bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Single-cell Atlas of Human Fetal Cerebellum Decoding Medulloblastoma Origin and Oncogenesis

Zaili Luo[1]*, Mingyang Xia[2]*, Wei Shi[2]*, Chuntao Zhao[1]*, Jiajia Wang[3]*, Dazhuan Xin[1], Xinran Dong[2], Yu Xiong[4], Feng Zhang[1], Kalen Berry[1], Sean Ogurek[1], Xuezhao Liu[1], Rohit Rao[1], Rui Xing[1], Lai Man Natalie Wu[1], Siying Cui[2], Lingli Xu[2], Yifeng Lin[2], Hao Li[2], Wenkun Ma[3], Shuaiwei Tian[3], Qi Xie[5], Li Zhang[1], Mei Xin[1], Xiaotao Wang[6], Feng Yue[6], Haizi Zheng[1], Yaping Liu[1], Richard J. Gilbertson[7], Kathleen J. Millen[8], Jie Ma[3, #], Wenhao Zhou[2,#], Michael D. Taylor[9,#], Q. Richard Lu[1,#]

1. Brain Tumor Center, Division of Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
2. Key Laboratory of Birth Defects, Children's Hospital of Fudan University, Institutes of Biomedical Sciences, Fudan University, Shanghai 201102, China.
3. Department of Pediatric Neurosurgery, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, China.
4. Obstetrics and Gynecology Hospital of Fudan University, Shanghai 200090, China
5. School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China.
6. Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine Northwestern University, Chicago, IL, USA.
7. Cancer Research UK Cambridge Centre, CRUK Cambridge Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK.
8. Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA, USA
9. Developmental & Stem Cell Biology Program, The Hospital for Sick Children, Toronto, Ontario, Canada.

* These authors contributed equally.

#Correspondence: Dr. Richard Lu, Department of Pediatrics, Brain Tumor Center, Cincinnati Children's Hospital Medical Center; email: richard.lu@cchmc.org; or Dr. Michael Taylor, The Hospital for Sick Children; email: mdt.cns@gmail.com, Jie Ma, Xinhua Hospital, email; majie@xinhuamed.com.cn; or Wenhao Zhou, Children's Hospital of Fudan University; email: zwhchfu@126.com

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


## Abstract

Medulloblastomas (MBs) are the most common malignant childhood brain tumors, yet the origin of the most aggressive subgroup-3 form remains elusive, impeding development of effective targeted treatment strategies. Previous cell-type analyses of mouse cerebella or human counterparts from frozen tissue nuclei have not fully defined the compositional heterogeneity of MBs. Here, we undertook an unprecedented single-cell profiling of freshly-isolated human fetal cerebella at different developmental stages to establish a reference map for delineating the hierarchical cellular states in MBs. We identified a unique transitional cerebellar progenitor connecting neural stem cells to neuronal lineages in human fetal cerebella. Intersectional analysis revealed that the transitional progenitors were enriched in aggressive MB subgroups, including group-3 and metastatic tumors. Integrated single-cell multi-omic profiling revealed underlying regulatory networks in the transitional progenitor populations, including transcriptional determinants HNRNPH1 and SOX11, which are correlated with clinical prognosis in aggressive group-3 MBs. Genomic profiling and Hi-C analyses identified *de novo* long-range chromatin loops juxtaposing HNRNPH1/SOX11-targeted super-enhancers to cis-regulatory elements of MYC, an oncogenic driver for group-3 MBs. Targeting the transitional progenitor regulators inhibited MYC expression and MYC-driven group-3 MB growth. Together, our integrated single-cell atlases of human fetal cerebella and MBs reveal important cell populations predisposed to transformation and regulatory circuitries underlying tumor cell state evolution and oncogenesis, highlighting hitherto unrecognized transitional progenitor intermediates predictive of disease prognosis and potential therapeutic vulnerabilities.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Introduction

Medulloblastomas (MBs), the most common malignant childhood brain tumors[1,2], are classified broadly into four major subgroups: Sonic Hedgehog (SHH), Group-3 (G3), and Group-4 (G4) MBs, which all arise from the cerebellum [1,3], and WNT MB, which originates from the lower rhombic lip in the dorsal brainstem [1,4]. The developing cerebellum is composed of distinct neural cell types which arise from 3 distinct progenitor zones with spatiotemporally specific developmental hierarchies [5,6]. GABAergic neuronal lineage cells arise from the cerebellar ventricular zone (VZ) adjacent to the fourth ventricle with Purkinje cells accounting for the large majority of neuronal populations in the early cerebellar anlage[5,6]. Glutamatergic lineages arise from the dorsally located rhombic lip (RL). RL-derived granule cell progenitors (GCP) exit the RL and migrate rostrally over the developing cerebellar anlage to form the external granule cell layer (EGL), where they proliferate and differentiate to granule neurons. The RL also gives rise to additional glutamatergic populations including unipolar brush cells (UBC) which exit the RL and migrate directly into the cerebellar anlage [5,6].

Our understanding of the origins of cerebellar tumorigenesis is largely informed by molecular, cellular, and histological analyses from rodent models[7-9]. However, compared to the mouse cerebellum, the human cerebellum has a 750-fold larger surface area with higher cell numbers and increased foliar and lineage complexity [10-12]. The developing human cerebellum has expanded primary progenitor populations relative to rodents including the expansion of the cerebellar VZ to include a sub-ventricular zone ($VZ^{SVZ}$) and compartmentation of the RL into a $RL^{VZ}$ and $RL^{SVZ}$ [13,14]. The reactivation of embryonic developmental programs and emergence of stem-like cell plasticity have been implicated in tumorigenesis and treatment resistance in various types of cancers including MB[15-17]. Despite recent bulk and single-cell transcriptomic profiling of the developing human brain [18,19], prenatal development of the human cerebellum remains incompletely characterized. Histological and single-nucleus RNA-seq (snRNA-seq) analyses of human fetal cerebella revealed RL progenitors distinct from those in the mouse cerebellum [13,14]. This suggests that the developing mouse cerebellum may not fully recapitulate the cells-of-origin or cell types in human cerebellar tumors. However, snRNA-seq is limited by lower amounts of mRNA in nuclei

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

compared to whole cells, which may present a challenge for identifying rare or transient cell populations [20,21].

Deconvolution analysis of mouse cerebellar cell types and MBs suggests that SHH and G4 MBs resemble granular cell progenitors (GCP) and unipolar brush cells (UBCs), respectively [7,8,22]. However, mouse analogs of human G3 MB, the most aggressive tumor among MB subgroups with MYC amplification and the worst prognosis [23-25], have not been identified based on mouse cerebellar cell types, and the cell-of-origin for G3 MB remains elusive. Here, we carried out an unparalleled single-cell transcriptomics profiling of whole cells from freshly isolated human fetal cerebella to define cell identities, dynamics of transitional cell states, and their lineage trajectories during early human fetal cerebellar development. We further compared them to single-cell transcriptomes of MB subgroups to elucidate the developmental programs and cell populations that are vulnerable to tumor transformation. We identified a previously unrecognized transitional intermediate progenitor population in the human fetal cerebellum as a potential cell of origin for aggressive MBs such as G3 tumors. Moreover, integrative single-cell multi-omics (scRNA-seq/snATAC-seq) together with deep-sequencing 3D chromosome Hi-C analyses revealed unique tumor driver networks and enhancer hijacking events correlated with the growth of aggressive G3 MBs, pointing to potential therapeutic avenues for treatment of these deadly pediatric brain tumors.

## Results

### Single-cell transcriptomic analysis reveals previously uncharacterized transitional intermediate progenitors in the human fetal cerebellum

To characterize the cellular diversity of prenatal progenitors during human cerebellar development, we isolated fresh cerebellar tissues from aborted fetuses at post-conception weeks (PCW) 8 to 17 of gestation and profiled approximately 95,542 cells by 10x Genomics single-cell RNA sequencing (scRNA-seq) after quality control and doublet removal [26]. Higher numbers of genes or read counts per cell were obtained in the scRNA-seq data from fresh fetal cerebellar tissues than those from the snRNA-seq and Split-seq experiments[13] (Extended data Fig. 1a). Unsupervised clustering of individual cell transcriptomes visualized by t-SNE [27] identified 23 clusters in fetal

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cerebella (Fig. 1a and Supplementary Table 1). Clustering was independently verified with UMAP plotting (Extended data Fig. 1b). Similar to previously identified mouse and human cell types [7,13], we annotated the major cell types in human fetal cerebella by interrogating the expression patterns of canonical markers for different cell lineages including neural stem cells (NSCs), GCP lineage cells, unipolar brush cells (UBC) and their progenitors (UBC_P), Purkinje cells, GABA interneurons, microglia, monocytes, glial progenitors, oligodendrocytes, ependymal cells, vascular cells, and endothelial cells; the frequencies of these cell types varied over time post-conception (Fig. 1a,b).

To focus on neural cell types, the presumed origin of MB, in more depth, we investigated cell lineage trajectories by utilizing Monocle analysis[28], which can robustly infer branching trajectories along with developmental decisions in an unsupervised manner. We predicted the NSC population based on the stemness score as the start point and showed a trajectory through an intermediate population that we refer to as transitional cerebellar progenitors (TCPs) to the three main neuronal lineage branches, GCP, UBC and Purkinje cell lineages (Fig. 1c). Similarly, cell trajectory analysis by STREAM [29] also predicted the TCP state in the developmental hierarchy of cerebellar neural precursor cells (Extended data Fig. 1c). To define the expression dynamics along the trajectory, we used Slingshot pseudo-time analysis [30] through a diffusion map approach and found a successional gene expression pattern from NSCs (e.g., *SOX2*, *NES*, *HES5*) to TCPs (e.g., *HNRNPH1*, *SOX11*, *CTNNB1*) (Fig. 1d and Extended data Fig.1d). From PCW 8 to PCW 17 there was an increase in the proportion of Purkinje, then GCP and UBC lineage cells (Fig. 1e), consistent with known dynamic changes in the cellular composition of these major cerebellar cell types [13]. Importantly, TCP cells were abundantly derived from samples at PCW 12 and 13, a transitional period from the first to second trimester, but their frequency gradually decreased after PCW 14 (Fig. 1e), suggesting that they are a transient cell population in fetal development. Further sub-clustering analysis of cell cycle indicated that a TCP population also expressed cell-cycle genes in G1/S and G2/M phases and a proliferative marker Ki67 (Extended data Fig. 1e), suggesting that a subset of TCPs is in a mitotic state.

To validate the robustness of our cluster-specific gene expression observations, we plotted the

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expression of signature genes that are enriched in each cell population and showed that the signature genes are highly cell-type-specific in distinct progenitor cells (Fig. 1f, g). Among them, the TCP population was enriched in expression of *HNRNPH1* and *SOX11*, and was distinct from previously defined NSC, GCP, and UBC lineage cells (Fig. 1f, g and Extended data Fig. 1f). The gene expression signature of TCP cells partially overlapped with but is distinct from that of RL cells, a unique population in the human fetal cerebellum [13] (Extended data Fig. 2a-d).

To determine if the human cerebellar progenitor populations had orthologs in the mice, we compared the populations we identified to published reference profiles of developing mouse cerebella [7] by using a Support-Vector-Machine (SVM) algorithm for two-group classification [31] and LIGER analysis [32]. The majority of cerebellar populations of human fetuses shared similarity with mouse counterparts (Fig. 1h and Extended data Fig. 2e,f). However, molecular features of the TCP population were not present in any known mouse cerebellar cell population (Fig. 1h and Extended data Fig. 2e-g), suggesting that TCPs are a newly identified neural progenitor population in human fetal cerebella. Our observations revealed that the human fetal cerebellum consists of diverse cell populations including NSC, GCP and UBC lineage cells, as well as a previously unrecognized transitional intermediate phenotype, TCPs. Thus, single-cell differential gene expression analyses allow us to categorize distinct neural lineage progenitors during a heterogeneous developmental window of human fetal cerebella.

**In vivo validation of the intermediate progenitors in the human fetal cerebellum**

To validate TCPs as a spatiotemporally specific progenitor population during human fetal cerebellar development, we examined the expression of HNRNPH1 and SOX11, the most highly enriched TCP signature markers, in human fetal cerebella. We found that HNRNPH1$^{+}$ and SOX11$^{+}$ TCP cells were increased in regions adjacent to the NSC (SOX2$^{+}$) niche in the VZ from PCW 9 to PCW 12, a transitional period from the first to second trimester (Fig. 2a), but reduced progressively after the second trimester beginning at PCW 14 (Fig. 2a, b). This is consistent with the transient expression pattern of TCP cells based on lineage trajectory analysis.

In the developing posterior lobule, a RL progenitor cell pool was identified in the human RL

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

region that has not been detected in other organisms [13,14]. Markedly, we found that the TCP cell population with robust expression of HNRNPH1 and SOX11 was highly enriched in the transitional zone adjacent to the cell dense RL region at PCW 12 compared with other stages (PCW 9, 14 ad 16) (Fig. 2c,d), consistent with their transient cell states, and in the $RL^{VZ}$ region, while low expression of the TCP signature markers were detected in the $RL^{SVZ}$ region (Fig. 2e). Similar to the cerebellar VZ, expression of HNRNPH1 and SOX11 was detected in a population of SOX2+ NSCs in the RL transitional zone (Fig. 2f).

Our trajectory analysis predicted that the TCP population may give rise to GCP ($ATOH1^+$) and UBC ($EOMES^+$) populations (Fig. 1). To validate this in situ, we examined the transitional zone for each of these populations for TCP markers. We detected a sparse population of $HNRNPH1^+$ cells that co-labeled with EOMES and PAX6 in the transitional zone (Fig. 2g,h), wherein $PAX6^+$ progenitors can give rise to both GCPs and UBCs [33]. This is in line with a potential lineage trajectory from TCPs to UBC populations through $PAX6^+$ TCP intermediates. In addition, a population of $ATOH1^+$ GCP progenitors in the posterior EGL layer were also co-labeled with HNRNPH1 (Fig. 2i). Consistent with the hypothesis that a population of TCPs contributes to the development of GCPs, UBCs, and Purkinje cells, pseudo-temporal ordering of cell state evolution by Slingshot [30] and STREAM [29] analyses revealed that a cell lineage trajectory initiated in NSCs branches through the TCP subpopulation, which may serve as a precursor to generate GCP and UBC lineage cells (Fig. 2j,k and Extended data Fig. 3).

**Single-cell transcriptomic analyses reveal intermediate TCP-like cells in aggressive medulloblastoma**

Although GCP and UBC cells have been implicated as the cells of origin for SHH and G4 MBs, respectively [7,8], the cellular origin for G3 MBs has not been determined. To identify progenitor cells in the developing cerebellum that reflect the molecular features of cerebellar MBs (SHH, G3, and G4 MB), we first compared the human fetal cerebellar cell-type profiles to bulk transcriptomes of the MB cohorts from the Children Brain Tumor Tissue Consortium using CIBERSORTx [34]. Consistent with previous observations [7,8], the transcriptomic signatures of SHH-MB from children

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(>3 years old) and infants (≤ 3 years old) displayed strong similarity to GCPs (Fig. 3a), although child SHH tissues also exhibited weak similarity to the TCP signature. In addition, the transcriptome profile of G4 MBs resembled that of UBCs (Fig. 3a). Strikingly, of the MB subgroups evaluated, G3 MB tissues (including $MYC^{high}$ and $MYC^{low}$ tumors) had the strongest similarity to human fetal TCPs (Fig. 3a).

To further define the cellular identity of cerebellar MBs at the single cell level, we characterized a cohort of 26 MBs, including tumors from 18 patients newly diagnosed with MB, using scRNA-seq and snATAC-seq (the assay for transposase-accessible chromatin-seq) in matched tissues. These tumors included four SHH-MBs (45,731 cells), nine G3 MBs (82,834 cells), and thirteen G4 MBs (77,521 cells) with a median of 7,926 cells for scRNA-seq and 9,342 cells for snATAC-seq per tumor (Supplementary Table 2). We also included in our analysis previously reported transcriptomics data from 8 MB tumors [7]. Next, we determined the transcriptional similarity between cell populations in MB tissues and distinct normal cell populations of the developing human cerebellum. Analysis of the tumor cell landscape identified TCP-like populations in all G3, G4, and SHH MBs in addition to tumor subtype-specific cell clusters such as populations enriched in *MYC*, *NRL*, and *RPL* in G3 MBs (Fig. 3b), populations enriched in *EOMES*, *MYCN*, *NRL* in G4 MB cells (Fig. 3c), and populations enriched in *MYCN*, *GAP34*, and *MGP* in SHH MBs (Fig. 3d). Unsupervised VECTOR trajectory analysis [35] predicted that the TCP-like cells in SHH, G3, and G4 MB tumors were in an undifferentiated state along developmental trajectories toward different cell types in each MB subgroup (Extended data Fig. 4a-c). Consistently, reciprocal analysis of the overlaps between tumor cells and primary fetal cerebellar tissues using ProjecTILs[36], an algorithm for reference atlas projection, revealed that the MB tumor cells from G3, G4 and SHH subgroups also exhibited TCP cells/states present in the human fetal cerebellum (Extended data Fig. 4d-g).

Correlation between cell clusters observed in MB subgroups with human fetal cerebellar cell types showed that TCP-like cells in tumor tissues transcriptionally mimicked the normal TCP cell populations with high similarity scores, whereas neoplastic cells in G4 MBs (such as $EOMES^{+}$ and $MYCN^{+}$ clusters) and in SHH MBs (such as $MGP^{+}$ and $GAP43^{+}$ clusters) had gene signatures

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

similar to UBCs and GCPs, respectively (Fig. 3e). Consistent with this, using the expression signatures of normal human cerebellar cell populations to decompose the transcriptomes of a cohort of human MB subgroup tumors with CIBERSORTx [37], we found that TCP-like populations were present in the MB subgroups with higher abundances in G3 and G4 MBs than in SHH-MBs (Fig. 3f). As shown in Fig. 3g, all TCP-like cells in MB tumors exhibited the TCP signatures including *HNRNPH1*, *SOX11*, *RASGEFIB*, and *CTNNB1*. HNRNPH1 and SOX11 are elevated in a variety of cancers and involved in tumorigenesis [38,39], while CTNNB1 can drive tumor formation in the cerebellum if normal lineage restriction is lifted [40]. In addition to TCP signature genes, TCP-like populations in individual SHH, G3, and G4 MB tumors also expressed subgroup-specific signatures: The SHH-MB TCP-like cells expressed *PDLIM3*, *MGP*, *GAP43*, and *NHLH2*; G3-MB TCP-like cells expressed *MYC*, *PVT1*, *LMO3*, *HLX*, *IMPG2*, *NRL*, and *CRX*; and G4-MB TCP-like cells expressed *EOMES*, *LMX1A*, *GABRB3*, *NEUROD2*, *BARHL1*, and *KCNA1* (Fig. 3g and Supplementary Table 3). These observations suggest that there is a tumorigenic evolution of TCP cells into specific neoplastic TCP-like cells in individual MB subgroups, possibly caused by distinct driver mutations.

Next, we generated cell-state plots based on the marker gene expression [41] to analyze the distributions of tumor cell states in different MB subgroups. We observed specific enrichment of tumor cell states associated with MB subgroups (Fig. 3h). In contrast to the relatively restricted tumor cell state in SHH MB, we detected a $MYC^+$ and $NRL^+$ cell states as well as a state with a G4 MB-like signature in G3 MBs (Fig. 3h). The G4 MB tumors were enriched in G4 MB-specific states and also a cell state with an $NRL^+$ G3 signature (Fig. 3h). These observations suggest that a cohort of G3 and G4 tumors may have cells with either mixed or intermediate plastic states.

**Distinct tumor cell populations in G3 and G4 MBs**

G3 and G4 MBs share a certain degree of similarity in terms of signature genes based on bulk transcriptome and DNA methylation profiles [8,42]. It has been proposed that there may exist intermediate cell populations exhibiting characteristics shared between G3 and G4 tumors based on the analysis of selected G3- and G4-expressing gene sets [8]. To evaluate this hypothesis, we analyzed

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

two MBs (BT-309 and BT-325) that showed a mixture of G3 and G4 signatures based on a DNA methylation profiling array (Methylation EPIC; Fig. 4a). The analysis with the previously defined group 3/4 B/C scoring of selected marker genes [8] suggested a small population of cells with the profiles overlapping those of both G3 and G4 tumor cells (Extended Data Fig. 5a), while the majority of G3 and G4 cells were non-overlapping. However, unbiased single-cell clustering analysis of transcriptomes revealed distinct G3 and G4 tumor cell populations without appreciable intermediate states expressing both G3 and G4 signatures (Fig. 4b, c). This observation suggests that a mixture of distinct G3 and G4 tumor populations underlies the heterogeneity of the tumors with mixed G3 and G4 signatures. Moreover, distinct cell states were detected within both G3 and G4 MB tumors (Fig. 4d). Single cell copy number variation (CNV) analysis further confirmed that G3- and G4-like populations with distinct CNV patterns were present within the individual tumors (Extended data Fig. 5b). Together, these data suggest that a cohort of G3 and G4 MB tumors may harbor a mixture of G3- and G4-specific cell populations rather than a G3/G4 cell state continuum.

**Frequency of TCP-like cells and G3 MB gene signature increases during metastasis**

The primary BT-325 tumor, which harbored both G3-like and G4-like tumor cells, metastasized to the leptomeningeal surface of the brain. When compared with the primary tumor cells, the metastatic tumor had a substantial increase in frequencies of both TCP-like cells and $MYC^+$ G3-like tumor cell populations but a decrease in G4-like cells (Fig. 4e-g). Similarly, cell-state plot [41] indicated an enrichment in TCP-like and $MYC^+$ tumor cell states in the metastatic tumor coupled with a decrease in $EOMES^+$ G4-like cell states when compared to the primary tumor (Fig. 4h). The gene signatures of TCP-like and $MYC^+$ G3 tumor cells were enriched in the metastatic tumor (Fig. 4i). Furthermore, CNV analysis using a bulk DNA methylation array confirmed *MYC* gene amplification on chromosome 8 in the metastatic tumor (Fig. 4j), which is in accordance with the higher level of *MYC* expression in the metastatic tumor compared to the primary tumor (Fig. 4k). Similar increase in frequencies of TCP-like cells during metastasis was also observed in additional paired primary and metastatic G3 tumors analyzed by single-cell RNA-seq profiling or histochemistry (Extended data Fig.5c-f), which is in concordance with the high rate of metastasis in

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

G3 tumors [43]. These data suggest that TCP-like cells are enriched selectively in the metastatic tumors.

**Integrating single-cell omics identified core regulatory networks driving TCP transformation in aggressive MBs**

To decipher how dynamic accessibility at *cis*-regulatory elements (CREs) relates to the gene regulatory programs in TCP-like cells from aggressive G3 and G4 MBs, we performed single-nuclei ATAC-seq (snATAC-seq) to evaluate chromatin accessibility landscapes. High-quality nuclei were obtained from a set of fresh G3 and G4 MB tumors for which matched scRNA-seq data were also generated (Fig. 5a,b and Extended data Fig. 6a,b). By integrating scRNA-seq and snATAC-seq datasets using an ArchR software [44], we identified reproducible cell populations based on their epigenomic profiles that corresponded to gene expression profiles (Fig. 5c, d). Utilizing the marker gene integration matrix, we annotated all major cell populations including G3-like TCPs, *MYC*- and *NRL*-enriched clusters in G3 MBs, and G4-like TCPs, *EOMES*- and *KCNA1*-enriched clusters in G4 MBs (Fig. 5c, d). By correlation of accessibility of promoter and gene body elements with target gene expression [45], we found that the positively correlated peak-to-gene pairs were mostly subcluster-specific (Fig. 5e). The subcluster-specific genes associated with open accessible chromatin included *SOX11*, *HNRNPH1*, *RASGEF1B*, and *EIF5A* in TCP-like subpopulations of both G3 and G4 MBs. In $MYC^+$ G3 MB subpopulations, the genes were *MYC*, *YTHDF2*, *MARK4*, and *AKT2* while in $EOMES^+$ G4 MB subpopulations they were *EOMES*, *LMX1A*, and *MYCN* (Fig. 5e). To further validate these findings, we analyzed cluster-based pseudo-bulk ATAC-seq signals and showed locus-specific chromatin accessibility at the gene bodies linked to different MB subgroups (Fig. 5f). The signals of open chromatin peaks corresponded to MB subgroup-enriched genes such as the *MYC* and *NRL* loci in G3 tumor cells, *EOMES* and *LMX1A* in G4 tumor cells, and *SOX11* and *HNRNPH1* loci in TCP cells (Fig. 5f).

To determine the temporal relationship between chromatin accessibility and gene expression, we ordered the accessibility peak-to-gene pairs based on their CRE accessibility as a function of the integrated pseudotime. We identified paired accessible chromatin regions correlated to target gene

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expression in G3 and G4 MB tumor cells (Fig. 5g,h). This systematic analysis showed that the CRE accessibility of TCP-like cell states in G3 tumors appeared to precede that of the MYC$^+$ tumor cell cluster, which was followed by NRL$^+$ tumor cell cluster (Fig. 5g). Motif analysis indicated an enrichment of binding motifs for SOX11, HNRNPH1, and TWIST1 in the TCP-like subpopulation, whereas TCF3 and MYC were enriched in the MYC$^+$ G3 subpopulation, and NR2F1 and PAX5 motifs were enriched in the NRL$^+$ G3 subpopulation (Fig. 5g). Gene ontology (GO) analysis of ATAC-seq peaks and transcription factor (TF) motifs differentially enriched in G3 MB clusters showed an enrichment of epithelial cell development, epithelial-to-mesenchymal transition (EMT), and TGFβ/BMP signaling in the TCP-like population, while MYC$^+$ and NRL$^+$ subpopulations exhibited enrichment in cell-cycle phase transition, cell migration, and Notch signaling, and enrichment in photoreceptor cell development and Hippo signaling pathways, respectively (Fig. 5g). The enrichment in EMT, TGFβ signaling, and cell migration has been shown to contribute to the high metastatic potential of G3 MBs [46,47].

Similarly, we identified the subcluster-specific ATAC-seq peaks and TF motifs differentially enriched in G4 MBs. The accessibility of the TCP-like population emerged prior to that of the KCNA1$^+$ and EOMES$^+$ cell clusters. As in the TCP-like population in G3 tumors, an enrichment of epithelial cell development, neural progenitor cells, and UBC signatures such as LMX1A, was observed in TCP-like cells in G4 tumors (Fig. 5h). The KCNA1$^+$ clusters were enriched in cell-cell adhesion, regulation of neuronal progenitors, and MAPK signaling pathways, while the EOMES$^+$ G4 subpopulation was enriched in neuronal development, and HIF-1 and PI3K signaling pathways (Fig. 5h). The most enriched TF motifs in G4-like TCP cells were SOX11 and NEUROG1. In EOMES$^+$ G4 subpopulations, FOXG1 and LMX1A motifs were enriched, whereas RORA and PKNOX1/2 motifs were enriched in KCNA1$^+$ G4 subpopulations (Fig. 5h).

To identify positive transcriptional regulators that control gene expression restricted to TCP-like-cell populations, we integrated snATAC-seq and scRNA-seq data to identify TFs with gene expression scores that are positively correlated to changes in the accessibility of their corresponding motifs [44]. TCP-like cell populations in G3 and G4 MBs showed an enrichment of TFs such as HLX, CRX, OTX2, BARHL1, and LMX1A in addition to TCP markers HNRNPH1 and SOX11 (Fig. 5i;

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extended data Fig. 6c and Supplementary Table 3). We further identified the subtype-resolved CRE sites co-accessible with the promoters of potential drivers for MBs including *OTX2* and *HLX*, genes enriched in TCP-like cells in G3 tumors (Fig. 5j), as well as *BARHL1* and *PAX6*, genes enriched in TCP-like cells in G4 tumors (Fig. 5k). Thus, integrating scRNA-seq and snATAC-seq allowed us to map CREs to genes and to identify the transcriptional regulators enriched at tumor subcluster-specific CREs or putative enhancers. Co-accessibility of regulatory CREs and target gene loci might contribute to expression of MB subtype-specific drivers and their oncogenic programs.

**TCP-associated enhancer-hijacking promotes oncogenic programs in aggressive G3 MB**

Given that TCP signature markers SOX11 and HNRNPH1 were identified in CREs in TCP cell populations, we performed a Cut&Run genomic occupancy assay [48] in patient-derived MYC-driven G3 MB tumor cell lines (MB-004 and MB-002) compared with control non-transformed cells, human NSCs (hNSCs) and human astrocytes, to identify the direct targets of SOX11 and HNRNPH1 (Fig. 6a). We found that HNRNPH1 and SOX11 co-occupied the enhancer and promoter regions near transcriptional start sites marked by activating histone marks H3K27ac and H3K4me3 [49] (Fig. 6b). Signals of the activating H3K27ac mark were higher in G3 MB cells than hNSCs or human astrocytes (Fig. 6c and Supplementary Table 4), suggesting a potential higher level of gene activation in G3 MB cells than control cells. HNRNPH1 and SOX11 targeted common and unique genomic loci in G3 MB cells (Fig. 6a, boxed). GO analysis indicated that the unique targets in G3 MB cells were associated with the genes regulating G3 MB oncogenesis (*MYC*, *OTX2*, *PVT1*, and *CIP2A*), TGFβ/BMP signaling (*SMAD2* and *TGFBR3*), mitotic Ser/Thr kinase signaling (*PLK1* and *MAPK1*), and cell cycle regulation (*CHEK1*, *TOP2A*, *PCNA*, and *MCM2*) (Fig. 6d). Consistently, expression of these genes was higher in G3 MB cells than control cells (Fig. 6e). Notably, HNRNPH1 and SOX11 targeted unique enhancer/promoter elements of the G3 signature genes, *MYC* and *OTX2*, in G3 MB cells but not in hNSCs or human astrocytes (Fig. 6f). Analysis of the consensus sequence motifs associated with sites targeted by HNRNPH1 and SOX11 using the HOMER program revealed binding motifs for TFs for cell growth such as SMAD4, TCF3, SOX21, and TEAD4 (Fig. 6g), wherein TGFβ/SMAD and HIPPO/TEAD signaling pathways have been shown to regulate tumor growth and metastasis [46,47,50]. Thus, our results suggest that HNRNPH1 and

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SOX11 may cooperate with these TFs to regulate downstream gene expression that drives the G3 tumorigenic and metastatic programs.

To determine whether SOX11/HNRNPH1-occupied enhancers correspond to distal regulatory elements for activation of G3 MB driver genes, we performed Hi-C chromosome conformation capture in patient-derived G3 MB (MB-004) and G4 MB (UPN3550) tumor cells and detected unique genomic looping in the two cell lines (Extended data Fig. 7a, b). To further define the chromatin interactions induced by structural variations, we used a NeoLoopFinder algorithm [51] to reconstruct local Hi-C maps surrounding breakpoints. We identified unique structural variation (SV) events that may impact gene expression in G3 MB and G4 MB cells and revealed distinct sets of the interacting genomic loci involved in neo-loop formation at loop anchors (Extended data Fig. 7c, d). The neo-loop formation through inter-chromosomal translocation in G3 MB cells placed potential promoter/enhancer elements on a chromosome 11 segment close to the promoter of the genes on chromosome 19 (Extended data Fig. 7e) such as *PPP1R14A* (*CPI-17*), which has been shown to drive oncogenic RAS signaling in human cancers [52]. Consistently, expression of *PPP1R14A* was higher in the G3 MBs compared to SHH- and G4 MBs (Extended data Fig. 7f). These observations suggest that inter-chromosomal translocation or SV events might activate potential oncogenic drivers through enhancer-hijacking for tumorigenic programs in aggressive G3 MBs.

Hi-C analysis indicated that the topologically associated domains in G3-MB cells harbored unique long-distance interactions with the enhancer and promoter regions of *MYC* (Fig. 6h). Analysis of super-enhancers based on large clusters of active transcriptional enhancers with H3K27ac depositions [53] identified a set of potential super-enhancers at the regulatory elements near the gene loci such as *OTX2*, *DUX4*, *CASC8* and *MYC* in G3 MB cells (Extended data Fig. 8a). By integrating the long-range interacting sites with enhancer occupancy of HNRNPH1 and SOX11 in G3-MB cells and hNSCs, strikingly, we uncovered *de novo* chromatin interaction loops linking the distal super-enhancer clusters bound by HNRNPH1 and SOX11 upstream of *CASC8* to the active promoter/enhancer elements of the *MYC* locus in the G3 MB cells but not in hNSCs (Fig. 6h), suggesting that this long-range chromatin looping of super-enhancers juxtaposing to the *MYC* locus may promote *MYC* expression in G3 MB cells, in keeping with high MYC expression in G3 MBs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CRISPR interference (CRISPRi) targeting of SOX11/HNRNPH1 binding motifs in the distal super enhancers, but not a control enhancer site, resulted in significant downregulation of *MYC* expression in G3-MB cells (Fig. 6i), suggesting a critical role of the upstream super enhancers for *MYC* expression in G3-tumors. In addition, ChIP-qPCR assays showed that the level of H3K27ac occupancy on the enhancer/promoter in the loci of *MYC* and *OTX2*, another G3-MB-oncogenesis-associated gene, was reduced substantially when either SOX11 or HNRNPH1 was knocked-down in G3 MB cells to a level nearing that of hNSCs (Fig. 6j and Extended data Fig. 8b). In contrast, this long-range interaction loop conformation was not detected in G4-MB cells (Fig. 6h), suggesting a unique enhancer hijacking event in the *MYC* locus for driving the oncogenic process in G3 MB.

**TCP-like signature gene expression is critical for G3 MB growth**

Based on the TCGA dataset, the expression level of TCP marker *HNRNPH1* was higher in G3 MB than other MB subgroups (WNT, SHH, and G4), whereas *SOX11* expression was the highest in G4 MB (Fig. 7a). Notably, patients with high TCP scores (calculated for each patient as the weighted sum of expression of TCP signature genes) in G3 MB had worse prognosis than those with low TCP scores (Fig. 7b), while the prognostic impact was not observed in other MB subgroups (WNT, SHH and G4 MB) (Extended data Fig. 9a-c).

To verify that protein products of the TCP marker genes are also expressed, we performed immunostaining for HNRNPH1 and SOX11 in a set of human MB tissues. We detected expression of both HNRNPH1 and SOX11 in G3- and G4-MB tissues (Fig. 7c) and co-expression of HNRNPH1 and SOX11 in the nucleus of a population of cells in G3 MBs (Fig. 7d). Given that HNRNPH1 and SOX11 also co-occupied CREs in G3-MB cells, we performed a proximity ligation assay (in situ PLA) to determine if SOX11 and HNRNPH1 can physically interact. Expression of HNRNPH1 and SOX11 was detected in the same foci in the nuclei of G3-tumor cells, and depletion of HNRNPH1 abrogated the PLA signal (Fig. 7e). These data suggest that HNRNPH1 can directly interact with SOX11 and is critical for the complex formation in G3-MB tumor cells.

To determine the roles of HNRNPH1 and SOX11 in the growth of G3- and G4-MB tumor cells, we depleted tumor cells of HNRNPH1 or SOX11 using shRNAs expressed by lentivirus vectors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We found that silencing of *HNRNPH1* or *SOX11* in a patient-derived G3-MB cell line (MB-004) substantially reduced expression of MYC (Fig. 7f) and the genes related to G3 MB-associated signature, EMT, and TGF-beta signaling, while increasing expression of neuronal differentiation genes (Extended data Fig. 10a-c). In addition, knockdown of HNRNPH1 or SOX11 inhibited tumor sphere formation in G3-MB-derived cell lines MB-004 and D425 (Fig. 7g, h). Moreover, the growth of patient-derived G3-MB cell lines (MB-002 and MB-004) as well as a patient-derived G4-MB cell line (UPN3550) was substantially reduced upon *HNRNPH1* or *SOX11* silencing (Fig. 7i), in contrast, depletion of either factor did not substantially impair the growth of normal human NSCs or astrocytes. Silencing HNRNPH1 or SOX11 also significantly increased cell apoptosis (Fig. 7j). Furthermore, an *in vivo* xenograft analysis showed that SOX11 or HNRNPH1 depletion significantly inhibited the growth of tumors from the MYC-driven G3 MB-004 cells and prolonged animal survival (Fig. 7k, l). The tumors depleted of HNRNPH1 or SOX11 had decreased proportions of Ki67+ proliferative cells, while an increase in apoptosis detected by cleaved caspase 3 staining (Fig. 7m and Extended data Fig. 10d). These results suggest that expression of the TCP signature genes *HNRNPH1* and *SOX11* is critical for the growth of aggressive G3 MB tumors.

## Discussion

Spatiotemporal dynamics of cellular states, lineage transitions, and their functional plasticity during human cerebellar development and tumorigenesis are not fully understood [1,9]. By scRNA-seq of whole cells from freshly isolated fetal tissues and immunohistochemistry-based validations, we identified a previously unrecognized transitional progenitor population in the human fetal cerebellum. These cerebellar TCP cells were enriched during a narrow time window around the first-to-second-trimester transition stage PCW 12, a critical period critical for neuronal lineage specification, proliferation, and migration [14,54,55], and diminished thereafter. Lineage trajectory analyses predict that TCP cells exhibit features of undifferentiated and transitory cell states with the potential to give rise to different cerebellar cell types including GCP, UBC, and Purkinje cell lineages.

A recent study using single-nucleus snRNA-seq profiling from frozen fetal cerebellar tissues

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

identified different neural cell types but failed to identify the TCP subpopulation, although human RL cells were identified and characterized[13]. Our TCP gene signature partially overlaps that of RL cells[13], yet the populations are distinct. Whole-cell scRNA-seq from fresh tissues used in the present study has been shown to identify cell types more representative of cell populations of the starting tissues than does snRNA-seq [20,21], which may account for study differences. Further analysis is needed to determine differential study outcomes.

Our cross-species analyses of single-cell datasets between human and mouse cerebella do not detect an appreciable TCP-like population in mice. Although sparse SOX11+/HNRNPH1+ cells are distributed throughout mouse embryonic cerebella (Extended data Fig. 11), they are not enriched in the RL region, in stark contrast to robust developmental enrichment of TCP cells in the human cerebellar transitional zone of the RL, which is evolutionarily expanded in humans [14]. This suggests that TCP cells might represent cerebellar intermediate precursors similar to transit-amplifying progenitors for human neocortical expansion [56]. Such differences across species might explain why mouse model systems do not fully recapitulate human MB. Thus, the developmental hierarchy and plasticity of diverse progenitors we uncovered here may provide an important framework for deciphering the cell types critical for normal physiological functions and understanding perturbations that lead to MB.

Currently, the developmental origin and evolution of G3 MB remain elusive. By intersecting cellular states across developing fetal cerebella and MB subgroups, we discovered that TCP and tumor cell populations are interconnected by tumor-subtype-specific transitory states in G3, G4, and SHH MB, suggesting that the TCP-like cells serve as a transition state in these MB subgroups. Although it is possible that malignant transformation could occur in other lineage precursors through de-differentiation into a TCP-like cell as a cancer stem cell, our integrated single-cell omics and lineage trajectory analyses indicate that TCP-like cells may transition toward G3 tumors and might serve as a point of a tumor cell origin during G3-MB tumorigenesis. The subgroup-specific transitory TCP-like progenitors within different MBs may reflect intrinsic oncogenic mutations and cellular plasticity of TCP cells along distinct lineage trajectories, which may contribute to inter- and intra-tumoral heterogeneity as well as therapy resistance [15,16,57]. Our intersectional studies provide

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

new insights into the dynamics of cerebellar lineage transitions along with potential cell origins of aggressive MB, suggesting that TCP cells in early human fetal cerebella are predisposed to malignant transformation during MB tumorigenesis.

A recent study using SMART-seq suggests a cell-state continuum among G3 and G4 MBs based on a set of selected genes [8]. However, our unbiased single-cell clustering predicts distinct populations of prototypical G3 and G4 tumor cells within a cohort of G3 and G4 tumors, rather than a continuous spectrum between G3 and G4 tumor cells. The underlying discrepancy remains to be determined. The small set of the predefined genes might potentially diminish the distinction between cell types since they are also expressed in both G3 and G4 tumors to a certain extent [8], while a much higher number of cells and genes are assessed in our 10x Genomics sequencing platform. Our observation suggests that G3 and G4 MB cell populations might not interconvert, although our data does not exclude the possibility that a less differentiated progenitor cell is common to G3 and G4 MB tumors.

Metastasis is associated with poor prognosis and highly associated with G3 or G4 MBs[43,58]. The molecular and cellular nature of metastatic driving events in MB remain elusive. Our single-cell profiling of paired primary and metastatic MBs revealed a substantial increase in the proportion of transitional TCP-cells and *MYC*$^{+}$ tumor cells in the metastatic tumors, suggesting that the TCP-like subpopulation and G3 tumor lineage cells drive metastatic tumor formation. We found that a TCP-like cell signature is associated with poor prognosis in G3 MB but not G4 MB. The higher proportions of the TCP-like cells or MYC$^{+}$ TCP-like cells in G3 MBs compared to G4 MBs (Fig. 3e-h) might contribute to the difference in survival outcomes. In addition, the lower rhombic-lip-derived WNT MB, which has the best prognosis, does not appear to show an enrichment of the TCP-like cluster [59] (Extended data Fig. 12). Nonetheless, our findings at single-cell resolution suggest that tumor cells endowed with TCP-like characteristics might potentiate MB metastasis.

Integrated single-cell omics (scRNA-seq/snATAC-seq) analyses identified transcriptional regulatory networks in TCP-like populations and potential therapeutic targets for aggressive MBs. Targeting TCP-like cells through depletion of HNRNPH1 or SOX11 inhibited the growth of the aggressive G3 MB tumor cells. Moreover, 3D chromatin structure analysis revealed long-distance

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spatial looping of HNRNPH1/SOX11-bound super-enhancers juxtaposed to *MYC* promoter/enhancer elements, which was uniquely present in *MYC*-driven G3 tumor cells but not G4 tumor or hNSC cells. Thus, TCP cell identity determinants (HNRNPH1 and SOX11), which are upregulated in multiple cancers [38,39], not only define the TCP-like state but also can harbor long-range super-enhancers to promote expression of oncogenes such as MYC through enhancer hijacking events, suggesting that TCP cells might be highly susceptible to oncogenic transformation. Together, our data provide insights into the potential origin, lineage plasticity, and human-specific nature of MB subtypes, as well as their intra- and inter-tumoral heterogeneity in malignancy, while revealing a targetable vulnerability for therapeutic intervention of aggressive MB.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Acknowledgements**

The authors would like to thank Drs. Ed Hurlock, Tim Phoenix, and Emery Lu for comments and suggestions. We appreciate Arman Esshaghi Bayat for technical support. This study was supported in part by grants from the CancerFree Kids foundation, Pray-Hope-Believe foundation, and TeamConnor Childhood Cancer Foundation to QRL, and the grant from NIH R37NS095733 to KJM.

**Competing interest declaration:** The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Methods

## Human tissue collection

Human fetal and tumor tissues were obtained from the Children's Hospital of Fudan University and XinHua Hospital at the Shanghai Jiao Tong University School of Medicine, and Obstetrics and Gynecology Hospital of Fudan University. Informed consents for the use of tissues for research were obtained in writing from donors or the patients' parents. The study was approved by individual institutional review board at the Children's Hospital of Fudan University, XinHua Hospital at the Shanghai Jiao Tong University School of Medicine, and Obstetrics and Gynecology Hospital of Fudan University. Fresh cerebellar tissues from aborted fetuses and tumors after surgery were collected and digested by Collagenase IV (2μg/ml, Thermo Fisher, #17104019) enzymatic dissociation for 20 min at 37 °C after mechanically dissociation. Clinical information including age, sex, localization and MB subgroup are provided in Supplementary Table 1.

## Animal experiments

Immunodeficient NOD SCID gamma (NSG) mice were obtained from the Cincinnati Children's Hospital Medical Center (CCHMC) animal core. Mice of either sex were used and fed (4 or less mice per cage) in a vivarium with a 12 hrs light/ 12 hrs dark cycle. All studies complied with all relevant animal use guidelines and ethical regulations. The animal studies were approved by the IACUC (Institutional Animal Care and Use Committees) of the Cincinnati Children's Hospital Medical Center, USA. In the xenograft model, MB tumor cells were transduced with lentivirus targeting HNRNPH1 or SOX11 for 20 hours, and $2\times10^5$ cells suspended in 3μl PBS with 1μl Matrigel (Corning, #356234) were stereotactically injected into the cerebellum of NSG mice. Animals were monitored and bioluminescence images were captured weekly and harvested when they become moribund.

## Medulloblastoma cell line culture

Medulloblastoma cell lines D425, D283 and D458 were obtained from American Tissue Culture Collection (ATCC, Rockville, MD, USA). UPN3550 cells were isolated from a group 4 MB patient's primary tumor tissue, which was proved by the institutional reviewing board at the Cincinnati Children's Hospital. MB-002 and MB-004 cells were provided by Drs. Rachid Drissi and Martine Roussel at St Jude Children's Research Hospital. Cells were cultured in neurobasal medium (Sigma; SCM003) with 2% B-27, 1 μg/ml heparin, 2 mM L-glutamine and 1% penicillin/streptomycin, 25 ng/ml FGF and 25 ng/ml EGF at 37℃ in an atmosphere of 5% $CO_2$.

## Immunostaining, immunohistochemistry and immunoblotting

The immunostaining procedures were followed the method as previously described[60]. Briefly, cerebellar tissues were fixed with 4% PFA for 45 min and washed five times with PBS and dehydrated with 30% sucrose overnight, then blocked with OCT frozen embedding media (CRYO-4, Polarstat Inc.) and cryosectioned at 14 μm thickness. For adherent cells, cells were seeded on the coverslips and fixed with 4% PFA for 10 min and washed five times with PBS, then in blocking solution for 30 min. We used primary antibodies, including mouse anti-SOX2 (Santa Cruz Biotechnology; Cat#sc-365964), rabbit anti-SOX11 (Sigma, Cat#HPA000536; Thermo fisher, Cat#14-9773-82), rabbit anti-HNRNPH1 (Abcam, Cat#ab154894; Bethyl Laboratories, Cat#A300-511A), rat anti-EOMES (Invitrogen, Cat#14-4875-52), mouse anti-PAX6 (Santa Cruz

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Biotechnology, Cat#sc-81649), mouse anti-ATOH1 (Thermo Fishe,; Cat#H00000474-M09), rabbit anti-c-Myc (Cell Signaling, Cat#5605), rabbit anti-Ki67 (Thermo Fisher, Cat#MA5-14520), rabbit anti-Cleaved Caspase 3 (Cell Signaling, Cat#9661) and mouse-anti BrdU (BD Bioscience, 1:500) antibody with proper dilutions. For BrdU staining, BrdU pulse-labeled (10 μM, 2 hours at 37□) cells were denatured with 0.1N HCl for 1 h in water bath at 37 °C. After denaturation, cells were neutralized with 0.1 M Borax, pH 8.5 (Sigma) for 10 min. Cells were washed with PBS 3 times and blocked with 5% normal donkey serum (Sigma-Aldrich) in wash buffer for 1 h at room temperature. Secondary antibodies conjugated to Cy2, Cy3 or Cy5 were from Jackson ImmunoResearch Laboratories. Tissues or cells were mounted with Fluoromount-G (SouthernBiotech) for microscopy. Immunofluorescence-labeled images were acquired using a Nikon $C2^{+}$ confocal microscope. Cell images were quantified in a blinded manner.

For paraffin-embedded tissues, sections were dewaxed and hydrated using xylene and ethanol respectively. We performed antigen retrieval before permeabilization as previously described[60]. Slides were treated in 0.6% H2O2 in methanol for 30min at 37℃ and blocked in 5% normal donkey serum in PBST for 1h at room temperature. SOX11 and HNRNPH1-expressing cells in MB tissues were quantified using the described methods [61]. In brief, 0–5 denote different degrees (intensity and density) of IHC staining; 5 is the maximum and 0 is the minimum degree. The final score of the patients = SI (score of intensity) × SD (score of density).

For the western blot analysis, cells were lysed with RIPA lysis buffer (Millipore) supplemented with phosphatase and protease inhibitor cocktail (Roche). Protein concentration of each sample was determined by BCA assay using the BCA kit (Beyotime) according to manufacturer's instructions and equal amounts (5-15 μl) were loaded and separated by 12% SDS-PAGE gel. PVDF membrane (Millipore) was used for gel transfer and the membrane was probed with primary antibodies as indicated, followed by secondary antibodies conjugated with HRP. The signal was detected with Super Signal West Pico/Femto Chemiluminescent Substrate (Thermo Scientific).

**Generation and processing of DNA methylation data**

All single-cell MB patient samples sequenced in this study were analyzed using Illumina Infinium Methylation EPIC BeadChip arrays according to the manufacturer's instructions. Data were generated from total genome DNA isolated from freshly frozen tissue samples. Medulloblastoma subgroup predictions were obtained from a web-platform for DNA methylation-based classification of central nervous system tumors (https://www.molecularneuropathology.org/mnp). Resulting assignment of samples to SHH, Group 3 and Group 4 subgroups were used for all downstream analyses. CNV analysis from EPIC methylation array data was performed using the conumee Bioconductor package (http://bioconductor.org/packages/conumee/).

**scRNA-seq and scATAC-seq using 10x Genomics platform**

For scRNA-seq on the 10x Genomics platform, single cells were processed through the GemCode Single Cell Platform using the GemCode Gel Bead, Chip and Library Kits (10X Genomics) according to the manufacturer's instructions. The concentration of the single-cell suspension was assessed with a Trypan blue count and the sample will be used if there are more than 90% viable cells. Approximately 10,000–30,000 cells per sample were loaded on the Chromium Controller and generated single-cell GEMs (gel beads in emulsion). GEM-reverse-transcription, DynaBeads clean-up, PCR amplification and SPRIselect beads clean-up were performed using Chromium Single Cell 3′ Gel Bead kit. Indexed single-cell libraries were generated

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using the Chromium Single Cell 3′ Library kit and the Chromium i7 Multiplex kit. Size, quality, concentration and purity of the cDNAs and the corresponding 10×library was evaluated by the Agilent 2100 Bioanalyzer system. Amplified cDNA and final libraries were assessed on an Agilent BioAnalyzer using a High Sensitivity DNA Kit (Agilent Technologies).

For snATAC-seq on the 10x Genomics platform, single-cell libraries were generated using the GemCode Single-cell instruments and the Single Cell ATAC Library & Gel Bead Kit and ChIP Kit from 10x Genomics, according to the manufacturer's instructions. The samples were incubated at 37°C for 1 h with 10 µl of transposition mix (per reaction, 7 µl ATAC Buffer, and 3 µl ATAC Enzyme (10x Genomics)). Following the generation of nanoliter-scale GEMs, GEMs were reverse transcribed in a C1000 Touch Thermal Cycler (Bio-Rad) programmed at 72°C for 5 min, 98°C for 30 s, 12 cycles of 98°C for 10 s, 59°C for 30 s, and 72°C for 1 min, and held at 15°C. After reverse transcription, single-cell droplets were broken and the single-strand cDNA was isolated, cleaned up and amplified. Amplified cDNA and final libraries were assessed on an Agilent BioAnalyzer using a High Sensitivity DNA Kit (Agilent Technologies). All the libraries were sequenced on NovaSeq 6000 (Illumina) at a depth of approximately 400M reads per sample.

**scRNA-Seq processing and quality filtering**

For 10X genomics datasets, we used Cellranger v5.0.1 to align reads to the human reference sequence. The raw base call (BCL) files were demultiplexed into FASTQ files. The FASTQ files were aligned to the reference human genome GRCh38 (hg38) to generate raw gene-barcode count matrices. When clustering multiple samples together, we aggregated the multiple runs together to normalize on sequencing depth and re-computed the gene-barcode matrices.

For quality control and normalization of scRNA-seq**,** we utilize the Seurat program (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html) in R version 4.0.3 by reading in the data which are the reads in the output of the cellranger pipeline prom 10x, returning a unique molecular identified (UMI) count matrix. Low-quality cells were identified and removed from the datasets based on the cell with <200 genes expression and high mitochondrial gene content (5 s.d. above median). Doublets were detected and filtered using the R package DoubletFinder v.2.0.2 with default settings. The cells with low-abundance genes or genes expressed in <3 cells were also removed from the datasets. By defaulting in Seurat, we employ a global-scaling normalization method "LogNormalize" that normalizes the feature expression measurements for each cell by the total expression, multiplies this by a scale factor (10,000 by default), and log-transforms the result. Next, we apply a linear transformation ('scaling') that is a standard pre-processing step using all genes or variable genes. Then, we perform principal component analysis (PCA) to get the linear dimensional reduction after the data scaling.

**Clustering analysis, visualization and annotation**

Clustering analysis was performed with the R package Seurat (v.4.0.3). Highly variable genes were detected using Seurat's pipeline[62], calculating average expression and dispersion for each gene, diving genes into bins and computing a z-score for dispersion within each bin. We used a z-score of 0.5 as the cut-off of dispersion, and a bottom cut-off of 0.0125 and a high cut-off of 3.0 for average expression. Linear dimensionality reduction was performed using PCA, and statistically significant principal components were selected using the elbow and jackstraw methods from Seurat. The clusters of cells were identified by a shared nearest neighbor (SNN)-modularity-optimization based clustering algorithm from Seurat. We then visualized these clusters using t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), or Monocle 3. Cluster cell identity was assigned by manual annotation using known cell-type marker

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genes and computed differentially expressed genes (DEGs) using the FindAllMarkers function in the Seurat package (one-tailed Wilcoxon rank sum test, P values adjusted for multiple testing using the Bonferroni correction. For selecting DEGs, all genes were probed provided they were expressed in at least 25% of cells in either of the two populations compared and the expression difference on a natural log scale was at least 0.2. Manual annotation was performed iteratively, which included validating proposed cell clusters with known markers and further investigating clusters for which the gene signatures indicated additional diversity.

**Pseudo-time cell state trajectory analysis**

For the fetal cerebellum trajectory analysis, cells were grouped using 'UMAP' clustering algorithm. Cell state transition directions were inferred by Monocle 3, STREAM, or VECTOR programs which provides an unsupervised solution for determining the starting cells. For order_cells function in Monocle 3, the barcodes of selected clusters were normalized using Monocle dPFeature or Seurat to remove genes with low expression and perform PCA analysis on the remaining genes, for significant principal components (PCs) selection[63,64]. Differential gene expression analysis was performed using a generalized linear model, and the top 1,000 genes per cluster were selected and fit into a principal graph within each partition using the learn_graph function. For Slingshot cellular trajectory analysis [30] of fetal cerebella, the input matrix was filtered and normalized by the R package Seurat and cell types were annotated and provided as labels for Slingshot. For the single-cell pseudo-time trajectory in tumor tissues, cells were aggregated the multiple patients together to normalize on sequencing depth and re-computed the gene-barcode matrices using canonical correlation analysis (CCA).

**Deconvolution analysis**

CIBERSORTx[34] was applied to perform the deconvolution analysis of the bulk and scRNA-seq tumor data against the human cerebellar clusters except mitotic cells. The transcriptomes of the tumor data (bulk RNA-seq or clusters of scRNA-seq) were used as the input mixture online (https://cibersortx.stanford.edu/runcibersortx.php), and the signature matrix input was the human fetal cerebellum cluster expression matrix after removal of cell cycle related genes (~1,400), ribosome biogenesis genes (~300), mitochondrial and apoptosis-related genes (~100), to avoid bias in the deconvolution process. Quantile normalization was disabled and 100–500 permutations for significance analysis were run.

**Cell cycle analysis of human scRNA-seq tumor samples**

Cell cycle phase-specific annotations were used to define the cell cycle status for each individual cell[65]. We assign each cell a score using CellCycleScoring function in R version 4.0.5, based on its expression of G2/M and S phase markers. These marker sets should be anticorrelated in their expression levels, and cells expressing neither are likely not cycling and in G1 phase.

**Inferred CNV Analysis from scRNA-seq**

Malignant cells were identified by inferring large-scale chromosomal copy-number variations (CNVs) in each single cell based on a moving averaged expression profiles across chromosomal intervals by inferCNV [66]. We combined CNV classification with transcriptomic-based clustering and expression of non-malignant marker genes to identify malignant and non-malignant cells. Non-malignant cells displayed high expression of specific marker genes and no apparent CNVs.

**Filtering cells by TSS enrichment and unique fragments of the scATAC-seq**

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Enrichment of ATAC-seq accessibility at TSSs was used to quantify data quality without the need for a defined peak set. Calculating enrichment at TSSs was performed as previously described[67], and TSS positions were acquired from the Bioconductor package from 'TxDb. Hsapiens.UCSC.hg38.knownGene. Briefly, Tn5-corrected insertions were aggregated ±2,000 bp relative to each unique TSS genome-wide (TSS strand-corrected). The calculated TSS enrichment represents the maximum of the smoothed profile at the TSS. We then filtered all scATAC-seq profiles to keep those that had at least 1000 unique fragments and a TSS enrichment of 0.5. To minimize the contribution of potential doublets to our analysis, we removed snATAC-seq profiles that had more than 100,000 unique nuclear fragments.

**Gene regulatory network and motif enrichment analysis of scRNA-seq and scATAC-seq data**

To characterize underlying gene regulatory network and infer transcription factor activities in our scRNA-seq dataset, we used the single-cell regulatory network inference and clustering package to identify gene regulatory modules and retain those with a cis-regulatory binding motif for upstream regulators. We merged scRNA-seq and scATAC-seq datasets to create a common peak set, and quantify this peak set in each experiment. We load the peak coordinates for each experiment and convert them to genomic ranges using the GenomicRanges::reduce function to create a common set of peaks to quantify in each dataset. We use the detailed settings and parameters as default according to Signac (https://satijalab.org/signac/). ArchR package [44] was used for integrated scRNA-seq and scATAC-seq analyses according to default parameters, including quality control and cell filtering, dimension reduction, genome browser visualization, gene expression data preprocessing and cell annotation, DNA accessibility data processing, joint data visualization, differential accessibility and motif enrichment.

For nominating the marker genes and potential drivers, we utilized ArchR [44] to identify the enriched TFs whose inferred gene scores are correlated to their chromVAR TF deviation z-scores. The gene scores were calculated based on the summed chromatin accessibility and normalized across all genes to a user-defined constant (default of 10,000) according to the ArchR package [44]. Based on the gene scores and positive TF-regulators identified from ArchR, we nominated top 30 TFs or highly expressed genes (excluding non-coding genes or ribosomal proteins) as potential drivers or marker genes.

**Two-Dimensional Representation of SC-Derived Cell States and Group 3/4-B/C score**

Tumor cell clusters were used for computing subtype expression scores for each tumor cell in the datasets as previously described [41]. Cell clusters were separated into G3-MYC, G3-NRL, G4- and SHH-clusters. To visualize the enrichment of subsets of cells, across the two-dimensional representation, we calculated for each cell the fraction of cells that belong to the respective subset among its 100 nearest neighbors, as defined by Euclidean distance, and these fractions were displayed by colors. In addition, we used the previously defined group 3/4 B/C score system based on selected G3- and G4-expressing genes (top 30 genes from each metaprogram) [8] for the overlapping cell-state analysis in G3 and G4 MBs.

**Generation of Hi-C libraries and analysis**

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

MB004 (G3 MB) and UPN3550 (G4 MB) cells were processed for Hi-C library construction using the Arima Hi-C Kit following the manual (Arima Genomics, # A510008). Briefly, five million cells were cross-linked with 1% formaldehyde for 10 mins at room temperature and then quenched with 0.2M Glycine. Cell pellets were washed with cold PBS and lysed with lysis buffer to release nuclei and then permeabilized and in situ digested. KAPA Hyper Prep kit was used for library amplification (KAPA, KK2802). Hi-C libraries were sequenced 2×150bp on a NovaSeq 6000 instrument (Illumina). Juicer were used to process raw reads and generate Hi-C contact matrices (.hic files), aligning to reference genome hg38 to generate Hi-C contact matrices (.hic files). Contact matrices were visualized using Juicebox.

Bam files were used as input, with low-quality reads filtered out. We used Peakachu and diffPeakachu to call and compare loops in Hi-C data from both G3 MB and G4 MB cell lines in Hi-C data, then used diffPeakachu to compare one cell line with another cell line. Tumor subtype–specific loops were then merged using BEDTools pair to pair function with a negative slop of 10 kb. Hi-C breakfinder pipeline [68] was used to identify large structural translocations, deletions, and inversions. To identify neo-loops or enhancer hijacking events, we used a NeoLoopFinder pipeline as previously described algorithm [51].

**Targeting distal interacting enhancers using CRISPRi**

MB-004 cells were transduced with the enti-dCas9-KRAB-T2A-GFP virus (Addgene # 71237). sgRNAs targeting SOX11/HNRNPH1-binding motifs in the distal enhancers were designed using the CHOPCHOP program (https://chopchop.cbu.uib.no). GFP-reporter positive cells were flow-sorted 2 days transduction. DNA oligonucleotides were annealed and ligated into the lentiGuide-Cherry vector (Addgene # 170510) at the BsmBI restriction enzyme cutting sites. The sgRNA sequences were confirmed by Sanger sequencing. The lenti-vectors expressing each pair of gRNAs targeting distal enhancers were packaged in 293T cells using pMD2.G (Addgene # 12259) and psPAX (Addgene # 12260). The lentiGuide-Cherry lentiviruses carrying sgRNAs were concentrated from the virus-containing medium by ultracentrifuging and transduced into dCas9-KRAB-T2A-GFP-expressing MB-004 cells (MOI < 1). RNAs were then extracted from the GFP+/Cherry+ cells after 72 hr culture, and cDNAs were prepared using SuperScript® III First-Strand Synthesis System (Invitrogen) according to the manufacturer's instructions. qRT-PCR was performed to quantify gene expression using SYBR FAST qPCR Master Mix. All sgRNA and qRT-PCR sequences used for validation can be found in Supplementary Table 5.

**ChIP–qPCR**

ChIP-qPCR assays were performed as previously described [69]. Briefly, cells (2 x $10^6$) were transduced with lentivirus expressing non-targeting shRNA control, shHNRHNP1, or shSOX11 for 60 hours (selected with puromycin) and then fixed with 1% formaldehyde for 10 min at room temperature and quenched with 0.2M glycine. Sonicated chromatin was prepared in buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA and protease inhibitor cocktail) and then incubated with 4 μg H3k27ac antibody (Abcam, ab4729) overnight at 4 C. Magnetic protein A/G beads were incubated to each ChIP reaction for 1 hr at 4 C. ChIP DNAs were eluted into 200 μl elution buffer at 65 °C for 20 min and extracted with phenol/chloroform. Purified DNAs were subjected to qRT-PCR assay for quantifying H3K27ac occupancy on the enhancers. Sequences of ChIP-qPCR primers are listed in Supplementary Table 5.

**Cut & Run-seq and data processing**

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cut & Run-seq was performed as previously described[60]. Briefly, 200,000 cells were harvested and washed twice and captured by the addition of 10 µl pre-activated concanavalin A coated magnetic beads (Bangs Laboratories-BP531). Cells were then resuspended in 100 µl cold Antibody Buffer and 1 µl antibody (H3K4me3, EpiCypher #13-0041; H3K27Ac, Active Motif #39133; SOX11, Sigma #HPA000536 or HNRNPH1 Abcam #ab154894) was added for incubating on nutator overnight at 4 ℃. Cells were washed twice in 1 ml Digitonin Buffer (20 mM HEPES-KOH pH 7.5; 150 mM NaCl; 0.5 mM Spermidine; 1×Roche cOmpleteTM; 0.05% digitonin), and then resuspended with CUTANA pAG-MNase in 50 µl Digitonin Buffer. After wash twice, samples were quickly mixed with 100 mM $CaCl_2$ to a final concentration of 2 mM and incubation at 4℃ for 30 min and reaction was quenched by Stop Buffer. The cleaved fragments were released by incubating the tube for 30 min at 37 ℃ and then purified by Qiagen MinElute PCR Purification Kit. Libraries were prepared with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (E7645) and sequenced by NovaSeq PE150.

Cut & Run-seq reads were aligned to the reference human genome version hg38 with the program BOWTIE v2.3.4.1. Aligned reads were stripped of duplicate reads with the program sambamba v0.6.8. Peaks were called using the program MACS v2.1.2 with the narrow and broad peaks mode for Cut & Run-seq. Motif enrichment analysis was performed for both HNRNPH1- and SOX11-bound sites using HOMER findMotifsGenome function with -size 1000 -mask settings. TFs with high-expression level in G3 MB cell lines and significant H3K27Ac enrichment in G3 MB-specific enhancers were identified as G3 MB-specific active motifs.

**Statistical and Survival Analysis**

All analyses in this research were performed using Microsoft Excel, GraphPad Prism 6.00 (San Diego California, https://www.graphpad.com) or RStudio (https://www.rstudio.com/ and R v.4.0.3, R Development Core Team, 2016). We use the “cor” function in R to calculate the Pearson correlation coefficient. Statistical significance was determined using two-tailed Student’s t tests as indicated. One-way ANOVA test was performed by multiple comparisons following Turkey’s ranking tests when comparing multiple groups. Data are shown as mean ± SEM. Values of $p < 0.05$ denoted a statistically significant difference. Quantifications were performed from at least three experimental groups in a blinded fashion. The n value was defined as the number of experiments that were repeated independently with similar results. Animal survival time is the date of euthanasia when the mouse showed signs of death based on the suggestions of veterinarian. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those generally employed in the field. No randomization was used to collect all the data, but data were quantified with blinding.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## References:


1. Wang, J., Garancher, A., Ramaswamy, V. & Wechsler-Reya, R.J. Medulloblastoma: From Molecular Subgroups to Molecular Targeted Therapies. *Annu Rev Neurosci* **41**, 207-232 (2018).
2. Beccaria, K., Padovani, L., Bouchoucha, Y. & Doz, F. Current treatments of medulloblastoma. *Curr Opin Oncol* **33**, 615-620 (2021).
3. Cavalli, F.M.G., *et al.* Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **31**, 737-754.e736 (2017).
4. Gibson, P., *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095-1099 (2010).
5. Leto, K., *et al.* Consensus Paper: Cerebellar Development. *Cerebellum* **15**, 789-828 (2016).
6. Haldipur, P., Dang, D. & Millen, K.J. Embryology. *Handb Clin Neurol* **154**, 29-44 (2018).
7. Vladoiu, M.C., *et al.* Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67-73 (2019).
8. Hovestadt, V., *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74-79 (2019).
9. Roussel, M.F. & Hatten, M.E. Cerebellum development and medulloblastoma. *Curr Top Dev Biol* **94**, 235-282 (2011).
10. Lange, W. Cell number and cell density in the cerebellar cortex of man and some other mammals. *Cell Tissue Res* **157**, 115-124 (1975).
11. Van Essen, D.C. Surface-based atlases of cerebellar cortex in the human, macaque, and mouse. *Ann N Y Acad Sci* **978**, 468-479 (2002).
12. Sillitoe, R.V. & Joyner, A.L. Morphology, molecular codes, and circuitry produce the three-dimensional complexity of the cerebellum. *Annu Rev Cell Dev Biol* **23**, 549-577 (2007).
13. Aldinger, K.A., *et al.* Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci* **24**, 1163-1175 (2021).
14. Haldipur, P., *et al.* Spatiotemporal expansion of primary progenitor zones in the developing human cerebellum. *Science* **366**, 454-460 (2019).
15. Yuan, S., Norgard, R.J. & Stanger, B.Z. Cellular Plasticity in Cancer. *Cancer Discov* **9**, 837-851 (2019).
16. Manoranjan, B., *et al.* Medulloblastoma stem cells: where development and cancer cross pathways. *Pediatr Res* **71**, 516-522 (2012).
17. Meacham, C.E. & Morrison, S.J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328-337 (2013).
18. Johnson, M.B., *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494-509 (2009).
19. Miller, J.A., *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. Slyper, M., *et al*. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802 (2020).
21. Fischer, J. & Ayers, T. Single nucleus RNA-sequencing: how it's done, applications and limitations. *Emerg Top Life Sci* (2021).
22. Jessa, S., *et al*. Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet* **51**, 1702-1713 (2019).
23. Robinson, G., *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* (2012).
24. Kool, M., *et al.* Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. *Acta Neuropathol* **123**, 473-484 (2012).
25. Pugh, T.J., *et al*. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* (2012).
26. McGinnis, C.S., Murrow, L.M. & Gartner, Z.J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019).
27. Macosko, E.Z., *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
28. Qiu, X., *et al*. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).
29. Chen, H., *et al*. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat Commun* **10**, 1903 (2019).
30. Street, K., *et al*. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics* **19**, 477 (2018).
31. Abdelaal, T., *et al*. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol* **20**, 194 (2019).
32. Welch, J.D., *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873-1887 e1817 (2019).
33. Yeung, J., Ha, T.J., Swanson, D.J. & Goldowitz, D. A Novel and Multivalent Role of Pax6 in Cerebellar Development. *J Neurosci* **36**, 9057-9069 (2016).
34. Newman, A.M., *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nature biotechnology* **37**, 773-782 (2019).
35. Zhang, F., Li, X. & Tian, W. Unsupervised Inference of Developmental Directions for Single Cells Using VECTOR. *Cell Rep* **32**, 108069 (2020).
36. Andreatta, M., *et al*. Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun* **12**, 2965 (2021).
37. Steen, C.B., Liu, C.L., Alizadeh, A.A. & Newman, A.M. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol* **2117**, 135-157 (2020).
38. Liu, M., *et al.* HNRNPH1 Is a Novel Regulator Of Cellular Proliferation and Disease Progression in Chronic Myeloid Leukemia. *Front Oncol* **11**, 682859 (2021).
39. Oliemuller, E., *et al.* SOX11 promotes epithelial/mesenchymal hybrid state and alters tropism of invasive breast cancer cells. *Elife* **9**(2020).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

40. Patmore, D.M., *et al.* DDX3X Suppresses the Susceptibility of Hindbrain Lineages to Medulloblastoma. *Dev Cell* **54**, 455-470 e455 (2020).
41. Neftel, C., *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849 e821 (2019).
42. Northcott, P.A., *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311-317 (2017).
43. Zapotocky, M., *et al.* Differential patterns of metastatic dissemination across medulloblastoma subgroups. *J Neurosurg Pediatr* **21**, 145-152 (2018).
44. Granja, J.M., *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature genetics* **53**, 403-411 (2021).
45. Pliner, H.A., *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858 (2018).
46. Massague, J. & Ganesh, K. Metastasis-Initiating Cells and Ecosystems. *Cancer Discov* **11**, 971-994 (2021).
47. Cavalli, F.M.G., *et al.* Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **31**, 737-754 e736 (2017).
48. Meers, M.P., Bryson, T.D., Henikoff, J.G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**(2019).
49. Creyghton, M.P., *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010).
50. Lamar, J.M., *et al.* The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. *Proc Natl Acad Sci U S A* **109**, E2441-2450 (2012).
51. Wang, X., *et al.* Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. *Nat Methods* **18**, 661-668 (2021).
52. Riecken, L.B., *et al.* CPI-17 drives oncogenic Ras signaling in human melanomas via Ezrin-Radixin-Moesin family proteins. *Oncotarget* **7**, 78242-78254 (2016).
53. Hnisz, D., *et al.* Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* **58**, 362-370 (2015).
54. Volpe, J.J. Cerebellum of the premature infant: rapidly developing, vulnerable, clinically important. *J Child Neurol* **24**, 1085-1104 (2009).
55. de Graaf-Peters, V.B. & Hadders-Algra, M. Ontogeny of the human central nervous system: what is happening when? *Early Hum Dev* **82**, 257-266 (2006).
56. Lui, J.H., Hansen, D.V. & Kriegstein, A.R. Development and evolution of the human neocortex. *Cell* **146**, 18-36 (2011).
57. Barkley, D., Rao, A., Pour, M., Franca, G.S. & Yanai, I. Cancer cell states and emergent properties of the dynamic tumor system. *Genome Res* **31**, 1719-1727 (2021).
58. Chaffer, C.L. & Weinberg, R.A. A perspective on cancer cell metastasis. *Science* **331**, 1559-1564 (2011).
59. Riemondy, K.A., *et al.* Neoplastic and immune single-cell transcriptomics define subgroup-specific intra-tumoral heterogeneity of childhood medulloblastoma. *Neuro Oncol* **24**, 273-286 (2022).
60. Zhang, L., *et al.* Single-Cell Transcriptomics in Medulloblastoma Reveals Tumor-Initiating

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Progenitors and Oncogenic Cascades during Tumorigenesis and Relapse. *Cancer Cell* **36**, 302-318 e307 (2019).

61. Luo, Z.L., *et al.* A splicing variant of Merlin promotes metastasis in hepatocellular carcinoma. *Nat Commun* **6**, 8457 (2015).
62. Stuart, T., *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
63. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
64. Trapnell, C., *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).
65. Whitfield, M.L., *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**, 1977-2000 (2002).
66. Tirosh, I., *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309-313 (2016).
67. Corces, M.R., *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**(2018).
68. Wang, J., *et al.* Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. *Sci Adv* **7**(2021).
69. Wu, L.M.N., *et al.* Programming of Schwann Cells by Lats1/2-TAZ/YAP Signaling Drives Malignant Peripheral Nerve Sheath Tumorigenesis. *Cancer Cell* **33**, 292-308 e297 (2018).
70. Luo, Z., *et al.* Genomic and Transcriptomic Analyses Reveals ZNF124 as a Critical Regulator in Highly Aggressive Medulloblastomas. *Front Cell Dev Biol* **9**, 634056 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Figure Legends

### Fig. 1. Single-cell atlas of early developing human fetal cerebella

**a**, *t*-SNE visualization of transcriptionally distinct cell populations from 95,542 single cells in PCW 8-17 human fetal cerebellar tissue. TCPs are circled.

**b**, *t*-SNE maps of cell populations at indicated developmental stages. One tissue at each stage was used for the scRNA-seq analysis.

**c**, Left: t-SNE map of human cerebellar populations with neural cell types indicated in red. Middle: UMAP visualization of neural cell types from fetal cerebella. Right: Prediction of fate transitions using Monocle 3. TCPs are boxed.

**d**, Left: A zoom-in view of trajectories, Right: Heatmap with spline curves fitted to differentially expressed genes. Pseudo-temporal trajectories grouped by hierarchical clustering (k=3).

**e**, Upper: Proportion of each cell population at each developmental stage. Lower: Graph of proportion of TCPs at each developmental stage.

**f**, Median-scaled ln-normalized gene expression of selected markers for identified fetal cerebellum cell states.

**g**, Representative marker gene expression in NSC, TCP, GCP, UBC progenitor (UBC_P) or UBC, and Purkinje cells.

**h**, A matrix for cross-validation with an SVM classifier of the identified cell types between the mouse cerebellum and human fetal cerebellum.

### Fig. 2. Identification of a novel transitional progenitor intermediate in human fetal cerebella

**a**, Upper, hematoxylin and eosin (H&E) of midsagittal sections of PCW 9-16 human fetal cerebella. Scale bar: 500 μm. Middle and lower panels show immunostaining for SOX2/HNRNPH1 and SOX2/SOX11 from the boxed areas in the upper panels at the corresponding stages, respectively. Green arrows; the rhombic lip. Scale bar: 100 μm.

**b**, The percentage of cells positive for HNRNPH1 and SOX11 at indicted stages (two samples/stage). VZ, ventricular zone; SVZ, subventricular zone.

**c**, Midsagittal sections of PCW 12 human fetal cerebellum immunostained for HNRNPH1 and SOX11 in the RL transitional zone (TZ) (yellow boxed area). Scale bar: upper left, 500 μm; upper right and lower, 100 μm; inset, 30 μm. Arrows, co-labeled cells. RL, the rhombic lip. DAPI (blue), counterstain.

**d**, Quantification of TCP cells per 0.1 $mm^2$ in the TZ region from two samples/stage.

**e**, Immunostain for SOX11 and HNRNPH1 in the RL ($RL^{SVZ}$ and $RL^{VZ}$) from the red-boxed area in panel c at PCW 12. Scale bar: 100 μm.

**f**, Midsagittal sections of PCW 12 human fetal cerebellum immunostained for SOX11 and HNRNPH1 in the RL TZ region. DAPI (left), tissue counterstain. Scale bar: left, 500 μm, middle and right panels, 100 μm.

**g, h**, Representative images (g) and the quantification (h) showing PCW 12 fetal cerebella stained for HNRNPH1 and neuronal lineage markers PAX6 and EOMES. Boxed areas are shown at a high magnification in the right panels. Arrows indicated co-labeled cells. Scale bars: 500 μm (left); 50 μm (right panels). DAPI (blue), counterstain. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; two-tailed t-test; n = 3 tissues. Data are mean and s.d.

**i**, The EGL region of PCW 12 human fetal cerebellum stained for ATOH1 and HNRNPH1. Scale bar: left, 200 μm; right, 50 μm. Arrows indicate co-labeled cells.

**j**, Trajectory analyses from NSC to UBC (lower) and TCP to GCP (upper) lineages in the fetal cerebellum and representative marker gene expression in the different populations.

**k**, Schematic model showing a potential transition through the TCP population into GCP and UBC lineages during the development of human fetal cerebella.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Fig. 3. Single-cell transcriptomics reveals intermediate TCP-like progenitors in aggressive MB subgroups**

**a**, CIBERSORTx deconvolution analyses of human fetal cerebellar cell populations against bulk transcriptomes of human MB subgroups from CBTTC cohort.

**b-d**, UMAP visualization of b) G3 MB cells (n = 90,672 cells) from 9 patients, c) G4 MB cells (n = 45,731) from 13 patients, and d) SHH MB cells (n = 61,254 cells) from 4 patients. The cells are color-coded for indicated populations identified based on expression of representative marker genes.

**e**, Heatmap of cell populations in human MB subtypes identified based on comparisons to normal human fetal cerebella cell populations.

**f**, Predicted fraction of cells in human MB subgroup cohorts[42,70] using CIBERSORTx deconvolution.

**g**, Heatmap showing marker gene expression in pseudo-bulk normal fetal TCPs and TCP-like tumor cells in SHH, G3, and G4 MBs.

**h**, Two-dimensional cell-state plots of tumor subpopulations in indicated MB subgroups. Each vertex corresponds to one key cellular state along the transformation. The positions of tumor cell phenotypes (dots) indicate relative scores for the meta-modules, and color scales represent the count.

**Fig. 4. Distinct tumor cell populations in primary and metastatic medulloblastomas**

**a**, DNA methylation-based subgroup prediction scores for G3 and G4 MB samples. Samples with a mixture of G3 MB and G4 MB scores, samples G3 #7 (BT-309) and G4 #1 (BT-325), were further analyzed.

**b** and **c**, UMAP visualizations for the distributions of G3 MB and G4 MB cells from b) BT-309 and c) BT-325 tumors based on the subgroup-specific gene expression scores.

**d**, Heatmap of the marker genes in BT-309 and BT-325 MB samples showing mixed G3 and G4 cell populations.

**e**, UMAP of merged cells from the BT-325 primary and metastatic tumors.

**f**, Split plot of cell populations from BT-325 primary and metastatic samples.

**g**, Cell compositions between matched BT-325 primary and metastatic tumors.

**h**, Two-dimensional cell-state plots and proportions of tumor subpopulations in BT-325 primary and metastatic MB. Each vertex corresponds to one key cellular state along the transformation. The colors represent the count.

**i,** Dot plot displaying the expression of selected marker genes for the BT-325 primary and metastatic tumors. The size of the dot reflects the percentage of the cells that express the gene (Exp.). Average expression levels of the genes are color coded.

**j**, CNV analysis of chromosome 8 based on DNA methylation showing the *MYC* gene amplification in the metastatic tumor of BT-325.

**k**, Dot plot showing the expression levels of G3 and G4 marker genes in the BT-325 primary (x-axis) and metastatic (y-axis) tumors. Red and green indicate higher and lower expression, respectively, in the metastatic tumor compared to the primary tumor.

**Fig. 5. Single-cell omics reveals unique regulatory networks in transitional progenitor-like subpopulations in G3 and G4 MBs**

**a**, Workflow for integrating scRNA-seq and snATAC-seq data analysis using ArchR.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**b**, UMAPs of scATAC-seq from G3 MBs (n = 2 patients) and G4 MBs (n = 3 patients) with reference to scRNA-seq classification of cell clusters.
**c**, **d,** The gene integration matrix from scRNA-seq and snATAC-seq (left panel) and alignment for key TCP- and MB-related marker genes (right panels) in c) G3 MB samples (n= 2 patients) and d) G4 MB samples (n = 3 patients). The individual cell types are indicated by colors. The integrated gene activity of G3 or G4 signature genes is shown in the right panels.
**e**, Heatmap showing correlation of accessibility of promoter regions with target gene expression for tumor cell types.
**f**, Pseudo-bulk ATAC-seq tracks at the indicated gene loci in G3 and G4 MB cells.
**g**, **h**, Heatmaps of scaled accessible peak-to-gene links identified along the integrated pseudotime trajectory from g) TCP-like to G3 MB and h) TCP-like to G4 MB subclusters. Enriched transcription factor motifs are also shown.
**i**, Dot plot showing positive transcriptional regulators identified in the TCP-like cells from G3 and G4 MB compared to other tumor subclusters. x-axis, the correlation of TF motif enrichment to TF gene expression. y-axis, ΔTF deviation scores between TCP-like cells and other subclusters.
**j**, **k**, Genome accessibility track visualization of marker genes with peak co-accessibility in j) TCP-G3-like cells in G3 MB (*OTX2* and *HLX*) and k) TCP-like cells in G4 MB (*BARHL1* and *PAX6*).

**Fig. 6. TCP regulators mediate de novo oncogenic deregulation through enhancer hijacking in G3 MB**

**a**, Heatmaps of Cut&Run peaks detected using indicated antibodies in G3 MB cells and human NSCs showing ±1,000 bp around transcriptional start sites (TSS).
**b**, Signals of H3K4me3, H3K27ac, HNRNPH1, and SOX11 peaks plotted relative to the TSS.
**c**, Signals of H3K27ac peaks at TSSs in G3 MB cells and NSCs.
**d**, Enrichment analysis of pathway terms for genes co-occupied by HNRNPH1 and SOX11 in G3 MB cells.
**e**, Heatmap showing the relative density of indicated genes occupied by the H3K27ac, HNRNPH1, and SOX11 in G3 MB cells (MB-004) and human NSCs.
**f**, Genomic tracks for H3K27ac, SOX11 and HNRNPH1 occupancy at *MYC* (upper) and *OTX2* (lower) loci of human NSCs and G3 MB cells (MB-004).
**g**, Consensus motif analysis using HOMER for HNRNPH1 and SOX11 based on their genomic occupancy. The five most highly represented motifs are shown.
**h**, Hi-C map of long distance intrachromosomal interaction in G3 MB cells (MB-004) and G4 MB cells (UPN3550). H3K27ac, SOX11, and HNRNPH1 occupancy (middle) and pseudo-bulk genome co-accessibility (lower) in the *MYC* locus and long-distance regulatory elements in human NSCs and G3 MB cells are shown.
**i**, Proposed model for the contribution of enhancer hijacking to *MYC* expression.
**j**, *MYC* expression was significantly downregulated by CRISPRi targeting SOX11/HNRNPH1-binding motifs (I, II, II, IV except for V) relative to control sgRNAs (**$p < 0.01$; ***$p < 0.001$; n.s: non-significant; two-tailed t-test; n = 6 for all regions). Data are mean and s.d.
**k**, Relative enrichment of H3K27ac occupancy on the indicated enhancers (Enh1 and Enh2) or promoter (Pro) in MB-004 cells transduced with non-targeting shRNA control (shCtrl), shSOX11, or shHNRNPH1 and in control hNSCs (*$p < 0.01$; **$p < 0.01$; ***$p < 0.001$; two-tailed t-test; N = 3 for all regions). Data are mean and s.d.

**Fig. 7. Targeting TCP regulators inhibits aggressive MB growth**

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**a**, *HNRNPH11* (left) and *SOX11* (right) expression in subgroups of Cavalli's cohort dataset [47]. ***p < 0.001, two-tailed unpaired Student's *t*-test.
**b**, Kaplan-Meier analysis of overall survival of patients with cerebellar MBs (SHH, G3, and G4 MBs) based on the TCP signature gene expression in Cavalli's MB cohorts[47]. Significance was determined by log-rank test.
**c**, Representative images of human G3 and G4 MB tissue stained for HNRNPH1 (left) and SOX11 (right). Scale bars: 100 μm.
**d**, Representative images showed the colocalization of HNRNPH1 and SOX11 immunostaining (arrows) in G3 MB tissues. Scale bars: 2 μm.
**e**, Representative images showed the proximity ligation assay analysis of the HNRNPH1-SOX11 association (arrow) in G3 MB (MB-004) cells treated with lenti-shCtrl or lenti-sh HNRNPH1. Scale bars: 10 μm.
**f**, Representative immunoblots for SOX11, MYC, and HNRNPH1 in MB-004 cells treated with control or shRNAs targeting *HNRNPH1* (shHNRNPH1) or *SOX11* (shSOX11). For gel source data, see supplementary Fig. 1.
**g**, **h**, Images (g) and quantification (h) of tumor spheres formed by MB-004 or D425 cells treated with shCtrl, shSOX11, or shHNRNPH1. Scale bars: 200 μm. **p < 0.01, ***p < 0.001, one-way ANOVA with post hoc Tukey's test.
**i,** Cell viability of indicated cells transduced with shHNRNPH1 (left) or shSOX11 (right) measured by WST-1 assay. n=3 independent experiments; **p < 0.01, ***p < 0.001, one-way ANOVA with post hoc Tukey's test.
**j**, Representative immunostaining images of cleaved caspase 3 (upper) and quantification (lower) in MB-004 cells treated with shCtrl, shSOX11, or shHNRNPH1. **p < 0.01, ***p < 0.001, one-way ANOVA with post hoc Tukey's test. Scale bars: 100 μm.
**k**, Bioluminescent imaging of the NSG mice transplanted with MB-004 cells transduced with shCtrl, shSOX11, or shHNRNPH1 at indicated days.
**l**, Survival curves of NSG mice transplanted with MB-004 cells treated with shCtrl, shSOX11, or shHNRNPH1 (n=10 animals/group). **p< 0.01; log rank test.
**m**, Quantification of Ki67 (left) and cleaved caspase 3 (right) positive cells in xenograft tumors induced by MB-004 cells treated with shCtrl, shSOX11, or shHNRNPH1. **p < 0.01, ***p < 0.001, one-way ANOVA with post hoc Tukey's test.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

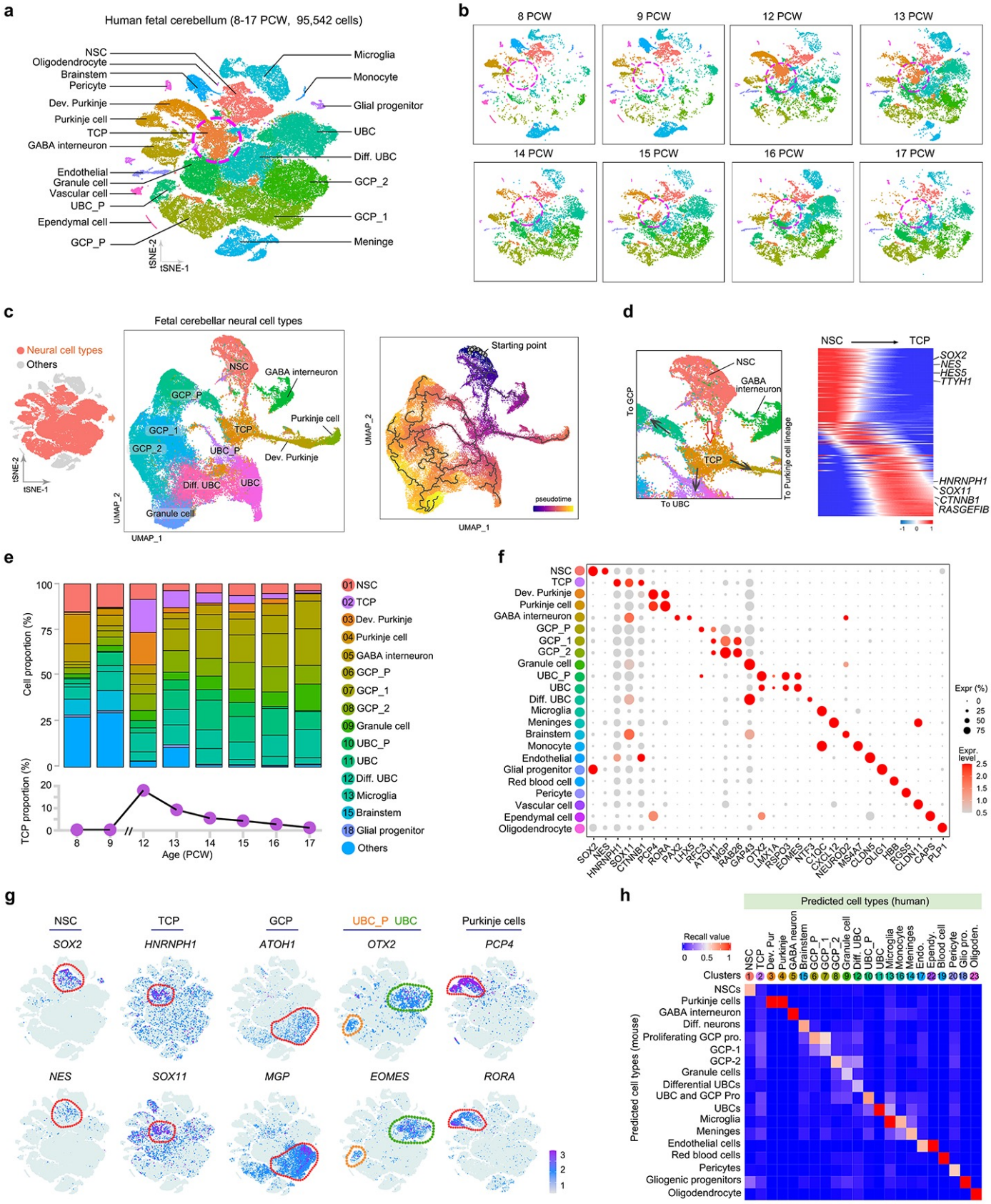


Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

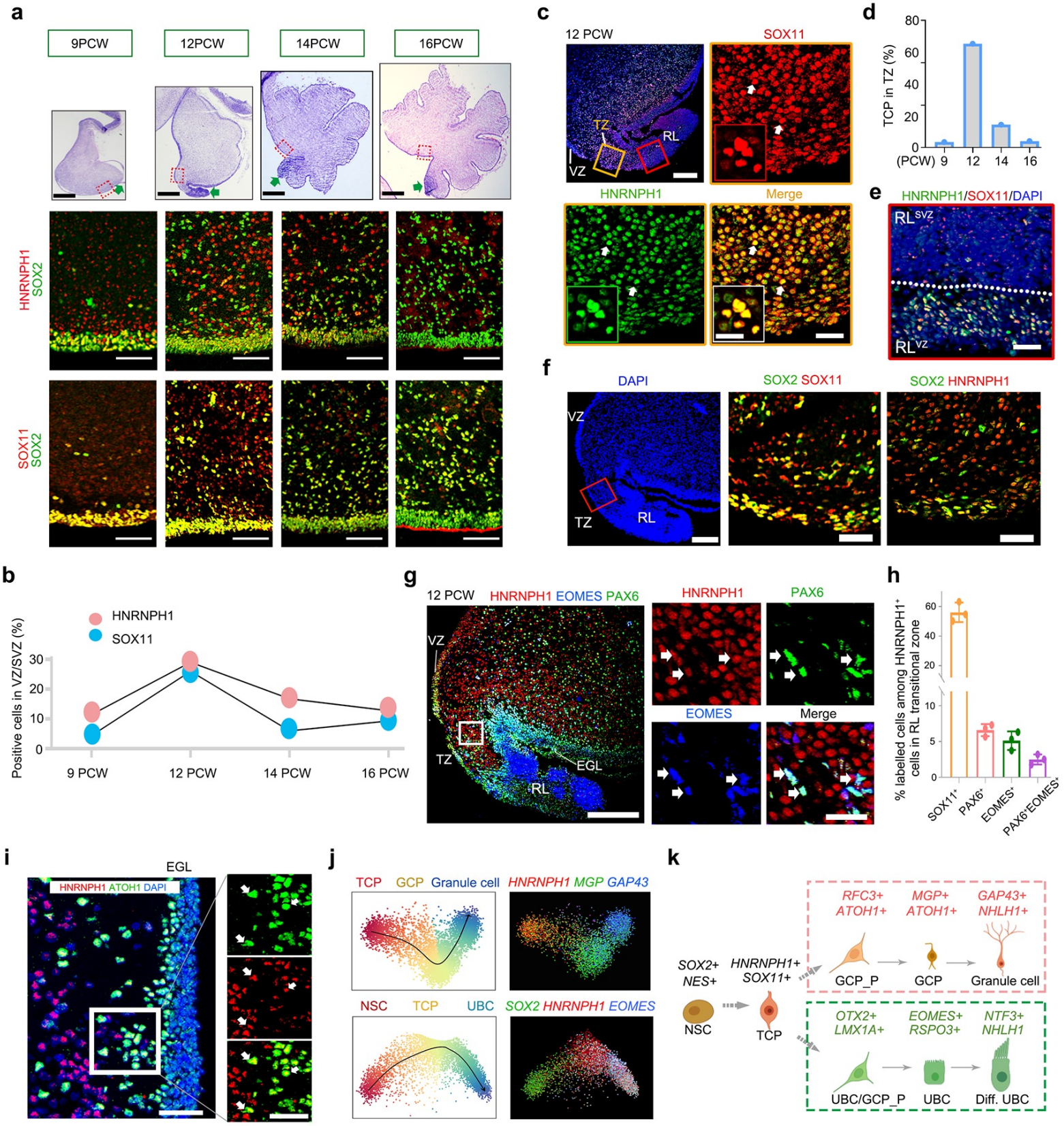


Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

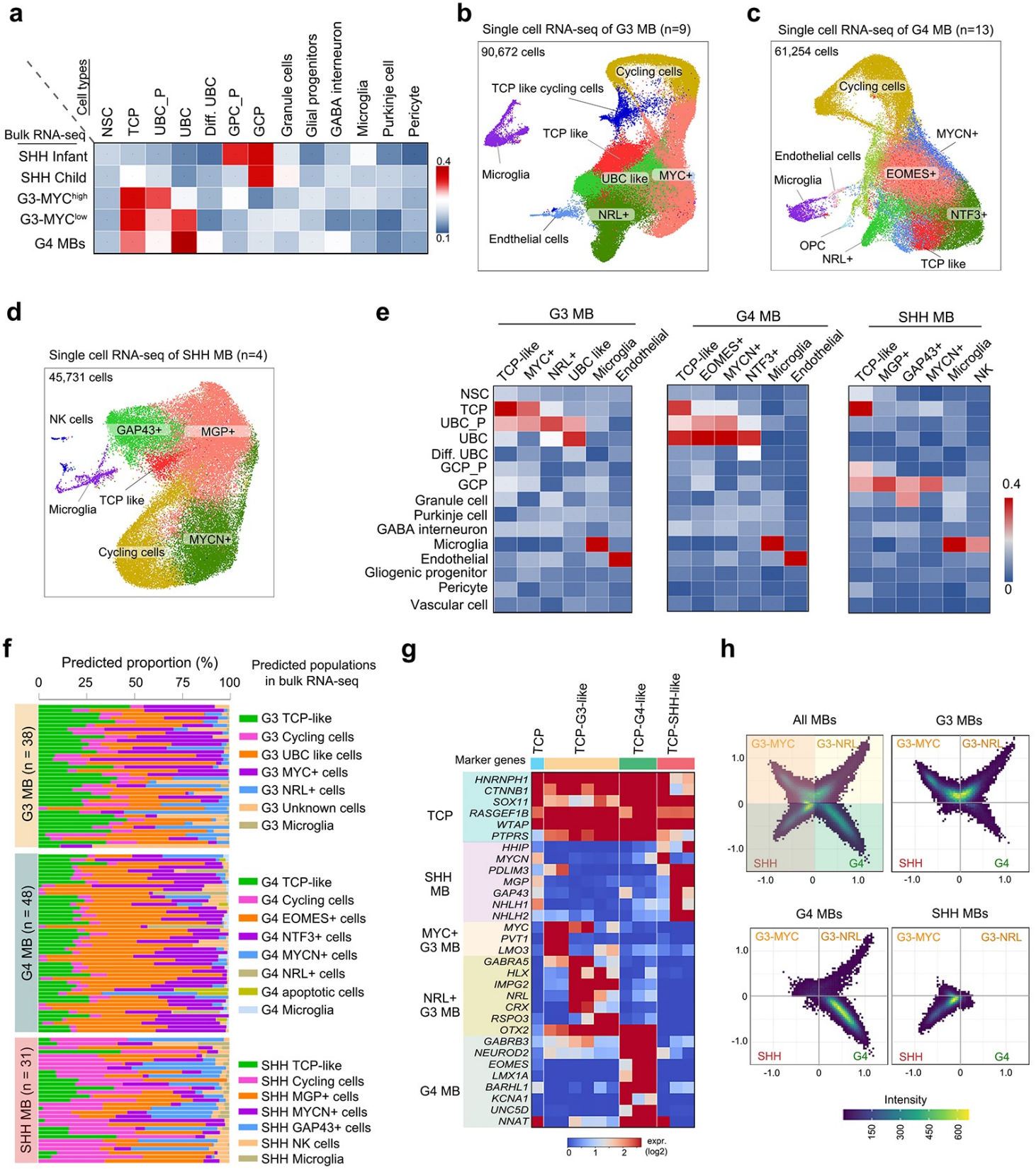


Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

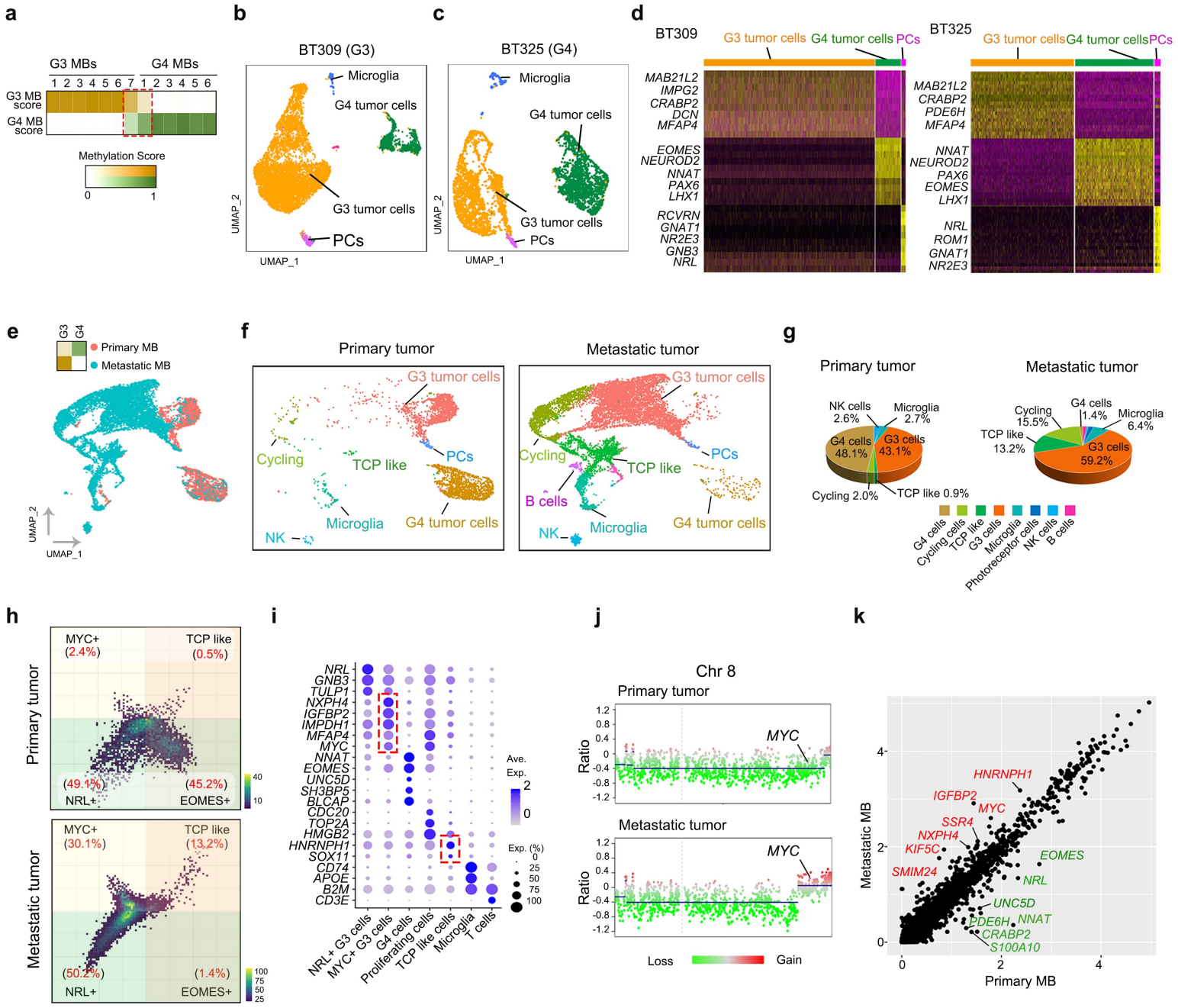


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

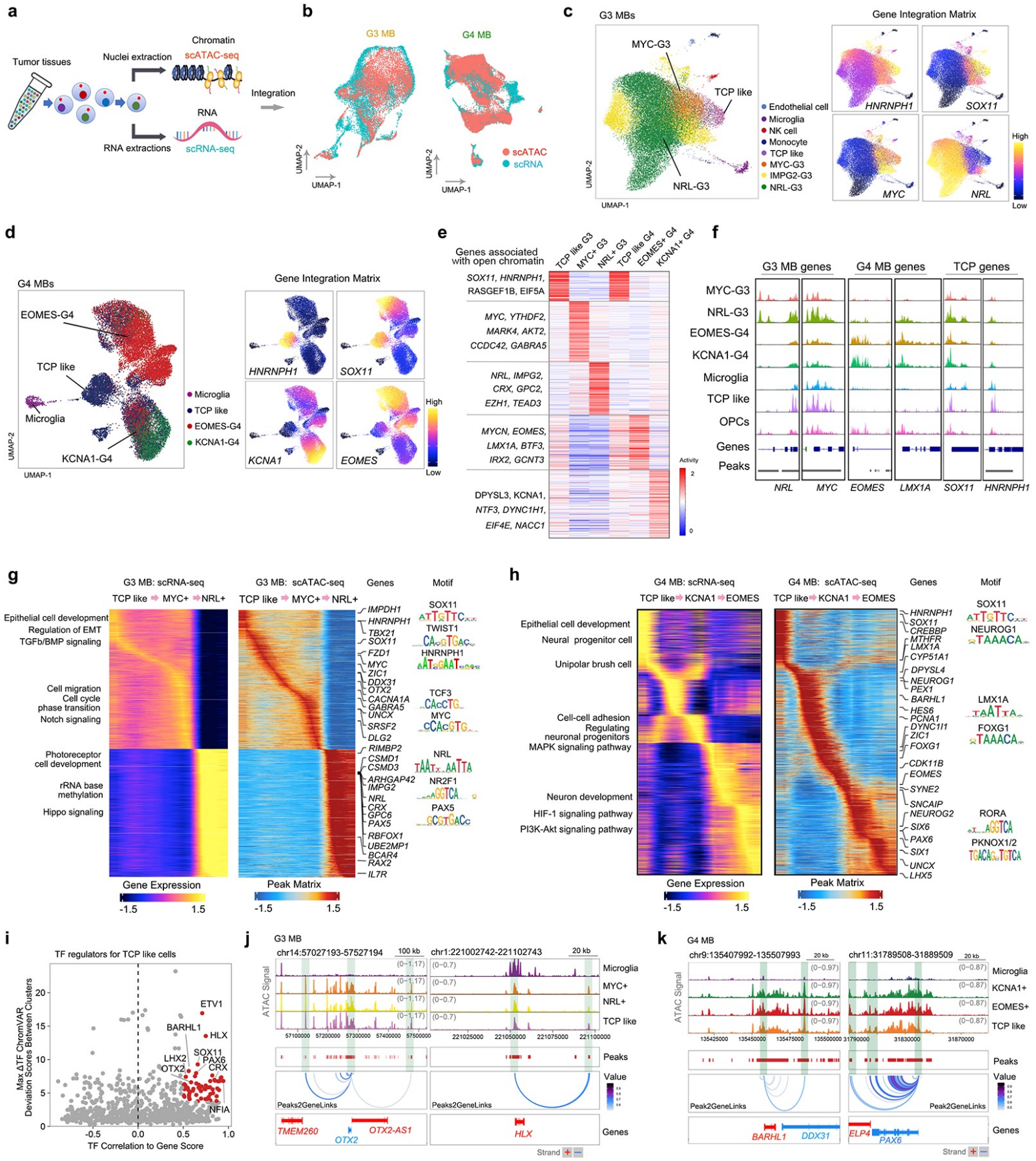


Figure 5

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

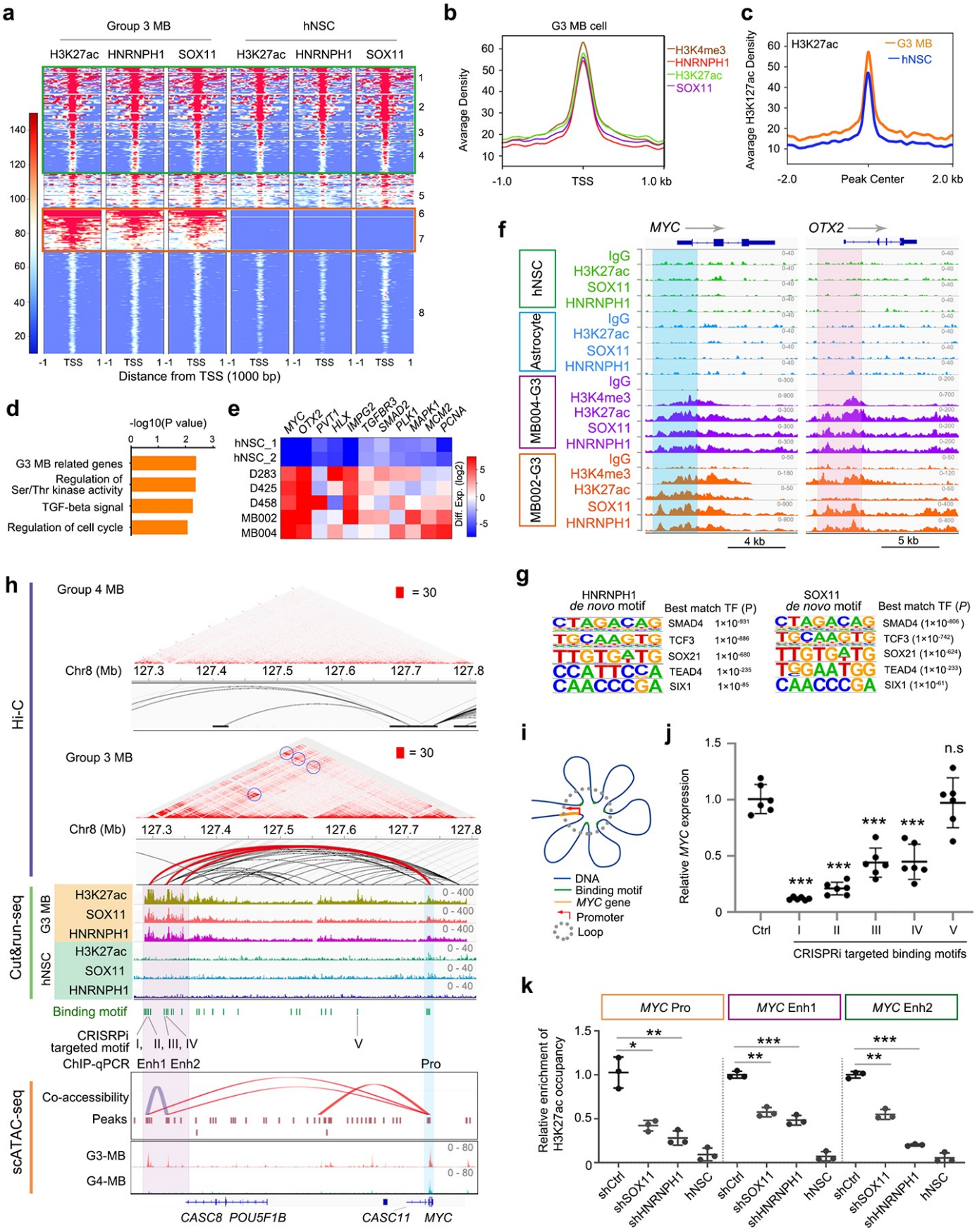


Figure 6

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.17.504304; this version posted August 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

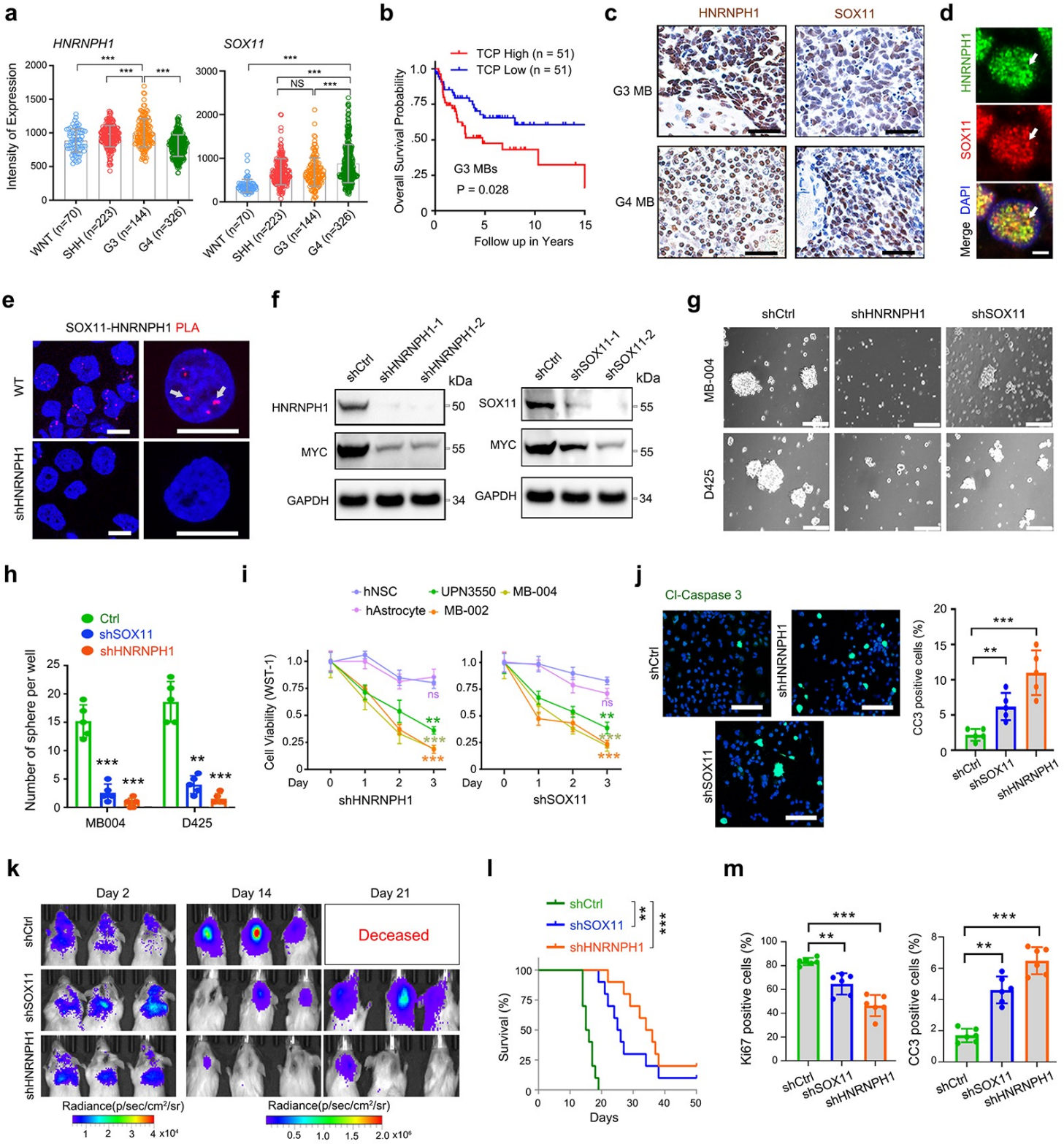


Figure 7