Identification of astrocyte-driven pseudolineages reveals clinical stratification and therapeutic targets in Glioblastoma

Leo Carl Foerster[1,2,9], Oguzhan Kaya[1,2,9], Valentin Wüst[2,3], Milica Bekavac[1,2], Kevin Chris Ziegler[1], Vuslat Akcay[1,2], Nina Stinchcombe[1,2], Noelia Gesteira Perez[1], Xiujian Ma[2,5], Ahmed Sadik[6], Phuong Uyen Le[7], Kevin Petrecca[7], Christiane Opitz[6], Haikun Liu[5], Christian Rainer Wirtz[8], Simon Anders[3], Angela Goncalves[4], and Ana Martin-Villalba[1]*

[1] Molecular Neurobiology, German Cancer Research Center, 69120 Heidelberg, Germany

[2] Combined Faculty of Mathematics, Engineering and Natural Sciences, University of Heidelberg, 69120 Heidelberg, Germany

[3] BioQuant Centre, University of Heidelberg, 69120 Heidelberg, Germany

[4] Somatic Evolution and Early Detection, German Cancer Research Center, 69120 Heidelberg, Germany

[5] Molecular Neurogenetics, German Cancer Research Center, 69120 Heidelberg, Germany

[6] Metabolic Crosstalk in Cancer, German Cancer Research Center, 69120 Heidelberg, Germany

[7] Department of Neurosciences, Montreal Neurological Institute-Hospital, McGill University, Montreal, QC, Canada

[8] Department of Neurosurgery, Ulm University Hospital, 89081 Ulm, Germany

[9] These authors contributed equally: Leo Carl Foerster, Oguzhan Kaya

*Corresponding author: a.martin-villalba@dkfz-heidelberg.de

Summary (151 words)

Cancer research has predominantly targeted genetic mutations, while only recently has attention shifted to understanding tumor cell-stages. However, the key organizational principles guiding tumor dynamics towards sustainable growth remained unexplored. By analyzing tumor cell ensembles from individuals with glioblastoma through the lens of the healthy adult stem cell lineage, we identified astrocytes as central to glioblastoma progression. We found dormant tumor cells resembling astrocytes progressing to active and differentiated stages, building tumor pseudolineages that ultimately influence patient survival. These tumor stages align with specific methylomes, offering potential for patient classification. Our study identifies the Wnt antagonist SFRP1 as a missing factor in glioblastoma that plays a crucial role in the transition from quiescence to activation in the healthy lineage. Excitingly, re-introduction of SFRP1 in glioblastoma halts tumor dynamics, enhancing survival in a PDX model. This fresh view on glioblastomas underscores the importance of understanding tumor dynamics and unveils novel therapeutic avenues.

Introduction

Glioblastoma multiforme (GBM), the most commonly diagnosed and fatal brain cancer in adults, is notorious for its dismal five-year survival rate of 7% and inevitable recurrence[1]. This malignant tumor wreaks havoc primarily due to its aggressive invasive properties and extensive connectivity throughout the highly vulnerable neural tissues[2–4]. Metaphorically speaking, GBM can be compared to the *de novo* development of an organ within the mature brain, complete with distinct cell types and microenvironments. This necessitates a structured organization of neoplastic cells into defined hierarchies that maintain the progression of the 'organ' or tumor. In a healthy brain, a similar hierarchy is seen in adult neural stem cells (NSCs) and their progeny, providing a model for comparison. These specialized astrocytes[5] reside in specific brain regions such as the ventricular-subventricular zone (v-SVZ), where they remain quiescent until signaled to activate and ultimately differentiate[6–9]. Aberrant activation of NSCs has been implicated in uncontrolled tumor progression[10–12], leading to the longstanding hypothesis that NSCs serve as the cell-of-origin for GBM[13,14]. However, dormant NSCs and parenchymal astrocytes share significant transcriptional similarities[15] and the ability to generate differentiated progeny upon injury[16,17]. This suggests that aberrant activation of these astrocytes may initiate progression toward GBM. Curiously, current single-cell sequencing studies often classify astrocyte-like tumor cells within the differentiated population[18,19], neglecting their potential as a GBM cell-of-origin. We therefore set out to explore GBM hierarchies through the perspective of healthy NSC lineages, placing astrocytes at the lineage onset. This approach provides a new perspective on GBM development, allowing us to uncover similarities and unveil vulnerabilities, shedding light on the complex mechanisms underlying the organization of GBMs.

Here, we directly compare GBM cellular hierarchies to those of adult mouse NSCs to interrogate the processes involved in tumor growth and organization. We present *ptalign*, an innovative approach that maps tumor cells onto a healthy pseudotime trajectory, creating tumor-specific pseudolineages. We characterize *ptalign* pseudolineages of 51 primary GBMs, further extrapolating pseudolineage-related outcomes and biomarkers from 399 bulk samples. This analysis reveals SFRP1, a secreted Wnt antagonist, to be recurrently dysregulated in GBM. Its reintroduction induces the generation of astrocyte-like tumor cells, significantly enhancing survival and offering a potential target for inhibiting malignant lineage progression in GBM. Collectively, this study offers a pioneering approach to study tumor organization via direct comparison to a healthy reference, enhancing understanding of tumor hierarchies and offering a pathway towards precision medicine strategies.

Results

NSC dynamics are captured by QAD stages

In order to compare GBM tumors to an equivalent healthy lineage in the adult brain, we first curated a murine v-SVZ NSC-lineage dataset comprising 14,793 single-cell transcriptomes from

n=6 replicates[20–22]. UMAP[23] embedding highlighted a NSC differentiation trajectory, which we isolated and integrated with the v-SVZ populations from [24] to harmonize cell type labels (SFig.1a-d). Major lineage transitions were visible in a two dimensional PCA (SFig.1e) and we fit a diffusion pseudotime[25] to capture their ordering. We excluded actively cycling cells from pseudotime inference steps as the cell cycle itself represents a branching event in pseudotime which disrupts the otherwise consistently one dimensional differentiation trajectory of NSCs (SFig.1f-g).

The maintenance and dynamics of the v-SVZ lineage can be reduced to the transitions between three stable stages: quiescence (Q), activation (A), and differentiation (D); together QAD. Here, we define v-SVZ QAD-stages by valleys in cell-density along pseudotime (SFig.1h). In this way, quiescence characterizes the early NSC lineage pseudotime, encompassing dormant and quiescent NSCs. We recently demonstrated[26] that dormant v-SVZ NSCs exhibit a specific astrocyte-methylome that distinguishes them from the stem-cell-methylome of quiescent NSCs. Thus, in the healthy v-SVZ, dormant astrocytes are included in the stem cell lineage and precede quiescent NSCs, consistent with the differentiation trajectory inferred from pseudotime. These stages are followed by activation, which comprises replication-poised non-cycling (A^{nc}) and cycling (A^{c}) NSCs, while differentiation includes neuroblasts (NBs) and neurons (Fig.1a).

Generalizing these stages, we derived a 242-gene pseudotime-predictive geneset (SVZ-QAD: Methods, S.Table 2) which enabled us to compare NSC pseudotime trajectories in human and mouse. We derived human NSC lineage pseudotime based in radial glia cells from four human cortical brain organoids (HBOs) (SFig.2a-d) and compared it to our mouse NSC lineage to find conserved expression patterns among individual QAD markers and globally using dynamic time warping (DTW) (SFig.2e-i). These results highlight the strong conservation of cell hierarchies in NSCs regardless of differences in species or ontogeny.

ptalign **unveils tumor pseudolineages in GBM**

To directly compare cellular hierarchies in health and disease, we borrowed principles from sequence alignment, where local similarity scores determine the placement of a query sequence within a larger reference. We employ a conceptually similar strategy in *ptalign*, scoring query-reference correlation dynamics to place cells along a reference pseudotime. Our approach involves training a neural network to predict pseudotime from query-vs-reference cell correlations over a defined geneset (e.g. SVZ-QAD; Fig.1b.I-II), assigning ‘aligned’ pseudotimes based on the best-matched reference increment (Fig.1b.III). Performance of our method is evaluated using three DTW-derived metrics (Methods), which gauge the ability of the supplied geneset to explain the query cell dynamics in a permutation paradigm (Fig.1b.IV). Finally, we assign lineage annotations to each cell based on their aligned pseudotimes, constructing what we term as the *pseudolineage* of the query cells. We benchmarked this method by comparing conventional dataset-derived human and mouse lineage pseudotimes with their *ptalign*-derived counterparts, finding strong correlation in our HBO dataset (Pearson=0.85, SFig.2j-l) and an unseen v-SVZ NSC lineage dataset (Pearson=0.95, SFig.2m-r).Thus, *ptalign* provides a robust

and scalable framework to identify lineage trajectories by comparison to a fixed lineage reference.

Applying *ptalign* to tumor-cell hierarchies, we conducted pseudotime alignment of healthy v-SVZ NSCs with 4 replicates of a GBM patient derived xenograft (PDX, pseudonymously T6) sequenced 5 months post-injection (mpi) by a miniaturized and automated Smart-seq2/3 protocol[27] (SFig.3a-d). This alignment is shown schematically in (Fig.1c). Scoring for SVZ-QAD genesets indicated the presence of all three lineage stages in the PDX (SFig.3e), which were consistently identified by *ptalign* (Fig.1d). The T6 pseudolineage was dominated by A^{nc}-cells which were complemented by a large Q- and smaller D- and A^{c}-cell population (Fig.1d and SFig.3f). We separately generated patient-derived T6 allografts (PDA) by injecting tumor spheres into HBOs and found that PDA pseudolineages recapitulated those of the PDX (SFig.3g-m), supporting the intrinsic organization of GBM cells along a distinct pseudolineage.

Next, we compiled 66 published primary GBM scRNAseq datasets[18,28–34], (S.Table 1, subsequently as 'cohort'). Malignant cells passing quality control thresholds and tumors with >500 cells were retained in a 55-tumor cohort (SFig.4a-b) for which we called GBM pseudolineages by *ptalign*. After applying *ptalign* permutation thresholds to eliminate GBMs without clear QAD lineage transitions, the remaining 51-tumor UMAP grouped GBMs by patient origin (Fig.1e), thus masking any potential shared biology. Instead, by projecting these tumors onto a common cell trajectory using *ptalign* (Fig.1f) we were able to discern a clear progression of Q-, A-, and D-dominated tumors (Fig.1g) closely following the v-SVZ NSC lineage continuum. Tumor pseudolineages were consistently dominated by A-stage cells while the D-stage appears dispensable in GBM. Q-stage cells, on the other hand, were present in virtually all tumors (50/51), supporting their essential role in fueling tumor lineages[35] (SFig.4c). Using inferCNV[36] to call copy-number variation (CNV) events from scRNA-seq, we additionally determined that CNV is not a major source of intra-tumor pseudolineage heterogeneity[19,37–40] (SFig.5a-e). Importantly, these insights are only discernable through systematic assessment of tumor trajectories delineated by *ptalign*, projecting tumor cells onto the healthy reference and facilitating their interpretation in that context.

Pseudolineages inform clinical outcomes

In our 51-GBM cohort, Q- and D-cell fractions did not show any significant associations with available patient metadata (SFig.5f-h). Thus, we retrieved clinically annotated cohorts of IDH-wildtype GBMs from The Cancer Genome Atlas (TCGA) (343 patients, microarray and RNA-seq) and Wu *et al*.[41] (56 patients, RNAseq) (Fig.2a and S.Table 1) and decomposed samples by pseudolineage using a 271-gene GBM-QAD signature (Methods, S.Table 2 and SFig.6a-b). We identified four main tumor classes, each enriched (e-) in predominant cell stages of their pseudolineage: eQ, eA, eQD, and eAD (Fig.2b-d and Methods). When evaluating patient survival by tumor class, only those comparisons spanning the Q-to-A transition yielded significant results ($p<0.1$, Fig.2e and SFig.7a), underscoring the critical role of the Q-to-A transition in directing tumor progression. Consistently, reduced hazard was linked to higher Q-scores in a Cox model, with the protective effect of Q notably surpassing the detrimental effects

of A (Fig.2f). This trend was also evident in progression-free intervals (SFig.7b). Despite both eQ and eQD tumors having a high degree of Q, the latter exhibited significantly improved progression-free intervals (SFig.7c), highlighting the prognostic advantage for a fully developed but inactive lineage over a truncated one. These findings underscore the significance of targeting the Q-to-A transition to achieve positive patient outcomes.

When we compared our identified tumor classes to previously published GBM signatures[18,28,42,43] using geneset scoring, we identified three cellstate-groups broadly corresponding to QAD-stages (SFig.6c-d). Intriguingly, D-stage cells were simultaneously enriched in the 'stem-like'-signature and depleted in the 'diff(erentiated)-like'-signature from [43]. This cell-state nomenclature aligns with historic associations of stem-like populations with cycling cells (SFig.6f) which assume astrocyte-like (Q-stage) cells to be a terminally differentiated population[18,19]. However, v-SVZ cycling cells relinquish their stem cell identity (Q-genes) and gain D-gene expression (SFig.6g), revealing that not stem- but differentiation-programs manifest during cell-cycle. This is a prime example of how the direct comparison of GBM and NSC cellular hierarchies by *ptalign* can enhance our understanding of tumor processes by leveraging the wealth of contextual knowledge available from NSCs.

Multimodal pseudolineage biomarkers

Next, we sought biomarkers for efficient patient stratification by tumor pseudolineage, singling out methylome arrays commonly used in WHO-classification of brain tumors[44]. In our prior work, we established that various stages of the healthy v-SVZ lineage (e.g. QAD) are characterized by distinct methylome profiles[26]. We hypothesized that differences in these stages' abundance would be evident in bulk tumor-methylome profiles. To test this, we used 83 IDH-wt GBM tumors with matched RNA-seq and methylation data from the TCGA and Wu *et al*. cohorts. PCA over highly-variable probes revealed that the first principal component (PC1) generally distinguishes alive and deceased patients[45] (SFig.7d-e), while tumors segregated into QAD-stages along PC2-3 (Fig.2g-h). As pairwise inter-tumoral RNA- and methylation-distances were correlated, we trained an ElasticNet model to predict RNA scores (cf. Fig.2b) from tumor methylomes and achieved a mean Pearson correlation of 0.70 on the holdout data (Fig.2i, SFig.7f-h, and S.Table 3). Thus, tumor pseudolineages can be assessed from bulk methylomes, facilitating cost-effective patient stratification and disease monitoring.

Further, we focused on A^{nc} cells to predict lineage bias, reasoning that given their universal presence across tumors they might harbor effective pseudolineage biomarkers. We compared expression in A^{nc} cells in 11 early-pseudotime (pt) biased and 16 late-pt biased tumors determined by the relative prevalence of dormant astrocyte-like Q-cells or D-cells, respectively (Methods). This analysis identified the secreted Wnt antagonist DRAXIN as a reliable predictor of tumor pseudolineage, being consistently higher-expressed in late-pt biased tumors compared to early-pt biased ones (Fig.2j-k). Despite being understudied, DRAXIN is known to function similarly to DKK in the mouse Wnt ecosystem[46], highlighting the pivotal role of Wnt signaling in regulating tumor stage transitions, particularly the Q-to-A transition highlighted above.

Aberrant cWnt activity in GBM pseudolineages

Collectively, our findings indicate that enhancing quiescence, rather than targeting activation, offers an effective means to improve patient outcomes. We hypothesized that this is due to the disruption of tumor dynamics and subsequently aimed to identify the key signaling pathways governing the shift from quiescence to activation. In the healthy adult NSC lineage we have shown that quiescence is governed through canonical Wnt (cWnt) signaling activity[20,47]. Thus, we directly compared cWnt activity levels in QAD stages of the healthy NSC lineage and its malignant counterpart using genetic reporters (TCF/Lef-EGFP, Methods) (Fig.3a).

Examination of reporter expression in the mouse v-SVZ lineage, we observed peak cWnt activity in the Q-stage which is abrogated in the A-stages and resurges in the late D-stage (Fig.3b-c). Of note, dormant parenchymal astrocytes exhibited lower cWnt activity than quiescent NSCs (SFig.8a). By contrast, in the malignant lineage, PDX and PDA GBM cells exhibited sustained cWnt activity throughout QAD-stages (Fig.3b), which was corroborated by spatial transcriptomics (S.Table 2; Fig.3d-e and SFig.8b-e). Intriguingly, tumor ventricular growths represented a special case, displaying no detectable cWnt activity in GBM cells (Fig.3f-g and SFig.8f), an observation consistent with *in vitro* cultured GBM spheroids (SFig.8g-h) and possibly attributable to the lack of contact with adjacent tissues. Nonetheless, tumor cells in ventricular growths organized along pseudolineage gradients (SFig.8i), emphasizing their capability to form lineages in this context. Together, we find that the Q-to-A transition is guided by tight regulation of cWnt signaling in the healthy lineage, which is dysregulated along the tumor pseudolineage. These results suggest that aberrant Wnt signaling licenses sustained tumor activation and growth.

To single out potential Wnt-related factors dysregulated in the Q-to-A transition, we first measured the pseudotime expression dynamics of a set of 37 Wnt pathway genes expressed in the v-SVZ NSC lineage. This revealed a cascading landscape of expression (Fig.3h) consistent with tight regulation of cWnt activity along the NSC lineage *in vivo*. We then fixed the v-SVZ expression dynamics and scored their deviation in tumors, deriving a pan-tumor Wnt-gene-dynamics consistency score (Methods). This identified DRAXIN, which peaks at the D-stage (Fig.3m-n), as having the best-conserved pseudotime expression dynamics, consolidating its use as a biomarker. Intriguingly, other secreted Wnt antagonists such as SFRP1 and NOTUM, which peak at the Q-stage in the healthy lineage, scored among the worst (Fig.3i-l). These results highlight the selective dysregulation of Wnt antagonists at the Q-to-A transition in GBM.

SFRP1 stalls malignant lineage progression

We have established the Q-to-A transition as an important modulator of GBM progression which is regulated by cWnt signaling. We reasoned that targeting this transition can lead to therapeutically relevant outcomes, including retaining cells in a Q-stage or forcing their activation. To test this, we reintroduced the Wnt antagonist SFRP1, which regulates this

transition in healthy NSCs[48], in GBM, cloning human SFRP1 downstream of the TCF/Lef promoter (Fig.4a; SFig.9a-b). Strikingly, mice bearing SFRP1-overexpressing (OE) tumors exhibited significantly improved overall survival (Fig.4b) while transcriptome and immunohistochemistry analysis revealed that SFRP1-OE stalled GBM cells in a dormant astrocyte-like stage (Fig.4c-d,f-g). GO-enrichments of SFRP1-OE tumors showed increased cytokine production and ion transport consistent with astrocyte function**.** Simultaneously, activation-linked terms including translation, spindle assembly and fatty acid catabolism were decreased (Fig.4e and S.Table 2). We observed a similar effect of SFRP1-OE on tumor cells in a PDA model (SFig.9c-d), while expression profiles of healthy cells remained unchanged (SFig.9e). These results were reproduced by the overexpression of NOTUM, another secreted Wnt antagonist (SFig.9f-h, n=1). Interestingly, Notum-OE and SFRP1-OE tumors exhibited similar transcriptional profiles, suggesting overlapping downstream effectors direct the generation of astrocytic tumor cells (SFig98i, Pearson=0.76).

Drawing upon our spatial transcriptomics data, we found Q-cells enriched across all brain regions in SFRP1-OE tumors (Fig.4h-i). Consistently, spatial neighborhood comparisons did not determine a significant association of QAD-stage with specific murine cell types in SFRP1-OE or control tumors (SFig.9j). Only ventricular outgrowths, which have no detectable cWnt activity (cf. Fig.3f), exhibited higher fractions of A-cells (Fig.4i). Interestingly, ventricular growths in both control and SFRP1-OE tumors exhibited QAD cells arranged along ordered gradients (SFig.8i and SFig.9l) and not intermingled as elsewhere in the brain.

To corroborate the astrocyte identity of SFRP1-OE tumor cells, we conducted whole-genome bisulfite sequencing (WGBS) on three technical replicates of SFRP1-OE and control PDX tumors (Methods). Indeed, we found our previously reported astrocyte-differentially methylated regions[26] (DMRs) selectively demethylated in SFRP1-OE (Fig.4j and S.Table 3). These insights were only discernible in WGBS data, as most v-SVZ DMRs had coverage of three or fewer probes on commercial methylation arrays (SFig.9n). Notably, one of the most divergently methylated DMRs overlapped an NFIB-promotor (Fig.4k). NFIB, a primary transcriptional activator of the astrocyte marker GFAP[49], is also used to reprogram iPSCs into astrocytes[50]. Expectedly, we identified a significant increase in NFIB+ and GFAP+ transcript-expressing cells in SFRP1-OE tumors (SFig.9o). Of note, the SFRP1 locus was methylated in both healthy NSCs and SFRP1-OE tumor cells (SFig.9m), suggesting that activated NSCs express SFRP1 to calibrate the activation of neighboring cells, as previously demonstrated for Notch ligands[15,51].

Collectively, these findings illustrate that the presence of SFRP1 influences the Q-to-A transition in tumor cells by promoting a shift toward a dormant astrocytic stage, thus halting their progression to the active stage. This supports the notion that astrocytes hold a foundational role in the pseudolineage, drawing parallels to the adult healthy lineage, where damaged or specialized astrocytes sustain lineage dynamics.

Discussion

In this work, we have clarified fundamental principles of tumor organization through pseudolineage analysis, conducting a direct comparison of adult stem cell lineages with tumor cells originating from the same tissue. By interpreting GBM processes through the lens of healthy NSC lineage transitions, we propose quiescent astrocyte-like cells as the progenitors of tumor pseudolineages. We identified pseudolineage-specific biomarkers which facilitate a more precise prognostic categorization of patients. Significantly, our targeted modulation of the Quiescence-to-Activation transition substantiates the functional role of these pseudolineages in driving tumor progression, highlighting it as a crucial transitional stage. The identification of Wnt activity as a pivotal regulator in this context not only introduces a fresh therapeutic avenue for brain tumors but also provides insights into the challenges facing Wnt-based treatments in other solid tumors. Overall, the unique perspectives offered by our comprehensive strategy not only hold implications for unraveling the organizational principles of various solid tumors, but also in guiding patient stratification and the identification of therapeutic targets.

Our unique perspective positions astrocytes as the initiators of GBM pseudolineages. Despite NSCs being recognized as the GBM cell-of-origin, the contribution of astrocytes, which resemble dormant NSCs[15], is underestimated. Existing single cell transcriptomics studies identify 'stem-like' tumor cells based on cycling activity[18,19], thereby overlooking the contribution of these dormant populations to tumor progression. Instead, by referring to the healthy NSC lineage, we reveal that cycling licenses differentiation, forming a sequence from dormancy through quiescence, cycling, and differentiation, which is also observed in tumors. Consistently, targeting the Q-to-A transition, for example by SFRP1 or NOTUM, reinforces the astrocytic origins of GBM pseudolineages by eliminating cycling cells while preserving the dormant state. Together, these results highlight a route to tumorigenesis through unsuccessful astrocyte-driven regenerative attempts as recently suggested in[52]. Thus, our approach sheds light on tumor dynamics and supports the repositioning of quiescent astrocyte-like cells at the beginning of the tumor pseudolineage, emphasizing their importance as therapeutic targets.

Applying *ptalign* offers valuable insights into the processes governing the origin and progression of tumor lineages. Unlike other trajectory alignment approaches[53–55], *ptalign* places tumor cells within a reference trajectory, enabling simultaneous comparison of tumors with both other tumors and healthy conditions. This innovation enabled the identification of SFRP1 as frequently dysregulated in human tumors[56] and predicted its action mediating the Q-to-A transition. Through pseudolineage analysis, we identified promising expression- and methylation-based biomarkers for patient stratification, the latter which could additionally benefit from improved detection of lineage-specific DMRs. Our findings support previous observations that CNV clones do not drive divergent cellular hierarchies[19,37–40], although we did not differentiate recurrent tumors undergoing subtype-transformations[57,58]. Notably, as opposed to local segregation of CNV clones[59], our spatial transcriptomics analysis shows that tumor pseudolineages exhibit no cell type or region bias, suggesting tumor biopsies should accurately represent each patient pseudolineage.

Pseudolineage analysis revealed the pivotal role of the Q-to-A transition in directing patient outcomes. Thus we targeted this transition by SFRP1 overexpression and achieved increased survival of tumor bearing mice. This strategy differs from therapeutic best practice by targeting a

transition over a stage. Whereas targeting e.g. cycling or dormant tumor cells individually leaves a population of tumor cells able to replenish pseudolineages, prior targeting of transitions stalls the tumor by stage, paving the way for combination therapies targeting that stage. Significantly, our proposed biomarkers effectively identify patients who would benefit from SFRP1-based therapies. We envision exciting prospects for SFRP1-based therapies that target the Q-to-A transition to stall tumor progression.

Within ventricular growths, we observed highly active pseudolineages devoid of cWnt activity and with QAD cells arranged along a gradient. Conversely, in other brain regions, infiltrating tumor cells were more intermingled across QAD stages and accompanied by ubiquitous cWnt activity. This contrast highlights the unique characteristics of ventricular outgrowths, potentially explaining the aggressive disease progression in patients with ventricular tumors[60]. Such a mechanism would be consistent with metastatic cells at the vessel lumen, where cWnt activity is only gained upon extravasation and triggers a proliferative-to-dormant switch[61]. Consequently, investigating Wnt activity in luminal outgrowths of colon cancer might reveal that these sites are insensitive to Wnt inhibition.

In sum, our study proposes a new perspective on glioblastoma organization with astrocytes as the foundational element of tumor pseudolineages. We highlight the instrumental role of targeting tumor dynamics to halt tumor progression and introduce pseudolineage analysis to decode tumor organization. The broader application of *ptalign* in a pan-cancer context offers the potential to uncover shared principles underlying stem cell activation that are exploited across tumors. Pseudolineage analysis emerges as a powerful tool for unraveling the complexities of cancer, guiding precision medicine approaches and improving patient outcomes.

References

1. Wen, P. Y. Malignant Gliomas in Adults. *N. Engl. J. Med.* (2008).

2. Carlsson, S. K., Brothers, S. P. & Wahlestedt, C. Emerging treatment strategies for glioblastoma multiforme. *EMBO Mol Med* **6**, 1359–1370 (2014).

3. Giese, A., Bjerkvig, R., Berens, M. E. & Westphal, M. Cost of Migration: Invasion of Malignant Gliomas and Implications for Treatment. *JCO* **21**, 1624–1636 (2003).

4. Venkataramani, V. *et al.* Glioblastoma hijacks neuronal mechanisms for brain invasion. *Cell* **185**, 2899-2917.e31 (2022).

5. Kriegstein, A. & Alvarez-Buylla, A. The Glial Nature of Embryonic and Adult Neural Stem Cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).

6. Furutachi, S. *et al.* Slowly dividing neural progenitors are an embryonic origin of adult neural stem cells. *Nat Neurosci* **18**, 657–665 (2015).

7. Lim, D. A. & Alvarez-Buylla, A. The Adult Ventricular–Subventricular Zone (V-SVZ) and Olfactory Bulb (OB) Neurogenesis. *Cold Spring Harb Perspect Biol* **8**, a018820 (2016).

8. Duan, X., Kang, E., Liu, C. Y., Ming, G. & Song, H. Development of neural stem cell in the adult brain. *Current Opinion in Neurobiology* **18**, 108–115 (2008).

9. Braun, S. M. G. & Jessberger, S. Adult neurogenesis in the mammalian brain. *Front. Biol.* **8**, 295–304 (2013).

10. Portela, M. *et al.* Glioblastoma cells vampirize WNT from neurons and trigger a JNK/MMP signaling loop that enhances glioblastoma progression and neurodegeneration. *PLoS Biol* **17**, e3000545 (2019).

11. Rheinbay, E. *et al.* An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. *Cell Reports* **3**, 1567–1579 (2013).

12. Nakada, M. *et al.* Aberrant Signaling Pathways in Glioma. *Cancers* **3**, 3242–3278 (2011).

13. Alcantara Llaguno, S. *et al.* Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model. *Cancer Cell* **15**, 45–56 (2009).

14. Lee, J. H. *et al.* Human glioblastoma arises from subventricular zone cells with low-level driver mutations. *Nature* **560**, 243–247 (2018).

15. Llorens-Bobadilla, E. *et al.* Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. *Cell Stem Cell* **17**, 329–340 (2015).

16. Magnusson, J. P. *et al.* A latent neurogenic program in astrocytes regulated by Notch signaling in the mouse. *Science* **346**, 237–241 (2014).

17. Zamboni, M., Llorens-Bobadilla, E., Magnusson, J. P. & Frisén, J. A Widespread Neurogenic Potential of Neocortical Astrocytes Is Induced by Injury. *Cell Stem Cell* **27**, 605-617.e5 (2020).

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

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

20. Kalamakis, G. *et al.* Quiescence Modulates Stem Cell Maintenance and Regenerative Capacity in the Aging Brain. *Cell* **176**, 1407-1419.e14 (2019).

21. Carvajal Ibañez, D. *et al.* Interferon regulates neural stem cell function at all ages by orchestrating mTOR and cell cycle. *EMBO Mol Med* **15**, e16434 (2023).

22. Kremer, L. P. M. *et al.* High throughput screening of novel AAV capsids identifies variants for transduction of adult NSCs within the subventricular zone. *Molecular Therapy - Methods & Clinical Development* **23**, 33–50 (2021).

23. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at http://arxiv.org/abs/1802.03426 (2020).

24. Cebrian-Silla, A. *et al.* Single-cell analysis of the ventricular-subventricular zone reveals signatures of dorsal and ventral adult neurogenesis. *eLife* **10**, e67436 (2021).

25. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**, 845–848 (2016).

26. Kremer, L. P. *et al.* Single-cell triple-omics uncovers DNA methylation as key feature of stemness in the healthy and ischemic adult brain. *bioRxiv* (2022) doi:10.1101/2022.07.13.499860.

27. Cerrizuela, S. *et al.* High-throughput scNMT protocol for multiomics profiling of single cells from mouse brain and pancreatic organoids. *STAR Protocols* **3**, 101555 (2022).

28. Richards, L. M. *et al.* Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nat Cancer* **2**, 157–173 (2021).

29. Yuan, J. *et al.* Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med* **10**, 57 (2018).

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

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

32. Wang, R. *et al.* Adult Human Glioblastomas Harbor Radial Glia-like Cells. *Stem Cell Reports* **14**, 338–350 (2020).

33. Jacob, F. *et al.* A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell* **180**, 188-204.e22 (2020).

34. Chen, A. X. *et al.* Single-cell characterization of macrophages in glioblastoma reveals MARCO as a mesenchymal pro-tumor marker. *Genome Med* **13**, 88 (2021).

35. Xie, X. P. *et al.* Quiescent human glioblastoma cancer stem cells drive tumor initiation, expansion, and recurrence following chemotherapy. *Developmental Cell* **57**, 32-46.e8 (2022).

36. inferCNV of the Trinity CTAT Project.

37. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).

38. Chaligne, R. *et al.* Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. *Nat Genet* **53**, 1469–1479 (2021).

39. Bhaduri, A. *et al.* Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* **26**, 48-63.e6 (2020).

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

41. Wu, Y. *et al.* Glioblastoma epigenome profiling identifies SOX10 as a master regulator of molecular tumour subtype. *Nat Commun* **11**, 6434 (2020).

42. Verhaak, R. G. W. *et al.* Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).

43. 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 (2021).

44. Louis, D. N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* **23**, 1231–1251 (2021).

45. Noushmehr, H. *et al.* Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell* **17**, 510–522 (2010).

46. Hutchins, E. J. & Bronner, M. E. Draxin acts as a molecular rheostat of canonical Wnt signaling to control cranial neural crest EMT. *Journal of Cell Biology* **217**, 3683–3697 (2018).

47. Seib, D. R. M. *et al.* Loss of Dickkopf-1 Restores Neurogenesis in Old Age and Counteracts Cognitive Decline. *Cell Stem Cell* **12**, 204–214 (2013).

48. Donega, V. *et al.* Single-cell profiling of human subventricular zone progenitors identifies SFRP1 as a target to re-activate progenitors. *Nat Commun* **13**, 1036 (2022).

49. Steele-Perkins, G. *et al.* The Transcription Factor Gene *Nfib* Is Essential for both Lung Maturation and Brain Development. *Molecular and Cellular Biology* **25**, 685–698 (2005).

50. Canals, I. *et al.* Rapid and efficient induction of functional astrocytes from human pluripotent stem cells. *Nat Methods* **15**, 693–696 (2018).

51. Dray, N. *et al.* Dynamic spatiotemporal coordination of neural stem cell fate decisions occurs through local feedback in the adult vertebrate brain. *Cell Stem Cell* **28**, 1457-1472.e12 (2021).

52. Mossi Albiach, A. *et al. Glioblastoma is spatially organized by neurodevelopmental programs and a glial-like wound healing response*. http://biorxiv.org/lookup/doi/10.1101/2023.09.01.555882 (2023) doi:10.1101/2023.09.01.555882.

53. Alpert, A. *et al.* Alignment of single-cell trajectories by tuMap enables high-resolution quantitative comparison of cancer samples. *Cell Systems* **13**, 71-82.e8 (2022).

54. Do, V. H. *et al.* Dynamic pseudo-time warping of complex single-cell trajectories. *bioRxiv* (2019) doi:10.1101/522672.

55. Cacchiarelli, D. *et al.* Aligning Single-Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of Myogenic Reprogramming Outcome. *Cell Systems* **7**, 258-268.e3 (2018).

56. Yu, J. *et al.* Association between SFRP promoter hypermethylation and different types of cancer: A systematic review and meta-analysis. *Oncol Lett* (2019) doi:10.3892/ol.2019.10709.

57. Varn, F. S. *et al.* Glioma progression is shaped by genetic evolution and microenvironment interactions. *Cell* **185**, 2184-2199.e16 (2022).

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

59. Erickson, A. *et al.* Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* **608**, 360–367 (2022).

60. Berendsen, S. *et al.* Adverse prognosis of glioblastoma contacting the subventricular zone: Biological correlates. *PLoS ONE* **14**, e0222717 (2019).

61. Jakab, M. *et al.* Lung endothelium instructs dormancy of susceptible metastatic tumour cells. *bioRxiv* (2022) doi:10.1101/2022.06.18.496680.

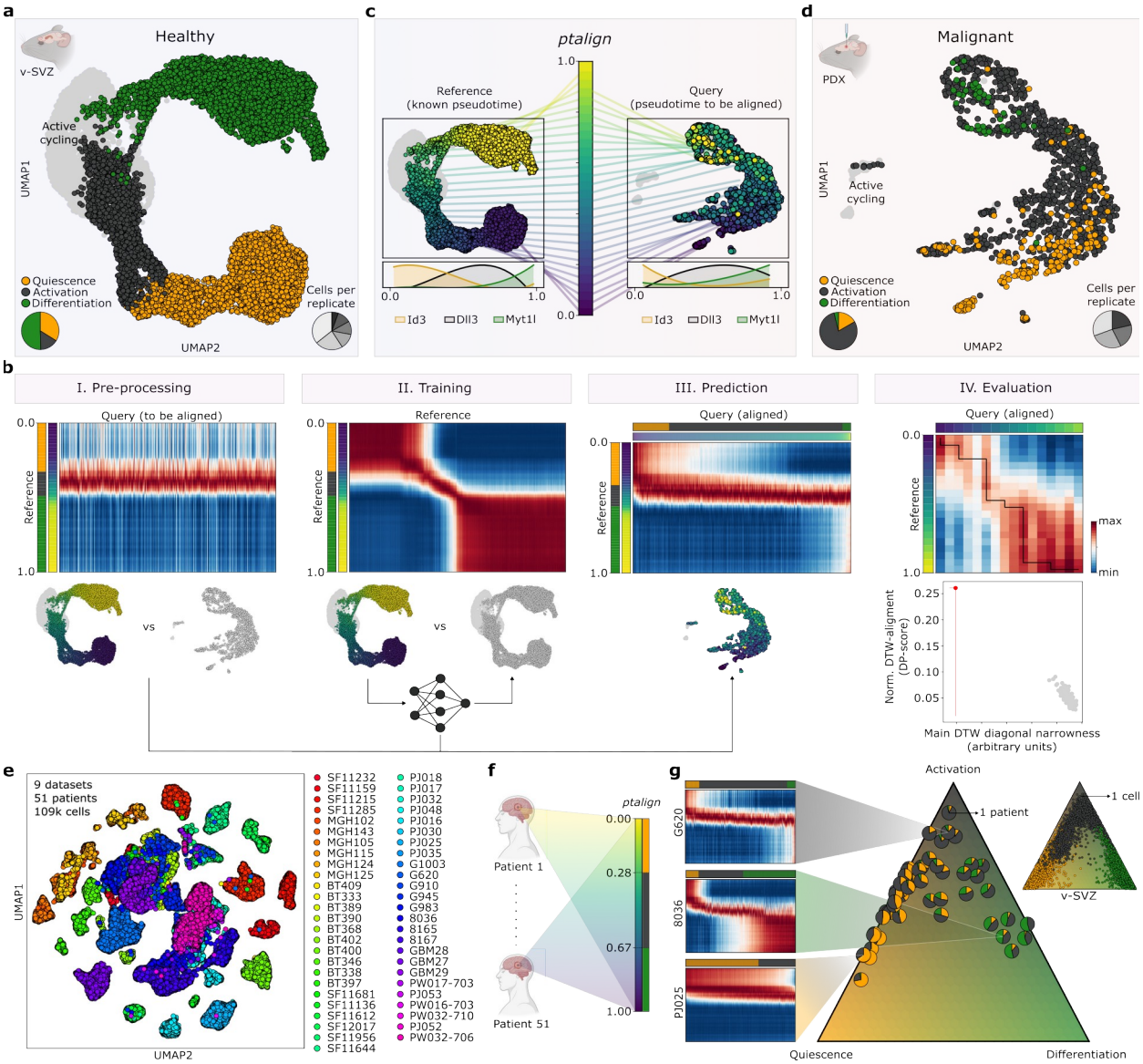


Fig. 1 | NSC lineage QAD states identified in a GBM cohort. a. UMAP embedding of integrated WT mouse v-SVZ NSC lineage scRNA-seq (n=14,793 cells; n=6 replicates), showing NSC differentiation trajectory comprising quiescence, activation, and differentiation (QAD) stages. Cycling cells are shown in gray and excluded in pseudotime analyses throughout. Pie charts indicate proportion QAD-stage cells and cells per replicate. **b.** Overview of *ptalign* pseudotime alignment for T6 tumor, from raw QAD correlation matrix of reference v-SVZ NSCs and unordered tumor cells (I), showing reference-reference pearson correlation matrix used to train pseudotime inference (II), and correlations from (I) ordered by aligned pseudotime (III) from which QAD labels are derived. The query-reference DTW (IV, top) with correlation-maximizing traceback from which *ptalign* pseudotime quality metrics are derived and used in a permutation paradigm to derive an empirical p-value (bottom). **c.** Reference v-SVZ NSC pseudotime (left) with scaled pseudotime expression splines of QAD-stage markers Id3, Dll3, and Myt1l. v-SVZ NSC pseudotime (center) extrapolated to GBM by *ptalign* (right), with pseudotime-binned cells linked to their average position in the UMAPs. **d.** T6 PDX UMAP of Smart-seq3 transcriptomes (n=1,176 cells, n=4 replicates) with *ptalign* derived QAD-stages, with cycling cells in gray. **e.** UMAP embedding of 109,079 cells from n=51 primary GBMs colored by patient origin. **f.** Schematic depiction of arbitrary patient samples projected onto v-SVZ NSC reference pseudotime by *ptalign*, enabling pseudolineage inference and QAD-stage assignment. **g.** Ternary plot of n=51 primary GBMs from (**e**) arranged by relative QAD-stage proportions from *ptalign*. Inset heatmaps show correlation matrices for representative Q, A or D-enriched tumors. Inset ternary relates QAD geneset scores for v-SVZ NSC cells from (**a**) with associated QAD state labels.

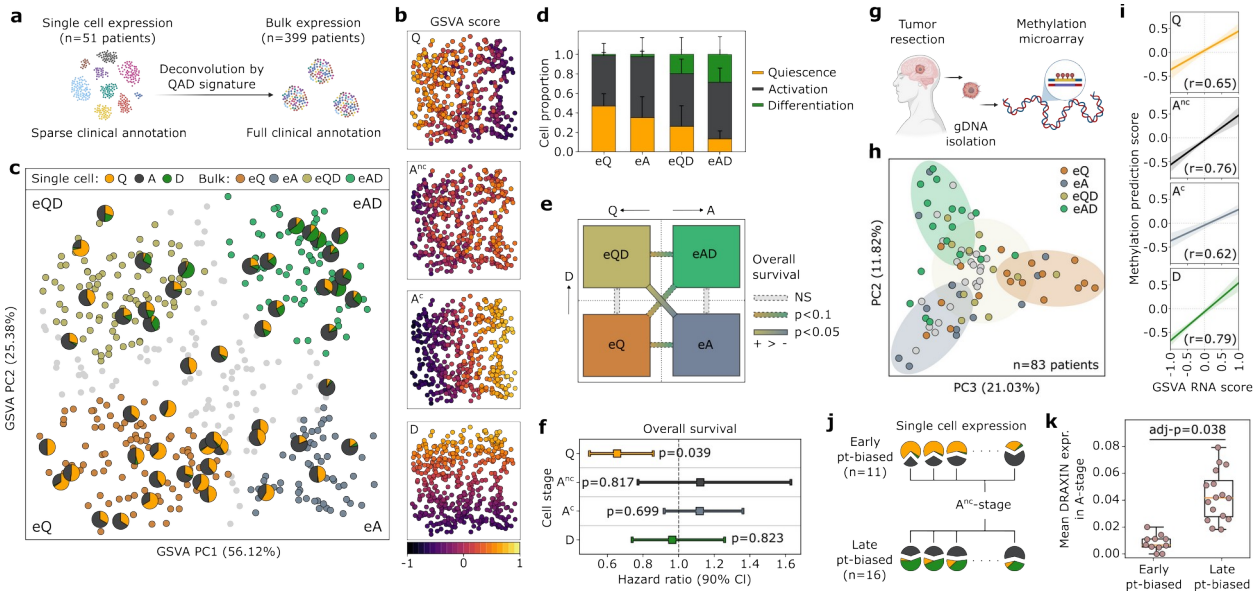


Fig. 2 | Patient stratification and biomarker assessment by GBM pseudolineage. a. Schematic depiction of extrapolating scRNA-seq data to fully annotated bulk TCGA samples using a GBM-QAD signature. **b.** TCGA and Wu *et al.* (ref[41]) tumors' (n=399) nonparametric GBM-QAD scoring by GSVA in 2D-PCA scatterplots of quiescence (Q), active non-cycling (A^{nc}) and active cycling (A^{c}), and differentiation (D) scores. **c.** GBM-QAD signature scores in PCA from (**b**) overlayed with *ptalign* QAD-stage piecharts from (Fig.1g). Four groups of GBMs enriched (e-) by stage are colored, making up eQ, eA, eQD, and eQD GBM classes. **d.** Mean proportion of QAD-stage cells per GBM class among scRNA-seq GBMs. Error bars denote standard deviation. **e.** Pairwise overall survival for eQ,eA, eQD, and eQD bulk GBMs. Significance was assessed by log-rank test in Kaplan-meier curve with patients up to 95th survival percentile included. Comparison color gradients indicate better (+) and worse (-) survival quadrants, respectively. **f.** Predicted hazards and 90% confidence interval from a univariate Cox proportional hazards model of overall survival by bulk GBM-QAD signature score from (**b**), with associated model p-values. **g.** Schematic depiction of gDNA isolation and bulk methylation quantification by Illumina 450k and EPIC 850k methylation microarrays. **h.** PCA embedding of variable methylation sites in (n=83) tumors with matched methylation and RNA profiling from (**c**). QAD-stage enrichments are underscored with transparent ellipses. **i.** Regression and 90% CI on a (n=28) holdout set for ElasticNet predictions on methylation data of GBM-QAD signatures from (**b**). Models were trained on a (n=55 GBMs) training cohort. Pearson correlations between truth and prediction (r) are indicated. **j.** Schematic depicting early pseudotime (pt) biased and late pt-biased GBMs compared by A^{nc}-stage expression levels. **k.** per-tumor pseudobulk mean proportion of DRAXIN A^{nc}-state expression. Statistical significance was determined by two-sided t-test with Benjamini-Hochberg correction.

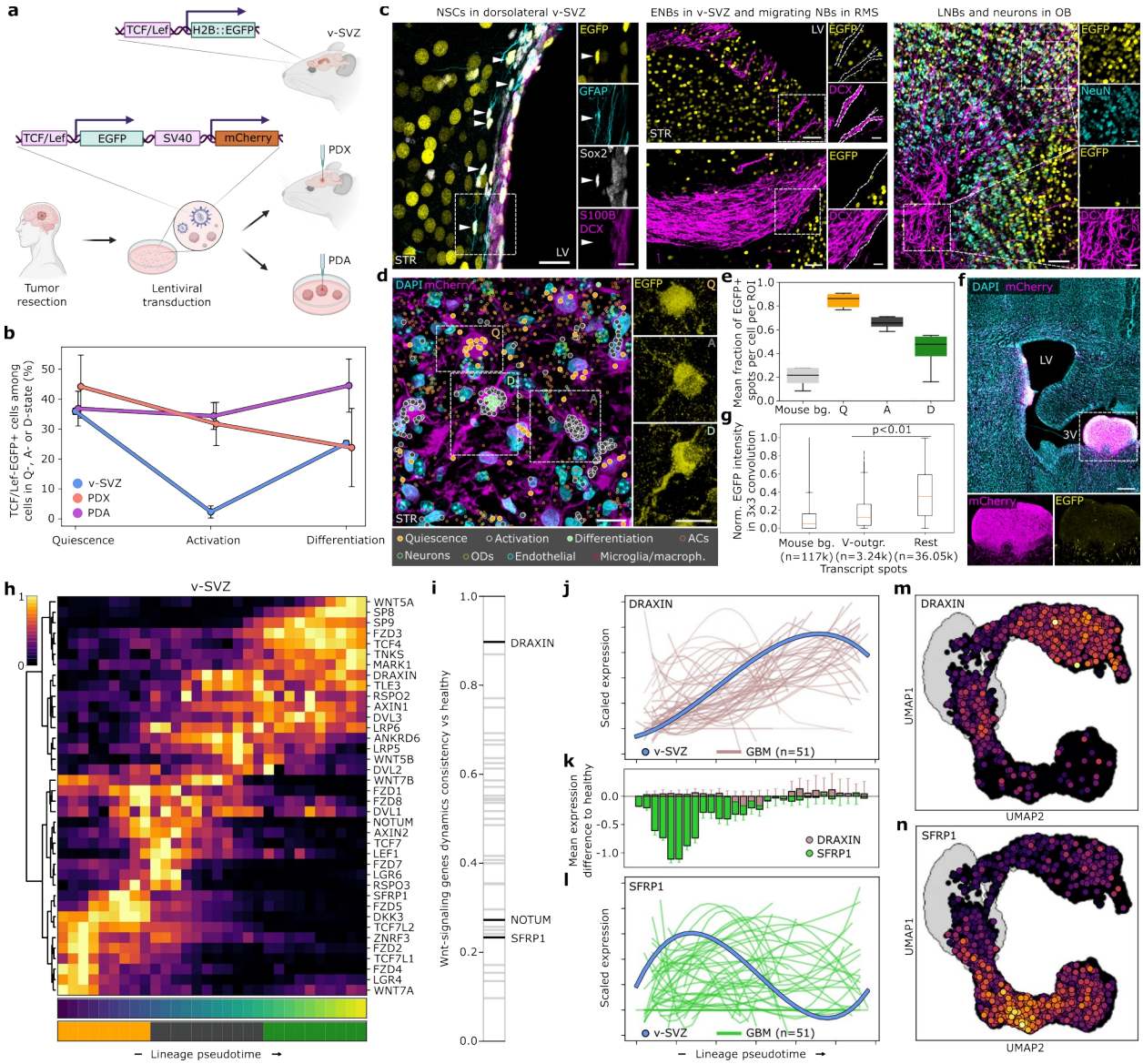


Fig. 3 | Tight regulation of Wnt signaling dynamics in v-SVZ NSCs is lost in GBM pseudolineages. a. TCF/Lef-EGFP constructs reporting canonical Wnt signaling activity in v-SVZ NSCs from a transgenic mouse line (top) and by lentiviral vector (bottom) in PDX and PDA tumors. mCherry ubiquitously labels tumor cells. **b.** TCF/Lef-EGFP activity quantified in QAD-stage v-SVZ NSCs (n=1,564 cells, n=5 replicates) and T6 PDX (n=1,176 cells, n=4 replicates) and PDA (n=12,378 cells, n=4 replicates) cells. Reporter activity was quantified by FACS and QAD-stage by scRNA-seq. Errorbars represent standard deviation in the normalized Wnt active cell-proportion. **c.** Representative immunofluorescence images of TCF/Lef-H2B::EGFP activity for v-SVZ NSCs (left; coronal view; scale bars, 20 and 10µm), early- and migrating-neuroblasts (NBs; center; sagittal view; scale bars, 50 and 20µm), and late-NBs and neurons (right; sagittal view; scale bars, 50 and 20µm). Arrows identify v-SVZ NSCs by presence of GFAP, SOX2 and absence of S100B and DCX. **d.** Representative spatial transcriptomics (Resolve biosciences) of QAD-stage cells in a T6 PDX brain. Left: striatal section of clustered cells from QAD-stages overlaid with DAPI and mCherry immunofluorescence. Tumor transcripts are colored by QAD-stage with a white outline; mouse transcripts are colored by cell type. Right: QAD-stage cells' EGFP fluorescence. Scale bars, 10µm. **e.** Mean fraction of EGFP+ spots for mouse and QAD-stage cells in n=6 spatial transcriptomics ROIs. **f.** Immunofluorescence image of a T6 PDX tumor GBM cells (mCherry) as a large ventricular growth in the third ventricle. Scale bar, 250µm. Insets highlight tumor cells devoid of Wnt activity, which is regained upon entry to the brain parenchyma (right). **g.** Normalized EGFP-intensity for mouse background (bg.) and tumor cells within (V-outgr.) and outside (rest) a ventricular growth seen in spatial transcriptomics (n=1, SFig.7f). Significance was assessed by permutation test. **h.** v-SVZ Wnt-signaling genes clustered by their scaled expression along pseudotime. **i.** Wnt-genes from (**h**) expression dynamics consistency

across n=51 GBM *ptalign* pseudotimes, with selected secreted Wnt antagonists highlighted. **j.** Splines of scaled DRAXIN and SFRP1 expression (**l**) in GBM *ptalign* pseudotimes, with v-SVZ expression dynamics in blue. **k.** Mean tumor deviation from v-SVZ reference for SFRP1 and DRAXIN expression. Errorbars represent standard deviation. **m.** Log-normalized expression of DRAXIN and SFRP1 (**n**) in v-SVZ NSC lineage UMAP, with cycling cells in gray. ENB: early neruoblast; LNB: late neuroblast; RMS: rostral migratory stream; STR: striatum; LV: lateral ventricle; OB: olfactory bulb; ACs: astrocytes; ODs: oligodendrocytes; 3V: third ventricle; Bg.: background; V-outgr.: ventricular-outgrowth.

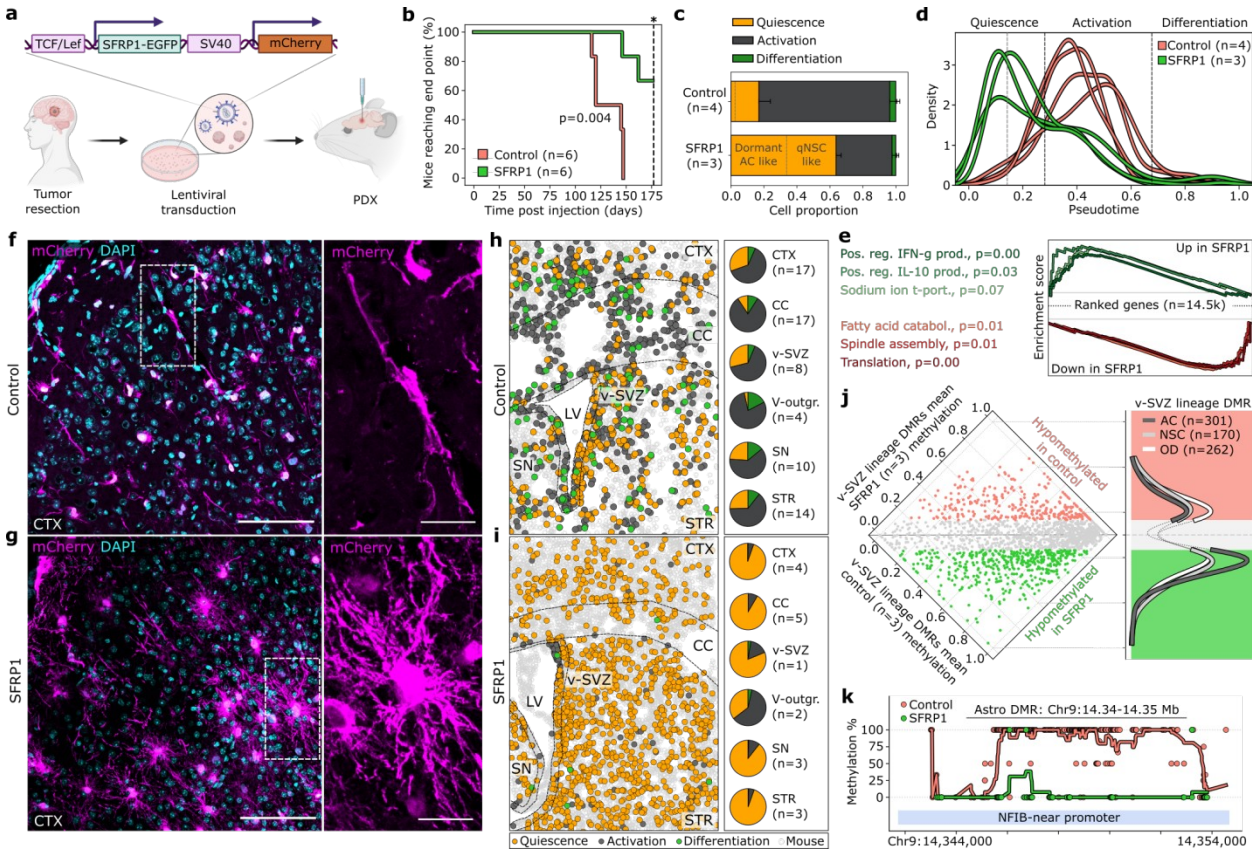
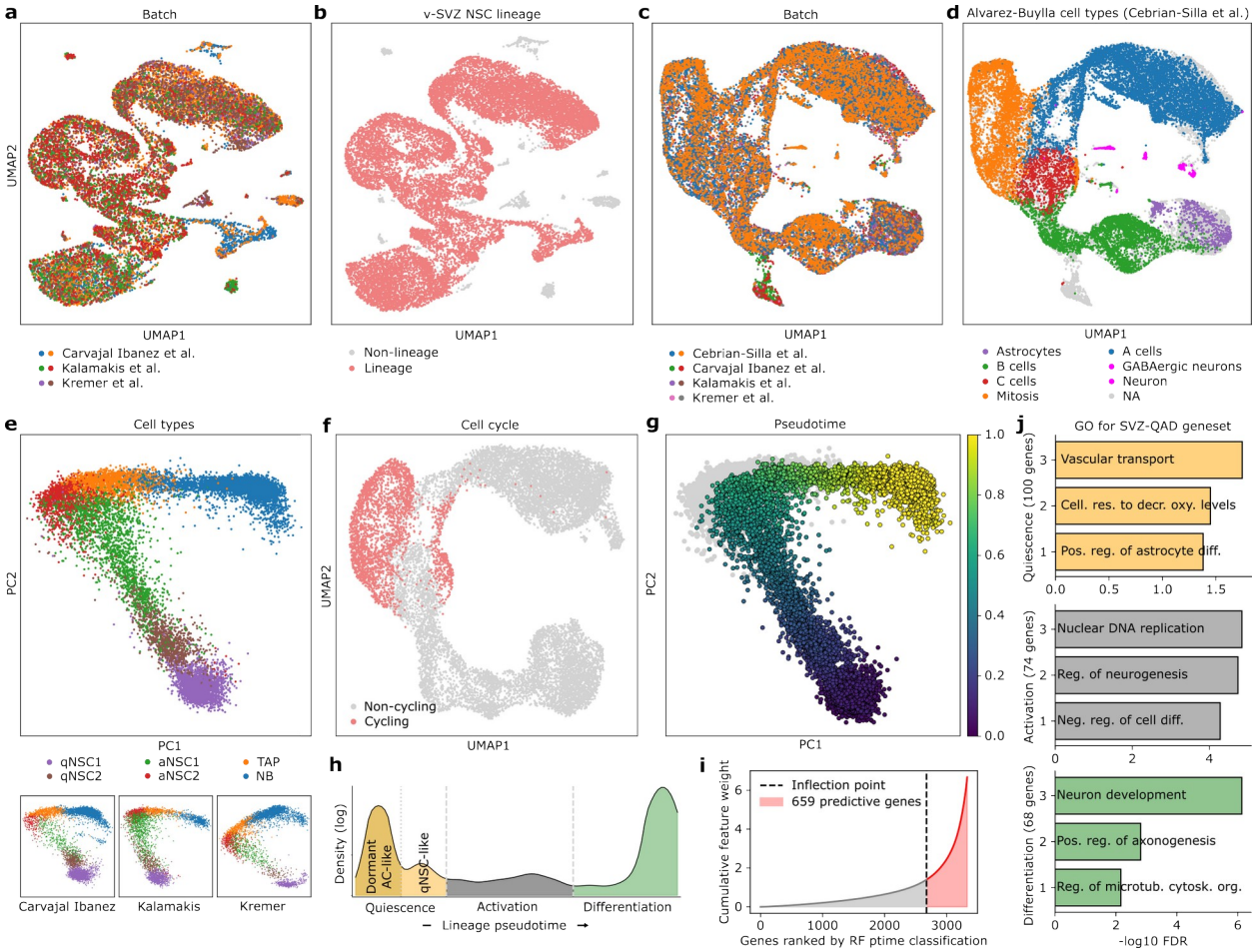
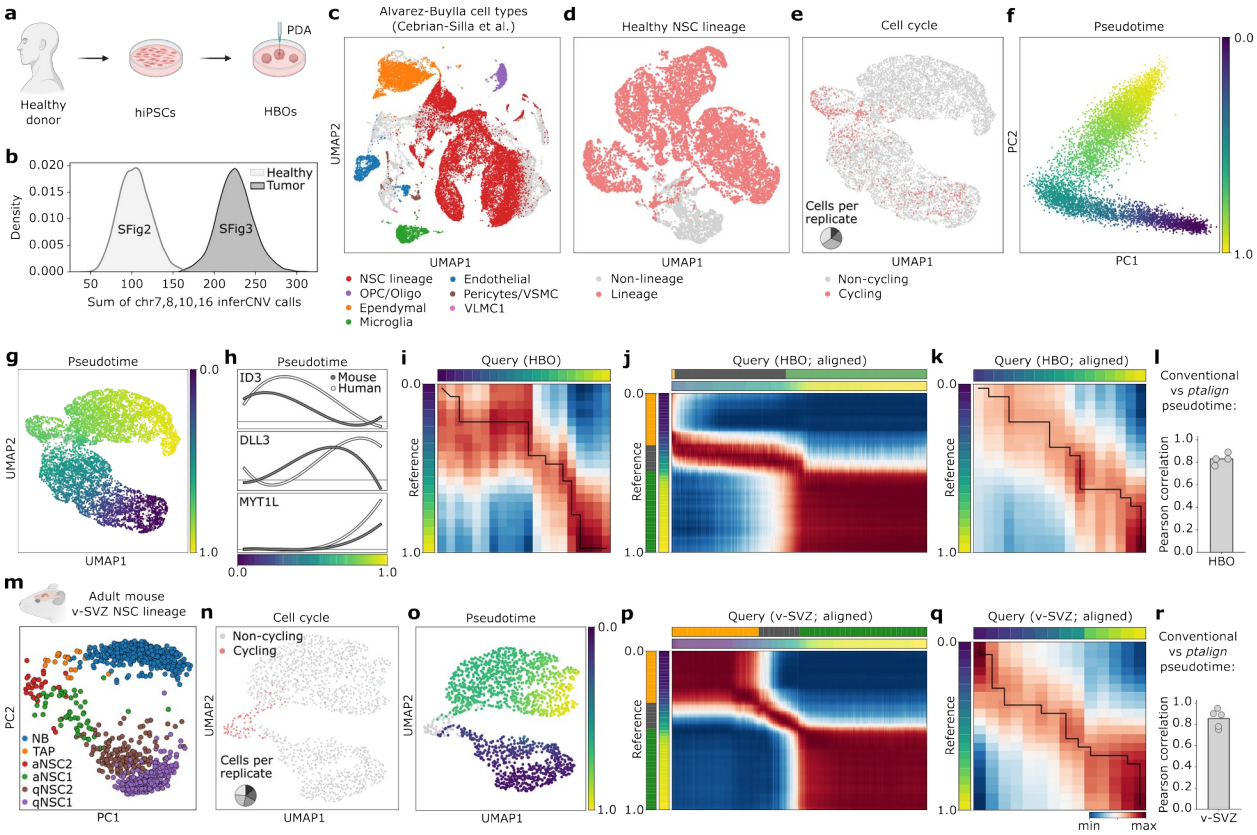


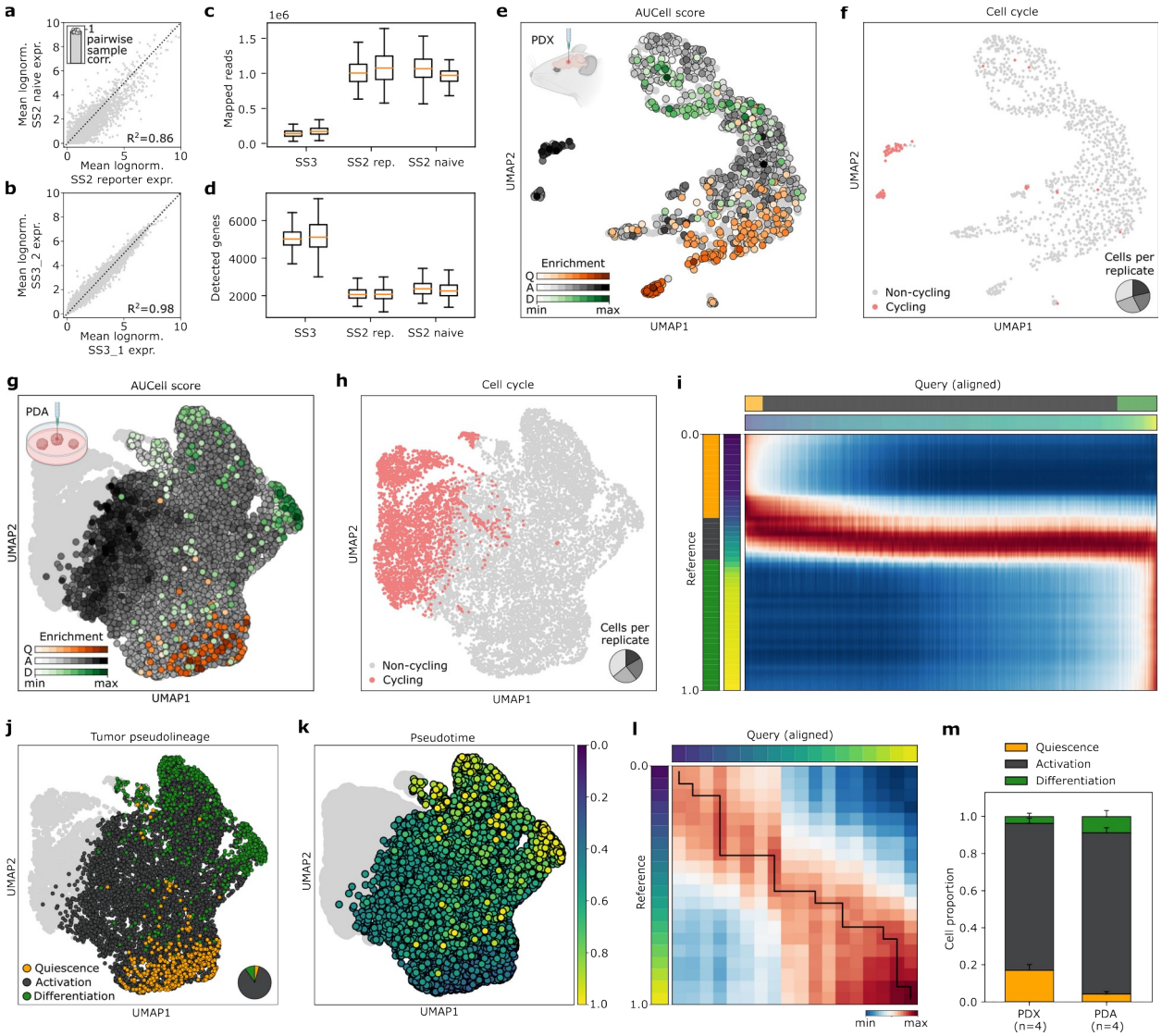
Fig. 4 | SFRP1 overexpression halts GBM pseudolineages in quiescence. a. Schematic depiction of SFRP1-overexpressing (OE) construct. mCherry ubiquitously labels tumor cells. **b.** Kaplan-Meier curve of mice reaching end point post injection for n=6 mice in control and SFRP1 cohorts. Significance was assessed by log-rank test. (*) denotes termination of experiment. **c.** Proportion of QAD-stage cells identified by *ptalign* in Smart-seq3 transcriptomes of SFRP1-OE cells (n=757, n=3 replicates) and control (n=1,176, n=4 replicates). Errorbars represent standard deviation. **d.** Cell densities along *ptalign* pseudotime by Gaussian KDE. v-SVZ QAD-stage boundaries are highlighted. **e.** Selected GSEA enrichments from genes ranked by DEseq2 Log fold-change between pseudobulked SFRP1-OE (n=3) and control (n=4 replicates). **f**. Representative immunofluorescence images of GBM cells in a control (**f**) and SFRP1-OE (**g**) PDX cortex. Scale bars 100µm, in insets 25µm. **h.** Entire spatial transcriptomics ROI depicting similar regions in SFRP1-OE (**i**) and control (**h**) PDX brains. Transcripts were associated to segmented nuclei to assign species and QAD-stage. Piecharts indicate sum of QAD-stage cells by brain region across ROIs. **j.** Left: scatterplot of mean methylation of v-SVZ celltype DMRs from Kremer *et al*. (ref[26]) in SFRP1-OE and control WGBS (n=3 technical replicates each). Differentially methylated regions are highlighted by genotype. Right: v-SVZ cell type DMRs in a Gaussian KDE over the vertical axis of the scatterplot. **k.** Selected v-SVZ DMR overlapping an NFIB-promotor in SFRP1-OE and control. Points correspond to CpG mean methylation between replicates. Lines comprise a 10-CpG methylation moving average. CTX: cortex; CC: corpus callosum; LV: lateral ventricle; V-outgr.: ventriclular outgrowth; SN: septal nuclei; STR: striatum.



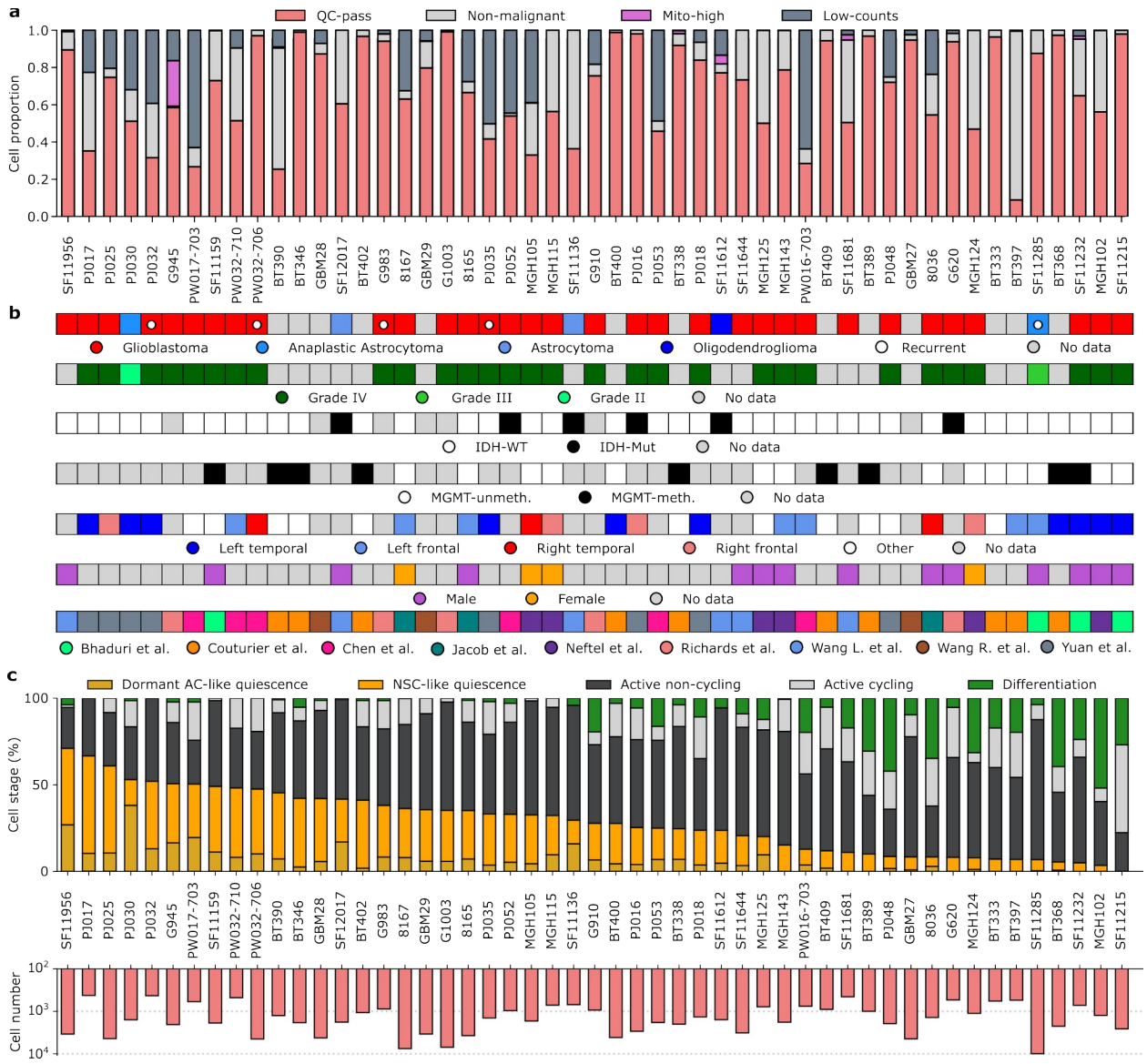
SFig. 1 | Isolation and characterization of v-SVZ NSC lineage into QAD stages. a. UMAP embedding of combined integrated v-SVZ NSC lineage datasets from n=3 published studies. **b.** Contiguous NSC lineage component isolated by DBSCAN clustering (red). **c.** UMAP embedding of integrated isolated v-SVZ NSC lineage datasets with cells from ref[24]. **d.** Cell types identified by label transfer in integrated dataset. **e.** Top: PCA of integrated v-SVZ NSC lineage cells with cell types obtained by label transfer. Bottom: isolated v-SVZ NSC lineage replicates show consistent lineage transitions captured by PCA. **f.** Cell cycle population identified in integrated v-SVZ NSC lineage (red). **g.** Diffusion pseudotime fit of integrated v-SVZ NSC lineage from (**e**), making the NSC reference pseudotime. Cycling cells are shown in gray. **h.** Cell density along NSC reference pseudotime by Gaussian KDE, with valleys denoting boundaries between QAD-stages. The Q-stage is subset into dormant astrocyte-like (AC-like) and NSC-like subsets based on the separation of qNSC1 and qNSC2 in (**e**). **i.** Selection of SVZ-QAD genes by kneedle algorithm in cumulative distribution of feature-weights in a random forest pseudotime prediction model. The 659 predictive genes were further decomposed by cell stage expression levels. **j.** Selected GO-enrichments for SVZ-QAD genes by cell stage.



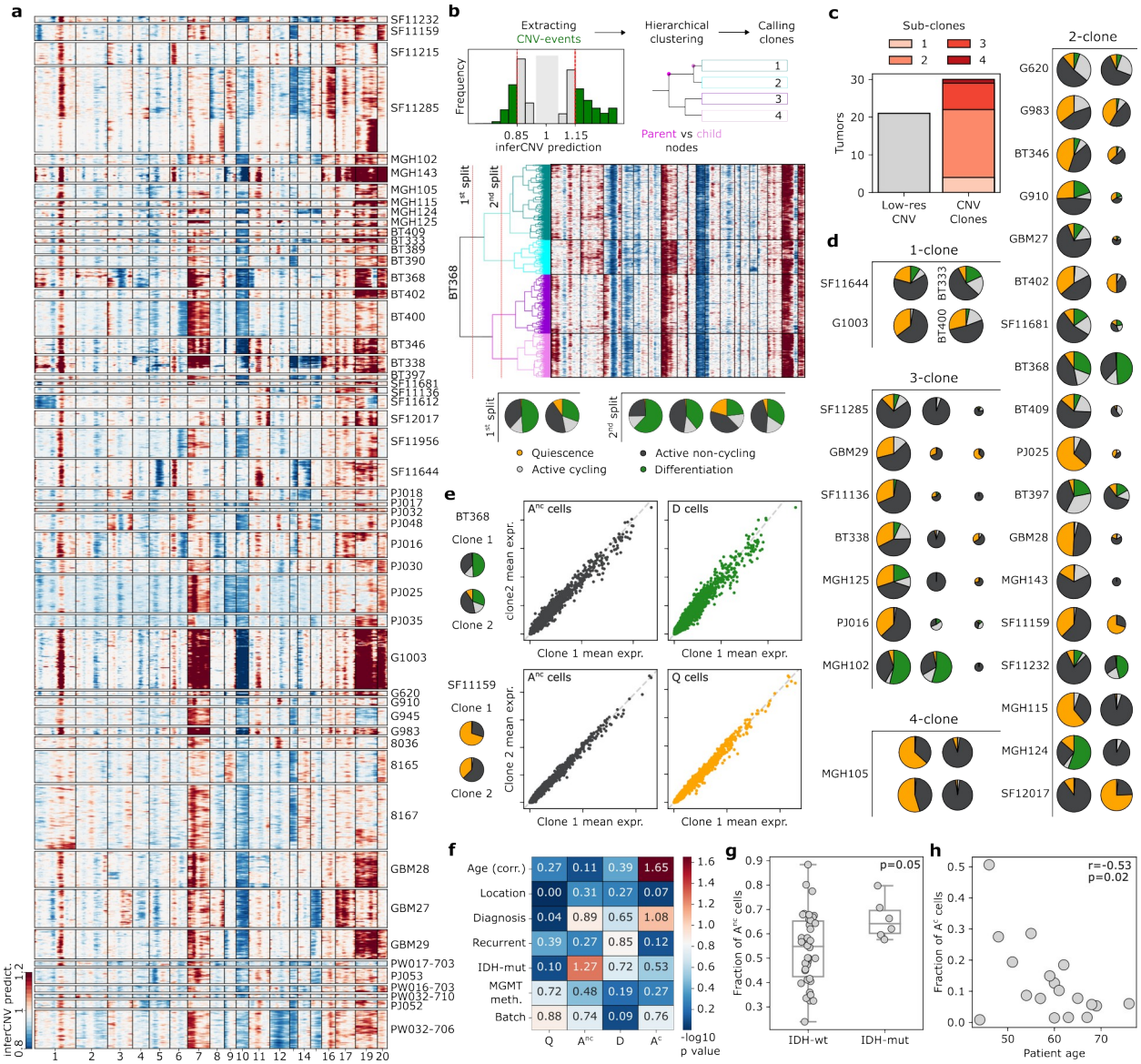
SFig. 2 | *ptalign* benchmarking on human brain organoids (HBOs) and unseen mouse NSCs. a. Schematic depicting the experimental setup (Methods). **b.** Identification of malignant and non-malignant cell populations in HBO datasets (Methods). Malignant tumor cells are addressed in SFig.3g-m. **c.** Isolation of non-malignant human organoid NSC lineage by integration with ref[24]. NSC lineage cells are highlighted in red, HBO cells in gray. **d.** UMAP of integrated NSC lineage HBO cells excluding non-lineage cells clustering with Pericytes and VLMCs in (**c**). HBO NSC lineage is highlighted in red and cycling cells are shown in (**e**). **f.** PCA and UMAP (**g**) of the integrated HBO NSC lineage reference pseudotime. Note that lineage states in PCA-space are highly concordant those in SFig.1g. **h.** Pseudotime expression splines of QAD-stage markers Id3, Dll3, and Myt1l in human and mouse NSC lineages. **i.** Human and mouse NSC lineage pseudotime DTW on SVZ-QAD geneset, with matrix traceback. **j.** *ptalign* correlation heatmap of HBO and v-SVZ NSC lineages, ordered by aligned pseudotime. **k.** DTW of human and mouse NSC lineage pseudotime dynamics inferred by *ptalign*, as in (**i**). **l.** Pearson correlation of n=4 replicates of conventional and *ptalign* pseudotimes for HBO NSC lineages. **m.** Expression PCA of n=1,564 cells from n=5 replicates of an unseen v-SVZ NSC lineage dataset with by label transfer cell types and cycling cells depicted in (**n**). **o.** UMAP of the unseen v-SVZ NSC lineage with pseudotime. **p.** *ptalign* correlation heatmap of the unseen v-SVZ NSC lineage with v-SVZ NSC lineage reference ordered by aligned pseudotime, with associated DTW (**q**). **r.** Pearson correlation in n=5 replicates of conventional and *ptalign* pseudotimes for the unseen v-SVZ NSC lineage.



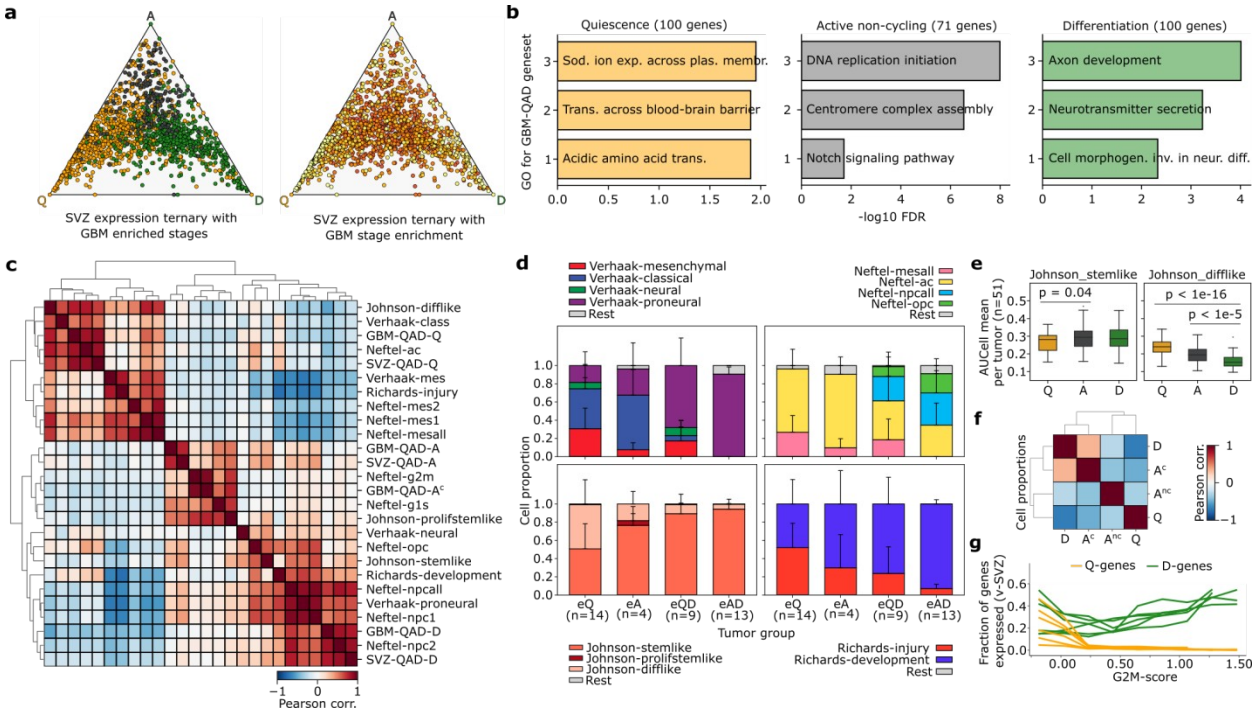
SFig. 3 | QAD stage inference in T6 PDX and PDA tumors. a. Comparison of mean log-normalized Smart-seq2 expression per gene in n=4 and n=2 replicates of Wnt-reporter and naive PDX cells, respectively. Inset barchart relates the Pearson correlation of expression between pairwise replicates by genotype. **b.** As in (**a**) for cells from n=2 replicates of T6 PDX tumors sequenced with Smart-seq3. **c.** Detected genes and mapped reads (**d**) by genotype and technology. Outliers are not shown. **e.** UMAP embedding of n=1,176 cells from n=4 replicates of integrated Wnt-reporter Smart-seq2/3 T6 PDX cells scored for SVZ-QAD genesets by AUCell and colored by the maximum-scoring cellstate, with color intensity relating the margin between the maximum and second-highest AUCell scores. **f.** Cycling cells in the integrated T6 PDX UMAP in red. **g.** AUCell scoring of T6 PDA cells (n=12,378, n=4 replicates) from SFig.2a as in (**e**), shown in UMAP. **h.** Cycling cells in the T6 PDA UMAP in red. **i.** *ptalign* correlation heatmap of T6 PDA cells and the v-SVZ NSC lineage reference ordered by aligned pseudotime, with QAD-stage assignment in (**j**) and aligned pseudotime in (**k**). The resulting *ptalign* DTW between T6 PDA cells and the v-SVZ NSC reference is given in (**l**). **m.** Barcharts depicting the mean proportion of QAD-stage cells in T6 PDX (n=4 replicates) and PDA (n=4 replicates) scRNA-seq datasets aligned to the v-SVZ NSC reference by *ptalign*. Errorbars denote standard deviation.



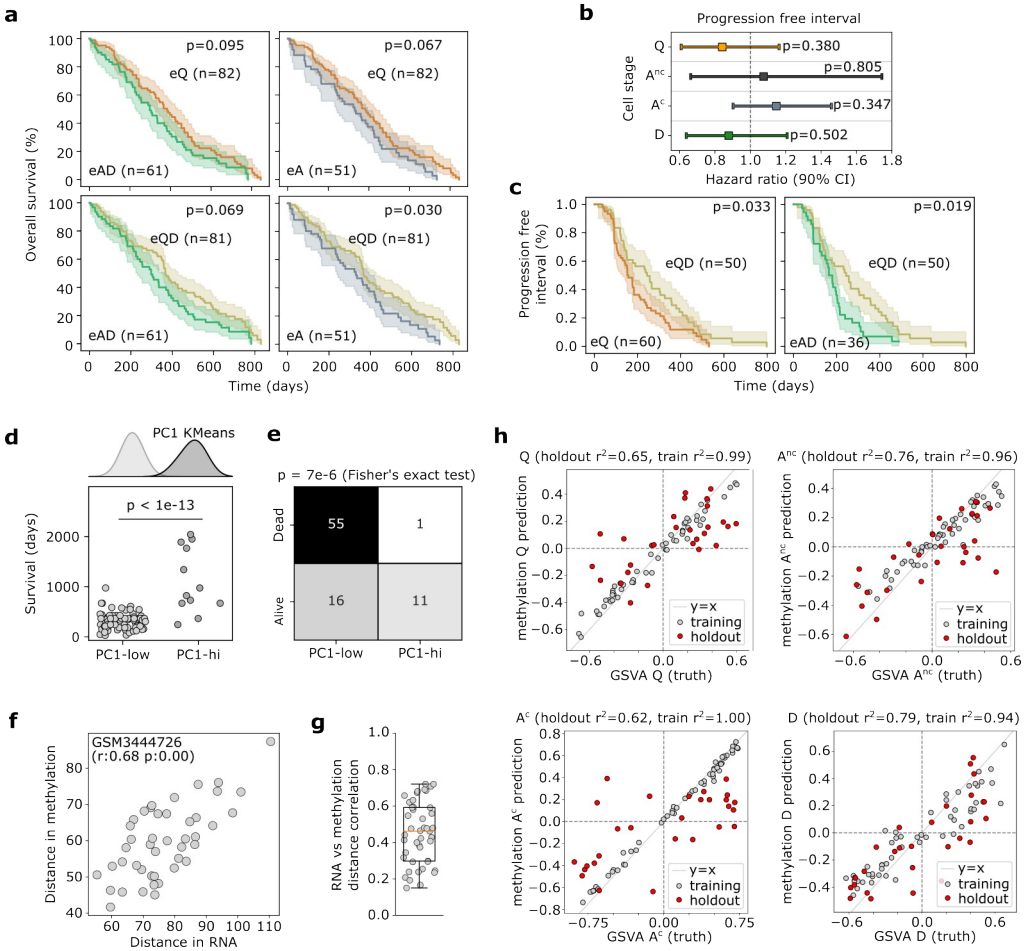
SFig. 4 | External GBM scRNA-seq datasets metadata and QAD-stages. a. Barcharts relating proportion of cells filtered in QC-steps for published primary GBM scRNA-seq cohort (n=51). **b.** Collated patient metadata across studies for tumor diagnosis, stage, IDH-mutation status, MGMT-methylation status, location, and sex. **c.** Top: stacked barcharts depicting the proportion of cells assigned to QAD stages and dormant astrocyte (AC)-like cells from *ptalign* pseudotimes, as well as cycling cells (gray), for (n=51) primary GBM scRNA-seq tumors. Bottom: number of cells assigned a *ptalign* pseudotime for each tumor. Each column in **a-c** relates information from the same tumor.



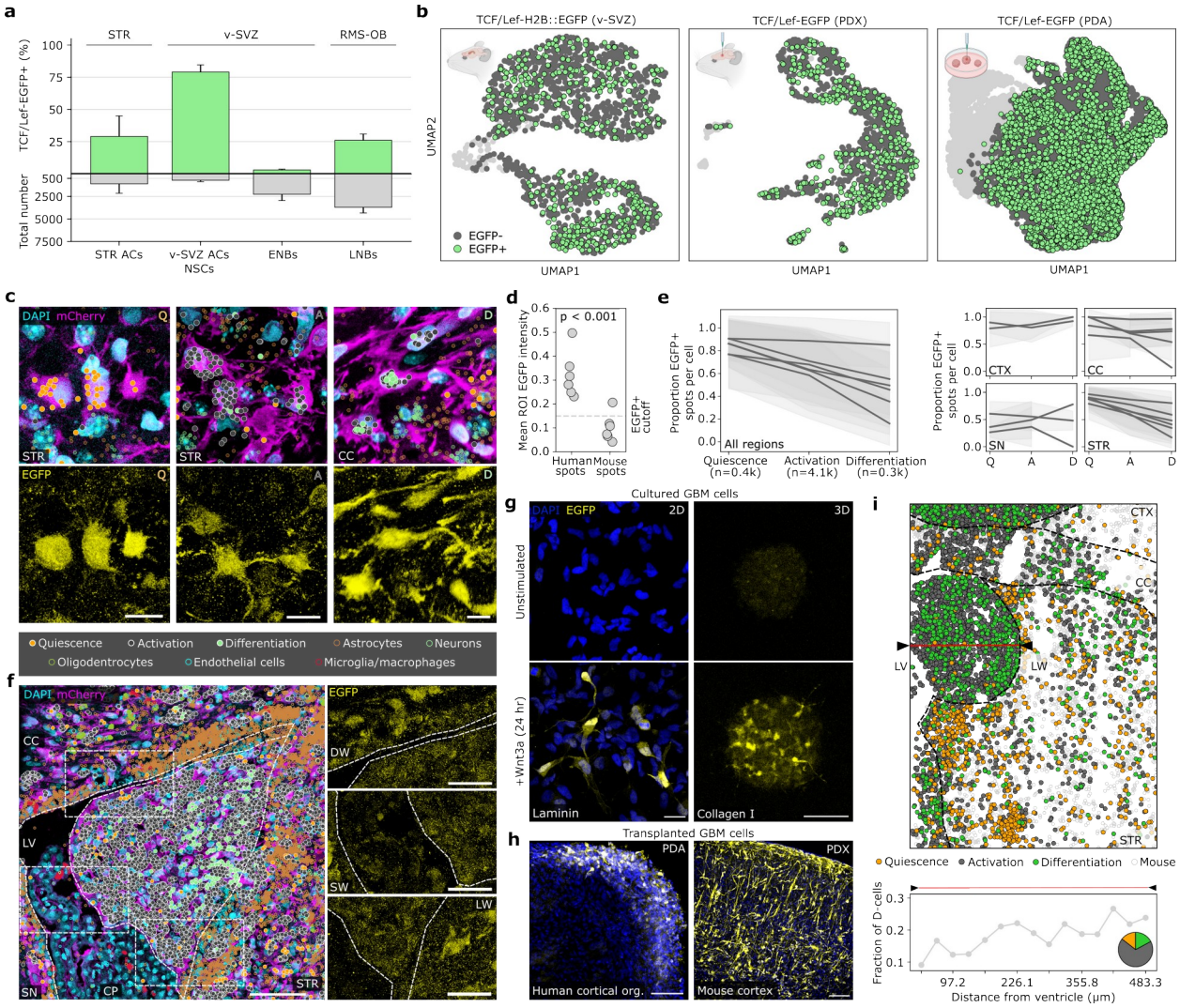
SFig. 5 | QAD-stage association with tumor features. a. Copy number variation (CNV) landscape by inferCNV of n=51 primary GBM tumors. CNV losses (blue) and gains (red) are arranged by chromosome (columns) and separated by tumor into cells (rows). Cells per tumor were hierarchically clustered by CNV-profile, and each 5th cell from the cluster-ordering is shown in the heatmap (of n=109,079 total). Sex chromosomes are not shown. **b.** Schematic representation of clone assignment for BT368. CNV-events are determined by thresholding inferCNV predictions and used to compare parent vs. child nodes in the 1st and 2nd split of a CNV-based hierarchical clustering (Methods). **c.** Barchart depicting GBM CNV-clone assignment. **d.** Proportion of QAD-stage and cycling cells per clone. Piechart size relates relative clone size. **e.** Gene expression in QAD-stage cells from two clones in SF11159 (top) and BT368 (bottom) denoted by piecharts. Scatterplots depict mean log-normalized expression of pseudobulk QAD-stage cells per clone, for all expressed genes. **f.** Heatmap of log-transformed p-values from one-way ANOVA testing for association between named feature and QAD-stage or cycling-cell proportion in n=51 primary GBM scRNA-seq datasets. **g.** Boxplots depicting the fraction of A-stage cells by IDH-mutation status. One-way ANOVA p-value (p) is indicated. **h.** Scatterplot of decreasing fraction of cycling cells with patient age for n=51 GBMs. Pearson correlation (r) and associated p-value (p) are indicated.



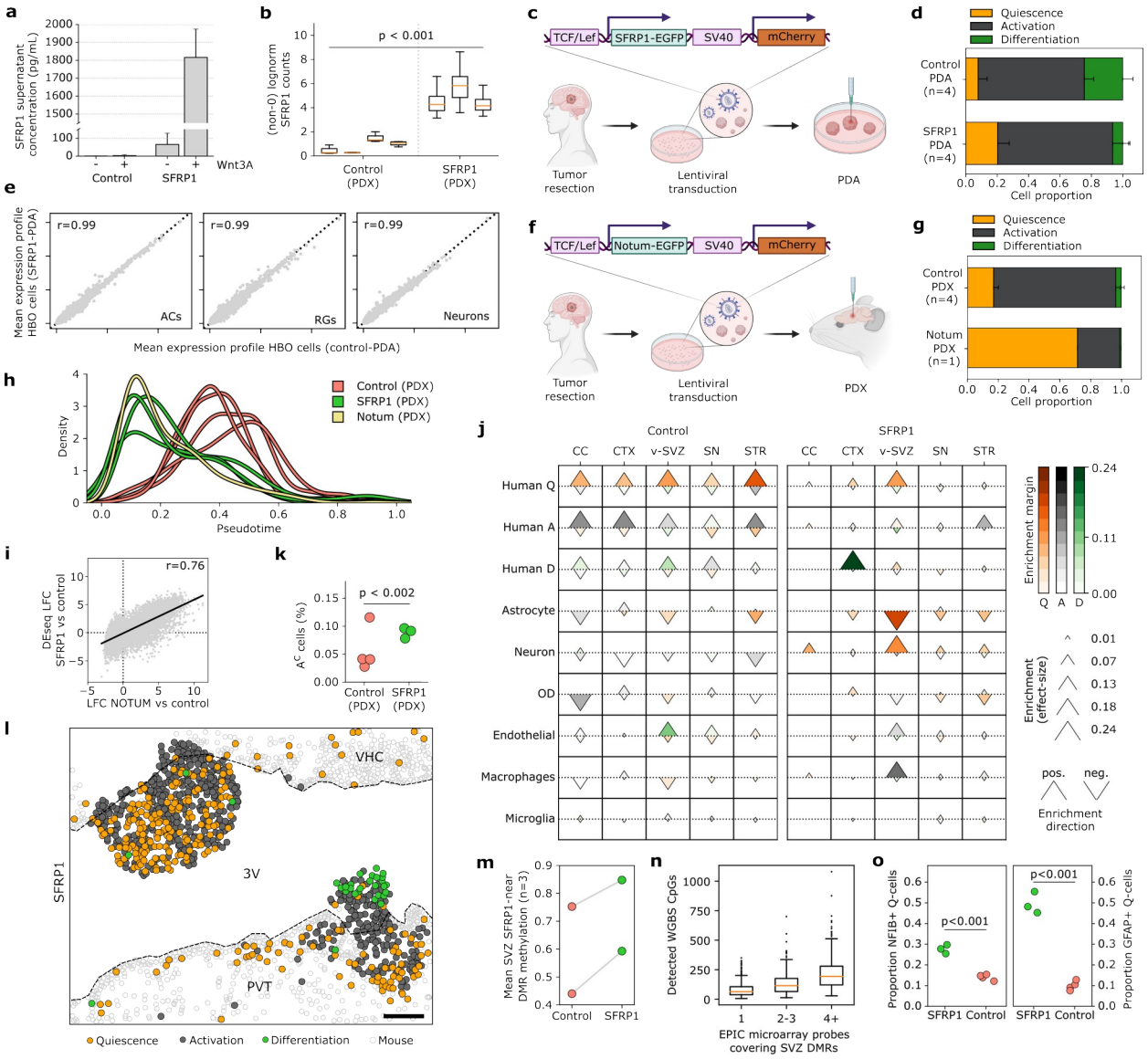
SFig. 6 | GBM-QAD geneset derivation and pseudolineage-signature comparisons. a. Setting GBM QAD-geneset by QAD-stage expression overlap between GBM and v-SVZ, in SVZ expression ternary. Genes are colored by QAD-stage (left) and stage-enrichment (right) among n=51 GBMs. **b.** Selected terms from GO-enrichment of GBM-QAD genes by stage. **c.** Hierarchical clustering of AUCell score Pearson correlations of GBM signatures on n=109,079 cells from n=51 tumors. **d.** Barcharts depicting proportion of GBM cells by state where (**c**) score>0.1 for GBM classes. Errorbars represent standard deviation. **e.** Mean AUCell score of stem-like (left) and diff-like (right) signatures from ref[43] for QAD-stages per tumor. Outliers are indicated. Significance was assessed by t-test. **f.** Hierarchical clustering of QAD-stage proportion Pearson correlations for n=51 scRNA-seq GBMs. **g.** Fraction of Q- and D-genes expressed by G2M-score among n=6 replicates of the v-SVZ NSC lineage from Fig.1a.



SFig. 7 | Clinical stratification and methylome-based inference of GBM pseudolineages. a. Pairwise overall survival for eQ,eA, eQD, and eQD bulk GBMs (cf. Fig.2c,e). Significance was assessed by log-rank test in Kaplan-meier curve with patients up to 95th survival percentile included. **b.** Predicted hazards and 90% confidence interval from a univariate Cox model of progression-free interval by GBM-QAD GSVA score, with associated p-values. **c.** Kaplan-Meier curves depicting progression-free interval for eQD compared to eQ (left) and eAD (right) GBMs. Significance was assessed by log-rank test. **d.** Methylation PCA (cf. Fig.2h) PC1 k-means groups overall survival compared by t-test (**d**) and mortality by Fisher's-test (**e**). **f.** Inter-tumoral euclidean distances in methylation- and RNA-modalities for selected tumor vs. other tumors (**f**) with Pearson correlation (r) and associated p-value (p) indicated; and per tumor (**g**). **h.** ElasticNet predictions of GBM-QAD signature scores from Fig.2b on n=55 training (gray) and n=28 holdout (red) samples for Q, A^{nc}, A^{c}, and D genesets individually. Pearson correlation (r) between prediction and target are indicated.



SFig. 8 | Canonical Wnt signaling in health and disease. a. FACS quantification of TCF/Lef- EGFP+ cells from sorted v-SVZ populations (n=3 mice). **b.** UMAPs of Wnt-reporter activity in v-SVZ (left), T6 PDX (center), and T6 PDA (right). Cycling cells are colored gray. **c.** Selected Q (left), A (center), and D (right) cells in spatial transcriptomics with transcripts overlaid with DAPI and mCherry (top) or EGFP (bottom) immunofluorescence. Scale bars, 10µm. **d.** EGFP+ spot threshold from n=6 spatial transcriptomics ROIs. Significance was assessed by t-test. **e.** Proportion of EGFP+ spots by QAD-stage for all (left) or individual (right) regions with >20 cells among spatial ROIs. **f.** Left: Wnt-reporter spatial transcriptomics ROI of a ventricular outgrowth. Transcriptomic markers are overlaid with DAPI and mCherry immunofluorescence. Tumor transcripts are colored by QAD-stage with a white outline, mouse transcripts are colored by cell type. Scale bar, 50µm. Right: contrasting EGFP fluorescence at dorsal- , septal-, and lateral-walls of the ventricle. Scale bars, 25µm. **g.** Wnt-reporter GBM cells were cultured on laminin coated chambers (left; fixed, stained; scale bar, 25µm) or as 3D spheroids in collagen matrix (right; live, unstained). Images were acquired 24h post recombinant-Wnt3a treatment. Scale bar, 100µm. **h.** Wnt-reporter GBM cells transplanted into HBOs (left, 10dpi) or mouse brains (right, 5mpi). Scale bars, 100µm. **i.** Top: large ventricular growth in a non-Wnt-reporter spatial transcriptomics ROI with transcripts associated to segmented nuclei to assign species and QAD-stage. Red line traces largest extent from LW to LV. Bottom: increasing D-cell fraction with distance from ventricle. ENB: early neuroblast; LNB: late neuroblast; STR: striatum; RMS: rostral migratory stream; OB: olfactory bulb; AC: astrocyte; CC: corpus callosum; CTX: cortex; SN: septal nuclei; LV: lateral ventricle; CP: choroid plexus; DW: dorsal-wall; SW: septal-wall; LW: lateral-wall.



SFig. 9 | GBM pseudolineage assessment for Wnt-based interventions. a. SFRP1 concentration by ELISA in supernatant of cells stimulated with/without Wnt3A. Errorbars represent standard deviation in n=3 replicates. **b.** SFRP1 expression levels in SFRP1-expressing cells, by genotype. Significance was assessed by t-test. **c.** SFRP1-OE construct used in T6 PDA tumors. mCherry ubiquitously labels tumor cells. **d.** Proportion of QAD-stage cells in SFRP1-OE (n=688 cells, n=4 replicates) and control (n=476, n=4 replicates) PDA cells. Errorbars represent standard deviation. **e.** Mean log-normalized expression for n=8,876 cells in n=4 replicates each of control and SFRP1-OE PDA astrocytes (left), radial glia (center) and neurons (right), with Pearson correlation (r) between genotypes. **f.** NOTUM-OE construct as in (**c**), with QAD-stage cell proportions (**g**) for n=1 replicates of NOTUM-OE PDX, as in (**d**), with pseudotime cell density (**h**). **i.** Mean NOTUM-OE log fold-change against SFRP1-OE DEseq2 log fold-change. Black line indicates best fit, Pearson correlation (r) is indicated. **j.** Neighborhood enrichment by permutation (Methods) for segmented cells in n=17 Wnt-reporter and n=6 SFRP1-OE spatial transcriptomics ROIs. Cell type enrichments are reported for QAD-stage cells with other cell types (rows), with enrichment direction by shape and strength by size while color relates the enrichment margin between the largest and second-largest enrichments. **k.** Proportion of cycling (A^c) cells in control and SFRP1-OE. Significance was assessed by t-test. **l.** Zoomed SFRP1-OE spatial transcriptomics ROI with QAD-stage cells in two ventricular outgrowths. Transcripts were associated to segmented nuclei to assign species and QAD-stage. Scale bar, 10µm. **m.** Mean methylation at SFRP1-overlapping DMRs for n=3 technical replicates of control and SFRP1-OE WGBS. **n.** Number of CpGs detected in WGBS for DMRs grouped into those covered by 1, 2-3, or 4+ probes in Illumina EPIC 850k methylation microarray. **o.** Proportion of NFIB+ (left) and GFAP+ (right) Q-cells among PDXs. Significance was assessed by t-test. AC: astrocyte; RG: radial glia; CTX: cortex; CC: corpus

callosum; SN: septal nuclei; STR: striatum; 3V: third ventricle; VHC: ventral hippocampal commissure; PVT: periventricular nuclei of the thalamus.

Data and Code Availability

The *ptalign* software and jupyter notebooks detailing the analyses and figures presented in this study will be made available on GitHub prior to publication. Our generated SmartSeq3 data from the mouse v-SVZ NSCs lineage as well as processed data from tumor PDX and PDA single-cell RNA-sequencing and WGBS datasets are available from GEO with accession GSE240676. Raw tumor data subject to DPA are submitted to EGA. Previously published scRNA-seq data of the v-SVZ which we re-analyzed in this study were retrieved from the following sources: Cebrian-Silla *et al*.[24] (https://cells.ucsc.edu/?ds=svzneurogeniclineage), Kremer *et al*.[22] (GSE145172), Carvajal-Ibanez *et al*.[21] (GSE197217), and Kalamakis *et al*.[20] (GSE115626). Spatial transcriptomics data we generated are submitted to Zenodo under https://zenodo.org/deposit/8186500. Analyzed GBM scRNA-seq datasets which had been previously published were retrieved as follows: Richards *et al*.[28] (https://singlecell.broadinstitute.org/single_cell/study/SCP503), Bhaduri *et al*.[39] (http://cells.ucsc.edu/?ds=gbm, with raw counts received in personal communication), Neftel *et al*.[18] (https://singlecell.broadinstitute.org/single_cell/study/SCP393), Yuan *et al*.[29] (GSE103224), Couturier *et al*.[31] (https://datahub-262-c54.p.genap.ca/GBM_paper_data/GBM_cellranger_matrix.tar.gz), Wang L. *et al*.[30] (GSE138794), Jacob *et al*.[33] (GSE141946), Wang R. *et al*.[32] (GSE139448), and Chen *et al*.[34] (GSE141383). Samples and metadata from the TGCA-GBM cohort are available through the NCI GDC Data Portal (https://portal.gdc.cancer.gov/), while curation of legacy TCGA data is detailed in the Methods. Bulk GBM data from Wu *et al*.[41] was retrieved from GSE121723, with raw counts received in personal communication. All data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

We gratefully acknowledge the following DKFZ core facilities for their support and expert assistance: Center for Preclinical Research, Small Animal Imaging Center, Genomics and Proteomics (Microarray Unit and Sequencing Open Lab), Single Cell Open Lab, Flow Cytometry and Light Microscopy. We are thankful to the Division of Biostatistics for statistical support. We further thank the following members of the Division of Molecular Neurobiology for their technical support: Santiago Cerrizuela, Damian Carvajal Ibanez, Jooa Hooli, Susanne Kleber, Katrin Volk, Kathrin Menzner, Mohammad Al-Shukairi, Heike Abendroth, Aylin Korkmaz and all the members of our group for continuously providing critical feedback on the project as well as the manuscript. Many thanks to the members of the Somatic Evolution and Early Detection group for their critical feedback, particularly Pablo Baeza Centurion for his insightful comments on the ptalign methodology. 7TGC plasmid was a gift from Roel Nusse; RKRR-P2A plasmid was a gift from Edward Green. O.K was funded by the Helmholtz International Graduate School for Cancer Research. This work was supported by the German Cancer Research Center (DKFZ), European Research Council (ERC; REBUILDCNS), and the German Research Foundation (DFG; SFB 1324). Last but most important, we are grateful to patients and their families for their generous donations to biomedical research.

Author contributions

O.K., L.C.F. and A.M.-V. developed the concept, designed the experiments, interpreted the data and wrote the manuscript. O.K, M.B., K.C.Z., V.A., N.S., and N.G.P. performed the experiments. L.C.F. performed computational data analysis and developed *ptalign*. X.M. and H.L. provided HBOs. A.S. and C.O. curated the TCGA data. V.W. and S.A. performed computational data analysis for the spatial transcriptomics experiments. C.R.W., K.P. and P.U.L. provided patient material and clinical data. A.M.-V., A.G. and S.A provided supervision and training. A.M.-V. provided oversight.

Methods

Cells

Patient tumors and tumor sphere culture

Primary tumor samples were received from the University Hospital Ulm upon obtaining informed consent prior to surgery. Experiments involving patient tumor biopsies were carried out in accordance with the Decleration of Helsinki and were approved by the ethics committees of Heidelberg University, Medical Faculty Mannheim (S-224/2021). All tumor specimens were examined by a neuropathologist to ensure that the tumors met GBM criteria defined by the World Health Organiyation. Upon arrival, fresh tissues were immediately dissociated using Brain Tumor Dissociation Kit (P) and expanded in culture. Tumor spheres were maintained in serum-free Neurobasal A medium supplemented with B27, heparin (2 µg/ml) and the stem mitogens EGF (20 ng/ml) and bFGF (20 ng/ml) at 5% CO_2 and 37°C. For passaging, tumor spheres were enzymatically dissociated into single cells using Accutase when sphere size reached approximately 100 µm in diameter approximately once in a week.

Culture of human 293T cells

293T cell line was used for production and titration of the lentiviral particles. 293T cell line is a fast growing, highly transfectable derivative of human embryonic kidney 293 cells. Due to the presence of SV40 large T antigen, they allow high levels of proteins to be expressed therefore are ideal for generating high-titer lentiviral particles.

293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FCS, 1% penicillin/streptomycin (10,000 units/mL) and 1% L-Glutamine (200 mM) at 37C and 5% CO_2.

Lentiviral Transduction

Human GBM cells were transduced with the 7TGC (Addgene plasmid #24304), 7TGC-SFRP1 or 7TGC-Notum lentiviral vectors at a multiplicity of infection (MOI) of 5. Expression of the transduction marker (mCherry) was confirmed by FACS analysis. Second-generation, self-inactivating, replication deficient lentiviral particles were produced, purified and titrated as described previously[1].

Immunocytochemistry

For 2D culture of human GBM cells, $3x10^5$ cells in 0.3 mL of Neurobasal A medium were seeded into each well of µ-Slide 8-well chamber (Ibidi). Prior to seeding, the wells were coated with 10 µg /ml Poly-D-Lysine overnight at the room temperature, washed 3 times with DPBS and subsequently coated with 10-20 µg/ml Laminin for 2 hours at 37 °C. The cells were treated with a recombinant Wnt3a (200 ng/mL dissolved in 0.1% bovine serum albumin (BSA) in distilled water). 0.1% BSA was used in carrier-only conditions as baseline controls. Following 24 hr post-treatment, the cells were washed 3 times with DBPS and fixed with 2% PFA for 20 minutes at the room temperature. Next, the fixed cells were incubated in permeabilization/blocking solution

(0.25% Triton-X 100 in PBS supplemented with 3% horse serum and 0.3% BSA) for one hour at room temperature to block unspecific binding sites. Subsequently cells were stained with anti-GFP antibody in blocking buffer overnight at 4°C. After washing 3 times 5 min with 0.25% Triton-X 100 in PBS, secondary antibody staining was carried out using appropriate Alexa-fluorophore antibodies diluted in blocking buffer for 1 hour at the room temperature. DAPI was added to the antibody cocktail for nuclear staining. Furthermore, cells were washed 3 times 15 min with 0.25% Triton-X 100 in PBS. After mounting with Fluoromount-G, slide was allowed to dry in dark at room temperature for 30 min. Samples were stored in the dark at 4°C until imaging.

For 3D culture of human GBM cells, individual tumor spheres of approximately 100 µm in diameter were seeded in 0.1 mL of Collagen matrix containing 0.27 mg of Collagen type I (Corning; rat tail) supplemented with 10% minimum essential medium (MEM; 10x) and 20% NaOH (0.1 M) for pH adjustment in a well of 10-well cell view cell culture slide (Greiner Bio-one). Collagen matrix was allowed to polymerize for 1 hr at 37°C and 0.3 mL of Neurobasal A medium with recombinant Wnt3a (200 ng/mL in 0.1% BSA in distilled water) or carrier-only (0.1% BSA in distilled water) was added for 24 hr before live-imaging.

Confocal images were acquired on a Leica TCS SP5 or SP8 confocal microscope at the light microscopy facility of the German Cancer Research Center. ImageJ was used for processing and images were only adjusted for brightness and contrast.

Molecular cloning of 7TGC-SFRP1 and 7TGC-Notum constructs

Human SFRP1 (Origene plasmid #RC207328) or Notum (Gateway ORF clone ID #164485821) was clonned in frame under 7xTCF promoter (7TGC; Addgene plasmid #24304) upstream of EGFP sequence using In-Fusion HD cloning kit according to manufacturer's instructions. A furine cleavage site (RKRR) as well as a self-cleaving peptite linker (P2A) was incorporated between SFRP1/Notum and EGFP to allow seamless protein translation avoiding fusions (S8.C and F). 7TGC plasmid was a gift from Roel Nusse. P2A-RKRR plasmid was a gift from Edward Green. Notum plasmid was obtained through the vector and clone repository[2] of the genomics and proteomics facility of the German Cancer Research Center. Constructs cloned as part of this study are available from the investigators upon request.

Enzyme-linked immunosorbent assay (ELISA) for SFRP1

T6 Wnt-reporter or SFRP1-OE GBM cells ($5x10^5$ per well) were seeded in a 6-well plate in Neurobasal A medium. The cells were treated with a recombinant Wnt3a (dissolved in 0.1% bovine serum albumin (BSA) in distilled water at a final concentration of 200 ng/mL). 0.1% BSA was used in carrier-only conditions as baseline controls. Following 24 hour post-treatment, the supernatant was collected and the ELISA protocol was followed according to manufacturer's instructions (Abcam; ab277082). The optical density (OD) of the samples was measured at 450 nm with the plate reader (Agilent Biotek). The respective concentration of each sample was calculated based on the standard curve generated by known protein concentrations and measured optical density (OD) at 450 nm following background subtraction according to manufacturer's instructions (Abcam; ab277082).

Culturing and maintenance of human iPSCs

The human iPSCs were acquired from Allen Institute for Cell Science (AICS-0036-006, https://www.allencell.org/cell-catalog.html). The AICS-0036-006 line was labeled with EGFP via CRISPR/Cas9 system, where mEGFP transgene was knocked in chromosome 19q13-qter. Their parental cell line (https://www.coriell.org, GM25256) was acquired from Coriell (UCSFi001-A, https://labs.gladstone.org/conklin/) and was derived from a healthy donor using an episomal reprogramming method[3].

iPSCs were feeder-independent and Matrigel (Corning, 354277) was employed to coat the plates or dishes. iPSCs were maintained on mTeSR1 medium (Stem Cell Technologies, 05850 or 100-0276) at 37˚C and 5% CO_2. ReLeSR™ (Stemcell Technologies, 05872) was used for passaging cells, with media change every other day.

Human brain organoid generation

The procedure was adapted from a previously published method[4,5,6]. Briefly, on day 0, hiPSCs were first dissociated into single cells and seeded in ultra-low attachment round-bottom 96-well plates (Corning, 7007), containing human embryonic stem cell (hESC) medium supplemented with 4 ng/ml bFGF (PeproTech, AF-100-18B-50) and 50 µM Rho-associated protein kinase (ROCK) inhibitor (Merck Millipore, SCM-075). On day 3, the medium was replaced with fresh hESC medium. From day 5, the embryoid bodies were transferred to ultra-low attachment 24-well plates (Corning, 3473) in Neural Induction Medium. On Day 11, organoids were embedded into droplets of Matrigel (Corning, 354234) and transferred into 6 well plates in another two-day Neural Induction Medium. Differentiation Medium without vitamin A was fed from day 13 to day 19. From day 20 on, organoids were cultured in Differentiation Medium with vitamin A. Agitation was employed from day 18 at 70 RPM and the medium was changed every 2-3 days.

Injection of GBM cells for PDA generation

For injection of lentivirally transduced primary GBM cells into human brain organoids, $5x10^4$ cells were cultured as single cell suspension overnight. Newly formed spheres were resuspended in 1 µl of Differentiation Medium with vitamin A, loaded into a NanoFil syringe and injected into the core of 2 month-old organoids under a dissection microscope. Tumor bearing organoids were maintained in Differentiation Medium with vitamin A on an orbital shaker (70 RPM) for 15 days at 5% CO_2 and 37°C.

<u>Animal Experiments</u>

Mouse strains

Male TCF/Lef:H2B/GFP mice[7] were bred at the Center for Preclinical Research at the German Cancer Research Center and were used for analyzing TCF/Lef-activity in mouse NSC studies. Male Fox Chase SCID Beige mice (CB17.Cg-Prkdc^{scid}Lyst^{bg-J}/Crl) were purchased from Charles River and were used to generate human-mouse xenograft tumors. Experimental mice had *ad libitum* access to food and water and were housed in specific pathogen-free, light (12 hr day/night cycle), temperature (21°C) and humidity (50–60% relative humidity) controlled

conditions. In addition, daily monitoring of mice for symptoms of disease determined the time of sacrificing for injected animals with signs of distress. All procedures were approved and conform to the regulatory guidelines of the official committee (Regierungspräsidium Karlsruhe, Germany; G19/21).

Orthotopic injection of GBM cells for PDX generation

For ortotopic injection of lentivirally transduced (or untransduced control) primary GBM cells into the mouse brain, $5x10^5$ cells were cultured as single cell suspension overnight. Newly formed spheres were resuspended in 2 µl of Matrigel, loaded into a NanoFil syringe and stereotactically injected into the striatum (2.0 mm lateral to the bregma at a depth of 3.0 mm) of 8-10 week-old Fox Chase SCID-Beige mice under general anesthesia and perioperative pain management. Tumor growth was longitudinally monitored by magnetic resonance imaging (MRI) at the Small Animal Imaging Center of the German Cancer Research Center. Upon reaching termination criteria outlined in the approved animal experiment application, mice were sacrificed and brains were collected after transcardial perfusion.

For perfusion, mice were anesthetized by intraperitoneal injection of perfusion solution. After opening the thoracic cavity exposing heart, transcardial perfusion was carried out with ice cold HBSS.

Flow cytometry analysis and fluorescence-activated cell sorting (FACS)

For quantification of TCF/Lef-EGFP reporter in the healthy v-SVZ-lineage and striatal astrocytes; striatum, v-SVZ, rostral migratory stream (RMS) and olfactory bulb (OB) (n=3 mice) were dissected and single cell suspensions were prepared using Neural Tissue Dissociation Kit (P) and gentleMACS Octo Dossiciator with Heaters (Miltenyi) according to manufacturer's instructions. To allow discrimination of different populations, single cell suspensions were stained with the following antibodies conjugated with appropriate fluorophores as previously reported[8]: anti-CD45, anti-Ter119, anti-O4, anti-Glast, anti-CD133 (Prom1) and anti-PSANcam including proper unstained, single stained and fluorescent minus one (FMO) controls. After excluding dead cells, doublets and CD45+/Ter119+/O4+ non-lineage populations, striatal astrocytes were gated based on Glast+vity; NSCs in the v-SVZ were gated based on Glast+vity and CD133+vity; early neuroblasts (ENBs) in the v-SVZ as well as late neuroblasts (LNBs) in the OBs were gated based on Glast-vity, CD133-vity and PSANcam+vity. FACS analysis was performed on a BD Fortessa at the Flow Cytometry Core Facility at the German Cancer Research Center.

For isolation of TCF/Lef-EGFP reporter healthy v-SVZ-lineage, single cell suspensions (n=5 mice) were prepared and stained as outlined above. Each population was index-sorted into 384 well-plates (Eppendorf Lobind) as single cells. Microplates containing cell lysates were briefly centrifuged, snap frozen on dry ice and were stored at -80°C until processed for Smart-seq3.

For isolation of the xenografted human GBM cells, tumor bearing mouse brains (n=6 mice) were dissected and single cell suspension was prepared using Brain Tumor Dissociation Kit (P) and gentleMACS Octo Dissociator with Heaters (Miltenyi) according to manufacturer´s instructions.

Transduction marker mCherry was used to discriminate transduced human GBM cells while a human specific-MHCI antibody was employed to discriminate untransduced human GBM cells. After excluding dead cells and doublets, MHCI+ or mCherry+ (EGFP+ and EGFP-) cell population was index-sorted into 384 well-plates (Eppendorf Lobind) as single cells. Microplates containing cell lysates were briefly centrifuged, snap frozen on dry ice and were stored at -80°C until processed for Smart-seq3.

For isolation of human brain organoid and allografted human GBM cells, tumor bearing organoids were dissociated using Brain Tumor Dissociation Kit (P) and gentleMACS Octo Dossiciator with Heaters (Miltenyi) according to manufacturer's instructions. Prior to multiplexing (n=4 HBOs), single cell suspensions were labelled with TotalSeqA anti-human hashtag oligonucleotide (HTO) antibodies (Biolegend, 1:50) to allow downstream de-multiplexing in silico as previously described[9]. After excluding dead cells and doublets, mCherry+ as well as mCherry- cell populations were sorted as bulk populations into 1.5 ml tubes (Eppendorf) and immediately processed for single cell RNA-sequencing using 10X Chromium 3' sequencing platform.

Sytox Blue (Life Technologies, 1:1000) was used in all experiments as a dead cell indicator. Index sorting of single cells as well as sorting of bulk samples was carried out using a 100 micron nozzle at a BD FACSAria II or BD FACSAria Fusion at the Flow Cytometry Core Facility at the German Cancer Research Center.

Single cell library preparation with Smart-seq2/3

Single cell RNA-sequencing (scRNA-seq) libraries from v-SVZ-lineage populations were prepared using Smart-seq3 protocol as previously described (Hagemann-Jensen, Nature Biotechnology, 2020) with minor modifications. The protocol was automated and miniaturized by incorporating liquid handling platforms including Mosquito HV (STPLabtech), Mantix (Formulatrix) and Viaflo 384 (Integra). Briefly, plates were incubated at 72°C for 10 min to facilitate lysis and denaturation of secondary structures in the RNA. Lysed cells were subjected to reverse transcription in 2 µl using Maxima H-minus reverse transcriptase (Thermo Scientific), an oligo(dT) primer and a template-switching oligonucleotide encompassing an 8-bp unique molecular identifier (IDT). In a randomly selected set of wells, we included ERCC Spike-ins (Ambion) at a 1:2500000 dilution. Full length cDNAs were amplified for 22 cycles of PCR using KAPA HiFi DNA polymerase (KAPA Biosystems). cDNA samples were purified with Ampure XP beads at a 1:0.8 ratio and cDNA quality in randomly selected 10 wells/plate was assessed on a High Sensitivity Bioanalyzer chip (Agilent). cDNA concentrations were quantified using Quant-it PicoGreen dsDNA Assay kit (Thermo Scientific) and Synergy LX multi-mode microplate reader (Biotek). cDNAs were normalized to 250-500 pg µl-1. 100-200 pg of cDNA per sample was used for tagmentation in 1.2 µl using Illumina XT DNA sample preparation kit. Libraries were finally amplified for 11 cycles of PCR in 4 µl using custom-designed Nextera index primers containing 8-bp index barcode sequences with a minimal Levenshtein distance of 4 as previously published[10]. Samples were purified with Ampure XP beads at a 1:0.8 ratio and DNA quality in randomly selected 10 wells/plate was assessed on a High Sensitivity Bioanalyzer chip (Agilent). Libraries were quantified as mentioned above and normalized libraries were equimolarly pooled and purified one last time at a 1:0.8 ratio. Prior to sequencing, final library concentration was

determined using Qubit dsDNA High Sensitivity Assay kit (Thermo Scientific) and Qubit fluorometer (Invitrogen) and the average fragment size was calculated on a High Sensitivity Bioanalyzer chip (Agilent).

scRNA-seq libraries from PDXs were prepared initially prepared using Smart-seq2[11] (n=4) and later using Smart-seq3 protocol (n=2) as outlined above. Libraries with 1 % spiked-in PhiX control (Illumina) were sequenced at the 75-bp paired end on a high output flow cell using an Illumina NextSeq550 instrument at a sequencing depth of ~1 M reads per cell at the sequencing open lab of the German Cancer Research Center.

Single cell library preparation by 10x Chromium 3´ sequencing

Single cell RNA-sequencing libraries from human brain organoids as well as patient-derived allografts were prepared using Chromium Next GEM Automated Single Cell 3´ Library and Gel Bead Kit v3.1 (10x Genomics) according to manufacturer's instructions at the single cell open lab at the DKFZ. In parallel, HTO libraries were prepared as previously described[9]. mRNA libraries with 2 % spiked-in PhiX control (Illumina) were sequenced at the 100-bp paired end on a P3 flow cell using an Illumina NextSeq2000 instrument at a sequencing depth of ~80 K reads per cell. HTO libraries with 40% spiked-in PhiX control were sequenced at the 75-bp paired end on a mid output flow cell using an Illumina NextSeq550 instrument at a sequencing depth of ~2 K reads per cell at the sequencing open lab of the German Cancer Research Center.

Illumina Infinium MethylationEPIC array

Genomic DNA isolation from primary GBM tumors or patient-derived GBM cells was carried out using QIAamp DNA Micro kit (Qiagen). The Illumina Infinium MethylationEPIC kit was used to analyze the DNA methylation status at >850000 5´CpG islands per sample according to the manufacturer´s instructions. The assay was run at the Genomics and the Proteomics Core Facility of the German Cancer Research Center.

Bulk whole-genome bisulfite sequencing (WGBS)

Tumor bearing mouse brains (5 mpi) were additionally perfused with 4% PFA and post-fixed in 4% PFA overnight at 4°C. 50 µm coronal sections were prepared using a Vibratom VT1200S (Leica). Samples were stored in 0.01% NaN3 in PBS at 4°C until processing. The genomic DNA (gDNA) from 3 whole-brain sections per group was isolated by QiaAmp DNA microkit (Qiagen 1048145) and the concentrations were measured by Qubit DNA HS kit. The protocol for bisulfite conversion and library preparation was adapted based on the genomic part of the scNMT-seq protocol[12] with the following modifications. The gDNA input was adjusted to be 20 ng per sample. The number of first-strand purification cycles was decreased to 4 runs, and library amplification reaction was performed in 10 cycles. The library concentrations were measured by Qubit DNA HS kit and the fragment distribution was analyzed by the Bioanalyzer DNA HS kit (Agilent). Before sequencing, libraries were pooled equimolarly. The final pool with 15% spiked-in PhiX control was sequenced at the 150 bp paired end on a P2 flowcell using and Illumina NextSeq 2000 at a sequencing depth of 40M read per sample at the sequencing open lab of the German Cancer Research Center.

Immunofluorescent staining and microscopy

To visualize healthy v-SVZ-lineage, mouse brains were additionally perfused with 4% PFA post-fixed in 4% PFA overnight at 4°C. 50 µm coronal or sagittal sections were prepared using a Vibratom VT1200S (Leica). Samples were stored in 0.01% NaN3 in PBS at 4°C until staining. For staining, sections were incubated in blocking solution for one hour at room temperature to block unspecific binding sites. Subsequently sections were stained with anti-GFP, anti-GFAP, anti-Sox2, anti-S100β, anti-DCX and anti-NeuN antibodies in blocking buffer overnight at 4°C on a rotator. After washing the sections 3 times 15 min with 0.25% Triton-X 100 in PBS, secondary antibody staining was carried out using appropriate Alexa-fluorophore antibodies diluted in blocking buffer for 1 hour at the room temperature. DAPI was added to the antibody cocktail for nuclear staining. Furthermore, sections were washed 3 times 15 min with 0.25% Triton-X 100 in PBS. After mounting with Fluoromount-G, sections were allowed to dry in dark at room temperature for 30 min. Samples were stored in the dark at 4°C until imaging.

To visualize PDX tumors, mouse brains were prepared, processed and imaged as outlined above, 5 mpi. To visualize PDA tumors, HBOs were washed and fixed with 4% PFA for 1 hr at the room temperature, 10 dpi. Both PDA and PDX were stained with the following primary antibodies: anti-RFP, anti-GFP and anti-human-specific Nestin prior to imaging.

Confocal images and tile scans were acquired in a Leica TCS SP5 or SP8 confocal microscope at the Light Microscopy Facility of the German Cancer Research Center. ImageJ was used for processing and images were only adjusted for brightness and contrast.

Spatially resolved transcriptomics by molecular cartography (Resolve Biosciences)

For molecular cartography (MC) experiments, tumor bearing mice were additionally perfused with 4% PFA and post-fixed with 4% PFA overnight at 4°C. Samples were incubated in 15% and subsequently 30% sucrose solution overnight each at 4°C, respectively. OCT-embedded specimens were cryo-sectioned into 10 µm thick coronal sections onto MC slides. Permeabilization, hybridization and automated fluorescent microscopy imaging were performed according to the manufacturer's instructions for fix-frozen samples with minor modifications. Briefly, the slides were thawed at the room temperature and dried at 37°C in a thermal cycler with MC slide holder. Sticky wells were attached to the MC slide to create the MC observation chamber. The sections underwent various treatments excluding post-fixation but including permeabilization, rehydration, and application of TrueBlack autofluorescence quencher, which was diluted according to the instructions provided by Biotium. Following thorough washing and priming, specific probes designed by Resolve Biosciences' proprietary algorithm for our transcripts of interest (Sequences not listed here) were hybridized overnight at 37°C. The hybridized sections were washed again, and the MC observation chamber was placed in the MC machine for eight automated cycles of coloring and imaging to determine the transcript localization of our panel of 98 transcripts in the tissue. Regions of interest (ROIs) were selected for each section (Control vs sFRP1) based on a brightfield overview scan. To aid the ROI selection process, we additionally scanned consecutive sections of the MC sections following immunofluorescent staining with tumor specific markers. In the final imaging round, the nuclei were stained with DAPI to create a reference image for nuclei segmentation. After the run, the

MC software registered the raw images, assigned transcripts to detected combinatorial color codes, and combined individual tiles to create ROI panoramas. The outputs consisted of text files containing the 3D coordinates of the transcripts and maximum projections of DAPI images for each ROI.

In order to complement the MC data with a panel of relevant proteins, processed sections were used to perform immuno-fluorescent labelling (as outlined above) of mCherry, EGFP and were counter-stained for DAPI. The same ROIs were tile-scanned using Leica SP5 and SP8 confocal microscope at the Light Microscopy Facility of the German Cancer Research Center and images from respective RNA-protein modalities were registered using the common DAPI staining in each panorama.

Computational methods

Derivation of integrated adult WT SVZ NSC lineage

Our SVZ NSC reference lineage is comprised of adult WT SVZ NSCs and progeny from three studies[8,13,14]. We incorporated cells from [15] to supplement our SVZ lineage and compare celltype labels. All datasets were reduced to the set of 16.598 human-mouse 1:1 orthologs reported in Ensembl 102[16] and human gene names are used to denote their mouse counterparts throughout. Counts from the above datasets excluding [15] were integrated using the seurat integration pipeline outlined in the 'Introduction to scRNA-seq integration' vignette (https://satijalab.org/seurat/articles/integration_introduction.html). The main NSC lineage component was extracted by DBSCAN with eps=0.65 on the UMAP coordinates and integrated with [15] to conduct label transfer and identify neuronal clusters. DBSCAN once again identified the main component in this embedding and we re-integrated to arrive at a SVZ NSC reference lineage.

Assigning an NSC lineage pseudotime capturing differentiation

To generate an NSC lineage pseudotime we compared PC1+2 loadings across all replicates. A loading Z-score was computed and genes kept where the absolute loading Z-scores across replicates summed to >5, arriving at 1,287 putative lineage genes, 943 of which were present in the SVZ reference integration. In the integrated dataset, the PC2-lowest cell was selected as the root for a diffusion map computed over PC1-3, and a branchless diffusion pseudotime[17] computed over two diffusion components.

Exclusion of cycling cells in the SVZ reference

Cellcycle scoring was carried out following the scanpy '180209_cell_cycle' example notebook (https://notebook.community/theislab/scanpy_usage/180209_cell_cycle/cell_cycle) with the there-supplied genes. Filtering for 1:1 orthologs we scored 51 G2M- and 42 S-phase genes. We set a G2M-score cutoff at 0.1 to determine cycling activity, which was consistent with expression of known cellcycle markers including MKI67, TOP2A, and UBE2C (not shown). We identified and excluded 2.769 cycling cells, leaving 12.024 noncycling cells for which we computed a noncyling lineage pseudotime as described above.

Human cortical organoids processing and lineage pseudotime assignment

Four of these organoids were sequenced by 10X, distributing tumor and healthy FACS-sorted populations by lane. Fastq files were processed by Cellranger v6.0.0 and aggregated with cellranger aggr without normalization. Organoid-hashtags were counted by CITE-seq-count[18] and cells assigned based on frequency of the most-frequent compared to the second most-frequent hashtag per cell, applying a 1.5-fold threshold. Cells with >500 genes, >2000 UMIs, and with <20% mitochrondrial transcripts were retained.

We separated healthy and malignant cells by inferCNV[19] using the 10X brain nuc-seq cells (https://www.10xgenomics.com/resources/datasets/1-k-brain-nuclei-from-an-e-18-mouse-2-standard-2-1-0) as a healthy reference. inferCNV was run on genes having a mean lognorm expression value >0.25. Cells clustered according to their 10X lane, and a gaussian mixture model was trained on CNV calls from chromosomes 7,8,10, and 16 to distinguish healthy and malignant cells. Cells with concordant 10X lane and GMM-CNV assignment were retained for downstream analysis.

This human NSC lineage was annotated by integration with [15] reference, revealing NSC lineage celltypes along with a group of organoid cells near brain pericytes, vascular smooth muscle cells, and vascular or leptomeningeal cells which were excluded from downstream analysis. DBSCAN-isolated healthy organoid cells were reclustered in a human NSC-lineage UMAP. We identified and exclude 847 cycling cells from this human NSC lineage, and derived a lineage diffusion pseudotime as described above.

Comparison of mouse and human NSC lineages by pseudotime

We compared human and mouse NSC pseudotime gene expression for individual cell stage markers by spline regression, as well as globally by DTW. For DTW we bin cells into equal-sized pseudotime increments with bin edges defined from the SVZ reference, then bin-wise Pearson correlation over human and mouse coexpressed genes is computed for all pairs. We additionally compute a dynamic-programming traceback in the DTW matrix.

NSC lineage QAD state assignment and pseudotime-predictive QAD geneset

We identify and call QAD celltypes through their frequency in pseudotime by fitting a gaussian KDE over the noncycling NSC lineage pseudotime and identifying density peaks by argrelextrema from scipy.stats[20]. A pseudotime from 0 to 0.282 denotes Quiescence (with 0 to 0.141 representing dormant astrocytic stages), 0.282 to 0.676 denotes Activation, and pseudotimes greater than 0.676 denote Differentiation.

To capture QAD cellstates across datasets, we prefilter genes expression by binning cells into 3333 equal-frequency pseudotime bins and excluding those genes which don't reach a summed expression threshold of 100 lognorm units, or those for which the mean counts per QAD state don't exceed 40% of the total. Recursive feature elimination was carried out using a RandomForestClassifier from sklearn[21] on a 3-fold cross validated 70/30 split of cells into six equal-width pseudotime bins, to further reduce the considered feature-set to 3333 genes. Finally, 500 permutations of randomly selecting 500 genes and training a

RandomForestClassifier to predict the above six pseudotime bins were carried out and the mean feature weights across permutations used with the kneedle algorithm[22] to select pseudotime-predictive genes. We eliminated ribosomal genes and compared these genes across QAD states to selected the top state-enriched genes passing a mean lognorm expression threshold of 60% relative to that state's maximum, taking up to 100 genes from each state. This left 242 genes which were assigned to QAD states according to their maximum expression. GO-term enrichment was carried out using the enrichr class from gseapy[23] with expressed genes as background.

ptalign for pseudotime alignment

The *ptalign* program requires four central inputs: a query counts table, a reference counts table, a reference pseudotime mapping, and a set of genes capturing the reference pseudotime trajectory. Query cells are placed along the reference pseudotime according to the dynamic of the correlation of their gene expression along the trajectory genes. Briefly, reference cells are binned into equal-sized bins along pseudotime and the mean expression taken per bin. Pearson correlations are computed between the query cell transcriptomes and the pseudotime-binned reference over the supplied trajectory genes. Resulting correlation matrices are optionally normalized by row (reference bin), then scaled by column (query cell) to a value between 0 and 1. These scaled correlations represent a given cell's distance to different parts of the reference pseudotime, and in the next step is reduced to a single value representing pseudotime. To this end, *ptalign* uses the supplied reference counts and reference pseudotime to compute a reference-reference correlation matrix and train a small multi-layer perceptron network to predict the known reference pseudotime from the correlation dynamic in the supplied matrix. A 5-fold cross-validated grid search is performed over a number of relevant parameters to find the best network hyperparameters for the supplied reference and geneset. Then, query cell's pseudotimes are fed into the trained network to predict pseudotime values. Cellstates are optionally inferred based on cutoffs derived from the reference pseudotime. The quality of an assigned pseudotime is determined via the computation of a DTW matrix over the pseudotime-binned reference and the equally binned query. Counts are log-normalized and Pearson correlations computed for all combinations of reference and query bins. A matrix traceback is computed by dynamic programming to determine a path of maximal correlation through the DTW matrix. *ptalign* performance is estimated by comparing the length of the DTW traceback, the average correlation along the traceback, and by extracting the narrowness-parameter from a parabolic fit to the mean correlation along the diagonals of the DTW matrix. These three metrics are compared to reference-query matrices computed according to equally sized and equivalently expressed permuted genesets to determine an empirical p-value.

TCF-Lef SVZ dataset

Five 384-well plates of sorted mouse brain populations from SVZ or OB were prepared using the SmartSeq3 protocol (see above). Reads were trimmed by trim_galore[24] and mapped against the mm10 mouse reference genome using STAR v2.5.3a[25]. SmartSeq3 mappings were processed using a custom script based on HTSeq count[26]. Cell counts were QC-processed and log-transformed, with the NSC lineage isolated by DBSCAN in UMAP space and cycling cells eliminated to produce a non-cycling lineage psuedotime as described above. Pseudotime

alignment was carried out with default parameters and QAD cellstates assigned by the above pseudotime cutoffs. The non-cycling lineage diffusion pseudotime and ptalign pseudotimes were compared by Pearson correlation per replicate to benchmark ptalign. This process was repeated for the HBO NSC lineage and pseudotime described previously.

Human GBM PDX and PDA QAD states and pseudolineage

Six 384-well plates of sorted T6 PDX tumor cells were prepared, two using the SmartSeq3 protocol and four by SmartSeq2, as described above. Reads were trimmed and mapped against both the Ensembl 102 human and mouse reference genomes separately using STAR v2.5.3a. Cells were assigned to one of human, mouse, or QC-fail if they passed 250.000 reads and 65% alignment-rate and called human if the ratio of human reads was at least 1.3-fold. The 1,176 QC-passing cells from T6 Wnt-reporters were integrated by the Seurat pipeline with cycling cells called as described above. QAD-stage signatures were scored by AUCell[27] with a 10% rank-cutoff. Pseudotime alignment was carried out with default parameters and QAD cellstates assigned by pseudotime cutoffs. This process was repeated for T6 PDA cells isolated by the malignant calling described above.

Pseudolineage inference on published external GBMs datasets

We retrieved published primary GBM single-cell RNA-seq counts tables where available. Precise datasets and cell numbers are related in XXtable. Samples were processed into an annotated counts table and subset to 1:1 orthologs between human and mouse. All 66 datasets were merged into an AnnData object with cellnames appended their sample names to avoid collisions. GBMs were filtered to exclude those with >30% mitochondrial reads and <631 UMIs, corresponding to the smallest cell in the SVZ reference. Tumor samples comprising fewer than 500 cells were excluded. Counts were log-normalized per sample and merged.

Malignant cells were identified by inferCNV with downstream malignant classification as in[28]. A healthy CNV reference was built by randomly sampling 300 cells from our healthy HBO reference dataset. CNV genes were determined by comparing expression levels in healthy HBOs, our SVZ reference dataset, and the collated GBM samples. inferCNV was run over CNV genes for individual tumors samples with the cutoff parameter set to 0.1 as recommended for 10X datasets. Malignant calling was prefaced by identification of nonmalignant celltypes in the UMAP space. Microglia, immune cells, and oligodendrocytes were identified using the markers from[28] and excluded. We further computed the Pearson correlation of individual cells to each sample's mean CNV calls. Outlier cells with a correlation <0.4 were excluded as well as tumors with <400 malignant cells.

The curated set of primary GBM scRNAseq datasets was scored for cycling activity and cycling cells excluded, as previously described. Pseudotime alignment was carried out with default parameters and QAD assignment carried out for each tumor sample individually.

A GBM-derived QAD geneset

To derive a GBM QAD geneset, expression in GBM QAD stages with >20 cells was compared to QAD-expression in the healthy SVZ. Genes seen in >41 tumors were filtered by the mean of

QAD-maximum lognorm expression. We computed normalized ranks of each gene by expression strength per state per tumor, relativising ranks across states and keeping 1.604 genes with >60% overrepresented ranking in a single state. These genes were then compared by their state expression in the SVZ reference by subsetting to those genes having a 60% enrichment in a given QAD-stage. DEseq was carried out on QAD-pseudobulks comprising state- vs non-state groups and cycling vs noncycling cells. A 92-gene cycling geneset was derived by LFC>2 and a GBM-QAD geneset selected by taking up to 100 of each of the QAD-stage enriched genes by LFC, arriving at a 271 gene signature comprising 100 Q-, 71 A-, and 100 D-genes.

Comparison of published GBMs geneset classifications

We compared published transcriptional signatures by AUCell. Published genesets were reduced to human-mouse 1:1 orthologs and AUCell scoring carried out with a 10% rank-cutoff. Cells were clustered by geneset score correlations and compared by mean AUCell score per QAD-stage. Cells were assigned their maximum signature and labels kept where that score was >0.1, thus facilitating cell state frequency comparisons per publication source.

QAD state association with tumor clones by CNV

Clones were called by hierarchical clustering in the inferCNV predicted CNV values. A linkage matrix was computed for each tumor and the first three splits used to define clones per tumor. If linkage splits after the root were imbalanced, ie indicated a branch whose length extended past the daughters of the other split, this branch was isolated and the daughters of the other split treated separately. In this way, tumors were reduced to three or four putative clones, which were further screened for bona fide CNV differences in order to select high-confidence clones. Among the 2.725 CNV genes, CNV calls >1.15 were considered gains and <0.85 losses, together 'CNV events'. An average of 100 CNV events were required per cell, and tumors where clones did not surpass this threshold were considered CNV-low and excluded. The ratio of CNV events in the left and right daughters of each hierarchical-split was computed and tumors where clones did not differ by >10% were excluded. Finally, for each of the remaining clones, we took the ratio of the average number of CNV events per cell per clone and compared to the value at the parent node. A zscore was computed over these ratios for all clones and a threshold of 0.5 selected to determine clones with truly diverging CNV. This way, we call 21 CNV-low tumors, four single-clone tumors, 18 two-clone tumors, 7 three-clone tumors, and one four-clone tumor.

GBM microarrays and clinical data curation

We compiled microarray and clinical data from the TCGA-GBM cohort as well as a recent publication by Wu *et al*[29]. TCGA clinical metadata was retrieved from the TCGA portal (https://portal.gdc.cancer.gov/), merged on the case_id field, and assigned a unique internal ID. Recurrent GBMs were excluded. We downloaded and manually curated TCGA-GBM data of the two channel Agilent 244K Custom Gene Expression (https://portal.gdc.cancer.gov/legacy-archive/search/f). We created a new annotation file by aligning the 60 k-mer probes to the non-redundant nucleotide database (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz; reference build hg38; downloaded on 21.07.2016) by using BLAST+ v2.2.30 and the blast result was

annotated using mygene v1.8 and additional gene information was added using the NCBI gene-info file (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_info.gz; downloaded on 15.08.2016). Background correction was performed using the “normexp” method with a setoff value of 50, followed by within array normalization using the LOESS-smoothing algorithm. Only probes that were successfully annotated as described above were retained. For every gene, probes were summarized into a single value (gene-set). The resulting normalized matrix was saved into an MA-list object also including the curated meta- and clinical data. Finally, on patient samples that reported no IDH mutation, had no missing clinical data, and were of age above 30 years were retained and mapped to the verysame set of internal IDs by associating case and submitter IDs. Low quality arrays were determined by the proportion of zero-count probes and excluded, as were recurrent, treated, and arrays with nonstandard notes. In the case of Wu et al data patient data was downloaded from GSE121722 and processed in parallel with the raw readcounts received in personal communication. Survival had been reported in months and transformed to days by multiplying by 30.5.

GBMs QAD scoring, projection, and class assignment

Pseudobulk counts for scRNA-seq GBMs were log-normalized with a size-factor determined per publication batch and the R GSVA package[30] used to score the GBM-QAD genesets and cellcycle. Results projected into two dimensions by PCA. We used the ground-truth QAD-stages to define four reference points at the corners of the PCA, for which we computed euclidean distances of individual tumors. Tumors were assigned to the nearest reference point based on a distance threshold combined with a middle band of variable diameter which we used as an exclusionary criteria to focus cluster-assignment to the four corners. The reference point for the bottomleft mesenchymal cluster was moved to closer coincide with the density maximum of mes-like Q-biased scRNA-seq tumors, and non-scRNAseq tumors assigned based on a distance threshold of 0.65 and a midband width of 0.5. In similar fashion, proneural, QAD, and QA clusters are assigned with a distance threshold of 1, 1.1, and 0.8 and a midband width of 0.6, 0.65, and 0.7, respectively.

Overall survival was compared by tumor pseudolineage group, including right-censoring by patient death and considering survival intervals up to the 95th percentile. Kaplan-meier curves were computed for all pairwise pseudolineage group comparisons using the lifelines package[31]. The same criteria were applied for comparisons of progression-free interval. A Cox proportional hazards (lifelines) regression was carried out on the continuous QAD- and cellcycle GSVA-scores individually.

Methylation datasets acquisition and RNA states prediction

We retrieved TCGA samples processed by EPIC450k array from the TCGA-GBM cohort via the data portal and matched these to the in-house identifiers used in our microarray dataset. We similarly retrieved the EPIC450k methylation values from the Wu et al cohort from GSE121722. These compiled probe data were filtered to ‘cg’ probes not present on the sex chromosomes, and subset to those having a MAF<0.01. We computed variable methylation sites with the kneedle algorithm, selecting a cutoff in the cumulative distribution of variance over mean across

probes and computed a four-component PCA. K-means (sklearn) was used to cluster samples on PC1 into PC1-high and -low groups.

We next compared intra-tumoral distances in the Wu et al tumor cohort. 45 tumors were selected which did not belong to the PC1-high group, variable methylation and RNA sites computed separately as described above, and pairwise euclidean distances in methylation and RNA space computed for all tumor pairs. We then computed the Pearson correlation of these distances between tumors. Prediction of GSVA QAD scores from methylation data was done by elasticnet regression (sklearn) in a grid-search paradigm. Train and test sets were compiled from 66 and 33% of the data, respectively, and variable features selected by taking the 5000 probes with the largest zscore in the methylation PC2-4 loadings, as well as 2000 probes with the highest correlation in QAD scores, resulting in a 10.259 probe feature set for training. Methylation betas at this probeset were normalized and scaled using the StandardScaler and used in a GridSearchCV with 5-fold cross-validation to optimize the mean squared error and determine the optimal l1_ratio, alpha, and fit_intercept parameters. The grid-search selected parameters were used to train models on the training data and evaluated on the test data by pearson correlation of QAD and cellcycle scores individually.

Identification of GBM pseudolineage-predictive genes

We defined Q-biased and D-biased tumors based on the presence of at least 10% of dormant astrocytic or D-cells, respectively. By this metric we identified 16 D-biased and 11 Q-biased tumors, while no tumor could be assigned to both classes. Pseudobulk counts representing the mean proportion of log-normalized counts per gene per tumor were generated for the A^{nc}-cells of these 27 tumors and a t-test with multiple testing correction carried out to identify gene expression differences in the Activation stage.

Wnt activity by QAD state from TCF-Lef reporter

Wnt activity in the index sorted cells from the xenograft datasets was determined by mapping cells to wells and assessing the fluorescence readout of each well relative to a reporter-positive reference gate. Similarly for the TCF-Lef reporter SVZ reference lineage. The tumor allograft dataset was decomposed into reporter-positive and negative samples based on the 10X lane. Each of these three datasets was separated by replicate and the number of Wnt-active and –inactive cells determined per QAD-stage and averaged across replicates with error determined by the standard deviation per stage in the proportion of Wnt-active cells after subtracting the mean per replicate.

Wnt genes pseudotime expression dynamics in health and disease

We compiled 37 Wnt-related genes expressed in SVZ NSCs and computed their mean expression in 30 equal-size pseudotime bins which we scaled across bins and applied across identical pseudotime bins in our GBM scRNA-seq cohort. We modeled the healthy expression dynamics using the UnivariateSpline class (scipy) and computed the correlation as well as mean squared-error between the model fit and each tumor's binned and scaled counts, then fit a line through these metrics in every Wnt-gene of every tumor and projected each of these

points onto the closest point on the line. By splitting the point's coordinates on the line into three quantiles we classify those tumor and Wnt-gene combinations with low MSE and high correlation as the best-fit, and those with high MSE and low correlation as the worst-fit points. The Wnt genes dynamics consistency score is then derived from the proportion of the number of times a gene occurs in the best-fit vs worst-fit category.

SFRP1 and NOTUM overexpression datasets

Sorted cells from T6 SFRP1- and NOTUM-overexpressing PDXs were sequenced, trimmed, mapped, and cells assigned to species as described above. Cycling cells were removed and pseudotime alignment carried out to determine QAD assignments as described above. Kaplan-meier curves were generated by genotype as above. These analyses were repeated separately for the NOTUM sample. SFRP1 samples were further subject to pseudobulk DEseq analysis to determine genewise log fold-changes, with GSEA from gseapy applied to determine enrichment among the DEseq expression differences. Only a single NOTUM-overexpressing replicate was sequenced and consequently expression differences were quantified by pseudobulk LFC to Wnt-reporter SmartSeq3 replicates.

T6 SFRP1-overexpressing and Wnt-reporter cells injected into HBOs were sequenced on the 10X platform with hashtag antibodies labeling individual organoids (see above). These libraries were processed by cellranger and cell QC thresholds applied as described above. inferCNV was used to distinguish malignant and healthy cell populations, and cycling populations called as previously. Pseudotime alignment was carried out for tumor cells from each organoid using default parameters. Healthy cells from SFRP1 and Wnt-reporter conditions were isolated, a UMAP generated, and cell types identified by marker expression in a leiden clustering. Healthy cell type expression was compared by genotype by averaging expression between pseudobulk replicates per cell type.

WGBS data processing and SVZ DMR readout

WGBS libraries were prepared as described in [12] in triplicate from PDX samples. We generated a combined human-mouse reference genome per [32]to map reads from mouse and human cells using Bismark v0.20[33] with Bowtie2 v2.3.5.1[34]. We used bismark_genome_preparation with default parameters to prepare the merged genomes. Paired reads were trimmed by trim_galore, and single-end reads individually mapped to the combined human-mouse reference with Bismark, setting –q -N 1 -D 30 -R 3 -p 6 with --non_directional. PCR duplicates were removed using deduplicate_bismark, and mate-pairs recombined by merging deduplicated R1 and R2 BAMs with samtools merge. Finally, we ran the bismark_methylation_extractor tool with --bedGraph to quantify DNA methylation and counted CpG coverage using coverage2cytosine. The resulting CpG coverages were parsed and mouse and human methylation fractions per CpG determined separately. v-SVZ cell type DMRs were retrieved from [35] and mapped to the human genome using liftOver. Human CpGs were assigned to a v-SVZ DMR if they were within the liftOver coordinates +/-20%, and average methylation fraction determined per genotype per DMR. DMRs with a genotype methylation difference >1 st.dev from the mean were considered differentially methylated, and a scipy gaussian_kde was used to estimate the density of celltype

DMRs outside this bound. CpG coverage for Illumina EPIC array probes was based on detected CpGs among all sequenced samples in coordinate ranges defined by DMRs.

Spatial Methods

Alignment & Preprocessing

For every ROI, Resolve Biosciences provided both 3D transcript positions and identities and a 2D DAPI image, which first needed to be integrated with our 3D confocal DAPI images. Resolve captures data in a tiled manner within every ROI; since they still had some stitching issues, every tile was separately aligned to the corresponding confocal image along the xy axes, using a homography obtained by matching SIFT features between the 2D DAPI image and a max-projection of the 3D DAPI image. Different angles of the slide in both microscopes were corrected by finding the angle of the slide plane in both modalities and rotating the transcript positions accordingly. The plane was fitted to the center of mass of the DAPI intensity along the z axis for the 3D image, and to the tiled mean z position for the transcripts. Finally, a further shift along the z axis was added such that the share of transcripts within segmented nuclei was maximized, using the segmentation described in the next paragraph.

Segmentation

The 3D DAPI images were segmented into nuclei with mesmer[36], using the default pretrained model in the nuclear mode. Our images have a resolution of 0.142 microns per pixel, but to stay closer to their training data we used 0.3 microns per pixel as the parameter for the segmentation. Segmentation was performed separately for every z layer of the image, and afterwards stitched across the z axis with a modified version of the intersection over union based approach used in [37]. Segmentation of the ROIs into different brain regions was performed by hand.

The transcripts were then segmented with Baysor[38]. Baysor first clusters the transcripts by their neighborhood; we chose to demand 21 clusters as this was the lowest number where each major cell type had at least one clear corresponding transcript cluster. Afterwards, it segments the transcripts into cells and background noise. Where transcripts intersected directly with segmented nuclei, this was added as the prior segmentation with confidence 0.98. The scale parameter was set to 2 microns, and we chose a minimum of 3 transcripts per cell. We excluded the transcript mCherry since it appeared to contain a lot of technical artefacts.

The parameters for Baysor were chosen to encourage over-segmentation rather than under-segmentation to properly resolve dense areas like tumors. To finally arrive at real cells, we assigned the Baysor cells to their nearest segmented nuclei, taking care to only combine Baysor cells that correspond to the same cell type. Prior to this, we consolidated the Baysor clusters to get rid of cases where multiple clusters corresponded to the same cell type. The assignment was done iteratively, first assigning Baysor cells that have a clear overlap with nuclei, and afterwards assigning the remainder to their closest compatible nucleus, starting with the Baysor cells with the most transcripts. Transcripts that Baysor marked as noise, and Baysor cells that could not be assigned to a segmented nucleus, were dropped.

Spatial transcriptomics analysis

Cell types for mouse cells (defined as at least 60% mouse transcripts) were obtained by associating Baysor clusters with cell types, and then using the cluster identity Baysor provides for every cell. To assign human cells with a state, we for each state marked the transcripts that are enriched in it (e.g. Q+ transcripts are enriched in Q). Cells with less than 20% Q+ or A+, or 10% D+ transcripts were marked as generic human. Cells with at least 0.6 times as much Q+ as A+ or D+ were assigned Q, with at least 0.6 times as much D+ as Q+ or A+ were assigned D, and the remainder was assigned A.

Neighborhood enrichment was calculated using squidpy[39], using a radius of 20 microns to construct the neighborhood graph. The permutations were performed on the respective brain regions, combining all relevant regions from different ROIs. Squidpy reports the enrichment as number of actual neighbors compared to number of expected neighbors, normalized by the standard error of the expected number. This was multiplied by the square root of the number of cells in the permutation to obtain z-scores in terms of the standard deviation, since this is a measure of the effect size that does not scale with the overall number of cells that entered the computation.

Statistical Analysis

Statistical analyses were performed in python using the scipy module. All p-values were calculated using a two-sided test, unless specified otherwise. Multiple testing correction was carried out using the python statsmodels library. Box plots each span the 25th and 75th percentile while the center lines indicate the median and the whiskers represent the absolute range. The handling of outliers is indicated in the respective figure legends.

References

1. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nat. Protoc.* **1**, 241–245 (2006).
2. The ORFeome Collaboration. The ORFeome Collaboration: a genome-scale human ORF-clone resource. *Nat. Methods* **13**, 191–192 (2016).
3. Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nat. Methods* **8**, 409–412 (2011).
4. Lancaster, M. A. *et al.* Guided self-organization and cortical plate formation in human brain organoids. *Nat. Biotechnol.* **35**, 659–666 (2017).
5. Lancaster, M. A. & Knoblich, J. A. Generation of cerebral organoids from human pluripotent stem cells. *Nat. Protoc.* **9**, 2329–2340 (2014).
6. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
7. Ferrer-Vaquer, A. *et al.* A sensitive and bright single-cell resolution live imaging reporter of Wnt/ß-catenin signaling in the mouse. *BMC Dev. Biol.* **10**, 121 (2010).

8. Kalamakis, G. *et al.* Quiescence Modulates Stem Cell Maintenance and Regenerative Capacity in the Aging Brain. *Cell* **176**, 1407-1419.e14 (2019).

9. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).

10. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

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

12. Cerrizuela, S. *et al.* High-throughput scNMT protocol for multiomics profiling of single cells from mouse brain and pancreatic organoids. *STAR Protoc.* **3**, 101555 (2022).

13. Kremer, L. P. M. *et al.* High throughput screening of novel AAV capsids identifies variants for transduction of adult NSCs within the subventricular zone. *Mol. Ther. - Methods Clin. Dev.* **23**, 33–50 (2021).

14. Carvajal Ibañez, D. *et al.* Interferon regulates neural stem cell function at all ages by orchestrating mTOR and cell cycle. *EMBO Mol. Med.* **15**, e16434 (2023).

15. Cebrian-Silla, A. *et al.* Single-cell analysis of the ventricular-subventricular zone reveals signatures of dorsal and ventral adult neurogenesis. *eLife* **10**, e67436 (2021).

16. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).

17. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).

18. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

19. inferCNV of the Trinity CTAT Project.

20. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

21. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Mach. Learn. PYTHON*.

22. Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a ‘Kneedle’ in a Haystack: Detecting Knee Points in System Behavior. in *2011 31st International Conference on Distributed Computing Systems Workshops* 166–171 (IEEE, 2011). doi:10.1109/ICDCSW.2011.20.

23. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).

24. TrimGalore.

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

26. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

27. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

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

29. Wu, Y. *et al.* Glioblastoma epigenome profiling identifies SOX10 as a master regulator of molecular tumour subtype. *Nat. Commun.* **11**, 6434 (2020).

30. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013).

31. Davidson-Pilon, C. lifelines: survival analysis in Python. *J. Open Source Softw.* **4**, 1317 (2019).

32. Callari, M. *et al.* Computational approach to discriminate human and mouse sequences in patient-derived tumour xenografts. *BMC Genomics* **19**, 19 (2018).

33. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

34. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

35. Kremer, L. P. *et al.* Single-cell triple-omics uncovers DNA methylation as key feature of stemness in the healthy and ischemic adult brain. *bioRxiv* (2022) doi:10.1101/2022.07.13.499860.

36. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 (2022).

37. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

38. Petukhov, V. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2022).

39. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).