Neuro-Oncology Advances
7(1), vdaf016, 2025 | https://doi.org/10.1093/noajnl/vdaf016 | Advance Access date 23 January 2025

# Inferred developmental origins of brain tumors from single-cell RNA-sequencing data

**Su Wang[†], Rachel Naomi Curry[†], Malcolm F. McDonald, Hyun Yong Koh, Anders W Erickson, Claudia L. Kleinman, Michael D. Taylor, Ganesh Rao, Benjamin Deneen, Arif O. Harmanci, and Akdes Serin Harmanci**

All author affiliations are listed at the end of the article

Corresponding Authors: Arif O. Harmanci, MSc, McWilliams School of Biomedical Informatics, University of Texas Health Science Center, 7000 Fannin St, Houston, Texas, 77030, USA (Arif.O.Harmanci@uth.tmc.edu); Akdes Serin Harmanci, PhD, Department of Neurosurgery, Baylor College of Medicine, 1250 Moursund St, Houston, Texas, 77030, USA (akdes.serinharmanci@bcm.edu).

[†]These authors contributed equally to this work.

**Abstract**

**Background**. The reactivation of neurodevelopmental programs in cancer highlights parallel biological processes that occur in both normal development and brain tumors. Achieving a deeper understanding of how dysregulated developmental factors play a role in the progression of brain tumors is therefore crucial for identifying potential targets for therapeutic interventions. Single-cell RNA-sequencing (scRNA-Seq) provides an opportunity to understand how developmental programs are dysregulated and reinitiated in brain tumors at single-cell resolution. The aim of this study is to identify the developmental origins of brain tumors using scRNA-Seq data.

**Methods**. Here, we introduce COORS (Cell Of ORigin like CellS), a computational tool trained on developmental human brain single-cell datasets that annotates "developmental-like" cell states in brain tumors. COORS leverages cell type-specific multilayer perceptron models and incorporates a developmental cell type tree that reflects hierarchical relationships and models cell type probabilities.

**Results**. Applying COORS to various brain cancer datasets, including medulloblastoma (MB), glioma, and diffuse midline glioma (DMG), we identified developmental-like cells that represent putative cells of origin in these tumors. Our method provides both cell of origin classification and cell age regression, offering insights into the developmental cell types of tumor subgroups. COORS identified outer radial glia developmental cells within $IDH^{WT}$ glioma cells whereas oligodendrocyte precursor cells (OPCs) and neuronal-like cells in $IDH^{Mut}$. Interestingly, $IDH^{Mut}$ subgroup cells that map to OPC show bimodal distributions that are both early and late weeks in development. Furthermore, COORS offers a valuable resource by providing novel markers linked to developmental states within MB, glioma, and DMG tumor subgroups.

**Conclusions**. Our work adds to our cumulative understanding of brain tumor heterogeneity and helps pave the way for tailored treatment strategies.

## Key Points

- COORS (Cell Of ORigin like CellS) applies a multilayer perceptron model for developmental-like cell type annotation.
- COORS predicts developmental-like analogs in various brain tumor scRNA-Seq datasets using a hierarchical approach.
- The identified markers pave new avenues for targeted therapies for brain tumors.

## Importance of the Study

This study offers important contributions to our understanding of brain tumor heterogeneity and demonstrates both conceptual and technical advances. COORS (Cell Of ORigin like CellS) uses multilayer perceptron models for tumor cell of origin classification and cell age regression based on developing brain scRNA-Seq datasets. Our models capture developmental features by learning from a large amount of healthy transcriptional profiles while reflecting similar programs in tumor cells. Distinct from the existing methods, COORS is not limited to the exact matching of healthy cell types, instead focusing on the identification of the origin states of tumor cells. The development and application of COORS represent a significant advancement in the accurate annotation of developmental-like cell states in brain tumor datasets. COORS serves as the first algorithm of its kind and can be potentially extended for use in other cancer types or disease contexts.

One of the greatest challenges to finding a cure for brain cancers is the robust inter- and intra-tumoral heterogeneity that characterizes these tumors.[1–4] This heterogeneity contributes to disease progression and is a key reason therapeutic approaches fail to prevent disease recurrence. Although the genetic evolution of cancer cells is a critical determinant, tumor heterogeneity is also influenced by nongenetic factors including varying developmental cellular programs, which include stem, progenitor, and senescent cell states.[5,6] Prior studies have demonstrated that aberrant expression of neurodevelopmental programs is pervasive in brain tumors and is largely driven by the reactivation of developmental transcriptional states that are acquired by genomic and epigenomic changes.[5] Given the complexity of cell types and an array of developmental states, isolating a single-cell type of origin poses a difficult task; however, a more thorough examination of brain tumor transcriptomics alongside transcriptional signatures of neurodevelopmental cell types may shed light on the origins of brain cancer. To gain a deeper understanding of which developmental cell types brain tumors most closely resemble, we hypothesized that tumor cell lineages can recapitulate cell lineages encountered in the developing brain. While tumors exhibit a multitude of dysregulated pathways, existing evidence, particularly in pediatric tumors, supports this hypothesis.[6–10] We, therefore, focused on employing developmental expression modeling trained on human brain atlases that span various developmental time points. This modeling approach allows us to characterize tumor cells by overlaying their gene expression patterns onto those of early neurodevelopmental stages. By identifying and studying these myriad cellular states from development, our goal is to uncover insights into the origins and behavior of brain tumors, ultimately paving the way for more effective treatment strategies and improved patient outcomes.

Single-cell RNA-sequencing (scRNA-Seq) provides an opportunity to dissect the complex cellular states during development and in health and disease.[11] However, it is computationally challenging to decipher the spectrum of heterogeneous developmental cell states in tumor cells using scRNA-Seq. Accurate identification of developmental-like cell states necessitates a comprehensive understanding of the interactions among all genes, which, in turn, requires a substantial amount of gene expression data. In this study, we developed COORS (Cell Of ORigin like CellS), a computational tool to annotate each developmental cell state in tumor cells at single-cell resolution. COORS uses a multilayer perceptron model (MLP) for cell of origin classification and cell age regression using developing brain scRNA-Seq datasets from previously published scRNA-seq datasets, comprising approximately 1 million cells from developing human and mouse brains.[6,12–16] Using COORS, we predicted developmental analogs in pediatric and adult tumors using public and in-house medulloblastoma (MB),[17] diffuse midline glioma (DMG),[14] and glioma scRNA-Seq data,[18,19] which revealed unique developmental cell types as putative cells of origin in each brain cancer subtype.

The primary motivation for the COORS algorithm is to leverage developmental cell types identified in normal brain development to annotate potential cell origins in tumor samples. This approach distinguishes COORS from other methods in 2 key ways: (1) Existing methods focus on the exact matching of healthy cell types, whereas COORS models developmental trajectories, and (2) most tumor cell type annotation tools focus on identifying tumor cells without attempting to annotate their developmental origins.

## Materials and Methods

### Data Preprocessing

We took advantage of multiple existing scRNA-Seq datasets as the reference for the COORS models, which contain in total around 1 million cells from developing human and mouse brains[6,12–16] (Supplementary Table 1). For each separate model trained on the reference dataset, we ranked the cell types based on the number of cells and set the lower 25% quantile as the baseline to which we subsampled those higher groups, collecting sufficient and unbiased reference samples. We excluded those cell types with cells fewer than 20, as the cell size was too small to serve as a reasonable training reference.

We conducted the standard pipeline of scRNA-Seq data preprocessing for both reference and testing datasets using Scanpy 1.7.2 in Python 3.6.13. For cell of origin classification, the preprocessing of reference and testing data was performed starting from the whole dataset. On the other hand, for cell age regression, we first grouped

the reference and testing datasets by cells of origin and then preprocessed each group separately. Each cell was normalized to have the same total read count and the matrices were transformed into a natural logarithm domain. We then annotated the top 2000 highly variable genes in the reference dataset.

We combined the reference marker genes and highly variable genes to identify the common gene set by intersecting them with the testing genes, and then subsetted both datasets to include only these common genes. We scaled both datasets to unit variance and zero mean and truncated them to 10. We kept all the other parameters in default values.

We randomly split each reference cell type into 2 subsets, one with 80% cells for model training and the other with 20% cells for model validation. The training subsets of all the cell types were concatenated as training data, and the validation subsets as validation data. We conducted one-hot encoding of cell types in both training and validation data. We scaled training, validation, and testing data into the range from 0 to 1 using Scikit-learn 0.24.2.

## MLP-Based Prediction Model

We developed multilayer perceptron networks for cell of origin classification and cell age regression.

*Cell of Origin Classifier*The cell of origin classifier has one input layer, variable numbers of hidden layers, and one output layer. The input layer has the same number of nodes as the input genes. The number of hidden layers varies from 1 to 4, and the number of nodes in one hidden layer is set to 256, 128, 64, or 32, which is determined after hyperparameter optimization. Following the dense connection within each hidden layer, there are batch normalization, activation, and dropout functions. We use the popular Rectified Linear Unit (ReLU) for hidden layer activation and set the dropout rate to be 0.1 or 0.2. The output layer uses the Softmax activation function so that each node outputs a nonnegative value smaller than 1 and all the values sum up to 1. Therefore, each output corresponds to the probability of one cell type. We compile the model using categorical cross-entropy as loss function, Adam as an optimizer, and accuracy as metrics.

*Cell Age Regressor*Similar to cell of origin classifier, cell age regressor consists of one input layer, a group of hidden layers, and one output layer. While the input layer and hidden layers are structurally the same as cell of origin classifier, the output layer of cell age regressor has only one node with a Sigmoid activation function that corresponds to the predicted cell age. The model is compiled using mean squared error as a loss function, Adam as an optimizer, and loss as metrics. Since more than one cell age regressors exist corresponding to each cell of the origin classifier, these regressors can have specific hyperparameters of hidden layers that are not necessarily the same.

*Model Training Prerequisites*We implemented cell of origin classifier and cell age regressor using Keras 2.6.0 with Tensorflow 2.4.0 as the backend in Python 3.6.13. Prerequisite packages for data preprocessing and model training include Numpy 1.19.5, Pandas 1.1.5, Scanpy 1.7.2, Anndata 0.7.8, Scipy 1.5.4, and Scikit-learn 0.24.2.

*Hyperparameter Optimization*We systematically optimized hyperparameters of cell of origin classifier and cell age regressor using grid search cross-validation implemented by Scikit-learn 0.24.2, focusing on tuning the number of hidden layers and nodes, the dropout rate of hidden layers, and learning rate of the optimizer. For each model, we varied the number of hidden layers from 1 to 4 and the number of nodes in each layer could be 256, 128, 64, or 32. We followed the convention of using the ReLU as an activation function in hidden layers. Training epochs were fixed to be 100 and batch size 32 as they did not show significant affections in our case. Along with the iteration of every possible hidden layer structure, we explored dropout rates of 0.1 or 0.2 and learning rates of 0.1, 0.01, 0.001, or 3 decaying learning rates that were scheduled to exponentially reduce during model training based on an initial rate of 0.1 or 0.01, final rate of 0.01 or 0.001, training epochs, and batch size.

*SHAP Evaluation*We took advantage of the game-theory-based approach SHAP (SHapley Additive exPlanations) to interpret the output of our COORS models with the Shapley values.[20] Specifically, we conducted the SHAP evaluation among cells classified to each cell of origin to investigate which genes with either high or low expression levels have the most significant impacts on the classification. Within each mapped cell of origin, we selected the top 50 tumor cells with the highest classification probability scores as typical samples and concatenated them for SHAP analysis. We summarized the training data with a set of weighted k-means ($k = 50$), which were used as background along with the COORS cell classifier to initiate the kernel explainer. We calculated SHAP values and generated the summary plot for each cell of origin, where the relative impacts of the top 20 genes over the cells are displayed, sorted by the sum of the absolute SHAP values across all cells. Besides the importance of genes on average, their impacts on a particular sample are also easy to access.

## Immunostaining

Immunostaining was performed on 10 μm paraffin-embedded human glioma sections, which were cut, deparaffinized, and treated with heat-induced epitope retrieval (HIER) using an antigen retrieval buffer (Sigma Aldrich Citrate Buffer, pH 6.0, 10X Antigen Retriever: C9999) when needed. Sections were blocked for 1 h at room temperature, followed by overnight incubation at 4 °C with primary antibodies: Rabbit anti-GABRG2 (1:200, Proteintech 14104-1-AP), Rabbit anti-NFIB (1:250, Sigma HPA003956), and Mouse anti-IDH1 (1:50, Dianova DIA-H09-L). Species-specific secondary antibodies tagged with Alexa Fluor dyes (488 nm, 568 nm, or 647 nm; 1:1,000, ThermoFisher) were used, followed by Hoechst nuclear counterstaining

(1:50,000, ThermoFisher H3570). Sections were mounted with Dako Fluorescence Mounting Media (Ref S3023).

## Results

### The Overall Workflow of COORS Algorithm

For training COORS models, we train a multilayer perceptron model for cell of origin classification and cell age regression using developing brain scRNA-Seq datasets[6,12–16] (Supplementary Table 1). For simplified illustration, assuming we have reference data with three origin like cell types A, B, and C, we train a neural network-based cell of origin classifier using this reference data, saving the model in our repository (Figure 1A and B). Concurrently, we train neural network-based cell age regressors, for each cell origin A, B, and C, also saving these trained models in the repository. In the assignment step, when processing tumor data, we select appropriate pre-trained models based on cancer region (eg, neocortex or cerebellum), apply cell type and age classifiers, and conduct SHAP analysis to identify critical tumor-specific developmental markers (Figure 1A and B). In addition to this process, COORS incorporates a developmental tree reflecting hierarchical brain cell type relationships, with mesenchymal lineage cell types not being the focus. Nodes in this tree represent cell types, and the node probabilities are modeled using beta distributions. We evaluate node significance using a combination of *t*-tests, beta tests, and Fisher's combined probability test, marking nodes with significant developmental or cell type assignments, further aiding in identifying tumor origin. For each developmental-like cell type, we further predict cell age using the corresponding pre-trained cell age regressors. Finally, we conduct SHAP analysis[20] to extract essential features from our machine learning neural network models, identifying tumor-specific developmental-like gene markers for each cell type and age within our training dataset (Figure 1A and B).

### Validation of COORS Algorithm on scRNA-Seq Medulloblastoma Data

Medulloblastoma is a pediatric brain tumor that is closely associated with early hindbrain development and can be classified into 4 molecular subgroups.[6,7,10,21–24] The WNT-activated subgroup is defined by mutations in the WNT signaling pathway and generally displays a favorable prognosis. In contrast, the Sonic Hedgehog (SHH) activated subgroup results from mutations in the SHH pathway and may have varying clinical outcomes. Group 3 (GP3) MBs have a distinct gene expression profile and are typically associated with a poorer prognosis. Group 4 (GP4) tumors, marked by a specific gene expression pattern, tend to have intermediate clinical outcomes. Understanding the origin of these 4 tumor subtypes might lead to the development of improved treatment strategies.

We applied the COORS algorithm to previously published MB scRNA-Seq data,[17] where subgroup annotations are available for each sample, containing 29 samples and approximately 40K cells in total. Using the COORS algorithm, we have used the pre-trained cell type and cell age models, derived from scRNA-Seq data of developing cerebellum, to map tumor MB cells[13] (Figure 2A). We have not focused on WNT pathway activated subgroup tumor cells because the WNT subgroup is known to originate from the lower rhombic lip (LRL) adjacent to the brainstem, rather than from the upper rhombic lip (URL) in the cerebellum.[17,25]. Figure 2B and C present the developmental tree inferences generated by COORS, which accurately represent the hierarchical relationships among cell types in the cerebellum. The tree reflects differentiation trajectories from earlier progenitor cells such as rhombic lip (RL)-derived cells and granule cell precursors (GCPs) to more mature granule neurons (GNs) and to unipolar brush cell (UBC-CN), showing how tumor cells align with these stages of normal development. COORS maps GP4 subgroup tumor cells to UBC-CN, while SHH subgroup tumor cells map predominantly to GNs, GCPs, and RL cells, consistent with previous studies on medulloblastoma cell origins (Figure 2B–F). However, COORS did not significantly map GP3 subgroup tumor cells to any neuronal cell type lineages, suggesting a less well-defined origin for this subgroup (Supplementary Figure 1).

We conducted SHAP analysis[20] to extract critical features from our machine learning neural network models, which returned new and known marker genes associated with URL-derived cell types and the MB subgroups to which they correspond (Supplementary Table 2). For example, SHH-GCP mapping SHAP markers include those associated with the external granule layer formation and GCP identity such as *CBFA2T2*,[26] *NDST3*,[27] *UNC13C*,[28] and *UNC5C*,[29] as well as markers linked to GCP proliferation during tumorigenesis, such as *BOC*.[30] Sonic Hedgehog-granule neurons mapping SHAP markers include those associated with GN identity and its developmental maturation such as *CNTN1*,[31] *GRIK2*,[26] *PAM*,[32] and *RBFOX*.[33,34] The impact of *GRID2*[35] on GNs has been shown in previous studies. Paralog *ROBO1 is crucial for* GN migration with *MSI2* likely marking GCPs in the external granule layer but not postmitotic granule cells in the internal granule layer. In addition, other SHAP features associated with UBC include *CACNA2D1* and *GRIA2*,[36,37] which contribute to neuronal signaling and excitatory functions critical for cerebellar development. Genes such as *LMX1A* and *RELN* are also involved in the differentiation and positioning of neuronal cells. *ERBB4*, a known oncogene, has been identified as a GP4-associated SHAP marker, showing significant activity and replication in various studies.[38] Furthermore, *KHDRBS2*[39] has been implicated in post-transcriptional regulation within this subgroup. *JMJD1C/KDM3C*, a histone demethylase, is frequently altered in various cancers and might be contributing to epigenetic regulation and chromatin remodeling in tumorigenesis within the GP4 tumor subgroup.[40] These findings emphasize the importance of these markers in medulloblastoma pathogenesis (Supplementary Figure 2, Supplementary Table 2).

Next, we predicted the cell age of each identified developmental cell type within MB tumor cells using our pre-trained cell age regressor models (Figure 2G). Interestingly, GP4 subgroup cells mapping to UBC mostly correspond to the later weeks in development by

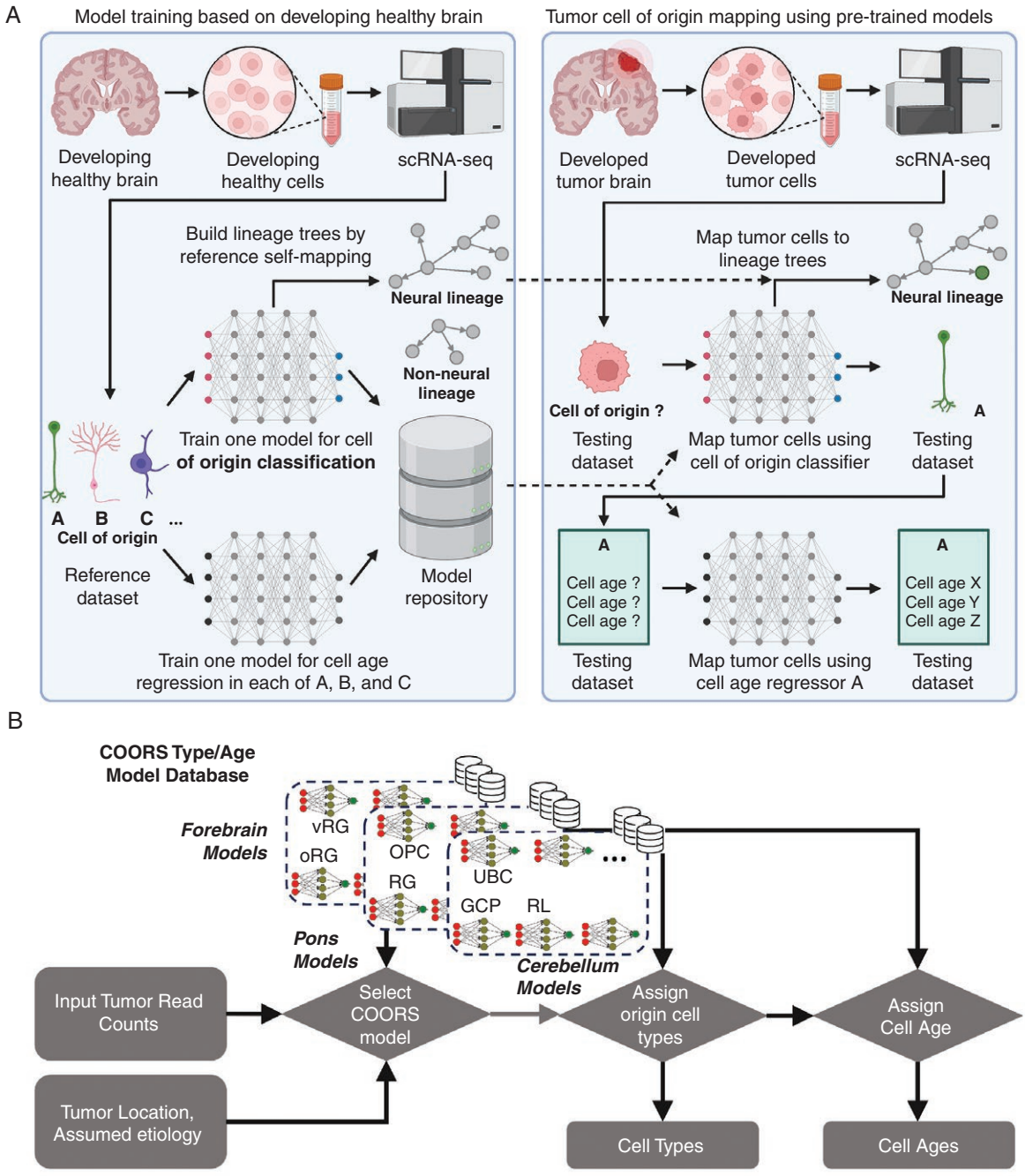


**Figure 1.** Overview of COORS algorithm. (A) In the first step, neural network models are trained for cell of origin classification and cell age regression using developing brain scRNA-Seq datasets, and the models are saved in the repository. In the second step, these pre-trained models are used to map scRNA-Seq tumor cells to developing brain cells, predicting cell origin and age while conducting SHAP analysis to identify tumor-specific gene markers. COORS also incorporates a developmental tree reflecting hierarchical brain cell type relationships. (B) Tumors are matched with specific origin like cell type assignment models based on their region of origin within the brain (eg, neocortex vs. cerebellum vs. pons), enhancing the precision of the COORS application. Post model selection, the mapping of tumor cells to developing healthy brain cells is performed through the application of these pre-trained models, as depicted in the schematic. Abbreviations: COORS, Cell Of ORigin like CellS; scRNA-Seq, single-cell RNA-sequencing; SHAP, SHapley Additive exPlanations.

17 post-conceptional weeks (PCW), while GP3 and GP4 subgroup cells mapping to RL mostly correspond to the earlier weeks by 11 PCW in development (Figure 2G, Supplementary Figure 3). The differentially expressed genes (DEGs) between tumor cells and their respective cells of origin, as well as the DEGs between the cell of origin and the following developmental stage cell type are listed in Supplementary Table 2.

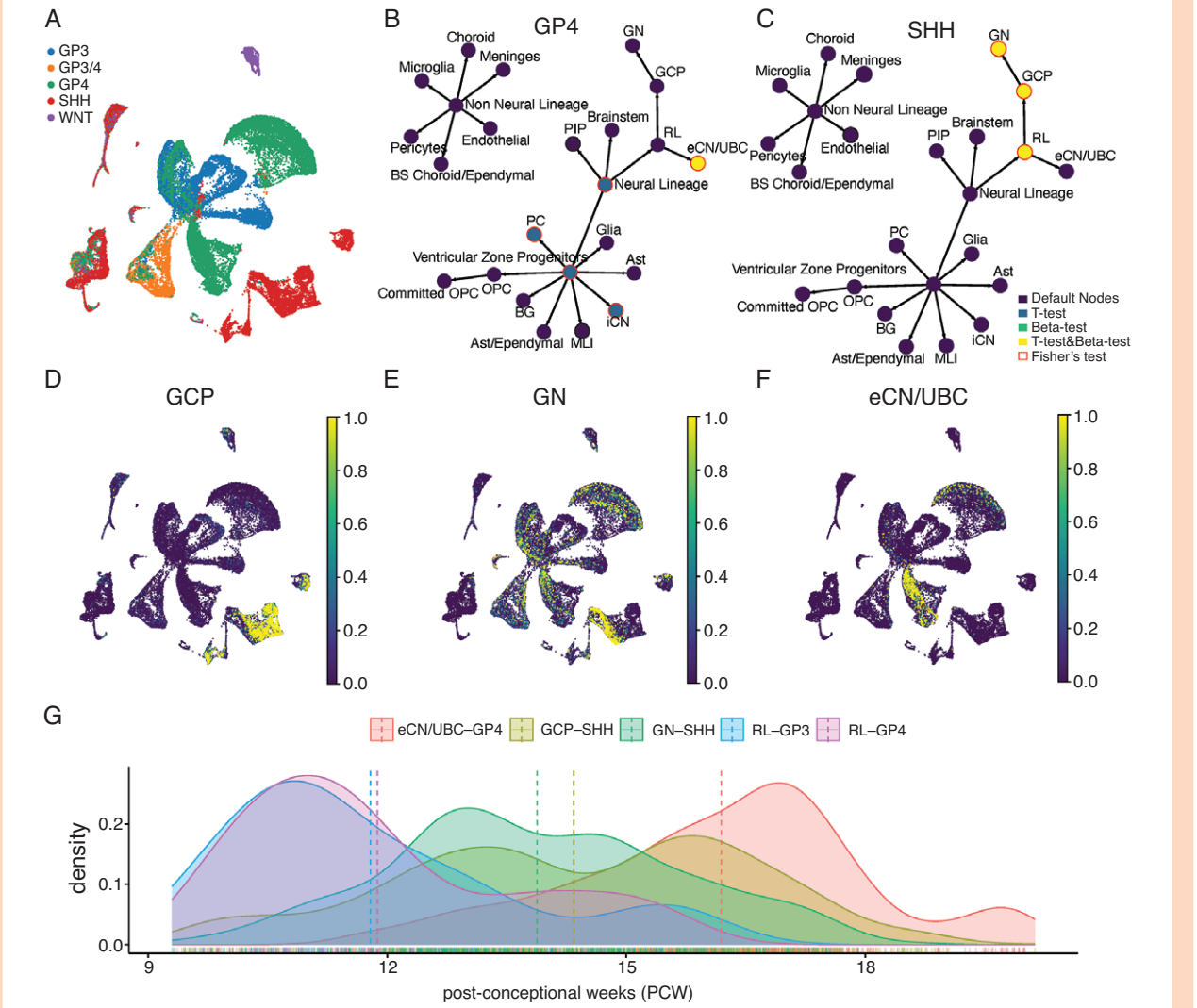


**Figure 2.** Characterization of developmental-like cell states in MB scRNA-Seq data. (A) Tumor subgroups are shown for the MB scRNA-seq dataset. (B and C) Developmental tree inference by COORS showing the hierarchical relationship between tumor cells and normal developmental cell types of GP4 and SHH tumor subgroups respectively. Nodes are color-coded based on their statistical significance: default nodes are insignificant (*P*-value > .05). Definitions of cell types are as follows: PC, Purkinje cells; RL, rhombic lip; GCP, granule cell progenitors; GN, granule neurons; eCN/UBC, excitatory cerebellar nuclei neurons/unipolar brush cells; iCN, inhibitory cerebellar nuclei neurons; PIP, PAX2+ interneuron progenitors; BG, Bergmann glia; Ast, astrocytes; Glia, glia; OPC, oligodendrocyte precursor cells; Committed OPC, committed oligodendrocyte precursor cells; MLI, molecular layer interneurons; Ast/Ependymal, astrocytes/ependymal cells; Choroid, choroid plexus; BS Choroid/Ependymal, brainstem choroid plexus/ependymal cells. (D–F) Developmental cell type probability scores are shown for GCP, GN, and UBC cell types. (G) Distribution of age mapping within each tumor subgroup and their respective mapped cell of origin pairs. Abbreviations: COORS, Cell Of ORigin like CellS; scRNA-Seq, single-cell RNA-sequencing.

## Application of COORS Algorithm on Pediatric DMG

Next, we applied the COORS algorithm on previously published H3.1/H3.2 and H3.3 histone 3 K27M-mutant DMG scRNA-Seq data containing 13 samples and approximately 47K cells in total.[6,14] Using the COORS algorithm, similar to our previous application in MB data, we have used the pre-trained cell type and cell age models, derived from scRNA-Seq data of developing mouse pons brain,[6] to map pediatric glioma cell origins (Figure 3A, Supplementary Figure 4). Additionally, Figure 3B and C shows inferred developmental origin trees for both H3.1/2K27M and H3.3K27M tumors. These figures highlight the significant β and *t*-test nodes associated with different developmental lineages, showing how the H3.1/2 tumors map predominantly to the ependymal lineage, while the H3.3 tumors map to neuronal intermediate progenitors (IPCs) (Figure 3A–C; Supplementary Figure 5).

In the case of H3.3K27M DMG, the association with intermediate progenitor cells reveals a more complex and dynamic cellular differentiation pattern by identified SHAP

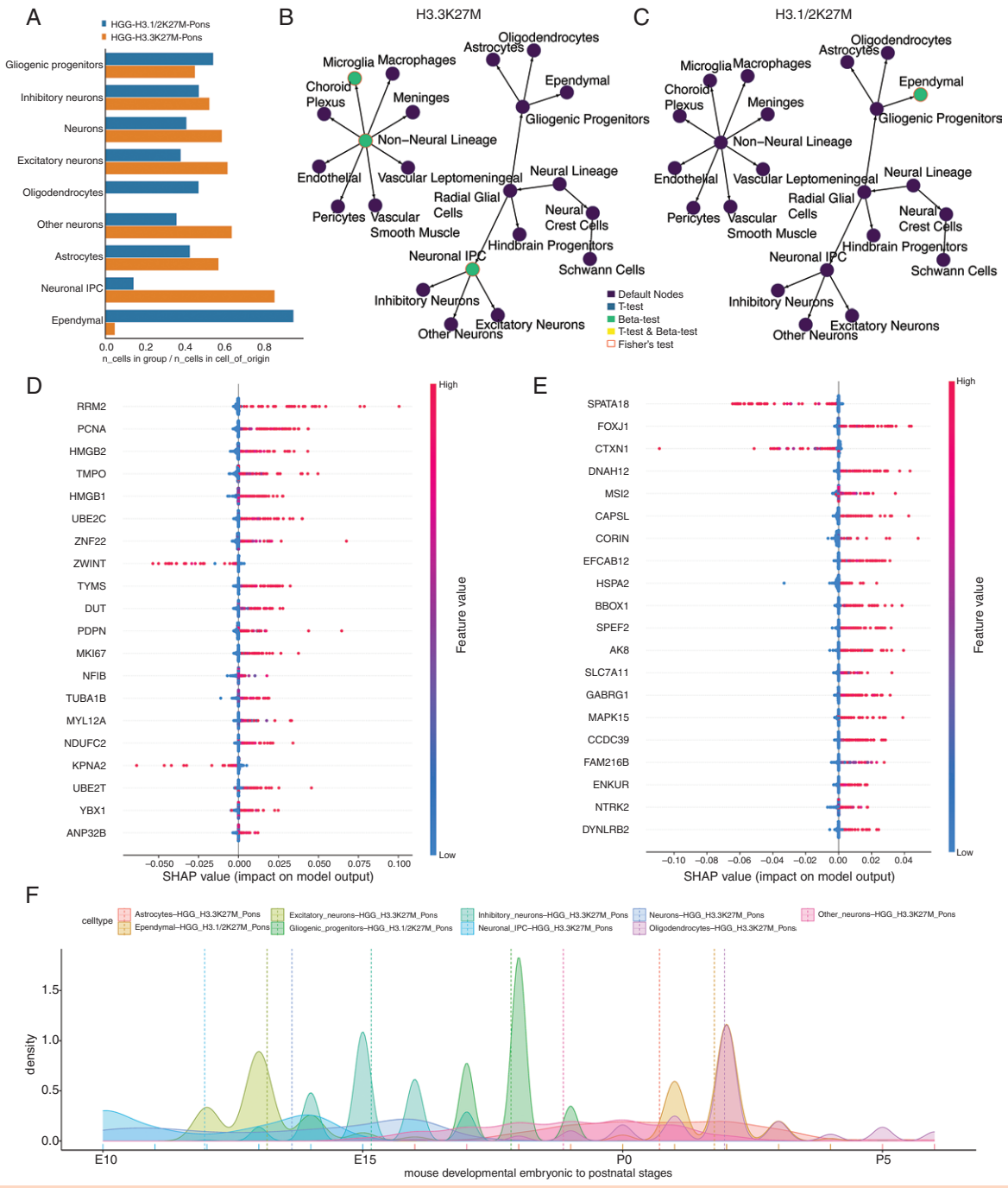


**Figure 3.** Characterization of developmental-like cell states in DMG scRNA-Seq data. (A) Barplots showing the distribution of DMG cells within each tumor subgroup mapped to individual developing cell types. (B and C) Developmental tree inference by COORS showing the hierarchical relationship between tumor cells and normal developmental cell types of H3.1/2K27M and H3.3K27M tumor subgroups respectively. Nodes are color-coded based on their statistical significance: default nodes are insignificant (*P*-value > .05). (D and E) The figure displays the results of SHAP analysis, showing the top impactful genes from each cell type Neuronal IPC and Ependymal respectively, in our training dataset. (F) Distribution of age mapping within each tumor subgroup and their respective mapped cell of origin pairs. Abbreviations: COORS, Cell Of ORigin like CellS; DMG, diffuse midline glioma; IPC, intermediate progenitors; scRNA-Seq, single-cell RNA-sequencing; SHAP, SHapley Additive exPlanations.

markers. As its rapid growth rate and high tumor invasiveness, many SHAP features were noticeable in DNA synthesis and proliferative markers, such as *RRM2* and *PCNA*, and subsequent transcriptional activities such as *HMGB1* and *HMGB2*. *NFIB*, a recognized transcription factor for neuronal progenitors,[41] contributed to the assignment of H3.3 tumor cells to neuronal IPCs. Our COORS model, which shows a negative SHAP value for KPNA2, effectively

captures genes with negative relationships, reflecting unbiased balanced multicollinearity. For brain development, it has been demonstrated previously that precise downregulation of *KPNA2* is essential during neural differentiation of embryonic stem cells.[42] In neural progenitor cells (NPCs), *YBX1* plays a critical role in controlling self-renewal and neuronal differentiation.[43] The evidence of neurodevelopmental cortical malformation caused by the *TUBA1A* variant underscores the significant impact of these abnormalities on human brain development, particularly affecting various stages of cell development, such as NPCs as well as IPCs, and others. This underscores the importance of accurately modeling both gene upregulation and downregulation at various developmental stages (Figure 3D and E; Supplementary Table 3).

Moreover, as demonstrated in previous studies,[14] SHAP in COORS highlighted *DNAH12* (a ciliary gene), which is specifically overexpressed in the H3.1/2K27M group, along with *FOXJ1* a well-known ependymal transcription factor.[44] *DNAH12* is a target gene of *FOXJ1* and plays a role in the progression of cancer by facilitating the expression of genes encoding dyneins and proteins associated with motile cilia, a critical process for these cells.[44] Additionally, the SHAP model identified *DYNLRB2*, a marker for ependymal cells,[45] which was not emphasized in previous research, demonstrating its significant contribution. While earlier studies have investigated *FOXJ1* target genes based on known interactions, our SHAP model revealed that not all downstream of *FOXJ1* contribute equally, showing that *DNAH12* and *DYNLRB2* have a higher impact specifically in the H3.1/2K27M setting through an unbiased approach. *CCDC39* is selectively expressed in embryonic choroid plexus and ependymal cells, and there is functional genomic evidence that mutations in this gene can lead to hydrocephalus.[46] Therefore, this provides indirect evidence that the H3.1/2K27M DMG group is associated with ependymal lineage and differentiation (Supplementary Table 3).

Next, we predicted the cell age of each identified developmental cell type within DMG tumor cells using our pre-trained cell age regressor models (Figure 3F). H3.3 tumor cells mapping to neuronal IPCs mostly correspond to the earlier weeks in development, whereas H3.1/2 tumor cells mapping to ependymal-like cells mostly correspond to the later weeks in development (Figure 3F). The DEGs between tumor cells and their respective cells of origin, as well as the DEGs between the cell of origin and the subsequent developmental cell type, are listed in Supplementary Table 3.

## Application of COORS Algorithm on In-House scRNA-Seq Glioma Data

Next, we applied the COORS algorithm to our in-house glioma scRNA-Seq data, containing 21 samples and approximately 234K cells in total.[18,19] Using COORS model we mapped $IDH^{Mut}$ oligodendroglioma, $IDH^{Mut}$ astrocytoma, and $IDH^{WT}$ tumor cells to pre-trained models derived from developing human brain dataset by Poliodakis et al.,[12] Bhaduri et al.,[16] and Jessa et al.[6] developing mouse forebrain datasets. $IDH^{Mut}$ oligodendroglioma and astrocytoma tumor cells map to oligodendrocytes and also to interneuron cell subtypes by Poliodakis et al.,[12] Bhaduri et al.,[16] and Jessa et al.[6] datasets (Figure 4A–C; Supplementary Figures 6–8). Previous studies also suggest that GABAergic neurons and oligodendrocyte precursor cells (OPCs) are derived from common neurodevelopmental origins; predominantly, they both originate from Nkx2.1-expressing precursors located in the medial, lateral, and caudal ganglionic eminences (CGE).[47,48] Moreover, GABAergic neurons and OPCs converge at a shared transcriptional state with expression of *OLIG2*,[49] *GABARs*,[50] and *PDGFRA*.[51] In our recent study, we demonstrated that a subset of $IDH^{Mut}$ glioma cells fire single, short action potentials (APs) and are defined by mixed characteristics of GABAergic neurons and OPC.[19] Interestingly, $IDH^{Mut}$ oligodendrogliomas map to interneuron cell origins from both the CGE and the medial ganglionic eminence (MGE), whereas $IDH^{Mut}$ astrocytomas map solely to the MGE. On the other hand, $IDH^{WT}$ tumor cells map to the outer radial glia (oRG) in Poliodakis et al.[12] dataset and dividing cells from the Bhaduri et al.[16] dataset. Previously researchers identified a subpopulation of cancer stem cells in GBMs that show a resemblance to oRG.[52]

In addition, we conducted SHAP analysis to extract critical features from our machine learning neural network models, enabling the identification of developmental-like gene markers specific to glioma for each mapped cell type (Figure 4D–I; Supplementary Figure 9; Supplementary Table 4). SHapley Additive exPlanations analysis identified that markers commonly associated with OPCs, oligodendrocytes, and GABAergic neurons such as *OLIG1*,[53] *PDGFRA*,[51] *NKX2-2*,[54] *OLIG2*,[55,56] and *GAD1*[19] predominantly contributed to the mapping of IDH mutant tumor cells to OPCs or GABAergic neurons. *GAD1* expression is confirmed through immunostaining in human $IDH^{mut}$ glioma in our previous study.[19] To estimate the cell age of each identified developmental cell type within glioma tumor cells, we applied our pre-trained cell age regressor models. Age mapping was performed on the dataset from Bhaduri et al.[16] due to its wide range of developmental age data and analysis revealed that $IDH^{Mut}$ cells corresponding to OPCs exhibit a bimodal age distribution, indicating stages early and late in development (Figure 4J). The DEGs between tumor cells and their respective cells of origin, as well as the DEGs between the cell of origin and the following developmental cell type, are listed in Supplementary Table 5.

In another dataset by Zeng et al.,[15] finer detailed subtypes of developing human brain lineages were characterized, including the GABAergic lineages (Figure 5A–C). We utilized this dataset to further explore and understand the subtypes of neuronal subtypes in gliomas. *$IDH^{Mut}$* tumor cells predominantly mapped to specific populations of GABAergic cells in the developing brain, GABA-cluster 10. Interestingly, $IDH^{WT}$ tumor cells mapped to the highly proliferative GABA-cluster 9, characterized by high expression of *CENPF*, potentially indicating NPC-like and OPC-like properties in these cells. Additionally, $IDH^{WT}$ tumor cells specifically mapped to NSC-cluster 12 which exhibits high expression of proliferative and neural stem cell marker genes (Supplementary Table 4). We confirmed the SHAP markers *GABRG2* and *NFIB*, identified from the Zeng et al. and Poliudokis et al. datasets, through immunostaining of a 1p/19q co-deleted IDH mutant Oligodendroglioma (Figure 5D; Supplementary Figure 10). *GABRG2* is a component of the GABA-A receptor complex,[57] and *NFIB* plays a critical role in neural development.[58]

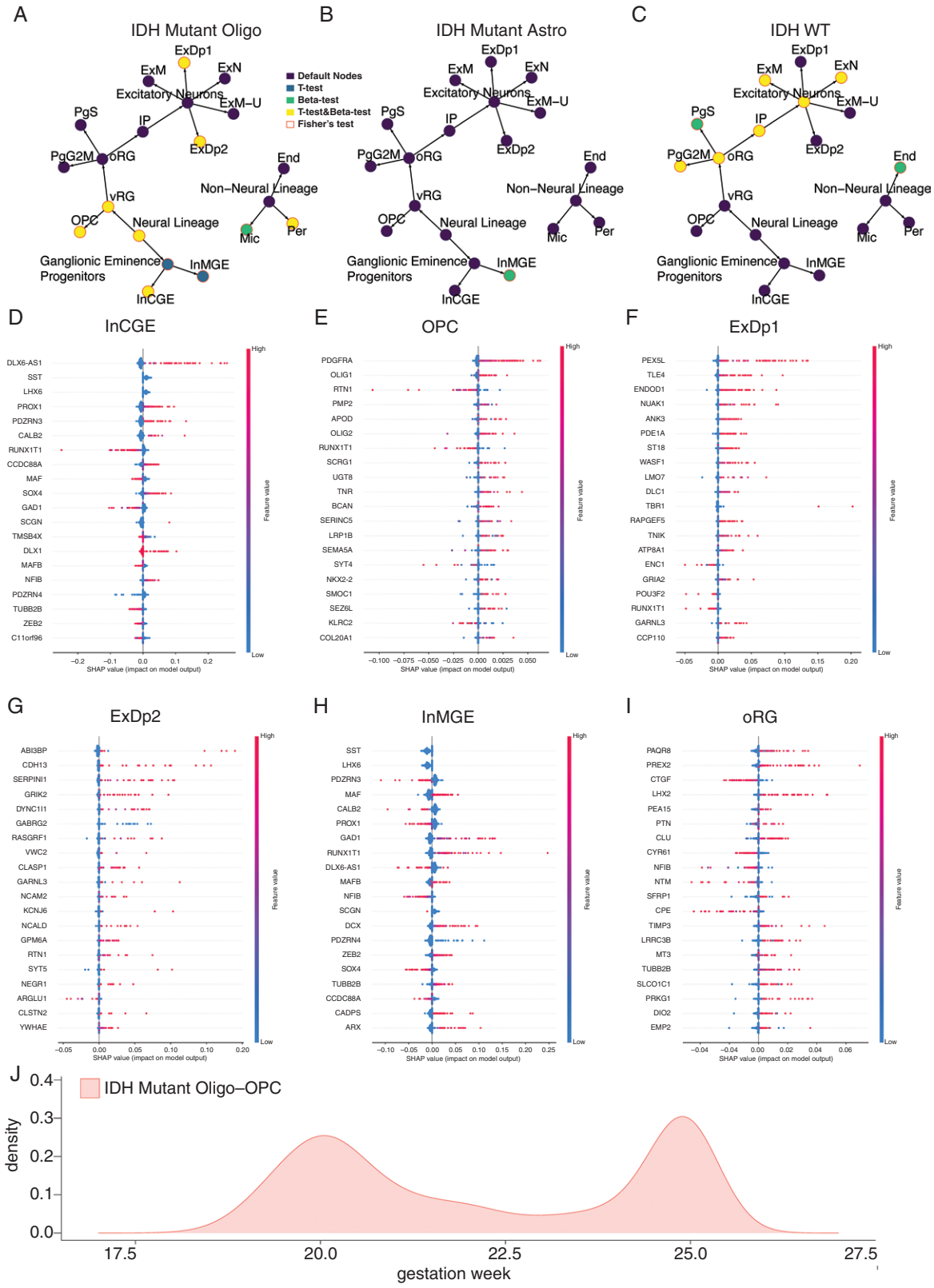


**Figure 4.** Characterization of developmental-like cell states in glioma scRNA-Seq data predicted from Poliudokis et al. pre-trained models.[12] (A–C) Developmental tree inference by COORS showing the hierarchical relationship between tumor cells and normal developmental cell types of IDH Mutant Oligo, IDH Mutant Astro, and IDH WT tumor subgroups respectively. Definitions of cell types are as follows: vRG, ventricular radial glia; oRG, outer radial glia; IP, intermediate progenitors; PgS, proliferating progenitors (S-phase); PgG2M, proliferating progenitors (G2/M-phase);

ExN, excitatory neurons; ExM, excitatory neurons mature; ExM-U, excitatory neurons mature-unspecified; ExDp1, excitatory deep layer neurons 1; ExDp2, excitatory deep layer neurons 2; OPC, oligodendrocyte precursor cells; InMGE, interneurons from medial ganglionic eminence; InCGE, interneurons from caudal ganglionic eminence; End, endothelial cells; Per, pericytes; Mic, microglia. (D–I) The figure displays the results of SHAP analysis, showing the top impactful genes from each cell type in InCGE, OPC, ExDp1, ExDp2, InMGE, oRG cell types predicted from Poliudokis et al. pre-trained models.[12] (J) Distribution of age mapping within IDH mutant oligodendrogliomas and their respective mapped OPC cell of origin pair. Abbreviations: COORS, Cell Of ORigin like CellS; scRNA-Seq, single-cell RNA-sequencing; SHAP, SHapley Additive exPlanations.

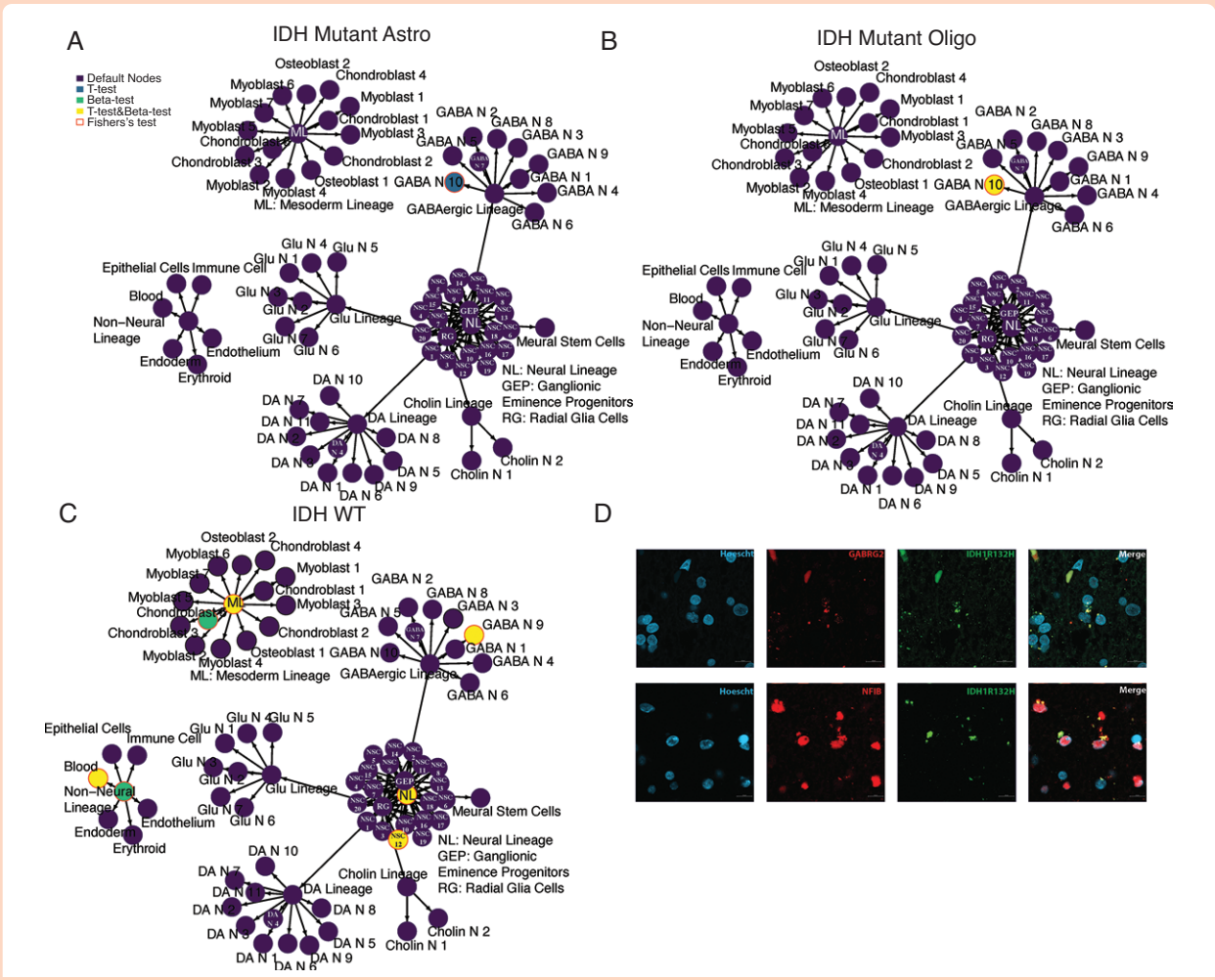


**Figure 5.** Characterization of developmental-like cell states in glioma scRNA-Seq data predicted from Zeng et al. pre-trained models.[15] (A–C) Developmental tree inference by COORS showing the hierarchical relationship between tumor cells and normal developmental cell types using Zeng et al. pre-trained models[15] of IDH Mutant Oligo, IDH Mutant Astro, and IDH WT tumor subgroups respectively. Nodes are color-coded based on their statistical significance: default nodes are insignificant (*P*-value > .05). Definitions of cell types are as follows (subclusters of the cell types are defined in Zeng et al. study): NSCs, neural stem cells; IPC, intermediate progenitor cells; Glu, glutamatergic neurons; GABA, GABAergic neurons; DA, dopaminergic neurons; Ch, cholinergic neurons. (D) Representative immunostaining for *GABRG2* and *NFIB* (red), and IDHR132H (green) in a human IDH$^{mut}$ oligodendroglioma tumor sample; scale bar = 20 μm.

## Discussion

Here we presented a hierarchical machine learning-based approach, named COORS, for the identification and characterization of tumor cells that exhibit gene expression patterns reminiscent of early developmental stages in the brain. COORS achieves this by employing NNMs trained on diverse scRNA-Seq datasets from developing human brain tissues. We applied our method to predict developmental-like cells in various brain cancer datasets, including MB, DMG, and glioma, with validation against well-characterized MB data, which aligns with recent publications on MB subgroup cell origins.[8,9,26,59] Earlier research

identified GN as progenitor cells for SHH-induced MB, while another study proposed that GP4 originates from the UBC lineage.[8] The mapping of DMG found that H3.1/2K27M tumors mapped to ependymal-like cells whereas H3.3 mapped to neuronal IPCs.

COORS identified oRG developmental-like cells within $IDH^{WT}$ glioma cells and OPC and neuronal-like cells in $IDH^{Mut}$. Interestingly, $IDH^{Mut}$ subgroup cells that map to OPC show bimodal distributions, that are both early and late weeks in development. Furthermore, COORS offers a valuable resource by providing novel markers for each tumor subgroup cell of origin mapping. Moreover, we provide information on the DEGs between tumor cells and their respective cells of origin, as well as between the cell of origin and the subsequent developmental cell type. These novel markers hold promise as potential therapeutic targets, offering new avenues for the development of targeted therapies for brain tumors.

In the past, efforts to induce the differentiation of cancer cells into more mature, less aggressive cell types, without damaging normal cells were met with limited success in solid tumors, likely due to insufficient understanding of the precise progenitor cells involved.[60] However, advancements in our comprehension of specific time points and cell types have paved the way for a more nuanced approach. By examining the subsequent steps in the lineage, such as the differentiation of OPCs into mature oligodendrocytes, we can identify key genes involved in this process. For instance, oligodendrocytes are characterized by decreased proliferative capacity compared to OPCs, suggesting a regulatory role for certain genes in cell fate determination. Targeting genes like *OLIG2*, which maintains OPC identity, while activating those involved in mature oligodendrocyte function, such as myelin basic protein (*MBP*), holds promise for directing OPCs toward differentiation into oligodendrocytes. This approach presents a potential avenue for differentiation therapy, wherein manipulating gene expression could drive tumor cells toward a more benign phenotype, offering a novel strategy for cancer treatment.

Overall, our approach relies on a separate model for each origin like cell type. While one could use a single model for the assignment of all cell types jointly (eg, multitask learning) to make use of all data at once, our experiments did not show significant benefit of one model compared to building simpler single models for each cell type. We hypothesize that this may be inherently a result of data size requirements and complexities of jointly learning hundreds of origin like cell types from unbalanced datasets. In addition, our approach provides more flexibility for selecting biologically meaningful models of origin like cell types in different tumors.

In summary, we developed COORS, a machine learning-based algorithm leveraging hierarchical cell type-specific multilayer perceptron models, to investigate the developmental origins of brain tumors capitalizing on the existing scRNA-Seq data. COORS enables origin like annotations in various brain tumors, flexibly mapping tumor cells from medulloblastoma, glioma, and DMG to developmental cell states that represent putative cells of origin, potentially extending this approach to other cancer types. This work adds to our cumulative understanding of the parallel biological processes in both normal development and brain tumors, paving the avenue for finding potential therapeutic targets.

## Supplementary material

Supplementary material is available online at *Neuro-Oncology Advances* (https://academic.oup.com/noa).


## Keywords

artificial neural network model | brain tumor | cell of origin | scRNA-Seq

## Funding

This work received no external funding.

## Acknowledgments

Schematics were created using Biorender.com.

*Conflict of interest statement*. None declared.


## Author contributions

The conception of the project, study and pipeline design, and interpretation of results: S.W., R.N.C., A.O.H., and A.S.H. Code development for the COORS pipeline: S.W. and A.S.H. Immunostaining: M.F.M. Manuscript preparation: S.W., A.O.H., and A.S.H. Manuscript feedback and contributions: R.N.C., H.Y.K., A.W.E., C.L.K., M.D.T., G.R., and B.D.

## Ethics statement

This study involved the analysis of previously published and publicly available datasets and did not involve the collection of new human or animal samples. As such, formal ethical approval was not required. However, all data used were obtained from sources that ensured compliance with ethical standards and obtained appropriate consent from participants where applicable.

## Data availability

Source code of COORS is publicly available at https://github.com/Su-Wang-UTH/COORS. All datasets used in this article were published in previous studies, and detailed information is provided in the main manuscript.

## Affiliations

Department of Neurosurgery, Baylor College of Medicine, Houston, Texas, USA (S.W., G.R., B.D., A.S.H.); Graduate School of Biomedical Sciences, Baylor College of Medicine, Houston, Texas, USA (R.N.C.); Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, Texas, USA (R.N.C., M.F.M.D., B.D.); Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas, USA (M.F.M.D.); Program in Development, Disease Models, and Therapeutics, Baylor College of Medicine, Houston, Texas, USA (M.F.M.D.); Department of Pediatrics and Neurology, Baylor College of Medicine, Houston, Texas, USA (H.Y.K.); The Arthur and Sonia Labatt Brain Tumor Research Center, The Hospital for Sick Children, Toronto, Ontario, Canada (A.W.E.); Developmental and Stem Cell Biology Program, The Hospital for Sick Children, Toronto, Ontario, Canada (A.W.E.); Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada (A.W.E.); Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada (C.K.); Department of Human Genetics, McGill University, Montreal, Quebec, Canada (C.K.); Texas Children's Cancer Center, Hematology-Oncology Section, Texas Childeren's Hospital, Houston, Texas, USA (M.D.T.); Department of Pediatrics, Hematology/Oncology and Neurosurgery, Baylor College of Medicine, Houston, Texas, USA (M.D.T.); Center for Cancer Neuroscience, Baylor College of Medicine, Houston, Texas, USA (G.R., B.D., A.S.H.); McWilliams School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas, USA (A.O.H.)


## References


1. Friedmann-Morvinski D. Glioblastoma heterogeneity and cancer cell plasticity. *Crit Rev Oncog.* 2014;19(5):327–336.
2. Patel AP, Tirosh I, Trombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science.* 2014;344(6190):1396–1401.
3. Filbin MG, Tirosh I, Hovestadt V, et al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science.* 2018;360(6386):331–335.
4. Venteicher AS, Tirosh I, Hebert C, et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science.* 2017;355(6332):eaai8478.
5. Curry RN, Glasgow SM. The role of neurodevelopmental pathways in brain tumors. *Front Cell Dev Biol.* 2021;9:659055.
6. Jessa S, Blanchet-Cohen A, Krug B, et al. Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet.* 2019;51(12):1702–1713.
7. Phoenix TN. The origins of medulloblastoma tumours in humans. *Nature.* 2022;609(7929):901–903.
8. Smith KS, Bihannic L, Gudenas BL, et al. Unified rhombic lip origins of group 3 and group 4 medulloblastoma. *Nature.* 2022;609(7929):1012–1020.
9. Selvadurai HJ, Luis E, Desai K, et al. Medulloblastoma arises from the persistence of a rare and transient Sox2+ granule neuron precursor. *Cell Rep.* 2020;31(2):107511.
10. Vladoiu MC, El-Hamamy I, Donovan LK, et al. Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature.* 2019;572(7767):67–73.
11. Jovic D, Liang X, Zeng H, et al. Single-cell RNA sequencing technologies and applications: a brief overview. *Clin Transl Med.* 2022;12(3):e694.
12. Polioudakis D, de la Torre-Ubieta L, Langerman J, et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron.* 2019;103(5):785–801.e8.
13. Aldinger KA, Thomson Z, Phelps IG, et al. Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci.* 2021;24(8):1163–1175.
14. Jessa S, Mohammadnia A, Harutyunyan AS, et al. K27M in canonical and noncanonical H3 variants occurs in distinct oligodendroglial cell lineages in brain midline gliomas. *Nat Genet.* 2022;54(12):1865–1880.
15. Zeng B, Liu Z, Lu Y, et al. The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell Stem Cell.* 2023;30(6):851–866.e7.
16. Bhaduri A, Sandoval-Espinosa C, Otero-Garcia M, et al. An atlas of cortical arealization identifies dynamic molecular signatures. *Nature.* 2021;598(7879):200–204.
17. Riemondy KA, Venkataraman S, Willard N, et al. Neoplastic and immune single-cell transcriptomics define subgroup-specific intratumoral heterogeneity of childhood medulloblastoma. *Neuro-Oncology* 2021;24(2):273–286. 10.1093/neuonc/noab135.
18. Curry RN, Aiba I, Meyer J, et al. Glioma epileptiform activity and progression are driven by IGSF3-mediated potassium dysregulation. *Neuron.* 2023;111(5):682–695.e9.
19. Curry RN, Ma Q, McDonald MF, et al. Integrated electrophysiological and genomic profiles of single cells reveal spiking tumor cells in human glioma. *Cancer Cell.* 2024;42(10):1713–1728.e6.
20. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *arXiv[csAI].* 2017:4768–4777. http://arxiv.org/abs/1705.07874
21. Northcott PA, Korshunov A, Witt H, et al. Medulloblastoma comprises four distinct molecular variants. *J Clin Oncol.* 2011;29(11):1408–1414.
22. Northcott PA, Buchhalter I, Morrissy AS, et al. The whole-genome landscape of medulloblastoma subtypes. *Nature.* 2017;547(7663):311–317.
23. Northcott PA, Korshunov A, Pfister SM, Taylor MD. The clinical implications of medulloblastoma subgroups. *Nat Rev Neurol.* 2012;8(6):340–351.
24. Ramaswamy V, Taylor MD. Medulloblastoma: from myth to molecular. *J Clin Oncol.* 2017;35(21):2355–2363.
25. Gibson P, Tong Y, Robinson G, et al. Subtypes of medulloblastoma have distinct developmental origins. *Nature.* 2010;468(7327):1095–1099.
26. Hendrikse LD, Haldipur P, Saulnier O, et al. Failure of human rhombic lip differentiation underlies medulloblastoma formation. *Nature.* 2022;609(7929):1021–1028.
27. Yabe T, Hata T, He J, Maeda N. Developmental and regional expression of heparan sulfate sulfotransferase genes in the mouse brain. *Glycobiology.* 2005;15(10):982–993.
28. Augustin I, Korte S, Rickmann M, et al. The cerebellum-specific Munc13 isoform Munc13-3 regulates cerebellar synaptic transmission and motor learning in mice. *J Neurosci.* 2001;21(1):10–17.
29. Kim D, Ackerman SL. The UNC5C netrin receptor regulates dorsal guidance of mouse hindbrain axons. *J Neurosci.* 2011;31(6):2167–2179.
30. Mille F, Tamayo-Orrego L, Lévesque M, et al. The Shh receptor Boc promotes progression of early medulloblastoma to advanced tumors. *Dev Cell.* 2014;31(1):34–47.
31. Berglund EO, Murai KK, Fredette B, et al. Ataxia and abnormal cerebellar microorganization in mice with ablated contactin gene expression. *Neuron.* 1999;24(3):739–750.
32. Yang H, Scholich K, Poser S, et al. Developmental expression of PAM (protein associated with MYC) in the rodent brain. *Brain Res Dev Brain Res.* 2002;136(1):35–42.
33. Vanner RJ, Remke M, Gallo M, et al. Quiescent Sox2+ cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. *Cancer Cell.* 2014;26(1):33–47.

34. Mullen RJ, Buck CR, Smith AM. NeuN, a neuronal specific nuclear protein in vertebrates. *Development.* 1992;116(1):201–211.
35. Hills LB, Masri A, Konno K, et al. Deletions in GRID2 lead to a recessive syndrome of cerebellar ataxia and tonic upgaze in humans. *Neurology.* 2013;81(16):1378–1386.
36. Sekerková G, Ilijic E, Mugnaini E. Time of origin of unipolar brush cells in the rat cerebellum as observed by prenatal bromodeoxyuridine labeling. *Neuroscience.* 2004;127(4):845–858.
37. Englund C, Kowalczyk T, Daza RAM, et al. Unipolar brush cells of the cerebellum are produced in the rhombic lip and migrate through developing white matter. *J Neurosci.* 2006;26(36):9184–9195.
38. Forget A, Martignetti L, Puget S, et al. Aberrant ERBB4-SRC signaling as a hallmark of group 4 medulloblastoma revealed by integrative phosphoproteomic profiling. *Cancer Cell.* 2018;34(3):379–395.e7.
39. Gao X, Zhuang Q, Li Y, et al. Single-cell chromatin accessibility analysis reveals subgroup-specific TF-NTR regulatory circuits in medulloblastoma. *Adv Sci (Weinh).* 2024;11(30):e2309554.
40. Sui Y, Gu R, Janknecht R. Crucial functions of the JMJD1/KDM3 epigenetic regulators in cancer. *Mol Cancer Res.* 2020;19(1):3–13.
41. Betancourt J, Katzman S, Chen B. Nuclear factor one B regulates neural stem cell differentiation and axonal projection of corticofugal neurons. *J Comp Neurol.* 2014;522(1):6–35.
42. Yasuhara N, Shibazaki N, Tanaka S, et al. Triggering neural differentiation of ES cells by subtype switching of importin-alpha. *Nat Cell Biol.* 2007;9(1):72–79.
43. Evans MK, Matsui Y, Xu B, et al. Ybx1 fine-tunes PRC2 activities to control embryonic brain development. *Nat Commun.* 2020;11(1):4060.
44. Jacquet BV, Salinas-Mondragon R, Liang H, et al. FoxJ1-dependent gene expression is required for differentiation of radial glia into ependymal cells and a subset of astrocytes in the postnatal brain. *Development.* 2009;136(23):4021–4031.
45. MacDonald A, Lu B, Caron M, et al. Single cell transcriptomics of ependymal cells across age, region and species reveals cilia-related and metal ion regulatory roles as major conserved ependymal cell functions. *Front Cell Neurosci.* 2021;15:703951.
46. Abdelhamed Z, Vuong SM, Hill L, et al. A mutation in Ccdc39 causes neonatal hydrocephalus with abnormal motile cilia development in mice. *Development.* 2018;145(1):dev154500.
47. Kessaris N, Fogarty M, Iannarelli P, et al. Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nat Neurosci.* 2006;9(2):173–179.
48. Wamsley B, Fishell G. Genetic and activity-dependent mechanisms underlying interneuron diversity. *Nat Rev Neurosci.* 2017;18(5):299–309.
49. Miyoshi G, Butt SJB, Takebayashi H, Fishell G. Physiologically distinct temporal cohorts of cortical interneurons arise from telencephalic Olig2-expressing precursors. *J Neurosci.* 2007;27(29):7786–7798.
50. Benamer N, Vidal M, Angulo MC. The cerebral cortex is a substrate of multiple interactions between GABAergic interneurons and oligodendrocyte lineage cells. *Neurosci Lett.* 2020;715(134615):134615.
51. Fruttiger M, Karlsson L, Hall AC, et al. Defective oligodendrocyte development and severe hypomyelination in PDGF-A knockout mice. *Development.* 1999;126(3):457–467.
52. Bhaduri A, Di Lullo E, Jung D, et al. Outer radial Glia-like cancer stem cells contribute to heterogeneity of glioblastoma. *Cell Stem Cell.* 2020;26(1):48–63.e6.
53. Lu QR, Sun T, Zhu Z, et al. Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. *Cell.* 2002;109(1):75–86.
54. Vallstedt A, Klos JM, Ericson J. Multiple dorsoventral origins of oligodendrocyte generation in the spinal cord and hindbrain. *Neuron.* 2005;45(1):55–67.
55. Zhou Q, Anderson DJ. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell.* 2002;109(1):61–73.
56. Zhou Q, Wang S, Anderson DJ. Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. *Neuron.* 2000;25(2):331–343.
57. Olsen RW, Sieghart W. GABA A receptors: subtypes provide diversity of function and pharmacology. *Neuropharmacology.* 2009;56(1):141–148.
58. Steele-Perkins G, Plachez C, Butz KG, et al. The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol Cell Biol.* 2005;25(2):685–698.
59. Schüller U, Heine VM, Mao J, et al. Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell.* 2008;14(2):123–134.
60. Hugues de Thé. Differentiation therapy revisited. *Nat Rev Cancer.* 2018;18(2):117–127.