Paediatric brain tumours are most commonly found in the posterior fossa, particularly the cerebellum, where medulloblastoma, ependymoma and pilocytic astrocytoma are most prevalent. Cerebellar tumours are currently treated using non-specific therapies. They have few somatically mutated driver genes, which has impeded the understanding of their biology and resulted in the development of targeted therapies lagging behind other major cancer types. Nevertheless, childhood cerebellar tumour types are known to have distinct molecular subtypes with different clinical behaviours. Medulloblastoma is now known to comprise four molecularly distinct diseases (subgroups), with further clinical and molecular heterogeneity within each subgroup. Sonic Hedgehog medulloblastoma subgroup transcriptionally mirrors the granule cell hierarchy as expected, while group 3 medulloblastoma resembles Nestin* stem cells, group 4 medulloblastoma resembles unipolar brush cells, and PFA/PFB ependymoma and cerebellar pilocytic astrocytoma resemble the prenatal gliogenic progenitor cells. Furthermore, single-cell transcriptomics of human childhood cerebellar tumours demonstrates that many bulk tumours contain a mixed population of cells with divergent differentiation. Our data highlight cerebellar tumours as a disorder of early brain development and provide a proximate explanation for the peak incidence of cerebellar tumours in early childhood.

Study of the origin and development of cerebellar tumours has been hampered by the complexity and heterogeneity of cerebellar cells that change over the course of development. Here we use single-cell transcriptomics to study more than 60,000 cells from the developing mouse cerebellum and show that different molecular subgroups of childhood cerebellar tumours mirror the transcription of cells from distinct, temporally restricted cerebellar lineages. The Sonic Hedgehog medulloblastoma subgroup transcriptionally mirrors the granule cell hierarchy as expected, while group 3 medulloblastoma resembles Nestin* stem cells, group 4 medulloblastoma resembles unipolar brush cells, and PFA/PFB ependymoma and cerebellar pilocytic astrocytoma resemble the prenatal gliogenic progenitor cells. Furthermore, single-cell transcriptomics of human childhood cerebellar tumours demonstrates that many bulk tumours contain a mixed population of cells with divergent differentiation. Our data highlight cerebellar tumours as a disorder of early brain development and provide a proximate explanation for the peak incidence of cerebellar tumours in early childhood.

1Developmental & Stem Cell Biology Program, The Hospital for Sick Children, Toronto, Ontario, Canada. 2The Arthur and Sonia Labatt Brain Tumor Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. 3Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. 4Computational Biology Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. 5Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. 6Division of Haematology/Oncology, Department of Pediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada. 7Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. 8Brain Tumor Program, Children’s Cancer Center and Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA. 9Department of Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada. 10Cancer Research Program, McGill University Health Centre Research Institute, Montreal, Quebec, Canada. 11Department of Cell and Cancer Research Institute, McMaster University, Hamilton, Ontario, Canada. 12Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada. 13Department of Surgery, McMaster University, Hamilton, Ontario, Canada. 14Charbonneau Cancer Institute, University of Calgary, Calgary, Alberta, Canada. 15Canada’s Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. 16Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. 17Division of Neurosurgery, The Hospital for Sick Children, Toronto, Ontario, Canada. 18Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. 19Ontario Institute for Cancer Research, Toronto, Ontario, Canada. 20Department of Neurosurgery, Duke University Medical Center, Durham, NC, USA. 21Department of Human Genetics, McGill University, Montreal, Quebec, Canada. 22Lady Davis Research Institute, Jewish General Hospital, Montreal, Quebec, Canada. 23Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. 24Division of Experimental Medicine, McGill University, Montreal, Quebec, Canada. 25Department of Surgery and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. 26These authors contributed equally: Maria C. Vladoiu, Ibrahim El-Hamamy, Laura K. Donovan. 27These authors jointly supervised this work: Nada Jabado, Lincoln Stein, Michael D. Taylor. *e-mail: nada.jabado@mcgill.ca; lincoln.stein@oicr.on.ca; mdt.cns@gmail.com

https://doi.org/10.1038/s41586-019-1158-7
Fig. 1 | Identification of cell types in the developing mouse cerebellum. t-distributed stochastic neighbour embedding (t-SNE) visualization of transcriptionally distinct cell populations from 62,040 single cells from nine developmental time points. Clusters of cells were identified using a shared nearest neighbour (SNN)-modularity-optimization-based clustering algorithm implemented using the Seurat R package. The cells are colour-coded by time point, as indicated by the legend on the right.

Cellular diversity in the developing mouse cerebellum

We isolated the mesial cerebellum (embryonic day (E)14–postnatal day (P)14) or hindbrain (E10–E12) from wild-type mice and performed scRNA-seq on more than 60,000 cells from five embryonic time points and four early postnatal time points (Fig. 1, Extended Data Figs. 1a–c, 2). Unsupervised clustering of individual cell transcriptomes yielded more than 30 distinct clusters, many of which were heavily populated by cells from specific time points in development (Fig. 1). Expression of known marker genes allowed the identification of clusters of progenitors belonging to glutamatergic (Atoh1), GABAergic (Ptf1a, Calb1, Pax2) and glial (Fabp7, Gdf10, Olig1) lineages (Extended Data Figs. 1d, 4). Stem cell-like clusters marked by Nestin expression were primarily seen early, whereas glutamatergic and GABAergic neuronal populations appeared in mid-development, and glial cells developed later overall. By contrast, non-central-nervous-system cells were found across all time points. Several distinct clusters of cells only appeared during restricted developmental time points, with many not found in the postnatal period (Extended Data Fig. 4). We conclude that it is possible to use scRNA-seq to identify biologically distinct cerebellar cellular populations.

Cerebellar tumours mirror embryonic cell populations

Using a carefully curated list of mouse and/or human orthologues and an algorithm to deconvolute complex RNA mixtures against a series of single-cell-type transcriptional profiles, we compared the transcriptomes of bulk human ependymomas (PFA and PFB, 43 tumours), as well as cerebellar pilocytic astrocytomas (C-PAs, 10 tumours) to distinct developmental clusters. PFA, PFB and C-PA are all transcriptionally most similar to the gliogenic progenitors-1 cell cluster (Extended Data Fig. 5), with some similarity to the proliferating ventricular zone progenitors, and to a previously unknown cluster of ‘roof-plate-like’ stem cells which has transcriptional similarity to the developing roof plate (Lmx1a, Msi1, Bmp7)45,46. The gliogenic progenitor cluster is observed initially at E12, peaks at E18 and has largely disappeared by P9 (Extended Data Fig. 4). The roof-plate-like stem cells are seen much earlier, appearing only between E10 and E12 (Extended Data Fig. 4). This is consistent with PFA, PFB and C-PA being classified as gliomas, and previous studies that suggest that PFA arises from the regional radial glia47.

Posterior fossa ependymomas and C-PAs are both histologically and clinically distinct from each other. Re-clustering of the gliogenic progenitors, early ventricular zone radial glia and roof-plate stem cells identifies eight distinct transcriptional clusters within this lineage (Extended Data Fig. 7). Both PFA and PFB remain transcriptionally best matched to the same developmental population (Extended Data Fig. 7). However, C-PAs now transcriptionally match to a very distinct sub-cluster of the cerebellar gliogenic progenitor cell cluster, supporting a model in which posterior fossa ependymomas and C-PAs have distinct cells of origin (Extended Data Fig. 7).

By deconvoluting bulk RNA-seq transcriptomes from human medulloblastomas (SHH, group 3 and group 4 MB, total 145 tumours) against transcriptionally defined cerebellar clusters (Fig. 3), we demonstrated that SHH MB shared most similarities to GCP clusters, as supported by previous publications22,23. Re-clustering of cells in the GCP lineage revealed additional heterogeneity, with the identification of seven distinct clusters (Extended Data Fig. 8). Comparison of bulk human SHH MB transcriptomes to these seven GCP cell lineage clusters revealed heterogeneity within SHH MB (Extended Data Fig. 8c). SHH MBs that transcriptionally resemble later time points in cerebellar development (SHH-1, similar to the postnatal GCPs 2.1) have a worse prognosis than those that resemble the earlier-arising GCPs (SHH-2, similar to the postnatal GCPs 1.1, P = 0.00442; Extended Data Fig. 8g). Furthermore, SHH-β subtype medulloblastomas are more similar to the earlier...
SHH-2 subset, demonstrating the similarity of this subtype to differentiation states in earlier GCP development (Extended Data Fig. 8).

The cell of origin for group 3 MB is not yet definitively known but has been suggested to be a Nestin⁺ cerebellar stem cell⁶⁵. Comparison of group 3 MB to developmental cerebellar cell clusters revealed a broad resemblance across group 3 MB to Nestin⁺ cerebellar early stem cells (Fig. 3). Of interest, subsets of group 3 MB transcriptomes also resemble developmental cell clusters in the GCP and UBC lineages,

Fig. 2 | Reconstruction of cerebellar developmental lineages through pseudo-temporal ordering of cells. a–c, t-SNE visualization and two-dimensional embedding showing constructed pseudo-time trajectories of different lineages in the developing cerebellum: astrocyte/Bergmann glia lineage (n = 12,304 cells), early glutamatergic lineage (n = 14,358 cells) and late glutamatergic lineage (n = 14,662 cells). Cells within specific lineage clusters were selected, visualized using t-SNE visualization (Seurat R package) and then ordered based on a reversed graph embedding method (Monocle 2). Heat maps demonstrate gene-normalized expression levels of cluster-specific markers, red being highest and blue being lowest. CN, cerebellar nuclei; NTZ, nuclear transitory zone; VZ, ventricular zone.

Fig. 3 | Deconvolution analyses of bulk human medulloblastoma tumour transcriptomes. Hierarchical clustering of patient samples of known molecular subgroups based on calculated relative abundance values of the mouse cell-type clusters in each sample, obtained from CIBERSORT (https://cibersort.stanford.edu). Expression signatures from 26 mouse cell types were selected to deconvolute bulk RNA-seq of human cerebellar tumours including SHH, group 3 and group 4 MBs (n = 145).
Similarly, C-PAs had a very strong match at both E16 and E18 (Fig. 4c).

Re-clustering of cells in the UBC lineage revealed a specific match to the E16 gliogenic progenitors (Fig. 4a, b). Deconvolution analysis of 45 samples from patients with group 4 MB against different developmental stages of the UBC cluster. f, In situ hybridization staining of medial sagittal slices for the marker gene Eomes during mouse cerebellar development (E15.5 and E18.5). g, Deconvolution analysis of 45 samples from patients with group 4 MB against different developmental stages of the UBC cluster. f, In situ hybridization staining of medial sagittal slices for the marker gene Eomes during mouse cerebellar development (E15.5 and E18.5).

Temporal mirroring of specific embryonic cell clusters

Different childhood tumours transcriptionally mirror specific clusters in defined cerebellar lineages. Many of these lineages are only detected over a defined period of development, whereas others persist into adulthood (Extended Data Fig. 4). We compared human tumour transcriptomes to their best-matched developmental cluster as a function of time. Comparison of bulk PFA and PFB transcriptomes to single mouse cells in the gliogenic progenitor cell lineage from E10 to P0 revealed a specific match to the E16 gliogenic progenitors (Fig. 4a, b). Similarly, C-PAs had a very strong match at both E16 and E18 (Fig. 4c).

The gliogenic progenitor cells form a discrete cluster from E14 to P0 (Fig. 4d), bracketing the period of highest transcriptional resemblance. Comparison of the bulk transcriptomes of human group 4 MB to mouse cells in the UBC lineage revealed that some group 4 MBs are transcriptionally similar to E16, whereas others are more similar to E18 (Fig. 4e). The UBC lineage is well defined and detected from E14 to P0 (Extended Data Fig. 4f). Comparison of group 4 MBs that more closely mirror E16 to those that mirror E18 demonstrated that group 4-β is largely restricted to ‘E16 similar’ tumours, whereas group 4-γ is completely restricted to ‘E18 similar’ tumours (P = 0.00004, Extended Data Fig. 9r). These data suggest that differences between group 4-β and group 4-γ could be secondary to their arising at different time points in the UBC lineage.

We did not attempt to temporally position group 3 MBs using bulk transcriptomics, as group 3 MB often transcriptionally match more than one cluster in the developing cerebellum. Although PFA, PFB, C-PA and group 4 MB are all transcriptionally best matched to cell clusters present during fetal development, SHH MBs are best matched to the GCP lineage in the early postnatal period at P5 (Fig. 4g, h). We conclude that, in addition to transcriptional mirroring of specific cell populations in the developing cerebellum, human cerebellar tumour transcriptomes are most similar to specific time points during development, predominantly during fetal life.

Single-cell heterogeneity in cerebellar tumours

Medulloblastomas exhibit well-characterized intertumoral heterogeneity, as well as geographic, spatial (metastases) and temporal (at recurrence) heterogeneity. Cerebellar ependymomas also show marked intertumoral heterogeneity, but the intertumoral heterogeneity amongst C-PAs is not well characterized. We undertook detailed analysis of the single-cell heterogeneity of C-PAs, revealing a spectrum of subtypes that differ in their genetic and transcriptomic profiles. This analysis highlights the need for further research into the biology of C-PAs, and suggests potential therapeutic targets for the treatment of these tumours.

The gliogenic progenitor cells form a discrete cluster from E14 to P0 (Fig. 4d), bracketing the period of highest transcriptional resemblance. Comparison of the bulk transcriptomes of human group 4 MB to mouse cells in the UBC lineage revealed that some group 4 MBs are transcriptionally similar to E16, whereas others are more similar to E18 (Fig. 4e). The UBC lineage is well defined and detected from E14 to P0 (Extended Data Fig. 4f). Comparison of group 4 MBs that more closely mirror E16 to those that mirror E18 demonstrated that group 4-β is largely restricted to ‘E16 similar’ tumours, whereas group 4-γ is completely restricted to ‘E18 similar’ tumours (P = 0.00004, Extended Data Fig. 9r). These data suggest that differences between group 4-β and group 4-γ could be secondary to their arising at different time points in the UBC lineage.

We did not attempt to temporally position group 3 MBs using bulk transcriptomics, as group 3 MB often transcriptionally match more than one cluster in the developing cerebellum. Although PFA, PFB, C-PA and group 4 MB are all transcriptionally best matched to cell clusters present during fetal development, SHH MBs are best matched to the GCP lineage in the early postnatal period at P5 (Fig. 4g, h). We conclude that, in addition to transcriptional mirroring of specific cell populations in the developing cerebellum, human cerebellar tumour transcriptomes are most similar to specific time points during development, predominantly during fetal life.
Fig. 5 | Cell-type deconvolution analyses of tumour-cell-specific clusters from human medulloblastoma scRNA-seq. a–c. Clustering analysis and t-SNE visualization of scRNA-seq data of samples from patients with SHH MB (n = 2), group 3 MB (n = 2) and group 4 MB (n = 4). Each patient’s sample is shown as a different colour. Each individual tumour cell cluster underwent a deconvolution analysis against 26 previously identified mouse cell populations using CIBERSORT, with each individual tumour cluster identified in the far left column of each heat map.

cRNA-seq of human cerebellar tumours including medulloblastoma (eight patients), PFA ependymoma (four patients) and C-PA (three patients). Analysis of clusters from scRNA-seq of human tumours revealed both tumour cell clusters and non-tumour cell clusters (that is, endothelium, monocytes/microglia, lymphocytes; Extended Data Fig. 6, Supplementary Table 4). In keeping with our aim to compare tumour cells to cell populations in the developing cerebellum, non-tumour cell clusters were removed from consideration.

SHH MB scRNA-seq clusters remain most similar to cells in the GCP lineage (Fig. 5a). Some SHH1 scRNA-seq clusters are most similar to the UBC and GCP progenitor cell cluster. Comparison to the seven re-clustered clusters in the GCP lineage revealed that SHH MBs contain cells that are transcriptionally similar to various different stages of GCP development (Extended Data Fig. 8c). These results are consistent with a model in which SHH MBs contain a variety of tumour cell types that represent different stages of GCP differentiation and that might exhibit distinct clinical behaviours and therapeutic responses.

Identification and transcriptional mapping of individual tumour cell clusters from group 3 MB revealed highly divergent lines of differentiation, with tumour clusters similar to multiple normal developmental clusters in the GCP, UBC, Purkinje cell and GABAergic interneuron lineages. This pattern is consistent with an origin from a very early, uncommitted cerebellar stem cell, followed by partial differentiation of transformed cells along diverse developmental lineages (Fig. 5b).

scRNA-seq of human group 4 MBs revealed discrete clusters that transcriptionally mirror the differentiated UBCs, as well as the UBC progenitors (Fig. 5c, Extended Data Fig. 9). Unexpectedly, we also observed tumour cell clusters from each human group 4 MB that are transcriptionally most similar to the GCP lineage (Fig. 5c). Comparison of group 4 MB single-cell tumour cluster transcriptomes to both the GCP and UBC lineages revealed that ‘GCP similar’ clusters are transcriptionally more similar to the GCP lineage than the UBC lineage (Supplementary Table 8, sheet 5). By contrast, ‘UBC similar’ clusters are more similar to the UBC lineage (Supplementary Table 8, sheet 4). Individual group 4 MBs contain highly variable percentages of differentiated versus less-differentiated cells (Extended Data Fig. 9f). Comparison of the group 4 MB scRNA-seq data revealed that UBC-like cells in group 4 MB transcriptionally mirror several time points in UBC development (Extended Data Fig. 9h). Two distinct types of UBCs have been described in the mammalian cerebellum22. Consistent with this, re-clustering of mouse cerebellar cells in the UBC lineage revealed two distinct types of UBCs (Extended Data Fig. 9c).

Group 4 MB is predominantly similar to only one of these subtypes (Calb2+ UBCs; Extended Data Fig. 9h). Individual cells from the UBC and GCP progenitor cluster simultaneously express both GCP and UBC marker genes, which is not observed in cells committed to a GCP or a UBC fate (Extended Data Fig. 10) but is observed in group 4 MB cells. These data are consistent with a model in which group 4 MB arises from a bipotential progenitor cell population (probably in cluster 8) that is capable of giving rise to cells in both the GCP and the UBC lineages.

Single-cell RNA-seq of PFA ependymomas revealed single-cell heterogeneity, with each tumour containing clusters that best matched sub-clusters of the gliogenic progenitor cell population or roof-plate-like stem cells (Fig. 6a, Extended Data Fig. 7). Of note, we did not observe clusters of more-differentiated cell types such as Bergmann glia or astrocytes within the PFA ependymomas, but only less-differentiated cell types. This lack of differentiated cell types within the tumour is unique to PFA among the childhood cerebellar tumours examined in this study. Some single-cell transcriptional clusters of human PFA tumours are more similar to single-cell clusters from the tumour of another patient than they are to other clusters from the same individual (Extended Data Fig. 7d). C-PA scRNA-seq also revealed clusters with transcriptional similarity to the relatively undifferentiated roof-plate-like stem cell cluster and the gliogenic progenitors (Fig. 6b).

Unlike PFA ependymoma, some C-PA tumour cell clusters demonstrate similarity to more differentiated cell types, such as astrocytes, Bergmann glia and oligodendrocytes (Fig. 6b).
and contaminate their tumours. However, we cannot exclude the possibility that tumors arise in other cell lineages and undergo trans-differentiation during transformation, or arise in cerebellar cells during the very early postnatal period. It is certainly possible that each of the cerebellar tumour types discussed above arises in a particular cell lineage and group 4 MB), as well as PFA ependymoma and C-PA, our scRNA-seq data demonstrated high levels of single-cell heterogeneity, with evidence of multiple lineages of differentiation and tumour cells matching different time points in the differentiation hierarchy. Childhood cerebellar tumour transcriptomes demonstrate high levels of similarity to discrete cell populations within the developing cerebellum, supporting a model in which discrete cells of origin have a profound influence on the transcriptome and biology of the observed mature tumours. Although the use of cell cycle genes is essential in defining developmental stages of cell populations in the cerebellum (that is, progenitor cell versus differentiated cell), we excluded cell cycle genes when comparing human tumour cells to developmental clusters in mice to avoid spurious comparisons based largely on cell cycle phenotypes (Supplementary Table 3); however, cell cycle content and genomic alterations were quantified independently in the human single-cell datasets (Extended Data Fig. 10, Supplementary Table 7). The relative biological significance of clusters that are well separated in the t-SNE plots, versus those that are spread out and partially overlapping, is uncertain, and a well-known issue in single-cell sequence analysis that will require future investigation.

Many of the normal mouse cerebellar cell populations that are transcriptionally most similar to human cerebellar tumours are only present in utero, or immediately postnatally; these populations are therefore not present in the brains of children at the time of presentation and probably could not contaminate their tumours. However, we cannot exclude the possibility that some of the single-cell tumour clusters in our study also contain small populations of infiltrating non-transformed differentiated cells. The ability of these different progenitor cell populations restricted to specific time points to give rise to different types of cerebellar tumours should ideally be tested functionally in vivo, as previously demonstrated4,14,18,23,24. Although there would be great value in comparing human cerebellar tumours to single-cell transcriptomes from normal human cerebellar cells from various time points in development, these types of samples are not readily available. The presence of multiple lineages and stages of differentiation within bulk medulloblastoma samples illustrates the difficulty of using the bulk tumour population as a tool to decipher tumour biology or to develop tumour diagnostics. The absence of differentiated cell types in PFA ependymomas, but not other tumour types, suggests the presence of a differentiation block in PFA ependymomas. The work presented here is focused on the relationship of posterior fossa tumours to cell types in the developing cerebellum. The current scRNA-seq datasets from posterior fossa tumours should allow further analyses and insights into both tumour-cell-autonomous and infiltrating non-tumour-cell biology. A more complete understanding of the biology and transcriptomes of the specific cerebellar hierarchies identified above, and their developmental timing, may allow a better comprehension of cerebellar tumour biology and promote the subsequent development of new mouse models, improved tumour diagnostics and eventually the development of new rational therapeutics based on the differences between tumour cells and their normal cells of origin.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1158-7.

Received: 20 June 2018; Accepted: 22 March 2019; Published online 01 May 2019.

1. Northcott, P. A. et al. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56 (2012).
2. Mack, S. C. et al. Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature* **553**, 101–105 (2018).
3. Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
4. Pugh, T. J. et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106–110 (2012).
5. Northcott, P. A. et al. The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
6. Jones, D. T. et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat. Genet.* **45**, 927–932 (2013).
7. Mack, S. C. et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450 (2014).
8. Cavalli, F. M. G. et al. Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell* **31**, 737–754 (2017).
9. Dubuc, A. M., Mack, S., Unterberger, A., Northcott, P. A. & Taylor, M. D. The epigenetics of brain tumors. *Methods Mol. Biol.* **863**, 139–153 (2012).
10. Dubuc, A. M. et al. Subgroup-specific alternative splicing in medulloblastoma. *Acta Neuropathol.* **123**, 485–499 (2012).
11. Northcott, P. A. et al. Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29**, 1408–1414 (2011).
12. Wechsler-Reya, R. J. & Scott, M. P. Control of neuronal precursor proliferation in the cerebellum by Sonic Hedgehog. *Neuron* **22**, 103–114 (1999).
13. Wallace, V. A. Purkinje-cell-derived Sonic Hedgehog regulates granule neuron precursor cell proliferation in the developing mouse cerebellum. *Curr. Biol.* **9**, 445–448 (1999).
14. Li, P. et al. A population of Nestin-expressing progenitors in the cerebellum exhibits increased tumorigenicity. *Nat. Neurosci.* **16**, 1737–1744 (2013).
15. Kawauchi, D. et al. A mouse model of the most aggressive subgroup of human medulloblastoma. *Cancer Cell* **21**, 168–180 (2012).
16. Pei, Y. et al. An animal model of MYC-driven medulloblastoma. *Cancer Cell* **21**, 155–167 (2012).
17. Swartling, F. J. et al. Pleiotropic role for MYCN in medulloblastoma. *Genes Dev.* **24**, 1059–1072 (2010).
18. Gibson, P. et al. Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095–1099 (2010).
19. Vanner, R. J. et al. Quiescent Sox2+ cells drive hierarchical growth and relapse in Sonic Hedgehog subgroup medulloblastoma. *Cancer Cell* **26**, 33–47 (2014).
20. Pajtler, K. W. et al. The current consensus on the clinical management of posterior fossa ependymomas of infancy. *Acta Neuropathol.* **133**, 5–12 (2017).
21. Witt, H. et al. Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. *Cancer Cell* **20**, 143–157 (2011).
22. Johnson, R. A. et al. Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature* **466**, 632–636 (2010).
23. Taylor, M. D. et al. Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell* **8**, 323–335 (2005).
24. Pomeroy, S. L. et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **415**, 436–442 (2002).
25. Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
26. Goldowitz, D. & Hamre, K. The cells and molecules that make a cerebellum. *Trends Neurosci.* **21**, 375–382 (1998).
27. Leto, K. et al. Consensus paper: cerebellar development. *Cerebellum* **15**, 789–828 (2016).
28. 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).
29. Fleming, J. T. et al. The Purkinje neuron acts as a central regulator of spatially and functionally distinct cerebellar precursors. *Dev. Cell* **27**, 278–292 (2013).
30. Parmigiani, E. et al. Heterogeneity and biopotency of astroglial-like cerebellar progenitors and the interneuron and glial lineages. *J. Neurosci.* **35**, 7388–7402 (2015).
31. Wojcinski, A. et al. Cerebellar granule cell replenishment postinjury by adaptive reprogramming of Nestin+ progenitors. *Nat. Neurosci.* **20**, 1361–1370 (2017).
32. Saunders, A. et al. Molecular diversity and specialization among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
33. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
34. Yeung, J. & Goldowitz, D. Wls expression in the rhombic lip orchestrates the embryonic development of the mouse cerebellum. *Neuroscience* **354**, 30–42 (2017).
35. Yeung, J. et al. Wls provides a new compartmental view of the rhombic lip in mouse cerebellar development. *J. Neurosci.* **34**, 12527–12537 (2014).
36. Garzia, L. et al. A hematogenous route for medulloblastoma leptomingeal metastases. *Cell* **173**, 1549 (2018).
37. Morrissey, A. S. et al. Spatial heterogeneity in medulloblastoma. *Nat. Genet.* **49**, 780–788 (2017).
38. Morrissey, A. S. et al. Divergent clonal selection dominates medulloblastoma at recurrence. *Nature* **529**, 351–357 (2016).
39. Wu, X. et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature* **482**, 529–533 (2012).
40. Tan, L. L. et al. Lateral cerebellum is preferentially sensitive to high Sonic Hedgehog signaling and medulloblastoma formation. *Proc. Natl Acad. Sci. USA* **115**, 3392–3397 (2018).
41. Kim, J. A., Sekerková, G., Mugnaini, E. & Martina, M. Electrophysiological, morphological, and topological properties of two histochemically distinct subpopulations of cerebellar unipolar brush cells. *Cerebellum* **11**, 1012–1025 (2012).

**Acknowledgements**

M.D.T. is supported by the NIH (R01CA148699 and R01CA159859), The Pediatric Brain Tumor Foundation, The Terry Fox Research Institute, The Canadian Institutes of Health Research, The Cure Search Foundation, b.r.a.i.n.child, Meagan’s Walk, The SWIFTY Foundation, Genome Canada, Genome BC, Genome Quebec, the Ontario Research Fund, Worldwide Cancer Research, V-Foundation for Cancer Research and the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. M.D.T. is also supported by a Canadian Cancer Society Research Institute Impact grant and by a Stand Up To Cancer (SU2C) St. Baldrick’s Pediatric Dream Team Translational Research Grant (SU2C-AACR-DT1113) and SU2C Canada Cancer Stem Cell Dream Team Research Funding (SU2C-AACR-DT19-15) provided by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, with supplementary support from the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research. M.D.T. is also supported by the Gannon Family Chair in Childhood Cancer Research at the Hospital for Sick Children and the University of Toronto. L.S. and I.E.-H. were supported by funding provided by the Government of Ontario. M.C.V is supported by The Canadian Institutes of Health Research Doctoral scholarship. A.L.J was supported by NIMH-R37MH085726, NCI-C1A92176 and NINDS-R01NS092096 and a National Cancer Institute Cancer Center Support Grant (P30 CA008748-48). This study was conducted with the support of the Ontario Institute for Cancer Research’s Genomics & Bioinformatics platform (https://genomics.oicr.on.ca/) through funding provided by the Government of Ontario.

**Reviewer information**

Nature thanks Kathleen Millen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions**

M.C.V, I.E.-H. and L.K.D. designed, performed and analysed the majority of the experiments in this study. H.F. contributed to the pre-processing of scRNA-seq data. B.L.H. contributed to timed mating and tissue isolation in developing mouse embryos. Y.S. and L.H. contributed to the SNV and CNV analysis of the scRNA-seq data. V.R. contributed to the clinical analysis of the primary tumour samples. L.D.H contributed to the differential gene expression analysis. S.K. contributed to primary tissue isolation, preparation and statistical analysis of scRNA-seq data. V.F. contributed to literature review, marker gene discovery and cell cluster annotations. P.S., H.S., A.S.M. and F.M.G.C. contributed to the analysis of the bulk RNA-seq data and the data availability submission. J.I.Y.L., H.J., D.P., A.M., B.L. and M.A.Q contributed to the collection and processing of primary tumour samples. K.N. optimized the 10X Genomics scRNA-seq protocols. C.D., X.W. and F.N. provided reagents, equipment and expert advice. S.C.M., L.G., S.K.S., J.A.C., M.A.M., D.M., P.D., T.P., F.N., E.M.T. and C.L.K provided clinical samples and helped design the study. A.L.J provided expert advice on identification of developing cerebellar cell populations. N.J., L.S. and M.D.T supervised the project and wrote the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1158-7.

**Supplementary information**

is available for this paper at https://doi.org/10.1038/s41586-019-1158-7.

**Reprints and permissions information**

is available at http://www.nature.com/reprints.

**Correspondence and requests for materials**

should be addressed to N.J., L.S. or M.D.T.

**Publisher’s note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Patient recruitment. Recruitment of participants was in compliance with the ethical regulations of the Hospital for Sick Children and McGill University Health Centre. Sample size of participants was determined based on consent availability and diagnosis. Available participant characteristic information is described in Supplementary Table 2. All collection and experiments involving samples from patients with tumours were approved by McGill University Health Centre (Montreal) and by The Arthur and Sonia Labatt Brain Tumour Research/Hospital for Sick Children (Toronto).

Animal experiments. All mouse breeding and procedures were approved by The Centre for Phenogenomics (Toronto). Mated C57BL/6 female mice were dissected in order to collect embryos from the following gestational time points: E10, E12, E14, E16 and E18. C57BL/6J pups were dissected to collect tissue from the following postnatal time points: P0, P5, P7 and P14.

Tissue handling and dissociation. Fresh tumour tissue was collected at the time of resection. The tumour tissue was mechanically and enzymatically dissociated using a collagenase-based dissociation method as previously reported. Early embryonic hindbrain structures were dissected from the gestational time points E10 and E12. An incision was made between the midbrain and hindbrain boundary, as well as between the preoptic hindbrain and pontine hindbrain, in order to isolate the isthmus and rhombomeres 1 and 2 at these early time points in development. Late embryonic cerebellar primordia were collected at embryonic time points E14, E16 and E18. Mouse dissections were performed under a Leica stereoscope with a pair of Moria ultra-fine forceps (Fine Science Tools). The tissue was transferred into ice-cold Leibovitz’s medium, followed by single-cell dissociation with the Papain Dissociation System (Worthington Biochemical Corporation). Postnatal cerebella were dissected from the following time points: P0, P5, P7 and P14. The central nervous system was fully dissected, then embedded in 2% low melting point agarose. One mid-sagittal slice of 300 μm was generated using the Leica vibratome. Under the stereoscope the cerebellum was isolated from the slice, followed by immediate single-cell dissociation as described above.

Reverse transcription, amplification and sequencing. The concentration of the single-cell suspension was assessed with a Trypan blue count. Approximately 10,000–14,000 cells per time point 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 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 10x library was evaluated by the Agilent 2100 Bioanalyzer system. The 10x library was loaded in the Illumina 2500 sequencing plates. Raw reads. Through 10x CellRanger pipeline, the raw base call (BCL) files were demultiplexed into FASTQ files. The FASTQ files were aligned to the reference mouse genome GRCh38 (mm10) to generate raw gene-barcode count matrices. When clustering multiple samples, we aggregated the multiple runs together to normalize on sequencing depth and re-computed the gene-barcode matrices. Alignment quality control matrices are found in Supplementary Table 1, sheet 1.

Quality control and normalization. Low-quality cells were identified and removed from the datasets. We considered low-quality cells as cells with <200–300 genes expressed, and cells with high mitochondrial content 4 (4 s.d. above median). We predicted doublets to be cells with relatively high library sizes (4–5 s.d. above median) and removed them from the analysis. Low-abundance genes were also removed from the datasets (genes expressed in <3 cells). The human single-cell datasets were normalized using methods adapted from Scran pipeline. Size factors were computed and were applied to normalize gene expression across the cells, to produce normalized log-expression values to each dataset individually. Mouse single-cell datasets were processed using CellRanger aggregation in order to account for sequencing depth variation, followed by quality control and normalization as described above.

Clustering analysis and visualization. Clustering analysis was performed with both the presence (Fig. 1) and absence (Extended Data Fig. 1) of cell cycle-related genes obtained from Ensembl BioMart. When implemented, removal of cell cycle genes was done before detection of highly variable genes (HVGs). HVGs were detected using Seurat’s pipeline, 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 principal component analysis (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 neighbour (SNN)-modularity-optimization based clustering algorithm from Seurat. We then visualized these clusters using t-distributed stochastic neighbour embedding (t-SNE).

Pseudo-time trajectory analysis. The barcodes of selected clusters were normalized using Monocle dPeat to remove genes with low expression and perform PCA analysis on the remaining genes, for significant PC selection. Cells were then grouped using ‘density peak’ clustering algorithm. Differential gene expression analysis was performed using a generalized linear model, and the top 1,000 genes per cluster were selected. Reversed graph embedding was then used to reduce the high-dimensionality data into lower dimensional space and build the trajectory.

The structure of the trajectory was plotted into two-dimensional space using the DRtree dimensionality reduction algorithm and order the cells in pseudo-time.

Creation of cell-type-specific signatures. For each cluster identified, the average expression of each gene was calculated. Differential gene expression was performed using Seurat’s likelihood ratio test method and we filtered out genes expressed in less than 25% of the cells. The top differentially expressed genes were used as markers to build a signature gene expression matrix. Genes involved in cell proliferation and ribosome biogenesis were obtained from Ensembl BioMart and omitted from the matrix. Human orthologues of mouse genes were identified and used to create the final matrix. Complete gene signatures and inputs can be found in Supplementary Tables 5 and 6.

Deconvolution analysis. CIBERSORT was used to perform the deconvolution analysis of the bulk and scRNA-seq tumour data against the mouse clusters. The full transcriptomes of the tumour data were used as the input mixture, and the signature gene input was the mouse cluster expression matrix after removal of all genes that could introduce bias in the deconvolution process, including 1,400 cell cycle genes, 300 genes associated with ribosome biogenesis, and around 100 mitochondrial and apoptosis-related genes. Quantile normalization was disabled and 100–500 permutations were run. To test CIBERSORT on our datasets, we created synthetic bulk mixtures from the mouse clusters, and selected known amounts of reads from various clusters. CIBERSORT roughly yielded the expected relative abundances. In order to generate reliable input expression profiles, tumour clusters with a very low number of cells were discarded from the analysis. To validate our mouse cluster signatures, we obtained human and mouse data of brain cell types from published datasets, and deconvoluted them against our mouse signatures to ensure that our expected abundances had similar values to our cell of origin matches.

Bulk RNA-seq human tumour samples. A total of 60 SHH, 40 group 3 and 45 group 4 human medulloblastoma bulk RNA-seq samples were obtained from MAGIC, Medulloblastoma Advanced Genomics International Consortium. The raw data were aligned to the reference human genome GRCh38 using STAR to generate raw counts. FPKMs were then obtained from DESeq2. We performed bulk RNA sequencing on 22 PFA, 4 PBC and 10 C-PA patient samples, and obtained FPKMs using the same strategy. Microarray expression data were obtained from 17 PBF samples.

Relative gene expression panels. Relative expression of genes within selected clusters was measured by calculating the average of the gene count from the log-normalized matrix in each cluster. The averages of specific genes were then scaled between the values of −1.0–1.0 among the selected clusters in order to reflect higher versus lower expression levels.

Cell cycle analysis of human scRNA-seq tumour samples. Cell cycle phase-specific annotations were acquired from a previous publication and used to quantify the percentage of phase-specific cell cycle genes in each individual cell. The percentages were then normalized from a scale 0–1, after which the ratio between G1/S and G2/M was calculated. Cells with low ratio values for the G2/M and G1/S ratios were labelled as in G0 phase, whereas cells with increasing values for G1/S but low ratio values were labelled as cells progressing into G1 phase. Cells with high ratio values for G2/M were labelled as cells progressing into S, G2 and M phase, respectively.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The datasets generated and analysed during the current study are available in the Gene Expression Omnibus (GEO) and European Genome-phenome Archive (EGA): https://www.ebi.ac.uk/ega/studies/ repositories: BAMs and filtered gene matrices of mouse developmental time points scRNA-seq (GSE118068), FASTQs of PBF bulk RNA-seq and microarray expression (EAG00001002696, GSE64415), BAMS of human tumour scRNA-seq and either BAMS or FASTQs of bulk PFA/C-PA RNA-seq (EAG00001003170) and FASTQs of MB bulk RNA-seq (EAG00001004453).

Code availability

The following packages were used for the data analysis: Cell Ranger v1.2.1, R v3.4.4, Seurat v1.4.0, v2.3.0 and v2.3.4, Monocle v2.6.3 and CIBERSORT (absolute mode beta).
43. Nguyen, Q. H. et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nat. Commun. 9, 2028 (2018).
44. De Simoni, A. & Yu, L. M. Preparation of organotypic hippocampal slice cultures: interface method. Nat. Protocols 1, 1439–1445 (2006).
45. Zheng, G. X. et al. Massively parallel digital transcripational profiling of single cells. Nat. Commun. 8, 14049 (2017).
46. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. 17, 75 (2016).
47. Zerbino, D. R. et al. Ensembl 2018. Nucleic Acids Res. 46, D754–D761 (2018).
48. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).
49. 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).
50. 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).
51. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982 (2017).
52. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods 12, 453–457 (2015).
53. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
55. Whitfield, M. L. et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol. Biol. Cell. 6, 1977–2000 (2002).
Extended Data Fig. 1 | Characterization of cell types in the mouse developing cerebellum. **a**, t-SNE visualization demonstrating 34 unique clusters of 62,040 single cells. **b**, Bar chart displaying the number of cells collected during each developmental time point \((n = 9)**. **c**, Bar plot displaying the number of cells within each identified cluster belonging to specific developmental time points. **d**, Circles showing the normalized average expression as indicated by the scale at the bottom right of established developmental lineage marker genes \((n = 24)** specific to each cell cluster.
Extended Data Fig. 2 | Clustering analysis of scRNA-seq data of mouse developing cerebellum from seven time points used for generating CIBERSORT expression signatures. a, b, t-SNE visualization (using the Seurat package) of transcriptionally distinct cell populations from 44,461 single cells from seven developmental time points annotated by cluster identity (n = 31) and by time point (n = 7).
Extended Data Fig. 3 | Reconstruction of cerebellar developmental lineages through pseudo-temporal ordering of cells. a–e, t-SNE visualization and two-dimensional embedding showing reconstructed pseudo-time trajectories of different lineages in the developing cerebellum: early germinial zones ($n = 6,096$ cells), GABAergic interneurons lineage ($n = 13,432$ cells), Purkinje cells ($n = 6,048$ cells), granule cells ($n = 15,011$ cells) and oligodendrocytes ($n = 1,433$ cells). Cells within specific lineage clusters were selected, visualized using t-SNE visualization (using the Seurat package) and then ordered based on a reversed graph embedding method (Monocle 2). Heat maps demonstrate gene-normalized expression levels of cluster-specific markers, red being highest and blue being lowest. MOs, myelinating oligodendrocytes; PCs, Purkinje cells.
Extended Data Fig. 4 | Diagram of developing cerebellar lineages showing relative abundance of cell type clusters across time. a, Line plot showing the number of cells of each glutamatergic lineage cluster at each collected time point. b, Line plot showing the number of glial population clusters at each collected time point. c, Line plot showing the number of GABAergic cells at each collected time point. d, Cartoon of individual cell clusters identified through unsupervised hierarchical clustering of single-cell transcriptomes from the developing mouse cerebellum. Cell clusters were arranged in their respective developmental hierarchies based on the expression of known marker genes as well as the results of pseudo-time analyses. Cluster annotations are found on the bottom right.
Extended Data Fig. 5 | Deconvolution analyses of bulk human PFA/ PFB ependymoma and C-PA tumour transcriptomes. Hierarchical clustering of patient samples of known molecular subgroups based on calculated relative abundance values of the mouse cell-type clusters in each sample, obtained from CIBERSORT. Expression signatures from 26 mouse cell clusters were selected to deconvolute bulk RNA-seq of human PFA (n = 22) and PFB (n = 21) ependymomas and C-PAs (n = 10).
Extended Data Fig. 6 | Clustering analysis and t-SNE visualization of human scRNA-seq data. a–e, t-SNE visualization of scRNA-seq data used as input for the CIBERSORT deconvolution analysis of SHH MB (n = 2), group 3 MB (n = 2), group 4 MB (n = 4), PFA (n = 4) and C-PA (n = 3) patient samples. Cluster annotations were established by expression of known marker genes unique to tumour and cell type and are defined as follows: SHH-1, tumour clusters: 1, 2, 3, 4, 5; monocyte/microglia: 6. SHH-2, tumour clusters: 1, 2, 3, 4, 5; monocyte/microglia: 6; T cells: 8. G3-1, tumour clusters: 1, 2, 3, 5, 6; microglia: 10. G3-2, tumour clusters: 1, 2, 3, 4, 5, 6, 7; microglia: 10. G4-1, tumour clusters: 1, 2, 3, 4, 5, 6, 7; microglia: 10. G4-2, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. G4-3, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 10. G4-4, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. PFA-1, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. PFA-2, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. PFA-3, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. PFA-4, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. C-PA-1, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9; T cells: 10. C-PA-2, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11; T cells: 12. C-PA-3, tumour clusters: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11; T cells: 12.
Extended Data Fig. 7 | See next page for caption.
Extended Data Fig. 7 | Re-clustering of the gliogenic progenitors and roof-plate-like stem cells with comparison to PFA/PFB ependymomas and C-PAs. a, t-SNE visualization of the eight sub-clusters obtained from combined re-clustering of roof-plate-like stem cells and gliogenic progenitor clusters (n = 2,525 cells). b, Gene expression of gliogenic progenitor and roof-plate-like stem cell marker genes onto t-SNE of sub-clusters (n = 2,525 cells). c, Pseudo-time trajectory analysis of the eight sub-clusters annotated by sub-cluster (top) and developmental time point (bottom) (n = 2,525 cells). d, Deconvolution analysis heat map of tumour single-cell PFA clusters (n = 9) (top) and tumour single-cell C-PA clusters (n = 6) (bottom) against expression signatures of the 8 mouse developmental sub-clusters. e–g, t-SNE visualizations of clustered populations of PFA (n = 4) and C-PA (n = 3) scRNA-seq patient samples used for CIBERSORT deconvolution analysis. t-SNE visualization of the six sub-clusters obtained from re-clustering of only the gliogenic progenitor sub-clusters (n = 1,709 cells). h, Pseudo-time trajectory analysis of the gliogenic progenitor sub-clusters (n = 1,709 cells) annotated by sub-cluster (top) and developmental time point (bottom). i, Deconvolution analysis heat map of samples from patients with bulk PFA (n = 22), PFB (n = 21) and C-PA (n = 10) against expression signatures of the six gliogenic progenitor sub-clusters.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | Re-clustering of the granule cell lineage with comparison to SHH MBs. a, t-SNE visualization showing seven distinct sub-clusters from re-clustering of the granule cell lineage (n = 15,011 cells). b, Pseudo-time trajectory analysis of the seven granule cell sub-clusters annotated by sub-cluster (top) and developmental time point (bottom) (n = 15,011 cells). c, Deconvolution analysis heat map of bulk SHH MB (n = 60) patient sample transcriptomes against expression signatures of the seven granule cell sub-clusters. d, Deconvolution analysis heat map of SHH MB scRNA-seq tumour-specific clusters (n = 10) against signatures of the seven granule cell sub-clusters. e, t-SNE plot of clustered populations of SHH MB scRNA-seq samples (n = 2). f, Comparison of clinical characteristics based on clustering by similarity to different points in GCP lineage of SHH-1 (n = 15) and SHH-2 (n = 45), comparing age at diagnosis. Box-plot centre lines show data medians; box limits indicate 25th and 75th percentiles; lower and upper whiskers extend to 1.5 times the interquartile range (IQR) from the 25th and 75th percentiles, respectively; outliers are represented by individual points; P value (P = 0.07) was determined by Wilcoxon test. g, Survival curve, corrected for metastatic dissemination and molecular subtype, of SHH-1 (n = 15) and SHH-2 (n = 45) identified through matching to a re-clustered granule cell lineage. P value (P = 0.00442) was determined by log-rank test and ‘+’ indicates censored cases. h–k, Comparison of additional clinical characteristics including histology, sex, molecular subtype affiliation and metastatic status of SHH-1 (n = 15) and SHH-2 (n = 45) patient samples. P values were determined using Fisher's exact test.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Re-clustering of the UBC lineage with comparison to group 4 MBs. a, t-SNE visualization of 6 distinct sub-clusters obtained from re-clustering of the UBC lineage (n = 9,605 cells). b, Gene expression of UBC lineage marker genes onto t-SNE of sub-clusters (n = 9,605 cells). c, Pseudo-time trajectory analysis of the six sub-clusters, showing clear branching of the GCP and UBC lineage annotated by sub-clusters (top) and developmental time point (bottom) (n = 9,605 cells). d, t-SNE visualization of the scRNA-seq clustered populations of group 4 MB human tumour samples (n = 4). e, t-SNE visualization of scRNA-seq clustering analysis of four group 4 MB patient sample tumours coloured by transcriptional match to both UBC and GCP gene expression signatures (9,895 cells positive out of n = 12,129 cells). f, Pie charts showing the percentage of cells at various states of differentiation in three G4 tumour samples based on their matches to UBC precursors, UBCs or postnatal GCPs. g, Deconvolution analysis heat map of group 4 MB (n = 45) bulk patient sample transcriptomes against expression signatures of the 6 UBC sub-clusters. h, Deconvolution analysis heat map of group 4 MB scRNA-seq tumour cell clusters (n = 15) against signatures of the 6 UBC sub-clusters. i–k, t-SNE visualization of re-clustered UBC and GCP progenitor cluster coloured by the number of cells expressing UBC transcriptional signature genes (573 cells positive out of n = 2,866 cells), the number of cells expressing GCP transcriptional signature genes (159 cells positive out of n = 4,607 cells) and the number of cells expressing both UBC and GCP gene signatures (75 cells positive out of n = 4,607 cells). l, Venn diagram showing that group 4 GCP-like clusters express 308 of 600 GCP signatures and 149 of 500 UBC signatures (n = 3,050 genes) (top) compared to group 4 UBC-like clusters which express 136 of 600 GCP signatures and 182 of 500 UBC signatures (n = 3,177 genes) (bottom). m, Comparison of clinical characteristics based on clustering by similarity to E16 and E18 time points in UBC lineage of group 4 MB labelled as group 4 (E16) (n = 17) and group 4 (E18) (n = 28), comparing age at diagnosis. Box-plot centre lines show data medians; box limits indicate 25th and 75th percentiles; lower and upper whiskers extend to 1.5 times the interquartile range (IQR) from the 25th and 75th percentiles, respectively; outliers are represented by individual points; P value (P = 0.45) was determined by Wilcoxon test. n, Survival curve, corrected for metastatic dissemination and molecular subtype, of group 4 (E16) (n = 17) and group 4 (E18) (n = 28) identified through matching to a re-clustered granule cell lineage. P value (P = 0.168) was determined by log-rank test and ‘+’ indicates censored cases. o–r, Comparison of additional clinical characteristics including sex, histology, metastatic status and molecular subtype affiliation of samples from patients with group 4 (E16) (n = 17) and group 4 (E18) (n = 28). P values were determined using Fisher’s exact test.
Extended Data Fig. 10 | Cell cycle analysis of human scRNA-seq data.

a–o, Dot plots showing the normalized ratio values of G1/S against G2/M ratios within each cell annotated by cluster identity (left) for SHH (n = 2), group 3 MB (n = 2), group 4 MB (n = 4), PFA (n = 4) and C-PA (n = 3). Re-clustering t-SNE visualization of the single-cell human tumours displaying cluster annotations (middle). Re-clustering t-SNE visualization with cell cycle phase ratios (G1/S, G2/M) projections (right).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- **n/a**
- Confirmed
  - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided
    - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted
    - *Give P values as exact values whenever suitable.*
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's d, Pearson's $r$), indicating how they were calculated

*Our web collection on statistics for biologists contains articles on many of the points above.*

Software and code

Policy information about availability of computer code

| Data collection | No software was used. |
|-----------------|-----------------------|
| Data analysis   | Cell Ranger v1.2.1, R v3.4.4, Seurat v1.4.0, v2.3.0 and v2.3.4, Monocle v2.6.3, CIBERSORT (absolute mode beta). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated and analysed during the current study are available in the following repositories: Mouse developmental time points single cell RNAseq (GSE118068), PFB bulk RNAseq (EGAS00001002696, GSE64415), Human tumor single cell RNAseq/PFA/C-PA bulk RNAseq (EGAS00001003170) and MB bulk RNAseq (EGAD00001004435).
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
Sample size was determined by the availability of the human samples. The mouse data set sample size was selected in function of important events occurring during mouse cerebellar development time line.

**Data exclusions**
All of the data acquired was utilized for analysis, unless specified otherwise in the figure legends or text.

**Replication**
No replicates were performed on the single cell RNA seq mouse data set within the individual time points. Replicates were used for the human RNA seq (bulk and single cell) as described in the figure legends.

**Randomization**
NA

**Blinding**
NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑ Antibodies |
| ☑ Eukaryotic cell lines |
| ☑ Palaeontology |
| ☑ Animals and other organisms |
| ☑ Human research participants |
| ☑ Clinical data |

### Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☑ ChIP-seq |
| ☑ Flow cytometry |
| ☑ MRI-based neuroimaging |

Animals and other organisms

Policy information about studies involving animals. ARRIVE guidelines recommended for reporting animal research.

**Laboratory animals**
The study involved collection of cerebellum and hind brain regions of wild type mice (C57BL6 strain) from multiple embryonic (Embryonic day 10, 12,14,16,18) and postnatal time (Post natal day 0, 5, 7 and 14) points. Both males and females mice were used with no discrimination. Litter mates, specifically for the embryonic time points were pooled when needed.

**Wild animals**
NA

**Field-collected samples**
NA

**Ethics oversight**
AUP 21-0100H approved by The Centre for Phenogenomics (Toronto).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants.

**Population characteristics**
Patients diagnosed with cerebellar pediatric brain tumors were recruited from McGill University Health Centre and the Hospital for Sick Children/The Arthur and Sonia Labatt Brain Tumour Research Centre Biobank.

**Recruitment**
NA
Ethics oversight

REB MCH003-26 approved by McGill University Health Centre (Montreal). REB 1000055059 approved by The Arthur and Sonia Labatt Brain Tumour Research/Hospital for Sick Children (Toronto).

Note that full information on the approval of the study protocol must also be provided in the manuscript.