

Research Report
Laidlaw Scholars Leadership and Research Program

Mapping the Mouse Brain Before It Becomes One: The Spatiotemporal Logic of Neural Development

Baran Bahari
Bachelor of Science in Microengineering, EPFL

Supervised by Yingjian Li, Alessandro Valente, and Gioele La Manno

La Manno Laboratory of Brain Development
and Biological Data Science

September 2026

Abstract

The embryonic mouse brain develops as a neural tube divided along its length into 19 segments called neuromeres. Within each segment, stem cells differentiate from radial glia into neuroblasts, into neurons. However, Which specific population produces which remains largely unknown, because measuring the genes a cell expresses destroys the tissue so no cell can be followed through time and every developmental stage must be read from a different embryos.

This research focused on developing a pipeline to reconstruct lineage relationships from data collected at different developmental stages. More than four million cells from 14 embryos were analyzed across six stages, from embryonic day 9.5 to day 15.5, each registered into the reference atlas of its own stage so that cells from different embryos share one anatomical frame. Common candidate parent-child relationships found by two independent methods were filtered based on different criteria.

As a result, the pipeline returns 17 complete lineage trees, of which eight are new and seven match an earlier doctoral study in the same laboratory. It also recovers all 11 individual parent-child relationships established in the published literature. Alongside this, distinct differentiation profiles across brain regions revealed the spatiotemporal trajectory of neural development, which was further supported by anatomical and functional evidence.

1 Introduction

The embryonic mouse brain develops over a pregnancy of 19 to 21 days. The brain begins as a developing neural tube, divided along its length into 19 segments called *neuromeres*, arranged sequentially along the anterior–posterior (rostral–caudal) axis, each giving rise to distinct brain regions.

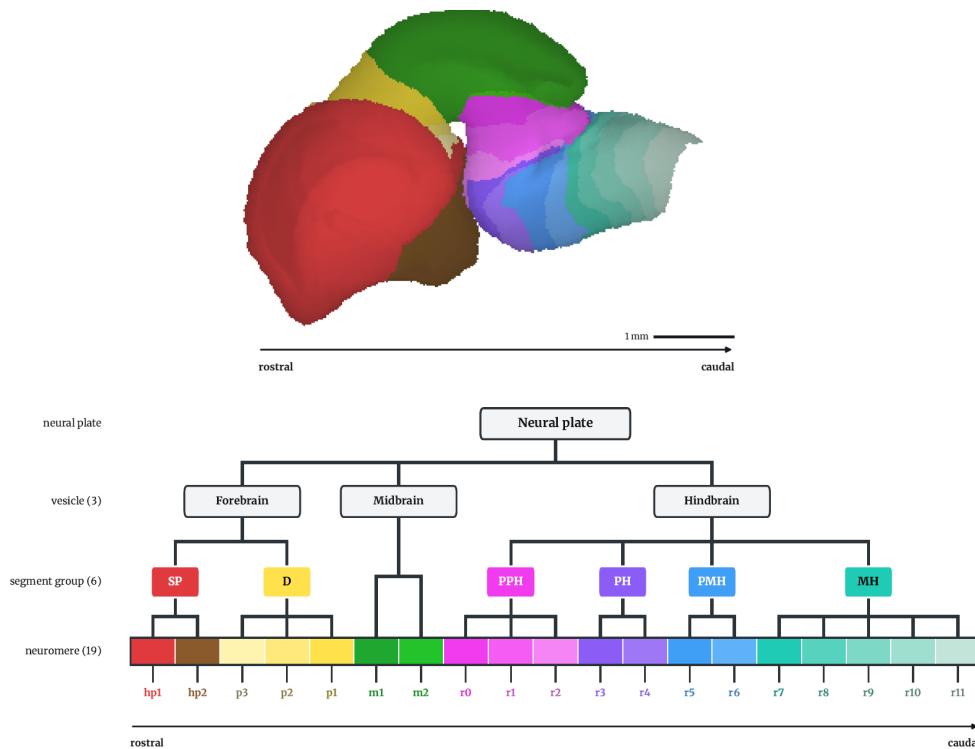


Figure 1. The mouse brain anatomical hierarchy at day 13.5 (embryonic day E13.5) by neuromere : Above, the brain coloured by neuromere, rostral at the left. Below, the same regions as a tree: neural plate, then forebrain, midbrain and hindbrain, then segment groups, then the 19 neuromeres from hp1 to r11.

Within each neuromere, cells evolve through the same three-stage developmental sequence, forming a lineage from progenitors to neurons. *Radial glia*, the stem cells of the neural tube, divide both to self-renew and to produce other cell types. They give rise to *neuroblasts*, partially differentiated but still immature cells that can either generate additional neuroblasts or further differentiate into *neurons*, the mature cells responsible for brain activity.

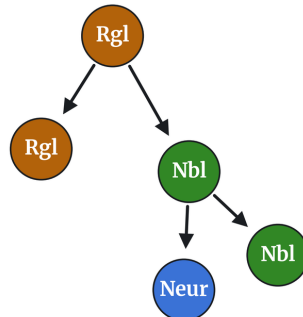


Figure 2. The lineage

That lineage has long been established. What remains unclear is not that radial glia become neurons, but which radial glial population gives rise to which neuronal population. The developing mouse brain holds dozens of transcriptionally distinct subcategories of each class, and the lineage tree connecting them remain largely unknown. As a matter of fact, reading which genes a cell expresses requires chemically processing the tissue, which destroys it, so each embryo can be analyzed exactly once. A radial glial cell cannot be labelled at day 10 and revisited at day 13, because the embryo examined at day 10 would have been destroyed and can no longer develop to reach day 13. Lineage must therefore be deduced by matching populations across samples from different individuals, which is the challenge this project aims to solve.

A small number of these developmental links have been established independently using experimental lineage-tracing approaches and can therefore serve as validation references. A method that cannot recover known links does not provide reliable new ones.

2 Research Design

The aim of this research was to develop a pipeline that takes data from embryonic brain tissue and returns lineage developmental trees, finding for a given radial glial population which neuroblasts it produces and which neurons those become. As each stage comes from different embryos, it has to be worked out from separate samples .

To achieve this aim, the project was divided in two major parts. First, I needed a database in which cells from different embryos could be compared. Second, I needed a way to propose candidate relationships and then, remove those that were not sufficiently reliable. While proposing candidate relationships can be done easily using established methodologies, the main challenge was to develop a filtering system that evaluates and removes candidates based on several independent criteria.

2.1 Building the database

The dataset is based on brain tissue images generated using HybISS, or hybridization-based in situ sequencing. This method uses fluorescent probes that bind to selected transcripts, which are then identified through successive rounds of imaging. This preserves the positions of each transcript,

allowing to map both the gene expression and the location within the tissue of each cell. Instead of removing cells from the tissue, HybISS detects RNA molecules directly within an intact brain section.

However, embryos differ in size, orientation, and how the tissue was cut. ABBA solves this by aligning each section to a reference atlas, through the four stages shown in Figure 3: the raw HybISS section from a specific timepoint (a); determining the x and y orientation of the cutting plane and finding the best-matching atlas section for each section (b); placing anchor points and either stretching or moving the section to match the atlas borders of each brain region (c); and, once the section has been aligned within these borders, every RNA molecule is now assigned to a named region (d).

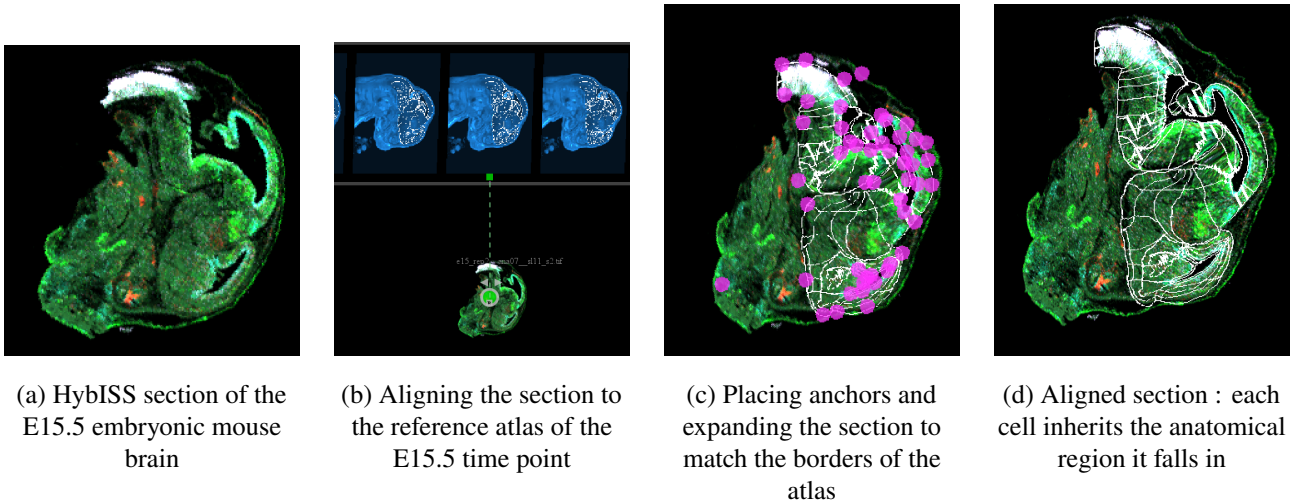


Figure 3. Aligning one section with ABBA

Each section is aligned to the atlas of its own timepoint. As the anatomy of the brain changes and new regions appear over time, each timepoint has its own atlas to which the sections can be aligned. Once all the sections of an embryo at a given timepoint are aligned, they are exported as a TIFF file, from which the coordinates of each RNA molecule can be read. As part of this project, 121 aligned sections from 14 embryos were used and analyzed.

2.2 The pipeline’s five steps

Several orders of steps were tried, and several methods at each one. They were judged on how much structure survived the filtering and how well what survived agreed with the published literature. The version below is the one that came out best.

1. **Aligning the sections and creating the dataset** After alignment, each section is a calibrated image in which cells are still represented as individual RNA reads. Pointillhist is a graph neural network developed in the lab. It assigns these cells to known populations. Because HybISS measures only about 200 genes and embryonic cells are small, individual cells often contain too little information to be identified on their own. Pointillhist therefore combines information at three scales: transcripts, cells, and local neighbourhoods, allowing each cell to be identified from both its own reads and the cells around it. The possible populations come from a single-cell RNA-seq reference atlas, and assignments are evaluated by how well they reproduce the observed gene expression while remaining coherent with known anatomy. The result is a table in which each cell is assigned one of about 300 population labels.
2. **Building and matching candidate transitions:** To create a first version of the database of parent–child candidates, we used two independent methods developed in previous studies. The

requirement for selecting these two methods was that they should rely on different processes and inputs to analyze the data and identify candidate transitions. Using independent methodologies means that when they agree on a candidate, that candidate has passed two separate and independent tests. This helps reduce errors and makes the results more reliable. Other methods were also reviewed as potential third confirmation but most of the other methods developed in previous studies shared part of their process or input with one or both of the two selected methods. These two are the most independent ones, providing the strongest confirmation when they agreed on a candidate transition.

The first method is *Optimal transport* that uses consecutive stages. It treats each stage as an amount of cell mass spread across its populations and sets a cost for moving mass from any population to any other. In other words, the cost is cheap between similar populations and expensive between dissimilar ones. Then, the cheapest overall way of moving the earlier stage onto the later one is selected. The idea is that cells do not change arbitrarily between stages, so the least costly transition is the most coherent one. The result is a table of flows, where pairs receiving a large flow become candidate transitions. The first matrix gives the cost of linking each pair of populations, based on the distance between their gene expression similarity. The second matrix gives the flow assigned to each pair by the optimal transport solution, which aims to minimize the total cost. Pairs receiving a large flow are therefore considered candidate transitions. Figure 4 shows these two matrices: the cost of linking each pair and the resulting flow between pairs.



Figure 4. How optimal transport chooses candidate: An example with five populations per stage. The flows in the second matrix are the resulting optimal transport solution based on the first.

The second method, *RNA velocity with CellRank*, works within one stage, using a completely different process from optimal transport. When a gene is transcribed, newly produced RNA first contains introns and is therefore classified as *unspliced*; after the introns are removed, it becomes *spliced* RNA. The amounts of unspliced and spliced RNA provide information about recent changes in gene expression: an excess of unspliced RNA suggests that the gene is becoming more active, while a decrease in its production suggests that its expression is being reduced. Measuring that ratio across all genes gives each cell a direction of transition, and CellRank turns those directions into probabilities that one population transforms into another. Figure 5 shows the three steps: the timing that makes it work, how a single gene is read, and how per-cell arrows become statements about populations.

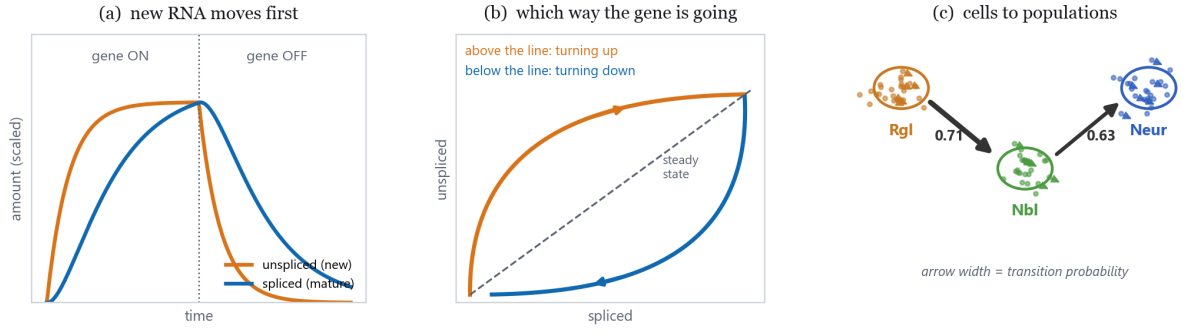


Figure 5. How RNA velocity and CellRank propose candidates: (a) When a gene becomes active, unspliced RNA increases first; when it becomes less active, unspliced RNA decreases first. (b) Comparing unspliced and spliced RNA shows whether a cell is increasing or decreasing the gene’s activity. (c) These changes are then combined to estimate transition probabilities between cell populations.

While Optimal transport uses population composition across stages, RNA velocity uses splicing within one stage and never compares two stages. Nothing makes them fail in the same way, so a transition found by both has passed two unrelated tests.

3. **Filtering, keeping only coherent transitions:** The candidates cannot be used as they are, for a simple reason: both methods always return an answer. Optimal transport will always find some cheapest way of moving one stage onto the next, whether or not a real relationship exists, and RNA velocity will always give every cell some direction, even when the signal is mostly noise. Neither has a way of saying no reliable transition was found. Furthermore, with hundreds of populations there are tens of thousands of possible pairs, so even a one-in-a-thousand coincidence will turn up dozens of times. The 4,360 candidates are a starting list for filtering, not a result. Therefore, the transitions are scored statistically. The first formula measures how strong a transition is:

$$E_{ij} = \frac{f_{ij}}{c_j}$$

- i and j are a population at one stage and a population at the next
- f_{ij} is the proportion of i ’s outgoing flow that goes to j
- c_j is the share of everything arriving at that stage that lands on j , in other words how popular j is as a destination
- $E = 1$ is chance level, and $E = 3$ means three times the expected flow

A big population receives cells from everywhere just because it is big, so counting raw arrivals would make the largest populations look like everyone’s children. Dividing by c_j asks instead whether i prefers j more than an average source would.

The second formula asks whether a transition like that could have come up by chance:

$$p_{ij} = \frac{1 + \#\{b : E_{ij}^* \geq E_{ij}\}}{B + 1}$$

- E_{ij}^* is the enrichment expected by chance, calculated after randomly reassigning the connections between parent and child cells while keeping the population sizes and total number of cells unchanged

- $B = 200$ is the number of randomised copies generated for each case
- $\#\{\cdot\}$ counts how many of those copies match or beat the real enrichment
- The added 1 is the result at $1/201$, as 200 draws cannot support a p-value of zero

Finally, transitions with less than 2% of a parent's weight are not considered, because a relationship can be statistically reliable and still be too small to mean anything, and the surviving p-values are corrected with the Benjamini–Hochberg procedure to hold the false discovery rate below 0.05. That means no more than one in twenty of the relationships kept is expected to be a false positive.

4. Checking and testing biological coherence:

A transition can be statistically strong but still make no biological sense. This step checks whether the transitions as a whole are coherent with the embryonic development. Each of four properties was compared with a random set of transitions where connections are reassigned, while each population keeps the same number of connections. This tests whether the observed patterns are stronger than expected from the lineage relationships alone.

- **Do parent and child share a neuromere?** Cells rarely cross between segments of the neural tube, so parent and child should belong to the same neuromere. They should also be close along the rostrocaudal axis, rather than being separated by a large distance.
- **Is the parent nearer the ventricle?** Cells move away from the ventricle as they mature, so the parent should be closer to the ventricle than the child. As a matter of fact you can see in this the Figure 6, the spatial distribution of cells while they mature. This also follows the normal lineage order from radial glia to neuroblasts to neurons.

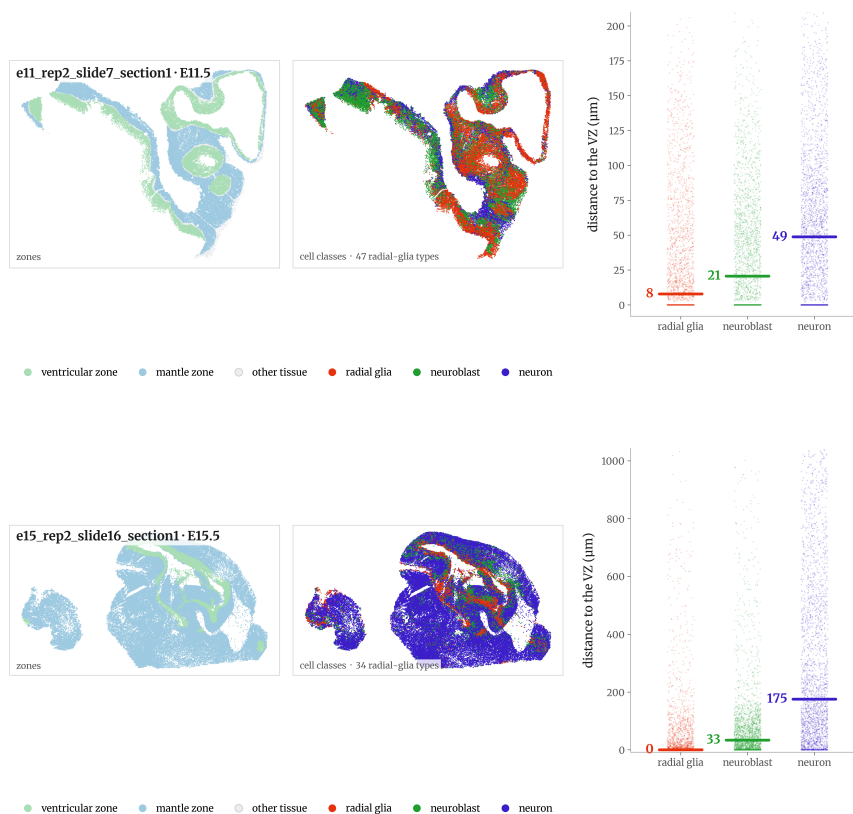


Figure 6. Lineages layer outward from the ventricle: Distance from the ventricular zone, by class, for two lineages.

- **Does the transition follow the developmental order?** A valid transition should connect consecutive stages and follow the order radial glia → neuroblast → neuron, without going backwards or skipping directly from radial glia to neuron. The child should also be less progenitor-like than the parent.
- **Does the parent peak no later than the child?** A parent population should increase and peak before its child. If the parent is still increasing after the child has peaked, the relationship is biologically unlikely.

5. **Validation against existing work.** Finally, what survived was compared against two outside sources: published research from other laboratories, which supplies the established relationships of mouse neural development, and the doctoral thesis of Irina Khven, a PhD student in this same laboratory who built her own lineage trees from the same tissue using different methods. Neither was used anywhere earlier in the pipeline.

3 Analysis and Findings

3.1 The final dataset

The aligned database has more than four million cells from 14 embryos and 121 sections, spread unevenly across the six stages. Cells were sorted into 60 subclasses from radial glia through neuroblasts to neurons, and 2.8% could not be correctly identified.

Figure 7 shows three of them, each cell drawn where the atlas places it and coloured by subclass along the scale underneath, warm for radial glia through green for neuroblasts to blue for neurons. At day 9.5 the brain is almost entirely orange and yellow. By day 12.5 the picture has split, the rostral end still warm while the caudal end has turned blue. By day 15.5 almost everything is blue or violet. Reading across gives the conversion of a progenitor sheet into a neuronal organ; reading within a panel gives the second result, that the change does not arrive everywhere at once.

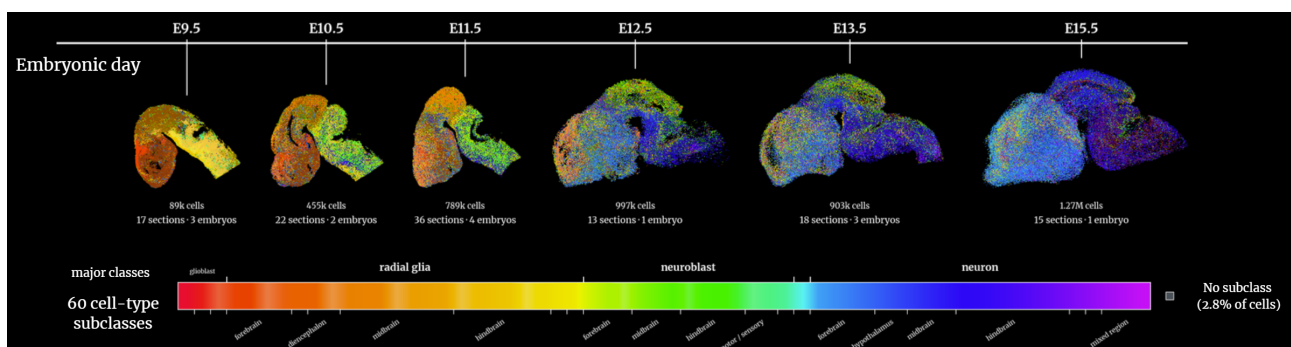


Figure 7. Cell-type distribution across the 6 developmental stages, obtained using Napari

3.2 The logic of neural development: a caudal-to-rostral wave

All segments follow the same lineage order, but at different rates. At each stage, the segment with the highest proportion of neurons is the most advanced one temporally, revealing the order of development that can be justified based on the spatial location of the neuromere or its functional or anatomical importance.

- **Day 9.5** Almost all cells are still radial glia. The leading segment (hp2) has reached only 6% neurons, and the segments barely differ from one another.

- **Day 10.5** The first leader appears at the caudal end, where r9 reaches 31% neurons. Its position explains why: it sits next to the developing spine and musculature and picks up early maturation signals from them, so what happens here is not decided inside the brain alone.
- **Day 11.5** Segment r5 reaches 33%. It makes neurons for eye movement, a function needed early in life, so finishing early makes sense.
- **Day 12.5** The hindbrain still leads, with r10 at the junction with the spinal cord furthest along at about 76% neurons.
- **Day 13.5** The leader is now p3, at 79%. It started late for a reason that can be named: the signalling centre that instructs it only switches on around day 10.5, two days after the equivalent centres further back. The delay is in the instructions, not the cells.
- **Day 15.5** The rostral end finishes last. Segment hp2 reaches 89% while its neighbour hp1 lags behind, which is informative rather than odd: hp1 contains the telencephalon, the future cerebral cortex, and is still holding on to a progenitor pool long after other regions have run theirs down.

Together these describe the order of maturation moving from the caudal side of the brain towards the rostral side over about six days.

3.3 Lineage trees returned by the pipeline

Optimal transport proposed 1,687 parent-child connections and RNA velocity with CellRank proposed 3,530 so only 857 were found by both.

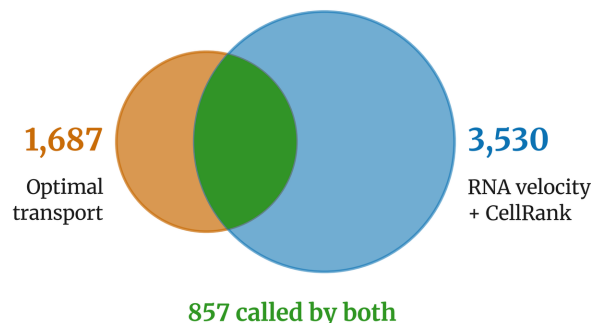


Figure 8. Relationships proposed by each method

The challenging part was finding reliable direction. Of the 4,360 candidate relationships, only 60 could be assigned a clear direction: from one population to another. In many cases, the data showed that two populations were related, but did not provide enough evidence to tell which one came first. This is because a connection can be detected even when the differences between the two populations are too small or too noisy to determine which is the parent and which is the child. Therefore, the main limitation of the results is not finding possible relationships, but finding their direction.

Filtering and checking brought the candidates down to 17 complete trees, each with a radial glial origin, the neuroblasts it produces and the neurons those become. Eight of them are new, and of the rest, seven match trees in Khven's thesis.

The 17 are whole trees, each assembled from eight to twelve individual parent-child relationships. The pipeline also recovered all 11 relationships established in the published literature, but those are single links sitting inside trees.

The eight new trees are situated in four neuromeres, and Figures 9 and 10 show them: the circles mark where each lineage is concentrated. Two are in hp1 at the rostral part and one in p2 behind it, three in the midbrain part m1, and two in r1 at the front of the hindbrain. The concentration of relationships in m1 and hp1 is mainly because we have more data from these regions.

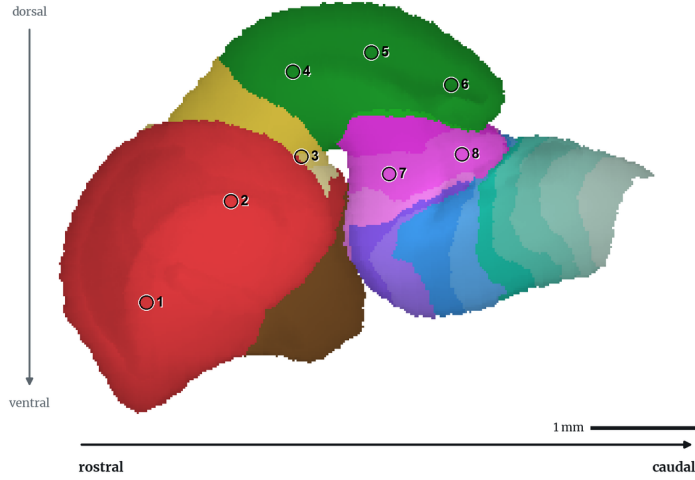


Figure 9. Where the eight new lineages are within the brain's neuromeres: Circles mark the neuromere each lineage is concentrated in

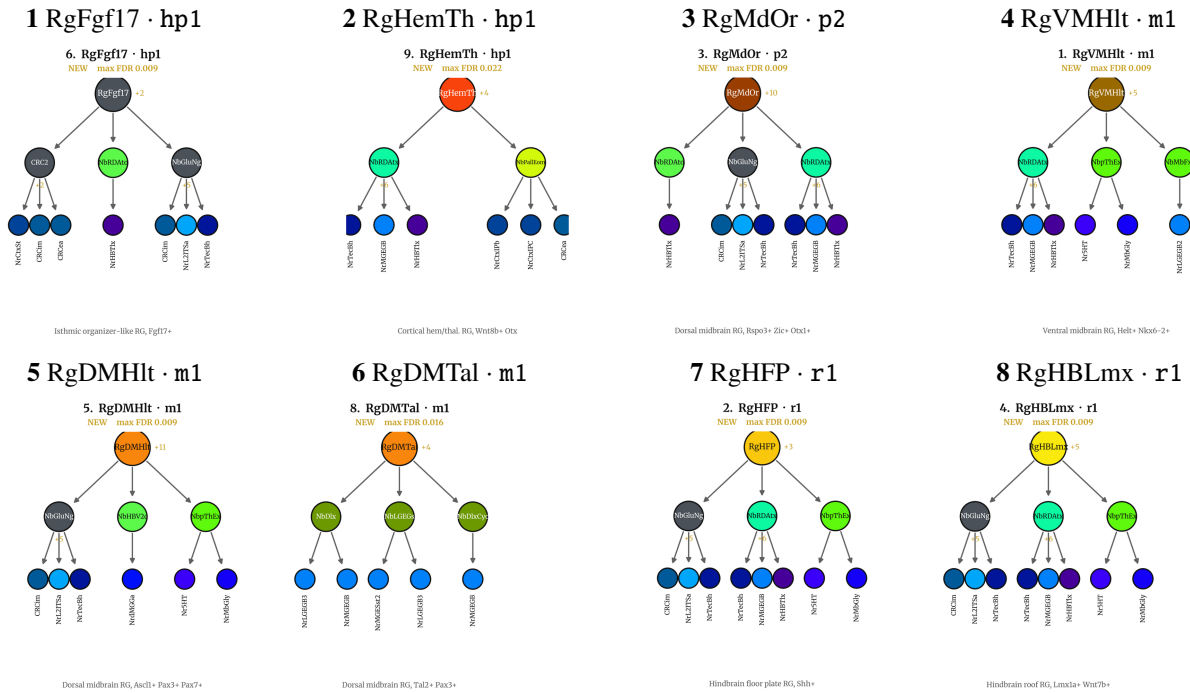


Figure 10. The eight new trees, numbered as in Figure 9: Weakest-link false discovery rate is written under each root

4 Conclusion

The result of this project a new method for reconstructing lineage relationships in the developing brain from spatial gene-expression data, tested against evidence it had no part in generating. It returns 17

complete trees, eight of them new and seven matching earlier work in the same laboratory, and recovers all 11 individual parent-child relationships established in the literature.

The analysis also established that the brain's segments differentiate on distinctly different schedules, progressing broadly from caudal to rostral across roughly six days. Those differences are not noise: they track the anatomical organisation of the neural tube, and in several cases the reason for a segment's timing can be named, from proximity to the developing spine for r9 to the prolonged construction of the cerebral cortex for hp1.

4.1 Limitations

Four limitations bound these conclusions, and each points at a specific improvement.

Too few transcripts per cell: Only about four genes are reliably detected in each cell. This is the binding constraint on everything else: populations that differ in a handful of genes cannot be told apart, so some of the 60 subclasses are certainly mixtures, and a tree rooted in a mixture is a tree rooted in an artefact. A deeper panel, or more probes per cell, would do more for the method's viability than any change to the algorithms.

Direction is missing for most pairs: For 55% of possible connections the pipeline can say a relationship exists but not which way it runs. The reason is the same shortage of transcripts: RNA velocity needs enough unspliced reads to measure a trend, and below that threshold it returns nothing usable. Until direction can be resolved for the majority of pairs, the trees are fragments rather than lineages.

No ground truth, and few benchmarks: Recovering 11 of 11 literature relationships sounds decisive, but eleven is a small number against 4,360 candidates, and those eleven are the best-studied lineages in the brain, not a random sample. A genetic barcoding experiment, in which dividing cells pass a heritable label to their progeny, would give real ancestry to test against and would turn the coherence checks from supporting evidence into a proper validation.

One embryo per stage, and no cell followed: Two of the six stages rest on a single animal, so individual variation cannot be separated from developmental change, and day 9.5 contributes only 89,000 cells. More replicates per stage would fix the first problem; nothing fixes the second, because lineage here is inferred statistically and never observed. Every parent-child link is drawn from matching snapshots of different animals. The trees are the most probable account of the evidence, not a record of events witnessed, and that distinction should survive any summary of this work.

4.2 Outlook

Beyond the fixes above, three directions follow. The first is to bring position into the lineage question, asking how a cell's neuromere and its location within it shape development. The second is to separate proliferation from differentiation, which this analysis treats together. The third is application: if the normal schedule can be mapped with confidence, departures from it become visible, and lineage trees turn into a tool for identifying where and when development goes wrong in neurodevelopmental disorders.

5 Resources

References

1. Khven, I. (2025) *Fine-grained mapping of neural progenitor subtypes in the embryonic brain*. PhD thesis, EPFL.
2. Klein, D. *et al.* (2025) ‘Mapping cells through time and space with moscot’, *Nature*, 638, pp. 1065–1075.
3. La Manno, G. *et al.* (2018) ‘RNA velocity of single cells’, *Nature*, 560, pp. 494–498.

Methods and tools

- **HybISS**: hybridization-based in situ sequencing, used to read gene activity within intact tissue sections so that each cell retains its anatomical position.
- **ABBA**: Aligning Big Brains and Atlases, used to register each tissue section to the reference atlas corresponding to its own developmental stage.
- **moscot**: the optimal transport framework used to propose candidate population-to-population transitions between consecutive stages (reference 2).
- **RNA velocity and CellRank**: used to infer the direction of transitions from the balance of newly transcribed and processed gene copies within individual cells (reference 3).