Glioblastoma Neurovascular Progenitor Orchestrates Tumor Cell Type Diversity

Elisa Fazzari[1], Daria J. Azizad[1], Kwanha Yu[2], Weihong Ge[1], Matthew X. Li[1], Patricia R. Nano[1], Ryan L. Kan[1], Hong A. Tum[3], Christopher Tse[3], Nicholas A. Bayley[3], Vjola Haka[3], Dimitri Cadet[3], Travis Perryman[10], Jose A. Soto[1], Brittney Wick[4], David R. Raleigh[5,6,7], Elizabeth E. Crouch[8,9], Kunal S. Patel[10], Linda M. Liau[10], Benjamin Deneen[2], David A. Nathanson[3], Aparna Bhaduri[1#]

[1]Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, California, Los Angeles, CA, USA.
[2]Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX, USA.
[3]Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[4]Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA
[5]Department of Radiation Oncology, University of California San Francisco, San Francisco, California, USA.
[6]Department of Neurological Surgery, University of California San Francisco, San Francisco, California, USA.
[7]Department of Pathology, University of California San Francisco, San Francisco, California, USA.
[8]Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA.
[9]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, CA, USA
[10]Department of Neurosurgery, David Geffen School of Medicine, University of California, Los Angeles, California, Los Angeles, CA, USA.
[#]Correspondence should be addressed to ABhaduri@mednet.ucla.edu

Abstract:

Glioblastoma (GBM) is the deadliest form of primary brain tumor with limited treatment options. Recent studies have profiled GBM tumor heterogeneity, revealing numerous axes of variation that explain the molecular and spatial features of the tumor. Here, we seek to bridge descriptive characterization of GBM cell type heterogeneity with the functional role of individual populations within the tumor. Our lens leverages a gene program-centric meta-atlas of published transcriptomic studies to identify commonalities between diverse tumors and cell types in order to decipher the mechanisms that drive them. This approach led to the discovery of a tumor-derived stem cell population with mixed vascular and neural stem cell features, termed a neurovascular progenitor (NVP). Following *in situ* validation and molecular characterization of NVP cells in GBM patient samples, we characterized their function *in vivo.* Genetic depletion of NVP cells resulted in altered tumor cell composition, fewer cycling cells, and extended survival, underscoring their critical functional role. Clonal analysis of primary patient tumors in a human organoid tumor transplantation system demonstrated that the NVP has dual potency, generating both neuronal and vascular tumor cells. Although NVP cells comprise a small fraction of the tumor, these clonal analyses demonstrated that they strongly contribute to the total number of cycling cells in the tumor and generate a defined subset of the whole tumor. This study represents a paradigm by which cell type-specific interrogation of tumor populations can be used to study functional heterogeneity and therapeutically targetable vulnerabilities of GBM.

Glioblastoma (GBM) is the most common and aggressive form of adult brain cancer, with a dismal prognosis of about 21 months of survival post diagnosis[1]. Standard of care therapy largely employs pan-cancer treatments including radiation and temozolomide chemotherapy[2,3] that inevitably result in tumor recurrence. Numerous other therapeutic approaches are in development, including immunotherapies[4,5], drugs targeting oncogenic driver mutations[6,7], and therapies to disrupt synaptic signaling between the tumor and its microenvironment[8-11]. While there is optimism about the efficacy of these approaches, foundational understanding of tumor composition and function is essential to expand our ability to develop personalized therapeutic approaches. Extensive molecular and spatial characterization has shown that GBM displays immense inter- and intra-tumoral heterogeneity[12-26]. Understanding the landscape of progenitor subtypes within GBM patient tumors offers a window to understand the ability for GBM to evolve and adapt. Characterizing how these subpopulations proliferate and contribute to tumor diversity would provide potential targets for cell type-based therapy that could be paired with additional treatments to extend patient lifespan more effectively.

Before the field can effectively design these cell type-based therapies, open questions remain regarding the functional role of GBM tumor heterogeneity, including granular descriptions of what progenitor

populations exist and their unique roles within the tumor. To characterize the landscape of progenitor cell types in human primary GBM, we compiled a meta-atlas of single-cell transcriptomes from seven published studies[12,14-16,18,24,25]. Using a new bioinformatic method for gene program analysis[27], we derived GBM gene expression meta-modules represented across multiple patient tumors. Our analyses validated previously described GBM gene programs and revealed novel tumor progenitor subtypes. Among these subtypes, we discovered a tumor-derived stem cell population exhibiting a mixed vascular identity and transcriptional programs reminiscent of neural stem cells, which we identified as a putative neurovascular progenitor (NVP). To our knowledge, this population has not been previously described in GBM nor in the human brain. Here, we validate the existence of NVP cells across tumors, describe their morphological and molecular features, and use both *in vivo* and direct-from-patient systems to characterize their function. We demonstrate that NVP cells have dual-fate capabilities in clonally generating neuron-like and vascular-like tumor populations, and that they generate a defined fraction of the overall tumor. Despite their low abundance in patient tumors, they disproportionately generate cycling cells and are competent to give rise to the majority, but not all, total tumor subtypes; elimination of NVP from an *in vivo* tumor model results in extended survival, highlighting the functional relevance of this clonal role within the tumor.

Results

Generation of meta-modules describes universal and specific GBM progenitor programs

Previous efforts have profiled GBM, and we reasoned that compiling these datasets could allow us to uncover nuanced gene programs essential to drive the tumor. We specifically focused on a gene module-centric approach that would enable the derivation of modules that underly these cell types and may enable the elucidation of rare but functionally significant populations within the tumor. Recently applied to datasets from the developing human brain, this gene module derivation strategy has the capability to identify cell type-specific and biologically dynamic transcriptional networks[27]. Thus, to highlight functional gene programs of GBM, we generated a meta-atlas of existing datasets[12,14-16,18,24,25]. While many of these datasets have been previously analyzed and aggregated, our approach of including only the IDH1 wild-type ($IDH1^{WT}$), adult, direct-from-patient samples accounts for the unique nature of GBM compared to other gliomas and removes any contribution of *in vitro* culture methods, cell type selection, or distinct pathologies that may interfere with gene module identification. We first performed uniform quality control on each individual tumor from every dataset, filtering for primary $IDH1^{WT}$ adult tumors and only neoplastic cells (Fig 1A - B, STable1) (Methods). We then performed tumor-specific unsupervised clustering using calculated cluster markers for each individual tumor. To focus on markers that were specific and enriched in each cluster, we leveraged our previously developed genescore metric that integrates the degree of cluster marker enrichment with marker specificity[12,28] to rank order these cluster markers across tumors (Methods). The genescores from each dataset were filtered based on dataset-specific thresholds and then aggregated for meta-module generation. The top genescores in this unified analysis represented the common sources of variation across datasets; additional analysis validated that each dataset contributed approximately equally to these top genescores (SFig 1A - B). Top genescores were further correlated to derive our meta-modules which represented groups of genes that are co-expressed across multiple tumors and datasets (Fig 1C). Gene meta-modules were defined by a correlation threshold and modules larger than 10 genes were retained for further analysis with no upper limit on module size. The resulting 152 modules ranged from 10 to 323 with median size of 56 genes (Fig 1D). The number of genes per module did not affect the likelihood of expression across cells (SFig 1C).

For the purpose of visualization, we generated an integrated UMAP on which module activity scores are represented. We identify that our modules correspond to previously described tumor cell states (Fig 1E, SFig 2) and that these integrated clusters can be annotated by adult and neurodevelopmental-like cell types, including those related to human cortical development such as radial glia (RG) and outer radial glia (oRG). Module activity scores show that modules associated with cell type or state identities such as oligodendrocyte, immune reactive, neuronal and cycling programs are expressed in the expected populations (Fig 1E-F). Amongst the 152 modules, however, biological programs not captured by these cell type designations were further identified. Examples of these program-specific meta-modules include cilia function, BMP signaling, and HIF1A signaling (STable2 - 3). Amongst these modules, numerous modules were widely expressed across all tumors, while others differentiated distinct groups of tumors (SFig 1D). Previous studies have published sets of gene modules for GBM, incorporating different aspects of the microenvironment and immune compartments[18,29-31]. Our gene module approach elicited modules that incorporate aspects of these published modules, while increasing granularity of the gene sets, which enables the study of programs that transcend cell type boundaries (SFig 2A-D).

Among the modules with cell type-specific expression but broad tumor representation, we noticed multiple modules resembling progenitor gene programs (Fig 2A). For example, module 6 which contains *NES*, was localized amongst radial glia and mesenchymal cell populations, whereas the canonical glioma stem cell marker module 24 containing *PROM1* (encoding for CD133[32]) was diffusely active in neural stem cell and oligodendrocyte precursor cell populations. Module 80 containing *SOX2* was broadly represented in the majority of cell types in our atlas, consistent with previous descriptions of *SOX2* in GBM[33]. Other modules including Module 41 (containing *YAP1*), module 43 (containing *SOX6*), and module 51 (containing neural progenitor marker *PAX6*) were expressed in a subset of tumor cells restricted to one or two annotated cell types. In each of these cases, the represented gene was expressed in expected cell type populations, whereas the module activity was more restricted, suggesting a specificity driven by the co-expression network and highlighting the value of examining the broader biological program. Scoring scRNA-seq datasets of the normal developing brain confirmed expression of our stemness meta-modules in a diverse array of human neural progenitor cell populations[27,34] (SFig 3).

Identification of a mixed-vascular population expressing stemness modules

We were particularly intrigued that multiple progenitor programs overlapped with a subclass of vascular cells that also uniquely expressed module 11 (Fig 2B). Module 11 is comprised of genes representing a mixed vascular identity including that of mural and endothelial cells (Fig 2B, STable2). Notably, a recently published vascular atlas orthogonally observed many module 11 genes were upregulated in glioma-associated vasculature compared to normal brain[35], though no module analysis nor function was interrogated in this study. Because all of our data were filtered to only include tumor populations, we determined this is a tumor-derived cell type. Intriguingly, this population also exhibited expression of markers reminiscent of neural progenitors (Fig 2B-C). Among these progenitor genes, markers of oRG identity including HOPX were expressed in cells with high module 11 activity (Fig 2B). oRGs are subsets of radial glia progenitors that are highly expanded in the developing human cortex compared to rodents, and although they are thought to disappear from the developing human cortex by birth, were identified by us and others to be reactivated in GBM, with functional roles in tumor migration[12,22]. Co-expression of oRG markers with mixed vascular identity was surprising as this has not been well described in the developing human brain, the adult human brain, nor GBM.

Previous work has convincingly shown that tumor associated vasculature plays a pivotal role in tumor progression. For example, pericytes derived from glioma stem cells *in vitro* and *in vivo* contribute to increased tumor proliferation[36]. Similarly, a recent study showed that cancer-associated fibroblasts in GBM promote glioma stem cell growth[37]. Additionally, introducing tumorigenic alterations in naïve human brain perivascular cells results in brain tumors[38]. Moreover, extensive work characterizing pericytes and PDGFRβ has been performed in GBM and has demonstrated a critical role for this signaling pathway in maintaining glioma stem cell renewal[39], increasing tumor cell proliferation[40,41], and contributing to treatment resistance[42]. We sought to build upon this literature and explore how our tumor-derived mixed vascular population differs from these previously described cell types by co-expressing features of progenitor identity as well. Additionally, we aimed to investigate whether the rare nature of this signature indicated a technical or dataset driven bias in our analysis. Arguing against this possibility, we observed that cells with high module 11 score were ubiquitous across tumors and datasets (Fig 2D). Sensitive to the fact that these tumors were dissociated without enrichment for this population in mind and that vascular cells typically require specialized isolation protocols[43], we next performed *in situ* immunofluorescence analysis to validate the presence of this hybrid cell type.

In situ validation of mixed vascular identity with canonical stemness markers

The majority of our module 11 genes are mural cell markers so we first stained for canonical pericyte markers and observed morphologies similar to those described in the literature, including large and small vessel associated mural cells[44] (Fig 2E). We also observed parenchymal co-localization of pericyte markers (and meta-module 11 members) PDGFRβ and NOTCH3 (Fig 2E, SFig 4B).

To examine the *in situ* co-localization of pericyte markers and neural progenitor markers, we performed immunostaining for PDGFRβ, NOTCH3, with NES (common glioma stem cell marker[45]) and HOPX (marker for oRG[46]). We hypothesized that these cells might be spatially arranged near canonical pericyte populations, and to explore this we performed tile scans to enable visualization of diverse niches within the tumor. PDGFRβ and NES double-positive cells have been previously described in the context of tissue regeneration after injury[47,48] and have been identified to be a characteristic of pericytes that can be reprogrammed into a neural lineage[49], however this identity has not been described in the context of GBM. Across these tumor sections, we found that a subset of pericyte markers co-localized with NES and HOPX, with parallel morphologies observed in

both co-stains, leading us to tentatively refer to these populations as “neurovascular progenitor” (NVP) cells (Fig 2E, SFig 4-5). Indeed, we observed that putative pericytes and NVP cells were seen in anatomical clusters or apart from other putative pericytes and in vessel-associated and parenchymal regions (Fig 2E). To validate that this was not a tumor-specific phenomenon, we repeated this staining across 3 patient tumor samples and noted similar morphologies in samples from each patient (SFig 4-5). These data gave us confidence that NVP cells exist across multiple tumors.

FACS enrichment from primary tumors allows for molecular characterization of NVP

Given the presence of putative NVP cells as a small fraction of the tumor population in the single-cell meta-atlas, but with validation *in situ*, we sought to more extensively characterize this population. To do so, we employed a sorting strategy that would allow us to enrich for tumor vasculature populations, including NVP. Direct-from-patient GBM samples were sorted for PDGFRβ expression immediately after surgical resection (without expansion or plating) using fluorescence activated cell sorting (FACS) (Fig 3A, STable4). The sorted and unsorted cells acquired from three patient tumor specimens were immediately captured for single-cell RNA-sequencing (scRNA-seq) and annotated for tumor identity with our cell types from the GBM meta-atlas (Fig 3A - C, SFig 6A). The resulting dataset consisted of 15,996 cells after quality control and copy number variation filtration for tumor cells (Methods). Using a combination of module 11 activity scoring and marker genes from our mixed vascular population, we molecularly defined a subset of our sequenced cells as NVP, and used the top marker genes of this cluster to define our NVP signature (STable5). Importantly, all of these annotated NVP cells contained copy number variation indicating they were tumor cells (SFig 6B), as opposed to tumor-associated vascular populations that are well described in the literature[50]. This enrichment and analysis provided confidence that NVP cells reliably exist across tumors and can be isolated for further characterization.

Interrogation of vascular and neural stem cell gene module co-expression

Within our annotated NVP population, we observed the expression of marker genes related to mural (*PDGFRβ, NOTCH3)*, neural stem cell (*HES1, NES, HOPX*) and endothelial identities (*TIE1, PECAM1*) (Fig 3C). Previous literature has demonstrated the presence of developmental marker genes in tumor vascular cells, and we sought to interrogate whether the presence of these markers was indicative of a larger gene program. Thus, we aggregated modules from published studies[51] and calculated our own modules from published datasets[52-54] to explore their activity in NVP cells and the rest of our data. NVP cells showed enrichment for adult vascular gene programs[54], developmental vascular gene programs[53], murine cancer stem cell[51], normal stem cell programs[55,56], and a recently identified meningioma NOTCH3+ stemness signature[52] (Fig 3D). These analyses highlighted that NVP cells co-express modules related to multiple vascular identities and stemness programs. However, NVP cells are neither the only vascular populations within the tumor nor the only progenitor populations, consistent with our meta-atlas analysis and previous studies[12,14,18]. Given the expression of oRG marker *HOPX* in our NVP cells, we explored representation of a curated oRG transcription network[46]. We observed that several additional canonical markers are expressed by the NVP cells (Fig 3D), further highlighting their potential to serve as progenitors within the tumor.

Identification and elimination of NVP in an in vivo *developmental GBM model extends survival*

Given the link to developmental cell types and progenitors in GBM, we explored scRNA-seq data from a published, tractable, developmentally derived *in vivo* model of GBM[57]. This approach uses *in utero* electroporation to knockdown three tumor suppressors, giving rise to GBM-like tumors postnatally. We re-analyzed these data from the lens of our human-annotated GBM cell types, finding numerous populations of interest including a mixed vascular population that resembled our NVP (Fig 3E). Excitingly, this model of GBM allows exploration of diverse progenitor populations that also exist in human patient tumors[57]. Using the same signatures as before, we scored the mouse tumor cells for these vascular and progenitor modules, finding that the mixed vascular population scored highly and specifically for our NVP score and similarly co-expressed various vascular and progenitor features (Fig 3E). Identifying NVP in this *in vivo* model of GBM provided orthogonal confirmation of the significance of NVP cells and provided a system in which to manipulate this population. Therefore, we performed *in utero* electroporation using this established *3xCr* approach which initiates tumors with CRISPR-mediated knockdown of *Pten*, *Nf1*, and *Trp53*.

Using this developmental *in vivo* tumor system, we sought to explore how NVP contributes to GBM composition. Our strategy was to leverage the mouse system to interrogate NVP in a way that would be cross applicable to the human GBM context. Although PDGFRβ is an excellent surface marker for enrichment of

NVP cells, in both our meta-atlas and our sorted cell analysis we noted that it was not entirely specific to the NVP population. When exploring a genetic manipulation, we were not restricted to surface markers and thus chose a handful of genes to target for knockout that were both specific to the NVP cluster and that were conserved between human and mouse NVP. These included highly specific transcription factors (*SOX18* and *FOXC1*), neural stem cell marker gene *FAM107A*, endothelial marker *CLDN5* and mural marker *RGS5* (SFig 7A). Our experimental condition applied the same *3xCr* approach but added in simultaneous knockdown of our five NVP targets (*3xCr+NVP^{null}*), enabling exploration of how the resultant tumors would differ without NVP (Fig 4A).

Our previous analysis of this model indicated robust and consistent representation of the NVP population at P70, thus we sacrificed 4 control *3xCr* and 4 experimental *3xCr+NVP^{null}* mice with equal numbers of male and female animals at this timepoint (P70) to explore the impact of NVP on tumor cell type composition (SFig 7B). We performed single-nuclei RNA-sequencing (snRNA-seq) and obtained data from 80,530 nuclei after quality control (Fig 4B). As before, we annotated cell types with our meta-atlas reference, observing that the *3xCr+NVP^{null}* resulted in a 96% reduction of the NVP cell fraction (Fig 4C). To further explore the differences between the *3xCr* and *3xCr+NVP^{null}* mice, we scored these data for the same modules that we observed to be activated in human NVP cells in the mouse tumor cells. We noted that a number of these modules were significantly downregulated in the *3xCr+NVP^{null}* condition, indicating that these gene programs are co-expressed in the mouse NVP cells, were specifically targeted by our depletion strategy, and may have functional roles in NVP maintenance (Fig 4D). These results validated the parallels between the mouse and human NVP populations, enabling us to further explore the impact of NVP on tumor composition. Thus, we calculated the cell type composition for each of the 8 profiled tumors and compared how they differed with and without NVP to explore how NVP functionally contributes to GBM tumor biology. In the *3xCr+NVP^{null}* mice, there was a decrease in the number of cycling cells, vascular cells, and a modest decrease in the number of tumor progenitors such as oligodendrocyte precursor cells (OPCs), and oRGs. There was also an increase in certain cell types including differentiated populations such as astrocytes and oligodendrocytes, as well as neuronal progenitor cells (NPC) (Fig 4C-E) that may be playing a compensatory role. These data describe a specific role for NVP within the tumor and led us to hypothesize that NVP may generate multiple cell types within GBM tumors. We additionally evaluated survival of a separate cohort of 3*xCr* (n=23) and *3xCr+NVP^{null}* (n=28) mice. The *3xCr+NVP^{null}* cohort showed an increase in median overall survival compared to the *3xCr* cohort (119 vs. 91.5 days, *log-rank p-value = .03*) (Fig 4F), indicating that the shifts in cycling populations and cell populations are sufficient to impact tumor growth and *in vivo* survival.

DNA barcoding-based clonal analysis of enriched NVP cells from human GBM samples demonstrates dual fate of NVP

The *in vivo* experiments provided direct evidence that NVP cells have essential roles in promoting GBM tumors. We sought to test our hypothesis that NVP generates tumor cell types in human GBM tumors, especially given recent findings that there exist species differences in brain vascular subtypes[54]. Studying human tumors involves additional challenges compared to mouse *in vivo* systems, including the observation that *in vitro* expansion of primary tumors alters tumor composition and may result in decreased progenitor complexity[58]. To address these challenges, we developed and optimized a human organoid tumor transplantation (HOTT) system that both preserves tumor heterogeneity and transcriptional fidelity to the parent tumor while also modeling key features of the tumor microenvironment[12,59]. Briefly, tumor cells directly from the patient can be labeled with lentivirus containing GFP markers and transplanted into an already existing cortical organoid that models key features of the developing human cortex. A unique feature of this HOTT system is that it can also enable the exploration of isolated subsets of tumor cells and allows for additional molecular modifications to be introduced. We took advantage of these features by repeating our enrichment for NVP by sorting for PDGFRβ+ cells and performed clonal tracking using the well described CellTag DNA barcoding system[60] (Fig 5A - B).

Upon transplantation of the NVP enriched tumor fraction into HOTT, we observed the recapitulation of complex cell morphology that we had previously seen *in situ* (Fig 5B). To perform this experiment, we first infected freshly dissociated GBM cells from 6 patients with a lentiviral library containing scRNA-seq compatible CellTag barcodes and a GFP tag before transplanting onto cortical organoids. We delivered the virus at a multiplicity of infection that would result in 3 - 4 barcodes introduced in each cell, enabling high confidence clone calling in subsequent analysis (Methods). We then allowed the tumor cells to proliferate before harvesting for scRNA-seq. The paired analysis of CellTag barcodes and transcriptional signatures provides the identity of tumor cells derived from the same clone.

As before, we performed quality control analysis and filtered for tumor cells using copy number variation analysis, yielding 20,153 cells across 6 unique tumors. Cells were uniformly labeled with CellTag-containing lentivirus (Fig. 5B), giving all cells the opportunity to be included in a clone. Our analysis showed that across our 6 tumors, we averaged a multiplicity of infection of 2-3, meaning that the probability of any two cells randomly receiving the same combination of barcodes is infinitesimally small, as has been previously described[60]. Using the CellTag clone calling pipeline[61], we identified 1465 clones ranging in membership from 2 to 28 cells, with an average size around 3. The transcriptional analysis was performed orthogonal to the clonal analysis, and identified a subset of cell types that we typically observe within GBM tumors. After performing clonal analysis on each tumor individually, we merged all tumors and clone identities in order to visualize trends in clone membership on the same UMAP space. Each clone represents a set of cells that are derived from a common progenitor in the HOTT system, though the parent cell may or may not remain or be captured by our system. As such, we present the clones from our analysis in sum (Fig 5E) and filtered for those that are restricted to containing one or more NVP cells in the clone (Fig 5F, bottom right, UCSC Genome Browser: https://gbm-nvp.cells.ucsc.edu). While here we pictorially represent select representations of clones with interesting and unexpected cell type membership, the vast majority of the total clones were comprised of cells in the same cluster, especially that of cycling cells. Because having multiple clone members of the same cell type is more likely than having clone members of different cell types, this provides confidence that the clone calling strategy indeed resulted in related cells being highlighted. Upon analyzing the cell types represented in these clonal relationships, we observed that cycling cells and OPCs were overrepresented. This was particularly interesting given the decrease in both the cycling population and OPCs observed when NVP was depleted in our *in vivo* experiment, suggesting an outsize role for NVP cells in generating cycling cells compared to their proportional representation in the tumor.

Our informatic analysis and *in situ* validation led us to hypothesize that NVP cells have the propensity to generate both neuronal and vascular cells within the tumor. Indeed, we found several clones across multiple tumors that contained an NVP cell and at least one neuronal and/or vascular cell, confirming the dual potency of this population (Fig 5D). Importantly, the direct clonal relationship between NVP cells, vascular cells, and neurons is permitted by this DNA-barcode clone analysis and is not accessible by traditional scRNA-seq approaches. While we previously demonstrated GBM cell type composition changes as a result of NVP depletion, this direct-from-patient clone interrogation advances our understanding of NVP function by providing a concrete relationship between multiple tumor cell types. We note that across canonical models of GBM, neuronal and vascular populations are thought to sit at opposite poles of GBM transcriptional identities[20,62], highlighting a unique aspect of NVP cells in that they are capable of giving rise to divergent states within the tumor.

NVP-enriched samples can generate a majority *of the parent tumor*

For a subset of our tumors (3 of 6), we performed a parallel transplantation into HOTT of the unsorted parent tumor in order to identify what fraction of the parent tumor can be generated by the NVP enriched fraction. We performed this control, as opposed to simply profiling the input tumor, in order to account for any cell type bias generated by the HOTT system. Across these 3 paired tumors, we recovered 74,234 cells. The NVP enriched fraction was able to give rise to 59% of parent tumor cell types, indicating a substantial contribution to the overall tumor landscape (Fig 5G). Notably, the NVP enriched fraction was deficient in numerous cell types present in the parent tumors, including mesenchymal, astrocyte, and transition state cells, validating our framework that although NVP has an essential role in GBM tumors, there are multiple progenitor cell types within any given patient tumor that generate these other populations. Again, these data are concordant with what we observed *in vivo*, showing that NVP has a functional role in promoting certain GBM cell types, but does not impact every cell in the tumor.

Discussion

Here, we utilized a new meta-atlas generation method to identify unified gene programs specifically within adult IDH1^{wt} GBM, leading to the discovery and functional characterization of a novel GBM neurovascular progenitor population (NVP). The initial goal of this meta-atlas approach was to effectively describe functional gene programs of GBM cell subpopulations with a focus on progenitor diversity. Indeed, the intersection of modules and marker genes across clusters gave us a robust cell type annotation that enabled the precise description of GBM cell types and progenitor identities, as well as provided us with coherence for our downstream analysis. Our modules and cell types are consistent with the major features of tumors used by others in the field to describe GBM biology (SFig 2). The diversity of progenitor modules that emerged from our

meta-atlas additionally provided us with nuance in being able to describe the 'flavors' of progenitor identity that exist across heterogeneous cell types and GBM tumors. This led to our initial interest in pursuing NVP as a potential tumor progenitor due to the surprising confluence of multiple progenitor modules and a cluster with mixed vascular identity. Importantly, these progenitor modules and their intersection with cell type annotations serve as a resource for the community to further explore additional tumor populations and their progenitor features, accessible through the cell browsers we have constructed for this study.

We discovered that although NVP cells comprise a small fraction of the total tumor, they disproportionately contribute to cycling cells and have dual fate potential in giving rise to neurons and/or vascular cells. This unexpected role of NVP initially seemed bizarre, but literature on the existence of pericyte progenitor cells in other tissue systems provide context for why NVP cells may exist in GBM. Pericytes across tissue systems have been described to be relatively undifferentiated cells and contribute to tissue repair after injury. Additionally the co-expression of pericyte markers with developmentally relevant markers such as HOPX[63] and NES have been described in mesodermal tissues including the heart and skeletal muscle[48] and reside in proximity to neurogenic cells where they participate in regenerative processes[47]. In the brain, primate pericytes have shown neurogenic potential after ischemia[64] and human pericytes can be directly reprogrammed into functional neurons[49]. This raises the hypothesis that dual fate NVP cells in GBM have aberrantly resurrected these injury response mechanisms contributing to cancer adaptation and plasticity.

In this study, we focused on the *in situ* and molecular validation of the existence of the NVP population, followed by functional characterization *in vivo* and in human patient GBM. *In vivo* experiments showed an expanded survival of mice with tumors lacking NVP, demonstrating that targeting NVP could be utilized as a therapeutic strategy. However, even in this experiment, compensatory mechanisms including the higher proportion of neural progenitor cells in these tumors reinforce that multiple progenitor populations exist within GBM tumors and multi-progenitor targeting will be required to effectively prevent tumor progression. This is further corroborated by our analysis that although NVP is disproportionately clonally linked to a subset of tumor cell types, entire populations within the parent tumor have no clonal relationship to NVP and therefore likely emerge from distinct progenitors. Additional work is needed to identify where in the progenitor hierarchy of GBM NVP cells may sit, as our clonal analysis does not provide historical information about either the tumor cell of origin or NVP emergence. The restriction of NVP potency to give rise to many but not all GBM tumor cell types presents a shift in our understanding of how direct-from-patient stem-like populations within the tumor function; this observation of restricted fate further bolsters the argument for why cell type-specific therapeutic approaches may be a valuable tool. Thus, our characterization of NVP cells serves as a paradigm that can be applied to all progenitor populations within GBM in order to advance our understanding of their role within the tumor to enable the development of effective combinatorial treatment strategies.

Acknowledgments

We would like to thank the members of the Bhaduri Lab for their insightful advice and comments on the study. We would like to thank the Broad Stem Cell Research Center Flow Cytometry core for their help in isolating cells for this project, and Charina Julian (UCSF) and Suhua Feng (UCLA) for help with running sequencing. We would like to thank Sergey Mareninov and others at the Brain Tumor Translational Research Core at UCLA for enabling tumor sample acquisition. We would like to thank Maximilian Haeussler (UCSC) for his help in compiling the genome browser. This study was generously funded by support to AB from: Swim Across America, Jonsson Comprehensive Cancer Center, Sloan Research Fellowship from the Alfred P. Sloan Foundation, NIH NCI P50CA211015 including a Career Enhancement Program Award, The Sontag Foundation (Distinguished Scholar Award), V Scholar Award from The V Foundation, The Uncle Kory Foundation, The American Cancer Society (CSCC-Team-23-980262-01-CSCC), The Margaret Early Medical Research Trust, The Rose Hills Foundation and the Broad Stem Cell Research Center. Funding for EF was provided by the David Geffen Scholarship and the UCLA-Caltech Medical Scientist Training Program (T32GM152342). BW was supported by CIRM funding DISC0-14514, a collaborative grant with AB. BW was additionally supported by NIH/NIMH RF1MH132662 and NIH/NHGRI U24HG002371. This material is based upon work by JAS supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2034835. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Contributions

The study was conceptualized by E.F. and A.B. Experimental design was performed by E.F., A.B., D.N, and B.D. P.R.N. and B.W. assisted with bioinformatic pipeline development. E.F., D.A., K.Y., M.X.L., W.G., R.L.K., A.T., N.A.B., C.S., V.H., D.C., T.P., J.A.S. performed experiments and informatics. Data interpretation was performed by E.F., A.B., E.E.C, and D.R.R. Primary tissue samples were provided by K.S.P. and L.M.L. Analysis was performed by E.F. The manuscript was written by E.F. and A.B. with input and edits from all authors.

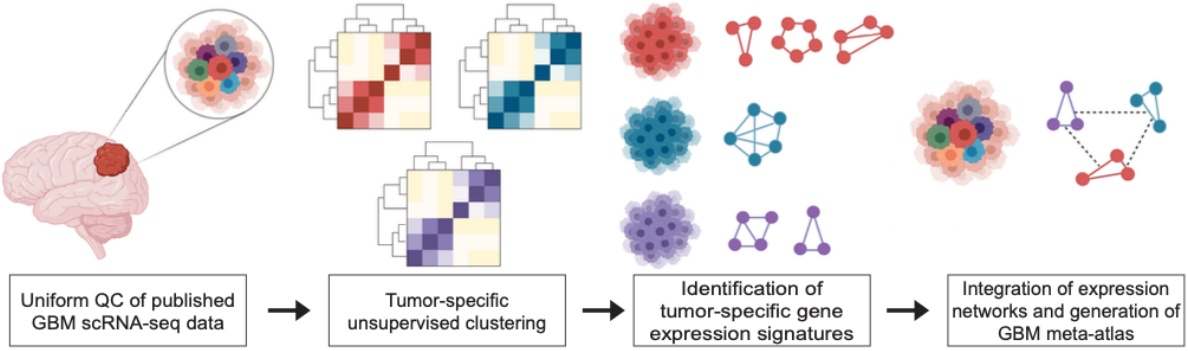
Data Availability

Our GBM meta-atlas and all original datasets generated in this study are browsable and downloadable in the UCSC Genome Browser: (https://gbm-nvp.cells.ucsc.edu), in which the data are also available as Seurat objects for download. New raw data collected from this study are also being deposited in dbGAP.

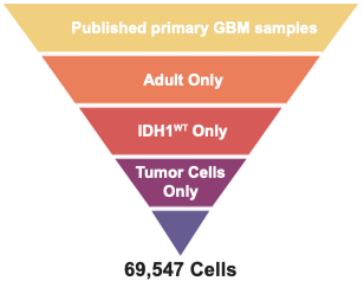
Figures and Figure Legends:

Figure 1. Meta-modules constructed from individual tumor analyses identify known and novel biological programs and cell states

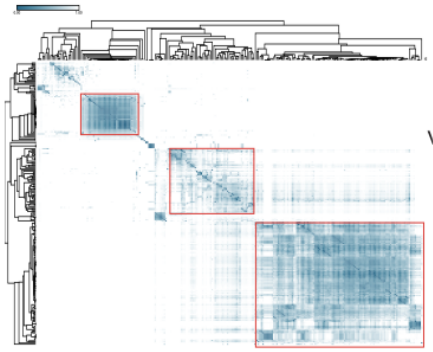
A. Workflow of meta-atlas generation



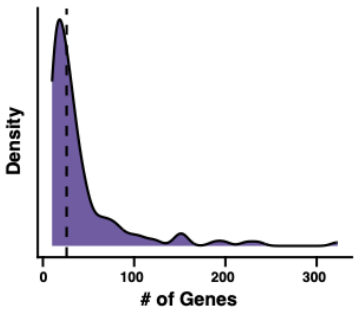
B. Exclusion criteria for meta-atlas



C. Representative heatmap of genescore correlation for meta-module generation

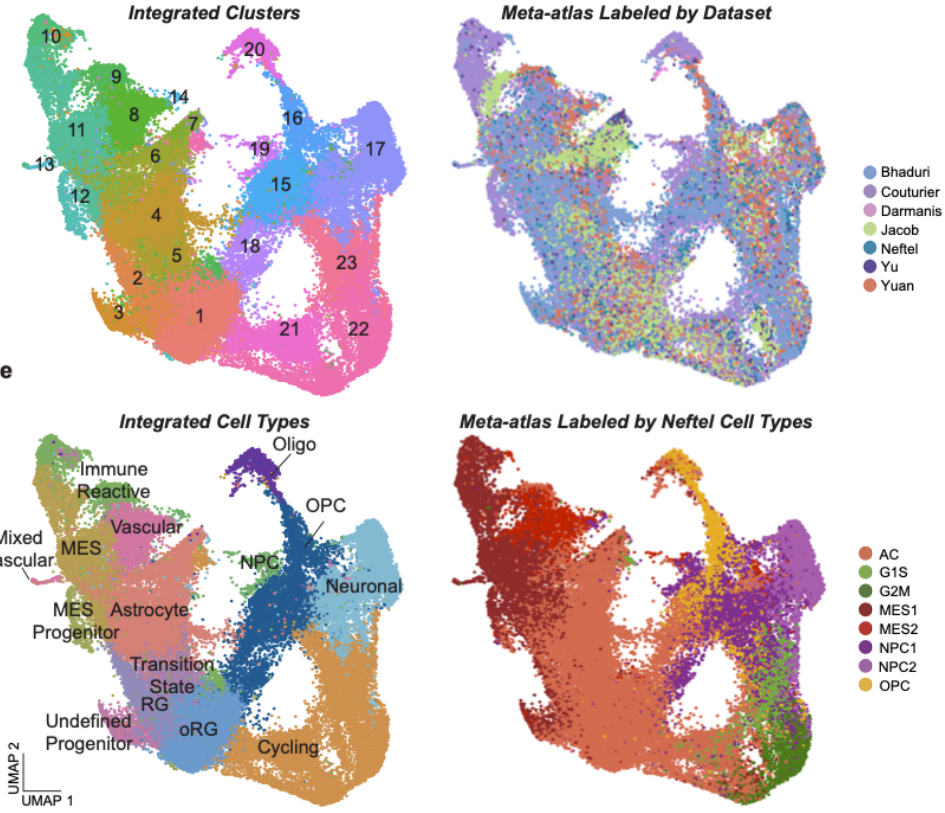


D. Distribution of meta-module sizes



Metamodule Count	152
Metamodule Sizes	10-323
Median Size	56

E. UMAPs of meta-atlas annotations



F. Module activity of cell type specific meta-module expression

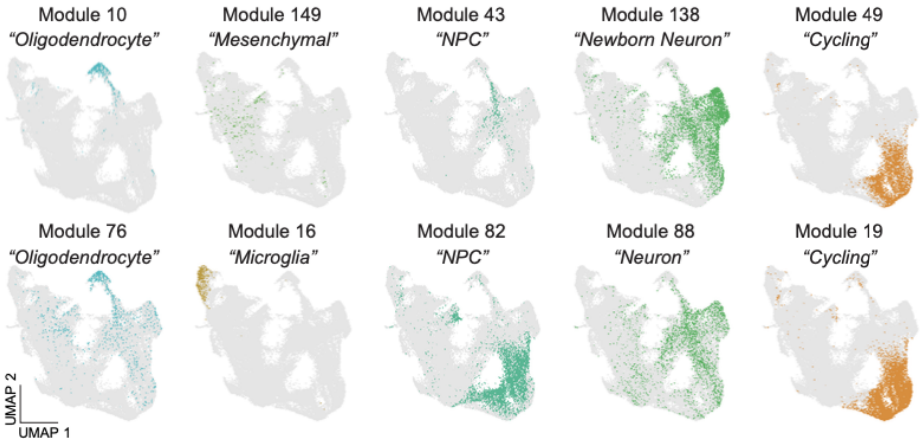


Figure 1. Meta-modules constructed from individual tumor analyses identify known and novel biological programs and cell states

A) A novel workflow of meta-atlas generation was developed to overcome technical and biological variation when interpreting gene networks in single-cell RNA sequencing (scRNA-seq) in patient glioblastoma (GBM) samples. We performed uniform quality control (QC) of each individual patient tumor sample. Hierarchical clustering was performed on each patient individually, followed by the calculation of cluster markers. Genescores were calculated to identify specifically enriched markers for each cluster. Top genescores were combined across datasets in order to identify preserved sources of variation. Hierarchical clustering of top genescores was performed to identify gene modules explaining biological processes across all datasets.

B) A stepwise exclusion process was implemented to curate the GBM meta-atlas from published primary GBM. The initial dataset included published primary GBM samples. The first criterion excluded pediatric samples, retaining only data from adult patients. Subsequently, samples harboring mutations in the isocitrate dehydrogenase 1 (IDH1) gene were excluded, focusing solely on IDH1 wild-type ($IDH1^{WT}$) samples. The final exclusion step involved the removal of non-tumor cells per the original authors' annotations. This filtration resulted in a final dataset comprising 69,547 cells.

C) Genescores were calculated on a tumor-specific basis, and scores were filtered and aggregated across the datasets. The distance matrix of the correlation of genescores between markers of individual tumors was calculated and hierarchically clustered. Red boxes highlight clusters of genes with high correlation, which are binned into meta-modules.

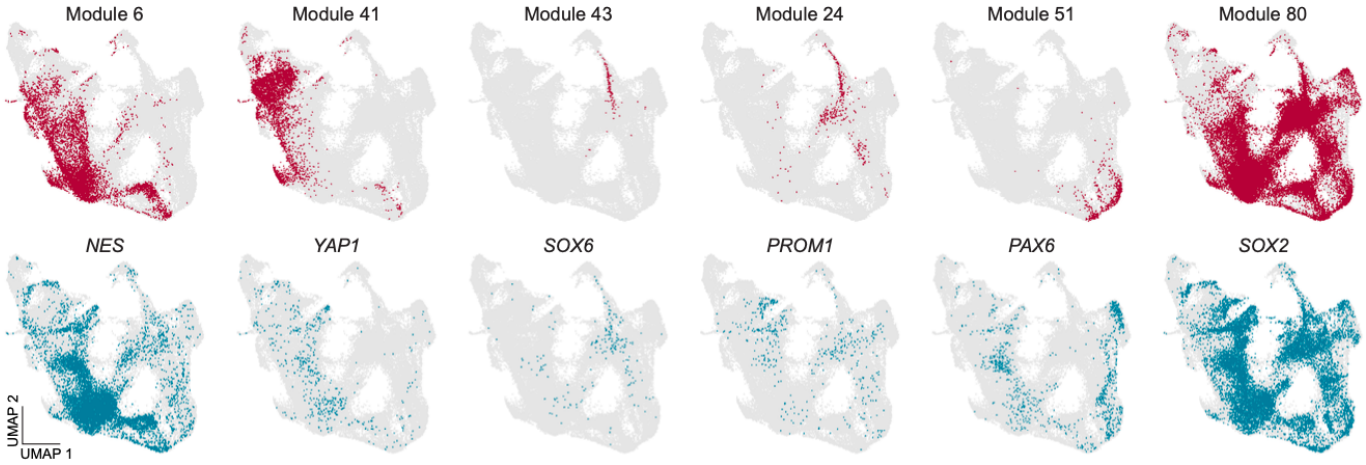
D) The distribution of meta-module sizes is shown with a density plot, indicating the frequency distribution of meta-modules by the number of genes they contain. The table provides key statistics of the meta-modules, showing a total of 152 meta-modules with sizes ranging from 10 to 323 genes and a median size of 56 genes.

E) For visualization purposes, an integrated uniform manifold approximation and projection (UMAP) of the cells in the meta-atlas was generated after uniform QC and filtration (Methods). Cells were integrated by conventional methods[65], which utilizes similar subpopulations to find common anchors between the datasets. UMAPs are labeled by cluster membership (top left), dataset (top right), cell type as annotated by cluster marker analysis (bottom left) and cell type as annotated by the highest module activity score from published meta-modules[18].

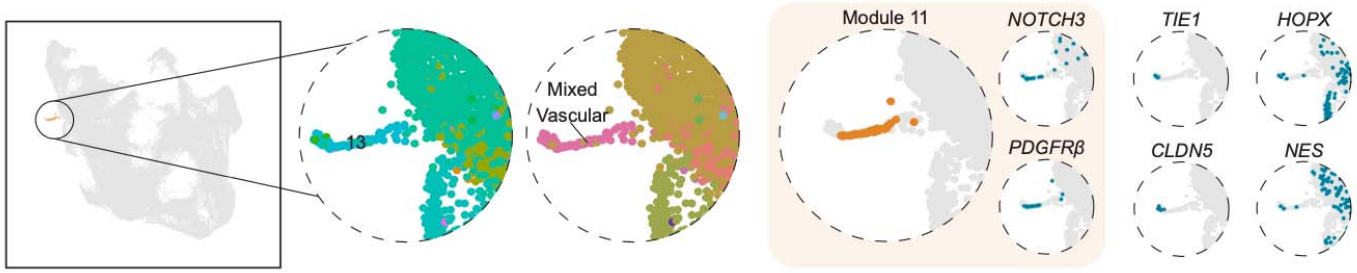
F) UMAPs displaying the module activity of cell type-specific meta-modules in the GBM meta-atlas. Each plot shows the expression pattern of a specific meta-module across the integrated dataset. The cell type specific meta-modules largely correspond to orthogonally annotated cell types (1E, bottom left and bottom right).

Figure 2. Meta-module analysis identifies progenitor features in cluster with mixed vascular identity

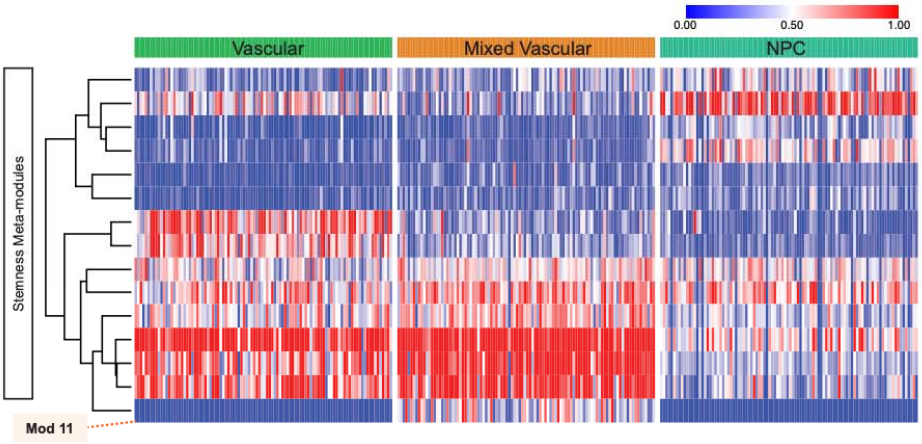
A. Feature plots of stemness modules and associated stemness genes



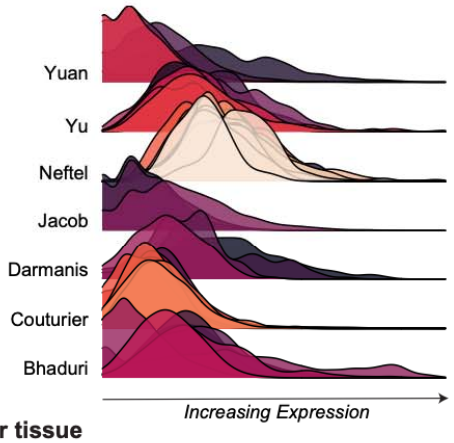
B. Expression of representative vascular and progenitor genes



C. Meta-module scores based on meta-atlas cell cluster



D. Meta-module 11 expression by tumor



E. *In situ* immunofluorescence of vascular and progenitor markers in primary tumor tissue

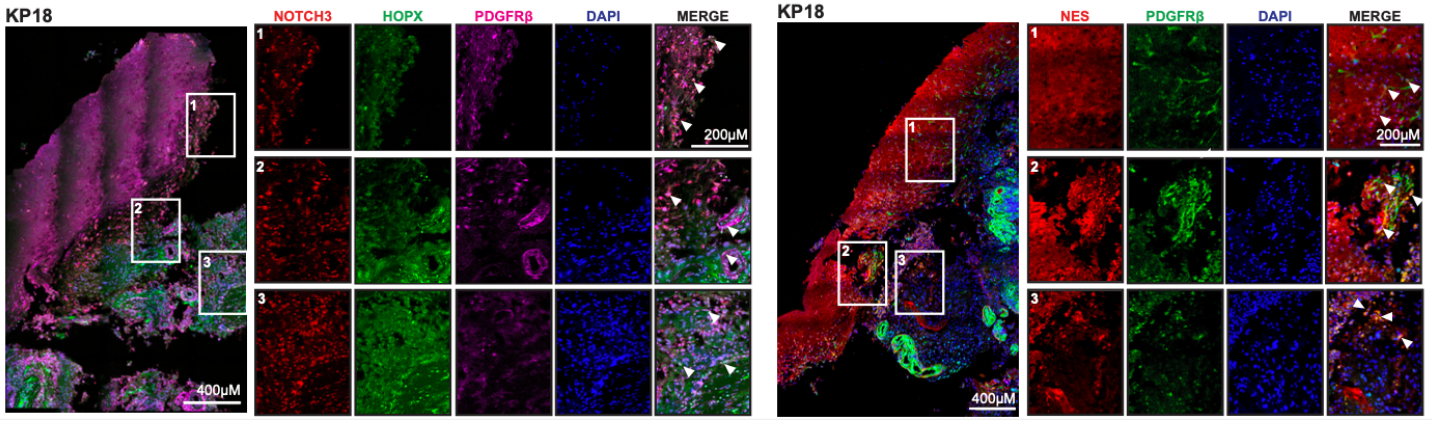


Figure 2: Meta-module analysis identifies progenitor features in cluster with mixed vascular identity

A) Our meta-modules contained multiple gene sets that appeared to correspond to different progenitor programs. UMAP plots depict the module activity of stemness-related meta-modules and the expression of associated stemness genes in the GBM meta-atlas. The associated stemness genes *NES, YAP1, SOX6, PROM1, PAX6*, and *SOX2* are shown in corresponding plots, highlighting the correspondence but incomplete overlap between module activity and gene expression.

B) Multiple progenitor programs overlapped in a small cluster annotated as “Mixed Vascular”, represented by cluster 13 on the integrated UMAP (Fig. 1E). Insets depict the expression of vascular-gene containing meta-module, meta-module 11 as well as member genes *PDGFRβ* and *NOTCH3* (in orange box). Also expressed in this cluster but not contained in meta-module 11 are neural progenitor maker *HOPX* and endothelial marker *TIE1*.

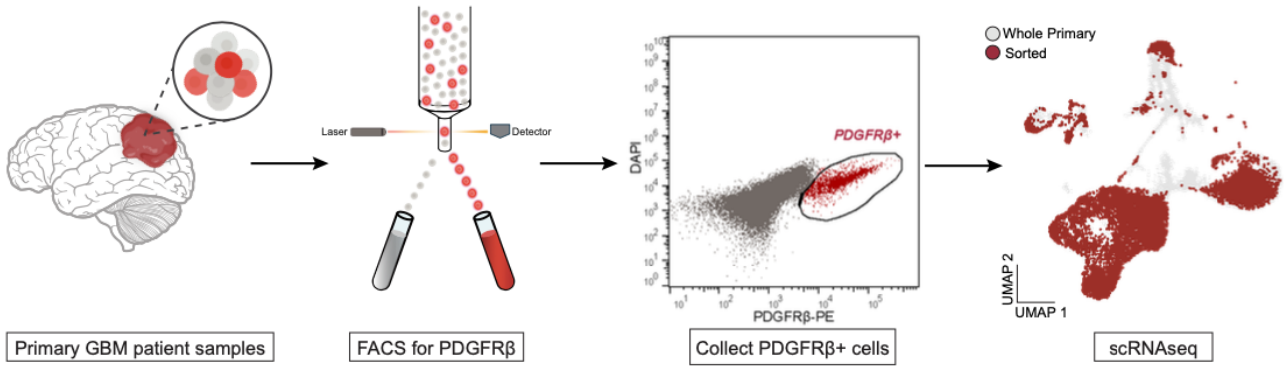
C) Module activity scores for meta-modules representing stemness programs were calculated for every cell in the meta-atlas. The heatmap depicts the expression profiles of vascular, mixed vascular, and NPC cell populations. The mixed vascular cell population exhibits a combination of stemness programs seen in both NPC and vascular cells. Expression of meta-module 11 uniquely distinguishes mixed vascular cells from vascular and NPC cells.

D) The distribution of meta-module 11 activity scores is plotted for individual tumors. Each ridge plot depicts the meta-module 11 scores of cells for one individual tumor, with tumors separated by dataset.

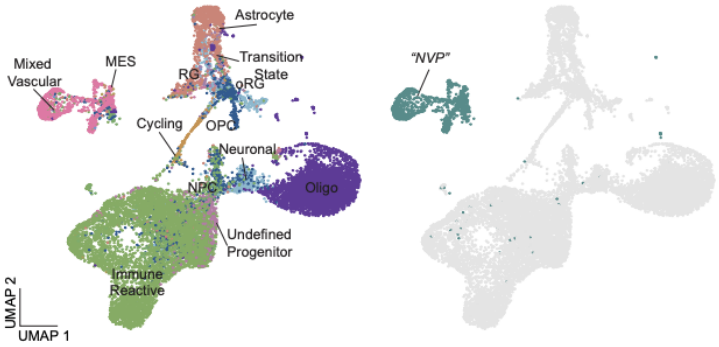
E) Immunostaining was performed to validate co-expression of vascular and progenitor markers in patient primary tumor tissue. Left panel: Tile scan of meta-module 11 members NOTCH3 (red) and PDGFRβ (magenta) co-localize with neural progenitor marker HOPX. Right panel: Tile scan of meta-module 11 member PDGFRβ (green) co-localized with neural progenitor marker NES (red) (co-localization denoted by arrows). Scale bar in large image = 400 μM, scale bar in insets = 200 μM (2x zoom). Staining was performed in total 3 tumors per panel. Representative images are shown.

Figure 3. Enrichment of PDGFRβ+ cells reveals primary GBM NVP cells represented across human and mouse tumors

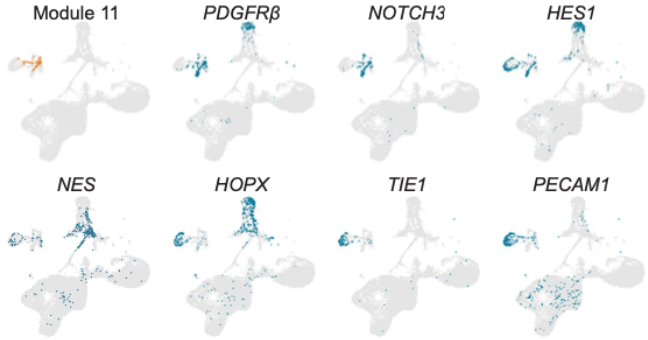
A. FACS and scRNA-seq enrichment strategy



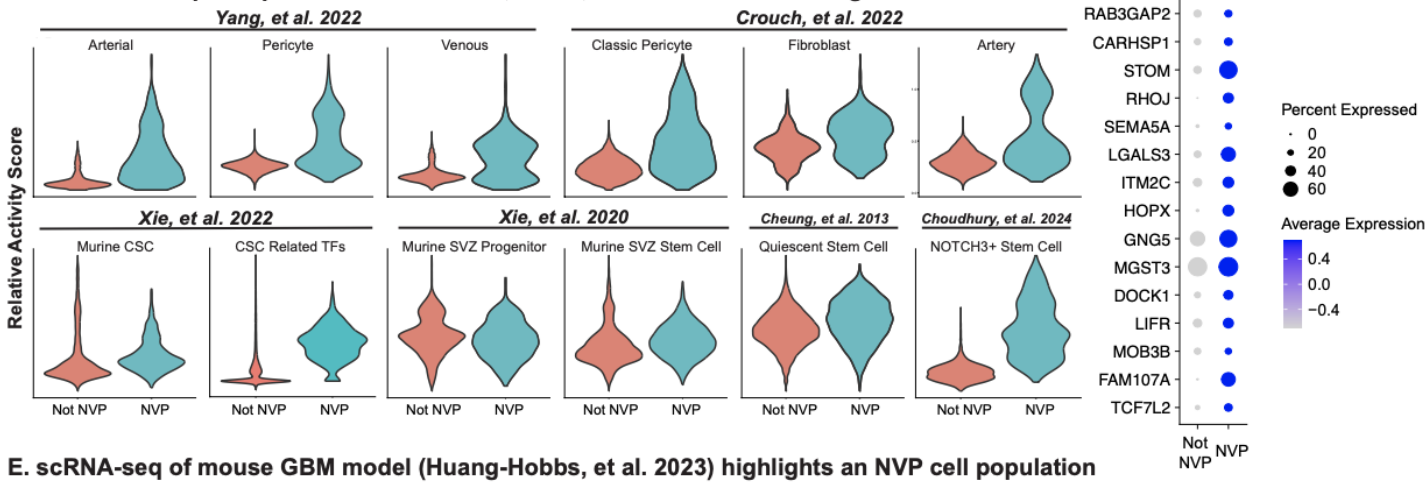
B. UMAP representations of scRNAseq clusters



C. Cell type marker expression



D. NVP cells express published endothelial, mural, and neural stem cell signatures



E. scRNA-seq of mouse GBM model (Huang-Hobbs, et al. 2023) highlights an NVP cell population

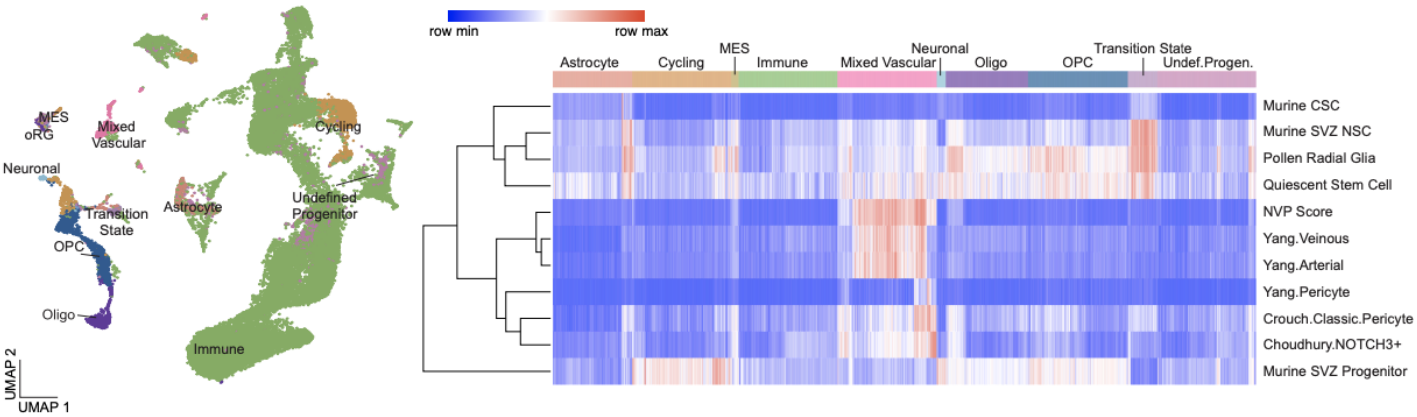


Figure 3. Enrichment of PDGFRβ+ cells reveals primary GBM NVP cells represented across human and mouse tumors

A) Fluorescence activated cell sorting (FACS) and single-cell RNA sequencing (scRNA-seq) enrichment strategy for primary GBM patient samples. Primary tumor samples were dissociated, and cells were labeled with PDGFRβ-PE and DAPI (live/dead). DAPI- PDGFRβ+ cells were collected via FACS. The FACS plot shows the separation of PDGFRβ+ (red) and PDGFRβ- (gray) cells. Sorted PDGFRβ+ cells were then captured with scRNA-seq along with the unsorted "Whole Primary" tumor. The UMAP displays the distribution of whole primary (gray) and sorted (red) cells in the scRNA-seq data.

B) Left: Cells from the Sorted and Whole Primary samples were annotated by via projection onto the annotated GBM meta-atlas[18]. Right: Based on cluster marker analysis and expression of vascular and neural progenitor genes (panel C), the cluster in turquoise was annotated as our putative Neurovascular Progenitor, or "NVP".

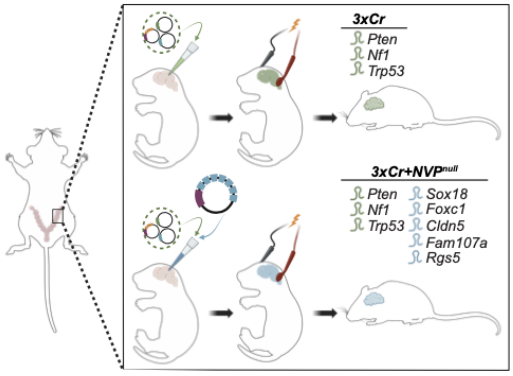
C) UMAP representations of meta-module 11 activity score and mural cell markers (*PDGFRβ, NOTCH3*), neural progenitor markers (*HES1, NES, HOPX*), and endothelial markers (*TIE1, PECAM1*) used to determine "NVP" annotation in panel B. Aforementioned marker genes and module activity coincide in the cluster ultimately annotated as "NVP."

D) Left: Gene modules were derived from published studies[51] or using in-house analysis of published scRNA-seq datasets. Gene modules included those for human adult vasculature[54], human developmental vasculature[53], mouse glioma cancer stem cell programs[62], mouse brain developmental types[55] and human stem cell types[52,56]. Violin plots compare the relative activity score of modules from each dataset between NVP and non-NVP cells. NVP cells are enriched for mural, endothelial, and neural stem cell program expression compared to non-NVP cells. Right: DotPlot depicting the enriched expression of outer radial glia (oRG) markers[46] in NVP cells.

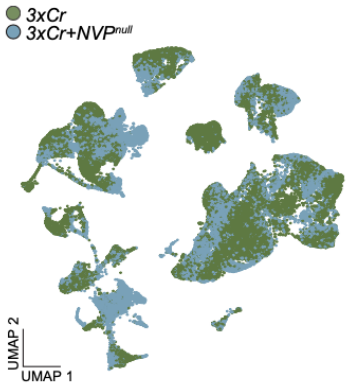
E) scRNA-seq data from a developmental model of mouse GBM[57] was processed according to our standard pipeline (Methods) and annotated via projection onto the GBM meta-atlas (Seurat, MapQuery[65]). Right: The dataset was scored for module activity for the previously described published modules, highlighting the expression of neural progenitor cell modules and vascular cell modules in the population annotated Mixed Vascular.

Figure 4. *In vivo* depletion of NVP decreases GBM cycling cells and extends mouse survival

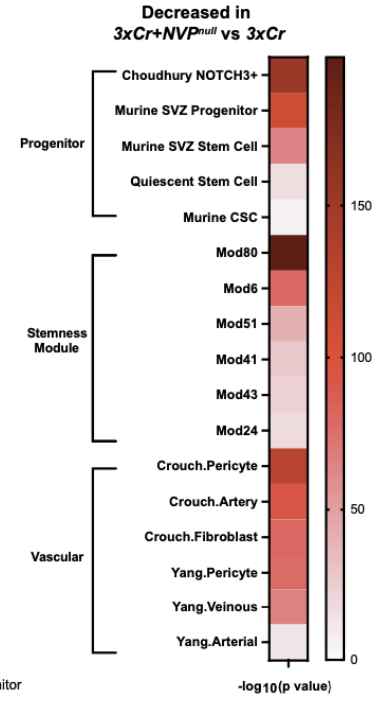
A. *In utero* electroporation model of mouse GBM ablates NVP cells



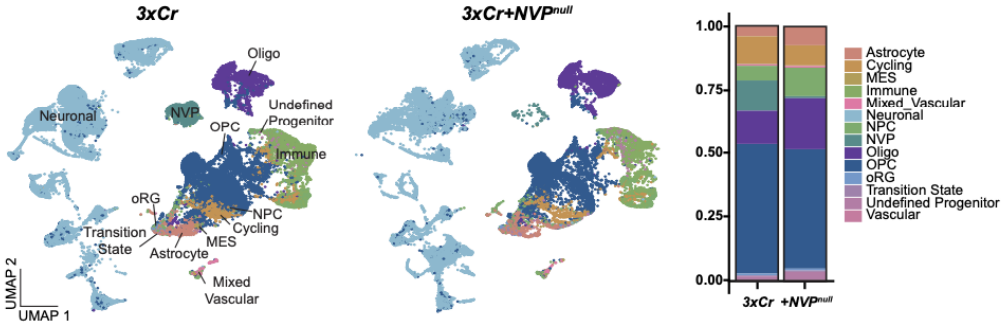
B. scRNAseq of *3xCr+NVP^{null}* and *3xCr* and show NVP depletion



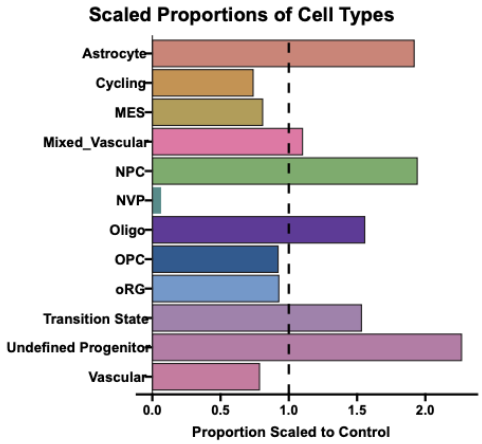
D. NVP module scores decrease in *3xCr+NVP^{null}* vs. *3xCr mice*



C. NVP depletion decreases changes GBM celltype representation



E. NVP depletion drives compensatory celltype changes



F. NVP depletion extends mouse survival

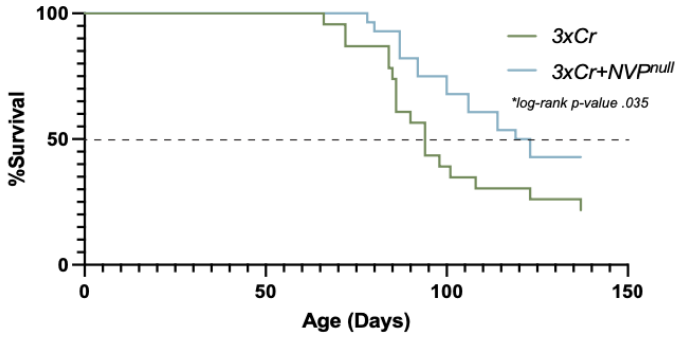


Figure 4. *In vivo* depletion of NVP decreases GBM cycling cells and extends mouse survival

A) We leveraged an *in utero* electroporation model of GBM to study the contribution of NVP cells to cell type composition and survival. The control group consisted of CRISPR-Cas9 mediated knockdown of *Pten, Nf1, Trp53* (*3xCr*). The experimental cohort received this *3xCr* plasmid with the addition of a plasmid which simultaneously expresses sgRNA sequences (*3xCr+NVP^{null}*) against five target genes (*Sox18*, *Foxc1, Fam107a, Cldn5, Rgs5*). Electroporation was performed at embryonic day 15 and tumors were harvested at postnatal day 70 then captured for single-nucleus RNA sequencing (snRNA-seq).

B) UMAP depicts the representation of *3xCr* and *3xCr+NVP^{null}* cells. snRNAseq data was processed using our standard QC metrics in addition to Cellbender to account for technical artifacts associated with ambient RNA[66].

C) Left: UMAP plots showing the impact of NVP depletion on GBM cell type representation. Cells were annotated by projection onto the GBM meta-atlas. “Neuronal” and “Immune” cells were excluded from downstream composition analysis, which shows the change in relative proportion of cell types from *3xCr* to *3xCr+NVP^{null}* mice. UMAP is split to label only *3xCr* cells (left) or *3xCr+NVP^{null}* cells (middle), allowing the visualization of reduced cluster representation in the cluster annotated as “NVP”. Right: difference in tumor cell type proportion between *3xCr* and *3xCr+NVP^{null}* mice.

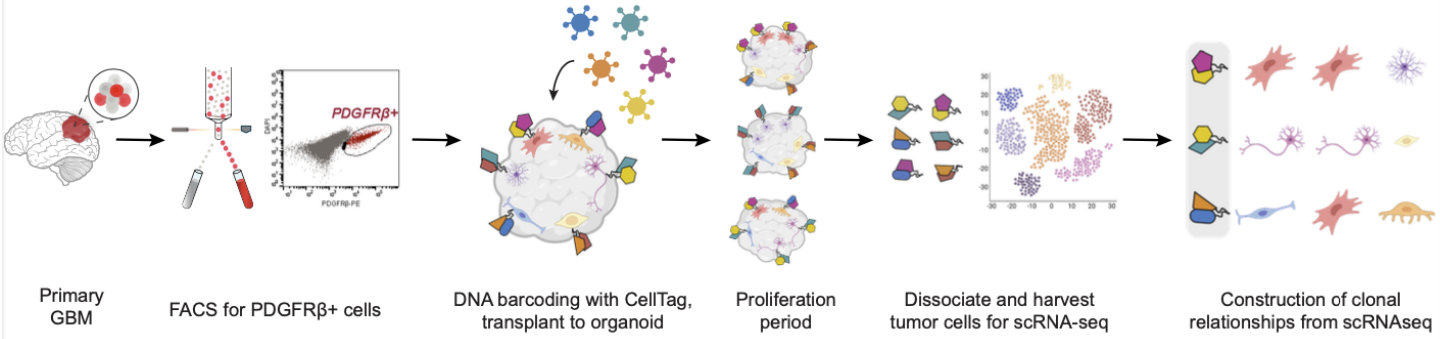
D) All cells in the dataset were scored for progenitor, stemness, and vascular modules (Fig. 2 and Fig 3). Log_2 fold change of module activity in *3xCr+NVP^{null}* cells compared to *3xCr* mice was calculated, and -log_{10} p-value is shown of the difference between the two conditions is shown. A t-test was performed for each variable to compare control and experimental conditions. The Benjamini-Hochberg procedure was applied to adjust the p-values for multiple comparisons, controlling the false discovery rate.

E) Differences in cell type proportion between *3xCr* and *3xCr+NVP^{null}* conditions were calculated by scaling the proportion of each cell type in the experimental condition relative to the control condition. The bar graph shows the scaled proportions of various cell types, with the dashed line indicating the control condition. Percent increases or decreases in cell type proportions are determined based on their relative position to the dashed line.

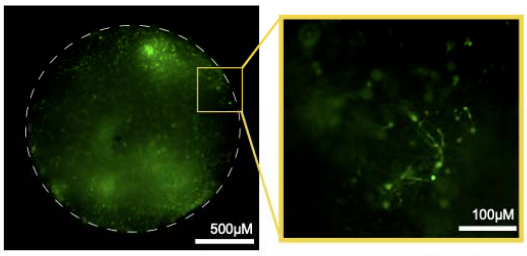
F) Kaplan-Meier curve showing survival of *3xCr* (n=23) and *3xCr+NVP^{null}* (n=28) mice. The *3xCr+NVP^{null}* cohort showed an increase in median overall survival (50% survival depicted by dashed line) compared to the *3xCr* cohort (119 vs. 91.5 days, *log-rank p-value = .03*).

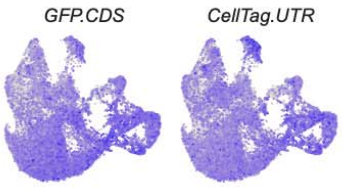
Figure 5. CellTag DNA barcoding of PDGFRβ+ cells from primary GBM demonstrates neuronal and vascular clonal relationships

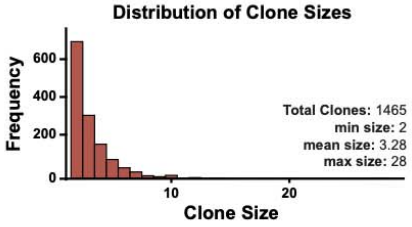
A. Clonal analysis of PDGFRβ+ cells from primary GBM samples



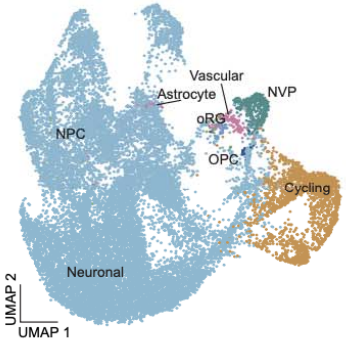
B. CellTag labeling of PDGFRB+ cells allows for clonal analysis

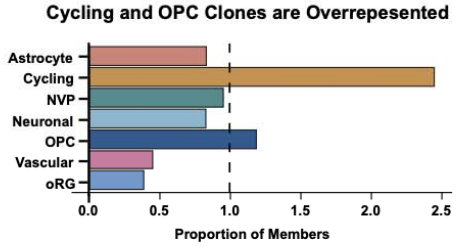




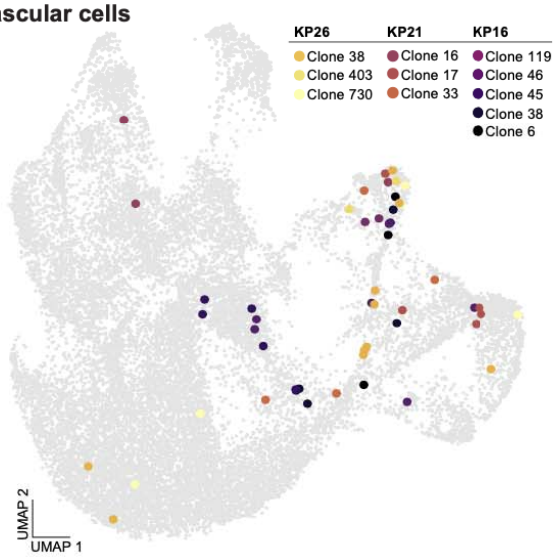


C. CellTagging clone analysis represents major tumor cell types



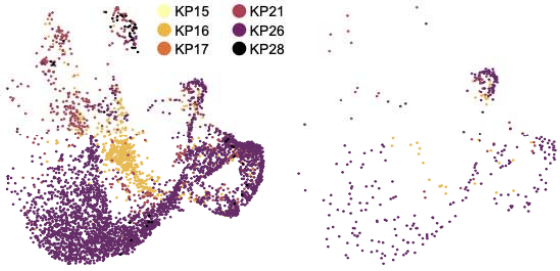


D. NVP clone partners include neuronal and vascular cells

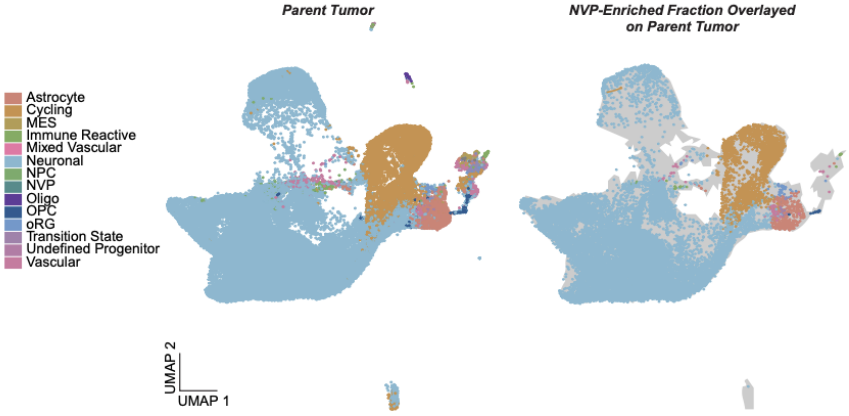


E. All Clone Members

F. NVP Clone Members



G. NVP generates a subset of parent tumor populations



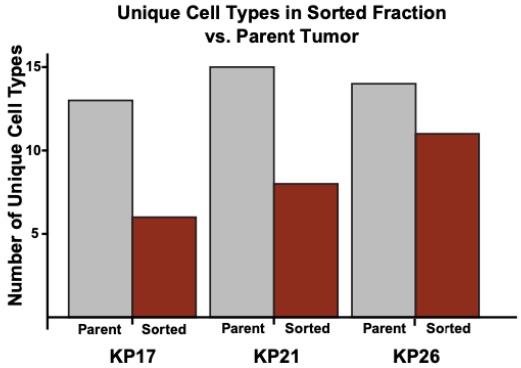


Figure 5. CellTag DNA barcoding of PDGFRβ+ cells from primary GBM demonstrates neuronal and vascular clonal relationships.

A) Six (6) primary GBM samples were subjected to FACS isolation of DAPI- PDGFRβ+ cells. These cells were then transduced with a lentiviral library containing plasmids which contain unique CellTag DNA barcodes and GFP. After transduction, cells were transplanted onto week 8-12 human cortical organoids. Following a proliferation period of 12-18 days, the tumor cells were dissociated from the organoid and sorted based on GFP. GFP+ cells are harvested for scRNA-seq. The resultant data include paired CellTag DNA barcode information with transcriptomic profiles, allowing the construction of clonal relationships, illustrating neuronal and vascular clonal dynamics.

B) Live-image of 3D cortical organoid with transplanted PDGFRβ+ CellTagged tumor cells in culture (4X magnification, scale bar = 500 μM). The inset depicts the cellular morphology of transplanted cells (20X magnification, scale bar = 100 μM). UMAPs depict the detection of CellTag associated genes GFP.CDS and CellTag.UTR, indicating that tumor cells were uniformly labeled by lentiviral transduction. Histogram displaying the distribution of clone sizes. Total clones: 1465, minimum size: 2, mean size: 3.28, maximum size: 28.

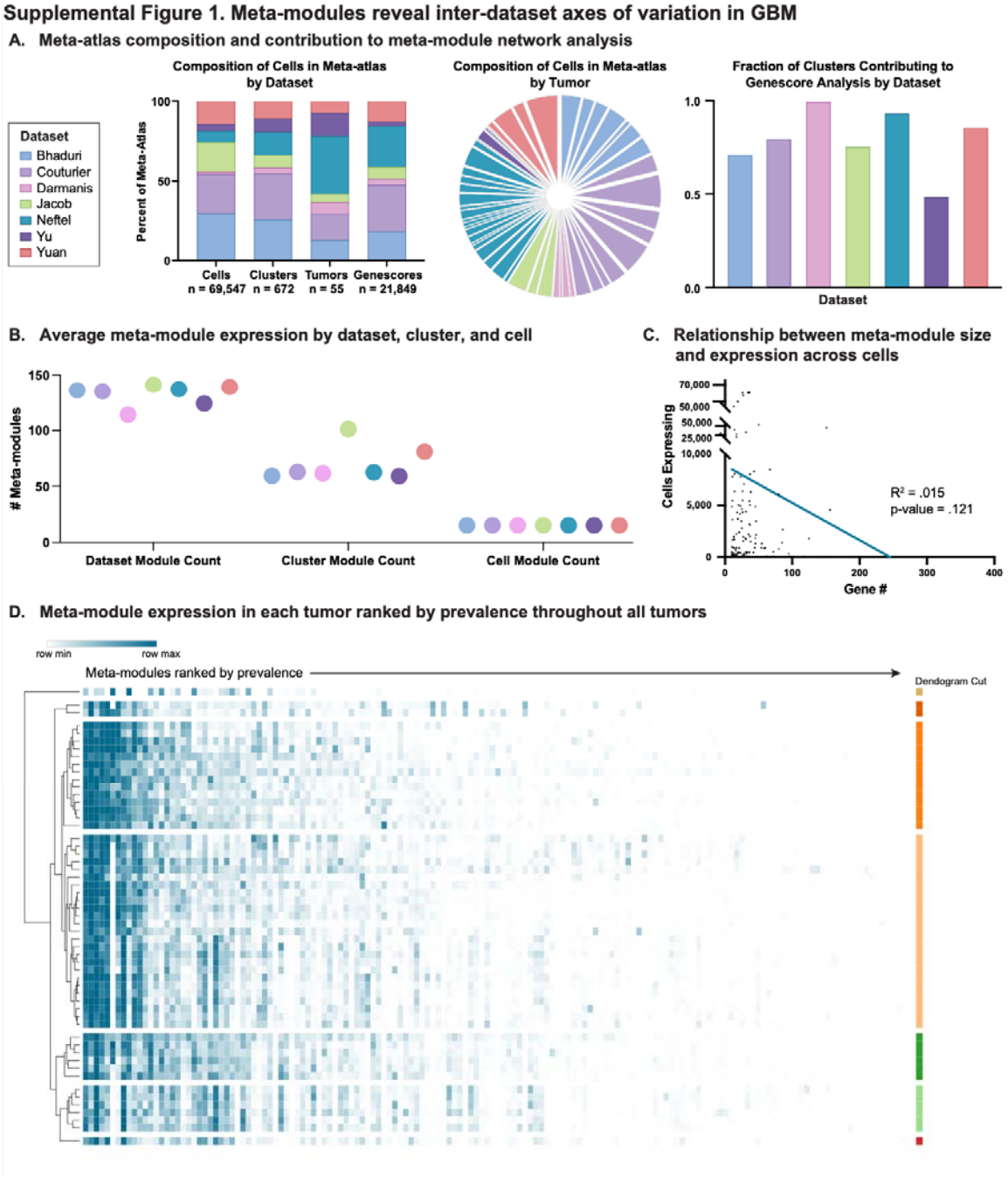
C) Cells were captured for scRNA-seq and subject to standard processing and QC (Methods). Cell identity was annotated via projection onto the GBM meta-atlas[65]. The bar graph depicts the likelihood of each cell type to be a member of a clone based on CellTag barcode analysis. Proportion of each cell type with a clone member was scaled to the proportion of cells in the entire dataset that are clone members.

D) Representative clones are depicted on the UMAP. Each dot represents one cell, and cells of the same color belong to the same clone. Clone members in NVP share clonal identity with cells of various cell types in the annotated UMAP, including vascular, neuronal, and other NVP cells.

E) UMAP highlighting cells that are members of a clone, colored by tumor. Clone members were distributed across cell types, with increased in the cycling cells cluster.

F) UMAP highlighting clones (n=90) with at least one NVP member, colored by tumor. NVP cells harbored clone partners in a diverse array of cell types, including other NVP cells, cycling cells, neurons, and vascular cells.

G) In parallel to PDGFRβ+ FACS sort and CellTag, in a subset of our tumors, we conducted a parallel infection, organoid transplantation, and scRNA-seq experiment. UMAP depiction of parent tumor scRNA-seq data is shown on the left. The right UMAP shows the PDGFRβ+ CellTagged cells overlayed on the UMAP space of the parent tumor, depicting that the majority of cell types could be derived from the sorted fraction. Notably, not all cell types are represented in the overlay. The bar graph depicts the number of unique cell types present in each tumor in either the parent sample or PDGFRβ+ CellTagged sample.



Supplemental Figure 1: Meta-modules reveal inter-dataset axes of variation in GBM

A) Overview of the meta-atlas composition by dataset and tumor, and the contribution to genescore analysis. Left: Stacked bar graphs showing the composition of cells (n = 69,547), clusters (n = 672), tumors (n = 55), and s (n = 21,849) in the meta-atlas, categorized by dataset. Middle panel: Pie chart depicting the composition of cells in the meta-atlas by tumor, colored by dataset. Right panel: Bar graph showing the fraction of clusters in each dataset contributing to analysis.

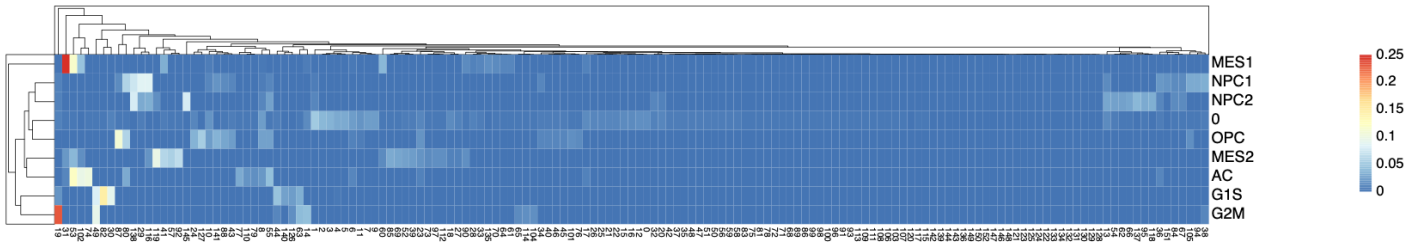
B) Number of meta-modules expressed by dataset, cluster, and cell show similarity in module number representation across datasets.

C) Relationship between size of the meta-module (gene #) and likelihood of cell expression depicts lack of correlation between module size and number of cells expressing the module.

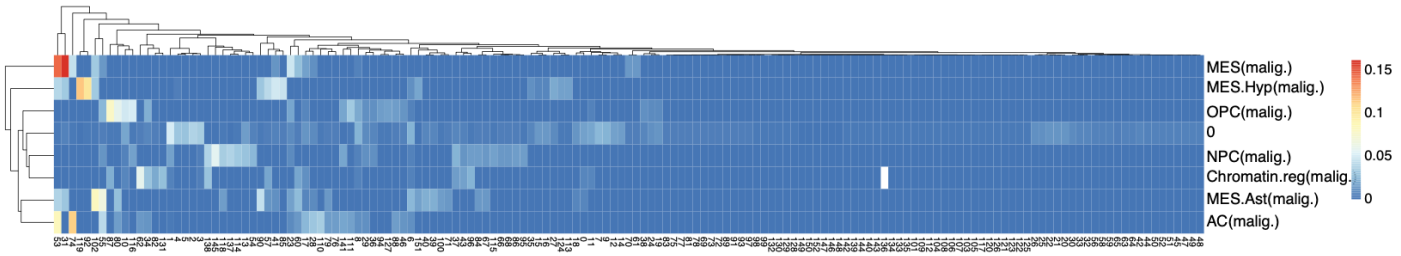
D) Heatmap displaying the prevalence of meta-modules across tumors in the meta-atlas. Each row represents a tumor, and each column represents a meta-module, ordered by overall prevalence. The color gradient indicates the expression level of each meta-module, with darker colors representing higher expression levels. The dendrogram on the left clusters tumors based on their meta-module expression profiles, and the color-coded bar on the right indicates the dendrogram cut. This illustrates the heterogeneity of meta-module expression across different tumors.

Supplemental Figure 2. Meta-modules recapitulate features of published GBM gene networks and contribute novel signatures

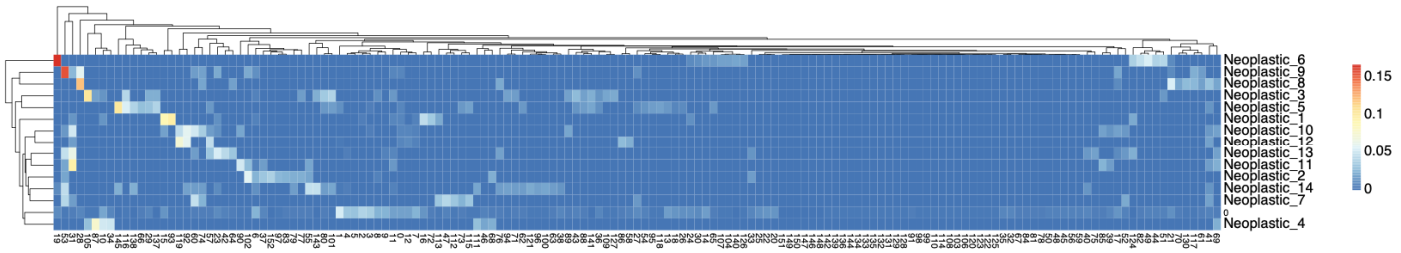
A. Jaccard Index Heatmap Between Meta-modules and Neftel, 2019



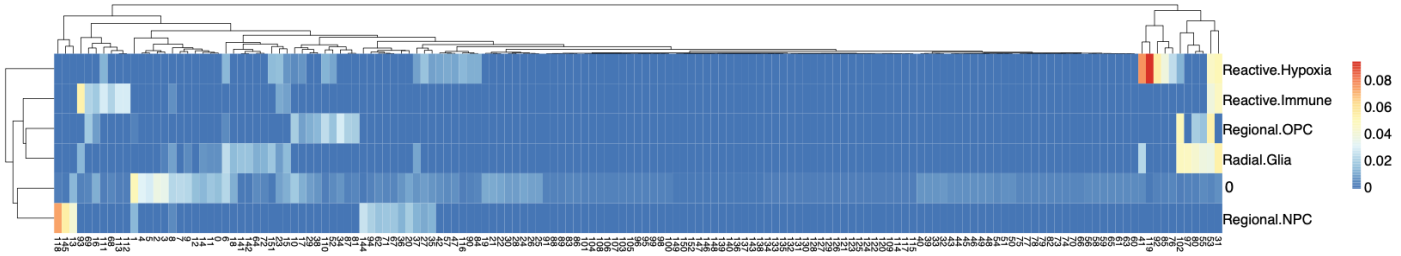
B. Jaccard Index Heatmap Between Meta-modules and Greenwald, 2024



C. Jaccard Index Heatmap Between Meta-modules and Ruiz-Moreno, 2022



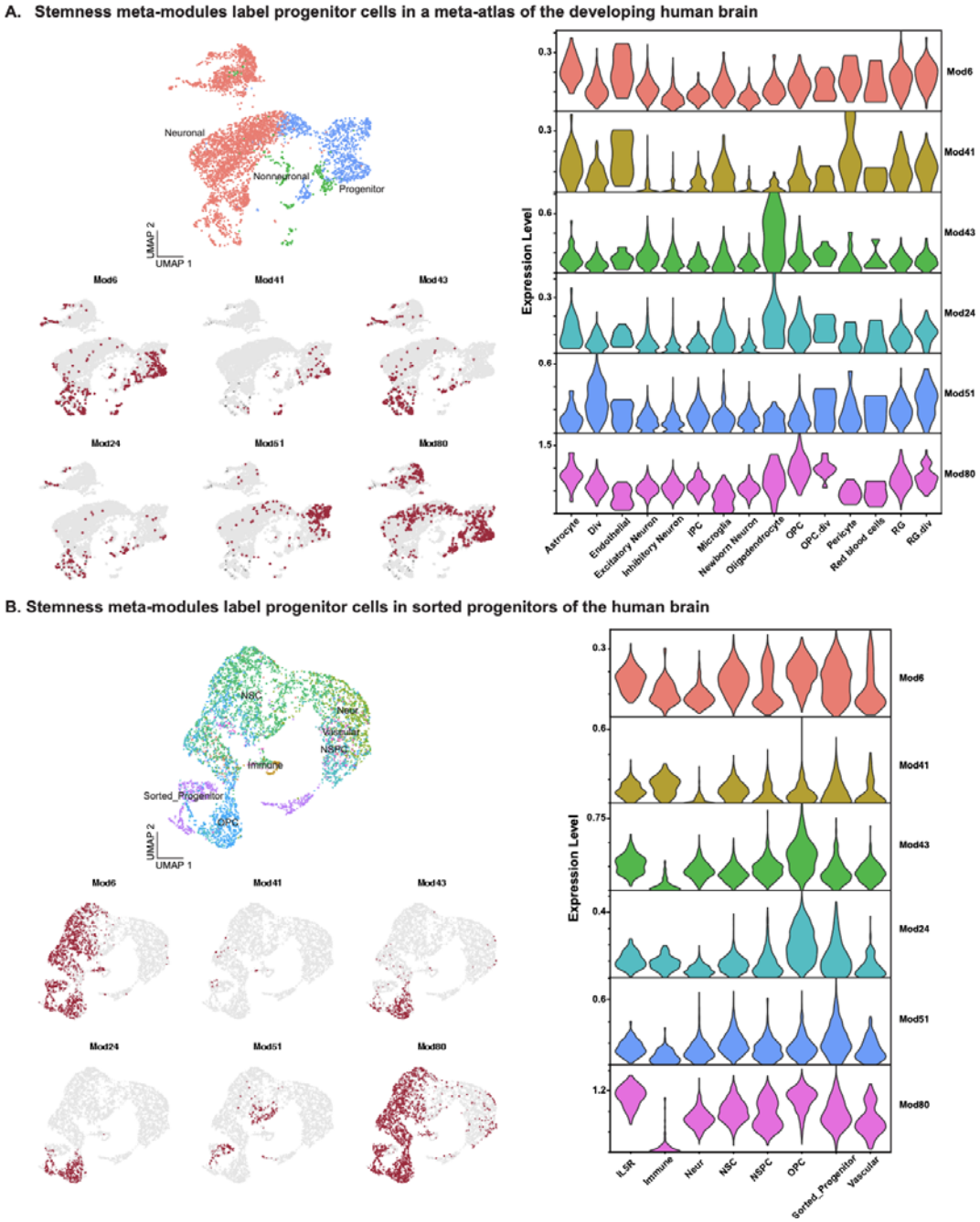
D. Jaccard Index Heatmap Between Meta-modules and Ravi, 2022



Supplemental Figure 2: Meta-modules recapitulate features of published GBM gene networks and contribute novel signatures

A-D) Jaccard index heatmaps showing the correlation between gene membership in the GBM meta-atlas meta-modules and modules from published datasets[18,29-31]. The y-axis includes module "0," representing genes not assigned to any published modules in that dataset. Color intensity indicates the strength of the Jaccard correlation, with darker colors representing higher correlations.

Supplemental Figure 3. Stemness meta-modules label annotated progenitor cell types in published normal neurodevelopmental datasets



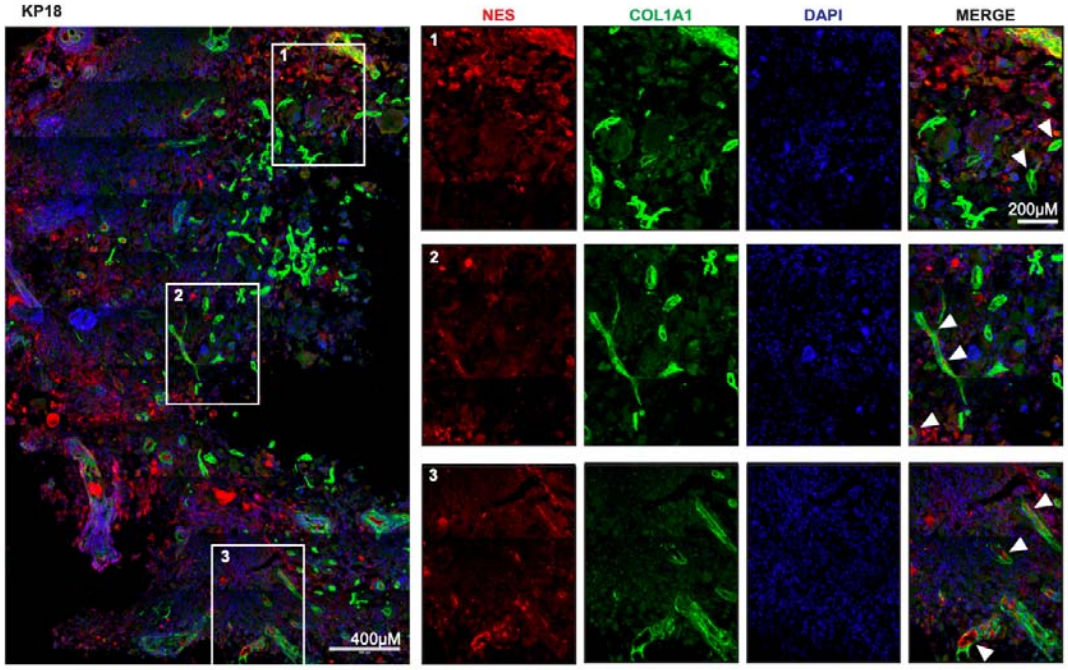
Supplemental Figure 3. Stemness meta-modules label annotated progenitor cell types in published normal neurodevelopmental datasets

A) UMAPs and violin plots showing the expression of stemness meta-modules from the GBM meta-atlas in a meta-atlas of the developing human brain[27]. Top UMAP depicts published annotation of cell types in the scRNA-seq dataset. Bottom UMAPs show module activity score of GBM meta-atlas stemness modules, depicting enrichment of stemness modules across developmental cell types. Violin plots of expression levels for each stemness meta-module across different cell types and are enriched in progenitor population (OPC, RG, RG.Div), indicating that stemness modules label progenitor cell types and neurons in the developing brain.

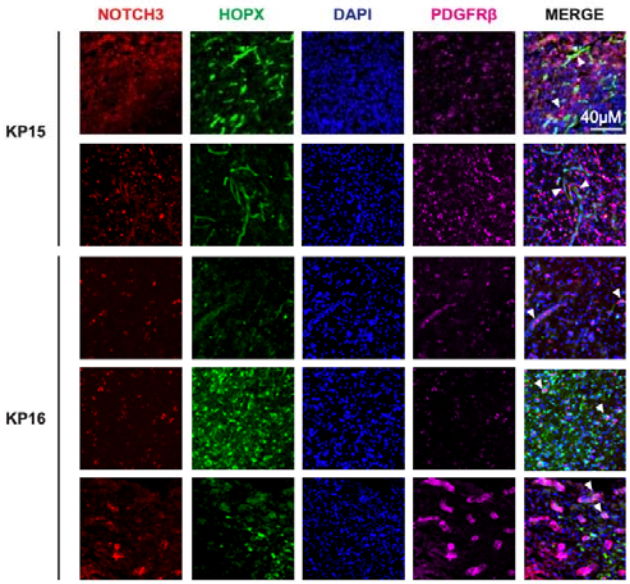
B) UMAPs and violin plots showing the expression of stemness meta-modules from the GBM meta-atlas in a dataset of sorted neural progenitor cell populations[34]. Top: cells from the dataset were annotated using the authors' annotations, with the exception that all of the sorted progenitor populations were grouped into one cell type. Bottom: Module activity for stemness meta-modules demonstrate representation across a variety of neural stem cell subtypes. Right: violin plots show stemness module representation in neural stem cell types, such as NSC, NSPC, OPC, and Sorted Progenitors.

Supplemental Figure 4. Co-expression of meta-module 11 genes with progenitor genes localize to vascular niches

A. Co-localization of meta-module 11 member COL1A1 with neural progenitor marker NES



B. NOTCH3, HOPX, PDGFRB co-localization in additional primary GBM samples

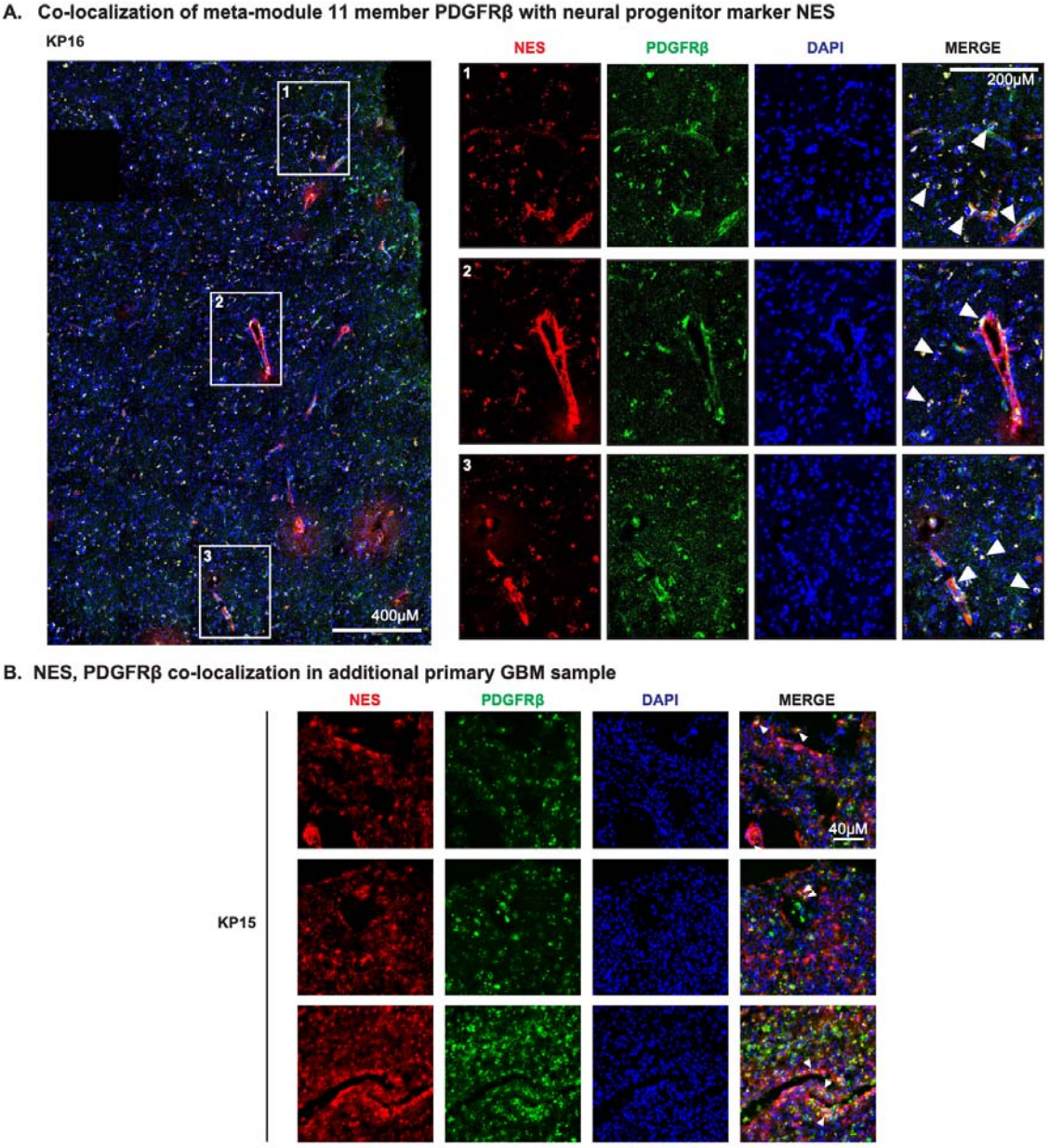


Supplemental Figure 4: Co-expression of meta-module 11 genes with progenitor genes localize to vascular niches

A) Immunostaining was performed to validate co-expression of additional and progenitor markers in patient primary tumor tissue. Left panel: Tile scan of meta-module 11 member and mural cell marker COL1A1 (green) and neural progenitor marker NES (red) (co-localization denoted by arrows). Co-staining appears to mostly localize to vascular morphologies while also existing in non-vascular associated tumor parenchyma. Scale bar in large image = 400 μM, scale bar in insets = 200 μM (1.7x zoom). Staining was performed in 3 tumors per panel. Representative images are shown.

B) Immunostaining in additional tumor samples demonstrates the co-localization of meta-module 11 members NOTCH3 (red) and PDGFRβ (magenta) with neural progenitor marker HOPX (green). Co-localization denoted by arrows. Staining was performed total 3 tumors per panel. Representative images are shown. Scale bar = 40 μM.

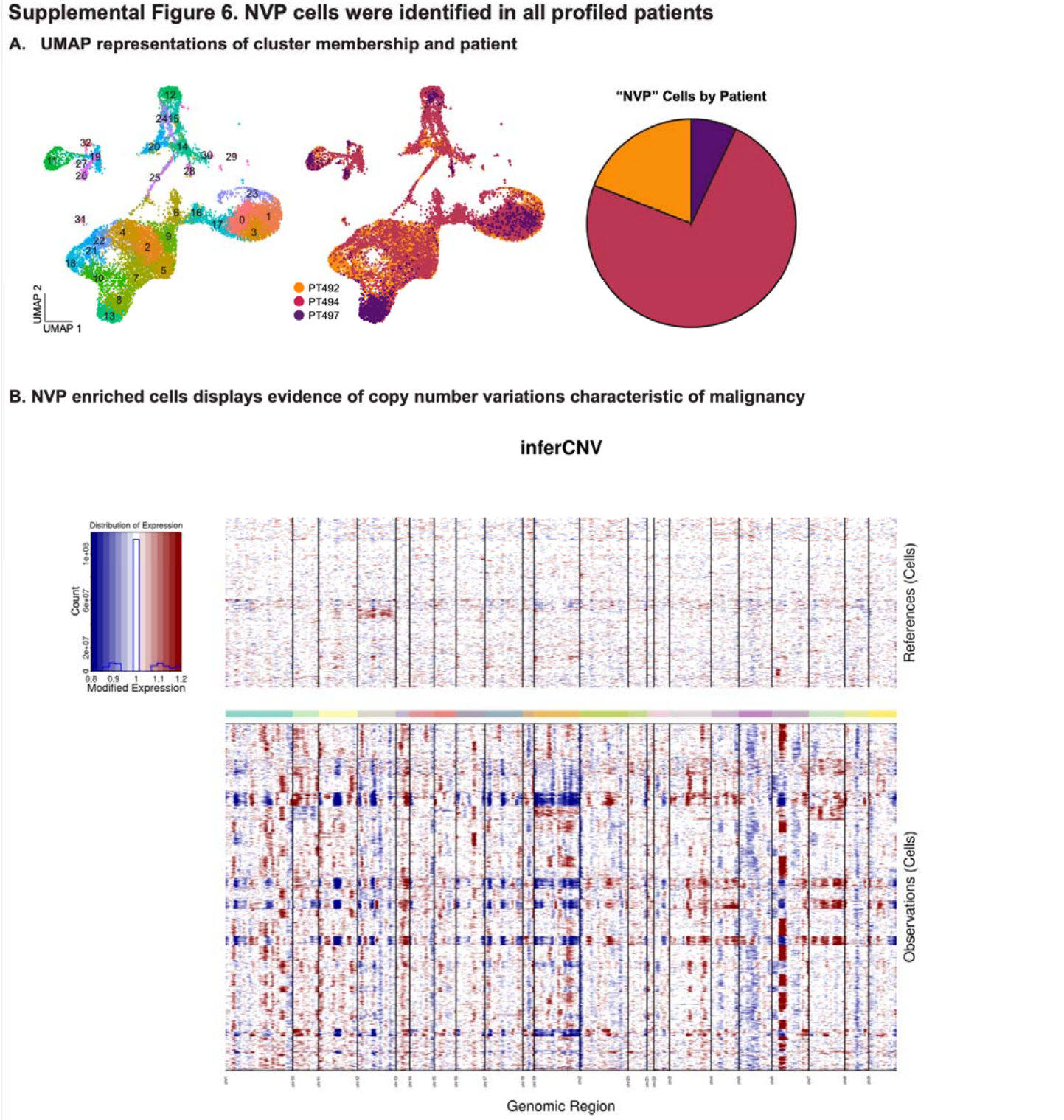
Supplemental Figure 5. Co-expression of meta-module 11 member PDGFRβ with neural progenitor marker NES in additional primary GBM samples



Supplemental Figure 5: Co-expression of meta-module 11 member PDGFRβ with neural progenitor marker NES in additional primary GBM samples

A) Immunostaining was performed to validate co-expression of vascular and progenitor markers in additional patient primary tumor tissue. Left panel: Tile scan of meta-module 11 member PDGFRβ (green) co-localized with neural progenitor marker NES (red) (co-localization denoted by arrows). Scale bar in large image = 400 μM, scale bar in insets = 200 μM (2x zoom).

B) Immunostaining was performed to validate co-expression of vascular and progenitor markers in additional patient primary tumor tissue. PDGFRβ (green) co-localized with neural progenitor marker NES (red) (co-localization denoted by arrows). Staining was performed in 3 tumors per panel. Representative images are shown. Scale bar = 40 μM.

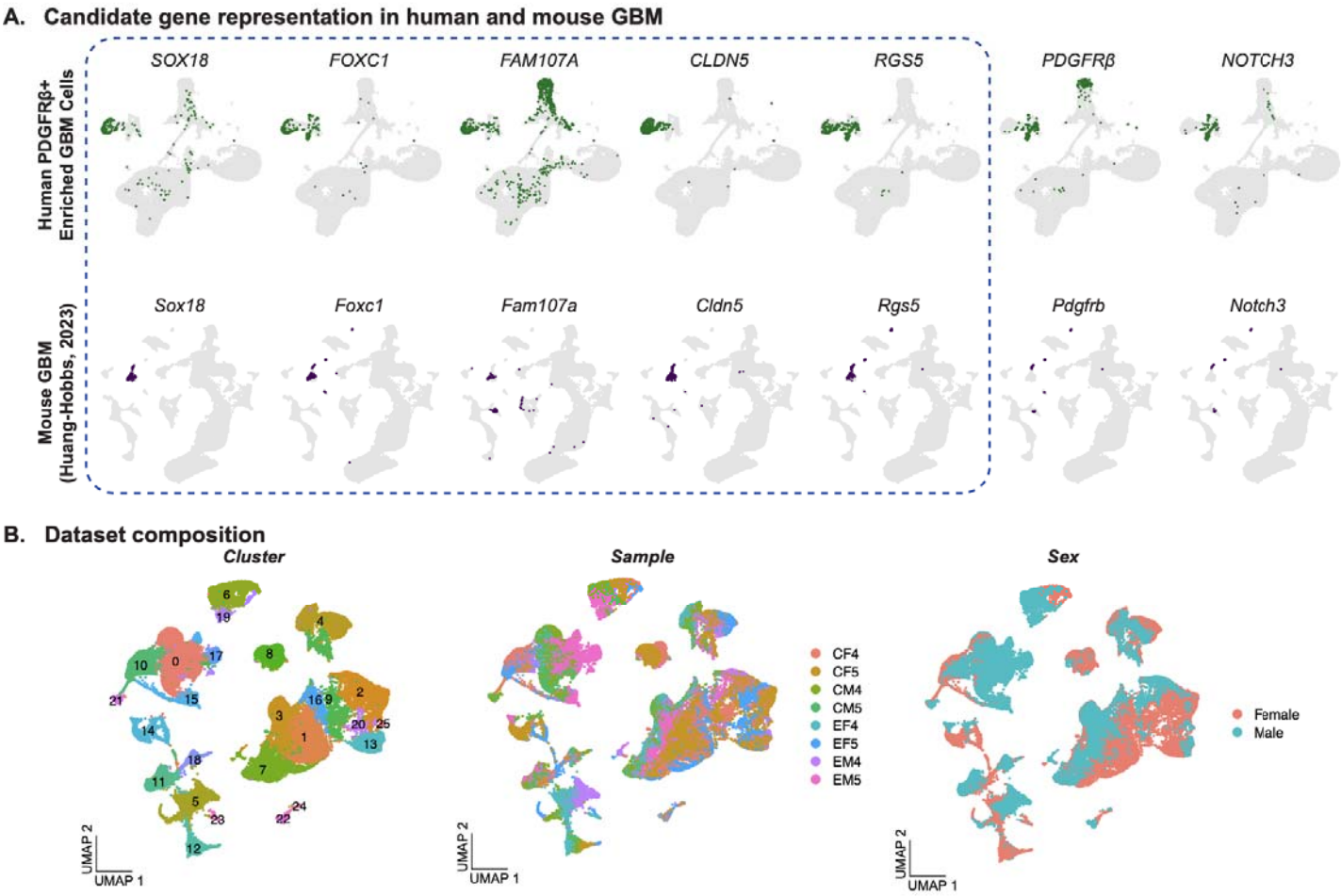


Supplemental Figure 6: NVP cells were identified in all profiled patients

A) UMAP representations of cluster membership and patient origin. Left: UMAP plot showing cluster membership from the Sorted and Whole Tumor samples. Right: UMAP plot displaying patient-specific clustering, with cells colored by patient (PT492, PT494, PT497).

B) InferCNV analysis of NVP-enriched cells displays evidence of copy number variations characteristic of malignancy. Heatmap showing the inferred copy number variations in reference cells (top) and observed cells (bottom), with variations indicated by red (gain) and blue (loss).

Supplemental Figure 7. Candidate selection in human and mouse GBM



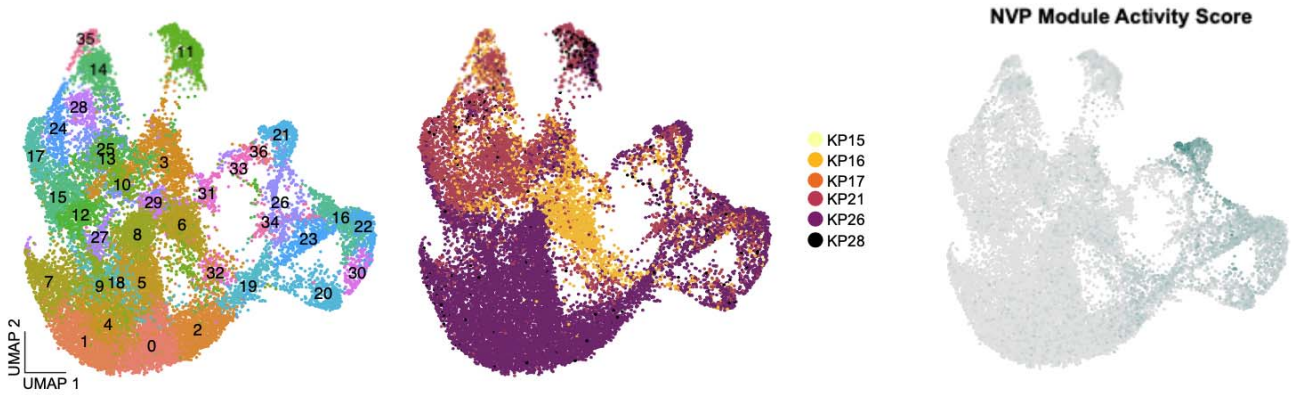
Supplemental Figure 7: Candidate selection in human and mouse GBM

A) Candidate gene representation in human and mouse GBM. UMAP plots show the expression of selected genes in human PDGFRβ+ enriched GBM cells (top row) and mouse GBM cells[57] (bottom row). Candidate genes were specific to the annotated "NVP" cluster in the human and "Mixed Vascular" cluster in the mouse models. CRISPR-Cas9 target genes are boxed in blue. *PDGFRβ* and *NOTCH3* and shown as examples of meta-module 11 gene members.

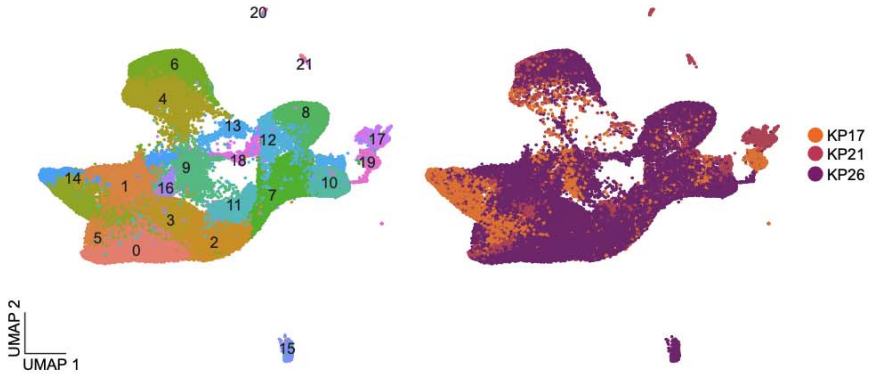
B) Dataset composition of snRNA-seq data from *3xCr* and *3xCr+NVP^{null}* cells. UMAP plots display the clustering of cells based on cluster identity (left), sample origin (middle), and sex (right).

Supplemental Figure 8. Clonal analysis of PDGFRβ+ human primary GBM cells

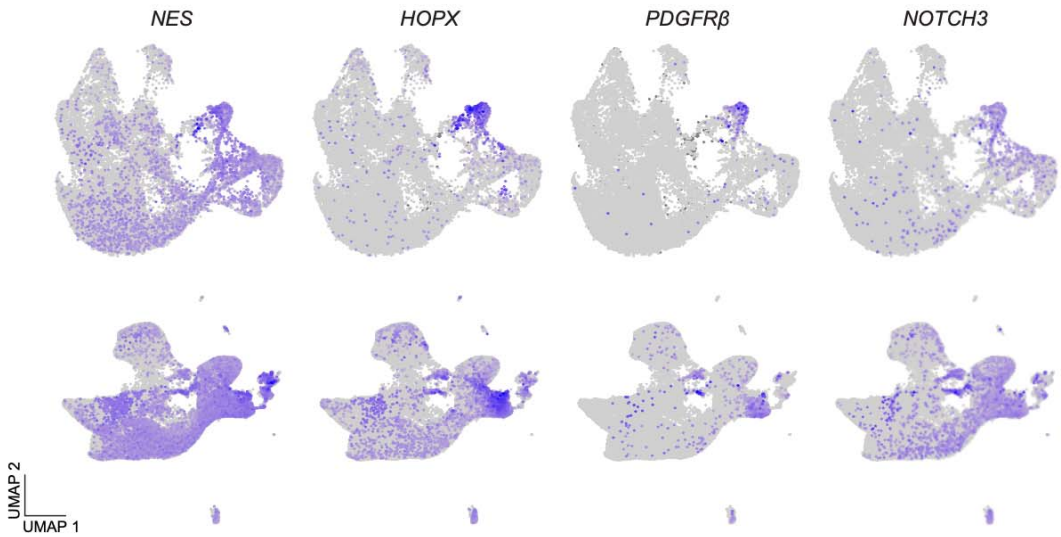
A. PDGFRβ+ cells from primary GBM samples labeled by cluster and patient



B. Matched parent tumor primary GBM samples labeled by cluster and patient



C. Cell type markers in PDGFRβ+ sorted samples and matched parent tumors



Supplemental Figure 8: Clonal analysis of PDGFRβ+ human primary GBM cells

A) UMAP plots showing CellTagged PDGFRβ+ cells after dissociation from cortical organoid labeled by cluster (left) or patient (n=6) (middle). Cells were scored for the NVP signature derived from enriched NVP cells (Fig 3).

B) UMAP plots of matched parent tumor primary GBM samples labeled by cluster (left) or patient (n=3, right).

C) Feature plots displaying the expression of cell type markers (NES, HOPX, PDGFRβ, NOTCH3) in PDGFRβ+ sorted (top row), organoid transplanted samples and matched, organoid transplanted parent tumors (bottom row).

Materials and Methods:

Meta-atlas dataset acquisition and uniform quality control

Gene count matrices for individual cells and corresponding metadata were obtained from various sources. For all datasets except those from Couturier and Bhaduri were downloaded from the Gene Expression Omnibus (GEO), and cells from the same individual were combined. All data not obtained from GEO were obtained through personal communication. While many of the datasets contained multiple tumor types, only direct-from-patient $IDH1^{WT}$ primary adult glioblastoma samples were incorporated into our analysis.

The 10X-derived count matrices and directories were processed using a standard pipeline to generate Seurat objects (Seurat version 4). Normalization of the counts was performed as needed, and cells with fewer than 500 detected genes and more than 10% of UMIs mapping to mitochondrial genes were filtered out. Genes detected in fewer than three cells were omitted. Datasets were subset to contain only "tumor" cells based on the paper's published annotations, or by cluster-based methods if published annotations were unavailable.

Generation of meta-modules from tumor cluster markers

For each individual tumor, we began by performing a self-correlation analysis on the normalized gene expression matrix to cluster cells based on their transcriptomic profiles. This procedure produced a correlation-based distance matrix, which was then hierarchically clustered to allocate cells into specific clusters. Most Seurat objects were clustered at the highest resolution (hclust deepSplit option set at 4). The resulting cluster assignments for each tumor were added as metadata into their respective Seurat objects. Conventional Seurat commands were subsequently employed to identify cluster markers. Genescores for these cluster markers were computed using the following equation:

Genescore of Gene A in Cluster 1 = [(% of cells in Cluster 1 expressing Gene A) / (% of cells not in Cluster 1 expressing Gene A)] * $\log_2$-fold-change of average Gene A expression between Cluster 1 cells versus all other cells in the dataset.

From the individual gene signatures, we derived gene modules exhibiting shared expression patterns across the meta-atlas. We compiled cluster markers from all tumors within the meta-atlas, retaining those in the 90th percentile of each dataset. The resulting table, containing genes, the meta-atlas clusters in which these genes were identified as markers, and their corresponding genescores, was used to create a distance matrix. Hierarchical clustering and division were performed using the dynamicTreeCut R Package with a minimum module size set to 10 genes. This strategy grouped genes based on their scores across all clusters in the meta-atlas, thereby forming meta-modules of genes with similar expression patterns across all 55 tumors represented in our meta-atlas.

Module activity score analysis

We evaluated meta-module activity within each cell by computing a module activity score based on the average expression level of each gene within a meta-module. Specifically, the activity score was determined by summing the normalized counts per million (CPM) for each meta-module gene and dividing this sum by the total number of genes in the meta-module to mitigate bias toward larger modules. Module activity scores for each module, for each cell were incorporated into Seurat objects as metadata, facilitating visualization using the FeaturePlot and DimPlot functions in Seurat. These scores proved to be highly versatile metrics, enabling the calculation of meta-module activity in datasets beyond our meta-atlas and allowing the measurement of activity across a wide range of gene networks.

Batch corrected integration for visualization and cell type annotation

For visualization purposes, we constructed the UMAP for our meta-atlas by integrating the specified datasets using conventional Seurat pipelines. Similarities between cells from different Seurat objects were identified and used as anchors to harmonize the data, employing the PCA method. FindVariableFeatures and ScaleData functions to prepare the data for principal component analysis (RunPCA function). Significant principal components were identified following the methods described by Shekhar et al., 2016[67].
Graph-based clustering was performed using the FindNeighbors and FindClusters functions, resulting in 23 clusters. Differential gene expression analysis (FindAllMarkers) was used to identify cluster markers, reporting

only genes positively enriched in a cluster. Annotation of cell identity based on integrated UMAP space was performed using a combination of annotation methods.

Immunofluorescent staining

Human primary GBM samples were obtained from patients undergoing surgery at Ronald Regan UCLA Medical Center. All patients signed informed consent forms for tissue collection under institutional IRB#10-000655. Samples were fixed in 4% paraformaldehyde for overnight at 4°C, rinsed with PBS, and equilibrated in 30% sucrose in PBS for 24-48 hours at 4°C. Post-equilibration, tumors were embedded in a 1:1 mixture of OCT and 30% sucrose, then frozen on dry ice. Frozen blocks were either stored at -80°C or processed on the cryostat into 10-16 µm-thick sections for immunofluorescence staining.

Sections were first rinsed with PBS for 15 minutes and subjected to antigen unmasking using a citrate-based solution (10 mM sodium citrate, pH 6) heated to 95°C for 20 minutes. Following this, sections were permeabilized and blocked with a buffer containing 5% donkey serum, 3% bovine serum albumin, and 0.1% Triton X-100 in PBS for 30 minutes at room temperature. Primary antibodies were incubated overnight at 4°C in the blocking buffer: rabbit anti-PDGFRβ (1:100), goat anti-PDGFRβ (1:100), mouse anti-Nestin (1:500), mouse anti-NOTCH3 (1:200), rabbit anti-COL1A1 (1:200), and rabbit anti-HOPX (1:500).

Following primary antibody incubation, sections were washed three times with PBS for 10 minutes each. Secondary antibody incubations, including DAPI (1:1000), were performed in blocking buffer for 2 hours at room temperature using the following secondary antibodies: AlexaFluor 555 donkey anti-goat (1:500), AlexaFluor 555 donkey anti-mouse (1:500), and AlexaFluor 488 donkey anti-rabbit (1:500).

Slides were then mounted with ProLong Gold antifade reagent and stored at 4°C for imaging using the ZEISS LSM 880 confocal system or EVOS M5000 digital-inverted benchtop microscope.

Tumor dissociation

Primary tumors were obtained from Ronald Reagan Hospital at UCLA, with patient consent and Institutional Review Board approval (IRB #10-000655 and #21-000108). Tumors were dissected into small fragments using a sterile scalpel and transferred to 5 mL microcentrifuge tubes containing 2.5 mL Papain and 125 µL DNase. Samples were incubated at 37°C for 45-60 minutes, with vigorous shaking by hand for 10 seconds every 5 minutes to facilitate dissociation. Post-incubation, the tissue was further dissociated by trituration and centrifuged at 300 x g for 5 minutes.

The cell suspension was then passed through a 40 µm filter to remove debris and counted. For further debris removal, cells were subjected to an ovomucoid density gradient according to the Papain Tissue Dissociation Kit instructions. Briefly, up to 20 million tumor cells were resuspended in a solution of ovomucoid inhibitor, DNase, and EBSS and carefully layered on top of an ovomucoid inhibitor cushion. The gradients were centrifuged at 70 x g for 6 minutes. The supernatant was discarded, and the tumor cells were collected, combined, and resuspended in media or FACS buffer for subsequent analyses.

Fluorescence-activated cell sorting

The cell suspension was incubated on ice for 30 minutes with Phycoerythrin (PE)-conjugated mouse anti-hPDGFRβ (1:100) and DAPI (1:2000). Following incubation, cells were washed and sorted using a Sony SH800, BD FACSAria, or Bio-Rad S3e cell sorter. All fluorescence-activated cell sorting (FACS) gates were set using unlabeled cells and single-color controls to ensure accurate gating and minimize background fluorescence. PDGFRβ+, DAPI- cells were collected into Sasai3 media.

Single cell capture and sequencing

Single cells, either sorted by FACS or derived from dissociated tumor cells, were captured using the 10X Genomics Chromium v3.1 3’ capture protocol on the 10X Chromium system. Targeting 10,000 cells for capture, all cells were used if fewer than 20,000 cells were retrieved post-sorting. The capture process was conducted according to the manufacturer’s instructions, and library preparation adhered to their guidelines. Sequencing was performed on an Illumina NovaSeq 6000 or NovaSeq X.

Single-cell analysis and quality control

Single-cell RNA sequencing (scRNA-seq) reads were aligned to a custom reference using a species-specific reference genome. For human samples, data was aligned to the GRCh38 human reference genome, incorporating sequences for the CellTag UTR and GFP.CDS. For mouse samples, data was aligned to the mm10 mouse reference genome, incorporating sequences for eGFP.

Cell-by-gene count matrices were generated using the 10X Genomics CellRanger pipeline with default parameters. Subsequent data analysis was performed using the Seurat R package in R Studio. Single-nucleus samples were run through CellBender[66], a package capable of removing background noise associated with ambient RNA in snRNA-seq data.

Cells expressing a minimum of 500 genes and exhibiting less than 10% mitochondrial gene content were retained. Genes expressed in at least three cells were also retained. Unique Molecular Identifier (UMI) counts were normalized via logarithmic transformation with a scaling factor of 10,000. Principal Component Analysis (PCA) was conducted on the scaled data using the top 2,000 variable genes. The number of dimensions (dims) was determined according to previously described methods[67]. Specifically, significant principal components (PCs) were selected based on the larger value between the square of the standard deviation of PCA scores (Seurat.Obj@reductions$pca@stdev^2) and the square root of the ratio of the number of genes to the number of cells plus one (sqrt(length(row.names(Seurat.Obj)) / length(colnames(Seurat.Obj))) + 1)^2.

Cells were clustered in the PCA space using Seurat's FindNeighbors and FindClusters functions with a resolution parameter set to 2.0. Visualization of cell clusters was performed using Uniform Manifold Approximation and Projection (UMAP) based on the previously defined dimensions. Doublets were predicted and excluded using the DoubletFinder R package[68] with default parameters.

Copy number variation analysis

To reduce the risk of false positives or negatives in tumor cell identification via FACS, copy number variation (CNV) analysis was conducted using InferCNV[69] with default settings. Batch-matched naïve organoids served as the reference for this analysis. Where appropriate, datasets were subset to contain tumor cells only.

Generation of modules from published datasets

Published gene lists were utilized for module activity score analysis when available. In cases where published gene lists were not available, gene modules were derived by performing differential gene expression analysis using the Seurat function FindAllMarkers, focusing exclusively on positively enriched genes based on the authors' published cell annotations. The resulting cluster markers were subsequently used to define the gene modules.

sgRNA candidate selection

Candidates for *in vivo* knockout were identified by intersecting human and mouse NVP signatures. The mouse signature was derived calculating cluster markers from clusters in the mouse single-cell dataset[57], annotated as "Mixed Vascular" following projection onto the GBM meta-atlas. These mouse cluster markers were intersected with those calculated from the human NVP-enriched cluster (Figure 3). Feature plots for all intersecting genes were generated, prioritizing genes specific to the putative NVP population in both mouse and human datasets. Additionally, genes with transcription factor functions were preferred due to their potential role in regulating critical gene programs within our clusters of interest. *Sox18* and *Foxc1* were selected based on these criteria. In addition to transcription factors, intersecting and specific genes that represent markers of mixed vascular and neural progenitor identities were also considered as candidates. *Fam107a* (outer radial glia), *Cldn5* (endothelial), and *Rgs5* (mural) were chosen for their representative roles in these cellular identities.

sgRNA design

For the design of single-guide RNAs (sgRNAs), we employed the CRISPR design tool available at CRISPOR[70] (http://crispor.tefor.net/). sgRNA sequences were designed against the mm10 genome, incorporating a 20bp-NGG sequence as the SpCas9 PAM site. Genomic sequences for the genes of interest were sourced from the NCBI database. Target sites were selected based on the following criteria: (i) the presence of a protospacer adjacent motif (PAM) sequence (NGG) required for SpCas9 binding, (ii) positioning within the coding region of

the gene, preferably targeting early exons to enhance the likelihood of generating loss-of-function alleles, and (iii) minimal off-target potential as predicted by the CRISPR design tool, prioritizing sgRNAs with high on-target activity scores and low off-target scores.

For genes with only a single exon (*Foxc1* and *Cldn5*), we designed two sgRNAs: one at the 5' end and one at the 3' end to maximize the chances of completely disrupting protein expression. For other targets, sgRNAs were designed against the second exon to achieve efficient gene disruption.

Cloning of an all-in-one CRISPR/Cas9 vector expressing multiple gRNAs

The cloning of an all-in-one CRISPR/Cas9 vector expressing multiple gRNAs was conducted using the Multiplex CRISPR/Cas9 Assembly system available from Addgene[71]. This approach utilizes a Golden Gate assembly system to sequentially assemble multiple inserts into a single plasmid. The final experimental plasmid contained sgRNAs targeting *Sox18*, the 5' end of *Foxc1*, the 3' end of *Foxc1*, the 5' end of *Cldn5*, the 3' end of *Cldn5*, *Fam107a*, and *Rgs5*.

***In utero* electroporation**

Mouse gliomas were generated using the CD1 IGS mouse background, following previously described protocols[57,72]. *In utero* electroporation was performed on embryonic day 15 with a plasmid cocktail containing constructs for CRISPR–Cas9-mediated knockout of tumor suppressor genes *Nf1*, *Trp53* (p53), and *Pten*, referred to as the triple CRISPR (*3xCr*) construct. The control group received a plasmid cocktail consisting of 1.5 µg/µl 3xCr, 2 µg/µl pGlast-PBase, and 0.25 µg/µl PBCAG-EGFP. The experimental group, termed *3xCr+NVPnull*, received an additional plasmid (also at 1.5 µg/µl) containing seven sgRNAs against NVP target genes. Mice were housed in a controlled environment with food and water available ad libitum, under a 12-hour light/dark cycle, at 20–22°C and 40–60% humidity. Both female and male mice were included in all experiments. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at Baylor College of Medicine and conform to the US Public Health Service Policy on Human Care and Use of Laboratory Animals.

Single-nucleus capture and sequencing of mouse tumors

At postnatal day 70 (P70), mice were euthanized, and tumors were dissected under a GFP fluorescent microscope. GFP-positive tumors were snap-frozen in liquid nitrogen for storage. Nuclei were isolated using the 10X Genomics Chromium Nuclei Isolation Kit with an 8-minute lysis time. Samples underwent two additional washing steps before proceeding to GEM (Gel Bead-in-Emulsion) capture, targeting the recovery of 10,000 cells. After standard QC, including Cellbender to remove technical artifacts associated with ambient RNA, data was processed using our Seurat pipeline and cells were annotated by projection onto the human GBM meta-atlas. "Neuronal" and "Immune" cells were excluded from downstream composition analysis, as these populations had low percentages of GFP positivity.

Stem cell culture and cortical organoid generation

The human embryonic stem cells (hESCs) UCLA6 was cultured as previously described[28,73,74]. Stem cells were maintained on Matrigel-coated 6-well plates in mTeSR Plus medium supplemented with 10% mTeSR Plus Supplement and 1X Penicillin/Streptomycin or Primocin. Media was changed every other day, and cells were passaged upon reaching >75% confluence.

For passaging, ReLeSR was applied at room temperature for 1 minute, aspirated, and cells were incubated at 37°C for 5 minutes. Cells were then dissociated into smaller clusters and replated at a 1:4 or 1:6 ratio on new Matrigel-coated plates. For cryopreservation, the passaging procedure was followed until the final step, where cells were resuspended in 1 ml of mFreSR per well. The cell suspension was transferred to cryovials and stored at -80°C for 24-48 hours before transfer to liquid nitrogen for long-term storage.

Cortical organoid generation

Cortical organoids were generated following an adapted protocol from Kadoshima et al[75], in alignment with other studies[28,73,74]. Briefly, human embryonic stem cells (hESCs) at >75% confluence in 6-well plates were treated with 1 mL Accutase per well and incubated at 37°C for 5 minutes. Cells were then washed with 1 mL of Sasai1, composed of GMEM, 20% KnockOut Serum, 0.1 mM β-mercaptoethanol, 1X Non-Essential Amino Acids (NEAA), 1X Sodium Pyruvate, and 1X Penicillin/Streptomycin or Primocin.

Cells were detached by scraping and collected in 15 mL tubes, followed by centrifugation at 300 x g for 5 minutes. The cell pellets were resuspended in Sasai1 supplemented with 20 µM Y27632 (Rock inhibitor), 5 µM SB431542 (TGF-β inhibitor), and 3 µM IWR1-endo (Wnt signaling inhibitor). A total of 1 million cells in 10 mL of Sasai1 with the small molecules were seeded into 96-well V-shaped low attachment plates to aggregate over 72 hours.
On Day 3, 50 µL of medium was replaced with 100 µL of fresh Sasai1 containing the small molecules. Media changes were performed every other day until Day 7, when the Rock inhibitor was omitted. On Day 18, organoids were transferred to ultra low attachment 6-well plates with media changes every other day. From Day 18 to Day 35, Sasai 2, consisting of DMEM/F-12 with Glutamax, 1X N-2 supplement, 1X Lipid Concentrate, and 1X Penicillin/Streptomycin or Primocin, was used.

From Day 35 onward, Sasai3 was applied, containing DMEM/F-12 with Glutamax, 1X N-2 supplement, 1X Lipid Concentrate, 1X Penicillin/Streptomycin or Primocin, 10% Fetal Bovine Serum, 5 µg/mL Heparin, and 0.5% Growth factor-reduced Matrigel. Throughout the culture period, live images were periodically taken to monitor organoid growth, and immunostaining was performed at Weeks 5 and 8 to confirm neuronal differentiation.

CellTag virus
Lentivirus containing the CellTag-V1 barcode library was either directly purchased from Addgene or manufactured by a commercial vendor (Vectorbuilder) using the DNA plasmid library purchased from Addgene. For commercial generation of CellTag virus, the libraries were sequenced using MiSeq and CellTags were filtered in order to create a whitelist based on the standard CellTag workflow. The Vectorbuilder lentiviral library displayed a complexity of 7,989 barcodes, while the Addgene lentiviral library had a complexity of 19.974 barcodes. Appropriate whitelists were applied to samples during data processing depending on which lentivirus was used in the experiment.

Transduction with CellTag and transplantation onto cortical organoids
Immediately following FACS, PDGFRβ+ cells were resuspended in Sasai3 medium. CellTag virus and polybrene (1:1000) were added to the cells, which were then incubated at 37°C for 60 minutes with gentle rotation. Amount of CellTag virus was based on a target MOI of 4 using this formula: TU_{total} = (MOI x Cell Number)/Viral titer (TU/µl).

Post-transduction, cells were washed three times with warm PBS and Sasai3 medium. The transduced tumor cells were resuspended in 20 µL of Sasai3 medium and transplanted onto cortical organoids using the hanging drop method.

For the hanging drop method, 8-12-week-old human cortical organoids were transferred to the lid of a 10cm dish using wide-bore 1000 µL tips. Excess media was removed, and 10-15 µL of the tumor single-cell suspension was added atop each organoid. The lid was then carefully inverted onto a 10 cm dish containing 10 mL of base culture medium to prevent evaporation during the hanging drop stage. These hanging drop co-cultures were maintained at 37°C for 12-16 hours, after which they were transferred to ultra-low attachment 6-well plates with Sasai3. Tumor cells typically surrounded the organoids and began migrating inward after a few days. These co-cultures were maintained for 12 – 18 days, with media changes performed three times per week, before being harvested for analysis.

Organoid-tumor transplant dissociation and collection of GFP+ cells by FACS
Tumor-transplanted organoids were transferred to a 1.7 mL Eppendorf tube containing 1 mL Papain and 50 µL DNase, and incubated at 37°C for 45-60 minutes. During the initial 10 minutes of incubation, the tube was shaken vigorously by hand for 10 seconds to aid dissociation, a process repeated every 5 minutes. Following incubation, the material was further dissociated by trituration and centrifuged at 300 x g for 5 minutes. The resulting pellet was resuspended in cold FACS buffer, filtered through a 40 µm mesh, and stained with DAPI to identify non-viable cells. DAPI-/GFP+ tumor cells were sorted and collected for downstream single-cell capture. Flow cytometry and cell sorting were performed using a Bio-Rad S3e. Gates were set to remove debris, eliminate doublets, and exclude dead cells using DAPI staining. Gates for GFP positivity were calibrated using control, batch-matched organoids without transplantation.

CellTag barcode analysis

scRNA-seq data was first subject to quality control as described above. Only quality-controlled cells were inputted into the CellTag barcode analysis. Barcode analysis was performed using the CellTagR package to support clone calling. First, BAM files were filtered using samtools to include only unmapped reads. A CellTag object was then created, initialized with the BAM file. To enhance the accuracy of clone identification, the starcode algorithm[76] was employed to cluster similar barcodes. This clustering corrects sequencing errors and allows for the more accurate identification of true clonal relationships by grouping barcodes based on sequence similarity. In order to investigate clones at different levels of stringency, we repeated the CellTag analysis with varying levels of strictness. The highest level of rigor involved inputting cells into the clone calling analysis which contained a minimum of two counts of at least two unique CellTag barcodes. High confidence clones were defined as clones identified by a minimum of two overlapping unique CellTag barcodes per cell. The next threshold involved inputting cells which contained a minimum of two counts of at least one unique CellTag barcode. Clones at this level were defined by cells that had an overlap of at least one unique barcode. For the analyses presented here, this distinction was not used.

Bibliography

1 Stupp, R. *et al.* Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial. *JAMA* **318**, 2306-2316, doi:10.1001/jama.2017.18718 (2017).

2 Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987-996, doi:10.1056/NEJMoa043330 (2005).

3 Stupp, R., van den Bent, M. J. & Hegi, M. E. Optimal role of temozolomide in the treatment of malignant gliomas. *Curr Neurol Neurosci Rep* **5**, 198-206, doi:10.1007/s11910-005-0047-7 (2005).

4 Bunse, L., Bunse, T., Kramer, C., Chih, Y. C. & Platten, M. Clinical and Translational Advances in Glioma Immunotherapy. *Neurotherapeutics* **19**, 1799-1817, doi:10.1007/s13311-022-01313-9 (2022).

5 Lim, M., Xia, Y., Bettegowda, C. & Weller, M. Current state of immunotherapy for glioblastoma. *Nat Rev Clin Oncol* **15**, 422-442, doi:10.1038/s41571-018-0003-5 (2018).

6 Lin, B. *et al.* EGFR, the Lazarus target for precision oncology in glioblastoma. *Neuro Oncol* **24**, 2035-2062, doi:10.1093/neuonc/noac204 (2022).

7 Tilak, M., Holborn, J., New, L. A., Lalonde, J. & Jones, N. Receptor Tyrosine Kinase Signaling and Targeting in Glioblastoma Multiforme. *Int J Mol Sci* **22**, doi:10.3390/ijms22041831 (2021).

8 Liu, R. *et al.* Glioblastoma recurrence correlates with NLGN3 levels. *Cancer Med* **7**, 2848-2859, doi:10.1002/cam4.1538 (2018).

9 Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573**, 532-538, doi:10.1038/s41586-019-1564-x (2019).

10 Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539-545, doi:10.1038/s41586-019-1563-y (2019).

11 Venkatesh, H. S. *et al.* Targeting neuronal activity-regulated neuroligin-3 dependency in high-grade glioma. *Nature* **549**, 533-537, doi:10.1038/nature24014 (2017).

12 Bhaduri, A. *et al.* Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* **26**, 48-63 e46, doi:10.1016/j.stem.2019.11.015 (2020).

13 Chen, X. *et al.* Single-cell RNA sequencing reveals intra-tumoral heterogeneity of glioblastoma and a pro-tumor subset of tumor-associated macrophages characterized by EZH2 overexpression. *Biochim Biophys Acta Mol Basis Dis* **1868**, 166534, doi:10.1016/j.bbadis.2022.166534 (2022).

14 Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat Commun* **11**, 3406, doi:10.1038/s41467-020-17186-5 (2020).

15 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, doi:10.1016/j.celrep.2017.10.030 (2017).

16 Jacob, F. *et al.* A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell* **180**, 188-204 e122, doi:10.1016/j.cell.2019.11.036 (2020).

17 Johnson, K. C. *et al.* Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. *Nat Genet* **53**, 1456-1468, doi:10.1038/s41588-021-00926-8 (2021).

18 Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849 e821, doi:10.1016/j.cell.2019.06.024 (2019).

19 Richards, L. M. *et al.* Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* **2**, 157-173, doi:10.1038/s43018-020-00154-9 (2021).

20 Wang, L. *et al.* The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. *Cancer Discov* **9**, 1708-1719, doi:10.1158/2159-8290.CD-19-0329 (2019).

21 Wang, L. *et al.* A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. *Nat Cancer* **3**, 1534-1552, doi:10.1038/s43018-022-00475-x (2022).

22 Wang, R. *et al.* Adult Human Glioblastomas Harbor Radial Glia-like Cells. *Stem Cell Reports* **14**, 338-350, doi:10.1016/j.stemcr.2020.01.007 (2020).

23 Wu, B. *et al.* Single-Cell Sequencing of Glioblastoma Reveals Central Nervous System Susceptibility to SARS-CoV-2. *Front Oncol* **10**, 566599, doi:10.3389/fonc.2020.566599 (2020).

24 Yu, K. *et al.* Surveying brain tumor heterogeneity by single-cell RNA-sequencing of multi-sector biopsies. *Natl Sci Rev* **7**, 1306-1318, doi:10.1093/nsr/nwaa099 (2020).

25 Yuan, J. *et al.* Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med* **10**, 57, doi:10.1186/s13073-018-0567-9 (2018).

26 Zhao, W. *et al.* Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. *Genome Med* **13**, 82, doi:10.1186/s13073-021-00894-y (2021).

27 Nano, P. R. *et al.* A Meta-Atlas of the Developing Human Cortex Identifies Modules Driving Cell Subtype Specification. *bioRxiv*, doi:10.1101/2023.09.12.557406 (2023).

28 Bhaduri, A. *et al.* Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142-148, doi:10.1038/s41586-020-1962-0 (2020).

29 Greenwald, A. C. *et al.* Integrative spatial analysis reveals a multi-layered organization of glioblastoma. *Cell* **187**, 2485-2501 e2426, doi:10.1016/j.cell.2024.03.029 (2024).

30 Ruiz-Moreno, C. *et al.* Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma. *bioRxiv*, 2022.2008.2027.505439, doi:10.1101/2022.08.27.505439 (2022).

31 Ravi, V. M. *et al.* Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. *Cancer Cell* **40**, 639-655 e613, doi:10.1016/j.ccell.2022.05.009 (2022).

32 Bao, S. *et al.* Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756-760, doi:10.1038/nature05236 (2006).

33 Berezovsky, A. D. *et al.* Sox2 promotes malignancy in glioblastoma by regulating plasticity and astrocytic differentiation. *Neoplasia* **16**, 193-206, 206 e119-125, doi:10.1016/j.neo.2014.03.006 (2014).

34 Liu, D. D. *et al.* Purification and characterization of human neural stem and progenitor cells. *Cell* **186**, 1179-1194 e1115, doi:10.1016/j.cell.2023.02.017 (2023).

35 Walchli, T. *et al.* Single-cell atlas of the human brain vasculature across development, adulthood and disease. *Nature*, doi:10.1038/s41586-024-07493-y (2024).

36 Cheng, L. *et al.* Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. *Cell* **153**, 139-152, doi:10.1016/j.cell.2013.02.021 (2013).

37 Jain, S. *et al.* Single-cell RNA sequencing and spatial transcriptomics reveal cancer-associated fibroblasts in glioblastoma with protumoral effects. *J Clin Invest* **133**, doi:10.1172/JCI147087 (2023).

38 Hu, Y. *et al.* Neural network learning defines glioblastoma features to be of neural crest perivascular or radial glia lineages. *Sci Adv* **8**, eabm6340, doi:10.1126/sciadv.abm6340 (2022).

39 Jiang, Y., Boije, M., Westermark, B. & Uhrbom, L. PDGF-B Can sustain self-renewal and tumorigenicity of experimental glioma-derived cancer-initiating cells by preventing oligodendrocyte differentiation. *Neoplasia* **13**, 492-503, doi:10.1593/neo.11314 (2011).

40 Kim, Y. *et al.* Platelet-derived growth factor receptors differentially inform intertumoral and intratumoral heterogeneity. *Genes Dev* **26**, 1247-1262, doi:10.1101/gad.193565.112 (2012).

41 Liu, T. *et al.* PDGF-mediated mesenchymal transformation renders endothelial resistance to anti-VEGF treatment in glioblastoma. *Nat Commun* **9**, 3439, doi:10.1038/s41467-018-05982-z (2018).

42 Cenciarelli, C. *et al.* PDGF receptor alpha inhibition induces apoptosis in glioblastoma cancer stem cells refractory to anti-Notch and anti-EGFR treatment. *Mol Cancer* **13**, 247, doi:10.1186/1476-4598-13-247 (2014).

43 Crouch, E. E. & Doetsch, F. FACS isolation of endothelial cells and pericytes from mouse brain microregions. *Nat Protoc* **13**, 738-751, doi:10.1038/nprot.2017.158 (2018).

44 Bergers, G. & Song, S. The role of pericytes in blood-vessel formation and maintenance. *Neuro Oncol* **7**, 452-464, doi:10.1215/S1152851705000232 (2005).

45 Jin, X., Jin, X., Jung, J. E., Beck, S. & Kim, H. Cell surface Nestin is a biomarker for glioma stem cells. *Biochem Biophys Res Commun* **433**, 496-501, doi:10.1016/j.bbrc.2013.03.021 (2013).

46 Pollen, A. A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55-67, doi:10.1016/j.cell.2015.09.004 (2015).

47 Birbrair, A. *et al.* Pericytes at the intersection between tissue regeneration and pathology. *Clin Sci (Lond)* **128**, 81-93, doi:10.1042/CS20140278 (2015).

48 Dore-Duffy, P., Katychev, A., Wang, X. & Van Buren, E. CNS microvascular pericytes exhibit multipotential stem cell activity. *J Cereb Blood Flow Metab* **26**, 613-624, doi:10.1038/sj.jcbfm.9600272 (2006).

49 Karow, M. *et al.* Reprogramming of pericyte-derived cells of the adult human brain into induced neuronal cells. *Cell Stem Cell* **11**, 471-476, doi:10.1016/j.stem.2012.07.007 (2012).

50 Wallmann, T. *et al.* Microglia Induce PDGFRB Expression in Glioma Cells to Enhance Their Migratory Capacity. *iScience* **9**, 71-83, doi:10.1016/j.isci.2018.10.011 (2018).

51 Xie, X. P. *et al.* Quiescent human glioblastoma cancer stem cells drive tumor initiation, expansion, and recurrence following chemotherapy. *Dev Cell* **57**, 32-46 e38, doi:10.1016/j.devcel.2021.12.007 (2022).

52 Choudhury, A. *et al.* Perivascular NOTCH3+ stem cells drive meningioma tumorigenesis and resistance to radiotherapy. *Cancer Discov*, doi:10.1158/2159-8290.CD-23-1459 (2024).
53 Crouch, E. E. *et al.* Ensembles of endothelial and mural cells promote angiogenesis in prenatal human brain. *Cell* **185**, 3753-3769 e3718, doi:10.1016/j.cell.2022.09.004 (2022).
54 Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885-892, doi:10.1038/s41586-021-04369-3 (2022).
55 Xie, X. P. *et al.* High-resolution mouse subventricular zone stem-cell niche transcriptome reveals features of lineage, anatomy, and aging. *Proc Natl Acad Sci U S A* **117**, 31448-31458, doi:10.1073/pnas.2014389117 (2020).
56 Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nat Rev Mol Cell Biol* **14**, 329-340, doi:10.1038/nrm3591 (2013).
57 Huang-Hobbs, E. *et al.* Remote neuronal activity drives glioma progression through SEMA4F. *Nature* **619**, 844-850, doi:10.1038/s41586-023-06267-2 (2023).
58 Gimple, R. C., Bhargava, S., Dixit, D. & Rich, J. N. Glioblastoma stem cells: lessons from the tumor hierarchy in a lethal cancer. *Genes Dev* **33**, 591-609, doi:10.1101/gad.324301.119 (2019).
59 Ge, W. *et al.* Human Organoid Tumor Transplantation Identifies Functional Glioblastoma - Microenvironmental Communication Mediated by PTPRZ1. *bioRxiv*, 2024.2005.2002.592055, doi:10.1101/2024.05.02.592055 (2024).
60 Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224, doi:10.1038/s41586-018-0744-4 (2018).
61 Kong, W. *et al.* CellTagging: combinatorial indexing to simultaneously map lineage and identity at single-cell resolution. *Nat Protoc* **15**, 750-772, doi:10.1038/s41596-019-0247-2 (2020).
62 Wang, Z. *et al.* Cell Lineage-Based Stratification for Glioblastoma. *Cancer Cell* **38**, 366-379 e368, doi:10.1016/j.ccell.2020.06.003 (2020).
63 Jain, R. *et al.* HEART DEVELOPMENT. Integration of Bmp and Wnt signaling by Hopx specifies commitment of cardiomyoblasts. *Science* **348**, aaa6071, doi:10.1126/science.aaa6071 (2015).
64 Yamashima, T. *et al.* Vascular adventitia generates neuronal progenitors in the monkey hippocampus after ischemia. *Hippocampus* **14**, 861-875, doi:10.1002/hipo.20001 (2004).
65 Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol* **42**, 293-304, doi:10.1038/s41587-023-01767-y (2024).
66 Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods* **20**, 1323-1335, doi:10.1038/s41592-023-01943-7 (2023).
67 Shekhar, K. *et al.* Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell* **166**, 1308-1323 e1330, doi:10.1016/j.cell.2016.07.054 (2016).
68 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324, doi:10.1016/j.cels.2019.03.003 (2019).
69 Tickle T, T. I., Georgescu C, Brown M, Haas B. inferCNV of the Trinity CTAT Project. (2019).
70 Concordet, J. P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**, W242-W245, doi:10.1093/nar/gky354 (2018).
71 Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* **4**, 5400, doi:10.1038/srep05400 (2014).
72 Yu, K. *et al.* PIK3CA variants selectively initiate brain hyperactivity during gliomagenesis. *Nature* **578**, 166-171, doi:10.1038/s41586-020-1952-2 (2020).
73 Pollen, A. A. *et al.* Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. *Cell* **176**, 743-756 e717, doi:10.1016/j.cell.2019.01.017 (2019).
74 Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523-527, doi:10.1038/s41586-019-1289-x (2019).
75 Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci U S A* **110**, 20284-20289, doi:10.1073/pnas.1315710110 (2013).
76 Zorita, E., Cusco, P. & Filion, G. J. Starcode: sequence clustering based on all-pairs search. *Bioinformatics* **31**, 1913-1919, doi:10.1093/bioinformatics/btv053 (2015).