular states (*11*), a single axis of variation between proneural and mesenchymal subtypes (*43*), and a recent report using a similar approach to ours that introduces another layer of complexity to glioblastoma heterogeneity by uncovering a mitochondrial subtype with unique vulnerabilities (*44*). Our study adds to this trend by providing a novel approach to condense tumoral heterogeneity to critical gene lists that can be used to identify upstream regulators. In conclusion, we propose a combinatorial approach where precision medicine will be composed of sample-specific drugs that also provide specific vulnerabilities to be exploited with metabolic and/or immune-activating approaches. The integration of different aspects of a cell or sample is paramount for the development of new therapeutics.

Materials and Methods

Patient and tumor datasets

Glioblastoma samples analyzed were composed of TCGA dataset (*5*), gliomasphere microarray dataset (*45*), and single cell RNAseq dataset (*17*) (GSE57872).

Gene set enrichment analysis, gene list generation, and target prediction

For TCGA and gliomasphere datasets, each sample was compared to the average of the whole dataset using the canonical (C2CP) and oncogenic (C6) gene set collections from the GSEA website (http://www.gsea-msigdb.org). Enrichment profiles were then used to generate principal component analysis plots, and the contribution of each gene set to a particular direction was extracted using R-package ‘FactorMineR’. The top 20 contributing gene sets in a particular direction were compared to one another and common elements (present in at least 5) were considered for the gene lists. The names are derived from the dataset (TCGA or GS), the component (PC1 or PC2), and the direction (positive or negative). Datasets were reanalyzed with these gene lists to obtain similar clustering and for downstream analyses. Gene lists were further explored via Ingenuity Pathway Analysis using the upstream regulator tool. Targets (molecules or drugs) were predicted to be activated or inhibited for each of the gene lists used as input.

Gene list correlation analysis

For both the TCGA and GS bulk transcriptomic datasets, Pearson correlation coefficient was computed using R (“corr.test” function in “psych” package) for each pair of gene lists using the sample-level enrichment scores previously generated. Hypothesis testing was performed for each pair to assess significance of correlation, resulting in a matrix of both raw and FDR adjusted p-values. Correlation plots were generated using the “corrplot” package in R. Displayed are the correlation coefficients for each pair and circles whose color and size reflect the coefficient value and magnitude, respectively. For pairs with non-significant correlation, the coefficients are displayed without circles. Significance was evaluated using α=0.05 for both raw (below diagonal) and FDR-adjusted (above diagonal) p-values. Correlation patterns were used to group the gene lists using hierarchical clustering, with black boxes marking the resulting clusters.

Single cell signature score analysis

Single cell expression data (n=430) from 5 primary human glioblastoma specimens were imported from GSE57872. For each gene set of interest, single cell enrichment scores were generated as described (*17*). Briefly, the enrichment score of a gene set was computed in each cell by taking the average expression of genes within the gene set and subtracting the average expression of all detected genes. Single cell enrichment scores were generated for 1) the 6 TCGA/GS gene lists discussed above, 2) the cell cycle meta-signature described in Fig 2B of (*17*), and 3) the stemness signature described in Table S1 of (*46*). These scores were used to visualize pair-wise gene set correlations across cells, specifically between each of the 6 TCGA/GS gene lists and cell cycle (Supp. Fig. 2-5A) or stemness (Supp. Fig. 2-5B). Single cell enrichment scores were then generated for two additional groups of gene sets: developing (*32*) and adult (*31*) brain cell type markers and glioblastoma cellular state markers (*11*). These scores were used in combination with the 6 TCGA/GS gene list scores to generate a correlation plot (Supp. Fig. 2-6) as described above in the “Gene list correlation analysis” section.

Patient-derived gliomasphere cultures

Established patient-derived gliomasphere lines were cultured and maintained as previously described (*47*). Experiments were performed only with lines that were cultured for less than 20 passages since their initial establishment and were tested for mycoplasma regularly.

In vitro functional analysis: sphere formation and cell proliferation

Cell proliferation experiments were conducted by plating cells at a density of 2,000 cells/well in a 96-well plate in quadruplicate. Cell number was measured after 3 and 7 days and normalized to the initial reading at day 0 using the CellTiter Glo Luminescent Cell Viability Assay (Promega). The experiments shown represent fold change at day 7 relative to day 0. For sphere formation assays, cells were plated at a low density (100, 50, 25, and 12 cells per well) in 96-well plates (24 wells per density). Cells were maintained for 10 days before sphere formation was evaluated. Spheres larger than 10 cells in diameter were considered for analysis. The numbers shown represent the number of cells per well or the stem cell frequency as calculated using the Walter and Eliza Hall Institute Bioinformatics Division ELDA analyzer (http://bioinf.wehi.edu.au/software/elda/) (Hu and Smyth, 2009). All sphere formation and proliferation experiments were repeated at least three times.

In vivo tumor xenografts and imaging

For tumor formation assessment, 8- to 12-week-old NOD-SCID null (NSG) mice were used in equal numbers of female and male. 5x10^4 tumor cells containing a firefly-luciferase-GFP lentiviral construct and either a scrambled or E2F1 shRNA vector were transplanted per mouse (n = 5), in accordance with UCLA-approved Institutional Animal Care and Use Committee protocols. Five mice were housed per cage, with a 12hr light/dark cycle, and were provided food and water ad libitum. Tumor growth was monitored every 2 weeks after transplantation by measuring luciferase activity using IVIS Lumina II bioluminescence imaging. ROIs were selected to include the tumor area, and radiance was used as a measure of tumor burden. Mice were monitored and sacrificed upon the development of neurological symptoms such as lethargy, ataxia, and seizures, along with weight loss and reduction in grip strength. Animals were sacrificed by CO2 asphyxiation and secondary cervical dislocation.

Lentivirus transduction in gliomasphere lines

Lentiviral vector particles containing E2F1 and scrambled shRNAs were purchased from Abmgood. Cells were transduced with the corresponding viruses for 48 hours and selected with puromycin (Sigma). Knockdown of E2F1 was confirmed using immunoblotting in treated samples.

Immunofluorescence analysis

Cells were plated on 24-well plates pretreated with laminin overnight. After two days of culture, cells were fixed in 4% paraformaldehyde for 15 min at room temperature, followed by blocking and overnight incubation at 4°C with γH2AX primary antibody (Cell Signaling Technology). Cells were then incubated with species-specific goat secondary antibody coupled to AlexaFluor dye (568, Invitrogen) and Hoechst dye for nuclear staining for two hours at room temperature. Plates were imaged using EVOS microscope, and quantification of positively stained cells was performed manually using ImageJ.

Immunoblotting analysis

Cells of interest were lysed using RIPA lysis buffer (Thermo Scientific), and protein concentrations were calculated using a BCA protein assay (BioRad). After denaturation with Laemmli buffer (BioRad), 10 mg of total protein was loaded on 4-12% polyacrylamide SDS-PAGE gels (NuPage, Themo Scientific), transferred to polyvinyl difluoride (PVDF) membranes (Millipore), and probed using the following antibodies: E2F1 (Santa Cruz Biotechnology, 1:500), CIP2A (Santa Cruz Biotechnology, 1:1000), and b-Actin (Cell Signaling Technology, 1:2000) for loading control. Species-specific horseradish peroxidase (HRP)-conjugated secondary antibodies were used for detection (Cell Signaling Technology, 1:5000). Membranes were developed using ECL-2 reagent (Pierce Biotechnology). All western blots were performed at least three times.

Irradiation of gliomaspheres

Cells were irradiated at room temperature using X-ray irradiator (Gulmay Medical Inc., Atlanta, GA) at a dose rate of 5.519 Gy/min for the time required to apply an 8Gy dose. The X-ray beam was operated at 300 kV and hardened using a 4mm Be, a 3mm Al, and a 1.5mm Cu filter, and calibrated using NIST-traceable dosimetry.

Statistical analysis

Reported data are mean values ± standard error of the mean for experiments conducted at least three times. Unless stated otherwise, one-way ANOVA was used to calculate statistical significance, with p-values detailed in the text and figure legends. P-values less than 0.05 were considered significant. Correlation analyses were performed using Pearson coefficient. Log-rank analysis was used to determine the statistical significance of Kaplan-Meier survival curves. Data analysis was done using R v 3.6.3 (*48*).

Acknowledgements

The authors thank the members of the Kornblum lab for insightful comments. We would like to acknowledge the UCLA pathology and flow cytometry cores for their technical assistance.

Funding:

UC President's Postdoctoral Fellowship (AGA)

UCLA/Caltech Medical Scientist Training Program (KT)

NIH Ruth L. Kirschstein National Research Service Award F31 MH122205 (KT)

Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (HIK)

National Institutes of Health, UCLA SPORE in Brain Cancer P50 CA211015-01A1 (HIK)

Author contributions:

Conceptualization: AGA, SAG, HIK

Methodology: AGA, KT, RK, AB, VS

Investigation: AGA, SDM, MS

Visualization: AGA, KT, HIK

Supervision: DAN, DHG, SAG, HIK

Writing – original draft: AGA, HIK

Writing – review and edits: AGA, KT, SDM, MS, AB, DAN, SAG, HIK

Competing interests: The authors declare no potential conflicts of interest.

Data and materials availability: All data are available in in the main text or supplementary materials.

Figures

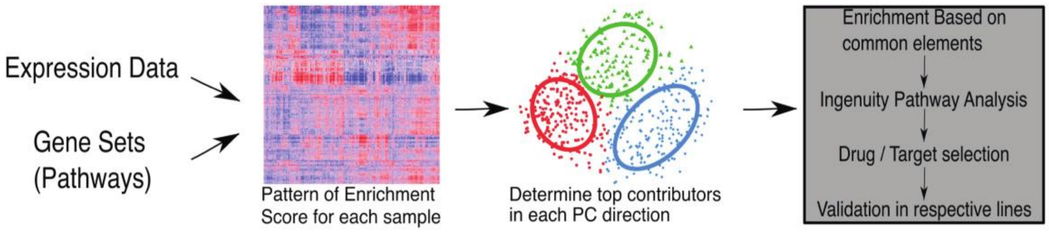


Fig. 2-1. Bioinformatics pipeline for pathway-based analysis. Expression data is analyzed through gene sets to generate enrichment score profiles that are then used to cluster samples. Top contributors to particular directions are obtained and run through Ingenuity Pathway Analysis to determine targets, which can be validated in corresponding samples based on their clustering.

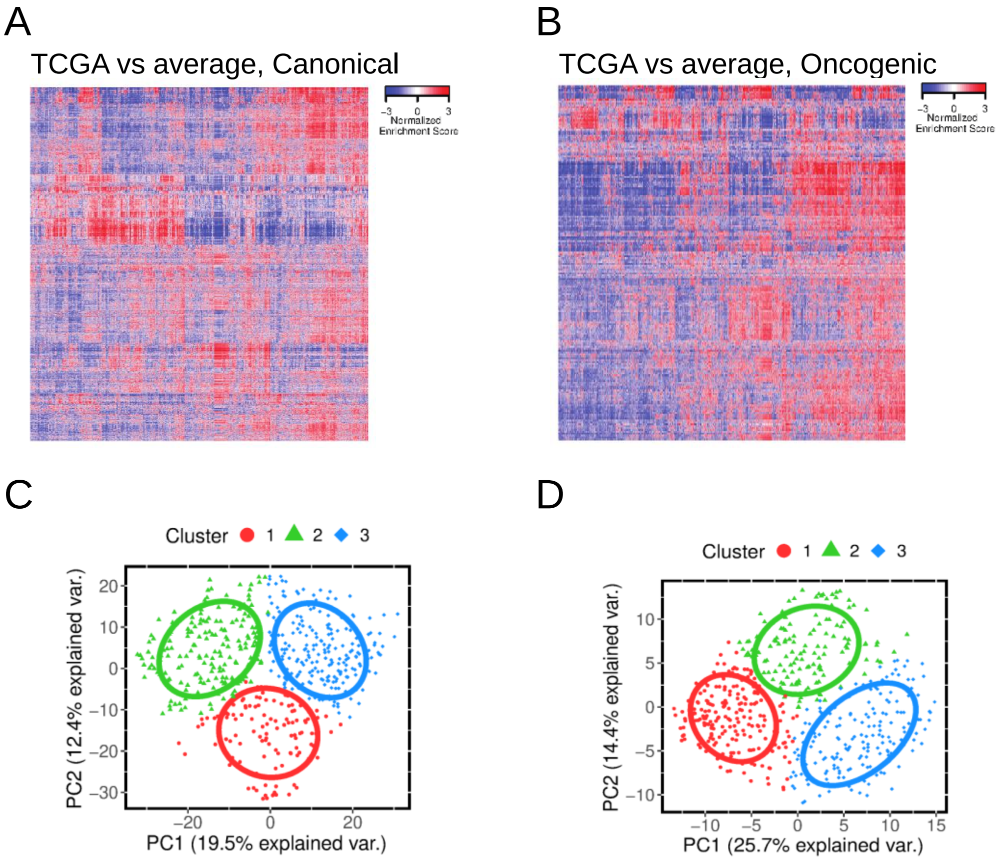
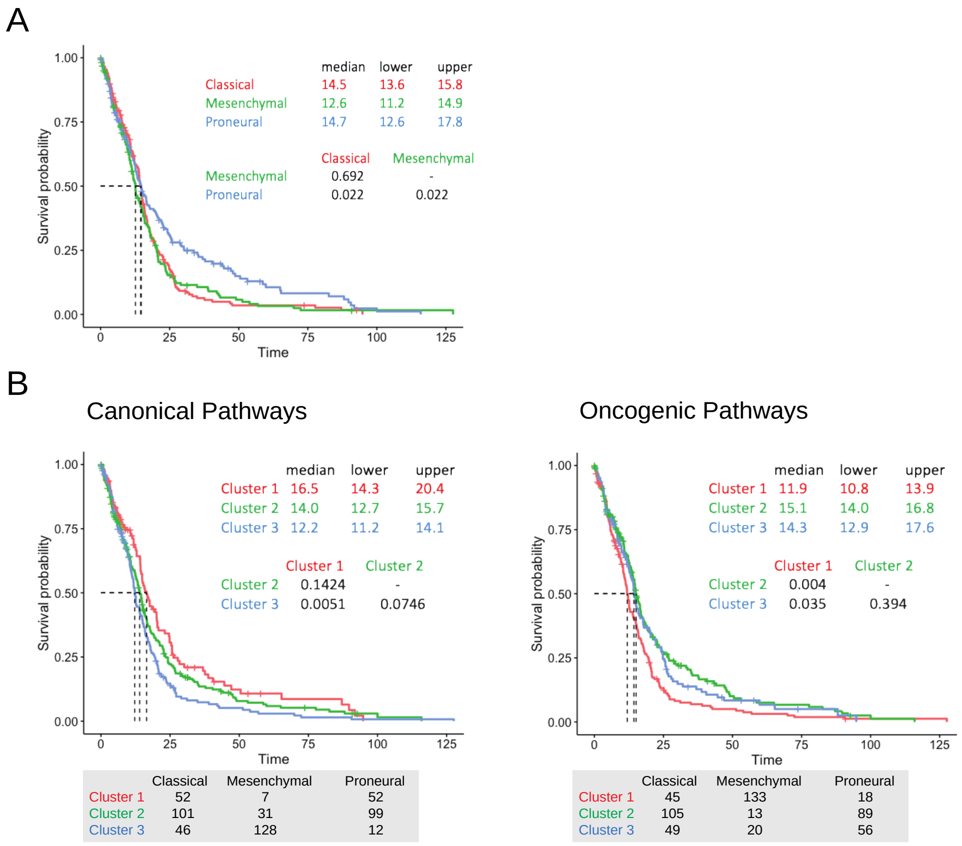


Fig. 2-2. Pathway-based analysis generates three distinct clusters based on enrichment profiles. Samples from the Cancer Genome Atlas were analyzed using either canonical pathways **(A)** or oncogenic pathways **(B)** from the Gene Set Enrichment Analysis to generate heatmaps based on the enrichment profile of each sample (column) with respect to each gene set (row) in both collections. **(C and D)** Profiles from **(A)** and **(B)**, respectively, were used to generate principal component analysis plots labeled by color and shape for each cluster. Circle lines represent the normal distribution of the samples in each cluster.



	Classical	Mesenchymal	Proneural
Cluster 1	52	7	52
Cluster 2	101	31	99
Cluster 3	46	128	12

	Classical	Mesenchymal	Proneural
Cluster 1	45	133	18
Cluster 2	105	13	89
Cluster 3	49	20	56

Fig. 2-3. Pathway-based clusters have clinical significance and do not overlap with molecular subtypes. (A) TCGA samples were clustered based on the original molecular subtypes described, and Kaplan-Meier curves were obtained. **(B and C)** Samples clustered based on enrichment profiles for canonical and oncogenic gene sets, respectively, were analyzed for survival using Kaplan-Meier curves. Tables at the bottom describe the distribution of the molecular subtypes for each cluster. Dotted lines represent median survival for each curve (also described in top tables). Time shown is in months. P-values after post-hoc analyses using Bonferroni-Hochberg correction.

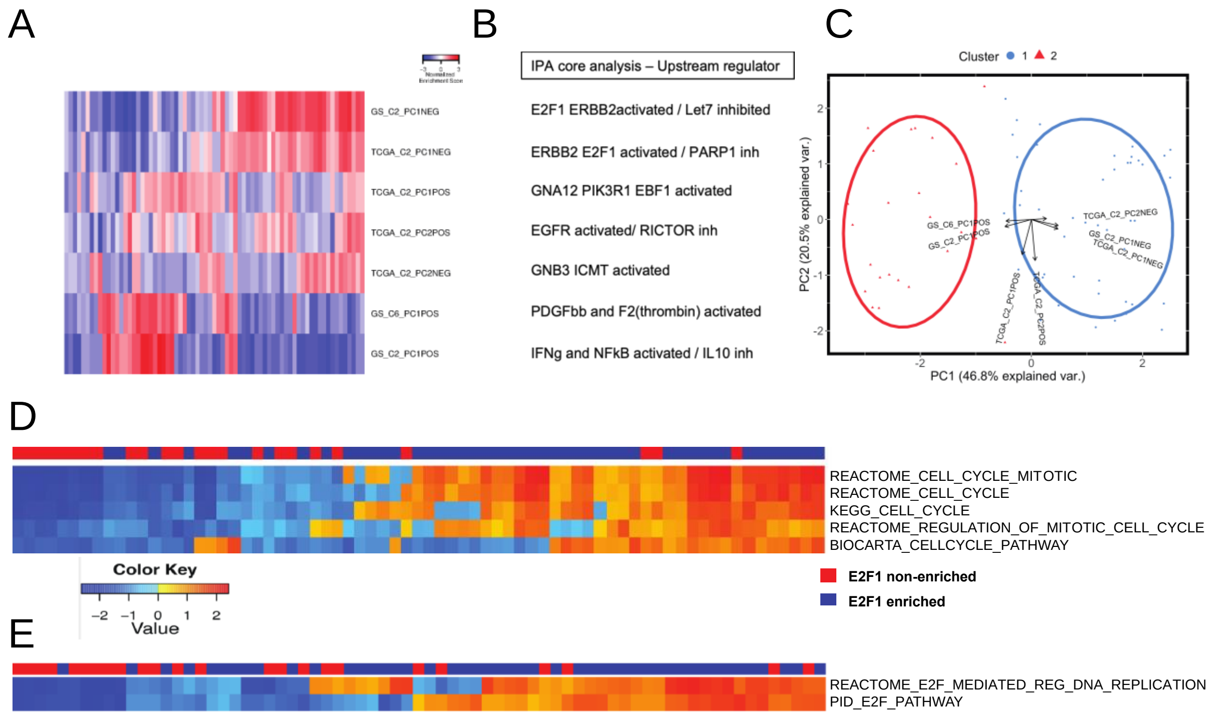
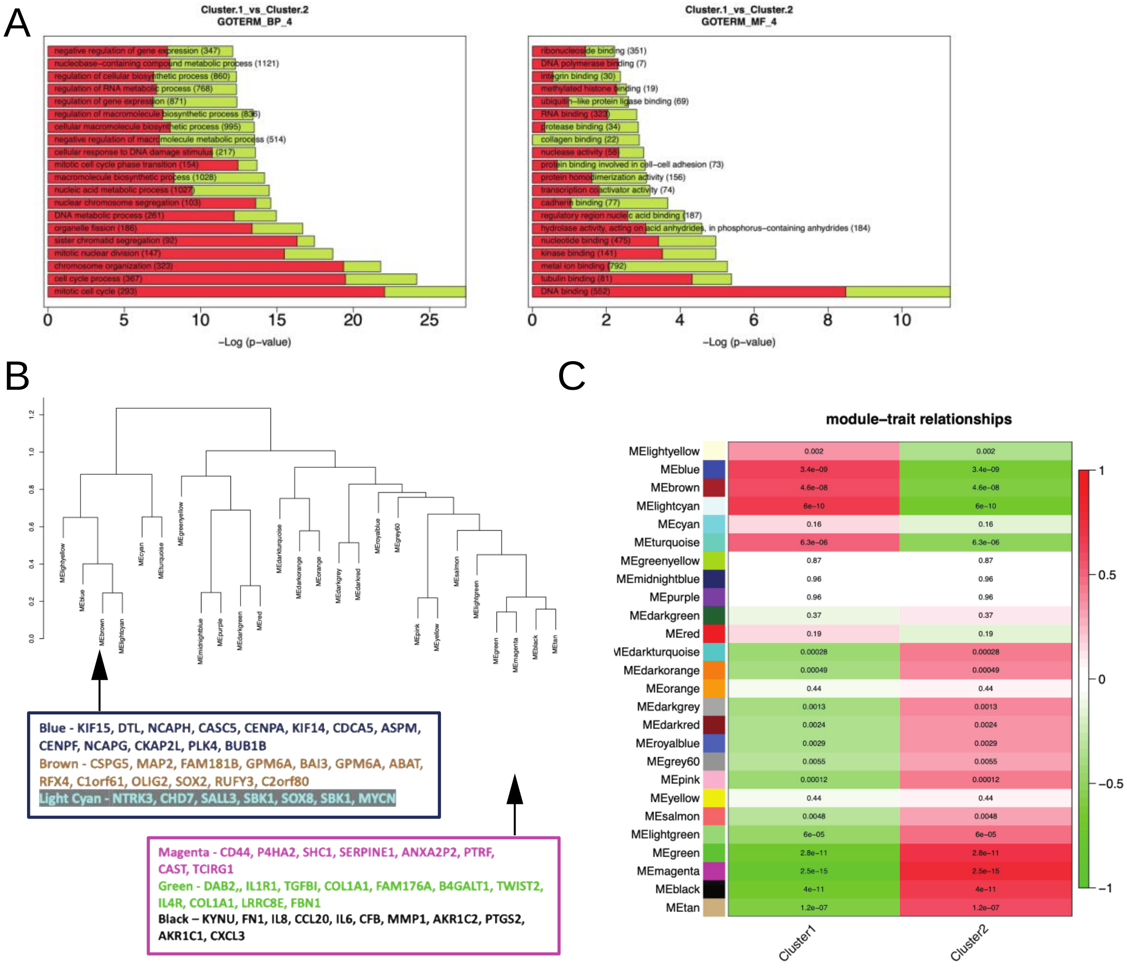


Fig. 2-4. Gene lists predict E2F1 as a main target in one of the clusters found in the gliomasphere dataset. (A) Enrichment profiles using gene lists were generated for gliomasphere samples. **(B)** Each gene list was evaluated using IPA and top predicted activated and inhibited upstream regulators are shown. **(C)** PCA plot from enrichment scores generated in **(A)** showing how each gene list contributes to a particular direction. **(D and E)** Samples from both clusters were evaluated for their enrichment of cell cycle-related **(D)** or downstream E2F1 target **(E)** gene sets from the canonical pathway collection.



Blue - KIF15, DTL, NCAPH, CASC5, CENPA, KIF14, CDCA5, ASPM, CENPF, NCAPG, CKAP2L, PLK4, BUB1B

Brown - CSPG5, MAP2, FAM181B, GPM6A, BAI3, GPM6A, ABAT, RFX4, C1orf61, OLIG2, SOX2, RUFY3, C2orf80

Light Cyan - NTRK3, CHD7, SALL3, SBK1, SOX8, SBK1, MYCN

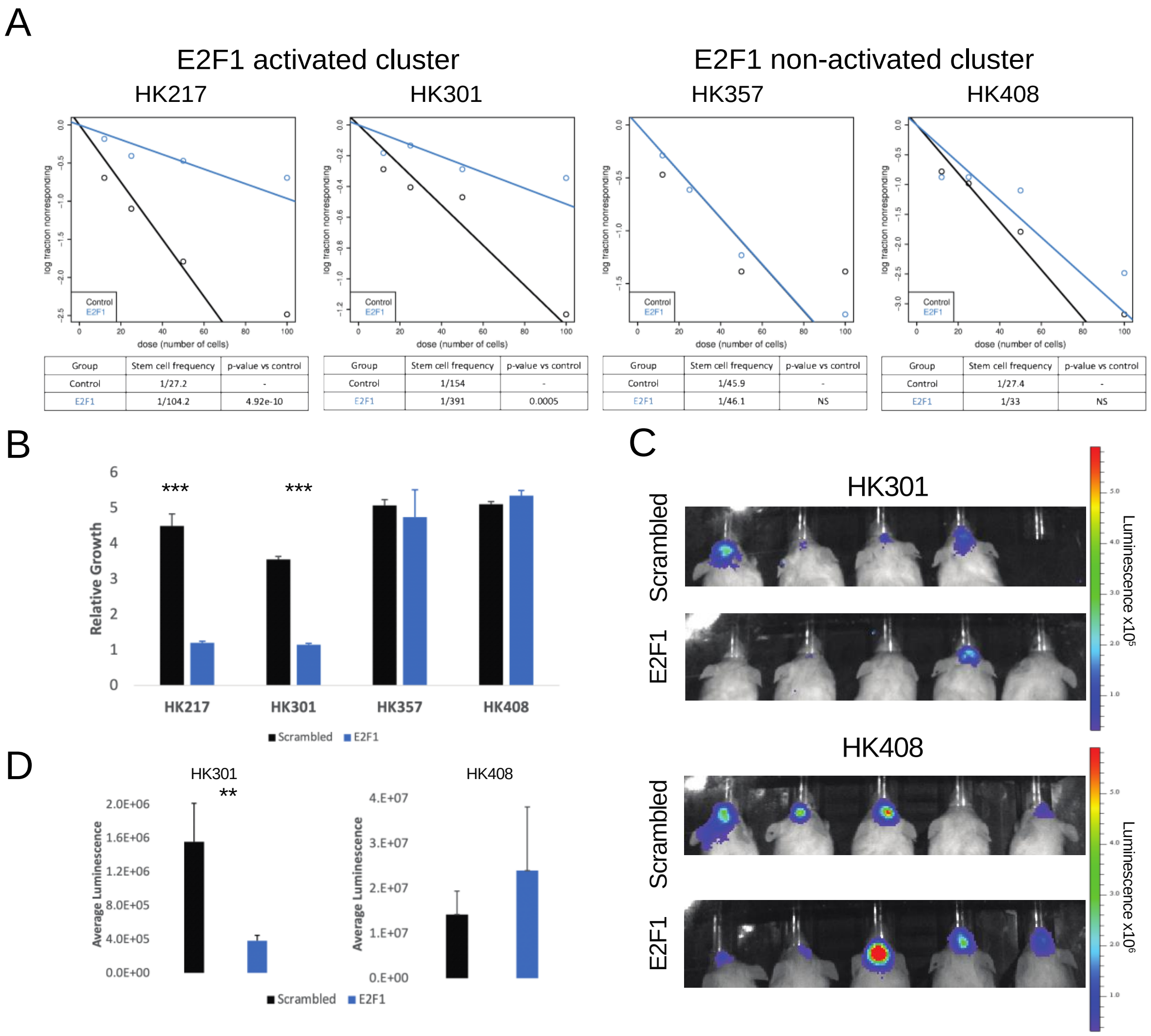
Magenta - CD44, P4HA2, SHC1, SERPINE1, ANXA2P2, PTRF, CAST, TCIRG1

Green - DAB2,, IL1R1, TGFBI, COL1A1, FAM176A, B4GALT1, TWIST2, IL4R, COL1A1, LRRC8E, FBN1

Black – KYNU, FN1, IL8, CCL20, IL6, CFB, MMP1, AKR1C2, PTGS2, AKR1C1, CXCL3

	Cluster1	Cluster2
MElightyellow	0.002	0.002
MEblue	3.4e−09	3.4e−09
MEbrown	4.6e−08	4.6e−08
MElightcyan	6e−10	6e−10
MEcyan	0.16	0.16
MEturquoise	6.3e−06	6.3e−06
MEgreenyellow	0.87	0.87
MEmidnightblue	0.96	0.96
MEpurple	0.96	0.96
MEdarkgreen	0.37	0.37
MEred	0.19	0.19
MEdarkturquoise	0.00028	0.00028
MEdarkorange	0.00049	0.00049
MEorange	0.44	0.44
MEdarkgrey	0.0013	0.0013
MEdarkred	0.0024	0.0024
MEroyalblue	0.0029	0.0029
MEgrey60	0.0055	0.0055
MEpink	0.00012	0.00012
MEyellow	0.44	0.44
MEsalmon	0.0048	0.0048
MElightgreen	6e−05	6e−05
MEgreen	2.8e−11	2.8e−11
MEmagenta	2.5e−15	2.5e−15
MEblack	4e−11	4e−11
MEtan	1.2e−07	1.2e−07

Fig. 2-5. Differential expression analysis and weighted gene co-expression network analysis underscores cell cycle enrichment and reveals distinct modules in both clusters. (A) Gene expression from the gliomasphere dataset was assessed through differential expression analysis. Cluster 1 represents samples in the E2F1-enriched cluster. Two examples from Gene Ontology are shown, depicting high enrichment for cell cycle terms. **(B)** WGCNA generated 26 modules when samples were analyzed based on their enrichment profiles for the gene lists. **(C)** Modules are ranked based on their abundance in both clusters. Modules at the top are highly enriched in the E2F1-activated cluster (cluster 1).



HK217

Group	Stem cell frequency	p-value vs control
Control	1/27.2	-
E2F1	1/104.2	4.92e-10

HK301

Group	Stem cell frequency	p-value vs control
Control	1/154	-
E2F1	1/391	0.0005

HK357

Group	Stem cell frequency	p-value vs control
Control	1/45.9	-
E2F1	1/46.1	NS

HK408

Group	Stem cell frequency	p-value vs control
Control	1/27.4	-
E2F1	1/33	NS

Fig. 2-6. E2F1 silencing compromises self-renewal and proliferation *in vitro* and tumor formation *in vivo*. (A) Samples from both clusters were treated with control (scrambled) or E2F1 siRNA and plated under limiting dilution in a 96-well plate. Graphs depict the number of wells that did not form spheres after 10 days vs. the number of cells plated (a vertical line implies all wells formed spheres). **(B)** Cells treated with scrambled or E2F1 siRNA were plated at a density of 2,000 cells per well in a 96-well plate in quadruplicate, and their growth was evaluated using luminescence. Relative growth is the fold change compared to basal measurement. **(C)** Cells from the E2F1-activated cluster were treated with two drugs identified using IPA, and their sphere formation capacity was evaluated after 10 days. **(C)** Cells treated

with scrambled or E2F1 shRNA were intracranially injected in NSG mice. Luminescence was assessed four weeks after transplantation. **(D)** Quantification for each group in **(C)** is shown. Experiments in **(A)** and **(B)** were performed at least three times. Data are represented as mean +/- SEM. **p<0.01 and ***p<0.001 as assessed by one-way ANOVA.

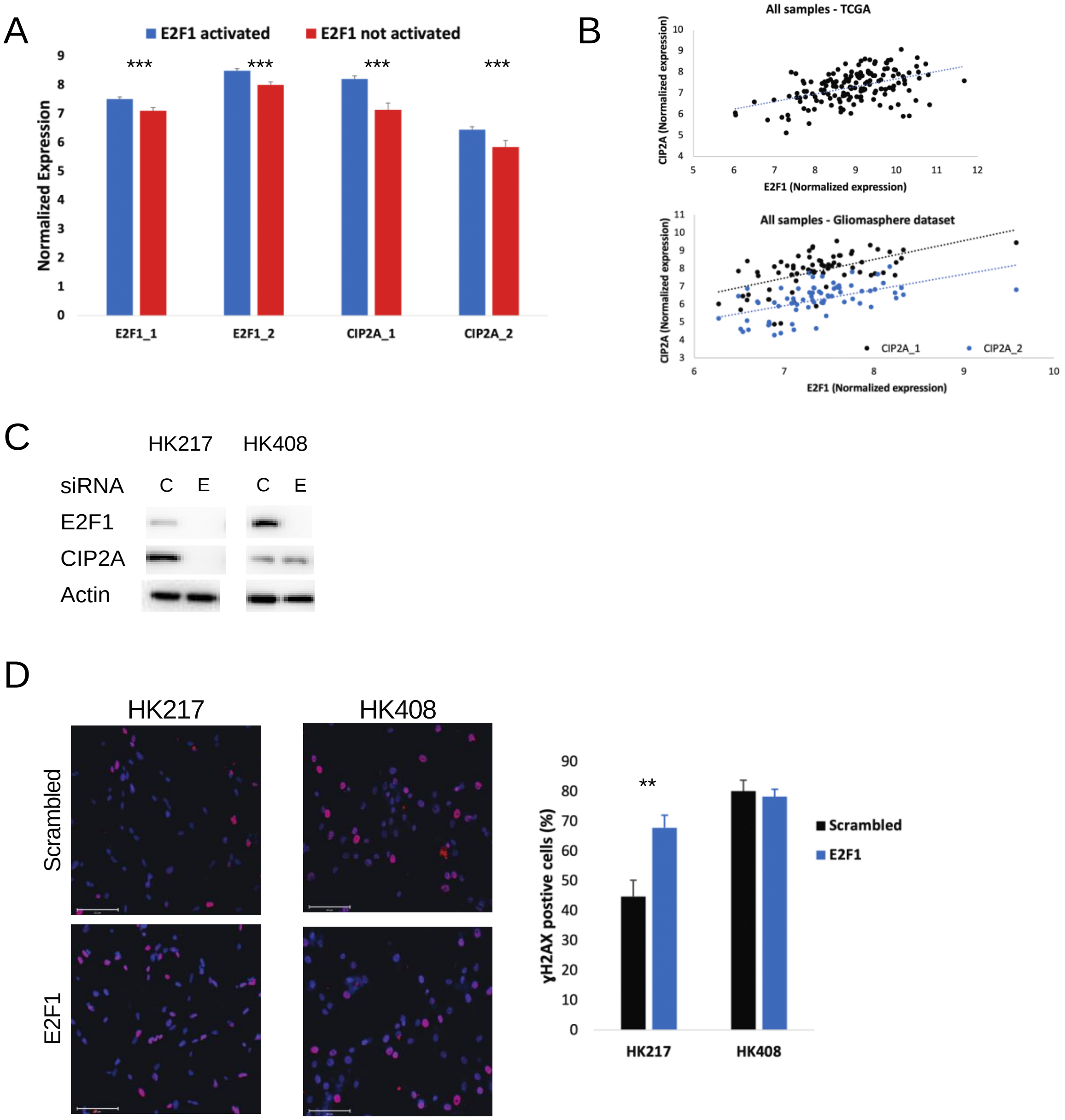
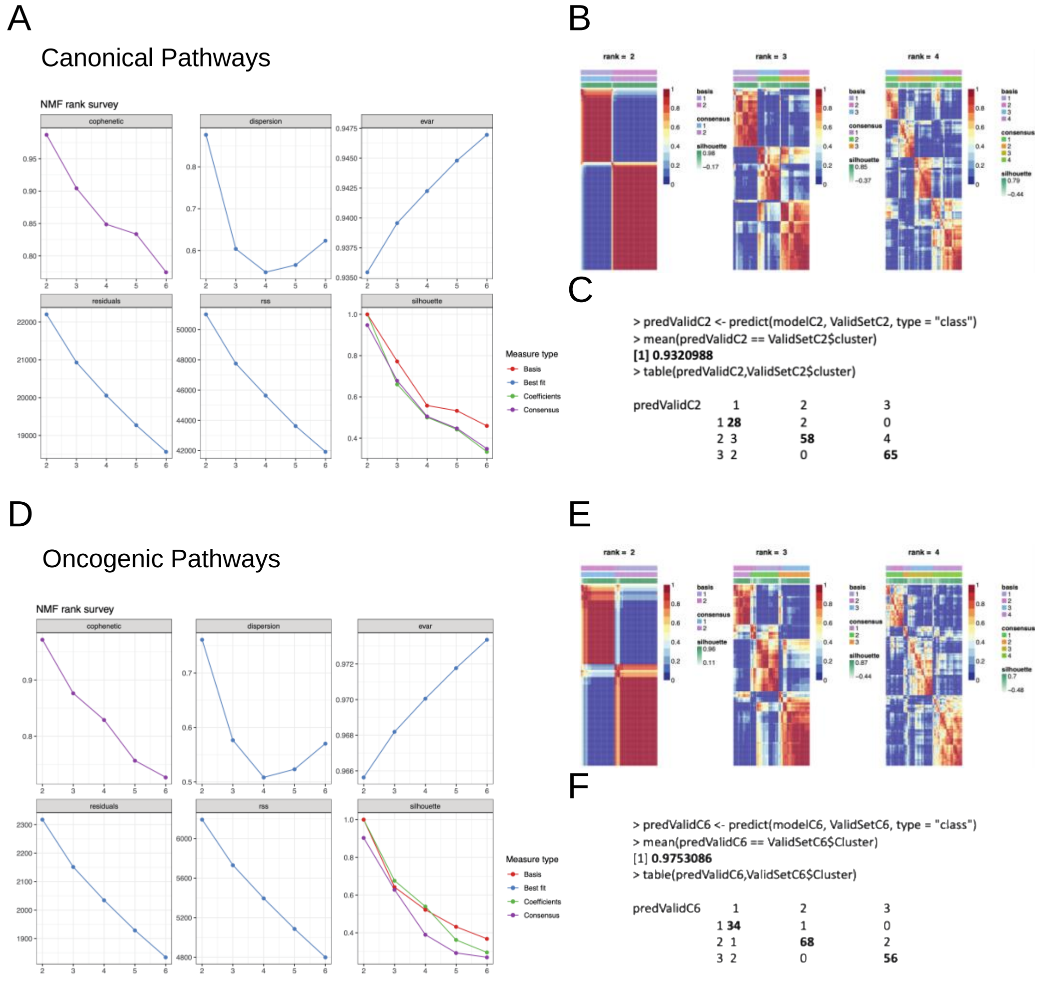


Fig. 2-7. E2F1 silencing compromises DNA damage response induced after irradiation. (A) Gene expression levels of *E2F1* and *CIP2A* in the gliomasphere dataset. **(B)** TCGA (top)

and gliomasphere (bottom) samples were plotted based on their expression levels of *E2F1* and *CIP2A* (two different probes in the gliomasphere dataset). Trend lines show significant positive correlation > 0.5. **(C)** Cells were treated with either C (control) or E (E2F1) shRNA, and protein was assessed for E2F1 and CIP2A levels. Actin serves as the loading control. **(D)** Treated cells were subjected to irradiation (8 Gy) and fixed after 12 hours for γH2AX staining (red). Nuclei were counterstained using DAPI. Quantification for each group is shown on the right. Experiments in **(C)** and **(D)** were performed at least two times. Data are represented as mean +/- SEM. **p<0.01 and ***p<0.001 as assessed by one-way ANOVA.

Supplementary Figures

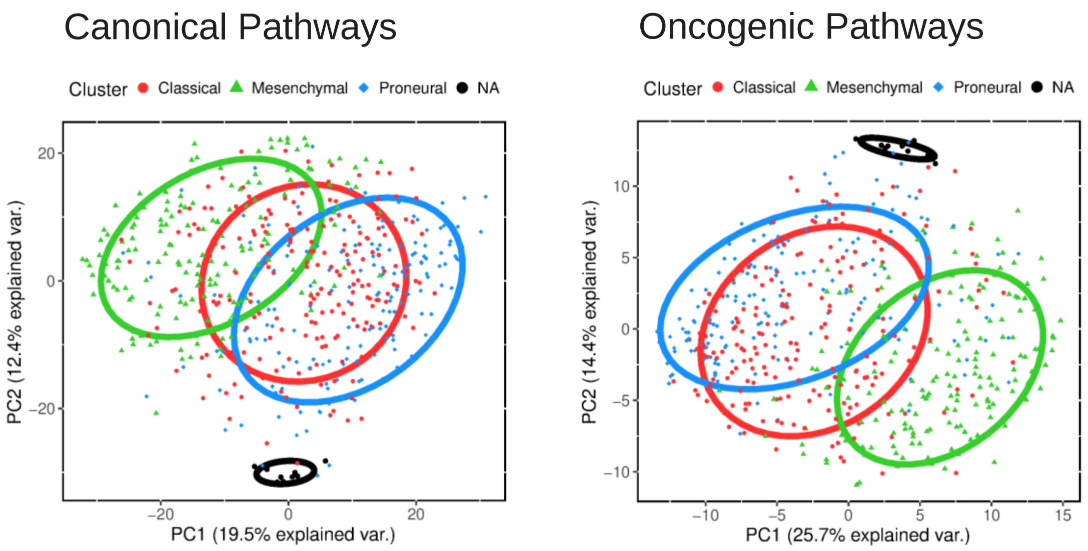


```
> predValidC2 <- predict(modelC2, ValidSetC2, type = "class")
> mean(predValidC2 == ValidSetC2$cluster)
[1] 0.9320988
> table(predValidC2,ValidSetC2$cluster)

predValidC2     1        2         3
            1 28        2         0
            2  3       58         4
            3  2        0        65
```

```
> predValidC6 <- predict(modelC6, ValidSetC6, type = "class")
> mean(predValidC6 == ValidSetC6$Cluster)
[1] 0.9753086
> table(predValidC6,ValidSetC6$Cluster)

predValidC6     1        2         3
            1 34        1         0
            2  1       68         2
            3  2        0        56
```

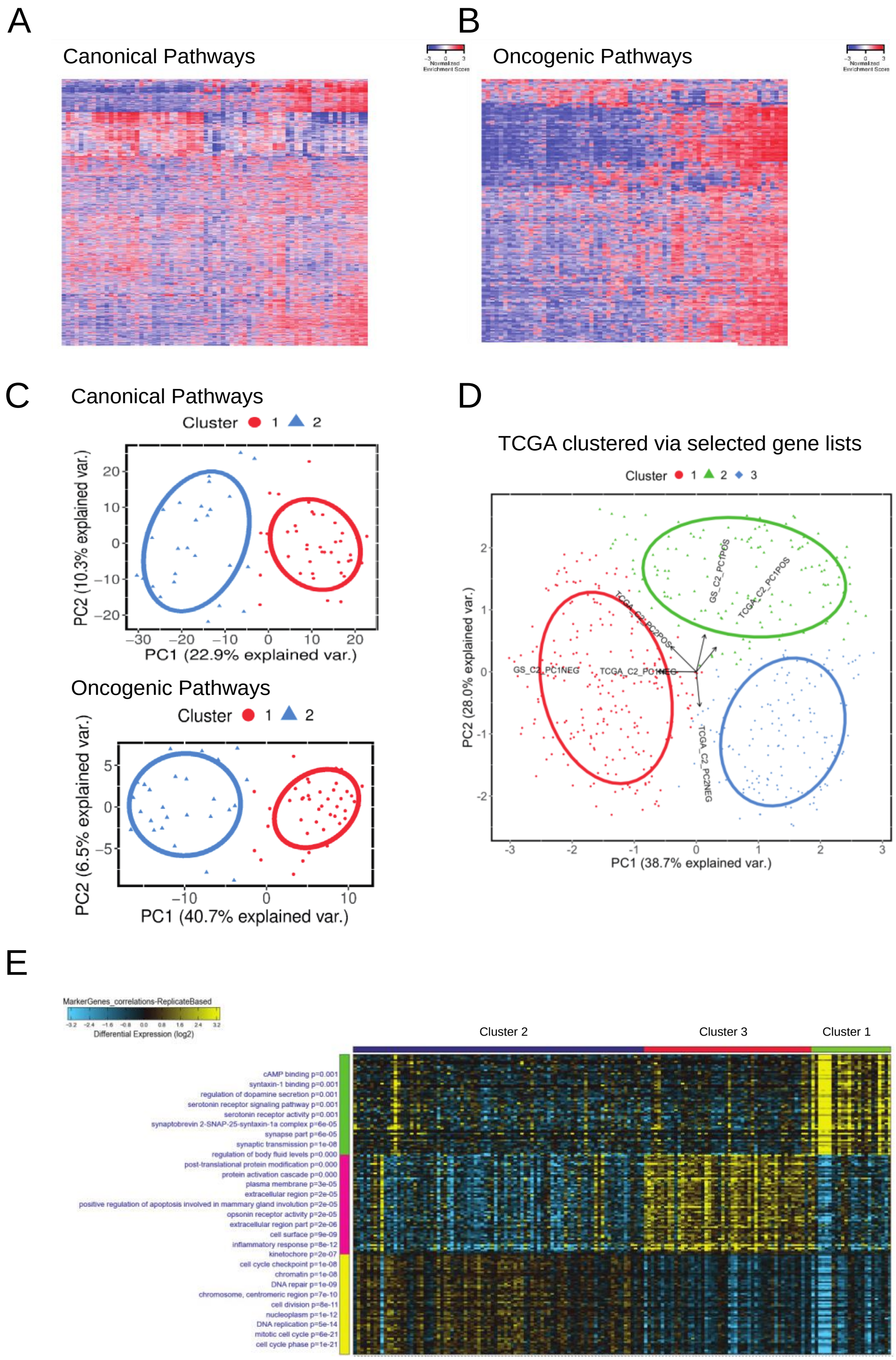
Supplementary Fig. 2-1.

Consensus clustering identifies three clusters for the TCGA dataset. (A and D) Non-negative matrix factorization (NMF) was performed with ranks ranging from 2 to 6. Various metrics are shown for canonical **(A)** and oncogenic **(D)** gene set collections. **(B and E)** Consensus clustering based on NMF with ranks 2 through 4, showing general structure for each rank in both canonical **(B)** and oncogenic **(E)** collections. **(C and F)** Random Forest classifier was used to validate the robustness of the number of clusters selected. For each analysis, canonical **(C)** and oncogenic **(F)**, training sets were generated and shown are the results for the correct clustering in the validations sets.



Supplementary Fig. 2-2.

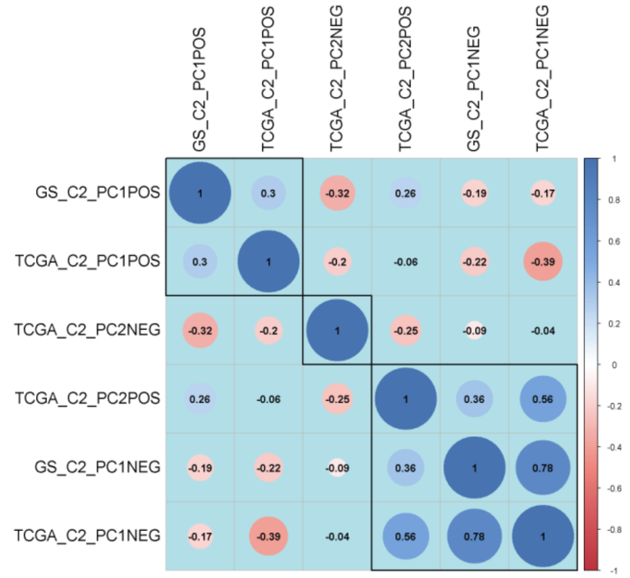
Pathway-based clustering does not overlap with molecular classification. Principal component analysis plots generated in Figure 2-2C and 2-2D were colored by molecular subtype in canonical (left) and oncogenic (right), respectively. Samples for which no subtype was determined are labeled as NA (black). Circle lines represent the normal distribution of the samples in each cluster.


A
Canonical Pathways
-3 0 3
Normalized Enrichment Scores
B
Oncogenic Pathways
-3 0 3
Normalized Enrichment Scores
C
Canonical Pathways
Cluster ● 1 ▲ 2
PC2 (10.3% explained var.)
20
10
0
−10
−20
−30 −20 −10 0 10 20
PC1 (22.9% explained var.)
Oncogenic Pathways
Cluster ● 1 ▲ 2
PC2 (6.5% explained var.)
5
0
−5
−10 0 10
PC1 (40.7% explained var.)
D
TCGA clustered via selected gene lists
Cluster ● 1 ▲ 2 ◆ 3
GS_C2_PC1POS
TCGA_C2_PC1POS
TCGA_C2_PC2POS
GS_C2_PC1NEG
TCGA_C2_PC1NEG
TCGA_C2_PC2NEG
PC2 (28.0% explained var.)
2
1
0
-1
-2
-3 -2 -1 0 1 2 3
PC1 (38.7% explained var.)
E
MarkerGenes_correlations-ReplicateBased
-3.2 -2.4 -1.6 -0.8 0.0 0.8 1.6 2.4 3.2
Differential Expression (log2)
Cluster 2
Cluster 3
Cluster 1
cAMP binding p=0.001
syntaxin-1 binding p=0.001
regulation of dopamine secretion p=0.001
serotonin receptor signaling pathway p=0.001
serotonin receptor activity p=0.001
synaptobrevin 2-SNAP-25-syntaxin-1a complex p=6e-05
synapse part p=6e-05
synaptic transmission p=1e-08
regulation of body fluid levels p=0.000
post-translational protein modification p=0.000
protein activation cascade p=0.000
plasma membrane p=3e-05
extracellular region p=2e-05
positive regulation of apoptosis involved in mammary gland involution p=2e-05
opsonin receptor activity p=2e-05
extracellular region part p=2e-06
cell surface p=9e-09
inflammatory response p=8e-12
kinetochore p=2e-07
cell cycle checkpoint p=1e-08
chromatin p=1e-08
DNA repair p=1e-09
chromosome, centromeric region p=7e-10
cell division p=8e-11
nucleoplasm p=1e-12
DNA replication p=5e-14
mitotic cell cycle p=6e-21
cell cycle phase p=1e-21

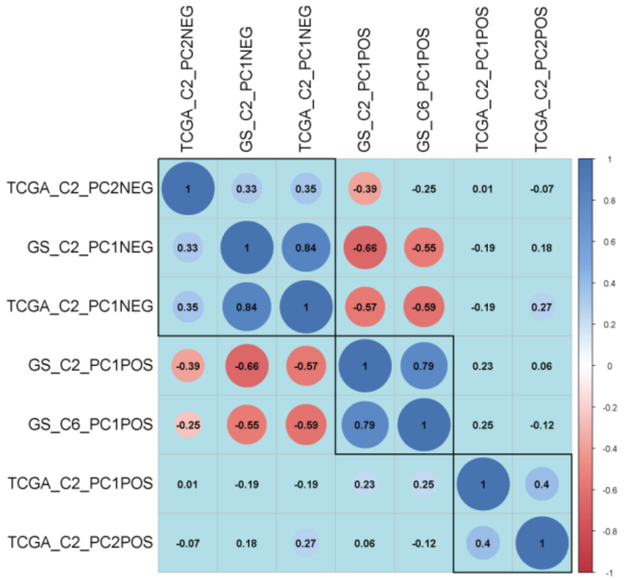
Supplementary Fig. 2-3.

Gliomasphere dataset analysis generates two clusters and TCGA gene ontology analysis of clusters shows differentially enriched terms. Samples from a 70-sample patient-derived gliomasphere dataset were analyzed using canonical **(A)** and oncogenic **(B)** pathways from the GSEA to generate heatmaps based on the enrichment profile of each sample (column) with respect to each gene set (row). **(C)** Profiles from **(A)** at the top and **(B)** at the bottom were used to generate PCA plots labeled by color and shape of each cluster. Circle lines represent the normal distribution of the samples in each cluster. **(D)** TCGA samples were reanalyzed utilizing the gene lists generated in Figure 2-4 to obtain enrichment profiles for each sample (not shown) and the corresponding shown PCA plot. The arrows show the contribution of each gene list to a particular direction in the plot. **(E)** The three clusters generated for the TCGA dataset were analyzed using differential expression analysis and then analyzed for the enrichment of gene ontology terms in each cluster. The plot shows the terms most highly enriched in a particular cluster compared with the other two. The expression of marker genes in each cluster is shown for every sample.

TCGA dataset

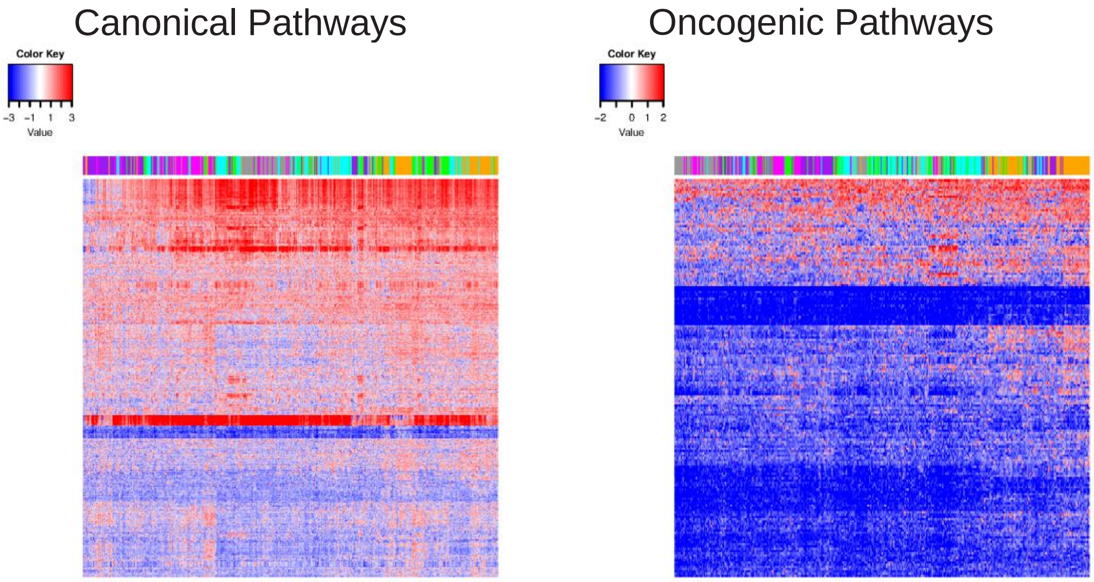


Gliomasphere dataset



Supplementary Fig. 2-4.

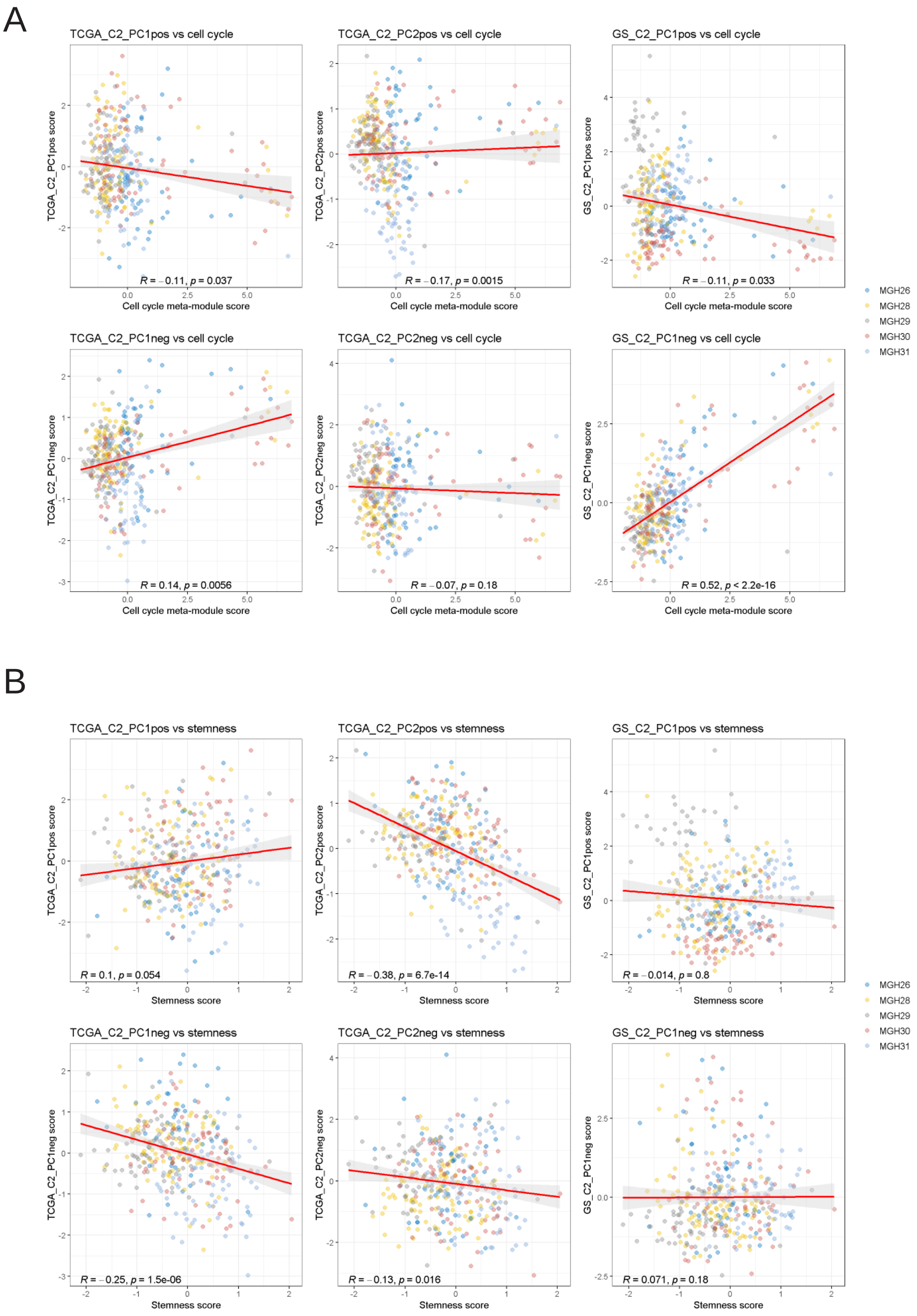
Correlation between gene lists in TCGA and gliomasphere datasets. The generated gene lists were used to obtain scores for each sample in both TCGA (top) and GS (bottom) datasets. Correlations between these scores in all samples for either dataset are shown, where the presence of a circle represents significant correlation and the size and color depict the strength of the correlation. Boxes mark groups of gene lists strongly correlated (based on hierarchical clustering).


Canonical Pathways

Oncogenic Pathways

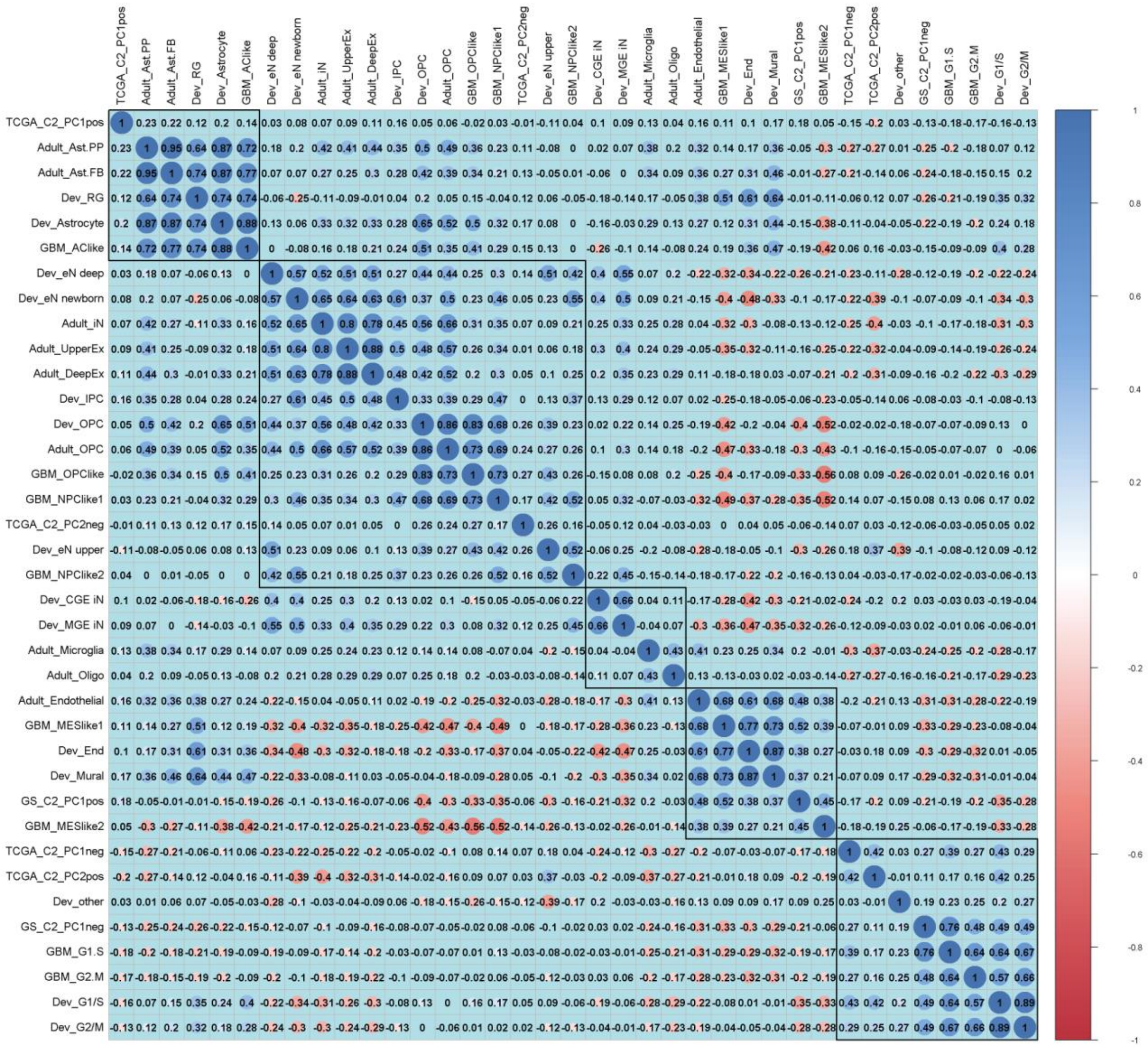
Supplementary Fig. 2-5.

Pathway-based clustering in single cell dataset reveals low intratumoral heterogeneity. Samples from a single cell RNA-seq database were analyzed using either canonical pathways (left) or oncogenic pathways (right) from the Gene Set Enrichment Analysis to generate heatmaps based on the enrichment profile of each sample (column) with respect to each gene set (row) in both collections. Colors at the top represent the sample to which each cell belongs (MGH26 = grey, MGH264 = magenta, MGH28 = cyan, MGH29 = orange, MGH30 = green, MGH31 = purple; grey and magenta overlap the most because MGH26 and MGH264 are different analyses from the same sample).


A
TCGA_C2_PC1pos vs cell cycle
R = −0.11, p = 0.037
TCGA_C2_PC2pos vs cell cycle
R = −0.17, p = 0.0015
GS_C2_PC1pos vs cell cycle
R = −0.11, p = 0.033
TCGA_C2_PC1neg vs cell cycle
R = 0.14, p = 0.0056
TCGA_C2_PC2neg vs cell cycle
R = −0.07, p = 0.18
GS_C2_PC1neg vs cell cycle
R = 0.52, p < 2.2e-16
Cell cycle meta-module score
TCGA_C2_PC1pos score
TCGA_C2_PC2pos score
GS_C2_PC1pos score
TCGA_C2_PC1neg score
TCGA_C2_PC2neg score
GS_C2_PC1neg score
MGH26
MGH28
MGH29
MGH30
MGH31
B
TCGA_C2_PC1pos vs stemness
R = 0.1, p = 0.054
TCGA_C2_PC2pos vs stemness
R = −0.38, p = 6.7e-14
GS_C2_PC1pos vs stemness
R = −0.014, p = 0.8
TCGA_C2_PC1neg vs stemness
R = −0.25, p = 1.5e-06
TCGA_C2_PC2neg vs stemness
R = −0.13, p = 0.016
GS_C2_PC1neg vs stemness
R = 0.071, p = 0.18
Stemness score

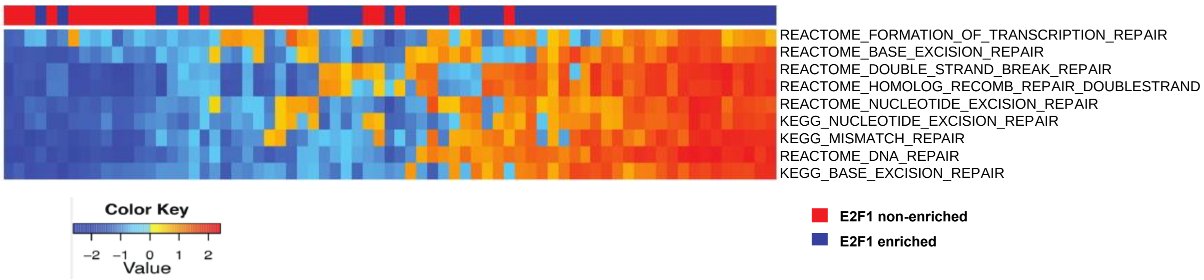
Supplementary Fig. 2-6.

Gene lists differentially correlate with cell cycle and stemness signatures. (A and B) Gene lists were used to obtain scores for each cell in the scRNA-seq dataset [17]. Similarly, stemness and cell cycle score scores were generated using the signatures reported in the same paper. For each cell, we plotted the scores for cell cycle **(A)** or stemness **(B)** signatures against the scores for all the gene lists. Cells are colored by tumor of origin and p-values and correlation coefficients are shown for each plot.



Supplementary Fig. 2-7.

Gene lists differentially correlate with cellular states and cell specific markers. Scores generated for each cell in Supplementary Figure 2-5 using gene lists were correlated with scores for cellular states ([11]) and specific cell type markers in development and adult brain ([31, 32]). The presence of a circle represents significant correlation, and the size and color depict the intensity of the correlation. Boxes mark groups of gene lists strongly correlated (based on hierarchical clustering).



Supplementary Fig. 2-8.

E2F1-enriched samples have high enrichment scores for DNA repair gene sets. Samples in the gliomasphere dataset from both clusters (determined in Figure 2-4) were evaluated for their enrichment of DNA repair gene sets from the canonical pathway collection.

References

1. D. R. Johnson, B. P. O'Neill, Glioblastoma survival in the United States before and during the temozolomide era. *J Neurooncol* **107**, 359-364 (2012).

2. R. Stupp *et al.*, Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* **10**, 459-466 (2009).

3. H. S. Phillips *et al.*, Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157-173 (2006).

4. W. A. Freije *et al.*, Gene expression profiling of gliomas strongly predicts survival. *Cancer Res* **64**, 6503-6510 (2004).

5. R. G. Verhaak *et al.*, Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98-110 (2010).

6. C. W. Brennan *et al.*, The somatic genomic landscape of glioblastoma. *Cell* **155**, 462-477 (2013).

7. H. Yan *et al.*, IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773 (2009).
8. H. Noushmehr *et al.*, Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510-522 (2010).

9. J. R. D. Pearson, T. Regad, Targeting cellular pathways in glioblastoma multiforme. *Signal Transduct Target Ther* **2**, 17040 (2017).

10. D. Akhavan *et al.*, De-repression of PDGFRβ transcription promotes acquired resistance to EGFR tyrosine kinase inhibitors in glioblastoma patients. *Cancer Discov* **3**, 534-547 (2013).

11. C. Neftel *et al.*, An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e821 (2019).

12. A. Bhaduri *et al.*, Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* **26**, 48-63.e46 (2020).

13. V. K. Mootha *et al.*, PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).

14. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

15. S. M. Johnson *et al.*, RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635-647 (2005).

16. S. C. Chiu *et al.*, Therapeutic potential of microRNA let-7: tumor suppression or impeding normal stemness. *Cell Transplant* **23**, 459-469 (2014).

17. A. P. Patel *et al.*, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396-1401 (2014).

18. B. Rollbrocker, A. Waha, D. N. Louis, O. D. Wiestler, A. von Deimling, Amplification of the cyclin-dependent kinase 4 (CDK4) gene is associated with high cdk4 protein levels in glioblastoma multiforme. *Acta Neuropathol* **92**, 70-74 (1996).

19. C. Hua, G. Zhao, Y. Li, L. Bie, Minichromosome Maintenance (MCM) Family as potential diagnostic and prognostic tumor markers for human gliomas. *BMC Cancer* **14**, 526 (2014).

20. H. Shen, K. Cook, H. E. Gee, E. Hau, Hypoxia, metabolism, and the circadian clock: new links to overcome radiation resistance in high-grade gliomas. *J Exp Clin Cancer Res* **39**, 129 (2020).

21. L. Giunti *et al.*, A microRNA profile of pediatric glioblastoma: The role of NUCKS1 upregulation. *Mol Clin Oncol* **10**, 331-338 (2019).

22. A. M. Knudsen *et al.*, Expression and prognostic value of the transcription factors EGR1 and EGR3 in gliomas. *Sci Rep* **10**, 9285 (2020).

23. N. Sakakini *et al.*, A Positive Feed-forward Loop Associating EGR1 and PDGFA Promotes Proliferation and Self-renewal in Glioblastoma Stem Cells. *J Biol Chem* **291**, 10684-10699 (2016).

24. H. D. Hemmati *et al.*, Cancerous stem cells can arise from pediatric brain tumors. *Proc Natl Acad Sci U S A* **100**, 15178-15183 (2003).

25. S. K. Singh *et al.*, Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63**, 5821-5828 (2003).

26. R. Galli *et al.*, Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res* **64**, 7011-7021 (2004).

27. T. N. Ignatova *et al.*, Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia* **39**, 193-206 (2002).

28. J. Chen *et al.*, A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522-526 (2012).

29. S. Bao *et al.*, Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756-760 (2006).

30. P. Guilhamon *et al.*, Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun* **4**, 2166 (2013).

31. D. Velmeshev *et al.*, Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).

32. T. J. Nowakowski *et al.*, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323 (2017).

33. L. Gutierrez, A. I. Magee, C. J. Marshall, J. F. Hancock, Post-translational processing of p21ras is two-step and involves carboxyl-methylation and carboxy-terminal proteolysis. *EMBO J* **8**, 1093-1098 (1989).

34. I. M. Ahearn, H. R. Court, F. Siddiqui, D. Abankwa, M. R. Philips, NRAS is unique among RAS proteins in requiring ICMT for trafficking to the plasma membrane. *Life Sci Alliance* **4**, (2021).

35. S. Li *et al.*, RAS/ERK signaling controls proneural genetic programs in cortical development and gliomagenesis. *J Neurosci* **34**, 2169-2190 (2014).

36. L. Ricci-Vitiani *et al.*, Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature* **468**, 824-828 (2010).

37. Y. Soda *et al.*, Transdifferentiation of glioblastoma cells into vascular endothelial cells. *Proc Natl Acad Sci U S A* **108**, 4274-4280 (2011).

38. L. Cheng *et al.*, Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. *Cell* **153**, 139-152 (2013).

39. S. Horvath *et al.*, Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A* **103**, 17402-17407 (2006).

40. R. M. Auvergne *et al.*, Transcriptional differences between normal and glioma-derived glial progenitor cells identify a core set of dysregulated genes. *Cell Rep* **3**, 2127-2141 (2013).

41. C. Park, I. Lee, W. K. Kang, E2F-1 is a critical modulator of cellular senescence in human cancer. *Int J Mol Med* **17**, 715-720 (2006).

42. A. Laine *et al.*, Senescence sensitivity of breast cancer cells is defined by positive feedback loop between CIP2A and E2F1. *Cancer Discov* **3**, 182-197 (2013).

43. L. Wang *et al.*, The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. *Cancer Discov* **9**, 1708-1719 (2019).

44. L. Garofano *et al.*, Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat Cancer* **2**, 141-156 (2021).

45. D. R. Laks *et al.*, Large-scale assessment of the gliomasphere model system. *Neuro Oncol* **18**, 1367-1378 (2016).

46. I. Tirosh *et al.*, Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313 (2016).

47. D. R. Laks *et al.*, Neurosphere formation is an independent predictor of clinical outcome in malignant glioma. *Stem Cells* **27**, 980-987 (2009).

48. R. C. Team. (2020).

Chapter 3:

Single nucleus RNA-sequencing reveals spatial and temporal signatures of enhancing and non-enhancing regions in high-grade glioma

Abstract

Both therapy for and molecular investigations into high-grade glioma largely focus on the contrast-enhancing (CE) portion of the tumor core. However, given the invasive nature of glioma, residual tumor cells responsible for recurrence likely exist in the peripheral, non-enhancing (NE) region. Single cell/single nucleus RNA-sequencing and subsequent bioinformatics analyses offer a flexible and scalable framework that allows powerful examination of tumor and non-tumor cell subpopulations across various contexts. In this study, we use pre-operative magnetic resonance images to prospectively identify biopsy targets in the CE region and NE sites that lie 0.5-2.0cm beyond the CE edge from multiple pathologically and molecularly distinct high-grade gliomas and leverage single nucleus RNA-sequencing to generate over 30,000 individual RNA profiles. We first establish three methods: an improved approach to accurately identify malignant glioma cells, a customizable wrapper function for visualization of marker expression profiles and hierarchies across clusters, and a new tool for generating copy number-based subcluster hierarchy plots that are annotated with pie charts underneath showing composition based on any given metadata variable. We then analyze multiple glioma cellular states, genetic subclones, and hierarchical relationships among malignant cells, all of which reveal both conserved and patient-specific features. There remains a significant tumor cell burden in the NE edge, and a subset of malignant cells in this region are actively proliferating. Systematic comparison of the CE and NE regions identifies altered malignant and non-malignant cell compartments, as well as cell type- and patient-specific differences in gene expression and pathway utilization profiles. Implementing an unsupervised machine learning approach for cell type prioritization reveals that MES-like and AC-like malignant cells are not only the most divergent populations between the two regions in the IDH-

WT GBMs, but also the only malignant populations whose intertumoral heterogeneity was significantly different between regions (lower in NE, implying convergence). Finally, we build an integrated atlas of GBM cell-cell communication networks among malignant and non-malignant subpopulations in the two regions and identify a significantly rewired connectome in the NE region, allowing us to pinpoint MES-like cells, altered vascular function, and neuro-immune interactions as potential drivers of regional divergence in glioblastoma.

Introduction

The intervention that provides the greatest contribution to overall survival in adult glioma remains surgical removal of tumor cells [1]. The current surgical standard of care is imaging-based resection of the contrast-enhancing (CE) region as depicted by T1-weighted magnetic resonance imaging (MRI) with gadolinium contrast [2]. The greater the extent of resection of the CE region, the greater the contribution to survival [3]. However, despite gross total resection of CE regions, high grade gliomas invariably recur, and this most commonly occurs at the edge of the resection cavity [4]. Given the invasive nature of glioma [5], these data suggest that glioma cells exist beyond the CE portion as seen on MRI, and that these cells are 1) not being removed by surgery and 2) capable of reconstituting the tumor despite standard of care adjuvant radiation and chemotherapy [2, 6, 7].

Strategies for molecular targeted therapy for glioma have largely been developed by studying bulk RNA or protein expression from cell lines or clinical surgical specimens, which, as described above, are typically derived from the CE portion of the tumor. This strategy fails to account for the mixed population of cell types in glioma and does not develop strategies aimed at the true targets for adjuvant therapy: residual cells beyond the CE portion of the tumor that are not resected. To counteract some of this criticism, several studies have utilized single cell sequencing to study intratumoral heterogeneity and hierarchical relationships between tumor cell types [8–14]. These studies have largely focused on single samples from CE regions of tumors, with very little data from regions beyond the CE edge that highlight the clinical relevance of intratumoral heterogeneity.

In this study, we prospectively select CE and NE targets from 6 high grade gliomas (4 IDH wild-type glioblastomas, 1 IDH mutant glioblastoma, 1 IDH mutant anaplastic oligodendroglioma) and 1 low grade (non-enhancing) oligodendroglioma. We couple intraoperative neuronavigation-guided targeted biopsy specimens from CE areas and areas beyond the edge of CE with single nucleus RNA-sequencing in order to evaluate a large number (32,914 after QC) of positionally localized cells. Our experimental design differs from previous single cell analysis in glioma in two fundamental ways: 1) we detail specific positional data to model cell populations on a spatial plane, and 2) we utilize nuclear sequencing and multiple biopsy specimens to obtain a high number of cells per sample. Our data provide a spatial landscape of tumor and non-tumor (glial, neuronal, immune, and endothelial) cell types as well as the gene expression, pathway activation, and interaction characteristics of these cell populations. We focus on the biopathology of glioma beyond what is currently the target of therapy, the CE region of the tumor. We reveal that these non-treated areas have significant tumor burden, proliferative capacity, differentially enriched genes and pathways, and rewired cell-cell interaction patterns. We specifically highlight MES-like cells as potentially undergoing a state switching event, with vascular remodeling enriched in the tumor core and proliferation enriched in the tumor periphery. Our analysis thus identifies specific target cells and pathways for future study and therapeutic development.

Results

Improved identification of malignant cells and subclones via snRNA-seq and CNA inference

We prospectively identified biopsy targets using pre-operative MRI in all patients undergoing surgery for glioma. For each of the 6 high grade gliomas (5 GBMs, 1 anaplastic oligodendroglioma), we selected 3 biopsy targets: 2 from the contrast-enhancing (CE) core of the tumor and 1 from the non-enhancing (NE) edge of the tumor, located 5-20mm from the CE edge (2 examples shown in **Fig. 3-1a**). Clinical pathology and fluorescence in situ hybridization (FISH) results identified: one IDH1 R132H 1p/19q co-deleted WHO III anaplastic oligodendroglioma, one IDH1 R132H EGFR-amplified WHO IV GBM, one IDH wild-type PTEN-deleted WHO IV GBM, three IDH1 wild-type WHO IV GBMs, and one IDH1 R132H 1p/19q co-deleted WHO II oligodendroglioma (**Supplementary Fig. 3-1a**). We refer to the IDH wild-type GBMs as “IDH-WT GBM,” the IDH mutant GBM as “IDH-mut GBM,” the grade III anaplastic oligodendroglioma as “GIII-O,” and the low grade oligodendroglioma as “LGG.” We isolated nuclei from each biopsy specimen and performed single nucleus RNA-sequencing (snRNA-seq) using the 10X Genomics workflow, yielding 32,914 individual nuclear transcriptomes that passed quality control filtering and that were included in the subsequent analyses (**Fig. 3-1b, Supplementary Fig. 3-1b-d**). To account for batch effects, we performed integration to generate one combined dataset, followed by graph-based clustering using the Louvain method for community detection (**Fig. 3-1c-d, Supplementary Fig. 3-1e**).

Tumor specimens typically comprise both malignant and non-malignant cells, so we first sought to identify malignant cells in our dataset. This is typically done by examining cluster-based canonical cell type expression patterns and/or copy number alteration (CNA) profiles. To build on these approaches, we developed a multi-step strategy that aims to improve accuracy and reduce noise (**Fig. 3-2**). In this workflow, we first examined canonical brain cell type marker expression patterns and hierarchical relationships across the clusters identified in each patient to assess whether putative normal brain cell types can be readily identified (**Fig. 3-2a**). Next, we performed an iterative sequence of single cell CNA inference analysis using the inferCNV R package [15] (**Fig. 3-2c**; see **Methods**). CNA inference was first performed without a reference group of cells, and the cells were grouped by original expression cluster to examine cluster-specific CNA profiles. Next, to resolve the observed heterogeneity within clusters, the cells were grouped by CNA profile via hierarchical clustering. In our dataset, canonical glioma CNA events were clearly visualized in each patient, and cells separated based on these events. Also, the resulting CNA clusters were mostly composed of cells belonging to similar expression-based clusters (second vertical color bar from the left in each heatmap), allowing confident identification of non-malignant cells as those belonging to both putative normal RNA clusters and non-malignant CNA clusters. Prior to the next step, myeloid cells (positive for myeloid markers and chromosome 6 amplification in MHC region) were removed to focus on brain lineages. CNA inference was then repeated using a normal cell type as reference, both to remove bias from composition differences across tumors and to identify CNA events relative to normal cells rather than to the average cell. Normal oligodendrocytes were selected as the reference cells since they were the only non-malignant, brain derived (i.e., non-immune) cell type present in all patient datasets. Cells were again grouped by CNA profiles, and cells belonging to clusters exhibiting canonical glioma CNA events were labeled as malignant. Finally, CNA inference was performed using only the malignant cells identified in the previous step and normal oligodendrocytes as reference. Consistent with previous studies [9-11], hierarchical clustering using the final malignant cell CNA profiles from each tumor identified multiple groups with specific CNA events, indicating the presence of distinct subclones (**Fig. 3-2c-d**). By developing this technique that integrates multiple rounds of CNA inference with the original transcriptome-based clustering information, we address several problematic types of cells: 1) malignant cells that, due to their atypical transcriptomes, belonged to putative normal RNA clusters (**Supplementary Fig. 3-2a-b**), 2) normal cells belonging to putative malignant or unclear RNA clusters, and 3) cells with noisy or unclear expression or CNA profiles, which we removed at each step by identifying small outlier groups (**Supplementary Fig. 3-2c**). For all subsequent analyses, malignant cells (**Fig. 3-2e**) were separated from non-malignant cells, and both datasets were processed again separately (see **Methods**).

Analysis of cellular states in glioma highlights intra- and inter-tumoral heterogeneity

We then sought to group malignant cells based on conserved drivers of intratumoral heterogeneity. Previous studies have identified multiple malignant cell types in gliomas, with marker expression profiles consistent with multiple differentiated brain lineages (e.g. oligodendrocytes, astrocytes, neural progenitor cells) [9-12, 16]. Neftel et al. (2019) presented an integrated model of glioblastoma malignant cellular states defined by distinct gene networks that robustly characterize intratumoral heterogeneity across multiple large glioblastoma cohorts. We implemented their proposed molecular classification strategy to stratify malignant glioma cells in our dataset based on their transcriptomic profiles. We generated single cell enrichment scores for each of the reported gene networks (meta-modules) and classified malignant cells according to the four cellular states described: astrocyte-like (AC-like), mesenchymal-like (MES-like), neural progenitor cell-like (NPC-like), and oligodendrocyte precursor cell-like (OPC-like) (**Fig. 3-3a-b**; see **Methods**). We also generated single cell enrichment scores for the two reported cell cycle programs, corresponding to G1/S and G2/M phases, and identified cycling cells based on the distribution of scores, as described (see **Methods**). To validate this classification scheme in our dataset, we then identified patterns of functional enrichment in each malignant cell population (**Fig. 3-3f**). Significantly enriched pathways generally corresponded to the cell types that each malignant population resembles, confirming that these specific gene lists are separating malignant cells based on expected functional patterns.

We then characterized intertumoral heterogeneity in malignant cell composition by performing differential abundance analysis across patients. Most tumors were composed of all malignant cellular states (**Fig. 3-3c**), but the proportion of each population was significantly different across tumors (**Fig. 3-3d**). As expected, the two oligodendrogliomas, TB76 and TB81, had the highest proportion of OPC-like cells; relative to the low grade tumor, the high grade oligodendroglioma had a higher proportion of AC-like and cycling cells and a lower proportion of OPC-like cells, suggesting the progression from low to high grade may be driven by OPC-to-AC conversion in oligodendrogliomas. MES-like cells were specific to glioblastomas (<5 cells in each oligodendroglioma). Interestingly, one tumor (PT95) was composed almost entirely (98%) of MES-like and cycling cells. The IDH-mut glioblastoma had the lowest proportion of NPC-like cells. The proportion of cycling malignant cells was highest in the four IDH-WT glioblastomas and was <10% in all tumors except PT90. In terms of which cellular states tend to be cycling, NPC-like and OPC-like cells had the two highest cycling proportions in all tumors (except PT95, which had <5 OPC-like cells) (**Fig. 3-3e**). However, we noted that each cell state had a cycling compartment in most, if not all, patients and thus may contain cells with the potential to initiate tumor recurrence.

We then sought to evaluate malignant cell populations more rigorously in terms of their contribution to intertumoral heterogeneity. While differential gene expression and pathway-level analyses offer tremendous insight when comparing cell populations, important complex patterns may not be readily identified when cell type abundances are not balanced across groups. More populous groups are likely to exhibit more differential expression that reaches statistical significance, even if they are not the most perturbed across conditions. To combat this, we reasoned that rather than comparing amount of differential expression, we could rank malignant cell populations based on broader transcriptomic patterns that may drive functional divergence across patients at the single cell level. To evaluate malignant cell populations based on variability across patients, we used Augur [17], which implements a machine learning algorithm that trains a classifier and determines how accurately cells from each cell type can be classified according to the experimental condition they belong to. Cell types are then ranked by their classification accuracy, which describes how "separable" they are by condition. To evaluate intertumoral heterogeneity, we instructed the classifier to split malignant cells based on cellular state and to evaluate how accurately the cells in each state could be labeled according to the patient that they were isolated from. Across the four cellular states, the classification accuracy was very high and nearly identical (**Supplementary Fig. 3-3a**, left). We then performed the reverse analysis, instructing the classifier to split the dataset by patient identity and attempt to predict the cellular states of cells within each patient. The resulting accuracies were all lower (**Supplementary Fig. 3-3a**, right) than the previous classification attempt, indicating greater separability (higher accuracy) when comparing across patients rather than across cell states within patients. This analysis clearly indicates that there is more inter-tumoral heterogeneity than intra-tumoral heterogeneity, at least with cellular state as the secondary grouping variable.

One area of controversy relates to whether there exists a phenotypic hierarchy amongst different malignant cell populations, with the neural stem cell-like cells potentially giving rise to the other cell types. Investigations in this area are often based on transcriptomic analysis. However, hierarchy can be more definitively delineated by examining CNA profiles across groups of cells, as groups exhibiting CNA events that are supersets of other groups (i.e., the same alterations and more) are likely downstream in the evolutionary hierarchy. To assess whether the different malignant cell populations in this cohort might exist along an evolutionary hierarchy, we first identified CNA subclusters in each tumor by repeating the CNA inference procedure described above using a tumor subcluster partitioning method based on permutation and tree construction (see **Methods**). We then sought to visualize the tree for each patient and superimpose other useful information, such as cellular state composition in each subcluster. However, we found that methods to visualize hierarchies based on these subcluster identities are currently lacking, with common tools requiring manual annotation and image editing in other software programs. We thus developed a customizable analytical tool to first convert tumor CNA subcluster information into a convenient nomenclature for plotting hierarchy trees, then visualize hierarchical relationships based on these subclusters with simultaneous superimposition of other types of metadata information present in the sequencing dataset. This method can be applied to any cancer sc/snRNA-seq analysis and can thus enable more advanced investigations in many research areas. We applied our tool to this glioma dataset to examine each patient's CNA subcluster composition and subcluster-specific compositions of malignant cellular states and biopsy regions (**Supplementary Fig. 3-3b-d**). We observed that, without any user bias, the random tree algorithm identified 8 CNA-based subclusters in the malignant compartments of each tumor, and that each of these subclusters was a terminal node in the branching hierarchy (after 3 branching events). This suggests that none of the malignant populations in this cohort lie upstream or downstream of other malignant populations in terms of evolutionary hierarchy. As these are recurrent tumors, this is potentially owing to malignant cells having advanced after fully diverged in the past. We also noted that the subclusters within each tumor had varying profiles in terms of cellular state composition. Strikingly, a vast majority of the subclusters identified across all tumors were dominated by one cellular state, with only a few being more evenly mixed. By contrast, the tumors analyzed in Neftel et al. (2019) did not show significant interaction between cellular state and CNA subclone. Again, this may be related to the difference between recurrent and primary gliomas, suggesting malignant populations may diverge significantly throughout the disease course and could exist in multiple stable groups by the time recurrence occurs, in contrast to the significant state plasticity observed in primary gliomas [10].

Beyond the CE core of glioma: significant tumor burden and altered cell type composition

We next used positional information to compare cellular and molecular profiles in the contrast-enhancing (CE)/core and non-enhancing (NE)/edge regions of glioma. For each tumor, there were two samples taken from CE regions of the tumor and one from an NE, FLAIR-positive region beyond the CE edge. We filtered our dataset to include only the 5 tumors for which we successfully recovered nuclear transcriptomes from both CE and NE biopsy specimens, which were TB76 (GIII-O), TB77 (IDH-mut GBM), TB78 (IDH-WT GBM), PT90 (IDH-WT GBM), and PT94 (IDH-WT GBM). We first grouped malignant cells based on the region they were isolated from and assessed tumor burden in the NE region by examining the presence and quantity of malignant cells. We observed that malignant cells were indeed present in the NE region in all patients, and that there were generally fewer malignant cells in the NE region than the CE region (**Fig. 3-4a**). Malignant cell proportion in the CE region was highest in the GIII-O, followed by the three IDH-WT GBMs, and lowest in the IDH-mut GBM (**Fig. 3-4b**). In the NE region, the IDH-WT GBMs had the three highest malignant cell proportions, followed by the GIII-O and lastly the IDH-mut GBM. Malignant cell proportion was similar or higher in NE compared to CE in the IDH-WT GBMs and lower in the IDH-mut GBM and GIII-O.

Given that malignant cells comprised a significant proportion of the samples from the NE regions, we sought to model the true tumor burden on MR imaging. We used three tumors for this analysis: TB76 (GIII-O), TB77 (IDH-mut GBM), and TB78 (IDH-WT GBM). Using the quantity of tumor cells from three points of interest in each tumor, we applied a standard quadratic equation (see **Methods**) to extrapolate the distance from the center of the tumor to where the tumor burden would approach 0 (**Fig. 3-4c**). This model predicted presence of tumor cells 2.26-3.09 cm from the center of the tumor, and a spheroid with this radius was constructed and overlaid on the corresponding pre-operative MR image (**Fig. 3-4d**). It was clear from this overlay that the predicted tumor edge was beyond the CE region of the tumor, and in the three tumors was predicted to be 1.41, 1.45, and 1.59 cm beyond the CE edge, still within the region of FLAIR positivity. The invasion and growth patterns of glioma are complex, with cases of multifocal and distant lesions [18], but our data suggest a consistent presence of tumor burden immediately surrounding the CE bulk that is often not clinically addressed.

Despite integration to correct for technical variation, we noted from the low dimensional plots (**Fig. 3-4a**) that there was altered cluster composition in the NE region compared to the CE region, which was confirmed by comparing the cluster breakdown across regions (**Supplementary Fig. 3-4a**), suggesting significant transcriptional and compositional differences across regions. Following up on these changes, we performed differential abundance analysis to identify regional differences in proportions of cellular states. In the NE regions of the three IDH-WT GBMs, the proportions of MES-like malignant cells were decreased, and the proportions of NPC-like malignant cells were increased. In the NE region of the IDH-mut GBM, OPC-like cells were decreased, and AC-like cells were increased; the opposite was observed in the GIII-O.

Since actively cycling cells are likely the main contributors to tumor progression and recurrence, we then compared the two regions in terms of cycling tendency using the previously generated annotations. The proportion of cycling malignant cells was not significantly different overall, and there were variable differences within each patient dataset (**Fig. 3-4f**). To characterize cycling via immunohistochemistry (IHC), we obtained image-guided biopsies in an additional set of 22 patients to analyze a total of 28 tissue specimens from NE regions and 31 specimens from CE regions (**Supplementary Fig. 3-4b**). Quantification of IHC images stained for Ki67 showed a wide distribution in both CE and NE regions, including several NE samples with high Ki67, but no significant difference between regions (**Fig. 3-4g**). In addition, we identified 5 patients where samples from both CE and NE regions were obtained during the same procedure (**Fig. 3-4h-i**). In these paired samples, Ki67 positivity was not significantly different between regions but was higher in the NE region of each tumor. While it is not clear if tumor cells in the NE region are significantly more proliferative than those in the CE region, our data indicate that the tumor cells in the NE region are actively proliferating, consistent with our hypothesis that they drive eventual recurrence.

To assess regional differences in terms of the cycling tendencies of individual cellular states, we split the cells by state and compared the cycling proportions between the NE and CE regions (**Fig. 3-4j**, top). In the NE regions, the cycling proportion of OPC-like cells was decreased in the IDH-WT GBMs and increased in the IDH-mut GBM and GIII-O. Interestingly, in PT90, the MES-like malignant cells overtake the NPC-/OPC-like cells in terms of cycling proportion in the NE region. Overall, MES-like cells exhibited the most consistent increase in cycling fraction in the NE region, a trend that was observed in 3 of 4 tumors that contained MES-like cells. We also performed the analysis in reverse to examine the state composition within the cycling compartment in each tumor (**Fig. 3-4j**, bottom) and found that in the IDH-WT GBMs, the cycling compartment was depleted of AC-like cells in the NE compared to CE regions. In the IDH-mut GBM, AC-like and NPC-like cells were depleted and MES-like and OPC-like cells were enriched in the NE cycling compartment. In the GIII-O, OPC-like cells were enriched and NPC-like cells were depleted in the NE cycling compartment.

Finally, we noticed that there was a connection between tumor burden (proportion of malignant cells) in the NE region (**Fig. 3-4b**) and proportion of malignant cells that are cycling (**Fig. 3-3c**), as the tumors with the highest (PT90) and lowest (TB77) tumor burden in the NE region exhibited the highest and lowest proportion of cycling malignant cells, respectively. Comparing the ranking with respect to these two metrics of all five tumors that had both CE and NE transcriptomes confirmed an identical rank order, and correlation analysis using the numerical values suggested a positive correlation (Pearson r=0.71, p=0.09; Spearman rho=1, p<0.01). Importantly, splitting the cells by region showed that this was driven by the cycling proportion rankings in the CE region (identical to NE tumor burden rankings) and not the cycling proportion in the NE region (not identical to NE tumor burden rankings). Thus, in this cohort, the proportion of malignant cells in the CE region that are cycling was predictive of relative quantitative tumor burden in the NE region, highlighting the power of this approach in finding biomarkers and drivers of tumor progression that may not be observed otherwise.

Region-specific molecular features of malignant cells in glioma

Because of their potential role in tumor recurrence, we next sought to investigate the gene expression profiles of malignant cells found in the NE regions of the tumor. We performed differential expression analysis to identify differentially expressed genes (DEGs) in malignant cells from the NE vs CE regions. The comparisons consisted of 1) all cells, 2) cells split by cellular state, and 3) cells split by cycling status. We analyzed each patient dataset individually to identify patient-specific regional differences, then examined whether there were conserved changes across patients (**Fig. 3-5a-e**). We observed several global differential expression patterns. As expected, the number of significant DEGs generally correlated with the number of cells included in the comparison, with one notable exception: MES-like cells had the highest number of conserved NE-enriched genes across tumors (**Fig. 3-5a**), despite having few, if any, upregulated genes within each tumor (**Fig. 3-5b**). Interestingly, in terms of directionality, we observed that all comparisons within an individual tumor that resulted in >5 DEGs were skewed in the same direction, with the only exception being NPC-like cells in TB78. Across nearly all comparisons, DEGs were mostly downregulated in PT94 and TB78, and mostly upregulated in PT90, TB76, and TB77. This suggests the presence of a global transcriptional activation or repression phenotype in the NE region that is conserved across cell populations (grouped by both cell state and cycling status) within each patient dataset.

To examine functional impact of these differential expression patterns, we then performed gene set enrichment analysis (GSEA) using the fgsea R package [19]. As input, we included the GO Biological Pathways, Canonical Pathways, Oncogenic Pathways, and Transcription Factor Target databases [20]. Using conserved DEGs across tumors in all malignant cells, we identified 686 significantly altered pathways between the two regions (Benjamini-Hochberg adjusted p-value < 0.05). Malignant cells in the NE region exhibited conserved upregulation in many pathways related to neuronal and synaptic function and downregulation in gene sets linked to extracellular matrix function, immune response, vascular development, EGFR signaling, and hypoxia (**Fig. 3-5f**). Analysis of all malignant cells in each individual patient dataset revealed a wide range of regional pathway enrichment magnitude. The GBMs all had at least 400 dysregulated pathway terms in NE compared to CE (1143 in PT94, 790 in TB78, 787 in TB77, 483 in PT90), while the GIII-O (TB76) had only 56 (**Fig. 3-5g, Supplementary Fig. 3-5a**). We noted that the rank order of GBMs based on number of differentially enriched pathways was identical to the order obtained when ranking the tumors by MES-like proportion in the CE region, as well as perfectly opposite of the ranking of overall OPC-like proportion, suggesting an antagonist relationship between these two cell types in terms of driving progression and biological divergence in the edge region of gliomas compared to the core.

By evaluating each tumor individually, we were able to identify patient-specific patterns of biological dysregulation in the NE regions. PT94, the tumor with the most dysregulated pathways, exhibited upregulation in many neuron-/synapse- related terms and downregulation in gene sets related to platelet function, translation elongation, EGFR signaling, immune function, and cilia motility (**Fig. 3-5g**). TB78 also exhibited upregulation in neuronal pathways, along with downregulation of terms related to p53/cell cycle regulation/tumor suppression, differentiation, and ECM, suggesting a transition to a proliferative state in the NE region (**Supplementary Fig. 3-5a**). In TB77, there was upregulation in pathways related to B cell function, angiopoietin, and STK33, which has been linked to tumor migration, EMT, and angiogenesis [21]; downregulation was observed in neuronal and synaptic signaling terms. PT90 had upregulation in ribosome- and translation-related terms and downregulation in immune function, transcription, and MAPK signaling. Finally, the GIII-O TB76 exhibited mostly upregulated pathways, most prominently related to ribosome and translation, along with the node of Ranvier compartment, likely owing to the expansion of the OPC-like malignant compartment.

Considering conserved regional DEGs across tumors in the individual malignant cellular states revealed via GSEA that AC-like and MES-like cells showed the most pathway dysregulation (103 and 46 significant terms, respectively). AC-like cells exhibited upregulation in neuron-, synapse-, and cell-cell interaction-related pathways and downregulation in antigen presentation, signaling via the oncogenic transcription factor CRX, immune response, and translation elongation (**Fig. 3-5h**). Interestingly, MES-like cells were the only malignant population that enriched strongly for specific transcription factor target gene sets, with upregulation in downstream targets of OCT1, HNF3, and MEF2 (**Fig. 3-5i**). OCT1 is known to increase proliferation and decrease migration in GBM, acting as a microenvironmental adaptation mechanism in response to high glucose [22]. HNF3A (FOXA1) and MEF2D have been shown to either promote (FOXA1) [23] or be necessary for (MEF2D) [24] proliferation in glioma. On the other hand, MES-like cells in the NE region experienced downregulation of translation and Slit-Robo signaling, which functions in angiogenesis (via VWF), regulation of microvessel density, and VEGF-induced migration/permeability [25]. These patterns suggest that MES-like cells are potentially in a distinct functional state in the NE region compared to the CE region, likely existing in a relatively more proliferative and less migratory/vascular-interactive state. OPC-like cells had just one conserved dysregulated pathway, which was upregulation in the node of Ranvier cellular compartment (**Supplementary Fig. 3-5b**). NPC-like cells had 32 dysregulated pathways, with PI3K complex being upregulated and neuron recognition, p53, and cell-cell adhesion being downregulated (**Supplementary Fig. 3-5b**). In non-cycling cells, HIF1, lymphocyte immunity, and ECM terms were downregulated in NE, while many neuronal and synaptic terms were upregulated (**Supplementary Fig. 3-5b**). Interestingly, in cycling cells, DNA replication was upregulated, and the transcriptional activator FOXJ2 was downregulated, which may relate to the global repression phenotype observed in PT94 and TB78 (**Fig. 3-5b**). Also downregulated in cycling cells was clotting factor X; as MES-like cells are more cycling in the NE region, this may relate to the apparent decrease in vascular remodeling driven by MES-like cells, since activation of coagulation can drive tumor progression and vascular invasion.

Finally, to complement the gene- and pathway-level comparisons, we sought to evaluate the altered cellular phenotypes more robustly by assessing which groups of cells are most “separable” when comparing the NE vs CE regions. We again leveraged the Augur package to identify the cell populations exhibiting the strongest differences between NE and CE. We first analyzed each tumor separately, grouping malignant cells by their cellular state or cycling status. This revealed heterogeneity across tumors in terms of the highest ranking cell group by separability, with AC-like and MES-like ranking at the top in IDH-WT GBMs and OPC-like and NPC-like ranking highly in the IDH-mut GBM and GIII-O (**Fig. 3-5j)**. We then asked whether there were differences across the two regions in terms of how readily different cell groups could be classified into their tumor of origin (i.e., intertumoral heterogeneity) or their cellular state (i.e., intratumoral heterogeneity). To compare the NE and CE regions in a way that allows evaluation of statistical significance, we first empirically generated a null distribution using 500 permutations, then compared observed results to the null distribution. Using this approach, we first evaluated regional differences in intertumoral heterogeneity within malignant populations by training the classifier to attempt to predict tumor of origin given cellular state or cycling status (**Fig. 3-5k**, top). This analysis revealed that AC-like and MES-like cells exhibited significantly more intertumoral heterogeneity in the CE region (AUC=0.839 and 0.837, respectively) compared to the NE region (AUC=0.765 and 0.749, respectively) (permutation test, BH-corrected p-value = 0.004 for both). No other malignant populations were significantly different in NE vs CE in terms of intertumoral heterogeneity. We also examined intratumoral heterogeneity by attempting to predict cellular state given either patient identity or cycling status, and there were no significant differences in NE vs CE (**Fig. 3-5k**, bottom). Taken together, our gene-, pathway-, and population-level analyses strongly implicate AC-like and MES-like cells as drivers of altered tumor biology across the core and edge regions in glioma.

Altered non-malignant cell microenvironment in NE region

We next turned our attention to the tumor microenvironment, an extremely important and complex factor in tumor biology and therapeutic development, to examine whether there are regional differences in the composition or transcriptional behavior of the non-malignant microenvironment. We began investigating non-malignant cells by annotating cell types via expression of canonical markers and hierarchical clustering, resulting in clusters assigned to astrocytes, oligodendrocytes, excitatory neurons, interneurons, endothelial cells, T cells, and myeloid cells; myeloid cells were further classified into blood-derived and microglial tumor-associated macrophages (referred to here as macrophages and microglia, respectively) using differential enrichment of a distinguishing gene list reported in Müller et al. (2017) (**Fig. 3-6a-c**). We first computed the proportion of the non-malignant microenvironment represented by each cell type (**Fig. 3-6d**). Oligodendrocytes and immune cells comprised the majority of all samples except for TB78, which was mostly neurons. The NE region exhibited several patient-specific composition changes, but was consistently depleted of T cells across patients, and had either a decreased or similar myeloid fraction, suggesting a potential immune-cold environment at the tumor periphery.

We then examined non-malignant cell type-specific transcriptomic differences between the NE and CE regions by performing differential gene expression analysis using the integrated dataset including cells from all patients (**Fig. 3-6e, Supplementary Fig. 3-6**). We found that the cell types exhibiting the highest proportion of differentially expressed genes (compared to total number of genes considered, which included those expressed in at least 5% of cells) were excitatory neurons (66%) and oligodendrocytes (54%), while all other cell types had a DEG percentage of <10%. As with the malignant cell populations, the DEG burden correlated with the number of cells in each cell type. To resolve this bias and to evaluate functional implications of the regional transcriptomic differences, we performed GSEA and compared NE vs CE within each cell type, yielding between 34 and 531 dysregulated pathways across cell types (**Fig. 3-6f-i**). In neurons, we observed conserved upregulation of oxidative phosphorylation and translation terms and conserved downregulation of neuron- and synapse-related terms, as well as excitatory neuron-specific downregulation of endothelin signaling. T cells similarly had downregulated neuron-and synapse-related terms. Microglia were found to experience upregulation in metabolism terms and downregulation in immune function terms, with macrophages interestingly exhibiting the opposite trend. Astrocytes had upregulated HSF1 and oxidative phosphorylation and downregulated receptor-mediated endocytosis and vascular wound healing. Oligodendrocytes also had upregulated HSF1 and downregulated translation-related terms. Endothelial cells exhibited upregulation in protein folding, collagen crosslinking, vessel development, and CD40, and downregulation in translation and neuron migration.

We identified several overlapping themes in the malignant and non-malignant compartments of the NE regions (compared to CE). Neuronal pathways were downregulated in T cells and both neuronal subtypes but were upregulated in malignant populations. Multiple OCT1 terms were downregulated in interneurons and excitatory neurons but upregulated in macrophages and MES-like malignant cells. EGFR signaling was downregulated when all malignant cells were combined and when all non-malignant cells were combined. Finally, many dysregulated genes and pathways linked to vascular function were identified. Robo signaling was downregulated in oligodendrocytes, T cells, and MES-like cells, but upregulated in neurons. Robo/VEGF crosstalk was downregulated in oligodendrocytes. VEGF signaling was downregulated overall. VEGFA expression was downregulated in macrophages. Multiple VEGF ligands were downregulated in oligodendrocytes and macrophages. Multiple VEGF and ANGPT ligands had upregulated expression in excitatory neurons, along with VWF and PTN. Proteins involved in blood-brain barrier (CLDN11, PDGFRb) were upregulated. These overarching themes implicate cell type-specific pathway dysregulation that converges on a limited number of themes, including vascular function, transcription and translation, metabolism, and normal brain cell function. Moreover, increased tumor burden and associated immune response in the CE region may stress non-malignant, non-immune cells and lead to a breakdown of normal function, including maintenance of vascular integrity, which in turn may explain certain MRI characteristics [27].

Integrated atlas of malignant-normal cell communication networks in the GBM core

Crosstalk between malignant cells and normal cells is a hallmark feature of cancer and is critical to tumor growth, invasion, and recurrence [28-30]. While observations made in the two compartments separately are important and insightful, connecting the analyses offers unparalleled insight into the cellular and molecular complexity underlying tumor biology. We thus sought to build an atlas of the glioblastoma connectome using the tumors in our dataset. We utilized CellChat [31], which uses a comprehensive signaling database (**Fig. 3-7a**) and performs quantitative inference and analysis of intercellular communication via network analysis, pattern recognition, manifold learning, and quantitative contrast approaches (see **Methods**). Malignant and non-malignant cells from each of the four GBM tumors containing both NE and CE transcriptomes were combined for this analysis, yielding a total of 13 cell types (malignant: cycling, AC-like, MES-like, NPC-like, OPC-like; non-malignant: astrocytes, oligodendrocytes, excitatory neurons, interneurons, endothelial cells, microglia, macrophages, and T cells). We first analyzed the CE regions to establish a baseline understanding of the glioblastoma connectome within the tumor core. CellChat identified 153 significant signaling pathways, with a rich network structure that involved communication between every possible pair of cell types (**Fig. 3-7b, Supplementary Fig. 3-7a**). In terms of overall signaling strength/probability, the top three pathways identified were NRXN, NCAM, and LAMININ, and we were able to fully resolve the relative contributions across cell types and across ligand-receptor pairs, highlighting the specific interaction patterns between malignant and normal cells in these pathways (**Fig. 3-7c-e, Supplementary Fig. 3-7b**). As expected, non-malignant neurons and NPC-like malignant cells were the dominant senders in the NRXN network, but interestingly, the malignant subpopulations exhibited similar or higher tendency to receive information in this network compared to non-malignant groups, consistent with studies showing that neurons can drive GBM growth through NRXN signaling [29, 30]. The NCAM network was dominated by normal neurons and glia, along with all of the malignant populations except MES-like cells. As expected, endothelial cells, along with astrocytes to a lesser extent, were the largest contributors of outgoing information flow in the LAMININ network. Within the malignant compartment, AC-like and MES-like cells were the primary senders and receivers of signaling information, consistent with their initial designation based on proposed similarity to astrocytes and endothelial cells. Overall, this analysis reveals a dense cell-cell communication network within the glioblastoma core, with specific pathway enrichment and structural and functional signaling patterns.

Rewiring of cell-cell communication networks in the NE region

After establishing a baseline of the cellular connectivity within the GBM tumor core, we asked whether there was altered cellular crosstalk in the NE region. Using CellChat, we performed a similar analysis using cells from the NE region, followed by a differential network analysis to compare signaling network differences between the two regions. In the NE region, CellChat identified a similar total number of significantly connected pathways (152 vs 153 in CE), and as in the CE region, there was significant communication observed between all possible pairs of cell groups (**Fig. 3-8a, Supplementary Fig. 3-8a**). The top significantly detected pathways were also similar, with 9 of the top 10 being shared between the two regions, though pathways had either differing or similar relative contributions from specific ligand-receptor pairs (**Supplementary Fig. 8b and c**, respectively). However, differential network analysis revealed that the overall connectivity, or signaling strength/probability, was markedly decreased in the NE region, especially in the relationships involving MES-like, AC-like, and cycling malignant cells (**Fig. 3-8b-c**). The only interactions that were stronger in the NE region were either exclusively composed of two normal cell populations or autocrine signaling among T cells (red lines in **Fig. 3-8b, right**). Overall, there was a strong decrease in the NE region in signaling within the malignant compartment (green box in **Fig. 3-8d**, right) and an increase in signaling within the non-malignant compartment (gold box, **Fig. 3-8d**, right). Since the analysis normalizes for number of cells in each group, this shift strongly suggests a significant difference between the two regions, with the malignant cells dominating the cell-cell signaling networks in the core and losing much of that influence in the periphery.

To investigate specific cell-cell interactions that were altered in NE vs CE, we visualized the differential interaction numbers and strengths from each sender and receiver group using unclustered (**Fig. 3-8d**) and clustered (**Fig. 3-8e**) heatmaps. We observed that bidirectional increases in interaction strength in the NE region (i.e., cell type A to B *and* B to A) consisted exclusively of immune cells and neurons. The only additional connections with increased interaction strength were those being sent to T cells from astrocytes, oligodendrocytes, and endothelial cells. The 5 most upregulated interactions involved either neuron-/astrocyte-to-T cell signaling or microglia-to-neuron signaling. Consistent with previous reports of macrophages playing a stronger role in the core of the tumor and microglia at the periphery [26], we observed a relative decrease in macrophage signaling compared to microglial signaling, both in the outgoing and incoming directions. Information flow decreased overall in the NE region, and this change was primarily driven by MES-like, AC-like, and cycling malignant cells (**Fig. 3-8e**, **Supplementary Fig. 3-8h-i**), along with normal astrocytes and, to a lesser extent, normal endothelial cells. Connections among these groups dominated the top 10 downregulated interaction pairs, with the strongest decrease exhibited by astrocyte-to-MES-like signaling. We also performed differential interaction strength analysis with three different levels of cell type grouping to confirm these shifts at multiple resolutions (**Supplementary Fig. 3-8e-g**).

To examine what specific pathways were driving these changes when comparing the NE and CE regions, we then performed joint manifold learning and classification of networks using

CellChat, which reduces the dimensionality of the connectivity results and identifies the pathways that experience the most change in terms of functional (cell types involved) and structural (network topology) re-wiring (**Fig. 3-8f-g, Supplementary Fig. 3-8j-k**). Consistent with several transcriptomic differences discussed above, we observed functional or structural divergence in the NE region in pathways related to vascular function/angiogenesis (VWF, PTN, ANGPT, HSPG, VEGF, APELIN, MSTN, CLDN, VCAM, CDH5), immune function (MHC-II, PDL2, BAFF, MIF, SELPLG, APRIL, CD40, BTLA, CD45, LIGHT, ACTIVIN, CD226, IL10, CD48, OX40), and cell-cell adhesion (NECTIN). These 3 groups almost exclusively comprised the top 20 functionally and structurally different pathways between the two regions. We then ranked the pathways by relative information flow between regions and identified the top NE-enriched and CE-enriched pathways, again pointing to pathways related to vascular development (e.g., EPO, VWF) and immune function (e.g., ICOS, PD-L1) (**Fig. 3-8h-i**).

Finally, after examining the patterns at the cell type and pathway levels, we sought to connect the two and determine what cell types were mediating the observed pathway alterations in the NE vs CE regions. Based on the patterns resulting from both the gene expression and cell-cell interaction analyses, we hypothesized that MES-like cells were driving the observed differences in terms related to vascular development and cell-cell adhesion. To evaluate this, we zoomed in on MES-like cells and identified the pathways with the largest differences in incoming and outgoing interaction strength in this population (**Fig. 3-8l**). As expected, MES-like cells in the NE region were found to experience downregulation in both vascular- and adhesion-related pathways, including LAMININ, COLLAGEN, CNTN, NCAM, FGF, and VEGF. We then focused on the VEGF pathway due to its established connection to many cancers, including

GBM. We visualized the cell type-specific expression levels of all ligands and receptors in the dataset that were part of the VEGF pathway and observed several dysregulated genes, including ligands being downregulated in macrophages, astrocytes, and MES-like cells and receptors being downregulated in endothelial cells (**Supplementary Fig. 3-8d**). We then examined interaction probabilities of specific ligand-receptor pairs and expanded the analysis to include TGFb and EGFR pathways, as they are both linked to vascular function and GBM progression. We visualized ligand-receptor interaction probability for all connections that were sent from malignant cells to non-malignant cells, noting a decreased interaction probability in all connections involving VEGF receptors on endothelial cells (**Supplementary Fig. 3-8l**, starred columns). These results strongly implicate VEGF as a candidate functional pathway that is significantly different between the core and edge regions, with varying contributions from each cell type and gene involved.

In addition to MES-like cells, we wanted to follow up on the observed increase in signaling strength in T cells from the NE region that was specific to non-malignant cell networks. We identified several pathways altered in T cells and examined the relevant cell type-specific expression patterns and ligand-receptor interaction probabilities. T cell-specific pathway dysregulation included both ECM and immune related networks, such as LAMININ, MHC-II, CD45, FN1, and TENASCIN (**Fig. 3-8k**). Part of the increased microglia-to-T cell signaling is likely driven by TNF-TNFRSF1B communication (**Fig. 3-8j**). Two of the five largest increases in interaction strength in the NE region were interneuron-to-T cell and excitatory neuron-to-T cell signaling (**Fig. 3-8e**). Following up on this surprising finding, we examined specific ligand-receptor pairs involving interleukin signaling in neuron-to-immune cell signaling (**Fig. 3-8m**).

We identified increased interactions that were shared among both neuronal subtypes (e.g., IL7 and IL18 signaling to T cells, IL1B signaling to microglia, IL1B signaling to macrophages) and others that were subtype-specific (e.g., IL10 and IL16 from excitatory neurons to T cells, IL4 from excitatory neurons to both microglia and macrophages, IL13 from interneurons to both microglia and macrophages).

In sum, while many signaling networks are conserved between the two regions, our higher resolution analysis reveals specific functional and structural network differences that suggest a significant degree of both rewiring and global reduction in connectivity compared to the core of the tumor, where the increased tumor burden likely takes advantage of these pathways to promote progression and invasion. The cell type- and interaction-specific pathway dysregulation that we identified strongly implicate candidate cell populations (MES-like, AC-like, endothelial, and immune) and functional pathways (VEGF, ECM, and immune) as targets for further analysis and therapeutic development.

Discussion

In this study, we leverage single nucleus RNA-sequencing and MRI-guided biopsy target selection to dissect spatial and temporal relationships between and within tumor and non-tumor cells in glioma. We use stringent image guidance to obtain specimens with quantifiable localization. Using nuclear sequencing and strict quality control measures, we are able to analyze on average ~4800 cells per tumor as well as ~1400 tumor cells from the NE region, an improvement on previous glioma sequencing studies that report either on CE alone or with few cells recovered from the NE region [32, 33]. We then characterize the cellular and transcriptomic landscape of the different tumor regions as described by standard MRI.

We present an improved and flexible framework to accurately identify malignant cells using iterative CNA inference analyses. Our strategy resolves several drawbacks related to traditional malignant cell identification workflows. Approaches that focus heavily on the original expression-based cluster identities suffer from general limitations related to clustering. For example, most clustering algorithms include an arbitrary parameter to determine the "resolution" of the clustering solution, which heavily influences whether the cluster annotations represent the true grouping of distinct biological types and/or states within the dataset. Moreover, due to the level of sparsity in single cell data, there is typically impurity in the clustering solution, such that cells of different types or states belong to "incorrect" clusters. Using these designations for labeling of malignant cells can thus lead to inaccuracy and is expected to result in both false positives and false negatives, especially since CNA events are being inferred from the expression data. As an example, we see in **Fig. 3-2c** that there are distinct CNA-based groupings within at least two of the original expression-based clusters (Cluster_7, fifth from the top, and Cluster_4, third from the bottom). Another challenge is the confident selection of a reference cell population for the estimation of CNA events, which is important because the final chromosome-level events (e.g., 1p deletion) are computed relative to a certain reference value. If no cell population is designated as the reference, the average of all cells will be used for comparison. This is not ideal for studies with multiple tumors, as the cell type composition across tumors is typically not similar, which will lead to a different “average cell” that will be used as the reference profile. However, designating a certain cell population as reference based on original expression-based clusters suffers from the clustering limitations mentioned above. Specifically, there may be malignant cells in non-malignant clusters or vice versa. We can see an example of this in **Supplementary Fig. 3-2**, where the dataset at step 3 of the workflow contains several reference cells with CNA profiles resembling malignant cells (green arrow in panel **a**), while the final step of the workflow (panel **b**) does not include these likely contaminating cells in the reference group. Similarly, we can address non-malignant cells that grouped with malignant cells in the expression-based clusters, as only the CNA-based cell groups exhibiting glioma CNA events are retained as malignant cells. Another issue is that workflows involving single cell/nucleus RNA-seq data are often prone to noise accumulating due to technical limitations (e.g., low quality cells) and multiple processing steps involving estimations and statistical models. Consequently, many raw findings include cells that appear to be outliers. In general, grouping individual cells by similarity using hierarchical clustering is a robust way to identify such cells, which can be considered a quality control step. In our CNA inference workflow, we remove cells that appear to be outliers upon performing hierarchical clustering, for example those falling under the branch next to the pink star in **Supplementary Fig. 3-2c** (step 2 result using the PT94 tumor). Finally, we improve the efficiency and accessibility of the currently available inferCNV workflow by automating the laborious setup procedure and implementing customization options that do not require manipulation of the original dataset or input files. Overall, we believe our workflow offers improved accuracy in malignant cell detection, in terms of both sensitivity and specificity, and increases dataset quality by identifying and removing extreme outliers. Using this strategy, we are able to confidently label malignant cells for downstream analysis.

Genetic subclones within gliomas are often either overlooked or not the focus of single cell molecular characterization. However, transcriptomic (**Fig. 3-2d**, right) and phylogenetic (**Supplementary Fig. 3-3b-d**) divergence suggest that these groups are likely functionally distinct. While previous glioma single cell sequencing studies have delineated specific hierarchical relationships [13, 34] or plasticity [10] among glioma cell types, these investigations used expression profiles for cell classification and evaluation of hierarchies. Thus, it remains unclear whether genetic (CNA-based) subclones lie along a rigid evolutionary hierarchy. To evaluate this, we present an improved tool for examining genetic subclones and potential hierarchical relationships among them. We first leverage the random trees algorithm implemented by the inferCNV package to identify CNA-based cell subclusters within each tumor. Visualizing the resulting subclusters is no easy task, and to our knowledge there does not exist a robust tool for automated generation of a clear visualization of subcluster phylogeny. Our method directly incorporates the inferCNV subclustering output file into the existing metadata table that contains information for each cell (e.g., number of genes, cell state, region of tumor), which enables plotting of a phylogenetic tree with superimposition of multiple layers of information about the relevant cells. Using this tool, we can readily identify patterns in compositions across CNA subclusters within individual tumors, in this case revealing stark differences in cellular state and regional composition across subclusters (pie charts in **Supplementary Fig. 3-3b-d**). Our tool includes several customizable features, including options for collapsing subclusters based on similarity of CNA locations, multiple ways to display subcluster sizes, and many parameters related to visualization details. Overall, this approach improves upon previous methods that focus on gene expression profiles (often using pre-defined lists of genes) and allows us to identify novel conserved and patient-specific genetic subclone-related features in glioma and their potential connections with regional divergence. In contrast to what has been observed in primary gliomas, we do not see a rigid genetic hierarchy among the malignant cells in this cohort of recurrent tumors, as the identified subclusters are all at the same branching level. This suggests that the tumors are potentially composed of multiple foci that were driven by different initiating cells, and further investigation of cell type and state patterns may help predict patient-specific cellular drivers of glioma recurrence.

The functional relevance of malignant cell states in glioma is poorly understood. Many of our findings in these highly different tumors are consistent with recent studies utilizing single cell sequencing in glioma [10, 16, 34], including the observation of multiple malignant cell types that are more different across than within tumors and whose proportions vary across tumors and correspond to overall pathological designation. These malignant cell states were initially described in Neftel et al. (2019) and were discovered by identifying gene networks that drive intratumoral heterogeneity in a large cohort of patients. Each cell state is thought to be biologically similar to a normal human cell type (AC-like cells = astrocytes, MES-like = mesenchymal origin, NPC-like = neural progenitors, OPC-like = oligodendrocyte precursors), which we validate via gene set enrichment patterns (**Fig. 3-3f**). Like the original TCGA GBM molecular subtypes, each cell state likely corresponds to specific functional drivers of tumor biology. Thus, cell state composition and state-specific molecular features in individual patients are an important metric for evaluating the relative importance and contributions of each of these functional patterns with respect to tumor progression, prognosis, and patient-specific therapeutic strategies.

Glioma recurrence is likely driven by malignant cells or altered non-malignant cells that remain after initial resection via surgery. By comparing the tumor core (CE) and edge (NE) regions, we identify significant differences on multiple levels that implicate multiple cell populations and functional networks. We first observe significant tumor burden in the NE region and build a model predicting that these cells may exist up to 1.5cm from the CE edge. In the NE region, multiple malignant cellular states are proliferating and thus have the potential to drive growth and recurrence. Furthermore, these malignant cells in the NE region exist in a relatively immune-cold environment with low T cell presence and activity. It remains unclear whether one or multiple specific malignant cell states must be present in the NE region to initiate recurrence after surgical resection, and further investigation may provide definitive evidence implicating a recurrence initiating cell.

Gene- and pathway-level comparisons between the two regions highlight the large degree of heterogeneity in molecular dysregulation across tumors, both in terms of overall magnitude of dysregulation and which functional networks are implicated. Interestingly, we observe a consistent direction of differential expression within each tumor when analyzing individual subpopulations, suggesting that the upstream drivers of transcriptomic dysregulation are shared or coordinately regulated across malignant cell groups. There is a stronger degree of pathway dysregulation in the CE vs NE regions in GBMs (between 483-1143 pathways) compared to GIII-O (56 pathways) using the same list of pathways as input. Importantly, GSEA only considers the rank order of the input genes and is thus not influenced by the number of genes that reached statistical significance in the differential expression analysis. Thus, unlike DEGs, the number of significantly dysregulated pathways is a robust indicator of true changes in biological function, suggesting significantly more functional divergence between the NE and CE regions in GBM than GIII-O. Running GSEA within each tumor identifies patient-specific regional differences, many of which are related to neuronal/glial function, proliferation, vascular function, immune response, ECM, and metabolism. This differential analysis highlights both common and patient-specific themes related to aberrant molecular and cellular function driven by tumor burden.

The large degree of biological differences between the core and edge regions prompts us to ask whether we can identify the cell populations driving specific changes. Overall, MES-like and AC-like cells exhibited more pathway dysregulation and transcriptional “separability” between regions than OPC-like and NPC-like cells. Regional pathway enrichment varies considerably across cell states. Compared to the CE region, AC-like cells in NE appear to function more like normal astrocytes, with upregulation in gene sets related to neuronal/synaptic function and downregulation in translation, immune response, and CRX signaling. Similarly, OPC-like cells in the NE region had upregulation in the node of Ranvier cellular compartment. NPC-like cells appeared to potentially function more like proliferative cells, with upregulation in PI3K and downregulation in p53, but it is unclear whether this is more similar to normal progenitor behavior or tumorigenic proliferation.

Like most aggressive and invasive cancers, high grade glioma depends on a well-orchestrated attack on the existing normal cell ecosystem, likely involving complex interaction networks among malignant and non-malignant populations. Using cell-cell communication inference, we first present an atlas of the GBM connectome in the core region, identifying a densely interconnected web of signaling involving all possible pairs of malignant and non-malignant populations (**Fig. 3-7**). Astrocytes and neurons are the top non-malignant communication senders, and cycling malignant cells are the top signaling senders in the malignant compartment. We identify the most significantly detected pathways and the relative contributions of individual ligand-receptor pairs within each detected pathway. We then build the connectome in the NE region and compare it to the CE region, revealing a marked decrease in overall interaction strength, especially in networks involving malignant cells, and an increase in interactions involving neurons/glia and immune cells. This suggests that the tumor burden in the core of the region is likely suppressing a large amount of neuron-/glia-to-immune cell signaling that is typically present in the brain. Among the malignant populations, while each of them had decreased signaling overall, MES-like cells exhibited the largest changes in both incoming and outgoing information flow, with AC-like cells ranking second. Comparing the two regions overall and within cell types implicates multiple pathways related to vascular and immune function (**Fig. 3-8, Supplementary Fig. 3-8)**.

Across all levels of analysis, MES-like cells appear to exhibit strong biological effects on the surrounding cellular infrastructure and are significantly different across regions. Comparing the transcriptomes of MES-like cells to the other malignant states reveals enrichment of ECM, collagen, and extracellular structure (**Fig. 3-3f**), suggesting a potential role in altering the structural integrity in the tumor microenvironment. This is consistent with the known GBM mesenchymal phenotype of aggressive growth, migration, vascular remodeling, and angiogenesis. In the tumor core, cell-cell interaction analysis reveals MES-like cells are significant players in multiple cell-cell signaling networks, with the corresponding signaling mainly flowing to and from normal astrocytes and endothelial cells. MES-like cells are important in laminin and collagen signaling, the third and fourth most significantly detected pathways, respectively. When comparing the NE and CE regions, MES-like cells and AC-like cells are the only malignant populations with different degrees of intertumoral heterogeneity (**Fig. 3-5k**). A decrease in this metric indicates that these two cell states are more similar across tumors in the NE region than the CE region, suggesting some degree of transcriptomic convergence across patients in the NE region that could be driven by the existence of conserved NE-specific functions in these cell states. Investigating the specific molecular changes in MES-like cells in the NE region reveals upregulation of proliferation-related pathways (OCT1, HNF3, MEF2, B-catenin) and downregulation of angiogenesis (Slit-Robo signaling, VEGF). MES-like cells also have a higher cycling proportion in the NE region in most tumors. In terms of cell-cell interactions, compared to the CE region, VWF and VEGF pathways are downregulated in NE when considering all cells, along with significant network rewiring (i.e., low functional/structural similarity) in multiple vasculature pathways (e.g., ANGPT, VEGF, VWF, CLDN, CDH5, VCAM). MES-like cells in the NE region exhibit the most downregulation in total communication strength, both in the incoming and outgoing directions, and have significantly weaker interactions with normal astrocytes and endothelial cells (**Fig. 3-8e**). Examining the specific pathways dysregulated in MES-like cells in the NE region identifies several vasculature-related networks, including VEGF, laminin, collagen, FGF, and PTN. Probing cell-cell molecular interactions and expression levels of ligand-receptor pairs mediating signaling specifically from MES-like cells to endothelial cells reveals downregulation of VEGFA, VEGFB, and VEGFC ligands and their receptor binding partners (**Supplementary Fig. 3-8d, l**). In sum, these analyses suggest that MES-like cells exist in a distinct functional state in the NE region compared to the CE region, likely in a relatively more proliferative and less pro-migratory/vascular-interactive state. On the other hand, normal endothelial cells in the NE region enrich for ECM and vessel development, suggesting that the vascular control is potentially transferred back to normal endothelial cells in the edge compared to the core. Further examination of this divergence could help identify strategies to “hack” malignant cells and manipulate their ability to influence the microenvironment around them.

By integrating multiple levels of information, we also identify several consistent cellular and molecular correlates that are predictive of large-scale tumor characteristics. In this cohort, the proportion of malignant cells in the tumor core that are cycling is strongly predictive of the quantitative tumor burden in the NE region. As the NE region is often overlooked both in research investigations and clinical assessment, this presents an opportunity to extend our understanding of tumor progression and stage within glioma patients. Moreover, there is an opposite relationship between MES-like proportion of malignant cells in the CE region and OPC-like proportion in both regions combined, which are positively and negatively, respectively, correlated with pathway dysregulation burden between the two regions in GBMs. This suggests an antagonist relationship between these two malignant cell types in terms of driving progression and functional dysregulation in the core region of GBMs compared to the edge, consistent with studies linking the mesenchymal GBM subtype and MES-like proportion to more aggressive and invasive tumors associated with poorer prognosis, with the opposite reported for the pro-neural subtype and OPC-like proportion [35, 36]. Again, this quantitative molecular evaluation presents a framework that allows deeper understanding of how these cell states relate to tumor biology.

Taken together, this study presents a thorough quantification and molecular characterization of both malignant and non-malignant cell populations and their functional roles in the NE region of gliomas. We implicate candidate cell groups and functional interactions in both the malignant compartment and non-malignant microenvironment that may serve as drivers of glioma regional divergence and recurrence, suggesting increased scientific and clinical emphasis on these cells and pathways in the NE region. These features are less likely to be addressed by current standard of care therapies and may be the link to an improvement in survival of glioma patients.

Acknowledgements

We would like to thank the personnel affiliated with key core services at UCLA, including the Jonsson Comprehensive Cancer Center (JCCC) Flow Cytometry Shared Resource and Technology Center for Genomics and Bioinformatics (TCGB). We also thank the lab of Dr. Inma Cobos, particularly Dr. Marco Otero-Garcia, for initial guidance in setting up the single nuclei isolation protocol.

Materials and methods

Human subjects

This study was approved by the University of California, Los Angeles Institutional Review Board #10-000655 and #14-001261. All patients provided preoperative informed consent for all medical and surgical procedures and involvement in research studies.

MRI-guided biopsy targets

Patients referred to the UCLA Department of Neurosurgery for intrinsic supratentorial brain tumor underwent 3 Tesla thin-cut standard pre-operative MRI with and without gadolinium contrast (gadopentetate dimeglumine; Magnevist; Bayer HealthCare Pharmaceuticals) imaging within 1 week of the scheduled surgery. Digital Imaging and Communications in Medicine (DICOM) images of the following sequences were downloaded: T1 three-dimension magnetization-prepared rapid gradient echo (MP-RAGE) with and without contrast. Using AFNI [37], a software for analysis and visualization of functional magnetic resonance neuroimages, three 5mm x 5mm x 5mm spherical biopsy targets were selected per patient. A multidisciplinary team including neurosurgery, radiology, and oncology performed biopsy target selection. Targets were chosen based on feasibility and anatomic landmarks for reference during surgery in order to minimize inaccuracy from brain shift. Biopsy targets were then transferred to Brainlab Curve (BrainLAB AG, Munich, Germany), a surgical neuronavigation software, for intraoperative image guidance and tissue acquisition. There was no change to surgical and post-surgical standard of care therapy [2, 38] to all patients. Specimens were immediately fresh frozen for storage until nuclear isolation or paraffin-embedded for immunohistochemistry.

Immunohistochemistry (IHC) analysis

MRI-targeted biopsy specimens were collected from 22 patients with malignant glioma, yielding 28 specimens from NE regions and 31 from CE regions. These tissue specimens were paraffin-embedded and stained with anti-Ki67 antibody. Image based quantification was carried out using QuPath [39] to calculate the percentage of sample pixels with positive staining. Samples were compared using independent samples t-tests as well as paired samples t-tests in the 5 patients from whom both CE and NE specimens were acquired.

Isolation of single nuclei from tumor specimens

Single nuclei were isolated from all frozen tumor specimens using a protocol based on Krishnaswami et al. (2016) that employs iodixanol-based density gradient centrifugation. Briefly, specimens were cut into small pieces, homogenized on ice using a Dounce tissue grinder, filtered using a 40uM Corning cell strainer, centrifuged at 1000g for 8 minutes at 4C in Eppendorf tubes, subjected to an iodixanol (OptiPrep; STEMCELL Technologies) gradient of 50% vs 29%, centrifuged at 13500g for 20 minutes at 4C, resuspended in PBS with 1% BSA, and labeled with the DNA stain Hoechst 33342 (Sigma). Nuclei were then purified based on size and Hoechst fluorescence using a microfluidic chip-based gentle cell sorter (On-Chip Sort; On-Chip Biotechnologies). The recovered nuclei were counted using an automated cell counter (Countess; ThermoFisher) and immediately submitted to the UCLA Technology Center for Genomics and Bioinformatics for library preparation and sequencing.

Single nucleus library preparation, sequencing, read alignment

Single nuclei from each tumor specimen were counted again and visualized at the TCGB core facility to evaluate quality. Single nucleus cDNA libraries were then generated using the Chromium Single Cell 3’ v3 kit (10x Genomics) following the manufacturer’s protocol, which consists of loading nuclei into the channels of a 10x Chromium chip, partitioning the nuclei into Gel Beads in Emulsion (GEMs), reverse transcription of RNA within the droplets, breaking the GEMs, amplification, fragmentation, and addition of adaptor and sample index. For single nucleus RNA-sequencing (snRNA-seq), each library was split across multiple lanes of the NovaSeq 6000 S2 platform (Illumina) and sequenced to a depth of approximately 600 million reads per library, with a read length of 2x50 base pairs. The resulting raw reads were then demultiplexed and processed using Cell Ranger (10x Genomics) for alignment to the human genome (GRCh38), filtering, barcode (i.e., cell) identification, and counting of unique molecular identifiers (i.e., transcripts). The output feature-barcode (i.e., gene-cell) matrices contained the quantitative transcriptomic information for all single nuclei isolated from each tumor specimen.

Pre-processing and quality control of snRNA-seq datasets

SnRNA-seq data were further processed and analyzed in R [41] using the Seurat package [42, 43], with exceptions noted below. All functions were run with default parameters except as specified. To perform pre-processing and quality control, the feature-barcode matrices generated by Cell Ranger were first imported into Seurat. Each biopsy specimen was imported and processed separately during this stage. Genes that were expressed in fewer than 5 nuclei and nuclei that expressed fewer than 500 unique genes were both removed. At this point, specimens with a markedly low average number of unique genes expressed per nuclei were removed from the analysis. In the remaining specimens, mitochondrial and ribosomal transcripts were identified with their respective prefixes, and the percent of transcripts belonging to these two categories was computed in each nucleus using the `PercentageFeatureSet` function. Nuclei with mitochondrial transcript content of 5% or greater or ribosomal transcript content of 2% or greater were removed. Raw count log normalization, scaling, variance stabilization, and variable gene identification were then performed for each specimen using regularized negative binomial regression via the `SCTransform` function [44], resulting in an additional “SCT” expression matrix containing normalized data for each specimen in addition to the original “RNA” expression matrix containing the raw data. For each patient, the datasets from all specimens were then combined using Seurat integration [42], which performs canonical correlation analysis (CCA) to align specimens based on matched biological states (i.e., to correct for technical variation across samples). Integration was performed as recommended by the Seurat developers, specifically by selecting integration features, then identifying anchors, then integrating the datasets. The `SelectIntegrationFeatures` (using 5,000 genes), `PrepSCTIntegration`, `FindIntegrationAnchors` (using the first 50 principal components), and `IntegrateData` functions were used for integration. Each integrated dataset (one per patient) was then prepared for further analysis by dimensional reduction via principal component analysis (PCA), uniform manifold approximation and projection (UMAP) embedding, and t-distributed stochastic neighbor embedding (t-SNE). These were performed using the functions `RunPCA`, `RunUMAP` (using the first 30 PCs and embedding into 3 spatial dimensions), and `RunTSNE` (using the first 30 PCs) functions, respectively. Finally, each patient dataset was clustered based on gene expression profiles with Seurat’s default graph-based clustering that implements the Louvain algorithm for modularity optimization. Clustering was performed via the functions `FindNeighbors` and `FindClusters`. Low dimensional graphs were generated using Seurat's `DimPlot` function with default parameters except as noted in the figure captions.

Identification of malignant cells and non-malignant cells

Cells were classified using a multi-step procedure integrating both gene expression profiles (see *Cell type marker visualization* section) and predicted copy number alteration (CNA) profiles (see *Malignant cell identification* section). This analysis was performed separately for the integrated snRNA-seq dataset derived from each patient. Malignant cells were identified first, then the remaining cells were analyzed separately to identify non-malignant cell types.

Cell type marker visualization

To assess expression of canonical cell type markers, we developed a tool for generating dot plots (e.g., **Fig. 3-6b**). This tool is based on Seurat's `DotPlot` function and introduces two improved features. First, the user has the option to automatically color code the cell type labels and corresponding marker genes, which is currently not possible using Seurat's `DotPlot` function. Second, the cell clusters are ordered by similarity via hierarchical clustering, with the resulting dendrogram shown to the left of the plot. Importantly, this hierarchical clustering procedure considers all detected genes rather than only the genes shown in the dot plot, which is what currently available tools use [43, 45]. The output of our tool is a figure containing both a dendrogram and a dot plot, with color coded cell types and expression information (average expression and percentage of cells expressing) for each marker. The two improvements allow for user-friendly visual customization and more accurate representation of the transcriptomic similarities among clusters. These plots were used to identify cell types based on marker expression profiles and the branching structure seen in the dendrogram.

Malignant cell identification

First, expression of canonical brain cell markers was examined across all Seurat clusters that were identified (see *Pre-processing* section above). To visualize marker expression patterns in each dataset, dot plots were generated as described above (*Cell type marker visualization* section). These plots were used to assess whether any clusters could confidently be labeled as normal brain cell types based on their expression profiles and the branching structure seen in the dendrogram. Clusters whose marker expression profiles did not clearly align with a normal cell type or aligned with multiple cell types were considered putative malignant cell clusters.

Next, inference of CNA profiles at single cell resolution was performed for each patient dataset individually. We developed a custom wrapper function that streamlines a workflow leveraging the R package inferCNV [15]. The purpose of this wrapper function is to make the CNA inference workflow more efficient and user-friendly by automating the laborious setup procedure required by inferCNV and adding customization options that do not necessitate manipulation of the input files. This function first takes a Seurat object (single cell expression dataset) and automatically generates all of the input files necessary for inferCNV: raw count data for each cell are obtained via `GetAssayData` (using the “RNA” assay and the “counts” slot); an annotation file with the Seurat cluster names corresponding to each cell is obtained by extracting the cell names and cluster information from the Seurat metadata; and a gene ordering file with the chromosomal location for all genes present in the dataset is generated using the R package biomaRt [46, 47] to access the GRCh38.p13 human genome assembly via Ensembl [48]. The function then passes these input files as arguments to the `CreateInfercnvObject` function to generate an inferCNV object for the dataset being processed, with an additional option allowing the user to exclude certain groups of cells (i.e., Seurat clusters) from the inference analysis. Finally, the function passes the inferCNV object to the `run` function, which performs the rest of the CNA inference analysis that attempts to identify large-scale chromosomal CNAs by assessing smoothed average gene expression across a moving window of genomic positions, followed by prediction of CNA regions and identification of cell clusters based on CNA profiles. For each patient in the study, an inferCNV object was generated with the X, Y, and mitochondrial chromosomes excluded, then analyzed with default parameters except a cutoff value of 0.1, Hidden Markov Model (HMM) prediction with the i6 model, denoising with a filter of 0.12, and 6 threads for parallelization. This yields several outputs, most importantly a heatmap displaying relative expression intensity in each cell (rows) across chromosomal positions (columns) and text files including the HMM CNA predictions across genes/chromosomes for all cells and cell clusters.

To identify malignant cells in each patient dataset, CNA inference was first performed as described above, without any clusters defined as reference cells (i.e., values were compared to the average across all cells). The resulting heatmap groups cells based on their CNA profiles but also displays the original Seurat cluster identities, allowing assessment of potential cell type identities using both pieces of information. For each patient, the heatmap was examined to label CNA-based clusters corresponding to myeloid cells (based on amplification in a portion of chromosome 6 containing the human leukocyte antigen (HLA) system), putative malignant cells (based on canonical glioma-specific CNAs), and putative normal cell types (based on both lack of glioma CNAs and correlation with clusters that were previously labeled as confident normal cell types based on marker expression). Groups of cells that could not confidently be labeled were designated as unclassified. Cells outside of the large clusters visualized on the dendrogram were designated as noise. Next, CNA inference was repeated for each dataset using a reference set of cells, which included cells belonging to the Seurat clusters confidently labeled as oligodendrocytes (based on marker expression), since this was the only normal, brain-derived/non-immune brain cell type present in all patient datasets. As in the previous step, the resulting heatmap was examined to label CNA-based clusters corresponding to immune cells, malignant cells, and other normal cell types based on the CNA profiles and similarities to other clusters. For each cell, the labels from each of these two steps were both used to generate a final set of labels with the following potential values: malignant, non-malignant, noise, and unclassified. The two sets of CNA inference results were integrated in this manner to resolve 1) malignant cells that, due to their atypical transcriptomes, belonged to Seurat clusters corresponding to normal cell types and 2) cells with noisy profiles. The cells labeled as noise or unclassified were removed, and CNA inference was performed again on the final set of malignant cells, including hierarchical clustering to visualize subclones. This final set of malignant cells was used for the downstream analyses related to the malignant cell compartment. (To confirm non-malignant cell identities, we also compared CNA profiles of non-malignant cell types across tumors, which showed conserved features in each cell type, e.g., chr6 amplification in myeloid cells and chr11 amplification in astrocytes.)

Subsetting and processing of malignant and normal cells

For each patient, the integrated expression object containing all cells was then split into two objects, one including malignant cells and the other including non-malignant cells. Pre-processing was repeated for the individual malignant and non-malignant datasets for each patient. Each dataset was slimmed down to the raw, unprocessed counts using the `DietSeurat` function to retain only the RNA assay. Datasets were then processed using `SCTransform` as described in the *Pre-processing* section above. The malignant and normal cell objects from all patients were then combined into two integrated objects using Seurat integration as outlined in the *Pre-processing* section above, with one integrated object containing malignant cells from all patients and one integrated object containing normal cells from all patients. (Note: the `SCTransform` and integration workflows in the previous section were applied to individual specimens from each patient, while this section is performing these functions on merged datasets representing all specimens from each patient.) Each of these two integrated datasets was then prepared for further analysis by dimensional reduction and Seurat clustering as described in the *Pre-processing* section above. Low dimension graphs were generated using Seurat's `DimPlot` function with default parameters except as noted in the figure captions.

Identification of CNA subclusters and related hierarchy

InferCNV [15] was used to identify CNA-based tumor subclusters within each tumor using the random trees algorithm. Parameters were as described above except `analysis_mode = ″subclusters″` and `tumor_subcluster_partition_method = ″random trees″`. We then used a custom tool to visualize subcluster hierarchy and other related features. First, we incorporate each cell's subcluster identity (from the "cell_groupings" output file) into the Seurat object's metadata, including the specific branching trajectory. Next, we repurpose the clustree R package [49] to generate a phylogenetic tree by considering each branching level of the subcluster identities as an individual "clustering resolution" (i.e., the divisions at the top of the tree are "low resolution" clustering solutions, while the branches at the bottom are "high resolution" clustering solutions). This allows visualization of the branching structure, number of terminal branches, and cell abundance within each specific trajectory. Finally, since the Seurat metadata contains other types of annotations for each cell (e.g., number of genes, cell state, region of tumor), we superimpose additional layers of information about the cells in each terminal branch, allowing identification of subcluster-specific cellular composition with respect to other variables (pie charts in **Supplementary Fig. 3-3b-d**).

Molecular classification of glioma cellular states

Malignant cells were grouped into subpopulations based on the integrated model of glioblastoma malignant cellular states presented in Neftel et al. (2019). The same classification workflow was used in this study. Single cells were classified as astrocyte-like (AC-like), mesenchymal-like (MES-like), neural progenitor cell-like (NPC-like), and oligodendrocyte precursor cell-like (OPC-like) based on enrichment scores corresponding to gene lists related to each of these four cell states. The gene lists corresponding to each cell state can be found in Table S2 of Neftel et al. (2019). Enrichment scores were generated for each malignant cell as described [11] using the Seurat function `AddModuleScore` with default parameters except `assay="SCT"`, `search=T`, and `nbin=30`. This function computes enrichment of a gene list (i.e., module score) in each cell by calculating the average expression of the genes contained in the gene list and subtracting the average expression of a control gene list. The control gene list is compiled by first binning all analyzed genes into a certain number of bins (in our case 30) based on aggregate expression levels, then randomly selecting 100 bin-mates for each gene in the query gene list, which results in a control gene set that is 100-fold larger than the query list and has a similar distribution of expression levels. Module scores were generated for all malignant cells using the meta-modules defining the AC-like, MES-like, NPC-like, and OPC-like malignant cellular states, and each cell was assigned to the state with the highest score. Two-dimensional graphs based on these module scores were generated as described in Neftel et al. (2019). Briefly, cells from each cellular state were placed into one quadrant of the plot by computing the strength of score separation based on two groupings: OPC-like/NPC-like vs AC-like/MES-like (y-axis) and AC-like/OPC-like vs MES-like/NPC-like (x-axis). The equations used to compute the magnitudes of the x and y coordinates for each cell were as follows: for NPC-like and OPC-like cells, $x = \log_2(|\mathrm{NPCscore} - \mathrm{OPCscore}| + 1)$; for AC-like or MES-like cells, $x = \log_2(|\mathrm{MESscore} - \mathrm{ACscore}| + 1)$; for all cells, $y = \mathrm{maximum}(\mathrm{OPCscore}, \mathrm{NPCscore}) - \mathrm{maximum}(\mathrm{ACscore}, \mathrm{MESscore})$. The x values were then multiplied by -1 for OPC-like and AC-like cells. This placed the OPC-like cells in the top left quadrant, NPC-like in the top right, AC-like in the bottom left, and MES-like in the bottom right. Cells were colored according to their cellular state assignment or other grouping variables as noted in the figure legends. The top enriched pathways in each cellular state and in cycling cells (see next section) were identified using enrichR [50-52] via the `DEenrichRPlot` function.

Identification of cycling cells

Cycling cells were identified as described in Neftel et al. (2019). Briefly, enrichment scores were generated for each malignant cell in the same manner as in the previous section using two cell cycle signatures, one corresponding to the G1/S cell cycle phases and the other to the G2/M phases. Cycling cells were identified by fitting each of the two cycling signature score distributions to a normal distribution and testing each cell's significance using a threshold of 0.05. Cells that were significant for either of the two signature scores were labeled as cycling cells.

Differential expression and gene set enrichment analyses

Differentially expressed genes (DEGs) were identified using the `FindMarkers` (within groups) and `FindConservedMarkers` (shared across groups) functions in Seurat and visualized using the `EnhancedVolcano` function [53]. In each comparison, genes were included only if they were expressed in at least 5% of cells in one of the two groups being compared. For the bar plots in **Fig. 3-5a-b**, significant DEGs were defined as those exhibiting at least a 10% change in expression between groups and $p < 0.05$ (Bonferroni-adjusted p-value for `FindMarkers` output or combined p-value for `FindConservedMarkers` output). To perform gene set enrichment analysis (GSEA), we first generated a comprehensive list of pathway terms by combining the following databases: GO Biological Pathways, Canonical Pathways, Oncogenic Pathways, and Transcription Factor Targets [20]. For each differential expression comparison, GSEA was performed using the fgsea R package [19] with fold-change as the gene ranking metric, yielding enrichment scores for all pathways in the combined list.

Cell-cell communication analysis

The R package CellChat [31] was used for communication network analysis. Default parameters were used except as recommended by the developers in the online vignettes.

Visualization

Except as noted above, all remaining plots were generated in R using the ggplot2 package [54].

Figures

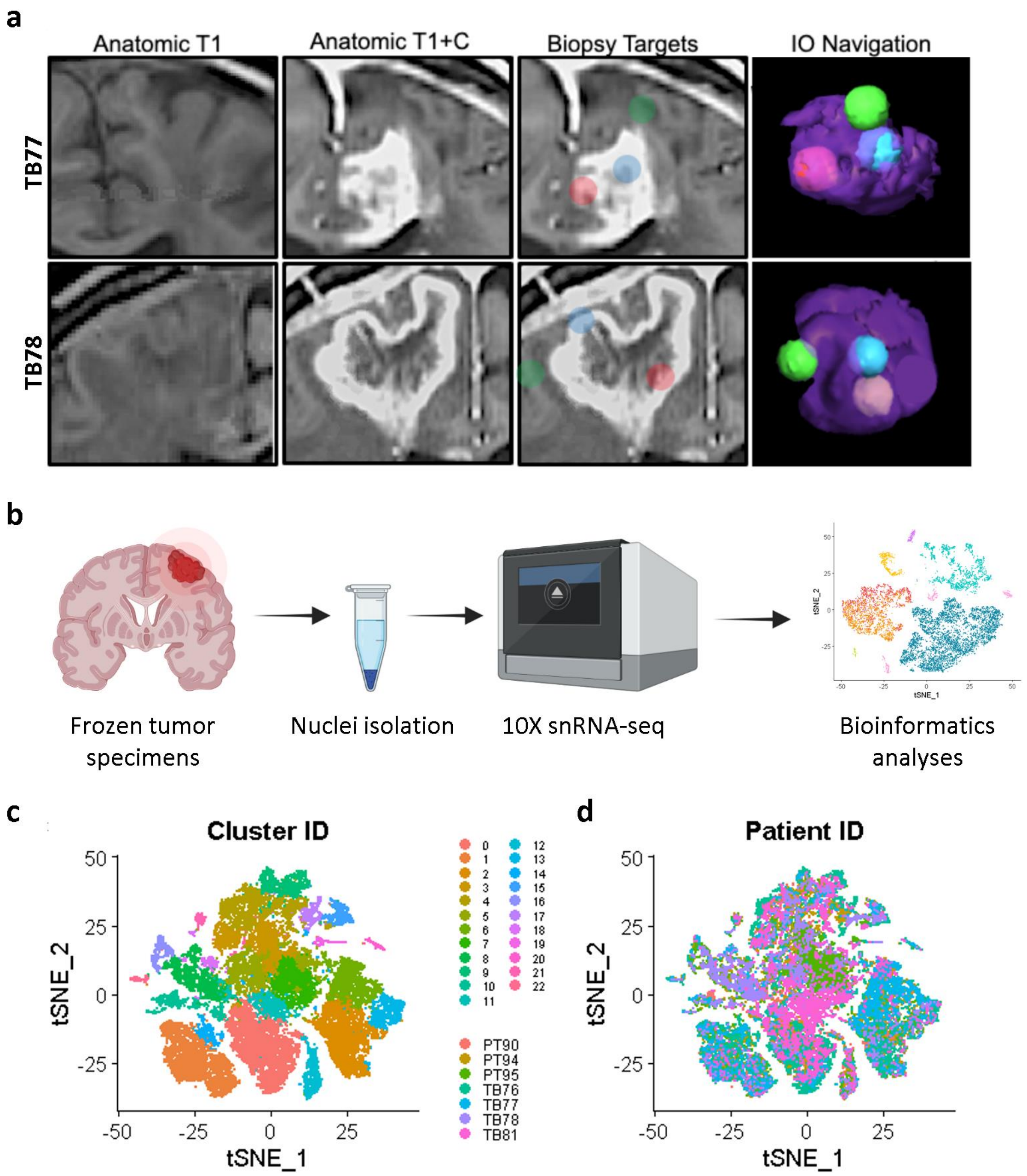
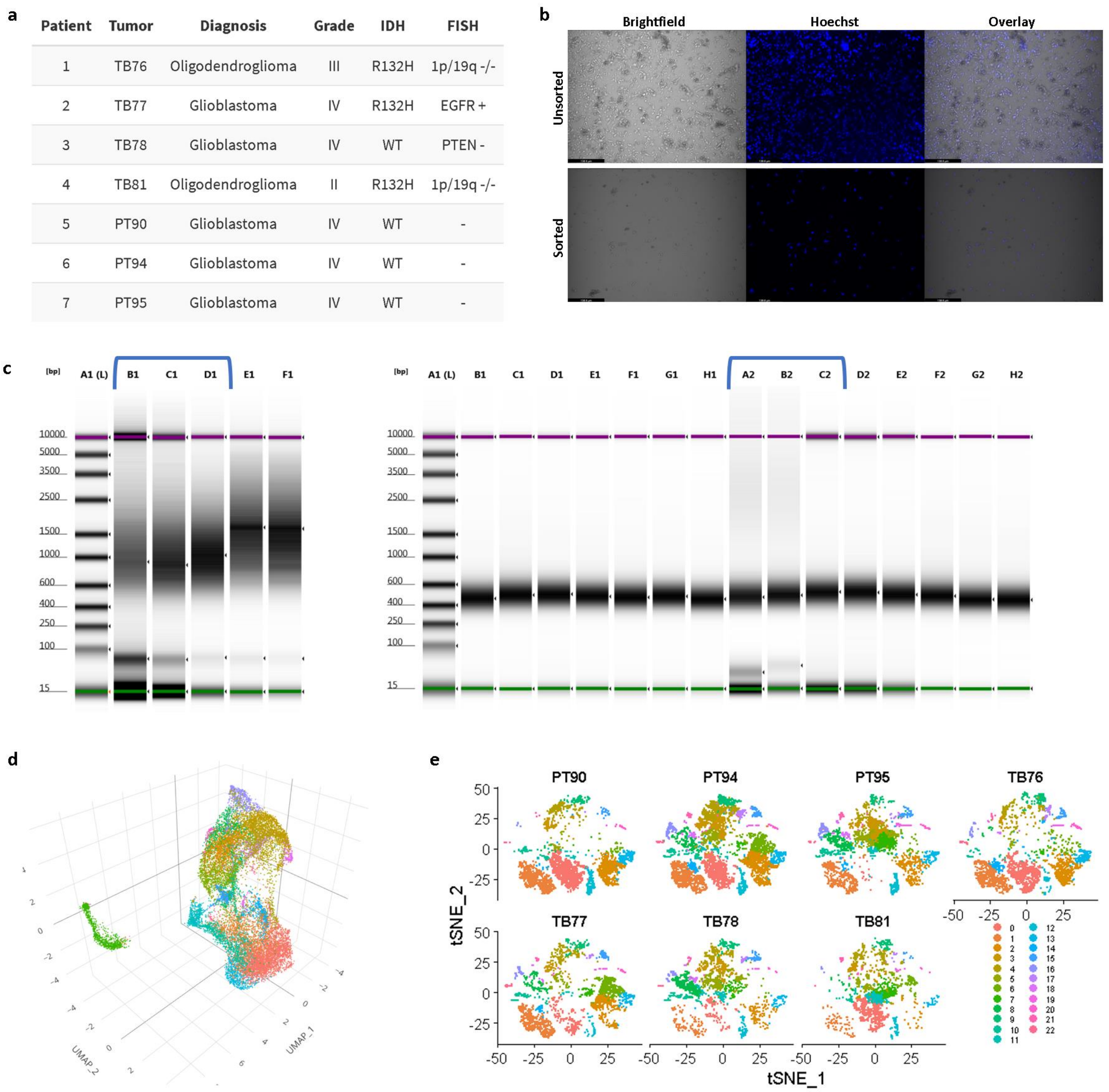


Figure 3-1. Single nucleus RNA-sequencing of MRI-guided glioma biopsy targets. **(a)** Targeted MRI: T1 with and without contrast preoperative magnetic resonance images, prospective biopsy targets, and intraoperative (IO) neuronavigation renderings of 2 contrast-enhancing and 1 non-enhancing region in TB77 (top; IDH-mut GBM) and TB78 (bottom; IDH-WT GBM). **(b)** Glioma single nucleus RNA-sequencing experimental workflow (created with Biorender.com). **(c, d)** Low dimensional plots via t-distributed stochastic neighbor embedding (t-SNE) showing the integrated snRNA-seq dataset with all sequenced nuclei, colored by expression-based clusters **(c)** or patient identity **(d)**.



Patient	Tumor	Diagnosis	Grade	IDH	FISH
1	TB76	Oligodendroglioma	III	R132H	1p/19q -/-
2	TB77	Glioblastoma	IV	R132H	EGFR +
3	TB78	Glioblastoma	IV	WT	PTEN -
4	TB81	Oligodendroglioma	II	R132H	1p/19q -/-
5	PT90	Glioblastoma	IV	WT	-
6	PT94	Glioblastoma	IV	WT	-
7	PT95	Glioblastoma	IV	WT	-

Supplementary Figure 3-1. Related to Fig. 3-1. (a) Clinical and pathological characteristics of glioma samples. **(b)** Isolation of single nuclei from glioma specimens (see **Methods**); shown is a microscopy image at 20x magnification, illustrating sample contents before and after sorting based on Hoechst fluorescence. **(c)** Example of cDNA (left, lanes B1-D1) and 10X library (right, lanes A2-C2) quality control via TapeStation automated electrophoresis analysis (Agilent Technologies, Inc) from the three biopsy specimens selected for PT90. QC was performed by the UCLA Technology Center for Genomics and Bioinformatics. **(d)** As in **Fig. 3-1c**, but with uniform manifold approximation and projection (UMAP) coordinates instead of t-SNE. Data were embedded into 3 dimensions. **(e)** Integrated snRNA-seq dataset with nuclei split by patient identity and colored by cluster.

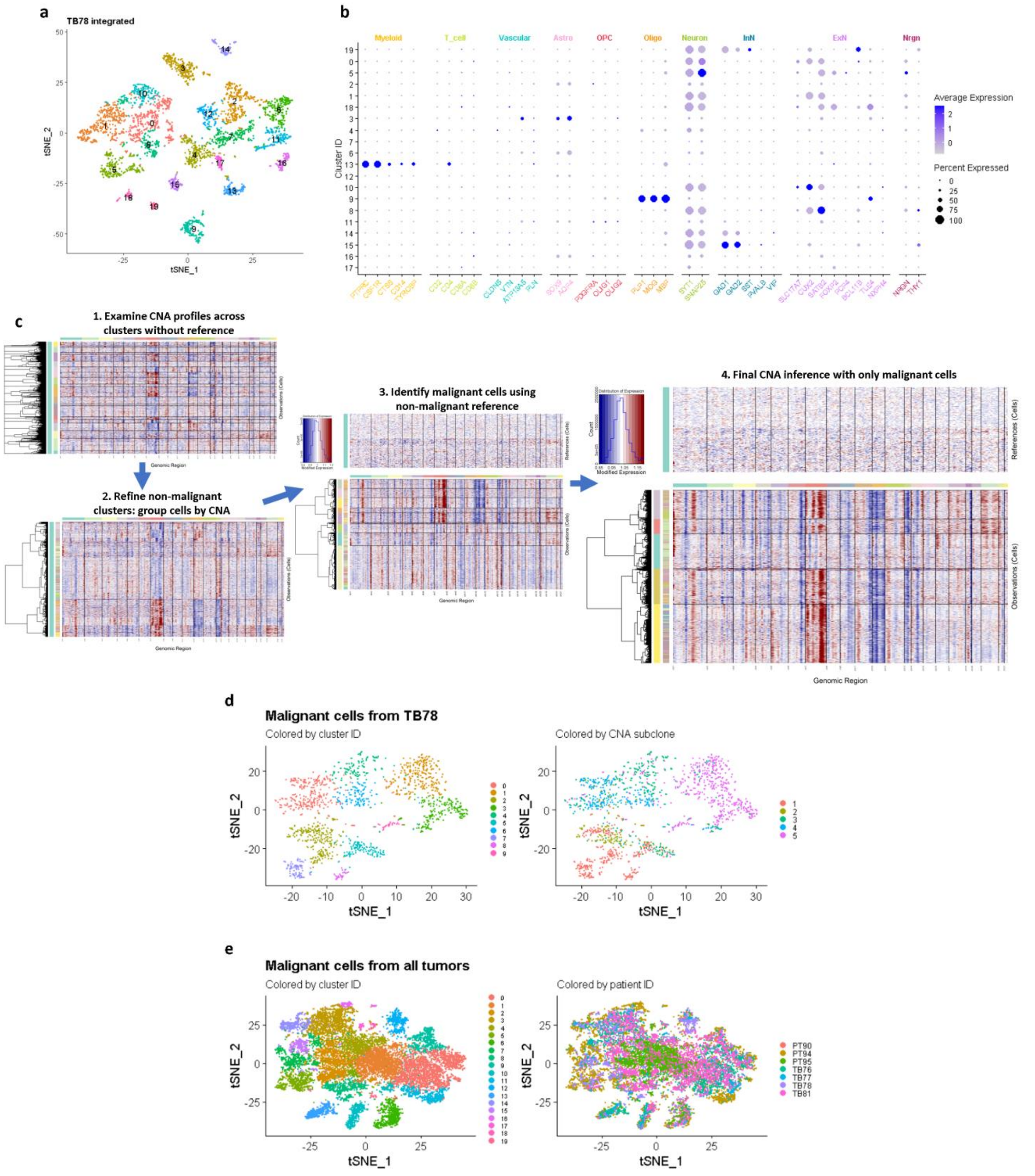
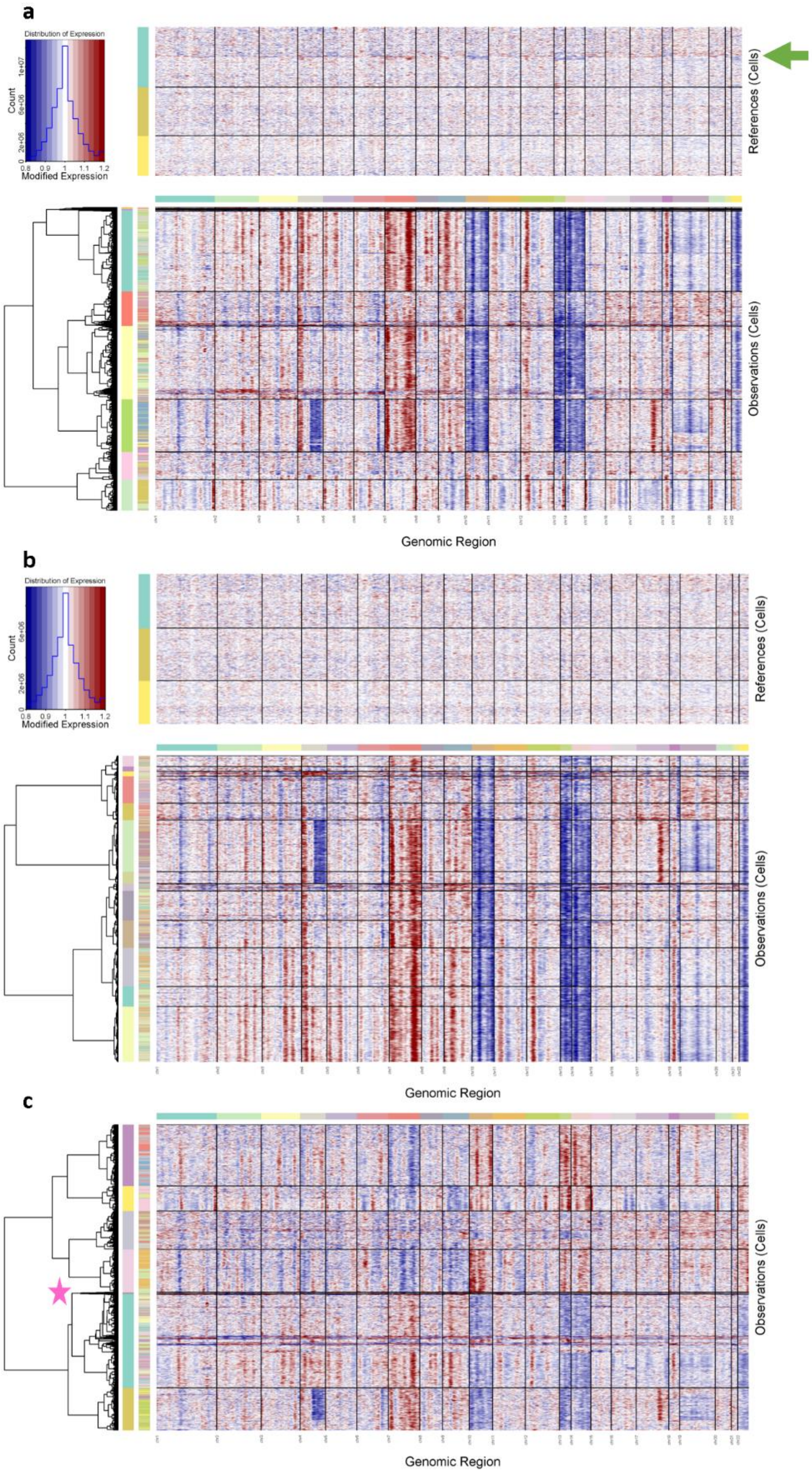
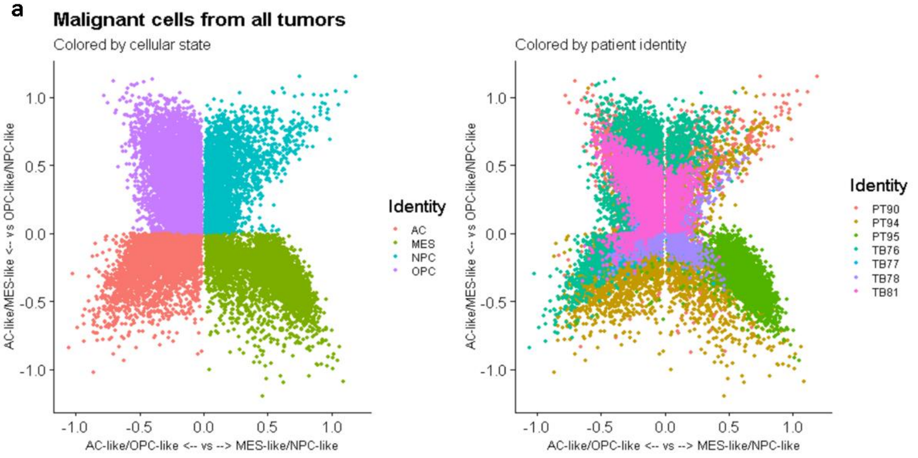


Figure 3-2. Improved identification of malignant cells in glioma specimens. Identification of tumor cells in TB78. Canonical cell type markers are assessed across expression clusters **(a, b)** and integrated with single cell CNA profiles obtained using iterative inference **(c)** (see **Methods**), yielding confident labeling of malignant and normal cells and identification of CNA subclones. In the final heatmap, the left-most vertical color bar indicates CNA-based subclone and the adjacent color bar indicates original expression-based cluster identity. **(d)** t-SNE plot showing 1655 malignant cells identified in TB78, colored by CNA subclone. **(e)** t-SNE plot showing all malignant cells integrated across patients. *(Astro = astrocytes, ExN and InN = excitatory and inhibitory neurons, Nrgn = neurogranin neurons, Oligo = oligodendrocytes)*

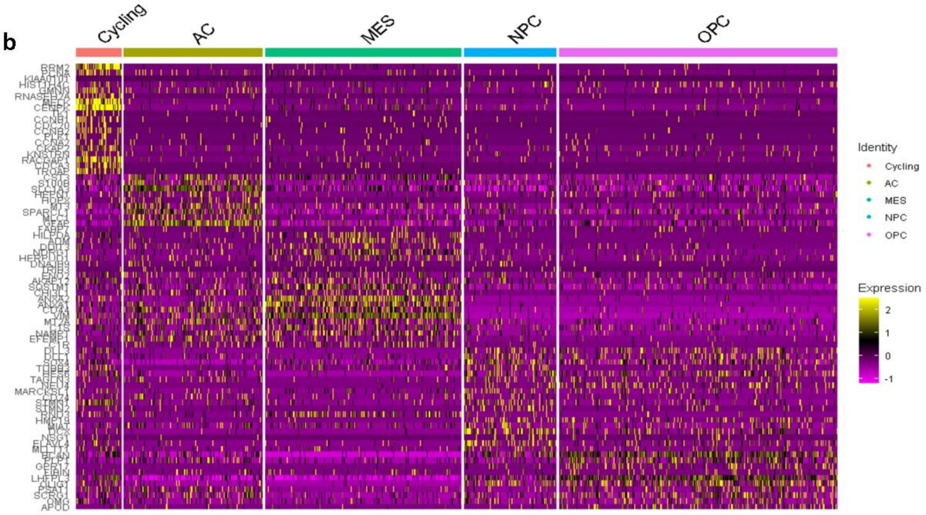


Supplementary Figure 3-2. Related to Fig. 3-2. Applying improved malignant cell identification strategy to the PT94 tumor. **(a)** Output from step 3, highlighting a group of reference cells appearing to be malignant (left of green arrow). **(b)** Final output shows absence of contaminating cells in the reference group. **(c)** Step 2 output with noisy group starred.

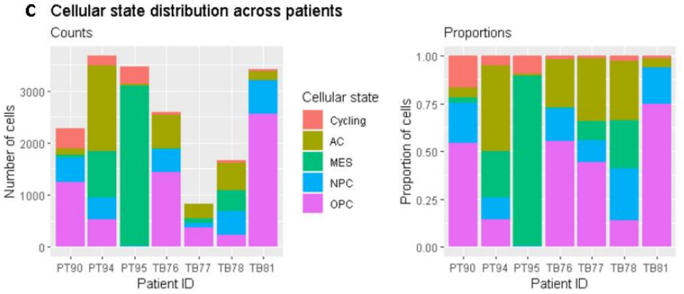
a


Malignant cells from all tumors

b



c


Cellular state distribution across patients

d

	P.Value	FDR
MES	0.0000000	0.0000000
OPC	0.0000000	0.0000000
AC	0.0000001	0.0000001
NPC	0.0002974	0.0003717
Cycling	0.0211318	0.0211318

e

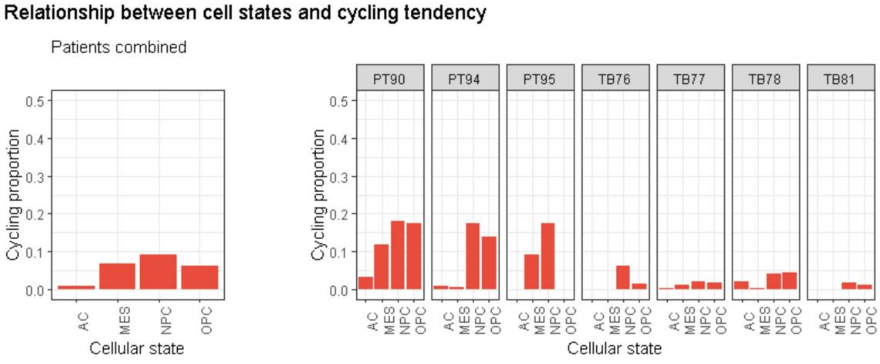
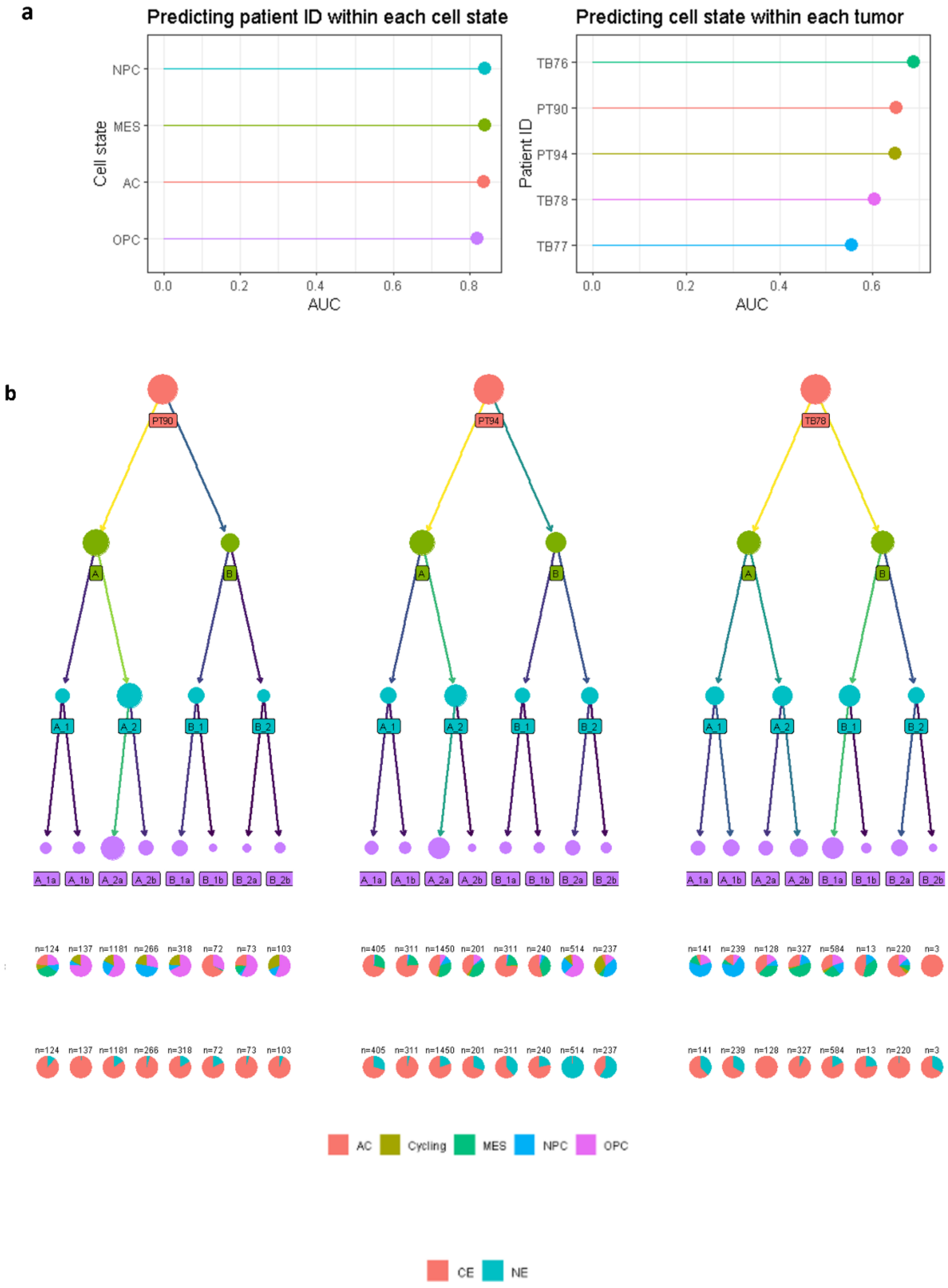
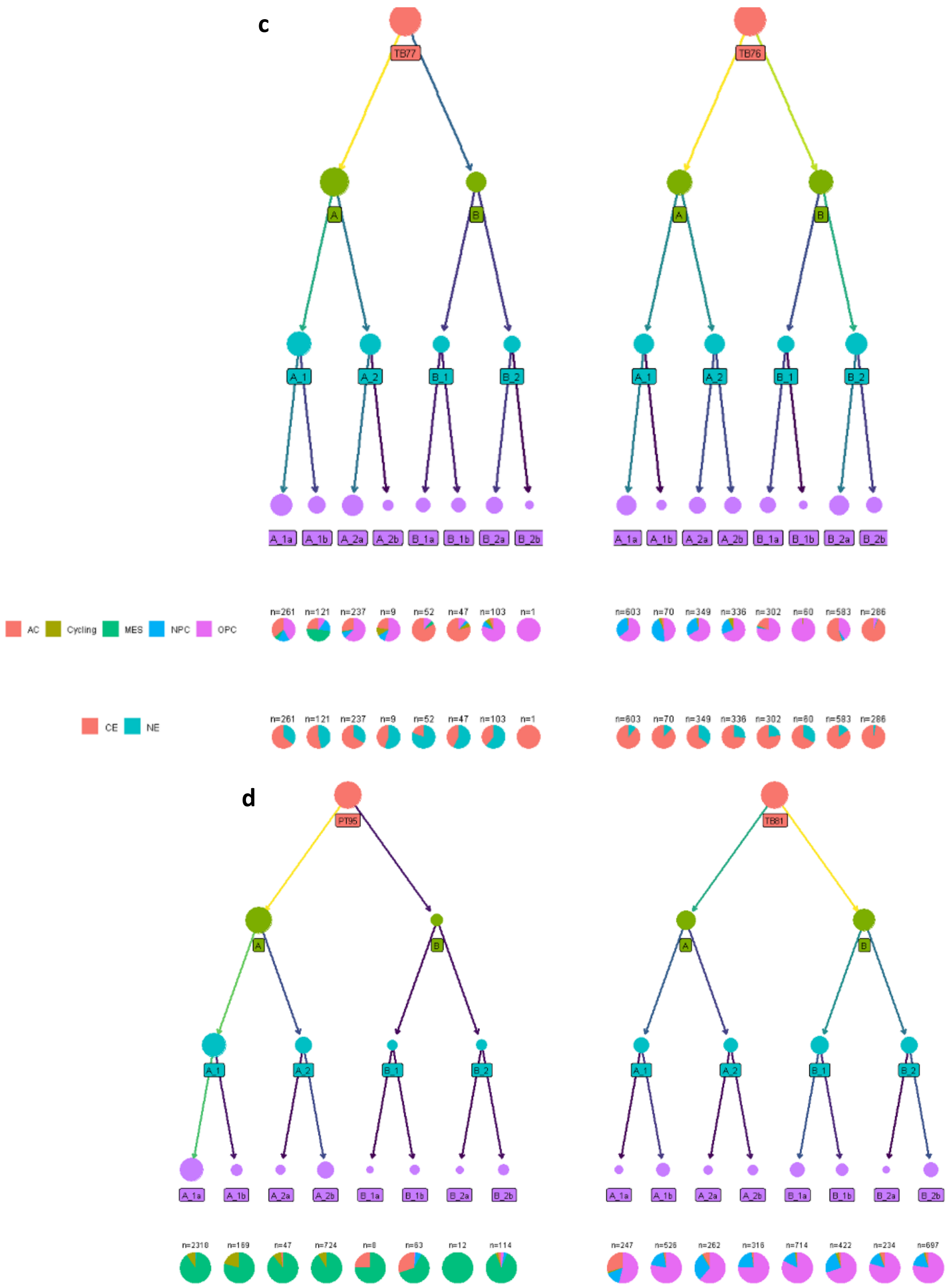

Relationship between cell states and cycling tendency

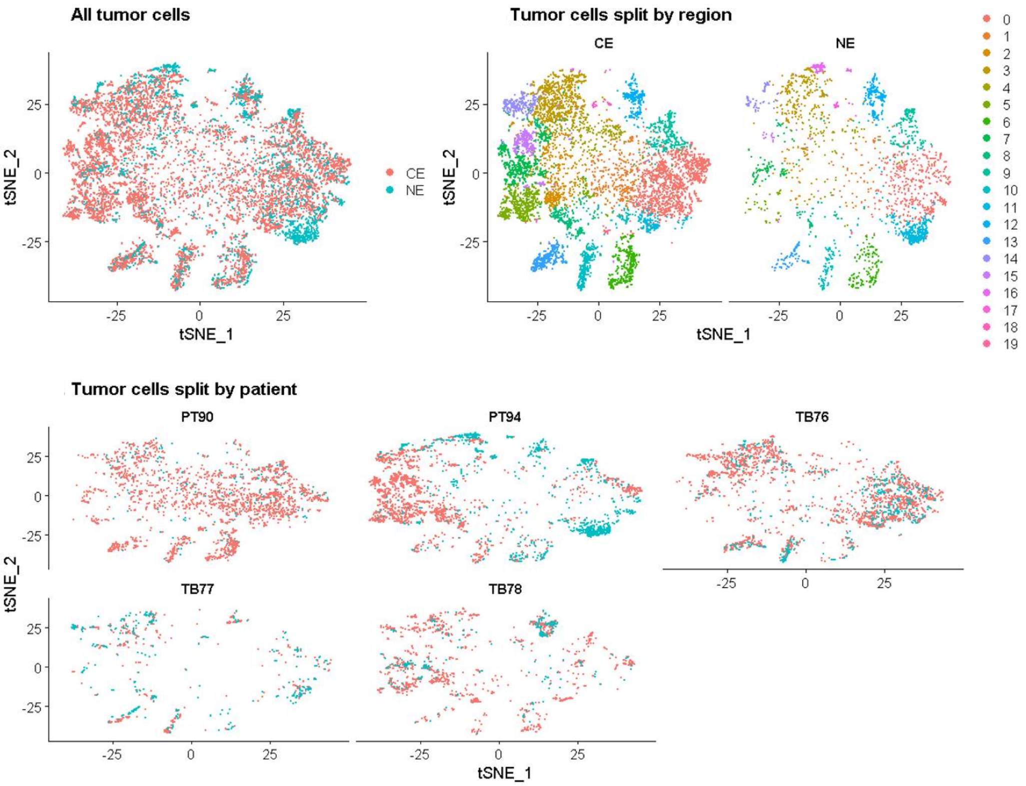
Figure 3-3. Recapitulation of established malignant cellular states in recurrent glioma. Glioma cells were classified by cellular state (AC-like, MES-like, NPC-like, and OPC-like) and cycling status (cycling, non-cycling) using gene network module scores as described by Neftel et al. (2019). Shown in **(a)** are all malignant cells integrated across tumors and plotted based on relative enrichment among the cellular state meta-modules (see **Methods**). **(b)** Heatmap depicting expression of top 10 genes across the gene lists obtained from Neftel et al. (2019). Rows correspond to genes and columns to cells. Cells are grouped according to cellular state, with cycling cells plotted separately. A random subset of 5000 malignant cells were included. Visualization **(c)** and evaluation via ANOVA **(d)** of intertumoral differences in malignant cell state composition. **(e)** Proportion of cycling cells detected in each cellular state. **(f)** Top enriched pathways in each cellular state and in cycling cells, generated using enrichR [50-52].


a
Predicting patient ID within each cell state
Cell state
NPC
MES
AC
OPC
0.0
0.2
0.4
0.6
0.8
AUC
Predicting cell state within each tumor
Patient ID
TB76
PT90
PT94
TB78
TB77
0.0
0.2
0.4
0.6
AUC
b
PT90
PT94
TB78
A
B
A_1
A_2
B_1
B_2
A_1a
A_1b
A_2a
A_2b
B_1a
B_1b
B_2a
B_2b
n=124
n=137
n=1181
n=266
n=318
n=72
n=73
n=103
n=405
n=311
n=1450
n=201
n=311
n=240
n=514
n=237
n=141
n=239
n=128
n=327
n=584
n=13
n=220
n=3
AC
Cycling
MES
NPC
OPC
CE
NE

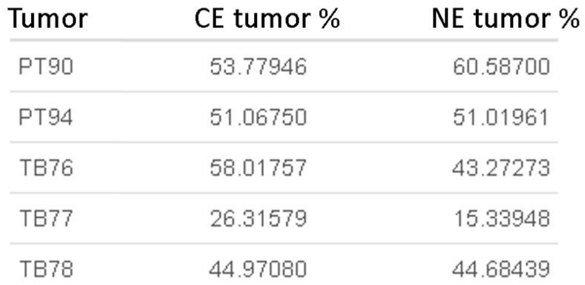


Supplementary Figure 3-3. Related to Fig. 3-3. (a) Cell type prioritization via Augur. Results from training the classifier to predict patient ID given cell state (left) or to predict cell state given patient ID (right), to evaluate inter- and intra-tumoral heterogeneity, respectively. **(b-d)** Using a novel tool to reconstruct hierarchies from CNA-based subclustering information. Top nodes are labeled by patient ID, edge color is scaled to number of cells traveling along that branch (yellow=high), node sizes are scaled to number of cells in that node or any daughter nodes, pie charts represent proportions of malignant cellular states (top chart) or region (bottom chart). Two of the tumors, PT95 and TB81, did not yield NE cells, so there is no region breakdown for those **(d)**.

a

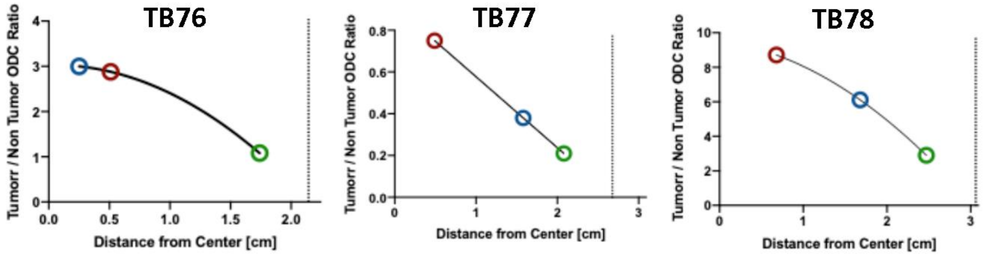


b

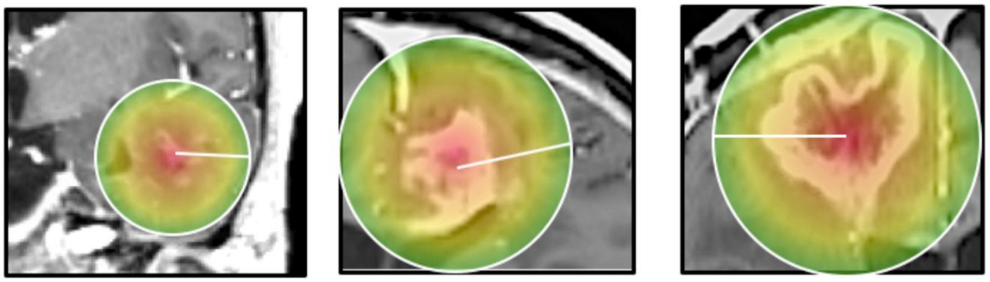


Tumor	CE tumor %	NE tumor %
PT90	53.77946	60.58700
PT94	51.06750	51.01961
TB76	58.01757	43.27273
TB77	26.31579	15.33948
TB78	44.97080	44.68439

c



d



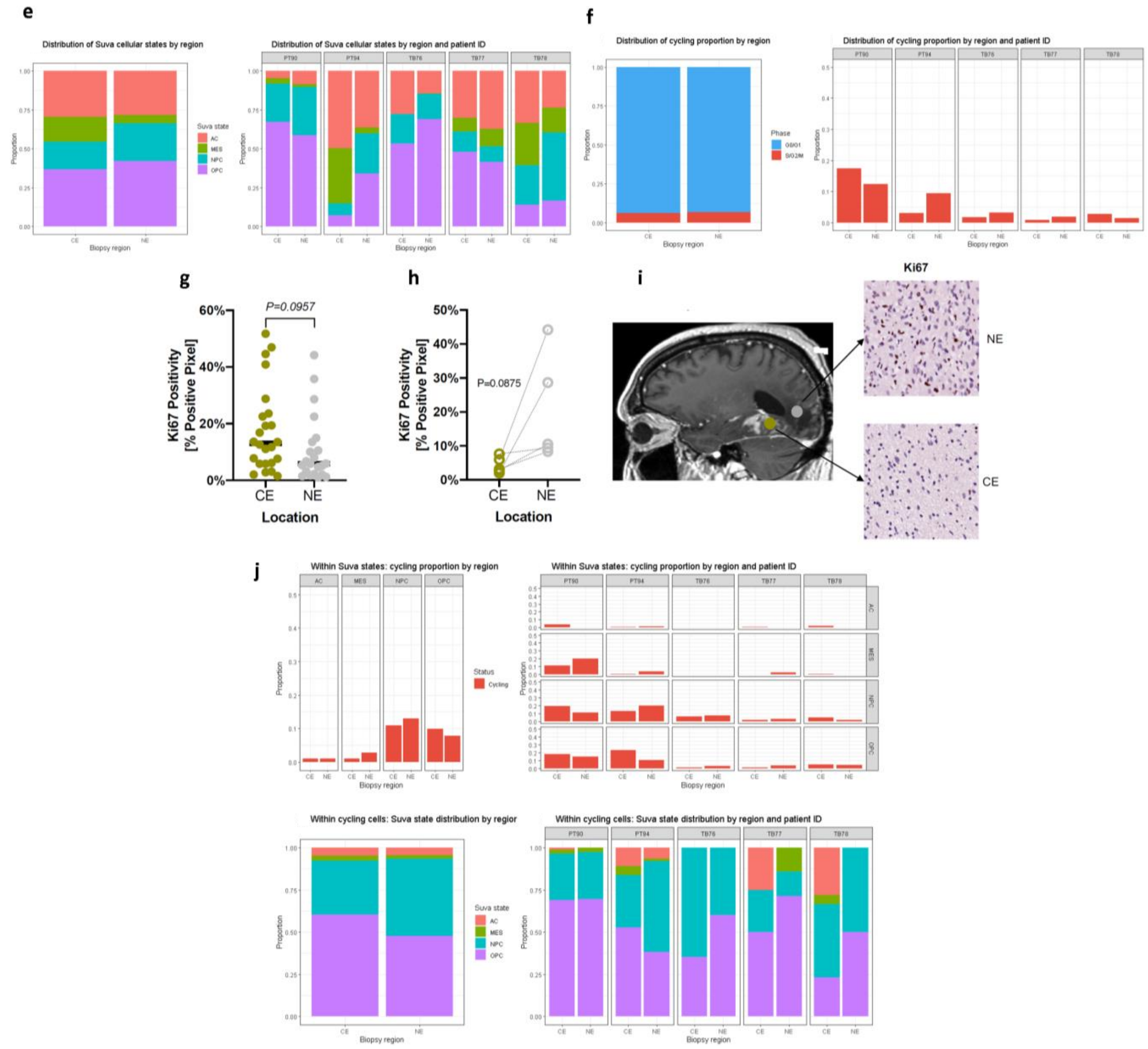
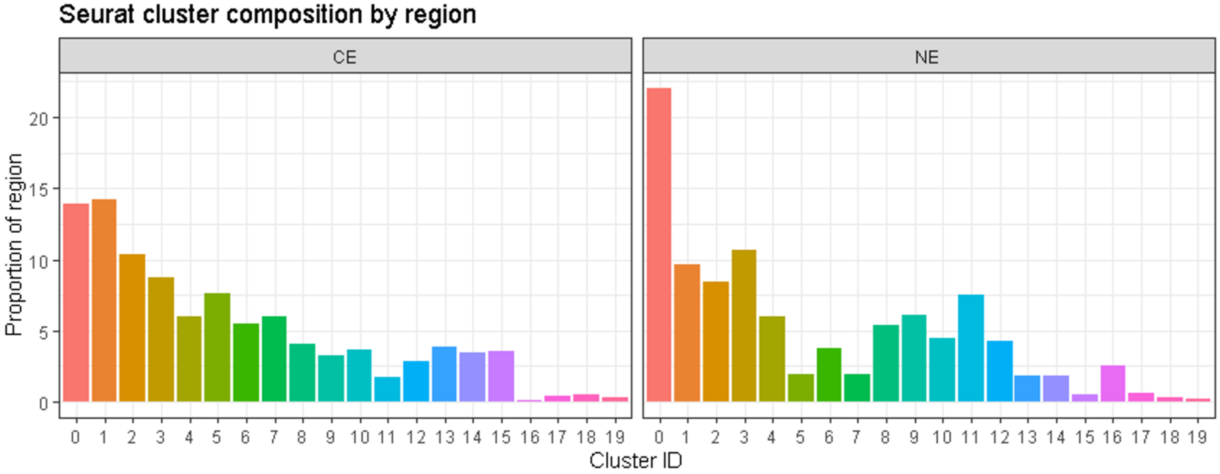
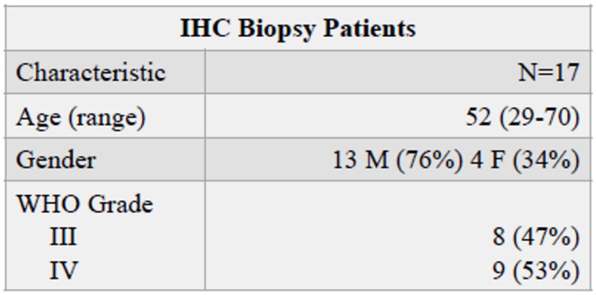


Figure 3-4. Characterizing tumor burden and composition in the NE region. (a) Malignant cells were labeled according to the region they were isolated from and visualized using t-SNE coordinates. **(b)** Quantification of malignant cell burden (proportion of all cells that are malignant) in the NE vs CE regions of each tumor. **(c)** Tumor burden relative to distance from center of tumor was modeled for the first 3 sequenced tumors by fitting the tumor:non-tumor oligodendrocyte ratio to a quadratic curve (solid line), with the x-asymptote (dotted line) representing predicted distance from the tumor center where tumor burden approaches 0. **(d)** Heatmap of tumor burden based on quadratic formula overlaid on pre-operative MR scan. **(e, f)** Proportion of malignant cellular states **(e)** and cycling cells **(f)** in the NE vs CE regions. **(g)** Percentage positive pixel quantification of immunohistochemistry staining for Ki67 stratified by region of acquisition. **(h)** Percentage positive pixel quantification of immunohistochemistry staining for Ki67 in enhancing and non-enhancing pairs from the same tumor. **(i)** Representative example of regions of Ki67 staining in enhancing and non-enhancing regions. **(j)** Cross-classification abundance analysis, combining cellular state groups with cycling status.

a

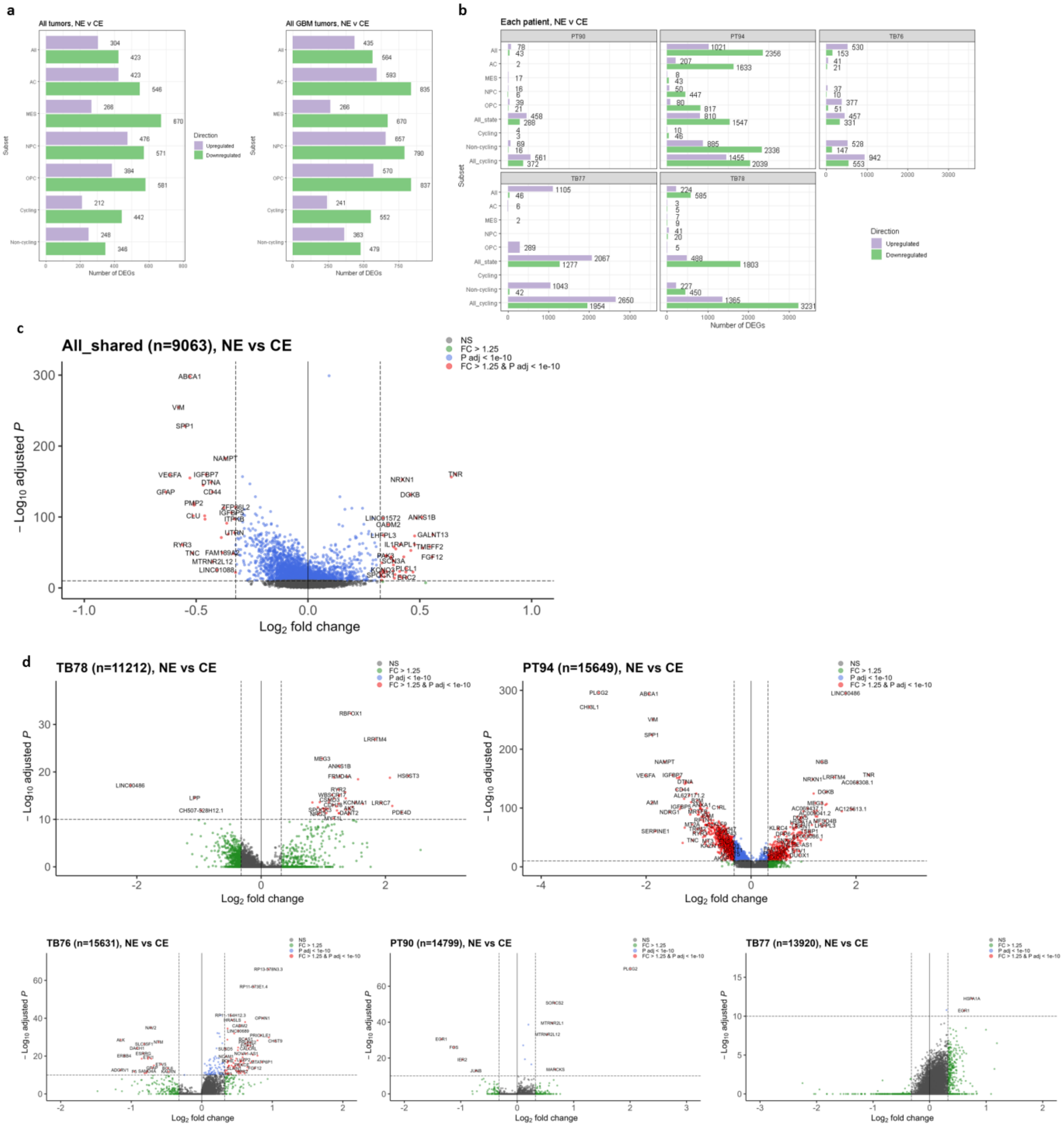


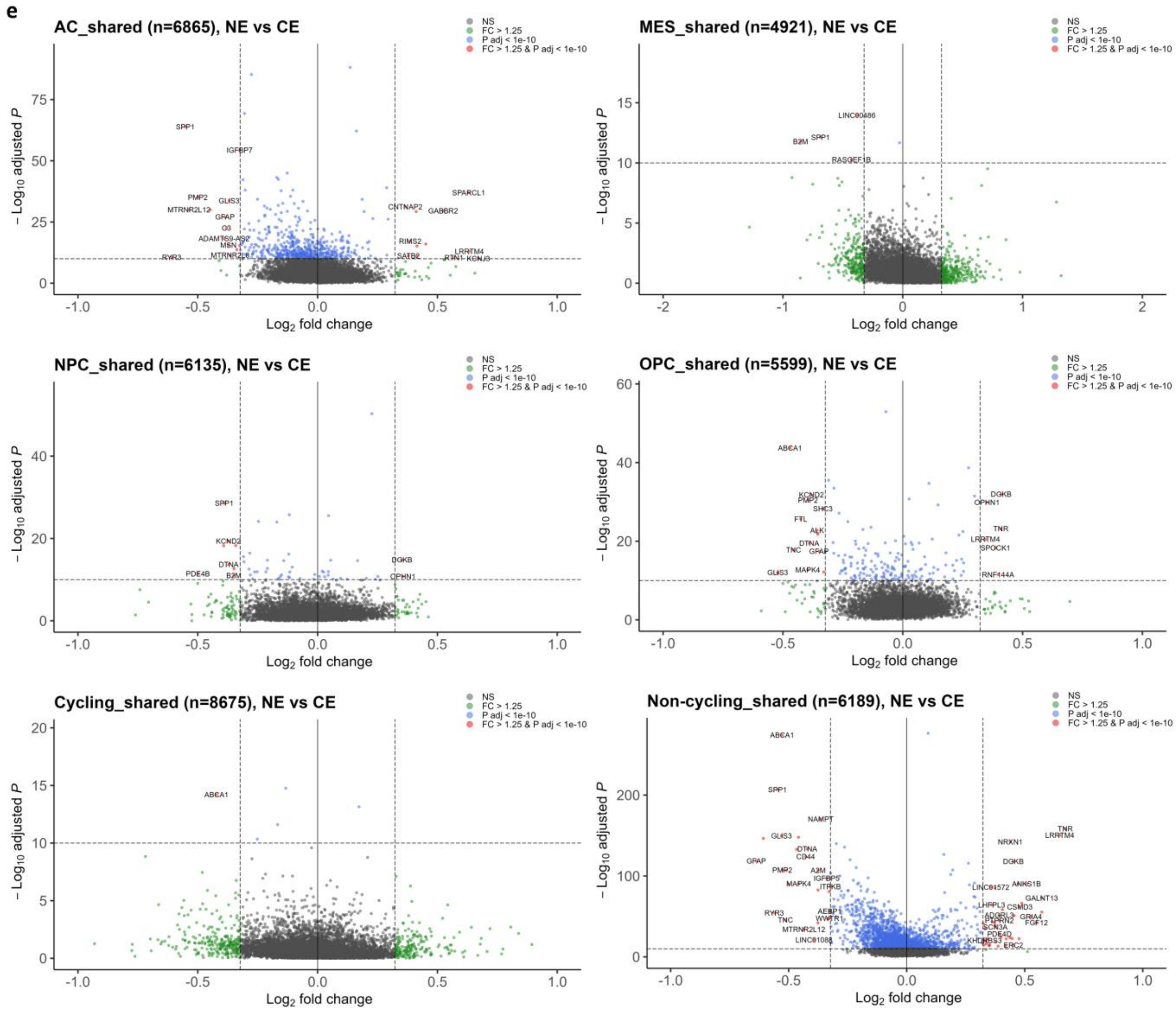
b



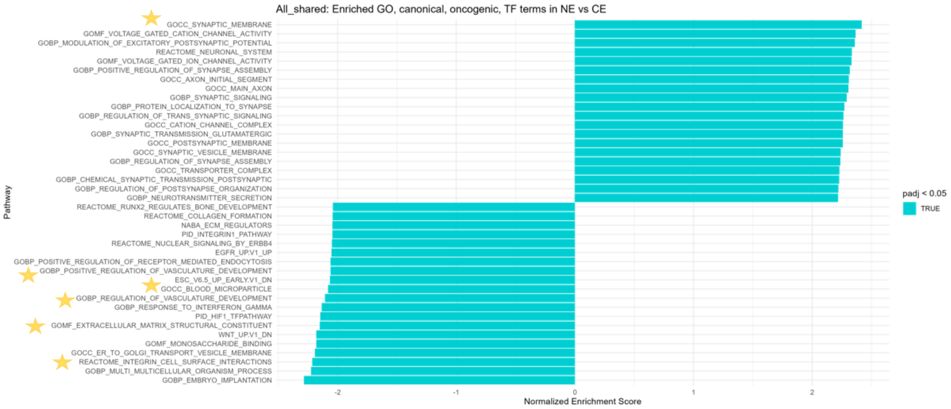
IHC Biopsy Patients	
Characteristic	N=17
Age (range)	52 (29-70)
Gender	13 M (76%) 4 F (34%)
WHO Grade III IV	 8 (47%) 9 (53%)

Supplementary Figure 3-4. Related to Fig. 3-4.
(a) Breakdown of expression-based cluster proportions by region. All 5 tumors from which we sequenced both regions were included. **(b)** Clinical and pathological characteristics of patients included in immunohistochemistry analyses (**Fig. 3-4g-i**).

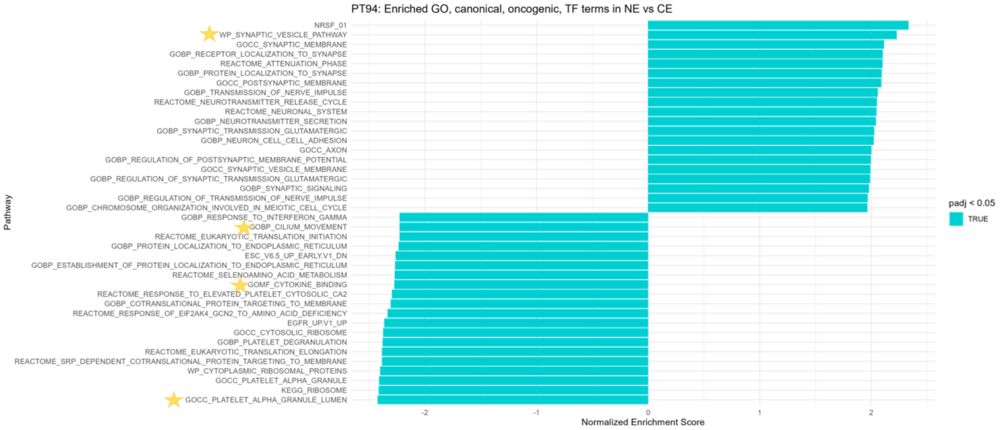

a
All tumors, NE v CE
All GBM tumors, NE v CE
Subset
All
AC
MES
NPC
OPC
Cycling
Non-cycling
Number of DEGs
Direction
Upregulated
Downregulated
304
423
423
546
266
670
476
571
394
581
212
442
248
346
435
564
593
835
266
670
657
790
570
837
241
552
383
479
b
Each patient, NE v CE
PT90
PT94
TB76
TB77
TB78
All_state
Non-cycling
All_cycling
Number of DEGs
c
All_shared (n=9063), NE vs CE
NS
FC > 1.25
P adj < 1e-10
FC > 1.25 & P adj < 1e-10
− Log10 adjusted P
Log2 fold change
ABCA1
VIM
SPP1
NAMPT
VEGFA
IGFBP7
DTNA
GFAP
CD44
PMP2
CLU
RYR3
TNC
MTRNR2L12
TNR
NRXN1
DGKB
ANKS1B
GALNT13
TMEFF2
FGF12
d
TB78 (n=11212), NE vs CE
PT94 (n=15649), NE vs CE
TB76 (n=15631), NE vs CE
PT90 (n=14799), NE vs CE
TB77 (n=13920), NE vs CE
RBFOX1
LRRTM4
MBG3
HS6ST3
LINC00486
LPP
PDE4D
PLCG2
CHI3L1
SERPINE1
NSG
DGKB
SORCS2
MTRNR2L1
MARCKS
EGR1
FOS
IER2
JUNB
HSPA1A


e
AC_shared (n=6865), NE vs CE
NS
FC > 1.25
P adj < 1e-10
FC > 1.25 & P adj < 1e-10
− Log10 adjusted P
Log2 fold change
SPP1
IGFBP7
PMP2
GLIS3
MTRNR2L12
GFAP
C3
ADAMTS9-AS2
MSN
RYR3
MTRNR2L8
SPARCL1
CNTNAP2
GABBR2
RIMS2
LRRTM4
SATB2
RTN1
KCNJ3
MES_shared (n=4921), NE vs CE
LINC00486
B2M
SPP1
RASGEF1B
NPC_shared (n=6135), NE vs CE
SPP1
KCND2
DTNA
PDE4B
B2M
DGKB
OPHN1
OPC_shared (n=5599), NE vs CE
ABCA1
KCND2
PMP2
SHC3
FTL
ALK
DTNA
TNC
GFAP
MAPK4
GLIS3
DGKB
OPHN1
TNR
LRRTM4
SPOCK1
RNF144A
Cycling_shared (n=8675), NE vs CE
ABCA1
Non-cycling_shared (n=6189), NE vs CE
ABCA1
SPP1
NAMPT
GLIS3
DTNA
CD44
GFAP
PMP2
A2M
IGFBP5
MAPK4
ITPKB
RYR3
TNC
AEBP1
WWTR1
MTRNR2L12
LINC01088
TNR
LRRTM4
NRXN1
DGKB
LINC01572
ANKS1B
GALNT13
LHFPL3
CSMD3
ADGRL3
PTPRN2
GRIA4
FGF12
SCN3A
PDE4D
KHDRBS3
ERC2

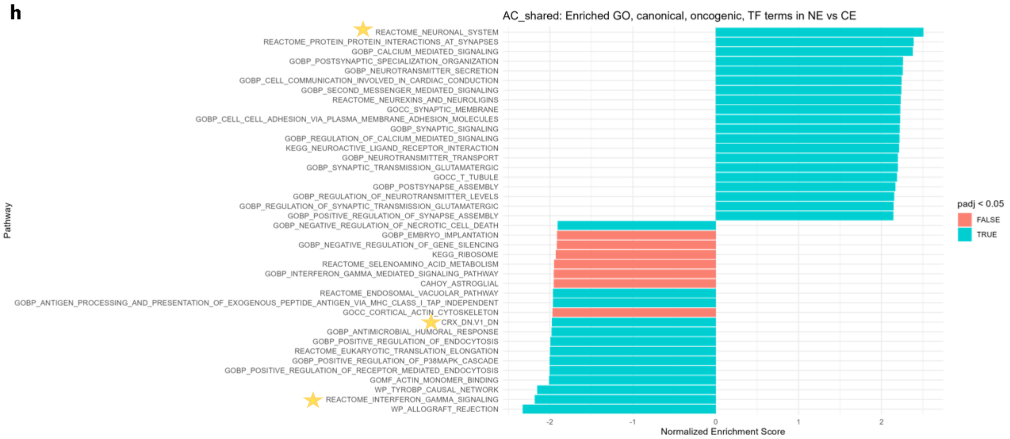
f


All_shared: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
Pathway
GOCC_SYNAPTIC_MEMBRANE
GOMF_VOLTAGE_GATED_CATION_CHANNEL_ACTIVITY
GOBP_MODULATION_OF_EXCITATORY_POSTSYNAPTIC_POTENTIAL
REACTOME_NEURONAL_SYSTEM
GOMF_VOLTAGE_GATED_ION_CHANNEL_ACTIVITY
GOBP_POSITIVE_REGULATION_OF_SYNAPSE_ASSEMBLY
GOCC_AXON_INITIAL_SEGMENT
GOCC_MAIN_AXON
GOBP_SYNAPTIC_SIGNALING
GOBP_PROTEIN_LOCALIZATION_TO_SYNAPSE
GOBP_REGULATION_OF_TRANS_SYNAPTIC_SIGNALING
GOCC_CATION_CHANNEL_COMPLEX
GOBP_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC
GOCC_POSTSYNAPTIC_MEMBRANE
GOCC_SYNAPTIC_VESICLE_MEMBRANE
GOBP_REGULATION_OF_SYNAPSE_ASSEMBLY
GOCC_TRANSPORTER_COMPLEX
GOBP_CHEMICAL_SYNAPTIC_TRANSMISSION_POSTSYNAPTIC
GOBP_REGULATION_OF_POSTSYNAPSE_ORGANIZATION
GOBP_NEUROTRANSMITTER_SECRETION
REACTOME_RUNX2_REGULATES_BONE_DEVELOPMENT
REACTOME_COLLAGEN_FORMATION
NABA_ECM_REGULATORS
PID_INTEGRIN1_PATHWAY
REACTOME_NUCLEAR_SIGNALING_BY_ERBB4
EGFR_UP.V1_UP
GOBP_POSITIVE_REGULATION_OF_RECEPTOR_MEDIATED_ENDOCYTOSIS
GOBP_POSITIVE_REGULATION_OF_VASCULATURE_DEVELOPMENT
ESC_V6.5_UP_EARLY.V1_DN
GOCC_BLOOD_MICROPARTICLE
GOBP_REGULATION_OF_VASCULATURE_DEVELOPMENT
GOBP_RESPONSE_TO_INTERFERON_GAMMA
PID_HIF1_TFPATHWAY
GOMF_EXTRACELLULAR_MATRIX_STRUCTURAL_CONSTITUENT
WNT_UP.V1_DN
GOMF_MONOSACCHARIDE_BINDING
GOCC_ER_TO_GOLGI_TRANSPORT_VESICLE_MEMBRANE
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
GOBP_MULTI_MULTICELLULAR_ORGANISM_PROCESS
GOBP_EMBRYO_IMPLANTATION
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
TRUE

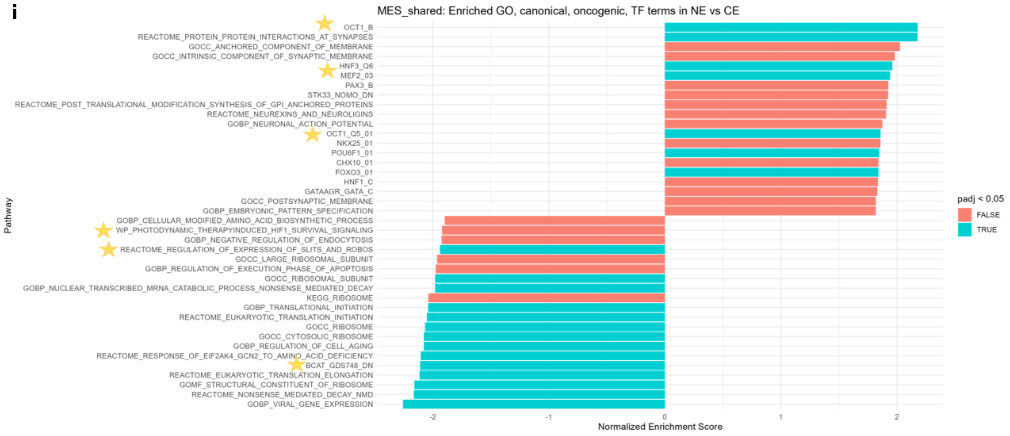
g


PT94: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
Pathway
NRSF_01
WP_SYNAPTIC_VESICLE_PATHWAY
GOCC_SYNAPTIC_MEMBRANE
GOBP_RECEPTOR_LOCALIZATION_TO_SYNAPSE
REACTOME_ATTENUATION_PHASE
GOBP_PROTEIN_LOCALIZATION_TO_SYNAPSE
GOCC_POSTSYNAPTIC_MEMBRANE
GOBP_TRANSMISSION_OF_NERVE_IMPULSE
REACTOME_NEUROTRANSMITTER_RELEASE_CYCLE
REACTOME_NEURONAL_SYSTEM
GOBP_NEUROTRANSMITTER_SECRETION
GOBP_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC
GOBP_NEURON_CELL_CELL_ADHESION
GOCC_AXON
GOBP_REGULATION_OF_POSTSYNAPTIC_MEMBRANE_POTENTIAL
GOCC_SYNAPTIC_VESICLE_MEMBRANE
GOBP_REGULATION_OF_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC
GOBP_SYNAPTIC_SIGNALING
GOBP_REGULATION_OF_TRANSMISSION_OF_NERVE_IMPULSE
GOBP_CHROMOSOME_ORGANIZATION_INVOLVED_IN_MEIOTIC_CELL_CYCLE
GOBP_RESPONSE_TO_INTERFERON_GAMMA
GOBP_CILIUM_MOVEMENT
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
GOBP_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
ESC_V6.5_UP_EARLY.V1_DN
GOBP_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
REACTOME_SELENOAMINO_ACID_METABOLISM
GOMF_CYTOKINE_BINDING
REACTOME_RESPONSE_TO_ELEVATED_PLATELET_CYTOSOLIC_CA2
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
EGFR_UP.V1_UP
GOCC_CYTOSOLIC_RIBOSOME
GOBP_PLATELET_DEGRANULATION
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
GOCC_PLATELET_ALPHA_GRANULE
KEGG_RIBOSOME
GOCC_PLATELET_ALPHA_GRANULE_LUMEN
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
TRUE

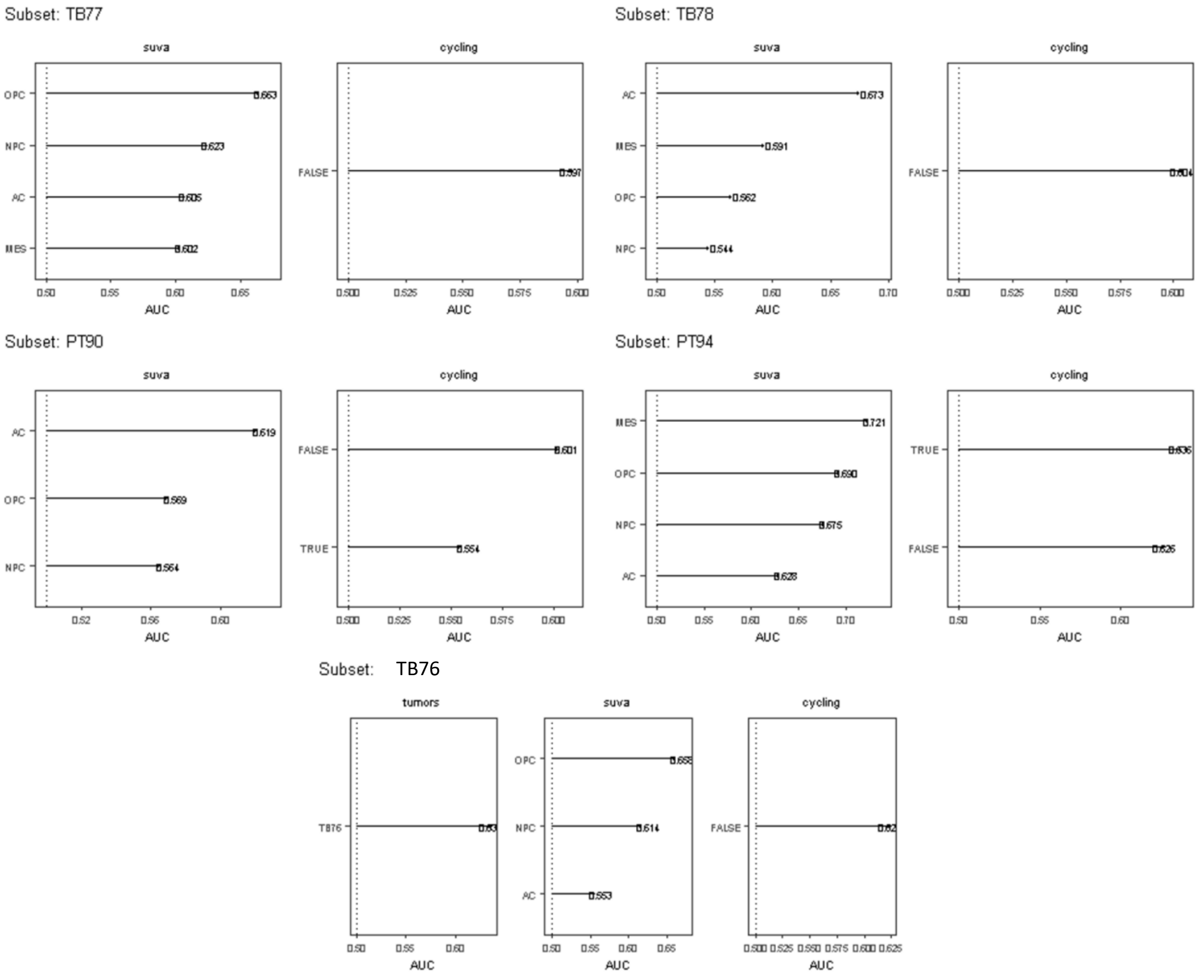
h


AC_shared: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
Pathway
REACTOME_NEURONAL_SYSTEM
REACTOME_PROTEIN_PROTEIN_INTERACTIONS_AT_SYNAPSES
GOBP_CALCIUM_MEDIATED_SIGNALING
GOBP_POSTSYNAPTIC_SPECIALIZATION_ORGANIZATION
GOBP_NEUROTRANSMITTER_SECRETION
GOBP_CELL_COMMUNICATION_INVOLVED_IN_CARDIAC_CONDUCTION
GOBP_SECOND_MESSENGER_MEDIATED_SIGNALING
REACTOME_NEUREXINS_AND_NEUROLIGINS
GOCC_SYNAPTIC_MEMBRANE
GOBP_CELL_CELL_ADHESION_VIA_PLASMA_MEMBRANE_ADHESION_MOLECULES
GOBP_SYNAPTIC_SIGNALING
GOBP_REGULATION_OF_CALCIUM_MEDIATED_SIGNALING
KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION
GOBP_NEUROTRANSMITTER_TRANSPORT
GOBP_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC
GOCC_T_TUBULE
GOBP_POSTSYNAPSE_ASSEMBLY
GOBP_REGULATION_OF_NEUROTRANSMITTER_LEVELS
GOBP_REGULATION_OF_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC
GOBP_POSITIVE_REGULATION_OF_SYNAPSE_ASSEMBLY
GOBP_NEGATIVE_REGULATION_OF_NECROTIC_CELL_DEATH
GOBP_EMBRYO_IMPLANTATION
GOBP_NEGATIVE_REGULATION_OF_GENE_SILENCING
KEGG_RIBOSOME
REACTOME_SELENOAMINO_ACID_METABOLISM
GOBP_INTERFERON_GAMMA_MEDIATED_SIGNALING_PATHWAY
CAHOY_ASTROGLIAL
REACTOME_ENDOSOMAL_VACUOLAR_PATHWAY
GOBP_ANTIGEN_PROCESSING_AND_PRESENTATION_OF_EXOGENOUS_PEPTIDE_ANTIGEN_VIA_MHC_CLASS_I_TAP_INDEPENDENT
GOCC_CORTICAL_ACTIN_CYTOSKELETON
CRX_DN.V1_DN
GOBP_ANTIMICROBIAL_HUMORAL_RESPONSE
GOBP_POSITIVE_REGULATION_OF_ENDOCYTOSIS
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
GOBP_POSITIVE_REGULATION_OF_P38MAPK_CASCADE
GOBP_POSITIVE_REGULATION_OF_RECEPTOR_MEDIATED_ENDOCYTOSIS
GOMF_ACTIN_MONOMER_BINDING
WP_TYROBP_CAUSAL_NETWORK
REACTOME_INTERFERON_GAMMA_SIGNALING
WP_ALLOGRAFT_REJECTION
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE

i


MES_shared: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
Pathway
OCT1_B
REACTOME_PROTEIN_PROTEIN_INTERACTIONS_AT_SYNAPSES
GOCC_ANCHORED_COMPONENT_OF_MEMBRANE
GOCC_INTRINSIC_COMPONENT_OF_SYNAPTIC_MEMBRANE
HNF3_Q6
MEF2_03
PAX3_B
STK33_NOMO_DN
REACTOME_POST_TRANSLATIONAL_MODIFICATION_SYNTHESIS_OF_GPI_ANCHORED_PROTEINS
REACTOME_NEUREXINS_AND_NEUROLIGINS
GOBP_NEURONAL_ACTION_POTENTIAL
OCT1_Q5_01
NKX25_01
POU6F1_01
CHX10_01
FOXO3_01
HNF1_C
GATAAGR_GATA_C
GOCC_POSTSYNAPTIC_MEMBRANE
GOBP_EMBRYONIC_PATTERN_SPECIFICATION
GOBP_CELLULAR_MODIFIED_AMINO_ACID_BIOSYNTHETIC_PROCESS
WP_PHOTODYNAMIC_THERAPYINDUCED_HIF1_SURVIVAL_SIGNALING
GOBP_NEGATIVE_REGULATION_OF_ENDOCYTOSIS
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS
GOCC_LARGE_RIBOSOMAL_SUBUNIT
GOBP_REGULATION_OF_EXECUTION_PHASE_OF_APOPTOSIS
GOCC_RIBOSOMAL_SUBUNIT
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
KEGG_RIBOSOME
GOBP_TRANSLATIONAL_INITIATION
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
GOCC_RIBOSOME
GOCC_CYTOSOLIC_RIBOSOME
GOBP_REGULATION_OF_CELL_AGING
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
BCAT_GDS748_DN
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
GOBP_VIRAL_GENE_EXPRESSION
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE

j


Subset: TB77
suva
OPC
NPC
AC
MES
0.663
0.623
0.605
0.602
cycling
FALSE
0.597
AUC
Subset: TB78
suva
AC
MES
OPC
NPC
0.673
0.591
0.562
0.544
cycling
FALSE
0.604
Subset: PT90
suva
AC
OPC
NPC
0.619
0.569
0.564
cycling
FALSE
TRUE
0.601
0.554
Subset: PT94
suva
MES
OPC
NPC
AC
0.721
0.690
0.675
0.628
cycling
TRUE
FALSE
0.636
0.626
Subset: TB76
tumors
TB76
0.63
suva
OPC
NPC
AC
0.668
0.614
0.553
cycling
FALSE
0.62
AUC

k

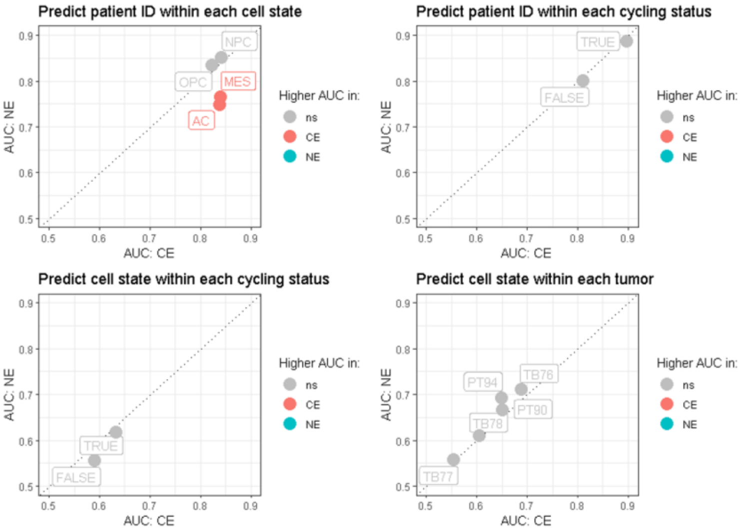
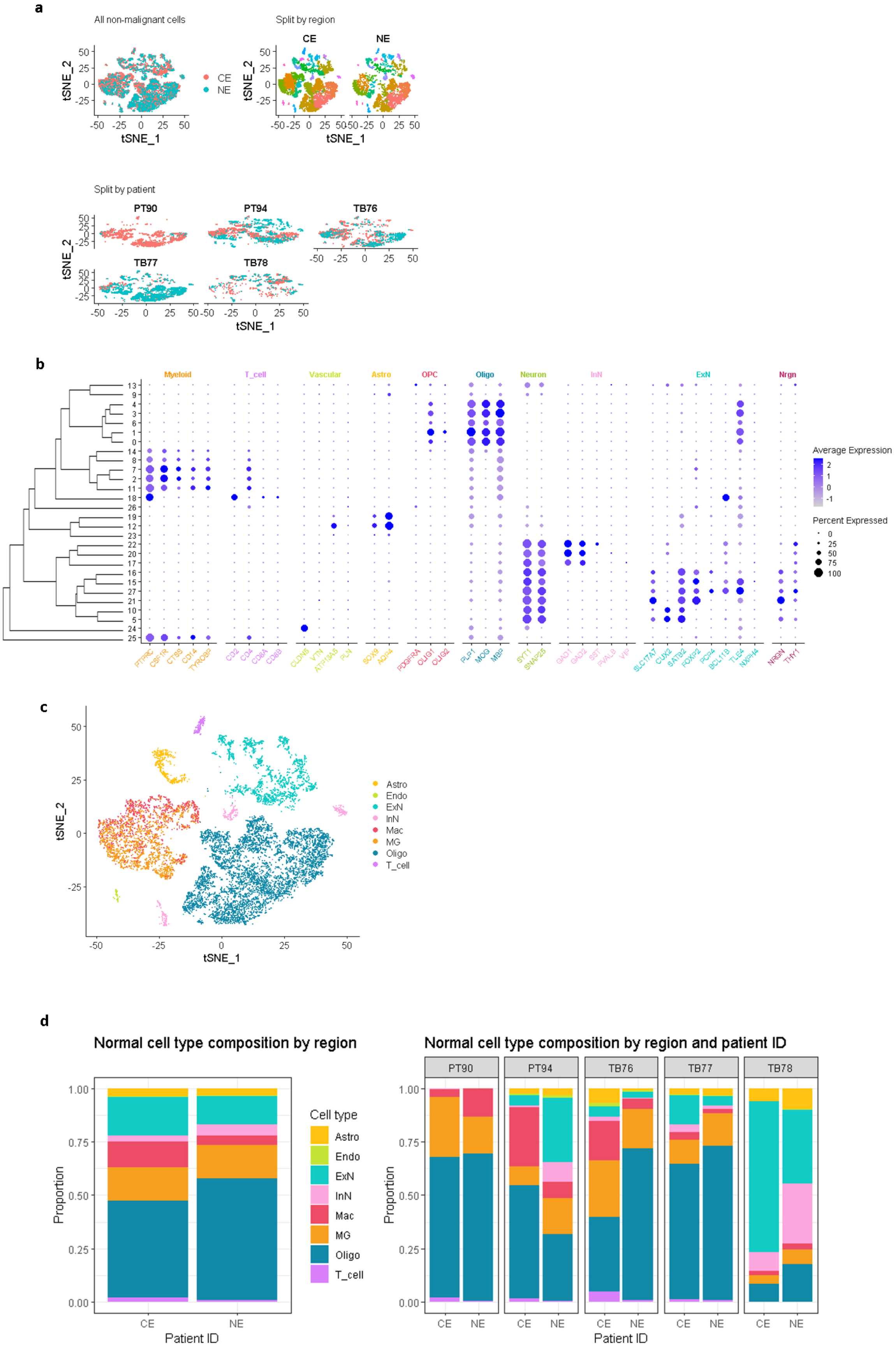

Predict patient ID within each cell state
NPC
OPC
MES
AC
Higher AUC in:
ns
CE
NE
AUC: NE
AUC: CE
Predict patient ID within each cycling status
TRUE
FALSE
Predict cell state within each cycling status
TRUE
FALSE
Predict cell state within each tumor
PT94
TB76
PT90
TB78
TB77

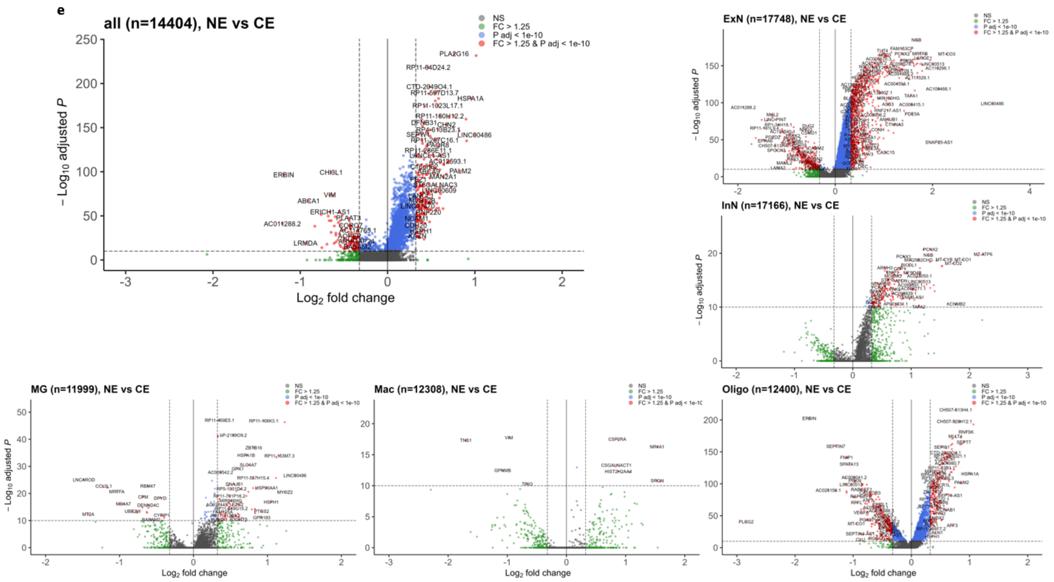
Figure 3-5. Region-specific molecular features of malignant cells.
Differential expression analysis was performed to compare the NE vs CE regions in the integrated dataset **(a)** and in individual tumors **(b)**. Numbers of differentially expressed genes (DEGs) are shown for each comparison. Cells were analyzed without grouping (“All”) and grouped by individual cellular states (“AC” through “OPC”) or cycling status (“Cycling” and “Non-cycling”). In each comparison that involved multiple tumors (e.g., AC-like cells in all tumors), conserved DEGs were identified across patients using the `FindConservedMarkers` function in Seurat, which performs the differential expression testing in each individual group and tests for conserved changes using p value combination techniques. In **(a)**, the grouping variable was patient identity, yielding conserved DEGs in the NE vs CE regions across patients. For the individual tumor analyses in **(b)**, conserved DEGs across cellular states and across cycling statuses were identified in a similar manner and correspond to the rows labeled “All_state” (all cellular states) and “All_cycling” (both cycling + non-cycling), respectively. **(c-e)** Volcano plots showing log fold change and significance for all genes that were considered in each differential expression comparison, which were genes expressed in at least 5% of the respective cell subsets. **(f-i)** The same differential expression comparisons were used for gene set enrichment analysis using a combination of the following databases: GO Biological Pathways, Canonical Pathways, Oncogenic Pathways, and Transcription Factor Targets [20]. Shown are the top up- and down-regulated pathways across all tumors **(f)**, the tumor with the most significantly dysregulated pathways, PT94 **(g)**, and the two cellular states with the most dysregulated pathways, AC-like **(h)** and MES-like **(i)**. **(j-k)** Cell type prioritization was performed using the Augur package to quantify and rank cell type-specific contributions to regional divergence. Prioritization was performed to assess NE vs CE separability in each tumor **(j)**, to assess change in intertumoral heterogeneity within each malignant populations across the two regions **(k, top)**, and differences across the two regions in cell state predictability (i.e., intratumoral heterogeneity) given cycling status or patient ID **(k, bottom)**.


a
TB76: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
REACTOME_SELENOAMINO_ACID_METABOLISM
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
GOCC_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_INFLUENZA_INFECTION
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
GOCC_RIBOSOMAL_SUBUNIT
GOCC_CYTOSOLIC_RIBOSOME
REACTOME_INTERACTION_BETWEEN_L1_AND_ANKYRINS
GOCC_LARGE_RIBOSOMAL_SUBUNIT
GOCC_NODE_OF_RANVIER
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_RRNA_PROCESSING
GOBP_SEMAPHORIN_PLEXIN_SIGNALING_PATHWAY_INVOLVED_IN_NEURON_PROJECTION_GUIDANCE
GOBP_PROTEIN_TRIMERIZATION
GOBP_INTERFERON_GAMMA_MEDIATED_SIGNALING_PATHWAY
GOCC_ANCHORED_COMPONENT_OF_PLASMA_MEMBRANE
WP_GENES_INVOLVED_IN_MALE_INFERTILITY
GOMF_SEMAPHORIN_RECEPTOR_ACTIVITY
GOCC_SEMAPHORIN_RECEPTOR_COMPLEX
GOBP_NEGATIVE_REGULATION_OF_LIPID_CATABOLIC_PROCESS
GOBP_MAMMARY_GLAND_ALVEOLUS_DEVELOPMENT
GOBP_REGULATION_OF_NUCLEAR_CELL_CYCLE_DNA_REPLICATION
PID_RXR_VDR_PATHWAY
GOMF_PROLINE_RICH_REGION_BINDING
GOMF_3_5_CYCLIC_AMP_PHOSPHODIESTERASE_ACTIVITY
GOBP_ORGAN_GROWTH
GOBP_CAMP_METABOLIC_PROCESS
GOMF_EPIDERMAL_GROWTH_FACTOR_RECEPTOR_BINDING
GOMF_TRANSMEMBRANE_RECEPTOR_PROTEIN_KINASE_ACTIVITY
GOBP_REGULATION_OF_SPROUTING_ANGIOGENESIS
REACTOME_HS_GAG_DEGRADATION
WP_COMPLEMENT_AND_COAGULATION_CASCADES
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE
TB77: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
GOBP_RESPONSE_TO_STEROL
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
PID_ANGIOPOIETIN_RECEPTOR_PATHWAY
STK33_UP
GOBP_B_CELL_DIFFERENTIATION
GOBP_RESPONSE_TO_FATTY_ACID
GOBP_PANCREAS_DEVELOPMENT
GOBP_REGULATION_OF_LEUKOCYTE_DIFFERENTIATION
GOBP_INTEGRATED_STRESS_RESPONSE_SIGNALING
GOBP_B_CELL_PROLIFERATION
GOBP_REGULATION_OF_B_CELL_PROLIFERATION
GOBP_REGULATION_OF_B_CELL_ACTIVATION
GOBP_REGULATION_OF_MACROPHAGE_DERIVED_FOAM_CELL_DIFFERENTIATION
STK33_NOMO_UP
GOBP_ER_NUCLEUS_SIGNALING_PATHWAY
GOBP_LIPOPOLYSACCHARIDE_MEDIATED_SIGNALING_PATHWAY
SIG_PIP3_SIGNALING_IN_B_LYMPHOCYTES
WP_LUNG_FIBROSIS
GOBP_CORONARY_VASCULATURE_MORPHOGENESIS
BIOCARTA_NKT_PATHWAY
GOBP_SYNAPTIC_SIGNALING
GOBP_VESICLE_MEDIATED_TRANSPORT_IN_SYNAPSE
GOCC_POSTSYNAPSE
GOCC_POSTSYNAPTIC_SPECIALIZATION_MEMBRANE
GOCC_TRANSPORTER_COMPLEX
GOBP_REGULATION_OF_TRANS_SYNAPTIC_SIGNALING
REACTOME_NEUROTRANSMITTER_RECEPTORS_AND_POSTSYNAPTIC_SIGNAL_TRANSMISSION
GOCC_PRESYNAPTIC_MEMBRANE
WP_DISRUPTION_OF_POSTSYNAPTIC_SIGNALLING_BY_CNV
REACTOME_PROTEIN_PROTEIN_INTERACTIONS_AT_SYNAPSES
GOMF_NEUROTRANSMITTER_RECEPTOR_ACTIVITY
REACTOME_NEUREXINS_AND_NEUROLIGINS
GOBP_REGULATION_OF_SYNAPTIC_PLASTICITY
GOCC_POSTSYNAPTIC_MEMBRANE
GOCC_NEURON_TO_NEURON_SYNAPSE
GOCC_CATION_CHANNEL_COMPLEX
CAHOY_NEURONAL
REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES
GOCC_SYNAPTIC_MEMBRANE
REACTOME_NEURONAL_SYSTEM
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
TRUE
TB78: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
GOCC_CATION_CHANNEL_COMPLEX
REACTOME_NEURONAL_SYSTEM
GOCC_TRANSPORTER_COMPLEX
GOCC_SYNAPTIC_MEMBRANE
GOMF_GATED_CHANNEL_ACTIVITY
GOCC_POSTSYNAPTIC_MEMBRANE
REACTOME_TRANSMISSION_ACROSS_CHEMICAL_SYNAPSES
GOMF_VOLTAGE_GATED_ION_CHANNEL_ACTIVITY
GOMF_CATION_CHANNEL_ACTIVITY
GOMF_VOLTAGE_GATED_CATION_CHANNEL_ACTIVITY
GOBP_SYNAPTIC_SIGNALING
GOCC_POSTSYNAPSE
GOCC_POSTSYNAPTIC_SPECIALIZATION_MEMBRANE
REACTOME_NEUROTRANSMITTER_RECEPTORS_AND_POSTSYNAPTIC_SIGNAL_TRANSMISSION
GOMF_PASSIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
GOBP_REGULATION_OF_TRANS_SYNAPTIC_SIGNALING
GOCC_POSTSYNAPTIC_DENSITY_MEMBRANE
GOCC_SYNAPSE
GOBP_REGULATION_OF_TRANSMEMBRANE_TRANSPORT
GOBP_REGULATION_OF_TRANSPORTER_ACTIVITY
GOBP_NEURAL_TUBE_DEVELOPMENT
WP_MIRNA_REGULATION_OF_DNA_DAMAGE_RESPONSE
PID_TCPTP_PATHWAY
REACTOME_RHOBTB2_GTPASE_CYCLE
KEGG_P53_SIGNALING_PATHWAY
REACTOME_RHOBTB_GTPASE_CYCLE
GOBP_CENTROMERE_COMPLEX_ASSEMBLY
GOBP_COLLAGEN_FIBRIL_ORGANIZATION
GOCC_PLATELET_ALPHA_GRANULE_LUMEN
GOBP_REGULATION_OF_CELL_CYCLE_G1_S_PHASE_TRANSITION
ATXN7L3_TARGET_GENES
GOBP_REGULATION_OF_MESENCHYMAL_CELL_PROLIFERATION
CTR9_TARGET_GENES
WP_ATM_SIGNALING_PATHWAY
GOBP_NEGATIVE_REGULATION_OF_CELL_CYCLE_G1_S_PHASE_TRANSITION
GOBP_REGULATION_OF_NEUROBLAST_PROLIFERATION
WP_TUMOR_SUPPRESSOR_ACTIVITY_OF_SMARCB1
WP_NEURAL_CREST_DIFFERENTIATION
GOBP_REGULATION_OF_NEURAL_PRECURSOR_CELL_PROLIFERATION
PSMB5_TARGET_GENES
Pathway
-2
-1
0
1
2
3
Normalized Enrichment Score
padj < 0.05
TRUE
PT90: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
REACTOME_SELENOAMINO_ACID_METABOLISM
GOCC_CYTOSOLIC_RIBOSOME
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
GOBP_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
GOBP_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
GOCC_RIBOSOMAL_SUBUNIT
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
GOCC_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
REACTOME_SIGNALING_BY_ROBO_RECEPTORS
GOBP_PROTEIN_TARGETING_TO_MEMBRANE
WP_HOSTPATHOGEN_INTERACTION_OF_HUMAN_CORONA_VIRUSES_MAPK_SIGNALING
MAP2K1_TARGET_GENES
SA_B_CELL_RECEPTOR_COMPLEXES
GOBP_MRNA_TRANSCRIPTION
PID_MAPK_TRK_PATHWAY
GOBP_POSITIVE_REGULATION_OF_PRI_MIRNA_TRANSCRIPTION_BY_RNA_POLYMERASE_II
REACTOME_SUMOYLATION_OF_UBIQUITINYLATION_PROTEINS
GOBP_PRI_MIRNA_TRANSCRIPTION_BY_RNA_POLYMERASE_II
PID_IL12_STAT4_PATHWAY
REACTOME_INTERLEUKIN_10_SIGNALING
SRF_C
PID_IL12_2PATHWAY
SRF_01
RELA_DN.V1_UP
PID_AP1_PATHWAY
PID_NFAT_TFPATHWAY
PID_AVB3_OPN_PATHWAY
REACTOME_NGF_STIMULATED_TRANSCRIPTION
PID_CD8_TCR_DOWNSTREAM_PATHWAY
REACTOME_NUCLEAR_EVENTS_KINASE_AND_TRANSCRIPTION_FACTOR_ACTIVATION
-2
-1
0
1
2
3
Normalized Enrichment Score
padj < 0.05
TRUE

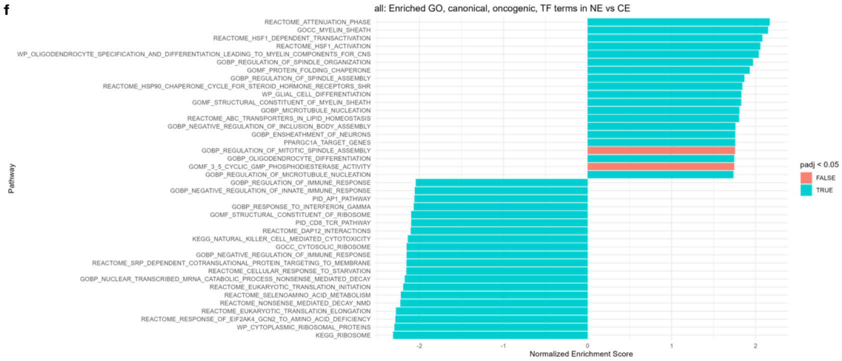
Supplementary Figure 3-5. Related to Fig. 3-5. Remaining GSEA results not visualized in **Fig. 3-5**. NE vs CE within patients **(a)** and within malignant cellular states **(b)**.


a
All non-malignant cells
CE
NE
Split by region
CE
NE
tSNE_1
tSNE_2
Split by patient
PT90
PT94
TB76
TB77
TB78
b
Myeloid
T_cell
Vascular
Astro
OPC
Oligo
Neuron
InN
ExN
Nrgn
PTPRC
CSF1R
CTSS
CD14
TYROBP
CD2
CD4
CD8A
CD8B
CLDN5
VTN
ATP13A5
PLN
SOX9
AQP4
PDGFRA
OLIG1
OLIG2
PLP1
MOG
MBP
SYT1
SNAP25
GAD1
GAD2
SST
PVALB
VIP
SLC17A7
CUX2
SATB2
FOXP2
PCP4
BCL11B
TLE4
NXPH4
NRGN
THY1
Average Expression
Percent Expressed
c
Astro
Endo
ExN
InN
Mac
MG
Oligo
T_cell
d
Normal cell type composition by region
Normal cell type composition by region and patient ID
Cell type
Proportion
Patient ID
CE
NE

e


all (n=14404), NE vs CE
NS
FC > 1.25
P adj < 1e-10
FC > 1.25 & P adj < 1e-10
− Log10 adjusted P
Log2 fold change
PLA2G16
RP11-94D24.2
HSPA1A
LINC00486
ERBIN
CHI3L1
ABCA1
VIM
ERICH1-AS1
AC011288.2
LRMDA
ExN (n=17748), NE vs CE
InN (n=17166), NE vs CE
MG (n=11999), NE vs CE
Mac (n=12308), NE vs CE
Oligo (n=12400), NE vs CE

f

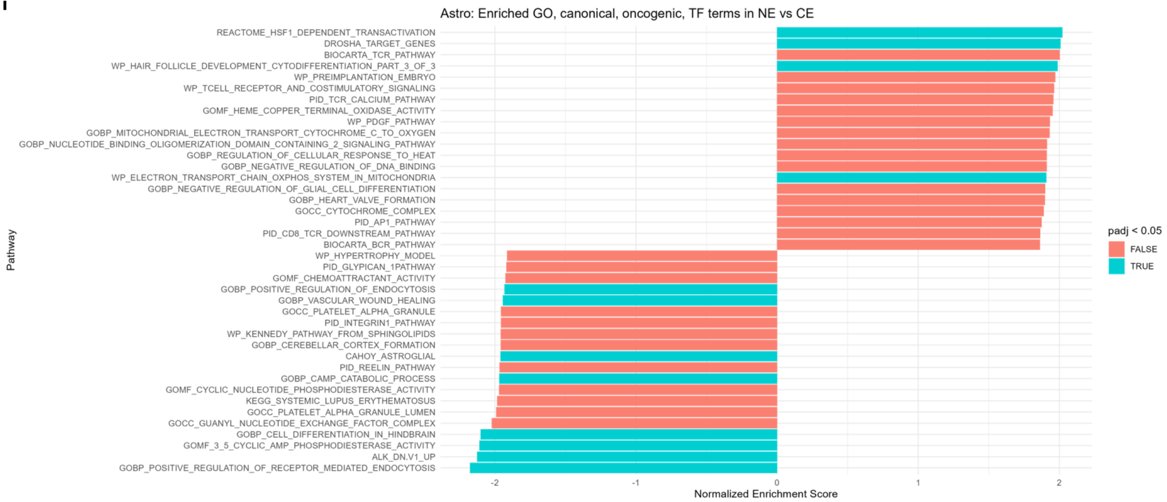

all: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
REACTOME_ATTENUATION_PHASE
GOCC_MYELIN_SHEATH
REACTOME_HSF1_DEPENDENT_TRANSACTIVATION
REACTOME_HSF1_ACTIVATION
WP_OLIGODENDROCYTE_SPECIFICATION_AND_DIFFERENTIATION_LEADING_TO_MYELIN_COMPONENTS_FOR_CNS
GOBP_REGULATION_OF_SPINDLE_ORGANIZATION
GOMF_PROTEIN_FOLDING_CHAPERONE
GOBP_REGULATION_OF_SPINDLE_ASSEMBLY
REACTOME_HSP90_CHAPERONE_CYCLE_FOR_STEROID_HORMONE_RECEPTORS_SHR
WP_GLIAL_CELL_DIFFERENTIATION
GOMF_STRUCTURAL_CONSTITUENT_OF_MYELIN_SHEATH
GOBP_MICROTUBULE_NUCLEATION
REACTOME_ABC_TRANSPORTERS_IN_LIPID_HOMEOSTASIS
GOBP_NEGATIVE_REGULATION_OF_INCLUSION_BODY_ASSEMBLY
GOBP_ENSHEATHMENT_OF_NEURONS
PPARGC1A_TARGET_GENES
GOBP_REGULATION_OF_MITOTIC_SPINDLE_ASSEMBLY
GOBP_OLIGODENDROCYTE_DIFFERENTIATION
GOMF_3_5_CYCLIC_GMP_PHOSPHODIESTERASE_ACTIVITY
GOBP_REGULATION_OF_MICROTUBULE_NUCLEATION
GOBP_REGULATION_OF_IMMUNE_RESPONSE
GOBP_NEGATIVE_REGULATION_OF_INNATE_IMMUNE_RESPONSE
PID_AP1_PATHWAY
GOBP_RESPONSE_TO_INTERFERON_GAMMA
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
PID_CD8_TCR_PATHWAY
REACTOME_DAP12_INTERACTIONS
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
GOCC_CYTOSOLIC_RIBOSOME
GOBP_NEGATIVE_REGULATION_OF_IMMUNE_RESPONSE
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
REACTOME_SELENOAMINO_ACID_METABOLISM
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
KEGG_RIBOSOME
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE

g

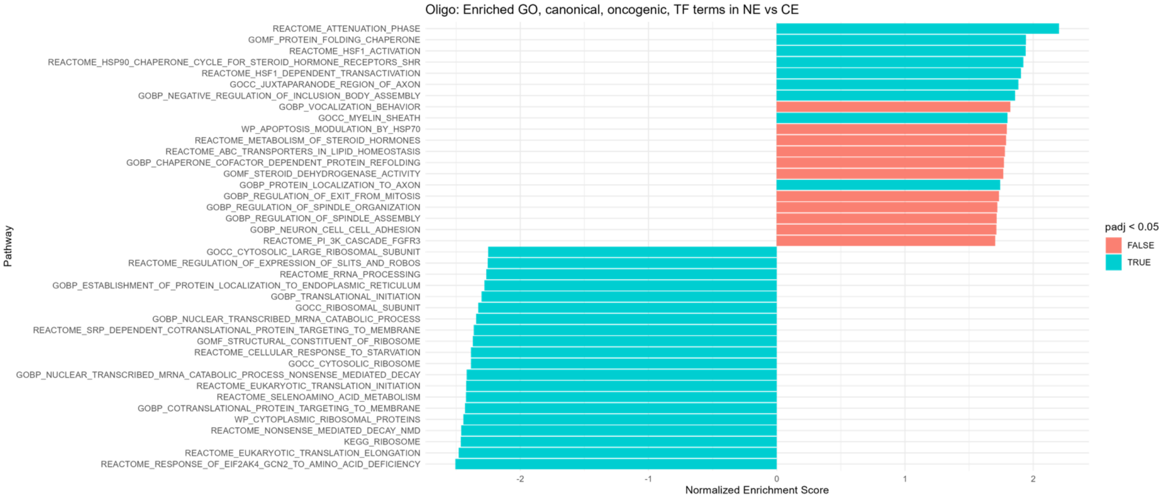

ExN: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOCC_CYTOSOLIC_RIBOSOME
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOBP_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
REACTOME_SELENOAMINO_ACID_METABOLISM
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
WP_ELECTRON_TRANSPORT_CHAIN_OXPHOS_SYSTEM_IN_MITOCHONDRIA
GOBP_OXIDATIVE_PHOSPHORYLATION
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
GOBP_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
KEGG_OXIDATIVE_PHOSPHORYLATION
GOCC_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT
KEGG_PARKINSONS_DISEASE
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS
GOBP_TRANSLATIONAL_INITIATION
REACTOME_PRESYNAPTIC_DEPOLARIZATION_AND_CALCIUM_CHANNEL_OPENING
GOMF_NEUROTRANSMITTER_RECEPTOR_ACTIVITY
GOBP_SYNAPTIC_MEMBRANE_ADHESION
PID_NFAT_TFPATHWAY
GOMF_GLUTAMATE_RECEPTOR_ACTIVITY
GOMF_VOLTAGE_GATED_ION_CHANNEL_ACTIVITY
WP_DISRUPTION_OF_POSTSYNAPTIC_SIGNALLING_BY_CNV
GOMF_GATED_CHANNEL_ACTIVITY
GOBP_CELL_CELL_SIGNALING_INVOLVED_IN_CARDIAC_CONDUCTION
GOBP_SYNAPSE_ASSEMBLY
GOBP_NEURON_CELL_CELL_ADHESION
GOMF_VOLTAGE_GATED_CATION_CHANNEL_ACTIVITY
GOBP_POSITIVE_REGULATION_OF_SYNAPSE_ASSEMBLY
GOCC_VOLTAGE_GATED_CALCIUM_CHANNEL_COMPLEX
GOMF_LIGAND_GATED_ION_CHANNEL_ACTIVITY
GOCC_NEUROTRANSMITTER_RECEPTOR_COMPLEX
PID_ENDOTHELIN_PATHWAY
GOMF_IONOTROPIC_GLUTAMATE_RECEPTOR_ACTIVITY
GOBP_REGULATION_OF_POSTSYNAPTIC_MEMBRANE_POTENTIAL
GOMF_LIGAND_GATED_CATION_CHANNEL_ACTIVITY
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
InN: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
WP_ELECTRON_TRANSPORT_CHAIN_OXPHOS_SYSTEM_IN_MITOCHONDRIA
KEGG_OXIDATIVE_PHOSPHORYLATION
KEGG_PARKINSONS_DISEASE
GOBP_OXIDATIVE_PHOSPHORYLATION
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS
REACTOME_THE_CITRIC_ACID_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT
GOBP_ATP_SYNTHESIS_COUPLED_ELECTRON_TRANSPORT
WP_OXIDATIVE_PHOSPHORYLATION
GOCC_RESPIRASOME
GOBP_ATP_METABOLIC_PROCESS
GOCC_RESPIRATORY_CHAIN_COMPLEX
GOMF_ELECTRON_TRANSFER_ACTIVITY
GOCC_INNER_MITOCHONDRIAL_MEMBRANE_PROTEIN_COMPLEX
GOBP_PROTON_TRANSMEMBRANE_TRANSPORT
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT
GOBP_RESPIRATORY_ELECTRON_TRANSPORT_CHAIN
DROSHA_TARGET_GENES
KEGG_HUNTINGTONS_DISEASE
GOBP_RIBONUCLEOSIDE_TRIPHOSPHATE_BIOSYNTHETIC_PROCESS
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
GOBP_PROTEIN_O_LINKED_GLYCOSYLATION_VIA_THREONINE
GOCC_NEUROTRANSMITTER_RECEPTOR_COMPLEX
CDP_02
CART1_01
WP_SPLICING_FACTOR_NOVA_REGULATED_SYNAPTIC_PROTEINS
CDPCR1_01
OCT1_B
WP_NO_METABOLISM_IN_CYSTIC_FIBROSIS
GOMF_TRANSMITTER_GATED_CHANNEL_ACTIVITY
OCT1_05
GOMF_IONOTROPIC_GLUTAMATE_RECEPTOR_ACTIVITY
OCT1_02
COMP1_01
GOCC_AXOLEMMA
GOMF_GLUTAMATE_RECEPTOR_ACTIVITY
GOBP_CELLULAR_RESPONSE_TO_CALCIUM_ION
GOBP_HEPARIN_METABOLIC_PROCESS
GOBP_CALCIUM_DEPENDENT_CELL_CELL_ADHESION_VIA_PLASMA_MEMBRANE_CELL_ADHESION_MOLECULES
GOBP_POSITIVE_REGULATION_OF_BLOOD_PRESSURE
GOBP_NEURON_CELL_CELL_ADHESION
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE


h
MG: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
WP_ELECTRON_TRANSPORT_CHAIN_OXPHOS_SYSTEM_IN_MITOCHONDRIA
REACTOME_HSF1_ACTIVATION
REACTOME_ATTENUATION_PHASE
REACTOME_HSF1_DEPENDENT_TRANSACTIVATION
GOCC_RESPIRATORY_CHAIN_COMPLEX
GOBP_PROTEIN_REFOLDING
KEGG_OXIDATIVE_PHOSPHORYLATION
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS
KEGG_PARKINSONS_DISEASE
BIOCARTA_FREE_PATHWAY
GOCC_RESPIRASOME
DROSHA_TARGET_GENES
GOBP_DE_NOVO_PROTEIN_FOLDING
TFAM_TARGET_GENES
GOBP_REGULATION_OF_CELLULAR_RESPONSE_TO_HEAT
GOBP_CHAPERONE_COFACTOR_DEPENDENT_PROTEIN_REFOLDING
GOBP_ATP_SYNTHESIS_COUPLED_ELECTRON_TRANSPORT
PPARGC1A_TARGET_GENES
WP_OXIDATIVE_PHOSPHORYLATION
GOBP_NEGATIVE_REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER_IN_RESPONSE_TO_STRESS
GOMF_TRANSFERASE_ACTIVITY_TRANSFERRING_PENTOSYL_GROUPS
GOBP_RESPONSE_TO_TYPE_I_INTERFERON
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
CORDENONSI_YAP_CONSERVED_SIGNATURE
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM
GOBP_FC_RECEPTOR_MEDIATED_STIMULATORY_SIGNALING_PATHWAY
REACTOME_FCGR3A_MEDIATED_IL10_SYNTHESIS
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
GOBP_MAST_CELL_ACTIVATION
BIOCARTA_ARAP_PATHWAY
GOMF_LIGASE_ACTIVITY_FORMING_CARBON_NITROGEN_BONDS
GOBP_NAD_METABOLIC_PROCESS
GOBP_NEGATIVE_REGULATION_OF_TYPE_I_INTERFERON_PRODUCTION
WP_WNT_SIGNALING_PATHWAY_NETPATH
GOBP_NEGATIVE_REGULATION_OF_OXIDATIVE_STRESS_INDUCED_NEURON_DEATH
GOBP_MIRNA_METABOLIC_PROCESS
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
REACTOME_FCERI_MEDIATED_CA_2_MOBILIZATION
GOCC_ACTIN_FILAMENT
KEGG_FC_GAMMA_R_MEDIATED_PHAGOCYTOSIS
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE
Mac: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
KEGG_TYPE_I_DIABETES_MELLITUS
KEGG_GRAFT_VERSUS_HOST_DISEASE
KEGG_ALLOGRAFT_REJECTION
KEGG_AUTOIMMUNE_THYROID_DISEASE
REACTOME_NUCLEAR_EVENTS_KINASE_AND_TRANSCRIPTION_FACTOR_ACTIVATION
GOCC_MHC_PROTEIN_COMPLEX
GOCC_CLATHRIN_COATED_ENDOCYTIC_VESICLE_MEMBRANE
GOCC_MHC_CLASS_II_PROTEIN_COMPLEX
GOBP_POSITIVE_REGULATION_OF_LEUKOCYTE_PROLIFERATION
GOMF_PEPTIDE_ANTIGEN_BINDING
REACTOME_NGF_STIMULATED_TRANSCRIPTION
KEGG_ASTHMA
GOBP_POSITIVE_REGULATION_OF_ALPHA_BETA_T_CELL_ACTIVATION
GOCC_CLATHRIN_COATED_ENDOCYTIC_VESICLE
GOBP_POSITIVE_REGULATION_OF_CELL_ACTIVATION
GOBP_POSITIVE_REGULATION_OF_CD4_POSITIVE_ALPHA_BETA_T_CELL_ACTIVATION
GOBP_REGULATION_OF_ALPHA_BETA_T_CELL_DIFFERENTIATION
GOBP_POSITIVE_REGULATION_OF_B_CELL_PROLIFERATION
REACTOME_PD_1_SIGNALING
GOBP_REGULATION_OF_CD4_POSITIVE_ALPHA_BETA_T_CELL_ACTIVATION
GOBP_MONOSACCHARIDE_CATABOLIC_PROCESS
BMI1_DN_MEL18_DN.V1_UP
KEGG_RIBOSOME
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
REACTOME_SELENOAMINO_ACID_METABOLISM
TGASTMAGC_NFE2_01
WP_GLYCOGEN_SYNTHESIS_AND_DEGRADATION
GOBP_PLASMA_LIPOPROTEIN_PARTICLE_CLEARANCE
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
BMI1_DN.V1_UP
GOBP_STEROL_TRANSPORT
GOBP_REGULATION_OF_PLASMA_LIPOPROTEIN_PARTICLE_LEVELS
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
WP_METABOLIC_REPROGRAMMING_IN_COLON_CANCER
KEGG_GLYCOLYSIS_GLUCONEOGENESIS
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX
KEGG_LYSOSOME
GOMF_CHOLESTEROL_BINDING
WP_GLYCOLYSIS_AND_GLUCONEOGENESIS
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
TRUE
T_cell: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
REACTOME_ATTENUATION_PHASE
KEGG_GLYCOSAMINOGLYCAN_BIOSYNTHESIS_KERATAN_SULFATE
REACTOME_HSF1_DEPENDENT_TRANSACTIVATION
GOBP_NEGATIVE_REGULATION_OF_INCLUSION_BODY_ASSEMBLY
GOBP_PHENOL_CONTAINING_COMPOUND_BIOSYNTHETIC_PROCESS
GOBP_PROTEIN_REFOLDING
BIOCARTA_CFTR_PATHWAY
GOBP_PEPTIDYL_SERINE_AUTOPHOSPHORYLATION
GOBP_RNA_SPLICING_VIA_ENDONUCLEOLYTIC_CLEAVAGE_AND_LIGATION
GOBP_INNERVATION
REACTOME_HSF1_ACTIVATION
GOBP_SUBPALLIUM_DEVELOPMENT
GOBP_STRIATUM_DEVELOPMENT
GOBP_DE_NOVO_PROTEIN_FOLDING
GOBP_POSITIVE_REGULATION_OF_ENDOPLASMIC_RETICULUM_UNFOLDED_PROTEIN_RESPONSE
BIOCARTA_GCR_PATHWAY
GOBP_NEGATIVE_REGULATION_OF_EPITHELIAL_CELL_DIFFERENTIATION
GOBP_POSITIVE_REGULATION_OF_STEM_CELL_PROLIFERATION
GOCC_ISWI_TYPE_COMPLEX
GOCC_PHOSPHATIDYLINOSITOL_3_KINASE_COMPLEX_CLASS_I
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS
GOCC_INTRINSIC_COMPONENT_OF_POSTSYNAPTIC_SPECIALIZATION_MEMBRANE
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
GOBP_PROTEIN_TARGETING_TO_MEMBRANE
GOCC_CELL_CELL_CONTACT_ZONE
GOBP_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
GOCC_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOBP_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOCC_CYTOSOLIC_RIBOSOME
REACTOME_SELENOAMINO_ACID_METABOLISM
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
KEGG_RIBOSOME
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE

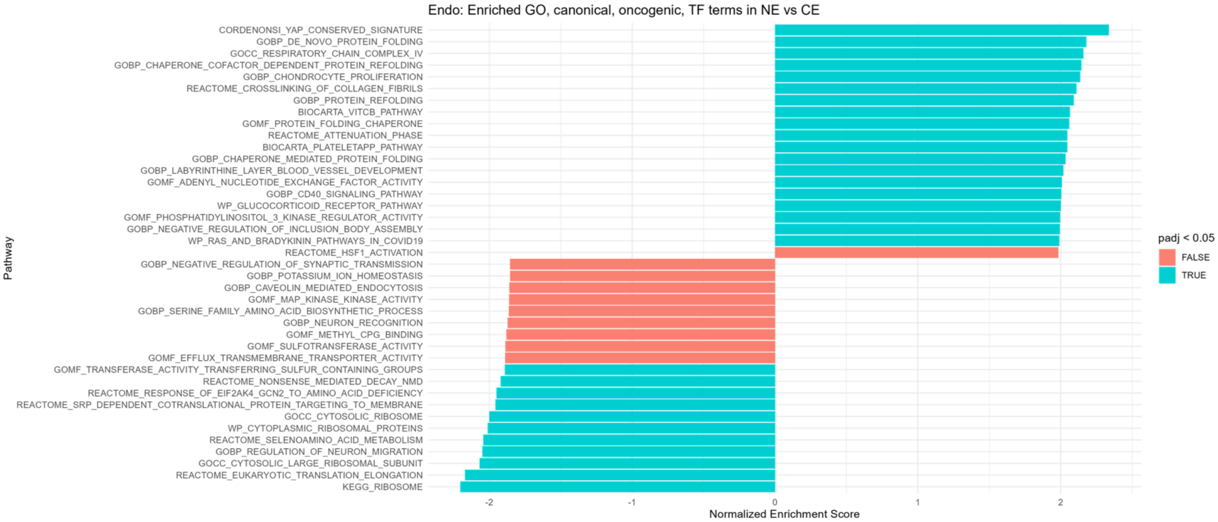
i



Astro: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
REACTOME_HSF1_DEPENDENT_TRANSACTIVATION
DROSHA_TARGET_GENES
BIOCARTA_TCR_PATHWAY
WP_HAIR_FOLLICLE_DEVELOPMENT_CYTODIFFERENTIATION_PART_3_OF_3
WP_PREIMPLANTATION_EMBRYO
WP_TCELL_RECEPTOR_AND_COSTIMULATORY_SIGNALING
PID_TCR_CALCIUM_PATHWAY
GOMF_HEME_COPPER_TERMINAL_OXIDASE_ACTIVITY
WP_PDGF_PATHWAY
GOBP_MITOCHONDRIAL_ELECTRON_TRANSPORT_CYTOCHROME_C_TO_OXYGEN
GOBP_NUCLEOTIDE_BINDING_OLIGOMERIZATION_DOMAIN_CONTAINING_2_SIGNALING_PATHWAY
GOBP_REGULATION_OF_CELLULAR_RESPONSE_TO_HEAT
GOBP_NEGATIVE_REGULATION_OF_DNA_BINDING
WP_ELECTRON_TRANSPORT_CHAIN_OXPHOS_SYSTEM_IN_MITOCHONDRIA
GOBP_NEGATIVE_REGULATION_OF_GLIAL_CELL_DIFFERENTIATION
GOBP_HEART_VALVE_FORMATION
GOCC_CYTOCHROME_COMPLEX
PID_AP1_PATHWAY
PID_CD8_TCR_DOWNSTREAM_PATHWAY
BIOCARTA_BCR_PATHWAY
WP_HYPERTROPHY_MODEL
PID_GLYPICAN_1PATHWAY
GOMF_CHEMOATTRACTANT_ACTIVITY
GOBP_POSITIVE_REGULATION_OF_ENDOCYTOSIS
GOBP_VASCULAR_WOUND_HEALING
GOCC_PLATELET_ALPHA_GRANULE
PID_INTEGRIN1_PATHWAY
WP_KENNEDY_PATHWAY_FROM_SPHINGOLIPIDS
GOBP_CEREBELLAR_CORTEX_FORMATION
CAHOY_ASTROGLIAL
PID_REELIN_PATHWAY
GOBP_CAMP_CATABOLIC_PROCESS
GOMF_CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS
GOCC_PLATELET_ALPHA_GRANULE_LUMEN
GOCC_GUANYL_NUCLEOTIDE_EXCHANGE_FACTOR_COMPLEX
GOBP_CELL_DIFFERENTIATION_IN_HINDBRAIN
GOMF_3_5_CYCLIC_AMP_PHOSPHODIESTERASE_ACTIVITY
ALK_DN.V1_UP
GOBP_POSITIVE_REGULATION_OF_RECEPTOR_MEDIATED_ENDOCYTOSIS
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE

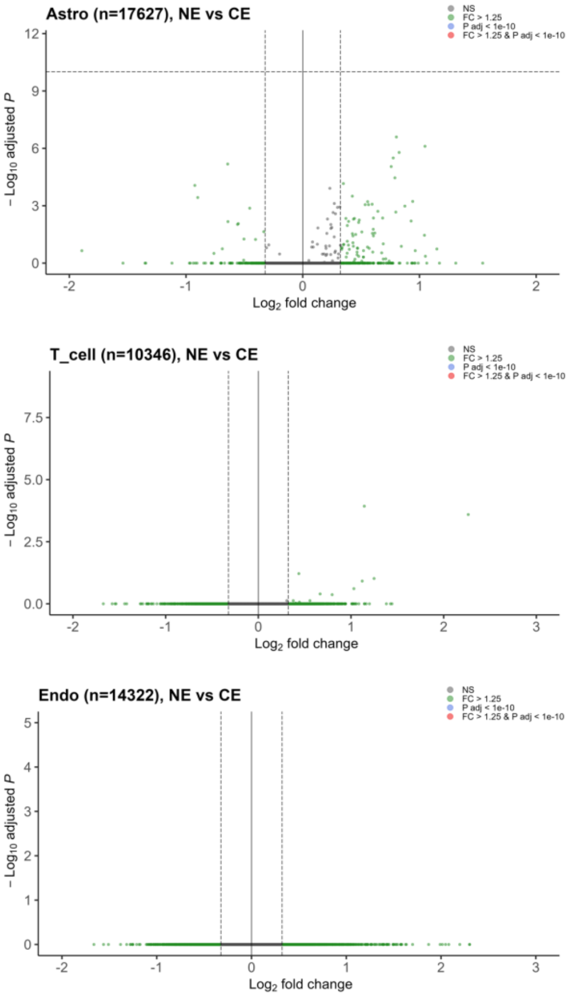


Oligo: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
REACTOME_ATTENUATION_PHASE
GOMF_PROTEIN_FOLDING_CHAPERONE
REACTOME_HSF1_ACTIVATION
REACTOME_HSP90_CHAPERONE_CYCLE_FOR_STEROID_HORMONE_RECEPTORS_SHR
REACTOME_HSF1_DEPENDENT_TRANSACTIVATION
GOCC_JUXTAPARANODE_REGION_OF_AXON
GOBP_NEGATIVE_REGULATION_OF_INCLUSION_BODY_ASSEMBLY
GOBP_VOCALIZATION_BEHAVIOR
GOCC_MYELIN_SHEATH
WP_APOPTOSIS_MODULATION_BY_HSP70
REACTOME_METABOLISM_OF_STEROID_HORMONES
REACTOME_ABC_TRANSPORTERS_IN_LIPID_HOMEOSTASIS
GOBP_CHAPERONE_COFACTOR_DEPENDENT_PROTEIN_REFOLDING
GOMF_STEROID_DEHYDROGENASE_ACTIVITY
GOBP_PROTEIN_LOCALIZATION_TO_AXON
GOBP_REGULATION_OF_EXIT_FROM_MITOSIS
GOBP_REGULATION_OF_SPINDLE_ORGANIZATION
GOBP_REGULATION_OF_SPINDLE_ASSEMBLY
GOBP_NEURON_CELL_CELL_ADHESION
REACTOME_PI_3K_CASCADE_FGFR3
GOCC_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS
REACTOME_RRNA_PROCESSING
GOBP_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM
GOBP_TRANSLATIONAL_INITIATION
GOCC_RIBOSOMAL_SUBUNIT
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOMF_STRUCTURAL_CONSTITUENT_OF_RIBOSOME
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
GOCC_CYTOSOLIC_RIBOSOME
GOBP_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
REACTOME_SELENOAMINO_ACID_METABOLISM
GOBP_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
KEGG_RIBOSOME
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE



Endo: Enriched GO, canonical, oncogenic, TF terms in NE vs CE
CORDENONSI_YAP_CONSERVED_SIGNATURE
GOBP_DE_NOVO_PROTEIN_FOLDING
GOCC_RESPIRATORY_CHAIN_COMPLEX_IV
GOBP_CHAPERONE_COFACTOR_DEPENDENT_PROTEIN_REFOLDING
GOBP_CHONDROCYTE_PROLIFERATION
REACTOME_CROSSLINKING_OF_COLLAGEN_FIBRILS
GOBP_PROTEIN_REFOLDING
BIOCARTA_VITCB_PATHWAY
GOMF_PROTEIN_FOLDING_CHAPERONE
REACTOME_ATTENUATION_PHASE
BIOCARTA_PLATELETAPP_PATHWAY
GOBP_CHAPERONE_MEDIATED_PROTEIN_FOLDING
GOBP_LABYRINTHINE_LAYER_BLOOD_VESSEL_DEVELOPMENT
GOMF_ADENYL_NUCLEOTIDE_EXCHANGE_FACTOR_ACTIVITY
GOBP_CD40_SIGNALING_PATHWAY
WP_GLUCOCORTICOID_RECEPTOR_PATHWAY
GOMF_PHOSPHATIDYLINOSITOL_3_KINASE_REGULATOR_ACTIVITY
GOBP_NEGATIVE_REGULATION_OF_INCLUSION_BODY_ASSEMBLY
WP_RAS_AND_BRADYKININ_PATHWAYS_IN_COVID19
REACTOME_HSF1_ACTIVATION
GOBP_NEGATIVE_REGULATION_OF_SYNAPTIC_TRANSMISSION
GOBP_POTASSIUM_ION_HOMEOSTASIS
GOBP_CAVEOLIN_MEDIATED_ENDOCYTOSIS
GOMF_MAP_KINASE_KINASE_ACTIVITY
GOBP_SERINE_FAMILY_AMINO_ACID_BIOSYNTHETIC_PROCESS
GOBP_NEURON_RECOGNITION
GOMF_METHYL_CPG_BINDING
GOMF_SULFOTRANSFERASE_ACTIVITY
GOMF_EFFLUX_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
GOMF_TRANSFERASE_ACTIVITY_TRANSFERRING_SULFUR_CONTAINING_GROUPS
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
GOCC_CYTOSOLIC_RIBOSOME
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
REACTOME_SELENOAMINO_ACID_METABOLISM
GOBP_REGULATION_OF_NEURON_MIGRATION
GOCC_CYTOSOLIC_LARGE_RIBOSOMAL_SUBUNIT
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
Pathway
-2
-1
0
1
2
Normalized Enrichment Score
padj < 0.05
FALSE
TRUE

Figure 3-6. Non-malignant cell microenvironment in NE vs CE. (a) t-SNE representation of non-malignant cells from both NE and CE regions. **(b)** Phylogenetic tree (left) and dot plot (right) showing marker expression patterns across clusters. **(c)** t-SNE representation of non-malignant cells labeled by cell type based on marker expression profiles. **(d)** Differential abundance analysis of normal cell types in the NE vs CE regions. **(e)** Differential expression analysis within cell types, NE vs CE. Gene set enrichment analysis was performed using each of the differential expression comparisons in **(e)** with a combination of the GO Biological Process, Canonical Pathway, Oncogenic Pathway, and Transcription Factor Target databases [20]. Top positively and negative enriched pathways are shown for all cells **(f)**, neurons **(g)**, immune cells **(h)**, and the remaining cell types **(i)**. *(Astro = astrocytes, Endo = endothelial cells, ExN = excitatory neurons, InN = interneurons, Mac = macrophages, MG = Microglia, Oligo = oligodendrocytes, T_cell = T cells)*



Supplementary Figure 3-6. Related to Fig. 3-6. Remaining cell type-specific differential gene expression results not visualized in **Fig. 3-6**. NE vs CE in astrocytes (top), T cells (middle), endothelial cells (bottom).

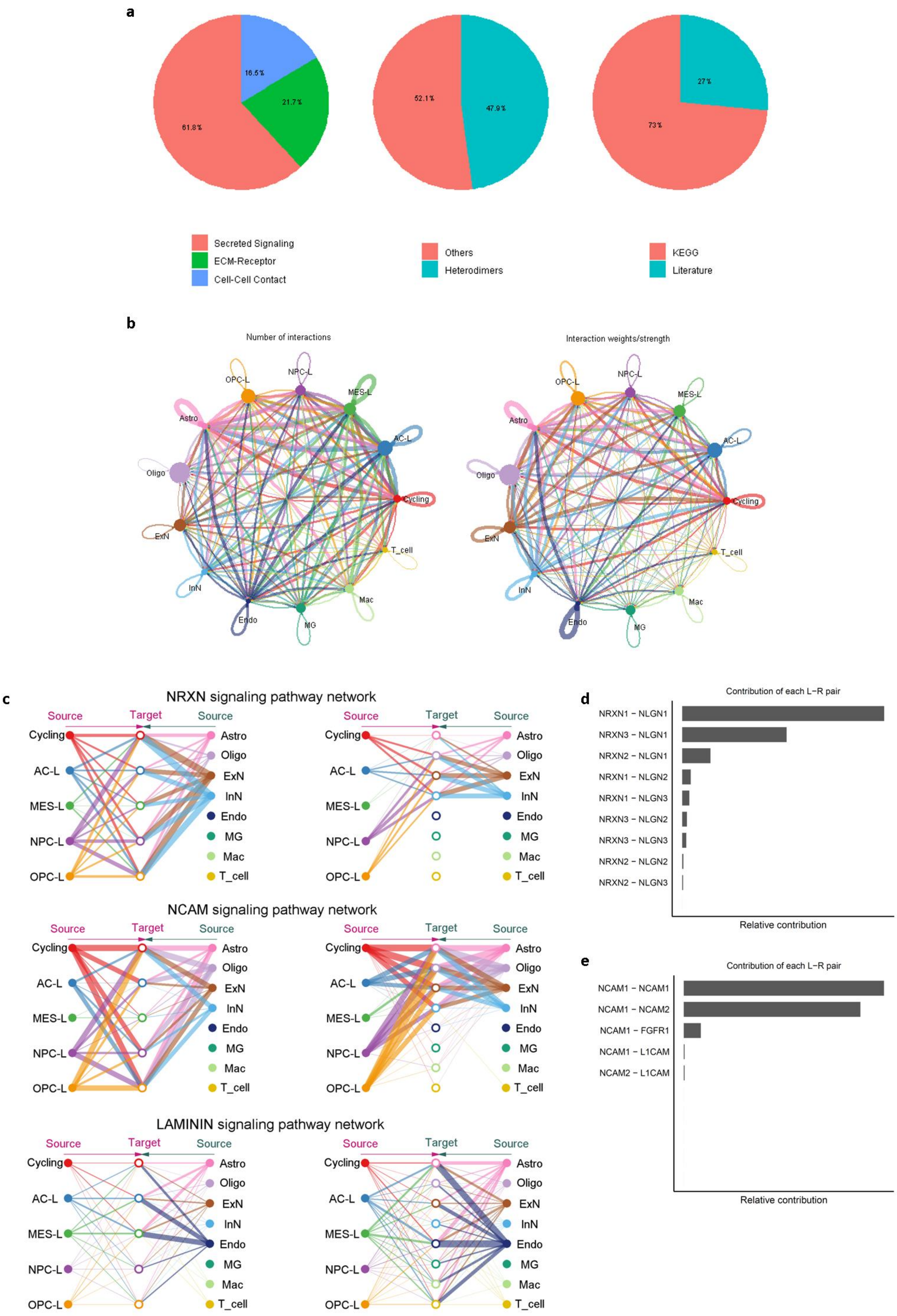
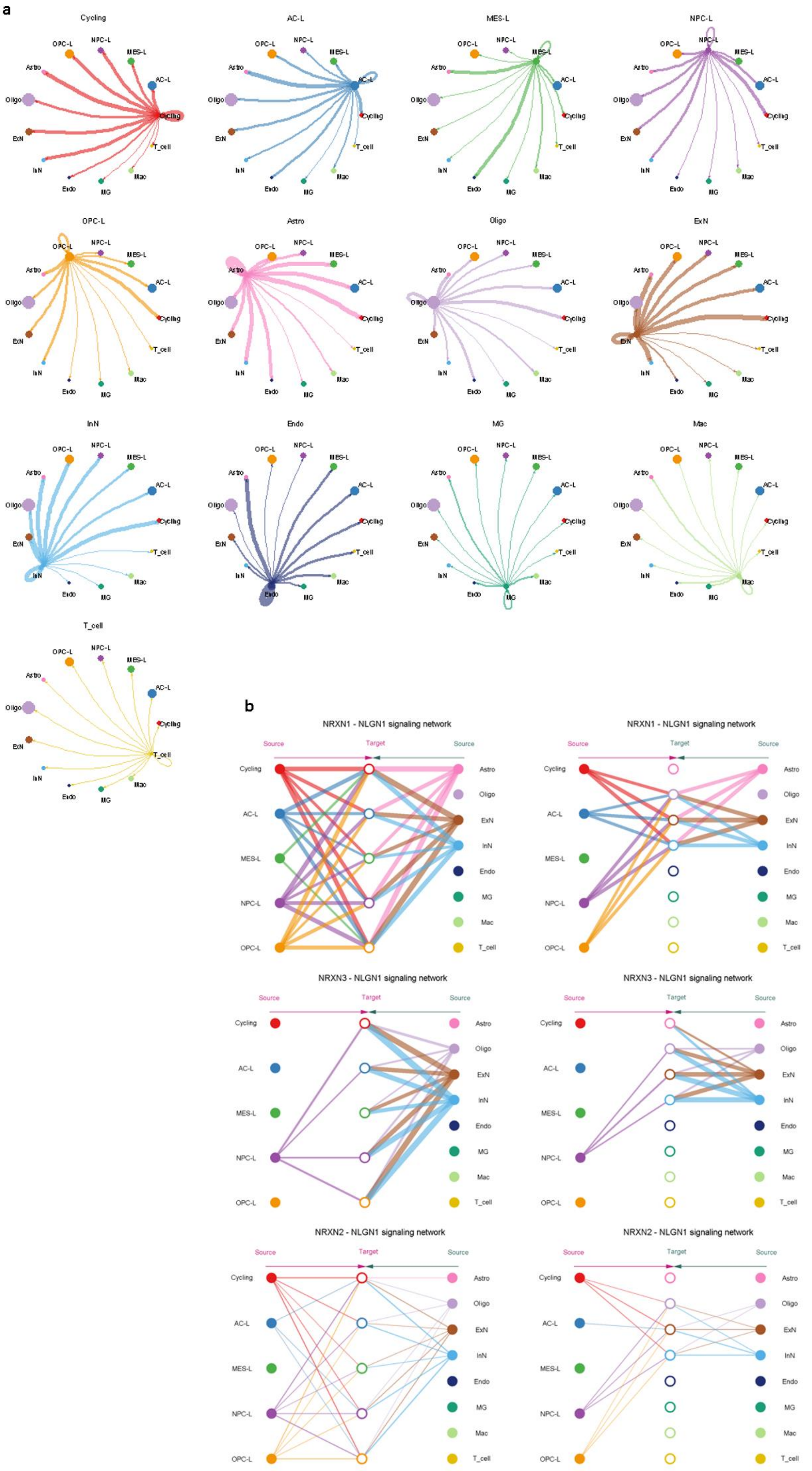
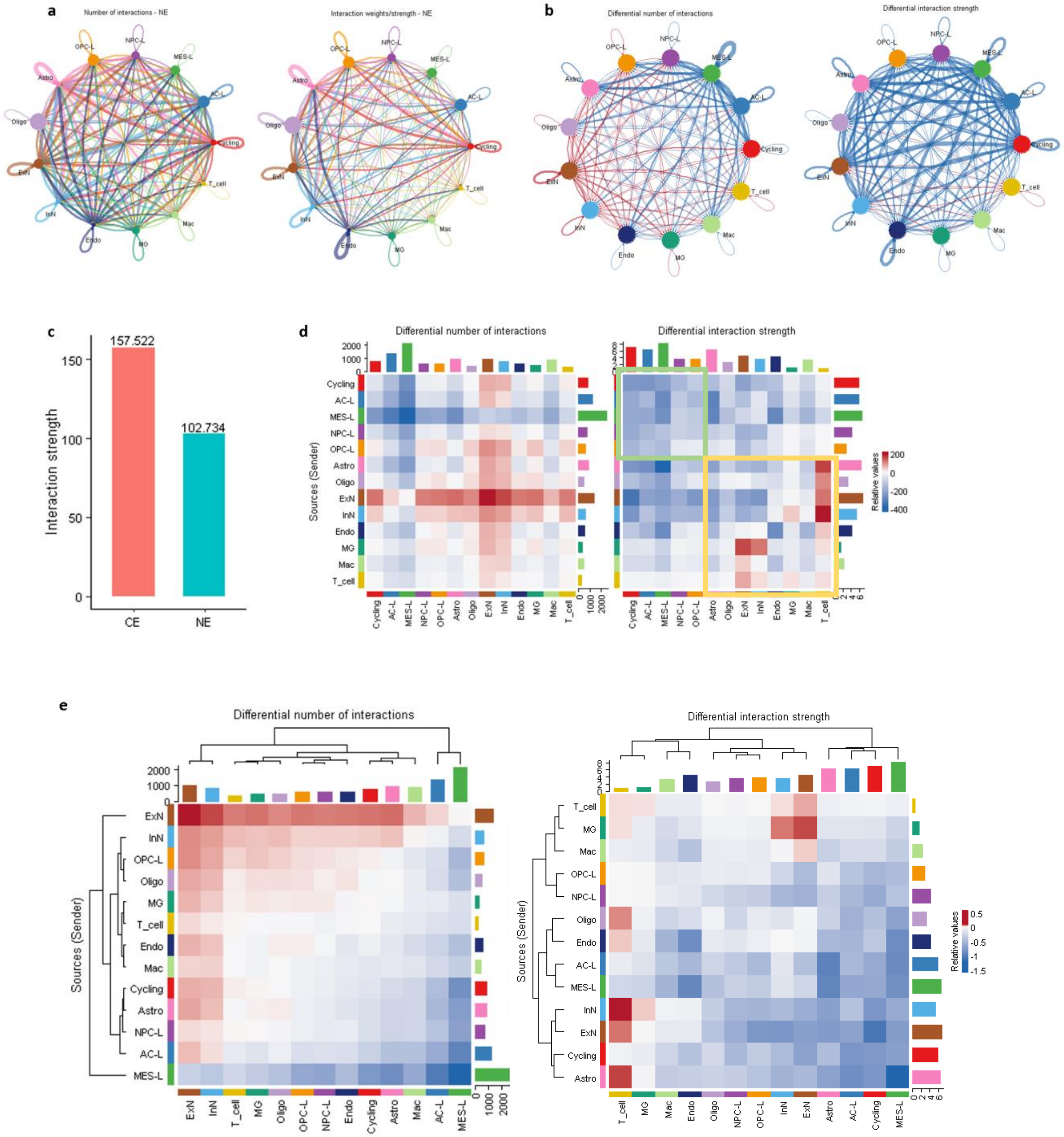

a
16.5%
21.7%
61.8%
52.1%
47.9%
27%
73%
Secreted Signaling
ECM-Receptor
Cell-Cell Contact
Others
Heterodimers
KEGG
Literature
b
Number of interactions
Interaction weights/strength
OPC-L
NPC-L
MES-L
AC-L
Astro
Oligo
Cycling
ExN
T_cell
InN
Mac
Endo
MG
c
NRXN signaling pathway network
NCAM signaling pathway network
LAMININ signaling pathway network
Source
Target
Cycling
AC-L
MES-L
NPC-L
OPC-L
Astro
Oligo
ExN
InN
Endo
MG
Mac
T_cell
d
Contribution of each L−R pair
NRXN1 − NLGN1
NRXN3 − NLGN1
NRXN2 − NLGN1
NRXN1 − NLGN2
NRXN1 − NLGN3
NRXN3 − NLGN2
NRXN3 − NLGN3
NRXN2 − NLGN2
NRXN2 − NLGN3
Relative contribution
e
Contribution of each L−R pair
NCAM1 − NCAM1
NCAM1 − NCAM2
NCAM1 − FGFR1
NCAM1 − L1CAM
NCAM2 − L1CAM
Relative contribution

Figure 3-7. Integrated atlas of malignant-normal cell communication networks in GBM.
(a) Summary of human CellChat interaction database used for communication analysis. **(b)** Total interaction number (left) and strength (right) in tumor core as identified by CellChat. **(c)** Hierarchy plots showing the signaling participants for the top three significant interaction pathways detected. Ligand-receptor pairs were ranked by contribution to entire signaling pathway. Shown are interaction pairs corresponding to the top 2 significant pathways in the CE region, NRXN **(d)** and NCAM **(e)**.

a
Cycling
AC-L
MES-L
NPC-L
OPC-L
Astro
Oligo
ExN
InN
Endo
MG
Mac
T_cell
b
NRXN1 - NLGN1 signaling network
NRXN3 - NLGN1 signaling network
NRXN2 - NLGN1 signaling network
Source
Target
Source
Cycling
AC-L
MES-L
NPC-L
OPC-L
Astro
Oligo
ExN
InN
Endo
MG
Mac
T_cell

Supplementary Figure 7. Related to Fig. 3-7.

(a) Strength/weight of signaling sent from each cell population, identifying strong and weak interaction partners. Line thickness corresponds to weight. **(b)** Hierarchy plot for each of the top three ligand-receptor interaction pairs contributing to the NRXN pathway. Circles in the middle column are signaling receivers, while those in the outer columns are senders. Malignant populations are on the left, and non-malignant populations are on the right. For each interaction, malignant cells are the receiver cells in the first plot, and non-malignant cells are the receiver cells in the second plot.



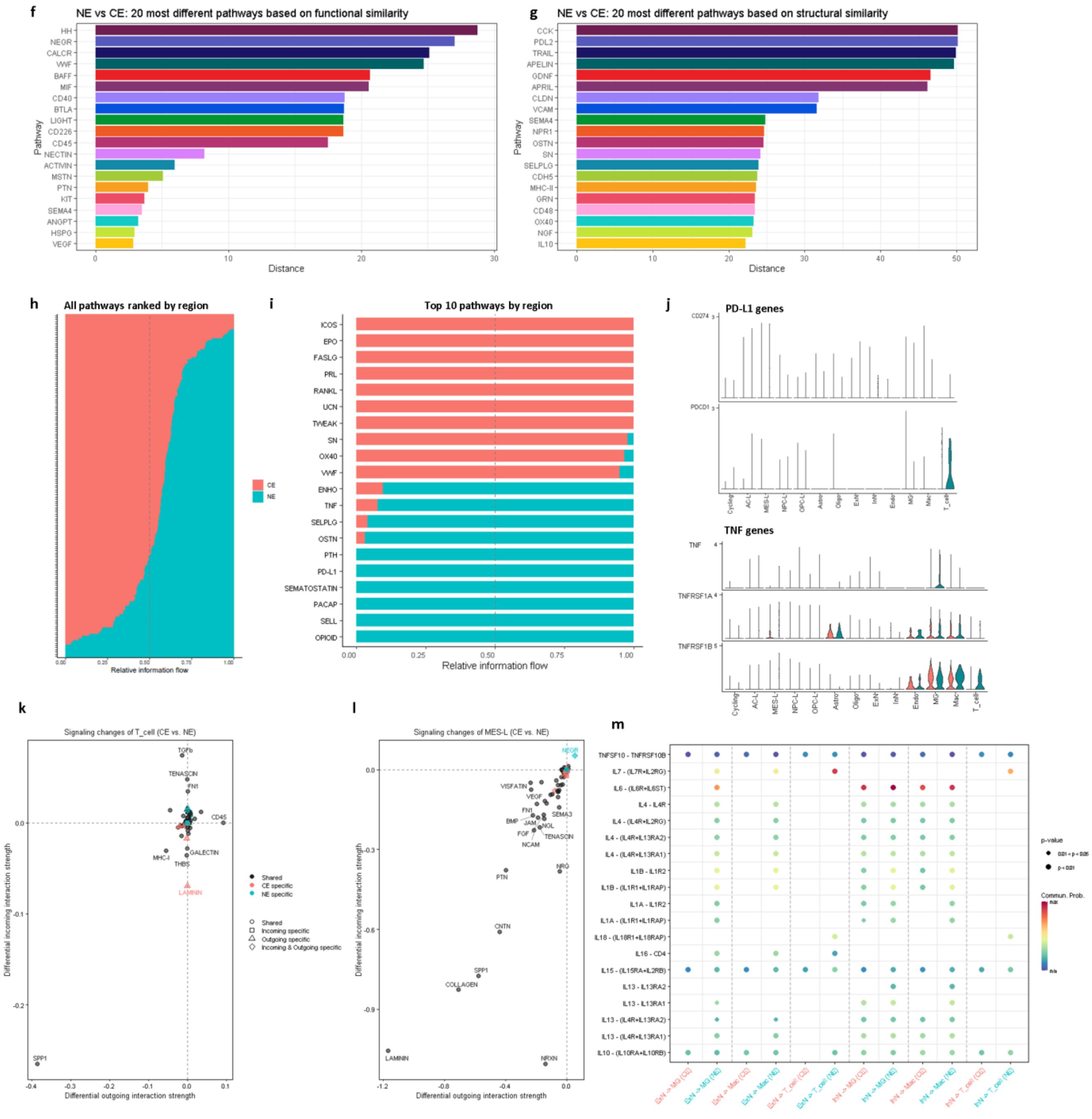
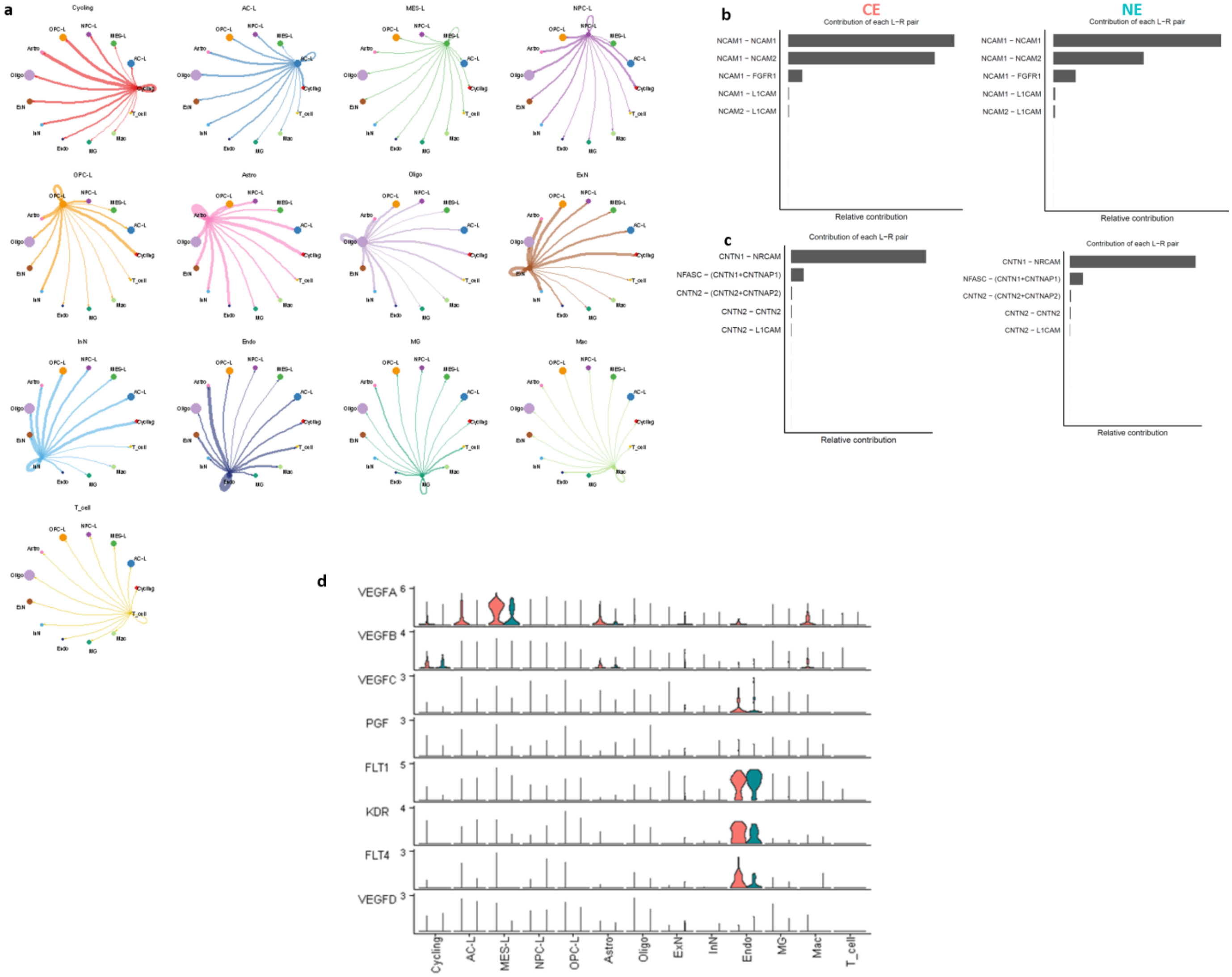
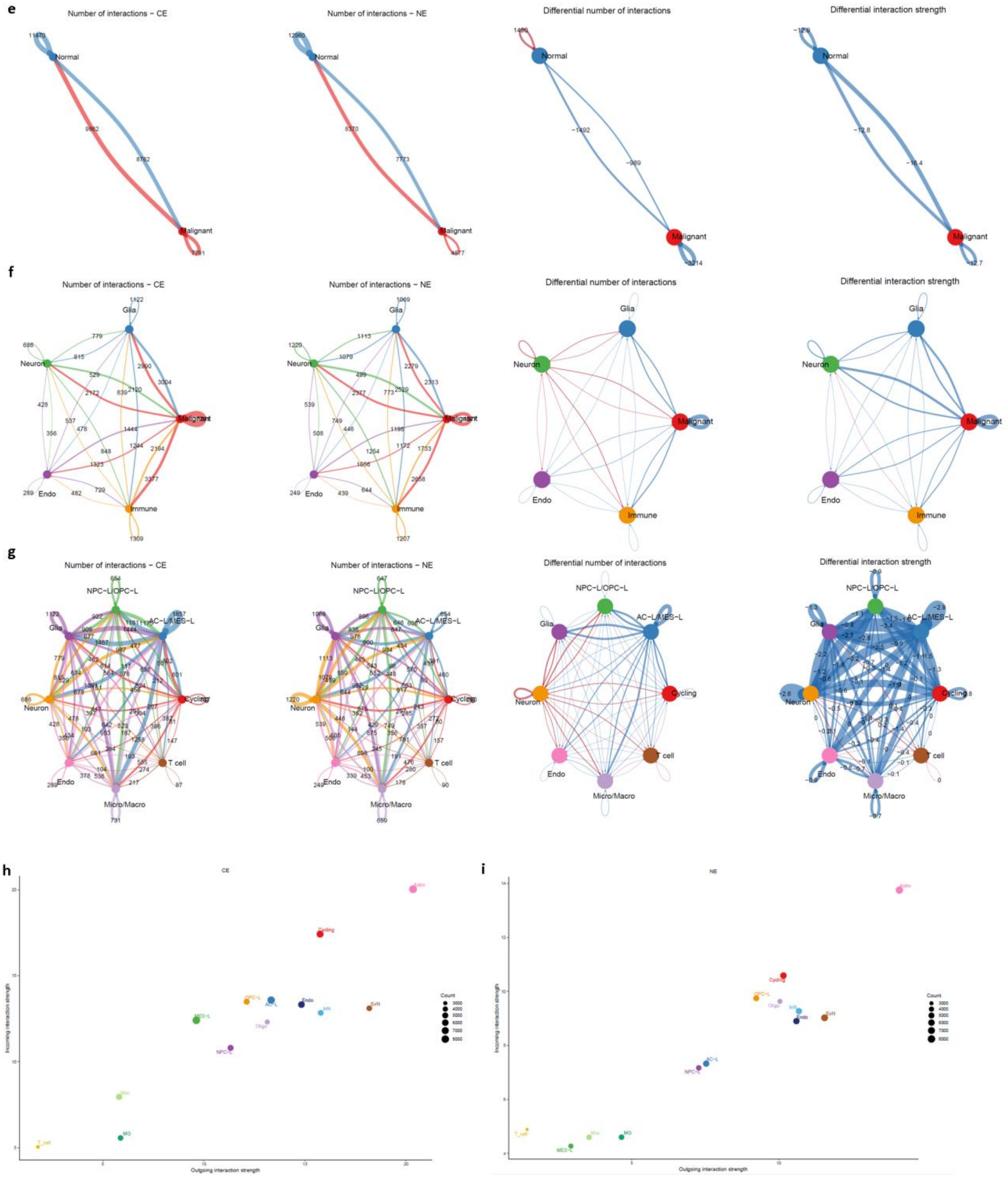


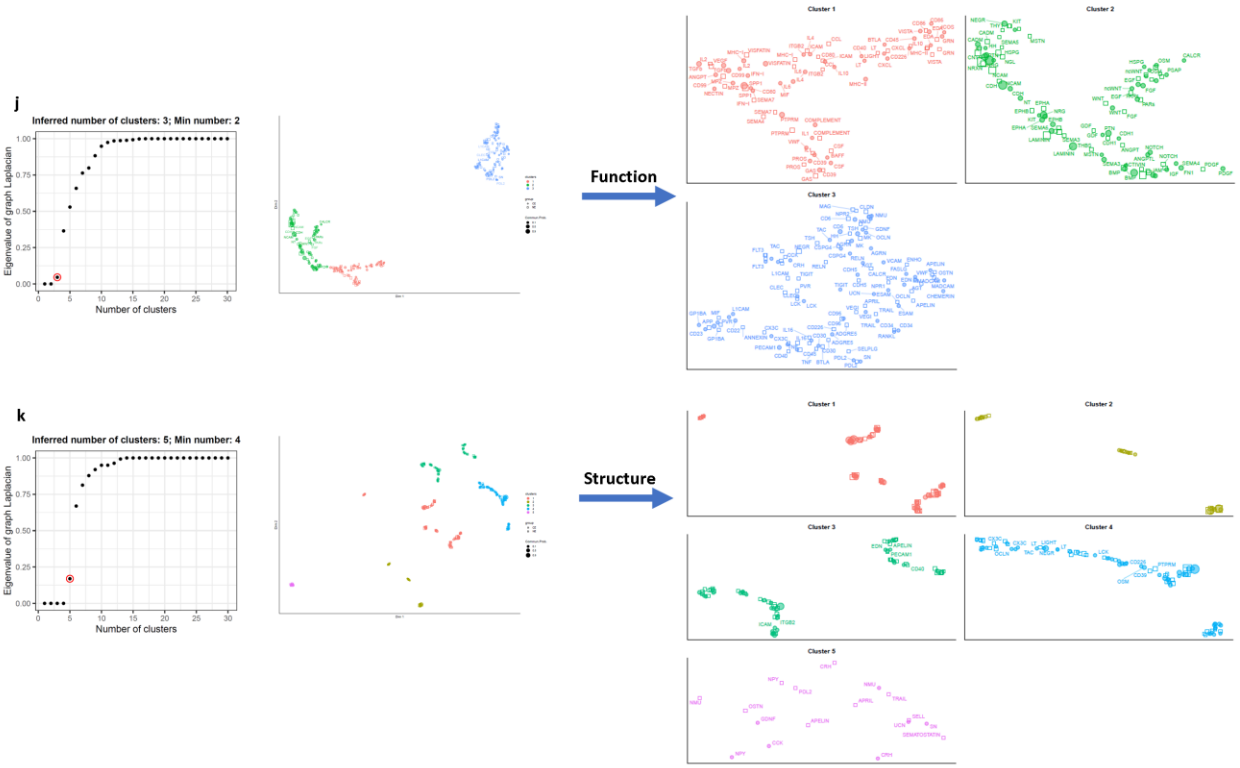
Figure 3-8. Rewiring of cell-cell communication network structure in NE region
(a) Total interaction number (left) and strength (right) in NE region as identified by CellChat. **(b)** Differential analysis of interaction number (left) and strength (right) between NE and CE regions. Red indicates increased communication in NE, while blue indicates decreased communication in NE. **(c)** Quantification of interaction strength by region. **(d)** Heatmap showing differential number (left) and strength (right) of pair-wise interactions across cell populations. Rows are senders of communication; columns are receivers. Color bars indicate total amount of signaling in the respective row or column. **(e)** Same as **(d)** but both rows and columns are clustered based on patterns of differential communication. Joint manifold learning and classification of networks was performed based on functional and structural

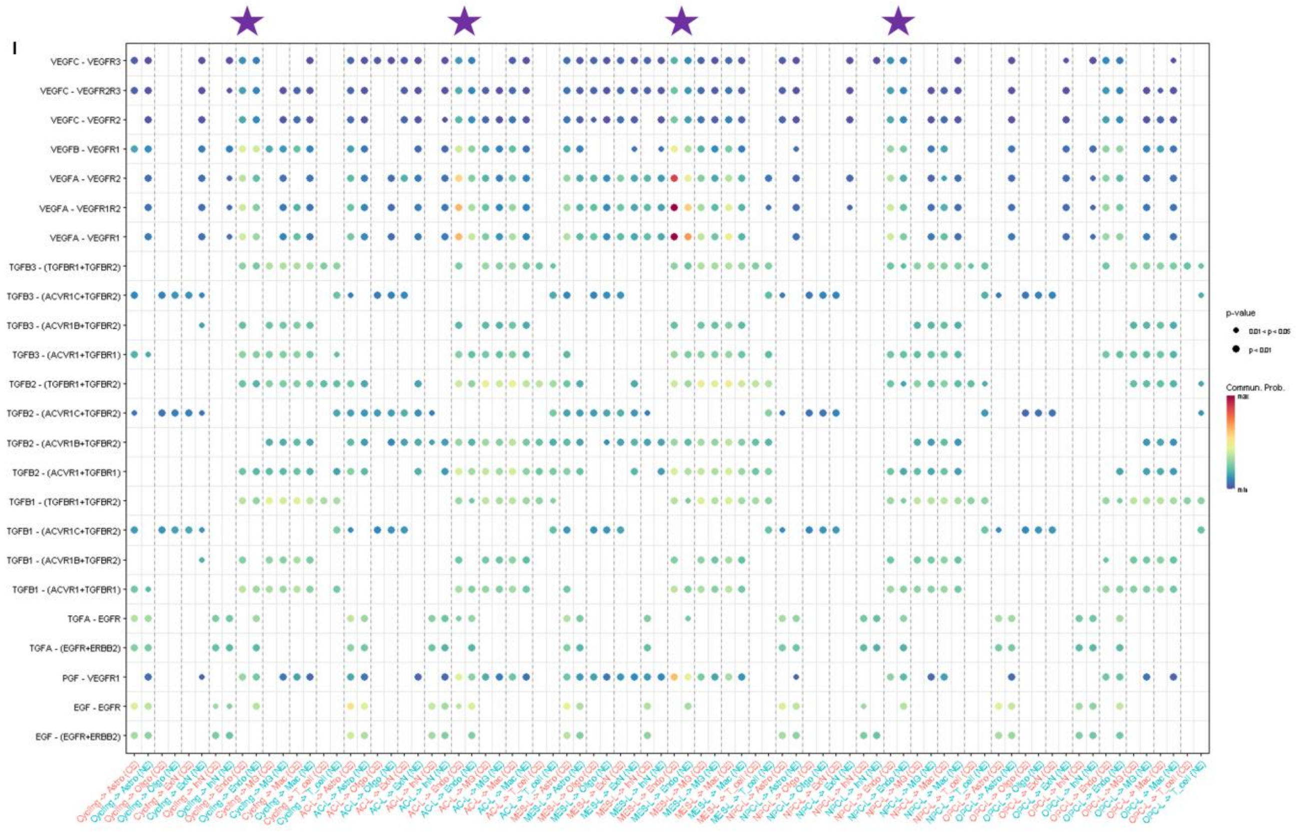
similarities, yielding two dimensional plots with 3 and 5 clusters respectively (see **Supp. Fig. 3-8j-k**). For each pathway, Euclidean distance between the NE and CE datasets was computed using both functional **(f)** and structural **(g)** similarity to examine communication differences. Pathways were also ranked by relative information flow across regions; all pathways **(h)** and the top 10 for each region **(i)**. Pathways were labeled by region using Paired Wilcoxon to test for significant difference between datasets. **(j)** Gene expression across cell types for proteins belonging to two NE-enriched pathways. **(k, l)** Quantification of differential incoming and outgoing interaction strength in NE vs CE, highlighting specific pathways in T cells **(k)** and MES-like malignant cells **(l)**. **(m)** Bubble plot showing region-specific interaction strength and significance with neurons as sources and immune cells as receivers. Interleukin-related pathways are visualized.





e
Number of interactions − CE
11470
Normal
9862
8762
Malignant
7781
Number of interactions − NE
12960
Normal
8370
7773
Malignant
4977
Differential number of interactions
1490
Normal
−1492
−989
Malignant
−3214
Differential interaction strength
−12.9
Normal
−12.8
−16.4
Malignant
−12.7
f
Number of interactions − CE
1122
Glia
686
Neuron
Malignant
289
Endo
Immune
1309
Number of interactions − NE
1069
Glia
1220
Neuron
Malignant
249
Endo
Immune
1207
Differential number of interactions
Glia
Neuron
Malignant
Endo
Immune
Differential interaction strength
Glia
Neuron
Malignant
Endo
Immune
g
Number of interactions − CE
654
NPC−L/OPC−L
AC−L/MES−L
Glia
Neuron
Cycling
T cell
Endo
Micro/Macro
781
Number of interactions − NE
647
NPC−L/OPC−L
AC−L/MES−L
Glia
Neuron
Cycling
T cell
Endo
Micro/Macro
669
Differential number of interactions
NPC−L/OPC−L
AC−L/MES−L
Glia
Cycling
Neuron
T cell
Endo
Micro/Macro
Differential interaction strength
NPC−L/OPC−L
AC−L/MES−L
Glia
Neuron
Cycling
T cell
Endo
Micro/Macro
h
CE
Incoming interaction strength
Outgoing interaction strength
Endo
Cycling
OPC−L
AC−L
Endo
InN
ExN
MES−L
Oligo
NPC−L
Mac
MG
T_cell
Count
3000
4000
5000
6000
7000
8000
i
NE
Incoming interaction strength
Outgoing interaction strength
Endo
Cycling
OPC−L
Oligo
InN
Endo
ExN
AC−L
NPC−L
T_cell
MES−L
Mac
MG
Count
3000
4000
5000
6000
7000
8000


j
Inferred number of clusters: 3; Min number: 2
Eigenvalue of graph Laplacian
Number of clusters
Function
Cluster 1
Cluster 2
Cluster 3
k
Inferred number of clusters: 5; Min number: 4
Eigenvalue of graph Laplacian
Number of clusters
Structure
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5


l
VEGFC - VEGFR3
VEGFC - VEGFR2R3
VEGFC - VEGFR2
VEGFB - VEGFR1
VEGFA - VEGFR2
VEGFA - VEGFR1R2
VEGFA - VEGFR1
TGFB3 - (TGFBR1+TGFBR2)
TGFB3 - (ACVR1C+TGFBR2)
TGFB3 - (ACVR1B+TGFBR2)
TGFB3 - (ACVR1+TGFBR1)
TGFB2 - (TGFBR1+TGFBR2)
TGFB2 - (ACVR1C+TGFBR2)
TGFB2 - (ACVR1B+TGFBR2)
TGFB2 - (ACVR1+TGFBR1)
TGFB1 - (TGFBR1+TGFBR2)
TGFB1 - (ACVR1C+TGFBR2)
TGFB1 - (ACVR1B+TGFBR2)
TGFB1 - (ACVR1+TGFBR1)
TGFA - EGFR
TGFA - (EGFR+ERBB2)
PGF - VEGFR1
EGF - EGFR
EGF - (EGFR+ERBB2)
Cycling -> Astro (CE)
Cycling -> Astro (NE)
Cycling -> Oligo (CE)
Cycling -> Oligo (NE)
Cycling -> ExN (CE)
Cycling -> ExN (NE)
Cycling -> InN (CE)
Cycling -> InN (NE)
Cycling -> Endo (CE)
Cycling -> Endo (NE)
Cycling -> MG (CE)
Cycling -> MG (NE)
Cycling -> Mac (CE)
Cycling -> Mac (NE)
Cycling -> T_cell (CE)
Cycling -> T_cell (NE)
AC-L -> Astro (CE)
AC-L -> Astro (NE)
AC-L -> Oligo (CE)
AC-L -> Oligo (NE)
AC-L -> ExN (CE)
AC-L -> ExN (NE)
AC-L -> InN (CE)
AC-L -> InN (NE)
AC-L -> Endo (CE)
AC-L -> Endo (NE)
AC-L -> MG (CE)
AC-L -> MG (NE)
AC-L -> Mac (CE)
AC-L -> Mac (NE)
AC-L -> T_cell (CE)
AC-L -> T_cell (NE)
MES-L -> Astro (CE)
MES-L -> Astro (NE)
MES-L -> Oligo (CE)
MES-L -> Oligo (NE)
MES-L -> ExN (CE)
MES-L -> ExN (NE)
MES-L -> InN (CE)
MES-L -> InN (NE)
MES-L -> Endo (CE)
MES-L -> Endo (NE)
MES-L -> MG (CE)
MES-L -> MG (NE)
MES-L -> Mac (CE)
MES-L -> Mac (NE)
MES-L -> T_cell (CE)
MES-L -> T_cell (NE)
NPC-L -> Astro (CE)
NPC-L -> Astro (NE)
NPC-L -> Oligo (CE)
NPC-L -> Oligo (NE)
NPC-L -> ExN (CE)
NPC-L -> ExN (NE)
NPC-L -> InN (CE)
NPC-L -> InN (NE)
NPC-L -> Endo (CE)
NPC-L -> Endo (NE)
NPC-L -> MG (CE)
NPC-L -> MG (NE)
NPC-L -> Mac (CE)
NPC-L -> Mac (NE)
NPC-L -> T_cell (CE)
NPC-L -> T_cell (NE)
OPC-L -> Astro (CE)
OPC-L -> Astro (NE)
OPC-L -> Oligo (CE)
OPC-L -> Oligo (NE)
OPC-L -> ExN (CE)
OPC-L -> ExN (NE)
OPC-L -> InN (CE)
OPC-L -> InN (NE)
OPC-L -> Endo (CE)
OPC-L -> Endo (NE)
OPC-L -> MG (CE)
OPC-L -> MG (NE)
OPC-L -> Mac (CE)
OPC-L -> Mac (NE)
OPC-L -> T_cell (CE)
OPC-L -> T_cell (NE)
p-value
0.01 < p < 0.05
p < 0.01
Commun. Prob.
max
min

Supplementary Figure 3-8. Related to Fig. 3-8.
(a) Within NE region: strength/weight of signaling sent from each cell population. Line thickness corresponds to weight. **(b, c)** Highlighting two of the top significant pathways identified, NCAM **(b)** and CNTN **(c)**, whose specific interaction pair contributions appear different and similar, respectively, in NE vs CE. **(d)** VEGF pathway: region-specific gene expression profiles across cell types. **(e-g)** Similar to **Fig. 3-8a-b** but using different cell grouping resolutions: **(e)** 2 groups - malignant and normal; **(f)** 5 groups - malignant, neurons, glia, endothelial cells, and immune cells; and **(g)** 8 groups - cycling malignant cells, NPC/OPC-like malignant cells, AC/MES-like malignant cells, neurons, glia, endothelial cells, microglia/macrophages, and T cells. Comparison of source and target tendencies of individual cell types in CE **(h)** and NE **(i)** regions. **(j, k)** NE vs CE differential analysis: joint manifold learning and classification based on functional **(j)** and structural **(k)** similarity, yielding 3 and 5 clusters, respectively. **(l)** Bubble plot showing region-specific interaction strength and significance with malignant cells as sources and non-malignant cells as receivers. Pathways visualized are PD-L1, VEGF, TGFb, TNF, and EGF. Purple stars indicate columns corresponding to malignant cell-to-endothelial cell signaling.

References

1. Brown, T. J. *et al.* Association of the Extent of Resection With Survival in Glioblastoma. *JAMA Oncol* **2**, 1460–1469 (2016).

2. Wen, P. Y. & Kesari, S. Malignant gliomas in adults. *N Engl J Med* **359**, 492–507 (2008).

3. Sanai, N., Polley, M.-Y., McDermott, M. W., Parsa, A. T. & Berger, M. S. An extent of resection threshold for newly diagnosed glioblastomas. *J Neurosurg* **115**, 3–8 (2011).

4. Bette, S. *et al.* Retrospective Analysis of Radiological Recurrence Patterns in Glioblastoma, Their Prognostic Value And Association to Postoperative Infarct Volume. *Sci Rep* **8**, 4561 (2018).

5. Paw, I., Carpenter, R. C., Watabe, K., Debinski, W. & Lo, H.-W. Mechanisms regulating glioma invasion. *Cancer Lett* **362**, 1–7 (2015).

6. Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756–760 (2006).

7. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth following chemotherapy. *Nature* **488**, 522–526 (2012).

8. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).

9. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* (2017).

10. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

11. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).

12. Weng, Q. *et al.* Single-Cell Transcriptomics Uncovers Glial Progenitor Diversity and Cell Fate Determinants during Development and Gliomagenesis. *Cell Stem Cell* **24**, 707-723.e8 (2019).

13. Wang, L. *et al.* The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. *Cancer Discov* **9**, 1708–1719 (2019).

14. Lan, X. *et al.* Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature* **549**, 227–232 (2017).

15. inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/inferCNV

16. Suvà, M. L. & Tirosh, I. The Glioma Stem Cell Model in the Era of Single-Cell Genomics. *Cancer Cell* **37**, 630–636 (2020).

17. Squair, J. W., Skinnider, M. A., Gautier, M., Foster, L. J. & Courtine, G. Prioritization of cell types responsive to biological perturbations in single-cell data with Augur. *Nat Protoc* **16**, 3836–3873 (2021).

18. Schaub, C. *et al.* Tumor growth patterns of MGMT-non-methylated glioblastoma in the randomized GLARIUS trial. *J Cancer Res Clin Oncol* **144**, 1581–1589 (2018).

19. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021) doi:10.1101/060012.

20. Igor Dolgalev (2021). msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version 7.4.1. https://CRAN.R-project.org/package=msigdbr

21. Liu, Y. *et al.* STK33 participates to HSP90-supported angiogenic program in hypoxic tumors by regulating HIF-1α/VEGF signaling pathway. *Oncotarget* **8**, 77474–77488 (2017).

22. Rooj, AK., Bronisz, A. & Godlewski, J. The role of Octamer binding transcription factors in glioblastoma multiforme. *Biochim Biophys Acta* **1859**, 805–811 (2016).

23. Zhang, C., Yang, M., Li, Y., Tang, S. & Sun, X. FOXA1 is upregulated in glioma and promotes proliferation as well as cell cycle through regulation of cyclin D1 expression. *Cancer Manag Res* **10**, 3283–3293 (2018).

24. Zhao, Y. *et al.* Myocyte enhancer factor 2D promotes tumorigenicity in malignant glioma cells. *Tumour Biol* **37**, 601–610 (2016).

25. Guo, S.-W., Zheng, Y., Lu, Y., Liu, X. & Geng, J.-G. Slit2 Overexpression Results in Increased Microvessel Density and Lesion Size in Mice With Induced Endometriosis. *Reprod Sci* **20**, 285–298 (2013).

26. Müller, S. *et al.* Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biology* **18**, 234 (2017).

27. Treiber, J. M. *et al.* Molecular physiology of contrast enhancement in glioblastomas: An analysis of The Cancer Imaging Archive (TCIA). *J Clin Neurosci* **55**, 86–92 (2018).

28. Caspani, E. M., Crossley, P. H., Redondo-Garcia, C. & Martinez, S. Glioblastoma: a pathogenic crosstalk between tumor cells and pericytes. *PLoS One* **9**, e101402 (2014).

29. Venkatesh, H. S. *et al.* Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion. *Cell* **161**, 803–816 (2015).

30. Venkatesh, H. S. *et al.* Targeting neuronal activity-regulated neuroligin-3 dependency in high-grade glioma. *Nature* **549**, 533–537 (2017).

31. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).

32. Darmanis, S. *et al.* Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep* **21**, 1399–1410 (2017).

33. Gill, B. J. *et al.* MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. *Proc Natl Acad Sci U S A* **111**, 12550–12555 (2014).

34. Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat Commun* **11**, 3406 (2020).

35. Fedele, M., Cerchia, L., Pegoraro, S., Sgarra, R. & Manfioletti, G. Proneural-Mesenchymal Transition: Phenotypic Plasticity to Acquire Multitherapy Resistance in Glioblastoma. *Int J Mol Sci* **20**, 2746 (2019).

36. Liang, Q. *et al.* Profiling pro-neural to mesenchymal transition identifies a lncRNA signature in glioma. *Journal of Translational Medicine* **18**, 378 (2020).

37. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**, 162–173 (1996).

38. Stupp, R. *et al.* Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *New England Journal of Medicine* **352**, 987–996 (2005).

39. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).

40. Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* **11**, 499–524 (2016).

41. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

42. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

43. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

44. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296 (2019).

45. Leonard, S., Rolland, A., Tarte, K., Chalmel, F. & Lardenois, A. *FlexDotPlot: a universal and modular dot plot visualization tool for complex multifaceted data*. http://biorxiv.org/lookup/doi/10.1101/2020.04.03.023655 (2020) doi:10.1101/2020.04.03.023655.

46. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

47. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184–1191 (2009).

48. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Research* gkz966 (2019) doi:10.1093/nar/gkz966.

49. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).

50. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

51. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).

52. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90 (2021).

53. Blighe, K, S Rana, and M Lewis. 2018. “EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.” https://github.com/kevinblighe/EnhancedVolcano

54. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Chapter 4: Conclusions

The overall aim of these studies is to establish a flexible and scalable workflow that we and others can leverage to dissect complex cellular and molecular relationships contributing to CNS disorders. Prior to the advent of single cell analytics, and even now to a lesser extent, untangling these complex relationships in a robust manner was extremely challenging. Recent progress in the development of single cell "-omics" technologies has enabled enormous growth in our understanding of many biomedical research contexts, including CNS disorders. Despite these advances, many questions remain unanswered, and progress in clinical management of severe diseases has been limited. This is in part due to technical limitations, as many components of bioinformatics workflows require a significant amount of manual manipulation, arbitrary decision-making, and inconsistency across experts regarding "best practice" guidelines. We argue that converging on unbiased, reproducible, creative, and efficient analytical workflows can significantly improve our ability to identify important and clinically relevant biological insights with high confidence. In these studies, we apply these principles to several open questions in glioma research and show that transcriptomic analysis with a focus on pathway activation and relationships among cells elucidates key drivers of multiple hallmark glioma features, most notably proliferation, vascular remodeling, and invasive potential. These findings present opportunities for confirming several novel pathways, cellular populations, and cell-cell interactions that are potential candidates for improving glioma therapeutics.

In Chapter 2, we focused on transcriptomic analysis at the pathway level and developed an alternative molecular classification scheme for glioblastoma tumors that aims to improve on the original TCGA categories [1]. We reasoned that analyzing gene expression alone was insufficient in terms of telling the full story about biological dysregulation in GBMs. To combat this, our method used gene set enrichment analysis (GSEA) and analyzed gene expression profiles based on differential enrichment of functional pathways, which yielded a stable clustering solution when considering a large number of bulk gene expression profiles generated from patient samples and patient-derived gliomasphere lines. We then probed these patterns of differential functional enrichment and identified gene networks driving the observed cluster divergence. Examining relevant pathways and upstream regulators of these gene networks pointed to E2F1 as a key driver of tumor cell proliferation. Importantly, despite widespread proliferative potential across the gliomasphere lines, E2F1 was only a driver of proliferation and tumor formation in the clusters that were predicted to be E2F1-activated. Moreover, E2F1-activated cells were sensitized to radiation by E2F1 inhibition. Finally, we discovered functional convergence and biological parallels in our gene networks by performing both bulk and single cell correlation analyses, identifying positive and negative connections within our gene networks and between our networks and other previously described function- and cell type-specific gene sets. This revealed a strong connection between one of our non-cell cycle networks and markers of MES-like tumor cells, adult endothelial cells, developing endothelial cells, and developing mural cells, suggesting vasculature-related dysregulation in tumors enriched for this gene network. These findings reflect our method's powerful ability to untangle heterogeneous tumor pathway utilization and validate therapeutic targets in a highly specific manner, as well as its adaptability to single cell RNA-seq analyses that are becoming more readily feasible.

In Chapter 3, we began to address a major challenge in glioma clinical management, namely the short survival period due to high likelihood of tumor recurrence. After initial glioma resection surgery, there are likely malignant cells left behind near the tumor periphery that later drive tumor recurrence despite adjuvant therapy. The peripheral region of the tumor has not been well-characterized at single cell resolution, as most single cell glioma studies either focus on the core of the tumor or examine a low number of cells from the periphery. We thus sought to comprehensively characterize the cellular and molecular landscape of the peripheral, non-enhancing (NE) region compared to the contrast-enhancing (CE) region of a cohort of recurrent glioma patients using MRI-guided biopsy and single nucleus RNA-sequencing. We first established an improved method for confidently identifying malignant cells in tumor sc/snRNA-seq datasets and used this tool to separate the malignant from non-malignant cells in our dataset. We then examined malignant cell composition and molecular characteristics using a previously described classification scheme [2], noting the presence of all malignant states in recurrent glioma, tumor-specific composition profiles, conserved proliferation ability across states, and functional enrichment that parallels the normal cell correlates of each cell state. Next, we identified copy number-based subclusters in each tumor and developed a flexible tool to visualize the subcluster hierarchies superimposed with other levels of information. We used this tool to inspect the identified subclusters and their malignant cell state compositions, finding that each tumor consisted of 8 terminally evolved subclusters, none of which were up/downstream from each other, and that many subclusters contained a dominant malignant cell state, suggesting a potential clonal stability with respect to cell states in recurrent glioma that was not observed in primary glioma [2]. We then focused on the NE region and observed a significant tumor burden in all patients, with patient-specific alterations in cell state composition, cycling fraction, and copy number subcluster composition. We noted here that NPC-like and, interestingly, MES-like cells were more proliferative in NE, potentially implicating these two states in recurrence-related proliferation. Comparing the molecular profiles of each cell state between the two regions revealed upregulation of several pro-proliferative terms and downregulation of vascular remodeling terms in MES-like cells in the NE region, suggesting specific pathways driving the increased MES-like cycling fraction and potentially identifying a functional transition from influencing vasculature in the tumor core to exhibiting proliferative capacity at the periphery. Analysis of non-malignant cells identified a decrease in the proportion of T cells and macrophages, suggesting a potential immune-cold environment near the periphery. Moreover, normal endothelial cells were enriched in vascular development terms in NE, implying that they may be taking back control of vasculature from MES-like malignant cells. Finally, cell-cell interaction analysis comparing the two regions revealed in the NE region a significant decrease in malignant cell communication, increase in normal-to-normal cell communication, rewiring of communication networks, and specific depletion of VEGFA-VEGFR2 signaling from MES-like tumor cells to normal endothelial cells, further implicating MES-like cells as significantly functionally altered. These findings present the first thorough characterization of single cells in the NE region in glioma with respect to functional transitions, tumor microenvironment interactions, and copy number-based hierarchical relationships.

Several of our findings are consistent with previous studies focusing on regional diversity in glioma. Similar to reports [3, 4] that macrophages and microglia play a stronger role in the tumor core and periphery, respectively, we observe a decrease in the macrophage:microglia ratio and a relative decrease in macrophage cellular communication compared to microglial communication, both in the outgoing and incoming directions, in the edge region.

Moreover, the observed functional transition in MES-like malignant cells (from being pro-angiogenic in the tumor core and proliferative in the periphery) parallels the reported 1) pattern of increased proliferation and decreased angiogenic factors [4] and 2) shift from a mesenchymal to a proneural phenotype in the tumor periphery [5]. Importantly, tumor cells from individual cell states [2] and from the edge region [6] exhibit tumor initiating potential, suggesting that these functional transitions may be critical for recurrence and may serve as candidates for targeted therapies.

We also identify multiple novel cell type-specific regional and functional signatures in glioma. Most notably, we reveal potential mechanisms underlying functional shifts driven by cell-cell communication among tumor cells and non-tumor cells. In terms of neuron-glioma interaction, previous studies have described glioma synapses, integration of glioma into neural circuits, increased excitability in glioma-infiltrated brains, and promotion of tumor growth driven by neurexin signaling following neuronal activity [7, 8]. We observe that the malignant cell states driving neuron-glioma interactions are the NPC-like, OPC-like, and cycling cells. On the other hand, the malignant cell states driving vasculature-glioma interactions are the MES-like, AC-like, and cycling cells. This dichotomy points to specific functional contributions of each cell state to tumor progression, implicating NPC-/OPC-like cells in growth, MES-/AC-like cells in migration, and cycling cells in both. Comparing the core and edge regions reveals that both interactions are decreased in the edge. In addition, the edge region exhibits re-wired vascular and immune communication pathways, along with increased interleukin signaling from neurons to immune cells. To our knowledge, these shifts in communication patterns at the tumor periphery have not yet been reported, and these axes serve as candidates for further investigation aiming to understand how to prevent or delay tumor recurrence by manipulating cell-cell interactions in the edge region.

One important limitation of this study is the cohort size (n=7 total, n=5 for NE vs CE comparisons). Although we attempt to combat this with strict significance filtering and careful selection of appropriate statistical tests and models, we are ultimately still limited by the overall sample size. We plan to address this by continuing to sequence newly acquired tumor specimens as they arrive, with the hope of releasing to the research community a publicly available glioma snRNA-seq atlas allowing visualization and analysis of the tumors that we have sequenced. Secondly, it will be important to perform this type of characterization using primary glioma tumors, which we also hope to achieve with future sequencing experiments. Third, beyond the identified cell populations and molecular networks, it would be of great use to delineate specific upstream mechanisms underlying the observed changes. One way to approach this would be to examine the chromatin accessibility profiles near strong candidate genes, which could reveal specific gene regulatory molecules driving the relevant changes. Another would be to experiment with gliomasphere cell lines and glioma mouse xenotransplant models to manipulate cells or pathways of interest. Fourth, we have not performed a targeted characterization of several typical aberrant features of glioma, such as hypoxia in the core. To address this, I plan to perform a comprehensive metabolomic profiling of the CE and NE regions, which could further implicate certain cell populations and functional pathways as candidate recurrence drivers that can be targeted with novel therapies.

Making significant progress in managing devastating diseases like glioma is clearly dependent on gaining a deeper understanding of the cellular and molecular complexity within the relevant compartments. We show here that leveraging integrative analytical frameworks focusing on complex features like functional enrichment and cell-cell interaction enable us to disentangle the relationships driving disease progression. The above two approaches both implicate specific cellular (MES-like, vascular, neuron, immune) and molecular (E2F1, VEGF) drivers of glioma progression in specific contexts, presenting a promising strategy for comprehensive and accurate deconvolution of high dimensional biological networks contributing to challenging CNS diseases.

References

1. Verhaak, R. G. W. *et al.* An integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1. *Cancer Cell* **17**, 98 (2010).

2. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

3. Müller, S. *et al.* Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biology* **18**, 234 (2017).

4. Darmanis, S. *et al.* Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep* **21**, 1399–1410 (2017).

5. Jin, X. *et al.* Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat Med* **23**, 1352–1361 (2017).

6. Bastola, S. *et al.* Glioma-initiating cells at tumor edge gain signals from tumor core cells to promote their malignancy. *Nat Commun* **11**, 4660 (2020).

7. Venkatesh, H. S. *et al.* Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion. *Cell* **161**, 803–816 (2015).

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