Multimodal Validation of the Existence of Transitional Cerebellar Progenitors in the Human Fetal Cerebellum

Zaili Luo*[1,2], Mingyang Xia*[3], Feng Zhang*[4], Dazhuan Xin[2], Rohit Rao[2], Karrie M. Kiang[2,5], Kalen Berry[2], Yu Xiong[6], Hongqi Liu[7], Yifeng Lin[7], Ming Hu[8], Mei Xin[2], Jie Ma[9], Hao Li[7], Michael D. Taylor[10,11], Wenhao Zhou[12,#], and Q. Richard Lu[2,#]

1. Division of Pediatric Hematology/Oncology, Children's Hospital of Alabama and the University of Alabama at Birmingham, Birmingham, Alabama, USA
2. Brain Tumor Center, Division of Experimental Hematology and Cancer Biology, Cancer and Blood Diseases Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
3. Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China.
4. Department of Histoembryology, Genetics and Developmental Biology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
5. Department of Surgery, School of Clinical Medicine, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong
6. Obstetrics and Gynecology Hospital of Fudan University, Shanghai 200090, China
7. Key Laboratory of Neonatal Disease, Ministry of Health, Children's Hospital of Fudan University, Shanghai 201102, China
8. Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Case Western Reserve University School of Medicine, Cleveland, OH, USA.
9. Department of Pediatric Neurosurgery, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China
10. Texas Children's Cancer and Hematology Center, Houston, TX 77030.
11. Department of Neurosurgery, Baylor College of Medicine, Houston, TX 77030.
12. Division of Neonatology and Center for Newborn Care, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

*These authors contributed equally.

[#]Coresponence: Dr. Q. Richard Lu, email: Richard.Lu@cchmc.org. Cincinnati Children's Hospital Medical Center. Dr. Wenhao Zhou, email: zhouwenhao@fudan.edu.cn, Fudan Children's Hospital.

Abstract

The developing human cerebellum comprises a series of transient progenitor states that are essential for generating diverse neural subtypes, yet the identity and validation of intermediate cell populations bridging stem-like and lineage-committed neuronal precursors remain limited. In our previous single-cell transcriptomic study, we identified a distinct transitional cerebellar progenitor (TCP) population enriched in specific progenitor zones such as the rhombic lip during human fetal cerebellar development. Here, we provide additional multimodal validations of this population. Rigorous reanalysis of our single-cell transcriptomic data, applying stringent quality control measures, validated the quality of TCP cells and their classification as a transcriptionally distinct population. Multiple orthogonal validations of TCP signature genes (*SOX11* and *HNRNPH1*) using RNAscope *in situ* hybridization, Xenium-based spatial transcriptomics, and immunohistochemistry on additional fetal cerebellar samples across different stages demonstrated the consistent presence of TCPs in the rhombic lip, transitioning from *PRTG*$^{+}$ stem-like zones in the ventricular zone at early developmental stages to the subventricular zone overlapping with *EOMES*$^{+}$ unipolar brush cell precursors at later stages. TCP-like populations were also independently identified in two fetal cerebellar single-nucleus transcriptomic atlases, and their gene signature was enriched in a cell population associated with aggressive medulloblastomas. Collectively, these multimodal validations confirm the existence of a transitional progenitor population in the human fetal cerebellum, with implications for cerebellar lineage progression and medulloblastoma origin.

Introduction

The development of the human cerebellum is a highly coordinated and complex process that involves not only the sequential and concurrent generation and specification of distinct progenitor populations, but also their spatial patterning, migration, and integration into evolving neural circuits[1,2]. Despite significant progress in understanding cerebellar development, the transitional states bridging early neuroepithelial progenitors and lineage-restricted neuronal precursors remain poorly defined. In contrast to post-mitotic neuronal cell types, which exhibit discrete and stable marker expression patterns, transitional progenitor intermediates are often characterized by graded, quantitative shifts in the relative expression of a combinatorial marker signature, reflecting their dynamic progression along a developmental continuum[3-5].

In our recent single-cell transcriptomic profiling of fresh human fetal cerebellar tissues, we identified a previously uncharacterized transitional cerebellar progenitor (TCP) population that exhibited unique transcriptional features distinct from known neural stem cells or committed neuronal lineages[6]. Immunostaining of human fetal cerebellar sections indicates that these TCP cells are spatially present in the progenitor-rich zones including the ventricular and transitional zones of the cerebellar anlage, including the rhombic lip, a critical germinal niche in the early developing human fetal cerebellum. Here, we provide additional evidence to further substantiate the existence of this progenitor population. First, we performed rigorous reanalysis of our previously published single-cell RNA-seq dataset using more stringent filtering criteria and quality control measures to reconfirm the identity and integrity of TCP cells. Moreover, our *in-silico* analyses of two independent cerebellar single-nucleus transcriptomic datasets[7,8] revealed cell populations resembling TCPs in the human fetal cerebellum. Second, we validated the presence of TCP cells using multiple orthogonal approaches, including RNAscope fluorescent *in situ* hybridization, Xenium-based spatial transcriptomics, and immunohistochemistry on fetal cerebellar tissue specimens from independent donors at different stages. These complementary approaches consistently identified TCP cells in the specific progenitor zones such as the rhombic lip, marked by TCP-signature gene expression in the ventricular zone during early fetal development and a shift to the subventricular zone at later fetal stages. Taken together, our additional analyses and orthogonal validations reinforce our previous findings identifying the TCP population as a bona fide transitional progenitor state within specific neurogenic geminal zones, such as the rhombic lip, during human fetal cerebellar development.

Re-analysis of single-cell transcriptomics validates the data quality for the TCP population

In our original study[6], by performing computational analyses of whole-cell single-cell

transcriptomics data from fresh fetal cerebellar tissues, we identified the TCP population, which is enriched in the post-conception week (PCW) 12 specimen. This analysis utilized the standard pipelines implemented in CellRanger and Seurat with commonly used filtering criteria and quality control measures as well as doublet removal. To address potential concerns regarding ambient RNA contamination or low data quality-driven cell identification, we conducted additional analyses using more stringent quality control measures. These analyses employed established pipelines CellRanger and Seurat, alongside specialized tools including DoubletFinder[9] and SoupX[10], which are designed to detect and remove doublets and ambient RNA contamination from droplet-based scRNA-seq data. These rigorous filters re-confirmed that the TCP cells consistently passed all major QC thresholds, supporting the robustness and validity of their identification.

To further assess cell quality, we carried out additional DropletQC[11] analysis to enhance the identification of empty droplets and damaged cells in single-cell RNA-seq data. We first performed a nuclear fraction analysis of the TCP cell population in the PCW12 sample using DropletQC. Kernel density estimation supported the use of a nuclear fraction cutoff of 0.1 as an appropriate threshold to distinguish low-quality cells from high-quality ones (**Fig. 1a**). The vast majority of TCP cells (95.2%) in the PCW 12 sample exceeded the nuclear fraction cutoff of 0.1 (**Fig. 1b**), confirming their quality. We therefore conclude that the TCP cell population is of high quality, meeting the rigorous filtering criteria and DropletQC quality controls, and should not be excluded from bioinformatics analysis.

To more rigorously evaluate the cell quality of the PCW 12 sample, we reanalyzed these data using widely used filtering cutoff criteria[12-16], which are supported by a machine learning framework[17], namely, removing cells with < 500 or > 5000 expressed genes and > 20% mitochondrial gene content, together with DropletQC nuclear fraction cutoff of 0.1 determined through density estimation (**Fig. 1c**). Under these filtering criteria, unsupervised clustering identified TCP cells as a distinct population located adjacent to *SOX2*+ neural stem cells (NSCs) (**Fig. 1d**). Cells in the TCP cluster have significantly higher expression of TCP signature genes than do all other cells (**Fig. 1e**). Similarly, density estimation across all human fetal cerebellar cells from PCW 8 to PCW 17 supported a consistent nuclear fraction cutoff of 0.1 as appropriate (**Fig. 1f**). Applying these stringent QC filtering criteria across all samples, we consistently detected TCPs as a distinct cell population near NSCs through unsupervised clustering (**Fig. 1g,h**), with the significant elevation of TCP signature gene expression as determined using a Wilcoxon rank sum test in our original study (**Fig. 1i,j**). Moreover, evaluating differential gene expression using additional six top-performing pseudo-bulk models of likelihood tests[18], edgeR−LRT, DESeq2−LRT, edgeR−QLF, limma−trend, DESeq2−Wald, and limma−voom, we further confirmed significant differential expression of TCP signature genes (e.g., *SOX11, HNRNPH1, CTNNB1*) in TCP cells

compared to non-TCP cell types at PCW 12 and at combined stages (**Fig. 1k**). Together, these analyses demonstrate that TCP cells possess a distinct transcriptional identity characterized by a set of significantly enriched marker genes. Notably, under the stringent filtering criteria (requiring cells to express >500 and <5000 genes and a nuclear fraction >0.1), TCP cells exhibit a higher average gene expression (1,193 genes per cell) compared to well-defined post-mitotic unipolar brush cells (UBCs) (1,080 genes per cell) and granule cells (784 genes per cell), supporting the biological relevance of the transcriptional profile observed in the TCP cells (**Fig. 1l**).

The TCP population in our dataset was most abundant at PCW12 and remained detectable at later stages, although its proportion relative to the total cerebellar cells gradually declined after PCW14. The decreased proportion of TCP cells among total cerebellar cells at later fetal stages may reflect the progressive transformation and absorption of the rhombic lip, along with the rapid expansion of other cerebellar structures. This conforms to the expected pattern for a transitional intermediate, whereas an artifact from contamination would be expected to exhibit a more random and uniform distribution over time. Debris would tend to reflect the most abundant mRNAs in the tissue, and there was no enrichment of housekeeping genes. Thus, the *in-silico* re-analysis of our single-cell transcriptomics of cerebellar samples, using more stringent QC measures and additional statistical tests, provide robust evidence re-confirming the existence of the TCP population in the human fetal cerebella. It is important to note that the assertion that results obtained using commonly applied filtering criteria and quality control measures are merely artifacts fundamentally challenges the validity of any analysis employing similar methodologies.

Multiple orthogonal experimental validations for the existence of TCP cells by RNAscope *in situ* hybridization, Xenium spatial transcriptomics, and immunohistochemistry

Validation of TCP existence by RNAscope fluorescent multiplex in situ assays. In addition to the immunohistology validation of the presence of TCP cells in the cerebellar and rhombic lip ventricular zone at early developmental stages in our previous study[6], we have further performed fluorescent mRNA *in situ* hybridization of TCP signature markers (*SOX11, HNRNPH1),* along with *Protogenin* (*PRTG*), a recently identified rhombic lip stem-like cell marker[12]*,* and a UBC progenitor cell marker *EOMES*, on the midsagittal sections of human fetal cerebella at PCW11, PCW14 and PCW17 using RNAscope, a widely recognized gold standard for validating gene expression and spatially mapping[19]. Cells labeled with *SOX11* and *HNRNPH1,* were notably enriched in the rhombic lip, especially the rhombic lip ventricular zone (RL^{VZ}) at the early developmental stage (**Fig. 2a-c**). At PCW11, cells expressing TCP signature genes are predominantly co-localized with $PRTG^{+}$ stem-like cells in the RL^{VZ}, with minimal

overlap observed with *EOMES*$^{+}$ UBCs at this early stage (**Fig. 2a**). At the later developmental stages PCW14 and PCW17, TCP cells appear to transition from co-localizing with *PRTG*$^{+}$ cells to displaying increased spatial overlap with *EOMES*$^{+}$ UBC cells in the RL^{SVZ} (**Fig. 2b, c**). These characteristics may represent the developmental state of TCP as transitional cells, that is, the transition from neural stem-like cells to differentiating neuronal progenitor cells. This expression pattern mirrors that of *PRTG* during early developmental stages near or prior to PCW11, suggesting that TCP progenitor cells may represent a downstream transitional population derived from *PRTG*$^{+}$ stem-like cells and potentially serve as a candidate cell of origin for Group 3 (G3) medulloblastoma (MB)[12]. Moreover, the positional shift of TCP cells toward the RL^{SVZ}, a primary source of UBCs, at later fetal stages aligns with our previous identification of TCP-like tumor cells in UBC-derived Group 4 MB[6], suggesting that the transitional identity of TCP cells may represent a shared developmental intermediate across distinct MB subgroups. Notably, the combined expression of TCP signature markers enables precise identification of this transitional cell population. The use of multiple markers to delineate a distinct cell type or state is a widely accepted approach, ensuring specificity within complex and dynamic developmental contexts. Importantly, the strong and regionally enriched expression of TCP signature genes in the rhombic lip supports the selective localization and existence of the TCP cell population.

Validation of the existence of TCP cells by Xenium spatial transcriptomic profiling in the human fetal rhombic lip. To independently validate the presence of TCP cell population, we performed high-resolution spatial transcriptomic analysis using the Xenium 5K platform on midsagittal sections of human fetal cerebellum at PCW 12.5, 14, and 16. Xenium is a next-generation *in situ* transcriptomics technology developed by 10x Genomics that enables multiplexed detection of thousands of genes with subcellular resolution while preserving the tissue's spatial architecture. Compared to RNAscope *in situ* hybridization, which is based on a branched DNA amplification system and double-Z probe design for single-gene detection with high sensitivity, Xenium relies on barcoded probe pairs and provides broader transcriptomic coverage to generate high-throughput gene expression maps in a spatial context.

Xenium spatial transcriptomic profiling revealed distinct spatial localization of TCP signature genes *SOX11* and *HNRNPH1* within the rhombic lip, supporting the regional enrichment of this transitional progenitor population (**Fig. 3a–c**). Co-localized expression of the UBC neuronal progenitor marker *EOMES* and the rhombic lip progenitor marker *OTX2* within the *SOX11*$^{+}$*/HNRNPH1*$^{+}$ cell population further delineated the TCP cells in the rhombic lip progenitor zone, providing additional transcriptional context for the TCP niche in the rhombic lip. These Xenium spatial mapping data corroborate our RNAscope *in situ* hybridization findings, demonstrating consistent enrichment of TCP

cells in the rhombic lip and increased spatial overlap with the rhombic lip subventricular zone (RL^{SVZ}) marked by *EOMES*$^{+}$ at later developmental stages (e.g., PCW14 and PCW16). Thus, these data further confirm the spatially and temporally restricted presence of TCPs in the human fetal rhombic lip and reinforce their identity as a transcriptionally distinct, intermediate population within the uniquely human developing cerebellar germinal zone.

Validation by immunohistochemistry in additional fetal cerebellar specimens. In our original study, we validated the presence of TCP cells marked by SOX11 and HNRNPH1 co-expression in the cerebellar and rhombic lip ventricular zone using immunostaining[6]. By analyzing additional human fetal cerebellar tissues, we re-demonstrate the presence of the TCP population marked by SOX11 and HNRNPH1 in the rhombic lip of the cerebellar samples at PCW 12 (**Fig. 4a**). The TCP population, marked by high SOX11 and HNRNPH1 expression, is enriched in the rhombic lip ventricular zone and overlaps with a subset of SOX2^{+} stem-like cells, while remaining distinct from EOMES^{+} UBCs at this stage (**Fig. 4b**), consistent with our original study.

Overall, multiple orthogonal lines of evidence, including RNAscope fluorescent multiplex in situ hybridization, Xenium spatial transcriptomics, and immunohistochemistry on additional specimens across developmental stages, reconfirm the presence of *SOX11*$^{+}$*/HNRNPH1*$^{+}$ TCP cells enriched in the rhombic lip, as reported in our original study. These findings support that this population represents a bona fide cell type rather than background noise. Together, these validation approaches using independent samples from distinct stages reaffirm the robustness of our original findings on the presence of TCP cells in specific germinal zones such as the rhombic lip during human fetal cerebellar development.

TCP-like population identified across two independent human cerebellar datasets

To determine whether the TCP population can be identified in independent cerebellar atlases, we analyzed the cerebellar single-nucleus transcriptomics dataset across cell types by Sepp et al.[8], spanning PCW 7 to PCW 20 into adulthood. Our independent analysis of exonic gene expression across cerebellar cell types within the Sepp dataset notably corroborates the presence of a TCP-like population. Since our whole-cell transcriptomic single-cell data primarily map to exonic regions derived from mature spliced mRNAs, we plotted the pseudo-bulk expression of exonic RNA transcripts based on the published website (https://apps.kaessmannlab.org/sc-cerebellum-transcriptome/). Exonic RNAs represent coding transcripts and closely correlate with protein expression, aligning well with our RNAscope mRNA *in situ* hybridization and immunohistology data. From the published Sepp dataset website, we observe that

the exonic RNA expression of TCP signature genes such as *SOX11* and *HNRNPH1* is highly enriched in the cell type classified as "VZ_neuroblast" (ventricular zone neuroblasts) at PCW11 (**Fig. 5a**), a developmental stage closely aligning with PCW12, where we identified an enrichment of TCP cells. Although the nomenclature used for "VZ-neuroblast" cell type defined in the Sepp dataset differs from that in our previous classification, their strong expression of TCP signature genes suggests that these cells are analogous to the TCP population identified in our earlier study.

Consistently, analysis of gene expression levels in different cell types classified by another independent single-nucleus RNA-seq dataset from the developing human fetal cerebellum at multiple stages, previously published by Aldinger et al. (2021)[7], also identified a population of TCP-like cells highly enriched for *SOX11* and *HNRNPH1* within the rhombic lip region such as the subventricular zone RL^{SVZ} (**Fig. 5b**). These findings are consistent with our RNAscope fluorescent multiplex *in situ* hybridization results, which demonstrate the presence of TCP cells marked by *SOX11* and *HNRNPH1* in the rhombic lip such as the RL^{SVZ} in the human fetal cerebellum (**Fig. 5c, d**).

Collectively, analyses of two independent cerebellar single-nucleus transcriptomic datasets reveal a 'VZ-neuroblast' population in the fetal cerebellum and a progenitor population in the rhombic lip subventricular zone, both of which transcriptionally resemble TCP cells, providing further support for the existence of the TCP population described in our original study.

4. Deconvolution re-analyses confirm TCP-like population in malignant MB tumors

In our original study, we used CIBERSORTx deconvolution analysis[20] using a set of the marker genes of each fetal cerebellar cell type, and showed the enrichment of the TCP signature in malignant MB tumors, especially the G3-MYC subgroup of MB[6]. We have further performed additional deconvolution analyses including MuSiC (MUlti-Subject SIngle Cell deconvolution)[21] on the bulk MB tumor dataset using default parameters and our annotated scRNA-seq fetal cerebellar cell-type object. Even under these conditions, we detected the enrichment in a TCP-like population in malignant MBs. This population was especially prominent in G3-MYC^{high} tumors (**Fig. 6a**). Consistent with previous literature[13], we also detected enrichment of GCP and UBC signatures specifically in SHH MB and G4 MB, respectively (**Fig. 6a**), validating the approach. Similarly, our analysis using the CIBERSORTx deconvolution approach[22], with all the gene sets of expressed genes in each cluster, also revealed enrichment in the TCP-like signature in malignant MB tumors including G3-MYC^{high} tumors (**Fig. 6b**). Therefore, both MuSiC and CIBERSORTx deconvolution analyses using the gene set from each cell-type cluster yielded consistent results that showed enrichment of the TCP signature in malignant MBs. It should be noted that the methods of deconvolution of bulk RNA-seq data are only predictive of

potential cell populations in tumor tissues, whereas our scRNA-seq transcriptomics analysis at single-cell resolution clearly identified the TCP-like population in individual G3 tumors (**Fig. 6c**, **d**).

In our original study, we compared the expression of TCP signature genes *HNRNPH1* and *SOX11* in MB subgroups using a normalized bulk microarray expression dataset from MBs (n = 763)[23], which approximated a normal distribution[24] (**Fig. 6e**, **f**). We therefore performed additional analyses using the actual expression values[23]. An unpaired *t*-test showed that TCP signature genes *HNRNPH1* and *SOX11* were expressed at significantly higher levels in G3 and G4 MBs, respectively, than in other subgroups in our original study. By using additional ANOVA analysis with Tukey's multiple comparisons, a rigorous statistical method for multiple group comparison, we reached the same conclusions (**Fig. 6e**, **f**). The intellectual disability gene *SOX11* is expressed in the cerebellum and medulloblastoma, and its mutations have been implicated in human neural developmental disorders[25,26]. Similarly, mutations in *HNRNPH1* are linked to a distinct syndromic neurodevelopmental syndrome[27], characterized by recurrent cerebellar structural abnormalities. *HNRNPH1* functions as an essential gene in 88% of tested cell lines according to DEPMAP[28], consistent with our original finding that it is critical for the growth of aggressive medulloblastomas, such as Group 3 MB.

Conclusion

Our re-analyses using more stringent filtering criteria and quality control measures, along with multiple orthogonal experimental validations—including RNAscope in situ hybridization, Xenium-based spatial transcriptomics, and immunohistochemistry in additional fetal cerebellar samples across developmental stages—reconfirm the presence of the TCP population in the human fetal cerebellum, particularly in regions such as the rhombic lip, as reported in our original study[6]. In addition, the independent identification of 'VZ-neuroblast' and rhombic lip subventricular progenitor populations resembling TCP cells across two cerebellar single-nucleus transcriptomic datasets[7,8] further supports the existence of the TCP population described in our original study. Our additional quality control evidence provided here shows that TCP cells were not detected due to “liberal” filtering criteria, ambient RNA contamination or empty/damaged droplets. Importantly, we provide RNAscope fluorescent multiplex *in situ* hybridization and high-resolution Xenium spatial transcriptomics for the existence of the TCP cells in specific geminal zones of the human fetal cerebellum such as the rhombic lip across different fetal developmental stages that we proposed in our original study. Thus, our findings are not an anomalous unique situation seen in a single donor. It is important to note that differences in the use of fresh versus frozen cerebellar tissues, along with variability in the anatomical regions sampled, can impact cell-type clustering and resolution, and gene expression profiles, potentially contributing to discrepancies

observed across single-cell transcriptomic studies[29-34].

Our multiple orthogonal validation approaches using RNAscope *in situ* labeling, Xenium spatial transcriptomics, and immunohistochemistry across multiple donors confirm the anatomical specificity of the TCP population—such as its localization to the rhombic lip—and support that it is not a bioinformatic artifact. Furthermore, we detect the TCP-like population in multiple malignant MB tumor tissues[6]. Thus, the consistent identification of TCP populations in both developing fetal cerebellar tissues and malignant MB samples strongly supports the biological validity of the originally described TCP population.

TCP cells in the rhombic lip may represent a transitional population that progresses and differentiates along the rhombic lip lineage trajectory, bridging early stem-like progenitors and committed neuronal precursors (**Fig. 4d**). At early fetal stages (e.g., PCW11–12), *SOX11*/*HNRNPH1*-expressing TCP cells are predominantly localized within the *PRTG*-expressing stem-like progenitor zone in the RL^{VZ}. By later fetal stages (e.g., PCW14 and PCW17), a population of these cells appears to transition from the RL^{VZ} to the RL^{SVZ}, a region marked by *EOMES*$^{+}$ UBC cells (**Fig. 4d**). This stage-dependent spatial distribution of TCP cells across distinct progenitor domains likely reflects their dynamic progression within the rhombic lip, transitioning from neural stem-like progenitors in the RL^{VZ} to differentiating neuronal lineages in the RL^{SVZ} during human fetal cerebellar development, in line with the model proposed in our original study. While *SOX11* and *HNRNPH1* may be expressed in other cerebellar cell types, our analyses indicate that these TCP signature genes exhibit specific and enriched expression in the rhombic lip region during critical developmental windows of the human fetal cerebellum. Thus, reanalysis of our single-cell transcriptomic data with more stringent quality control, combined with *in-silico* interrogation of independent fetal cerebellar single-cell datasets and additional orthogonal validations of gene expression by RNAscope and Xenium spatial transcriptomics across multiple developmental stages, collectively provides robust and convergent evidence supporting the existence of the TCP population in the developing human fetal cerebellum.

Acknowledgements:

The authors would like to thank Drs. Charles Stiles, Joseph Powell, William Weiss, Richard Gilbertson, Martine Roussel, Huda Zoghbi, William Dobyns, Alexandra Joyner, Mario Suva, Jiyang Yu, Saulnier Olivier, Tamra Ogilvie, Kenny Campell, and Yi Zheng for discussions and comments. The illustrative diagram was created using BioRender.com.

Figure 1

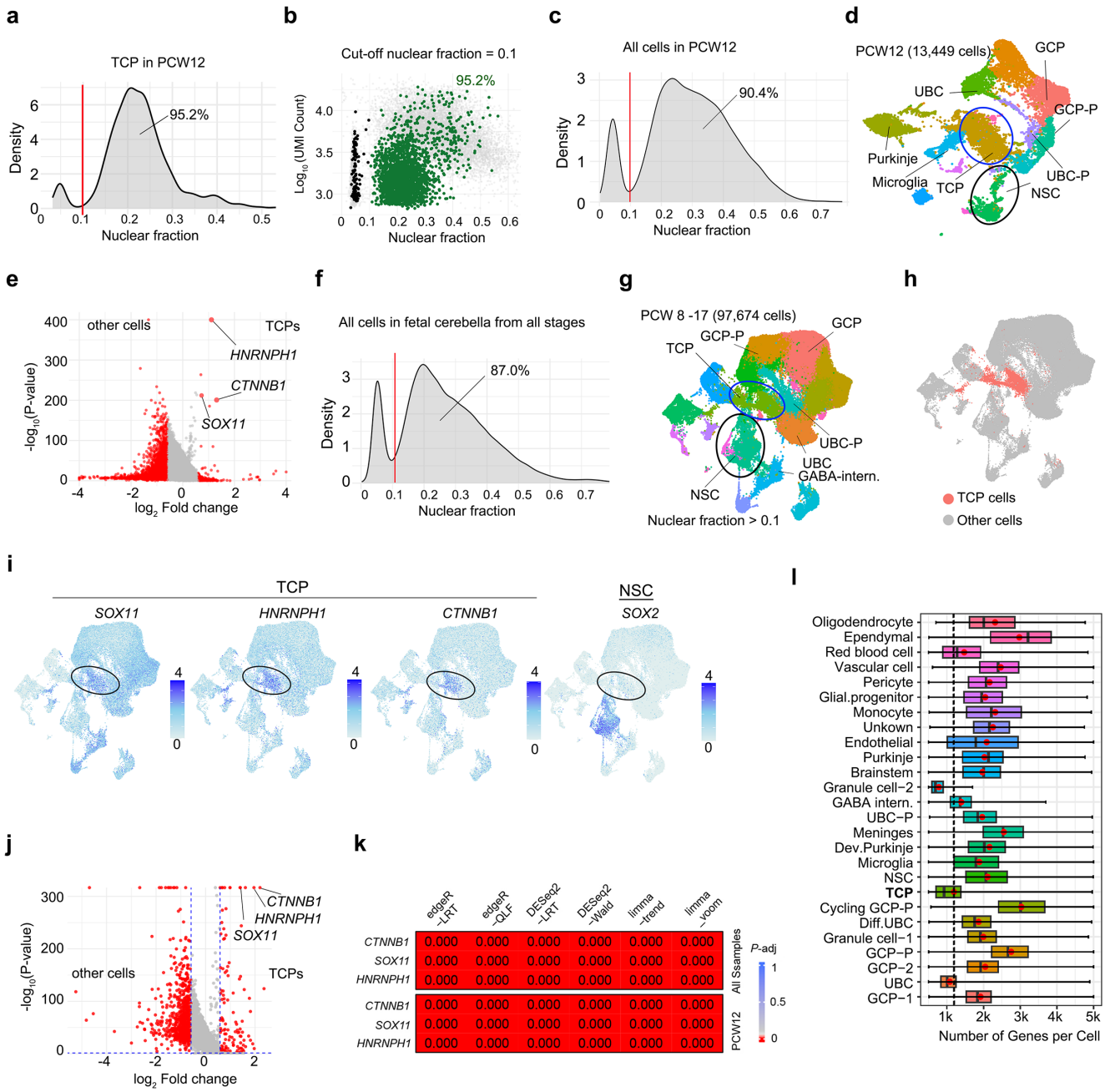

Figure 1. Additional quality controls including DropletQC re-confirm the presence of the TCP population in human fetal cerebella.

a, Kernel density estimation for TCP cells in the PCW 12 sample by DropletQC. Red line, nuclear fraction cutoff at 0.1. **b**, Scatter plots of nuclear fraction and UMI counts per cell in the TCP cells from PCW 12 using indicated cutoff scores. **c**, Density plot of nuclear fraction for all cells in the PCW 12 sample. The red line, nuclear fraction at 0.1. **d**, UMAP plots of scRNA-seq data at PCW 12 using more stringent filtering criteria (>500 and <5000 genes and nuclear fraction >0.1). Circles, NSC and TCP populations. **e**, Volcano plots of differentially expressed genes (red, fold change >1.5) in TCP cells versus non-TCP cells at PCW 12 using a Wilcoxon rank sum test. **f**, Density plot of nuclear fraction for all cells in all samples of human fetal cerebella from PCW 8 to PCW 17. The red line, nuclear fraction

0.1. **g**-**h**, UMAP plots of human fetal cerebellar cells from PCW 8 to PCW 17 with more stringent filtering criteria and nuclear fraction cut-off > 0.1 with g) all cell types labeled and h) TCP distribution shown. Circles, NSC and TCP populations. **i**, Expression of TCP signature genes and neural stemness gene *SOX2* in UMAP plot. Circles, TCP populations. **j**, Volcano plot of differentially expressed genes (fold change >2) in TCP versus non-TCP cells with significance determined using a Wilcoxon rank sum test. **k,** Significance of differential expression of the TCP signature genes determined using the indicated statistical models to analyze pseudo-bulk data of the samples from all stages and at PCW 12. **l**. Boxplots summarizing the number of genes expressed per cell across cerebellar cell types (cells with > 500 or < 5000 expressed genes; nuclear fraction cut-off > 0.1). Boxes represent the first, median, and third quartiles; whiskers extend to 1.5 x the interquartile range. Red dots indicate the mean number of expressed genes per cell for each cell type. Dash line: mean number of expressed genes per cell in TCPs.

Figure 2

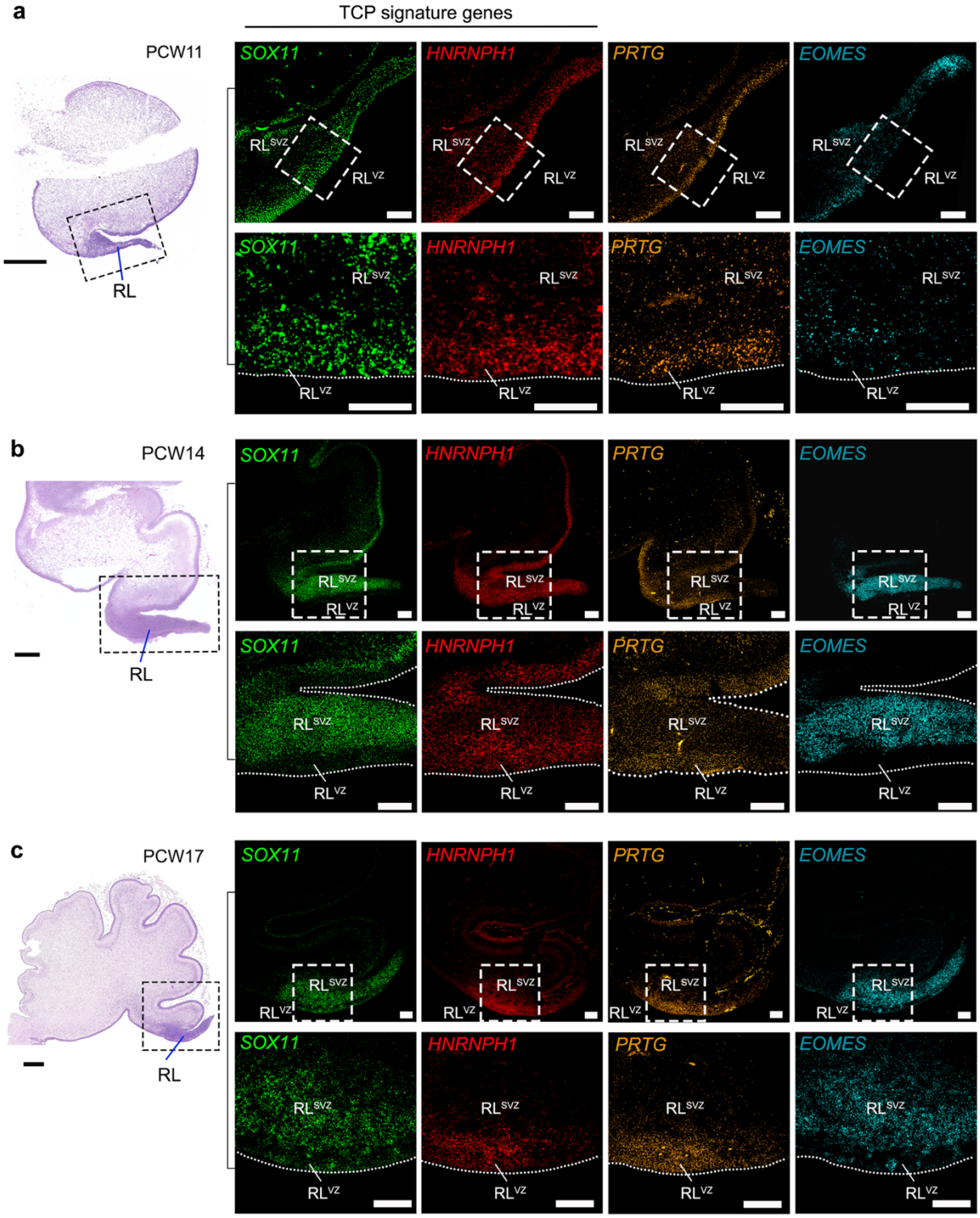

Figure 2. Orthogonal RNAscope fluorescent mRNA *in situ* hybridization validates the existence of the TCP population in the rhombic lip in the human fetal cerebellum.
a-c) Left, hematoxylin/eosin staining of midsagittal sections in the human fetal cerebellum at post-conception weeks (PCW)11 (a), PCW14 (b), and PCW17 (c). Boxed areas: the rhombic lip (RL). Scale bars, 500 µm. Right, RNAscope fluorescent mRNA labeling of TCP signature genes, *SOX11, HNRNPH1*,

and a stem-like cell marker *PRTG,* along with a UBC neuronal progenitor marker *EOMES* in the rhombic lip at PCW11, PCW14, and PCW17 as indicated. The boxed areas are shown in higher magnification in the corresponding lower panels. Scale bars, 100 µm. RNAScope probes: *SOX11* (Advanced Cell Diagnostics, Cat#443871-C2), *HNRNPH1* (Cat#1283271-C1), *PRTG* (Cat#1096421-C3), and *EOMES* (Cat# 429691-C3).

Figure 3

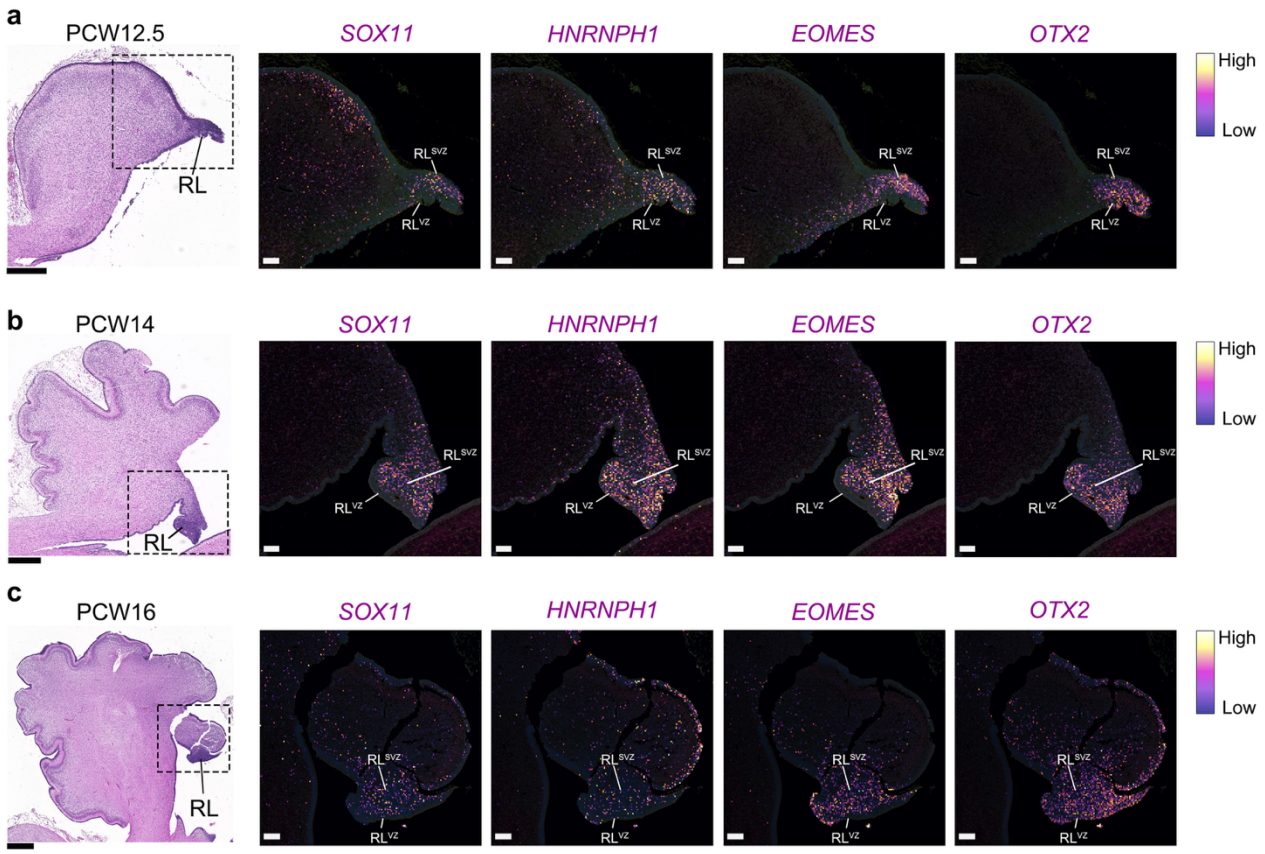

Figure 3. Xenium 5K spatial transcriptomics validates the existence of the TCP population in the rhombic lip in the human fetal cerebellum.
a-c, Left, hematoxylin/eosin (H&E) staining of midsagittal sections in the human fetal cerebellum at post-conception weeks PCW12.5 (a), PCW14 (b), and PCW16 (c). Boxed areas: the rhombic lip (RL). Scale bars, 500 µm. Right, spatial gene expression maps of TCP signature genes, *SOX11, HNRNPH1,* along with a UBC neuronal progenitor marker *EOMES* and rhombic lip progenitor marker *OTX2* in the rhombic lip at the developmental stages as indicated. Scale bars, 100 µm.

Figure 4

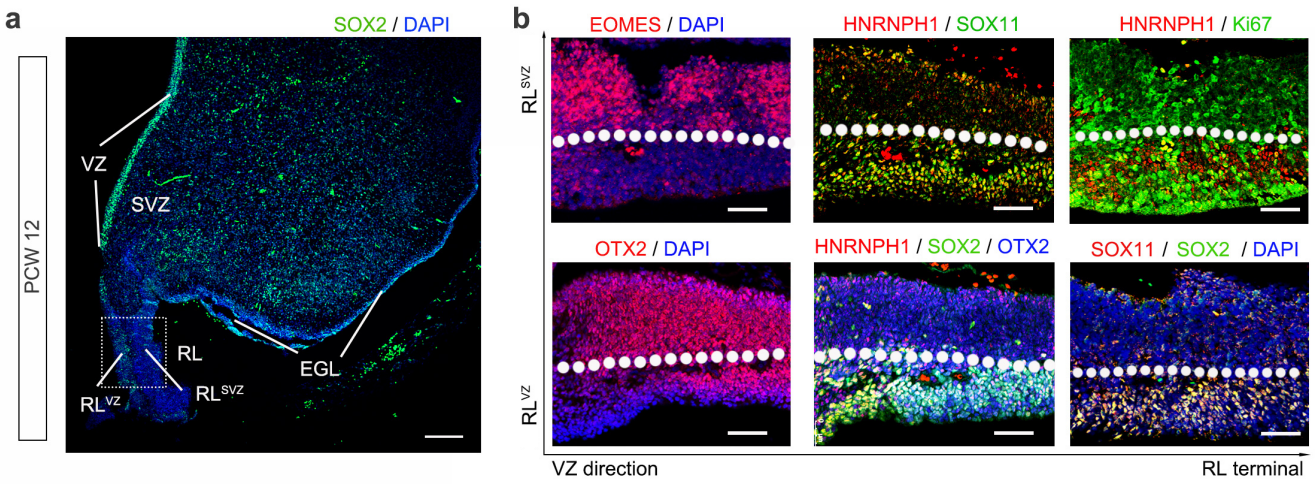

Figure 4. Immunohistochemistry of additional human fetal cerebella re-confirms the existence of the TCP population enriched in the rhombic lip.

a, **b**, Representative images of the rhombic lip (RL) region of human fetal cerebella at PCW 12 immunostained for the markers as indicated. The boxed area in panel a is shown in high magnification in panel b. The dash lines in panel b show the boundary between RL^{VZ} (ventricular zone) and RL^{SVZ} (subventricular zone). Ki67, cell proliferation marker; SOX2, a stem-cell marker; OTX2, rhombic lip marker; EOMES, UBC marker. Scale bars, 500 μm in a, 100 μm in b.

Figure 5

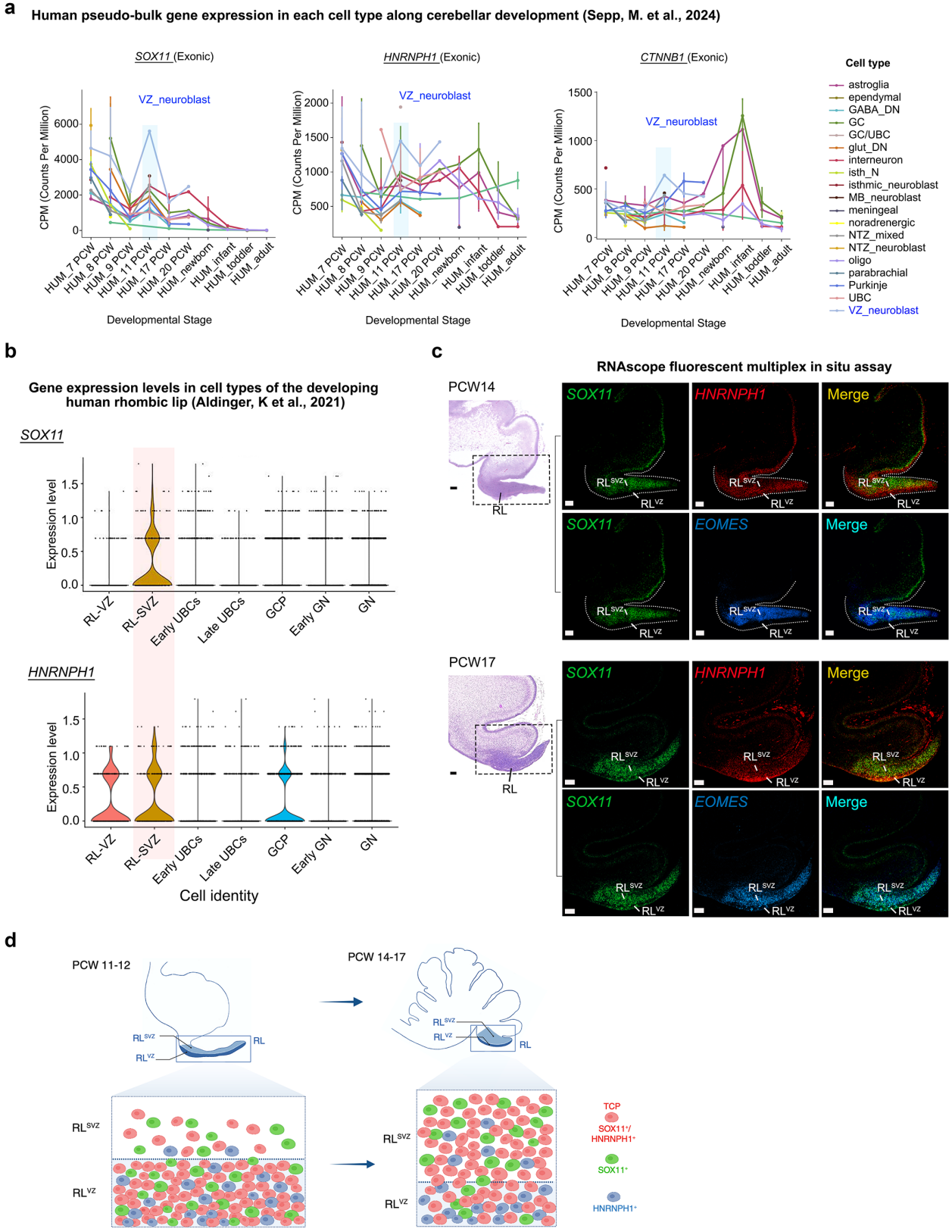

Figure 5. TCP-like cells identified across two independent human cerebellar single-nuclei transcriptomic datasets.

a. Pseudo-bulk expression of TCP signature genes *SOX11, HNRNPH1,* and *CTNNB1* across different

cell types and developmental stages defined in the Sepp study (2024)[8]. Notably, the high expression levels of these genes in VZ_neuroblast cells at PCW 11 are indicated. *SOX11* (exonic, ENSG00000176887), *HNRNPH1* (exonic, ENSG00000169045), *CTNNB1* (exonic, ENSG00000168036). **b**. Violin plots depicting the expression levels of TCP signature genes, *SOX11* and *HNRNPH1*, enriched in the rhombic lip (RL) progenitor cells such as RL-SVZ progenitors across the cell types of the developing human rhombic lip from different stages PCW 9, 10, 11, 14, 17, and 21 from the Aldinger dataset (2021)[7]. **c**, RNAscope fluorescent multiplex mRNA labeling of TCP signature genes (*SOX11, HNRNPH1*) with an SVZ-enriched UBC progenitor marker *EOMES* in the rhombic lip at PCW14 and PCW17 as indicated. Scale bars, 100 µm. **d**. Schematic diagram depicts *SOX11/HNRNPH1*-expressing TCP cells transition from the *PRTG*$^{+}$ stem-like progenitor zone in the RL^{VZ} at early stages (e.g., PCW11) to the *EOMES*$^{+}$ UBC-marked RL^{SVZ} at later stages (e.g., PCW14–17) during human fetal cerebellar development.

Figure 6

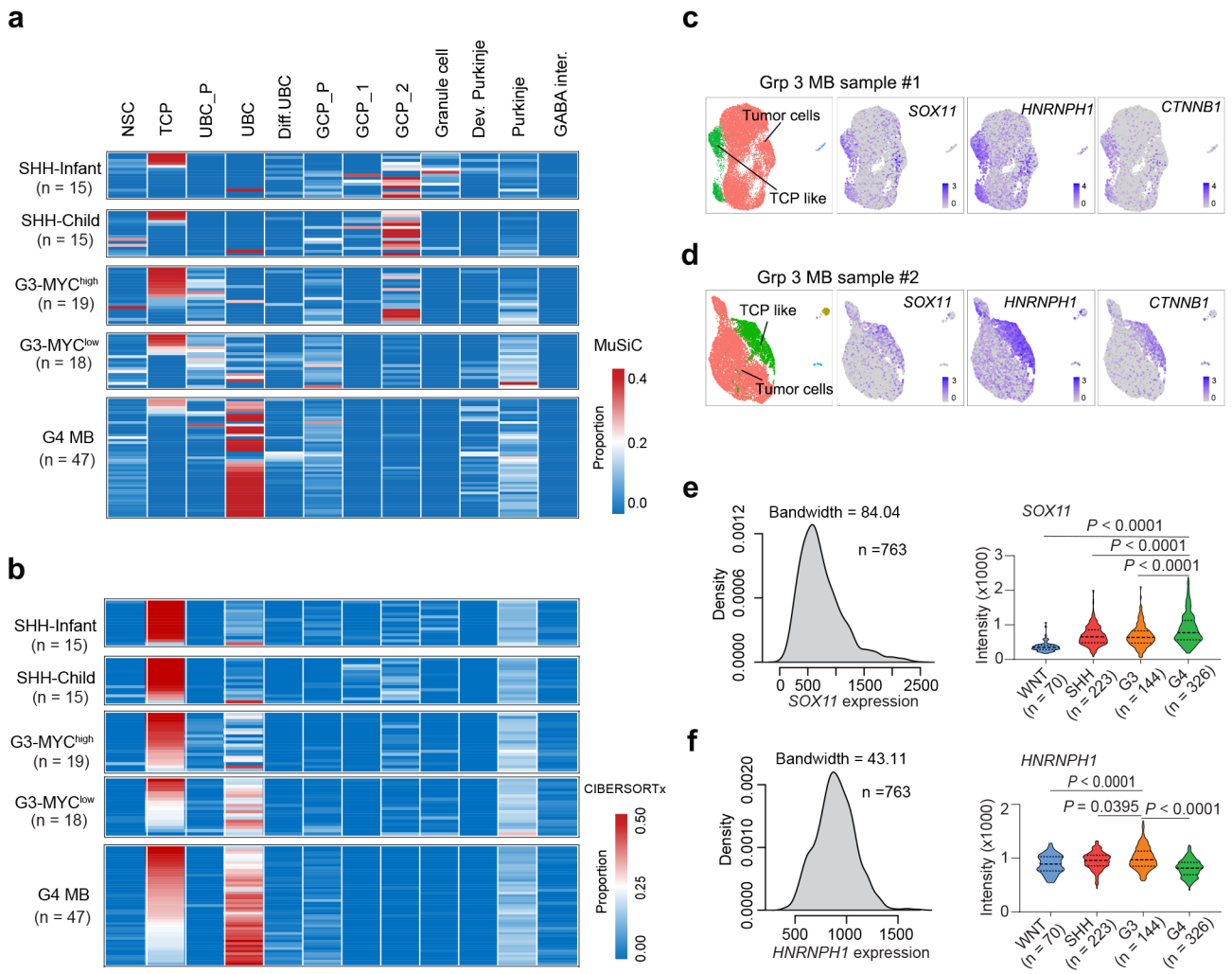

Figure 6. Additional deconvolution analyses validate the presence of TCP-like tumor cells within malignant MBs.

a-b, Heatmap showing the predicted cell-type proportions in individual MB patients using a) MuSiC or b) CIBERSORTx deconvolution. Single-cell expression matrices of all gene sets generated by the Seurat program in each cluster were used for MuSiC or CIBERSORTx deconvolution analyses. For the bulk RNA-seq matrix in MB tissues, we used 13,889 genes (FPKM >1) from the entire gene list. **c-d**, Representative UMAP of cell-type compositions (left) and expression distribution of TCP marker genes in the UMAP plot (right) for two independent G3 MB tumor samples. **e-f**, Density plots of expression of e) *HNRNPH1* and f) *SOX11* in the normalized bulk microarray dataset of 763 MB tissues[23] (left) and the violin plots of expression of these genes in tumor groups. Significance was determined by one-way ANOVA followed by Tukey's multiple comparison tests.

Methods

Human fetal cerebellar and tumor tissues

Human fetal and tumor tissues were obtained from the Children's Hospital of Fudan University and the Obstetrics and Gynecology Hospital of Fudan University. Written informed consent for the use of tissues in research was obtained from donors or the patients' parents. The collection of fetal tissues was approved by the Institutional Review Board of the Obstetrics and Gynecology Hospital of Fudan University and Guangzhou Women and Children's Medical Center.

RNAscope *in situ* hybridization and immunohistochemistry

RNAscope fluorescent in situ hybridization assays were performed using commercially available probes from Advanced Cell Diagnostics, following the manufacturer's instructions. The in situ hybridization probes used were: *SOX11* (Cat# 443871-C2), *HNRNPH1* (Cat# 1283271-C1), *PRTG* (Cat# 1096421-C3), and *EOMES* (Cat# 429691-C3). The immunostaining procedures followed the method as previously described[6]. For immunohistochemistry, we used the following primary antibodies: mouse anti-SOX2 (Santa Cruz Biotechnology, Cat# sc-365964), rabbit anti-SOX11 (Sigma, Cat# HPA000536; Thermo Fisher, Cat# 14-9773-82), rabbit anti-HNRNPH1 (Abcam, Cat# ab154894; Bethyl Laboratories, Cat# A300-511A), rat anti-EOMES (Invitrogen, Cat# 14-4875-52), and anti-OTX2 (Proteintech, Cat# 13497-1-AP).

Xenium-based spatial transcriptomics

Freshly collected cerebellar tissue was immediately immersed in 10% neutral buffered formalin (NBF), fixed at 4°C for 16–48 hours, and subsequently embedded in paraffin. Formalin-fixed paraffin-embedded (FFPE) sample preparation and sectioning (5 µm thickness) were performed according to the 10x Genomics User Guide (CG000578). Deparaffinization and decrosslinking were conducted following the 10x Genomics protocol (CG000580). Xenium Prime in situ analysis was performed using the Xenium Prime 5K Human Pan-Tissue and Pathways Assay Kit (10x Genomics, PN-1000671), following the experimental guidelines in the 10x Genomics User Guide (CG000760). Briefly, tissue sections were hybridized with target-specific probes at 50°C for 16–24 hours. Subsequent washing, probe ligation, and amplification steps were carried out using the Xenium Slides & Sample Prep Reagents Kit (10x Genomics, PN-1000460). Amplified transcripts were stained using the Xenium Cell Segmentation Staining Reagents (10x Genomics, PN-1000661) for cell segmentation. Prior to imaging, sections were treated with autofluorescence quenching reagents and stained with DAPI. Fluorescence detection and transcript decoding were performed on a 10x Xenium Analyzer (10x Genomics, PN-1000481). Following Xenium detection, hematoxylin and eosin (H&E) staining was carried out according to the Post-Xenium Analyzer H&E Staining User Guide (CG000613).

Human fetal cerebellar single-cell RNA-seq data processing

FASTQ files[6] were aligned using CellRanger v3.1. Additional quality control metrics for each sample were computed using the established pipelines in CellRanger and Seurat as well as such as DoubletFinder[9], SoupX[10], and DropletQC[11] with default or specified parameters.

Deconvolution of MB tumors

Estimation of cell-type proportions within public CBTTC cohort expression profiles (https://cbtn.org/) was conducted using a website-based deconvolution approach CIBERSORTx (https://cibersortx.stanford.edu/) and MuSiC (https://github.com/xuranw/MuSiC) run on default parameters. MB bulk RNA-seq expression value (PPKM) and average expression (signature gene matrix)

of each cell type from human fetal cerebellum were used as inputs in CIBERSORTx. MB bulk RNA-seq FPKM expression values and Seurat objects of neural stem cells and neuronal lineage cell types from the human fetal cerebellum were used as inputs for MuSiC.

Code availability

Custom code is available at https://github.com/Luostlac/public1 and https://github.com/jumphone/public.

References:

1 Haldipur, P. *et al.* Spatiotemporal expansion of primary progenitor zones in the developing human cerebellum. *Science* **366**, 454-460 (2019). https://doi.org:10.1126/science.aax7526

2 Ament, S. A. *et al.* A single-cell genomic atlas for maturation of the human cerebellum during early childhood. *Sci Transl Med* **15**, eade1283 (2023). https://doi.org:10.1126/scitranslmed.ade1283

3 Telley, L. *et al.* Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364** (2019). https://doi.org:10.1126/science.aav2522

4 Flames, N. *et al.* Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J Neurosci* **27**, 9682-9695 (2007). https://doi.org:10.1523/JNEUROSCI.2750-07.2007

5 Ruan, X. *et al.* Progenitor cell diversity in the developing mouse neocortex. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2018866118

6 Luo, Z. *et al.* Human fetal cerebellar cell atlas informs medulloblastoma origin and oncogenesis. *Nature* **612**, 787-794 (2022). https://doi.org:10.1038/s41586-022-05487-2

7 Aldinger, K. A. *et al.* Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci* **24**, 1163-1175 (2021). https://doi.org:10.1038/s41593-021-00872-y

8 Sepp, M. *et al.* Cellular development and evolution of the mammalian cerebellum. *Nature* **625**, 788-796 (2024). https://doi.org:10.1038/s41586-023-06884-x

9 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019). https://doi.org:10.1016/j.cels.2019.03.003

10 Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9** (2020). https://doi.org:10.1093/gigascience/giaa151

11 Muskovic, W. & Powell, J. E. DropletQC: improved identification of empty droplets and damaged cells in single-cell RNA-seq data. *Genome Biol* **22**, 329 (2021). https://doi.org:10.1186/s13059-021-02547-0

12 Visvanathan, A. *et al.* Early rhombic lip Protogenin(+ve) stem cells in a human-specific neurovascular niche initiate and maintain group 3 medulloblastoma. *Cell* **187**, 4733-4750 e4726 (2024). https://doi.org:10.1016/j.cell.2024.06.011

13 Vladoiu, M. C. *et al.* Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67-73 (2019). https://doi.org:10.1038/s41586-019-1158-7

14 Patmore, D. M. *et al.* DDX3X Suppresses the Susceptibility of Hindbrain Lineages to Medulloblastoma. *Dev Cell* **54**, 455-470 e455 (2020). https://doi.org:10.1016/j.devcel.2020.05.027

15 Hovestadt, V. *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74-79 (2019). https://doi.org:10.1038/s41586-019-1434-6

16 Heimberg, G. *et al.* A cell atlas foundation model for scalable search of similar human cells. *Nature* **638**, 1085-1094 (2025). https://doi.org:10.1038/s41586-024-08411-y

17 Bishara, I., Chen, J., Griffiths, J. I., Bild, A. H. & Nath, A. A machine learning framework for scRNA-seq UMI threshold optimization and accurate classification of cell types. *Front Genet* **13**, 982019 (2022). https://doi.org:10.3389/fgene.2022.982019

18 Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat Commun* **12**, 5692 (2021). https://doi.org:10.1038/s41467-021-25960-2

19 Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* **14**, 22-29 (2012). https://doi.org:10.1016/j.jmoldx.2011.08.002

20 Steen, C. B., Liu, C. L., Alizadeh, A. A. & Newman, A. M. Profiling Cell Type Abundance and

Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol* **2117**, 135-157 (2020). https://doi.org:10.1007/978-1-0716-0301-7_7

21 Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10**, 380 (2019). https://doi.org:10.1038/s41467-018-08023-x

22 Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019). https://doi.org:10.1038/s41587-019-0114-2

23 Cavalli, F. M. G. *et al.* Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **31**, 737-754 e736 (2017). https://doi.org:10.1016/j.ccell.2017.05.005

24 Ong, L. D. & LeClare, P. C. The Kolmogorov-Smirnov test for the log-normality of sample cumulative frequency distributions. *Health Phys* **14**, 376 (1968).

25 Lee, C. J., Appleby, V. J., Orme, A. T., Chan, W. I. & Scotting, P. J. Differential expression of SOX4 and SOX11 in medulloblastoma. *J Neurooncol* **57**, 201-214 (2002). https://doi.org:10.1023/a:1015773818302

26 Hempel, A. *et al.* Deletions and de novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. *J Med Genet* **53**, 152-162 (2016). https://doi.org:10.1136/jmedgenet-2015-103393

27 Reichert, S. C. *et al.* HNRNPH1-related syndromic intellectual disability: Seven additional cases suggestive of a distinct syndromic neurodevelopmental syndrome. *Clin Genet* **98**, 91-98 (2020). https://doi.org:10.1111/cge.13765

28 Sannigrahi, M. K. *et al.* A novel pipeline for prioritizing cancer type-specific therapeutic vulnerabilities using DepMap identifies PAK2 as a target in head and neck squamous cell carcinomas. *Mol Oncol* **18**, 336-349 (2024). https://doi.org:10.1002/1878-0261.13558

29 Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802 (2020). https://doi.org:10.1038/s41591-020-0844-1

30 Fischer, J. & Ayers, T. Single nucleus RNA-sequencing: how it's done, applications and limitations. *Emerg Top Life Sci* (2021). https://doi.org:10.1042/ETLS20210074

31 Oh, J. M. *et al.* Comparison of cell type distribution between single-cell and single-nucleus RNA sequencing: enrichment of adherent cell types in single-nucleus RNA sequencing. *Exp Mol Med* **54**, 2128-2134 (2022). https://doi.org:10.1038/s12276-022-00892-z

32 Danielsson, F., James, T., Gomez-Cabrero, D. & Huss, M. Assessing the consistency of public human tissue RNA-seq data sets. *Brief Bioinform* **16**, 941-949 (2015). https://doi.org:10.1093/bib/bbv017

33 McCall, M. N., Illei, P. B. & Halushka, M. K. Complex Sources of Variation in Tissue Expression Data: Analysis of the GTEx Lung Transcriptome. *Am J Hum Genet* **99**, 624-635 (2016). https://doi.org:10.1016/j.ajhg.2016.07.007

34 Nieuwenhuis, T. O. *et al.* Patterns of Unwanted Biological and Technical Expression Variation Among 49 Human Tissues. *Lab Invest* **104**, 102069 (2024). https://doi.org:10.1016/j.labinv.2024.102069