

Spatiotemporal Clonal Architecture of the Newborn Mouse Forebrain

Corresponding Author: Dr Tomasz Nowakowski

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)

In this manuscript Yuan et al. reported the creation of a comprehensive atlas of clonal cell lineage relationships across multiple subdivisions (cortex, thalamus, hippocampus, striatum, and olfactory bulb) of the newborn mouse brain by applying lentiviral barcoding and single cell transcriptomics. The results highlighted that clonally-biased glutamatergic neurons and astrocytes were predominantly generated and retained within their region of origin, while clonally-biased GABAergic neurons and oligodendrocyte precursor cells exhibited extensive dispersion across regions. The authors reported uncovering previously unrecognized lineage relationships in the cingulate cortex and identifying substantial postnatal glutamatergic neurogenesis in multiple forebrain regions. They also reported that Sp8 and Meis2 marked subtypes of GABAergic neurons with broadest clonal dispersion across brain regions. The authors concluded that their findings would serve as a foundational reference for studies of mammalian brain development and its disorders.

This is an interesting study with the goal to create a comprehensive atlas of clonal cell lineage relationships across a newborn mouse brain. Such an atlas will be a useful tool to understand how neural progenitor cells generate the large diversity of cell types in the mammalian brain. However, I have many concerns with this study.

1. There is no justification that the numbers of cells sequenced and the numbers of multi-cell clones with barcodes recovered are enough for the authors to make conclusion about cell lineage across the multiple brain regions analyzed together. The methods involved labeling progenitors with STICR lentivirus, sorting the labeled cells using FACS, single cell RNA-seq on a 10X platform, and data analysis. Cell loss will occur at multiple steps starting from cell dissociation. The percentage of the total cells that are recovered by FACS can vary significantly, often being 50% or less. The cell capture efficiency on a 10X platform range from 50-80%. The barcodes will not be recovered for all the cells, etc. The authors should consider the numbers of progenitor cells and all these factors, and calculate how many multi-cell clones they need to make meaningful conclusion about cell lineage in this study.
2. In multiple places in the manuscript, the authors make conclusions about fate-restricted populations (for example, lines 137-138, 142-144, 154-155, 169-170 etc) or region-restricted. However, the numbers of such clones were not explicitly reported. Given the reasons in 1, these so-called fate-restricted or region-restricted clones might be due to incomplete capture of the whole lineage, instead of fate-restriction or region-restriction.
3. Most of the main conclusions are already known, including glutamatergic neurons and astrocytes were predominantly generated and retained within their region of origin, while clonally-biased GABAergic neurons and oligodendrocyte precursor cells exhibited extensive dispersion across regions.
4. Although a comprehensive atlas of cell lineages in broad brain regions will be very useful for understanding brain development, the lineage data presented in the current study do not capture a complete lineage landscape, and most of the lineages are likely incomplete, which reduces the value of this study.

Referee #2

(Remarks to the Author)

The work by Yuan and colleagues is an impressive and ambitious study that generates a large-scale clonal atlas of the newborn mouse forebrain using high-complexity lentiviral barcoding (STICR), single-cell transcriptomics, and whole-brain MERFISH mapping. Across cortex, thalamus, hippocampus, striatum, and olfactory bulb (OB), the authors accurately reconstruct lineage relationships within and across regions and developmental timepoints (E11–P0 injections; P4 collection). The work is elegant and robustly carried out. Key findings include: (i) a “born together, stay together” principle for glutamatergic neurons and astrocytes (regional retention), contrasted with (ii) broad dispersion of GABAergic neurons and OPCs (“born together, rear apart”) across regions; (iii) an unidentified lineage relationship for neurons destined to the cingulate cortex; (iv) identification of a highly dispersed population of Meis2⁺Sp8⁺ GABAergic neurons across brain regions; (v) substantial postnatal glutamatergic neurogenesis in pallial regions; and (vi) lineage biases whereby glutamatergic progenitors preferentially associate with astrocytes while GABAergic progenitors preferentially associate with OPCs.

The originality of this work is high. The integration of scalable barcoding with multi-region scRNA-seq and whole-brain spatial transcriptomics is technically impressive and conceptually powerful. The demonstration of contrasting dispersion vs retention rules across major lineages, the explicit coupling between neuronal subtypes and glial lineages, and the introduction of clonal information into spatial mapping are highly significant contributions. The identification of fate-restricted populations, the postnatal glutamatergic lineage continuity in cortex (which is generally assumed to switch entirely to gliogenesis after birth), and the identification of Meis2⁺Sp8⁺ dispersed interneurons are novel and of broad interest to developmental and systems neurobiology. The addition of spatial mapping of clonal relationships is also an original and important feature of this study.

The methodology is carefully executed, and the dataset is of remarkable scale and quality. The barcode design, the very large number of cells analyzed, and the integration with MERFISH provide a uniquely comprehensive view of forebrain development. The mapping strategy (correlation-based alignment of MERFISH and scRNA-seq clusters) is well-justified and yields interpretable results. The study sets a strong benchmark for clonal tracing at scale.

The analyses are convincing and thoughtfully presented. While clarifications on null models, confidence intervals, and statistical tests for enrichment would make the conclusions even stronger, the overall the statistical treatment appears sound.

The conclusions are exciting and largely supported by the data. I particularly appreciate how the authors combine regional and temporal information to extract general principles (“stay together” vs “disperse apart”). Some statements would benefit from clarifications in timing (e.g., P0 injections vs actual neurogenesis) and anatomical definitions (habenula vs thalamus), but these are refinements rather than challenges to the main findings.

Here we list a few points that, if addressed, could further increase clarity and impact:

1. The spatial maps of MERFISH data nicely illustrate the anatomical distribution of major cell types. To further enhance interpretability, the author should please consider adding stacked bar plots showing the relative proportions of cell types across sections or regions to complement spatial localization.
2. The dot plot in Extended Data Fig. 2b appears to show a relatively higher proportion of inhibitory (GABAergic) neurons than excitatory (glutamatergic) neurons in cortex, whereas the expected ratio is closer to ~80% excitatory and ~20% inhibitory. Please, clarify whether values reflect proportions across all cell types or are calculated independently for each cell type across regions. If the latter, indicating this explicitly would help, as relative PN/IN proportions are otherwise not directly observable.
3. The average clone size decreases with later injections and is attributed to shorter lineage-tracing duration (Extended Data Fig. 2c). Would it be more accurate to also interpret this trend as progressive clonal restriction or loss of progenitor potency over developmental time? Could the authors clarify this point would better connect the observation to underlying mechanisms.
4. The manuscript reports early generation and broad early dispersion of microglia, with late-labeled microglial clones no longer dispersed across regions. Could the authors clarify whether endothelial, vascular, and epithelial lineages were similarly evaluated for clonal dispersion, and whether comparable patterns were observed? Additionally, did you detect transient-amplifying microglia or intermediate proliferative states, and if so, how were these distinguished (markers, spatial context, clone size distributions)?
5. Regarding the clonal analysis of microglia, could the authors clarify how these cells are labeled in their system? Since microglia originate from yolk-sac progenitors rather than ventricular neuroepithelium, their labeling likely reflects infection of already colonizing cells that subsequently amplify locally within the brain. It would be helpful to make this point explicit, and to comment on whether similar considerations apply to other non-neuroepithelial lineages such as vascular or endothelial cells.
6. The section describing thalamic clonal relationships includes habenular neurons alongside ventral/dorsal thalamus. As the habenula is part of the epithalamus (within the diencephalon) rather than thalamus as strictly defined, can the authors please clarify anatomical labels and ensure consistency across text and figures (a brief note in Extended Data Fig. 6 or Methods would help).
7. The finding that postnatally born glutamatergic neurons share clones with prenatally generated glutamatergic neurons and glia suggests continuity across the neurogenesis–gliogenesis transition. While postnatal neurogenesis is well established in the hippocampus, the evidence presented here for cortical postnatal glutamatergic neurogenesis appears

novel and important, as cortical development is generally thought to shift entirely from neurogenesis to gliogenesis after birth. This seems novel and important; consider emphasizing it more explicitly in the Results and discuss it.

8. The authors report CGE-derived clusters mostly generated at E12–E14, with MGE-derived clusters also within this window. Prior literature generally places MGE onset around E12 and CGE slightly later (~E14 onward). Could the authors clarify whether your data support an earlier onset of CGE-derived neurogenesis, or whether this may reflect clustering or labeling differences?

9. Please clarify the x-axis in Fig. 3e (e.g., “clone size” or “number of cells per clone”) for the cumulative proportion plot to aid interpretation.

10. Across Extended Data Figs. 6, 8, and 10, references to clusters and clonal groups are difficult to reconcile with the figures due to dense text annotations and lack of a clear color legend. Consider adding a color annotation with legend or other graphical markers (e.g., explicit labels for triangles in half-heatmaps) to map the described groups directly onto the plots.

11. The identification of Meis2⁺Sp8⁺ dispersed interneurons is striking — could their clonal analysis be linked to embryonic single-cell datasets to infer developmental origin more precisely?

Overall assessment: This is a very strong and exciting paper that makes a major contribution to developmental neurobiology. The dataset will be of great value as a reference for the community, and the conceptual advances are important. Our comments above are intended to help strengthen the clarity and impact of the presentation — the overall quality of the work is excellent.

Referee #3

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4

(Remarks to the Author)

Yuan et al. use lentiviral barcoding together with single cell transcriptomic to study cell lineage during mouse embryonic forebrain development including thalamus, cortex, olfactory bulb and striatum. They find that clones containing glutamatergic neurons are largely restricted to specific brain subregions (thalamus, cortex, ..) while those containing GABA-ergic neurons are more spread, as previously described. They also describe biased clones containing glutamatergic and astrocytic cells, and GABA-ergic neurons and OPC, and describe what is interpreted as postnatally (P0) generated glutamatergic neurons.

This study uses frontier and state of the art technologies aimed at an interesting topic. However from the data shown it remains unclear whether this study goes beyond what is already known using more classical tools. The results shown appear to be mostly confirmatory, not only in the cortex but also in striatum and thalamus. In particular it is unclear whether the new data bring more insights than previous data reported for thalamus and striatum using more conventional technologies (see for instance Kelly Neuron 2018; Wong PLOS Biol 2028)?

What is particularly missing here is a more in depth analysis with much higher spatial resolution, which would be required to disentangle clonal relationships in subregions of interest in the described regions, in particular thalamus (dorsal / ventral - all thalamic nuclei) and basal forebrain (including subregions of striatum, GP, CP ..). Indeed these important aspects of lineage relationships remain much less known (although for thalamus see recent report of Puche et al 2025).

The lack of description of the data for other structures of the forebrain such as hypothalamus and amygdala is also crucially missing here.

Moreover the description and presentation of the data is also often too succinct (very heavy figures and very limited legend explanations), making it difficult to grasp the data and understand the logic behind the proposed interpretations.

On the other hand, the study has the ambition to become a landmark resource, it would be important to make the data more readily available online using a user-friendly interface, as usually done remarkably by the Allen Institute.

Below are more specific comments related to technical issues that have to be solved and/or clarified.

- From Fig.1 and associated suppl data it seems like there are relatively few glutamatergic neurons in the dataset, compared with GABAergic cells (there seem to be more GABA-ergic cells in thalamus and cortex for instance). This points to a bias in the cell or transcriptome collection that could mislead the interpretation of the data.

- The authors describe fate restricted clusters of upper layer neurons in the cortex that could relate to previously described clonal analyses using Cux2-Cre lineage that had remained controversial. This is potentially interesting but the data shown should be explicated more clearly: to which time-points of viral injection do they correspond? If they consist of late generated

clones the interpretation would be quite different.

- The data claiming neurogenesis after birth should be substantiated further. Since the library used is lentiviral, how can the authors conclude that the cells shown are born at P0, why not using retroviruses instead? Also, what is the identity of the cortical neurons generated?

Minor comment :

- the study is focused on forebrain, so this should be clearly explicated in the abstract and introduction, in which 'brain' is often used instead of 'forebrain'. Unless the authors add their data on midbrain, hindbrain, cerebellum, ..

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)

The authors have sufficiently addressed my questions and concerns. The revised manuscript is much improved.

Referee #2

(Remarks to the Author)

We have reviewed the authors' responses to the referees' comments and the changes made in the revised manuscript. Overall, we find that the authors have addressed our concerns thoroughly and constructively. In particular, the authors have added useful analyses of MERFISH cell-type proportions, microglial and non-neuroepithelial lineages, postnatally generated glutamatergic neurons, interneuron developmental origins, and subregion-resolved lineage relationships. They have also improved figure readability and anatomical terminology, and have made the dataset available through a user-friendly browser. Importantly, the authors now more appropriately frame their lineage conclusions as population-level clonal biases rather than complete lineage reconstructions or strict fate restrictions. The added simulations and barcode-shuffling analyses provide a reasonable framework for interpreting the robustness of these conclusions despite incomplete clone recovery, which is an inherent limitation of scRNA-seq-based lineage tracing. The revised manuscript is substantially improved in clarity, scope, and interpretation. Importantly, the value of this study should not be assessed only by asking whether each broad developmental principle is entirely unprecedented. Several aspects of forebrain lineage progression have indeed been explored previously using classical birthdating, sparse clonal analyses, or regionally restricted approaches. However, the present study provides a substantially more comprehensive and modern framework by integrating clonal lineage tracing with single-cell transcriptional identity, developmental timing, and spatial mapping across multiple forebrain regions. This integrated dataset has considerable value as a resource for the field and enables lineage relationships to be interrogated at a scale and resolution that were not achievable with earlier approaches. The additional subregion-resolved analyses, validation experiments, improved anatomical annotations, and public browser further strengthen both the biological and community value and novelty of the manuscript.

The strength of this work lies not only in identifying individual novel observations, but also in providing a systematic framework to interrogate lineage relationships across multiple forebrain regions, cell types, developmental stages, and anatomical subdomains. The added subregion-resolved analyses and validation experiments further strengthen the biological insight of the study, while the public browser will make the dataset broadly useful to the developmental neuroscience community.

The revised text also more carefully frames the conclusions as population-level clonal biases rather than complete lineage reconstructions or strict fate restrictions, and the added simulations and barcode-shuffling analyses provide a useful framework for interpreting these lineage relationships despite incomplete clone recovery. Overall, the authors have responded thoroughly and constructively to all the comments. The revised manuscript represents an important contribution and resource for developmental neuroscience and at this stage is deemed suitable for publication.

Referee #3

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #4

(Remarks to the Author)

The authors have addressed the most critical points in the revised manuscript.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We greatly appreciate the reviewers' time and effort to evaluate the manuscript, and thank them for their thoughtful critique and recommendations. We have addressed their specific comments in revision and share our responses to specific points below.

Referees' comments:

Referee #1 (Remarks to the Author):

1. There is no justification that the numbers of cells sequenced and the numbers of multi-cell clones with barcodes recovered are enough for the authors to make conclusion about cell lineage across the multiple brain regions analyzed together. The methods involved labeling progenitors with STICR lentivirus, sorting the labeled cells using FACS, single cell RNA-seq on a 10X platform, and data analysis. Cell loss will occur at multiple steps starting from cell dissociation. The percentage of the total cells that are recovered by FACS can vary significantly, often being 50% or less. The cell capture efficiency on a 10X platform range from 50-80%. The barcodes will not be recovered for all the cells, etc. The authors should consider the numbers of progenitor cells and all these factors, and calculate how many multi-cell clones they need to make meaningful conclusion about cell lineage in this study.

Response:

We thank the reviewer for raising this important point. Incomplete recovery of barcoded cells is an important technical limitation of lineage-tracing approaches that use scRNA-seq as the readout. At its core, the goal of clonal lineage tracing is to uncover insights into progenitor cell dynamics and fates of the daughter cell. To provide insights into the relationship between the technical shortcomings of the methodology and statistical robustness of insights, we performed systematic modeling of simulated data and barcode shuffling analysis.

First, to estimate the range of capture efficiency, we examined the cell loss during the steps for which experimental evidence can be recorded. Based on recorded cell numbers, 51% of GFP-positive cells detected during FACS are recovered by 10X Genomics (**Extended Data Fig. 2a**), which is consistent with the reviewer's assessment. Among the 10x-captured cells, approximately 61% retained high-quality barcode information after doublet removal and low-quality filtering of both transcriptomics and barcodes (**Extended Data Fig. 2a**). To have a more accurate estimate of likely cell loss, we took advantage of the observed fraction of multicellular clones in our dataset (30.7% at E12) and compared the numbers to previously published data across comparable developmental timepoints and brain regions¹⁻⁷. In these studies, which relied on clonal analysis in intact tissues, the average clone size produced by a single progenitor cell starting from E12 was between 5 and 20 cells. With the multicellular clone ratio (30.7%) and an estimate of the "true" clone size from previous literature (5-20), we were able to calculate the capture rate (**see method**) to be approximately ~5.6–22.1%, with ~10% as a reasonable working estimate. It is worth noting the caveat of this analysis, which is that the prior 'reference' studies limited their clonal tracing to progenitor cells, whereas lentivirus infection can include barcoding of post-mitotic cells, which is likely to obscure accurate assessment of the exact recovery efficiency. Nonetheless, we note that our assessed efficiency is significantly higher than the ~0.5% reported in a previous study⁸, which reflects the optimized expression cassette in our study and differences in developmental stage.

Second, we performed a clonal sampling simulation within a ground-truth clone to determine how many multicellular clones would be required to robustly recover the major lineage features under realistic capture rates (**Extended Data Fig. 2b**). Briefly, we first simulate ground-truth clones with defined cell-type composition

and clone sizes. We then performed Monte Carlo simulations, repeatedly sampling cells from ground-truth clones under defined capture rates, to assess how the number of recovered multicellular clones influences the interpretation of the results (**see method**). Specifically, we evaluated: (i) whether all constituent cell types, including a rare cell type, could be detected; (ii) whether the relative cell-type ratios within a clone could be recovered; and (iii) whether trends in clone size could still be resolved.

(i) For recovery of all constituent cell types, we found that in the sampling regime most relevant to our experiments (~5.6–22.1%), the probability of detecting all cell types increases rapidly as the number of recovered multicellular clones increases. At a sampling rate of ~10%, recovery of approximately 10 multicellular clones was sufficient to achieve >80% probability of detecting all constituent cell types in the ground-truth clone (**Extended Data Fig. 2c**). (ii) For recovery of cell-type proportions, we compared the inferred cell-type ratios from sampled multicellular clones with the known ground-truth ratios and calculated the correlation between them. When more than 10 multicellular clones were recovered, the inferred cell-type ratios showed high concordance with the ground truth, with correlation coefficients exceeding ~0.8 (**Extended Data Fig. 2d**). (iii) For clone size, our method does not support precise inference of the absolute size of an individual clone in the same way as sparse imaging-based approaches such as MADM. However, the relevant question for this study is whether overall trends in clone size remain detectable under partial sampling. We therefore simulated multiple clones with different sizes and found that at a sampling rate of ~10%, recovery of 10 multicellular clones was sufficient to capture the overall trend in clone-size differences with good clarity (**Extended Data Fig. 2e**). Overall, we estimate that clonal fate assignments supported by more than 10 multicellular clones begin to reach a reliable threshold, with increasing clone numbers further improving robustness.

To further test the robustness of our major conclusions, we performed barcode shuffling analysis by randomizing the correspondence between clone identity and cell identity 200 times and repeated the main analyses (**Extended Data Fig. 21a**). In contrast to the real data, barcode-shuffled data did not reproduce the structured intra- and inter-regional organization of the clonal relationship matrix (**Extended Data Fig. 21b**), the strong differential-expression signals between regionally dispersed and non-dispersed clones (**Extended Data Fig. 21c**), or the lineage-bias pattern of excitatory neurons with astrocytes and inhibitory neurons with OPC (**Extended Data Fig. 21d**).

Taken together, because recovery is incomplete, we do not attempt to exhaustively reconstruct individual lineages or infer the full composition and absolute size of single clones. Instead, the strength of our dataset lies in repeated partial sampling across many independently recovered multicellular clones, enabling statistically robust inference of population-level lineage relationships. Simulations indicate that recovery of ~10 multicellular clones per fate is sufficient to support the major classes of lineage inferences under realistic sampling conditions. Thus, as the reviewer suggested, the number of multicellular clones should be considered when interpreting the results. In our dataset, the number of recovered multicellular clones substantially exceeded this threshold, with 25,943 clones in the cortex, 13,762 in the striatum, 7,869 in the hippocampus, 11,908 in the thalamus, and 11,281 in the OB.

We revised the main text and discussion to clarify the limitations imposed by incomplete recovery and to distinguish population-level lineage inference from exhaustive reconstruction of individual clones.

Addressing this comment and comment 2, We added the following paragraph into the main text:

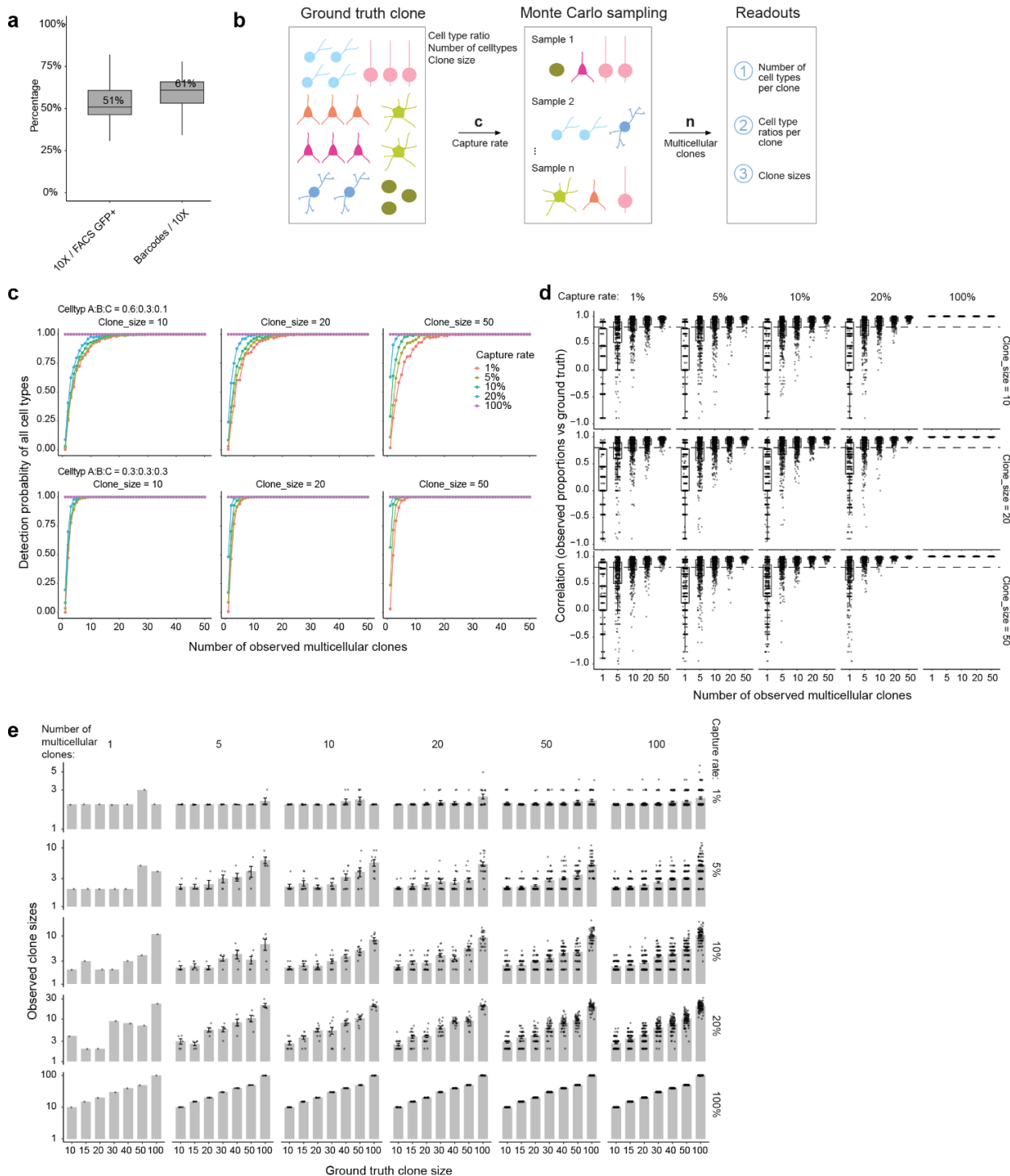
“Because scRNA-seq lineage tracing only partially captures each barcoded clone, we used simulations to assess the robustness of population-level lineage inference under repeated partial clone recovery. These simulations showed that, across 5–20% sampling rates, >10 recovered multicellular clones generally captured major clonal features, including cell-type composition, proportions, and clone-size

trends (Extended Data Fig. 2). We further tested whether incomplete sampling could generate apparent false clonal biases, and found that such false biases became uncommon above this threshold and fell below 5% once more than 20 multicellular clones were recovered (Extended Data Fig. 3). Thus, the large numbers of multicellular clones across regions in our dataset provide statistical power to resolve heterogeneous lineage relationships at the cluster level."

Addressing this comment and comment 2, we added the following paragraph into the discussion:

"Incomplete recovery of barcoded cells is an inherent limitation of scRNA-seq-based lineage tracing. We therefore do not reconstruct complete individual lineages, but infer population-level lineage patterns from repeated partial sampling of independently recovered multicellular clones. Simulations and barcode-shuffling analyses supported the robustness of these inferences, with major clonal features generally captured by >10 multicellular clones and false bias rates falling below 5% with >20 multicellular clones (Extended Data Figs. 2, 3, and 21)."

To aid the reader of the manuscript in assessing the data and statistical robustness, we added the analyses in supplementary figures. To aid the reviewer, we also include a copy of those figures below:



Extended Data Figure 2: Simulations evaluating the effects of capture rate and the number of recovered multicellular clones

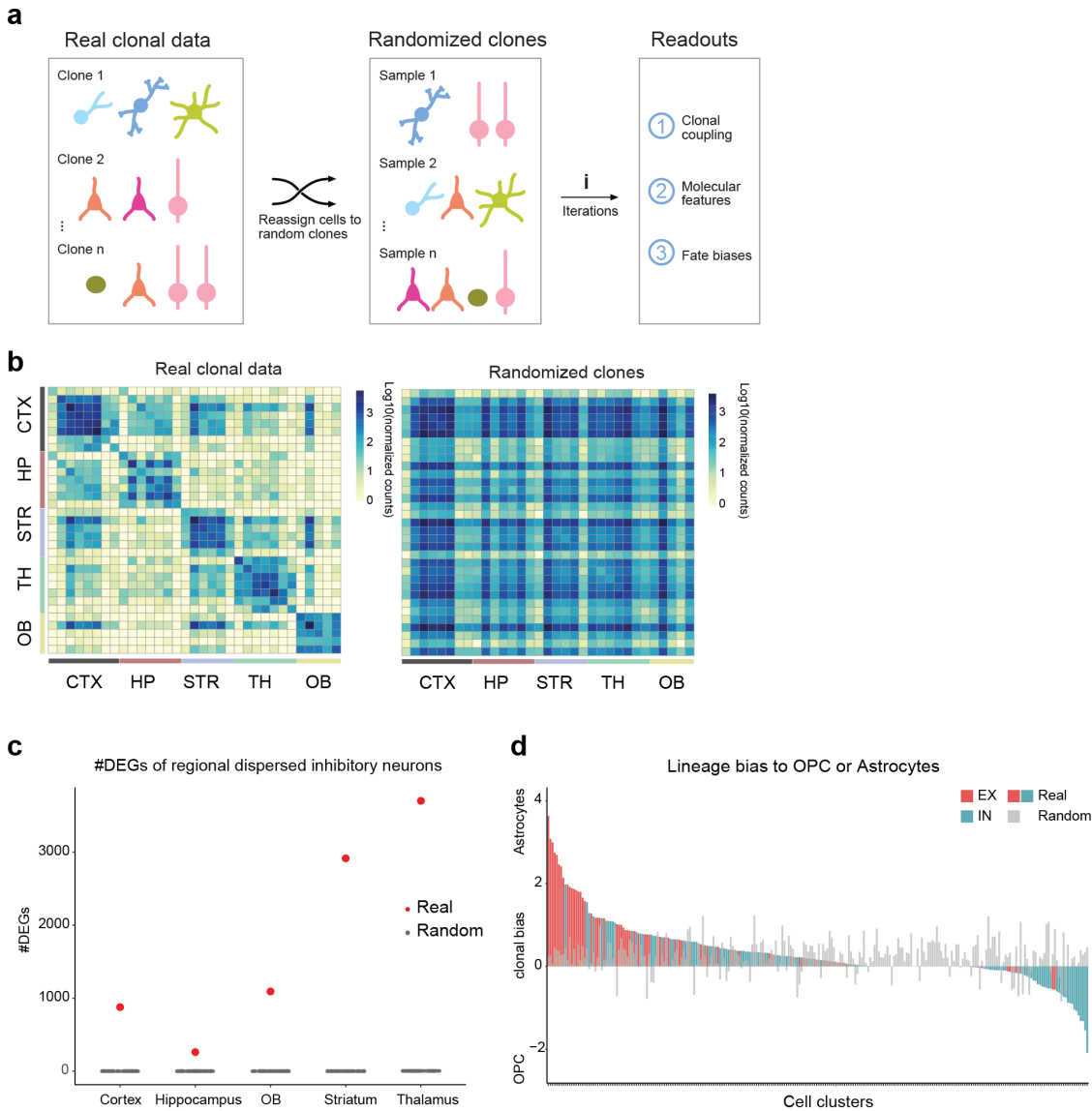
a, Box plots showing cell loss across experimental steps. Boxes indicate the median and interquartile range.

b, Schematic of the sampling simulation framework using ground-truth clones. The simulation models how clone features inferred from repeated partial sampling under different capture rates recover the ground-truth lineage structure.

c, Probability of recovering all cell types across different numbers of recovered multicellular clones. Panels represent different capture rates or ground-truth clone sizes. This analysis estimates how many observed multicellular clones are required to recover the constituent cell types of a lineage.

d, Correlation between cell-type proportions in simulated and ground-truth clones across different numbers of recovered multicellular clones. Panels represent different capture rates and ground-truth clone sizes.

e, Observed clone sizes for clones with different ground-truth sizes, illustrating how capture rate and the number of recovered multicellular clones influence recovery of clone-size differences.



Extended Data Figure 21: Random barcode reassignment abolishes major clonal features

a, Schematic of the barcode-reassignment analysis. Clonal barcodes were randomly reassigned to cells while preserving the overall barcode distribution, thereby disrupting the true correspondence between barcode identity and cell type. This analysis was used to test whether the major clonal features observed in the real dataset depend on the authentic barcode assignment.

b, Normalized counts of shared clones between cell types across regions under the real barcode assignment (left) and random barcode reassignment (right). Values were normalized to the total number of multicellular clones. Only multicellular clones were included in the analysis.

c, Number of DEGs identified for regionally dispersed inhibitory neurons using the real barcode assignment and random barcode reassignment. Differential expression was performed using a two-sided Wilcoxon test.

d, Biased enrichment of excitatory neurons or inhibitory neurons with astrocytes or OPCs under the real barcode assignment and random barcode reassignment.

2. In multiple places in the manuscript, the authors make conclusions about fate-restricted populations (for example, lines 137-138, 142-144, 154-155, 169-170 etc) or region-restricted. However, the numbers of such clones were not explicitly reported. Given the reasons in 1, these so-called fate-restricted or region-restricted clones might be due to incomplete capture of the whole lineage, instead of fate-restriction or region-restriction.

Response:

We thank the reviewer for this thoughtful comment. We agree that, given incomplete lineage recovery, apparently fate- or region-restricted clones could in principle arise from partial lineage capture. Although such patterns may suggest the existence of “restricted” clones, especially when many multicellular clones are recovered. Our original use of “fate-restricted” and “region-restricted” was not intended to imply definitive exclusion of alternative fates or regions at the level of individual progenitors. Rather, these terms were used to describe population-level trends in clonal output that are repeatedly observed across the dataset. Because partial sampling prevents us from making strict claims of absolute restriction, we have replaced “restriction” with “bias” in the revised manuscript to better reflect the level of inference. We now also explicitly report the number of multicellular clones for all places where clonal biases were discussed in the main text and figure legends.

One instance of the revised text is:

“Interestingly, we identified a fate-biased population of Tshz2 positive clusters 39 (Bcl11b+) and 36 (Satb2+) in 105 multicellular clones”.

Moreover, we further tested how incomplete sampling affects the appearance of “bias” and how many multicellular clones would be enough to make robust interpretation. As a direct extension of the analyses described in our response to Comment 1, we performed additional simulations specifically aimed at this issue. In these simulations, we constructed ground-truth scenarios in which cell types either truly arose from the same clone (**Extended Data Fig. 3a**) or instead belonged to two fully separated clonal fates (**Extended Data Fig. 3b**). We then asked how the number of recovered multicellular clones affects the probability of correctly identifying the underlying pattern, and conversely, how often incomplete sampling from a non-restricted ground truth can falsely generate an apparently fate-restricted pattern. These simulations showed that clone number is a critical parameter for interpretation. When only a small number of multicellular clones are recovered, such as less than 5, incomplete sampling can indeed give rise to an apparent pattern of fate-bias even when the ground truth is not. However, as the number of multicellular clones increases, the inferred pattern converges toward the true underlying state. Once the number of recovered multicellular clones exceeds 10, such false-positive “bias” patterns become much less likely; beyond 20 multicellular clones, the false fate bias rates become under 5% and robust to sampling rate (**Extended Data Fig. 3c**). We believe that this simulation and statistical framework will also be broadly useful for future sequencing-based lineage-tracing studies, which have similar partial sampling limitations.

We revised the manuscript to avoid describing these findings as definitive fate restrictions. As discussed in comment 1, in our dataset, the number of recovered multicellular clones substantially exceeded this threshold, with 25,943 clones in the cortex, 13,762 in the striatum, 7,869 in the hippocampus, 11,908 in the thalamus, and 11,281 in the OB. Importantly, the major findings reported here are supported by more than 20 multicellular clones, with the number explicitly reported.

As in the response of Comment 1, we revised the main text and discussion to clarify the limitations imposed by incomplete recovery and to distinguish population-level lineage inference from exhaustive reconstruction of individual clones.

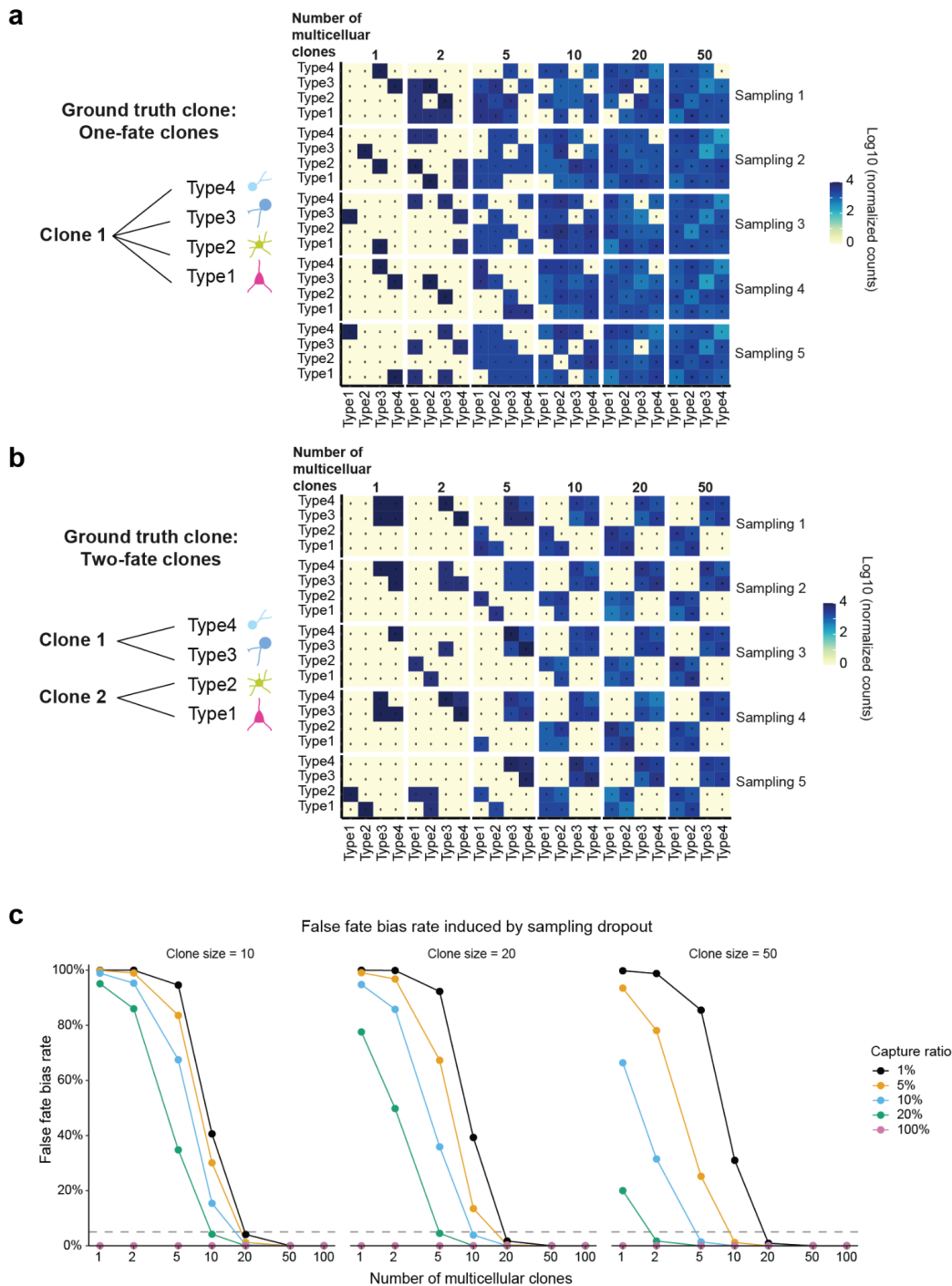
Addressing this comment and comment 1, We added the following paragraph into the main text:

“Because scRNA-seq lineage tracing only partially captures each barcoded clone, we used simulations to assess the robustness of population-level lineage inference under repeated partial clone recovery. These simulations showed that, across 5–20% sampling rates, >10 recovered multicellular clones generally captured major clonal features, including cell-type composition, proportions, and clone-size trends (Extended Data Fig. 2). We further tested whether incomplete sampling could generate apparent

false clonal biases, and found that such false biases became uncommon above this threshold and fell below 5% once more than 20 multicellular clones were recovered (Extended Data Fig. 3). Thus, the large numbers of multicellular clones across regions in our dataset provide statistical power to resolve heterogeneous lineage relationships at the cluster level.”

Addressing this comment and comment 1, we added the following paragraph into the discussion:

“Incomplete recovery of barcoded cells is an inherent limitation of scRNA-seq–based lineage tracing. We therefore do not reconstruct complete individual lineages, but infer population-level lineage patterns from repeated partial sampling of independently recovered multicellular clones. Simulations and barcode-shuffling analyses supported the robustness of these inferences, with major clonal features generally captured by >10 multicellular clones and false bias rates falling below 5% with >20 multicellular clones (Extended Data Figs. 2, 3, and 21).”



Extended Data Figure 3: Simulations evaluating how the number of recovered multicellular clones affects inference of clonal biases

a, Heatmap from simulations based on a ground-truth configuration in which all cells belong to a single clone, showing the number of shared clones between two cell types. Each row of panels represents an independent sampling simulation, and each column of panels represents the number of recovered multicellular clones.

b, Heatmap from simulations based on a ground-truth configuration in which cells belong to two separate clones, showing the number of shared clones between two cell types. Each row of panels represents an independent sampling simulation, and each column of panels represents the number of recovered multicellular clones.

c, False fate-bias rate, calculated from the ground-truth clones in panel a, across different capture ratios, clone sizes, and numbers of multicellular clones. False fate bias is defined as the probability that >25% of cell-type pairs exhibit either zero shared clones or a shared-clone fraction below 10% of the smaller of the two clone counts.

3. Most of the main conclusions are already known, including glutamatergic neurons and astrocytes were predominantly generated and retained within their region of origin, while clonally-biased GABAergic neurons and oligodendrocyte precursor cells exhibited extensive dispersion across regions.

Response:

We agree that individual findings of regional restriction of glutamatergic neurons and astrocytes versus dispersion of GABAergic neurons and oligodendrocytes are well established. What is surprising, however, are the degree of clonal bias of progenitors generating astrocytes and glutamatergic neurons segregating from progenitors biased towards generating GABAergic neurons and oligodendrocytes. Although individual examples have been reported, the brain-wide comprehensive sampling highlights this duality as a general principle, which we believe is an important concept to highlight, even if individual examples have been studied previously in individual brain regions.

At the same time, the study goes beyond confirming prior observations on a more systematic scale. We identify molecular signatures of GABAergic neurons with broadest cross-region dispersion, define extensive postnatal glutamatergic neurogenesis across multiple forebrain regions, and identify previously unrecognized lineage relationships in the cingulate cortex. These findings are enabled by the scale of the dataset and by the direct integration of clonal, transcriptomic, spatial, and temporal information.

In the revised manuscript, we followed the reviewers' recommendation, also in the response to the Comment 1 of the reviewer #4, to dig deeper into the data to provide more new biological insights that distinguish our study, particularly enabled by the integration of clonal lineage tracing and spatial transcriptomics. Briefly, we expanded the spatial analysis by mapping 416 clusters to subregional anatomical locations using the P4 MERFISH reference and re-examining lineage relationships at finer spatial resolution (**Extended Data Fig. 8**). This revealed additional subtype- and subregion-specific lineage structure across multiple brain regions, including finer temporal heterogeneity, anatomically localized clonal organization, and region-specific differences in neuronal–glial lineage relationships (**Figs. 3d-i, Figs. 4d-g; Extended Data Figs. 6,15,16,19, and 20**).

In the striatum, higher-resolution annotation resolved MSN populations with clear dorsal–ventral specificity and uncovered subtype-specific temporal differences in MSN production (**Figs. 3d-e**). Notably, we observed increased lineage separation between late-born neurons and late-generated astrocytes in the striatum (**Fig. 3f**).

In the thalamus, astrocyte–OPC clonal coupling was strongest at early embryonic stages and markedly reduced after E16, in contrast to the cortex, where astrocyte–OPC coupling persisted later, including after birth (**Extended Data Fig. 19**).

In the OB, spatial mapping of interneuron clusters resolved broad subregional groups and revealed both a superficial-to-deep temporal gradient and substantial heterogeneity within subregions, including subtype-specific restricted clonal patterns and distinct relationships with striatal cell types (**Extended Data Fig. 16, Fig. 3g**).

In the hippocampus, assigning neuronal and glial clusters to CA1, CA3, and DG (**Extended Data Fig. 20**) revealed subtype-specific lineage relationships, including selective clone sharing between particular OPC populations and hippocampal neuronal lineages (**Fig. 4g**), progressive loss of neuron–glia clonal coupling after E16, and shared clones between hippocampal CR cells, an ependymal-like astrocyte cluster, and ependymal cells (**Extended Data Figs. 6d-f**).

Together, these analyses illustrate the value of integrating cross-region lineage tracing with whole-brain spatial transcriptomic information to resolve developmental lineage relationships at finer anatomical and molecular resolution. We've also revised our abstract and discussion to highlight novel findings from our study.

We added these new results into our revised main text, figures, and supplementary figures as follows:

“Spatial mapping resolved striatal clusters into medium spiny neurons (MSN) and interneurons with dorsal–ventral subregional specificity (Fig. 3d, Extended Data Fig. 15a). MSN populations across striatal subregions remained broadly clonally related, but differed in their temporal production dynamics (Fig. 3e). For example, D1 MSNs were mainly recovered from E11–E15 labeling, whereas D2 MSNs were continuously labeled, consistent with previous studies⁶. Notably, ventral D2 MSNs showed more prolonged neurogenesis than dorsal D2 MSNs (Fig. 3e), while olfactory tubercle-enriched D3 MSNs were recovered predominantly from early labeling stages. We also identified an Oprm1 MSN population⁹, marked by the mu-opioid receptor associated with opioid-related reward behaviors, which was labeled at both early (E11–E15) and late (E16–P0) stages. Leveraging the temporal coverage of our dataset, we next asked whether the prolonged MSN neurogenesis coincided with astrogenesis by examining clonal relationships between MSN subtypes and striatal astrocytes labeled at E11–E15 or E16–P0. Interestingly, late-born MSNs and striatal astrocytes labeled after E16 were largely clonally separated (Fig. 3f), suggesting that the MSN and astrocyte lineages in the striatum diverge after E16.”

“Mapping OB interneuron clusters to subregional locations resolved three broad spatial groups: glomerular layer (GL), superficial granule cell layer (GCL), and deep GCL (Extended Data Fig. 16a). These results suggest that OB cell populations at P4 emerge according to a temporal order: superficial populations are generated first, whereas deeper populations emerge subsequently (Extended Data Fig. 16b). Within each anatomical domain, multiple transcriptomic subpopulations of OB interneurons with heterogeneous clonal patterns can be identified (Extended Data Fig. 16c–e). Extending the analysis to striatal populations, GL clusters showed less clonal sharing with striatal interneurons than GCL clusters. In contrast, most OB GCL clusters shared clones with MSNs, especially at E11–E15 labeling, with marked heterogeneity across specific OB and striatal subtypes (Extended Data Fig. 16f).”

“Analyzing the clones shared between striatal and OB GABAergic neurons, notably, we found D1 MSNs showed a lower proportion of shared clones with OB inhibitory neurons than D2 MSNs, whereas both D1 and D2 MSNs shared more clones with deep GCL than with superficial GCL neurons, consistent with previous findings¹⁰. Notably, the Oprm1 MSN cluster showed a significantly higher proportion of clone sharing with OB inhibitory neurons than other MSN clusters (~ 3 fold) (Fig. 3g). Unlike D1 and D2 MSNs, Oprm1 MSNs shared clones with superficial and deep GCL neurons at comparable proportions at early labeling (E11–E15), but biased towards deep GCL neurons at late labeling (E16–P0). Given the clonal separation between MSNs and striatal astrocytes after E16, we next asked whether these striatal populations were linked to the same OB populations. Both late-labeled MSNs and astrocytes shared clones with deep GCL neurons; however, only MSNs shared clones with GL neurons, whereas only astrocytes shared clones with superficial GCL neurons. These results suggest that late striatal neurogenesis and astrogenesis are associated with distinct, although partially overlapping, OB lineage relationships.”

“Higher-resolution spatial mapping further revealed subregion-specific lineage relationships. For example, astrocytes in the thalamus and epithalamus shared clones at early stages, but these shared clones were markedly reduced after E16 (Extended Data Fig. 19a,b). Thalamic astrocytes and OPCs shared clonal barcodes at early embryonic stages, with minimal clonal coupling observed after E16 (Extended Data Fig. 19b). In contrast, cortical astrocyte–OPC clonal coupling persisted later, including after birth, and cortical OPC–astrocyte clone sharing remained detectable after P0, although the OPCs

in these shared clones were predominantly located in white matter, making it unclear whether they represent cells migrating into the cortex or a distinct white matter-associated population (Extended Data Fig. 19c,d). In the hippocampus, assigning astrocyte, OPC, and neuronal clusters to specific subregions and layers of CA1, CA3, and DG revealed subtype-specific lineage patterns (Extended Data Fig. 20). Interestingly, most OPC clusters shared clones primarily with CA1 neurons, indicating that the majority of hippocampal OPCs are linked to the CA1 lineage, whereas one OPC cluster shared clones with neurons across CA1, CA3, and DG, suggesting transcriptomic convergence from distinct lineage origins (Fig. 4g). After E16, OPC–neuron clonal sharing was largely lost, consistent with divergence of neurogenesis and gliogenesis. By contrast, astrocyte–neuron clonal coupling persisted longer, particularly in DG, where DG-localized astrocyte clusters continued to share clones with DG neurons (Fig. 4g).”

“Notably, hippocampal CR cells were found sharing clonal lineage relationships with astrocytes and ependymal cells (Fig. 1d, Extended Data Fig. 6d-e), which were not detected in cortical CR cells (Extended Data Fig. 6f).”

4. Although a comprehensive atlas of cell lineages in broad brain regions will be very useful for understanding brain development, the lineage data presented in the current study do not capture a complete lineage landscape, and most of the lineages are likely incomplete, which reduces the value of this study.

Response:

We thank the reviewer for recognizing the value of building a comprehensive atlas of lineage relationships across the developing brain. To facilitate community use of this resource, we have generated a user-friendly UCSC Cell Browser interface that includes both the lineage-tracing dataset and the mapped whole-brain MERFISH data. We also acknowledge that the current technology cannot recover all descendants of every individual clone. In the revised manuscript, we now clearly state this limitation and highlight the new insights enabled by this dataset.

“An interactive browser of single-cell data and MERFISH data can be found at the University of California, Santa Cruz (UCSC) cell browser <https://p4-brain-lineage.cells.ucsc.edu>. ”

At the same time, we wish to underscore the value of identifying reproducible lineage relationships among transcriptomically defined cell types. A major strength of sequencing-based lineage tracing lies not in exhaustively reconstructing complete clonal architectures, which are themselves partly shaped by stochasticity², but in systematically sampling large numbers of clones across developmental stages, brain regions, and molecularly defined cell populations. Although individual clones may be only partially recovered, recurrent patterns observed across many independent clones can reveal underlying principles of progenitor behavior.

In response to Comments 1 and 2, we incorporated a statistical framework to account for the technical limitations of sampling-based lineage reconstruction (**Extended Data Figs. 2, 3, and 21**). We now focus on reproducible enrichment or depletion of lineage relationships across many independent clones and cell populations, while interpreting rare or unobserved relationships with caution. Together, these revisions strengthen the major conclusions and highlight the value of this resource despite incomplete recovery of individual clones.

A key advance of this study is the direct integration of clonal relationships with comprehensive single-cell molecular identities across broad spatial and temporal contexts of brain development. Leveraging this

framework, our additional analyses revealed subtype- and subregion-specific lineage structures across multiple brain regions, including finer temporal heterogeneity in the olfactory bulb, anatomically localized clonal organization in the striatum, and region-specific differences in neuronal–glial lineage relationships in the hippocampus, as described in our response to Comment 3.

We have revised the Abstract, main figures, and Discussion to highlight the novel findings and resource value of this study, while clearly acknowledging the limitations of the approach.

As in Abstract:

“We uncover previously unrecognized differences in generation of ventral and dorsal medium spiny neurons (MSNs), as well as unexplored lineage relationships between late-born MSN, striatal astrocytes and distinct subpopulations of the OB neurons. We show that clones of GABAergic neurons with broadest regional dispersion are enriched for genes implicated in neurodevelopmental disorders. Our results highlight a developmental principle whereby glutamatergic neurons and astrocytes are clonally related and retained within their region of origin (“born together, stay together”), whereas GABAergic neurons are clonally related to oligodendrocyte precursor cells (OPCs) but disperse extensively across brain regions (“born together, rear apart”). Together, our findings will serve as a foundational reference for studies of mammalian forebrain development and its disorders.”

As in Discussion:

“Incomplete recovery of barcoded cells is an inherent limitation of scRNA-seq–based lineage tracing. We therefore do not reconstruct complete individual lineages, but infer population-level lineage patterns from repeated partial sampling of independently recovered multicellular clones. Simulations and barcode-shuffling analyses supported the robustness of these inferences, with major clonal features generally captured by >10 multicellular clones and false bias rates falling below 5% with >20 multicellular clones (Extended Data Figs. 2, 3, and 21).”

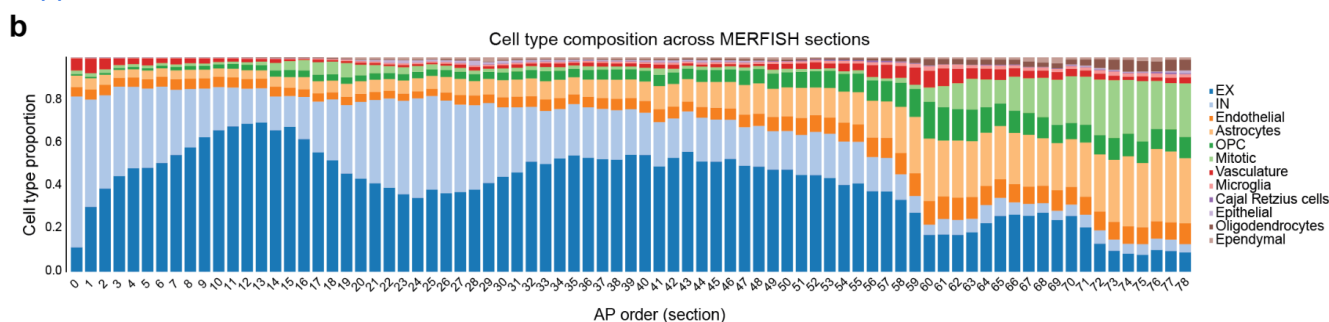
Referee #2 (Remarks to the Author):

Here we list a few points that, if addressed, could further increase clarity and impact:

1. The spatial maps of MERFISH data nicely illustrate the anatomical distribution of major cell types. To further enhance interpretability, the author should please consider adding stacked bar plots showing the relative proportions of cell types across sections or regions to complement spatial localization.

Response:

We thank the reviewer for this suggestion. We have added stacked bar plots showing cell-type proportions across sections in the MERFISH data to complement the spatial maps. This figure provides a high-resolution view of the dynamics of cell-type distribution along the rostral-to-caudal axis. We've added this figure to the supplement.



Extended Data Figure 5b: Cell-type proportions across MERFISH sections along the rostral–caudal axis. Sections are ordered from rostral to caudal, with AP = 0 denoting the most rostral section. Cell-type proportions were normalized to the total number of cells in each section.

2. The dot plot in Extended Data Fig. 2b appears to show a relatively higher proportion of inhibitory (GABAergic) neurons than excitatory (glutamatergic) neurons in cortex, whereas the expected ratio is closer to ~80% excitatory and ~20% inhibitory. Please, clarify whether values reflect proportions across all cell types or are calculated independently for each cell type across regions. If the latter, indicating this explicitly would help, as relative PN/IN proportions are otherwise not directly observable.

Response:

We thank the reviewer for this careful observation. We would like to clarify that the values shown in revised **Extended Data Fig. 4b** represent the proportions of recovered cell types within each region, rather than values independently normalized for each cell type across regions. We have revised the figure legend to clarify this point. The relatively lower apparent proportion of excitatory neurons in the cortex likely reflects a combination of biological and technical factors. First, the cells shown are derived from lentivirus infections generated on or after the day of viral injection rather than all cells present in the corresponding anatomical region at P4. Since many of our injection time points fall into the late neurogenesis, one would expect that the relative frequency of neurons versus glia and other cell types would be lower, leading to a pattern that mirrors a drop in excitatory neuron frequency. Second, excitatory neurons are generally more fragile and prone to loss during processing, leading to their relative underrepresentation among recovered cells. We have included a brief discussion of this point in the figure legend of **Extended Data Fig. 4b** and clarified that the values should not be used to derive values of absolute excitatory-to-inhibitory ratios.

“Dot plot showing the proportion of cell types in each region, with color and dot size indicating the ratio. The proportions were calculated by normalizing the count of each cell type to the total number of cells across all cell types within each region. These proportions were derived from the GFP+ cells collected and not the absolute cell-type ratios in the brain.”

3. The average clone size decreases with later injections and is attributed to shorter lineage-tracing duration (Extended Data Fig. 2c). Would it be more accurate to also interpret this trend as progressive clonal restriction or loss of progenitor potency over developmental time? could the authors clarify this point would better connect the observation to underlying mechanisms.

Response:

We thank the reviewer for this insightful suggestion. We agree that the decrease in average clone size with later injection times likely reflects not only the shorter duration of lineage tracing, but also progressive clonal restriction and reduced progenitor output over developmental time.

This interpretation is supported by our observation that later time points are associated with decreased cell-type diversity and stronger fate restriction. For example, in the striatum, clonal sharing between medium spiny neurons and astrocytes is evident at earlier developmental stages, but becomes much less frequent at later stages and is not detected after E16 (**Fig. 3f**). Similar stage-dependent restriction is also observed in thalamic OPC and astrocyte lineages (**Extended Data Fig. 19**). In addition, cross-regional clonal sharing among OPCs becomes more limited at later stages (**Fig. 4f**). Overall, these restrictions become apparent after E16 and are further strengthened after P0.

We have clarified this point in the revised main text to better connect the observed decline in clone size with the progressive restriction of progenitor competence during development.

“The average clone size progressively decreased with later injections (Extended Data Fig.4c), consistent with the shorter duration of lineage tracing and reduced progenitor output as development proceeds¹¹”

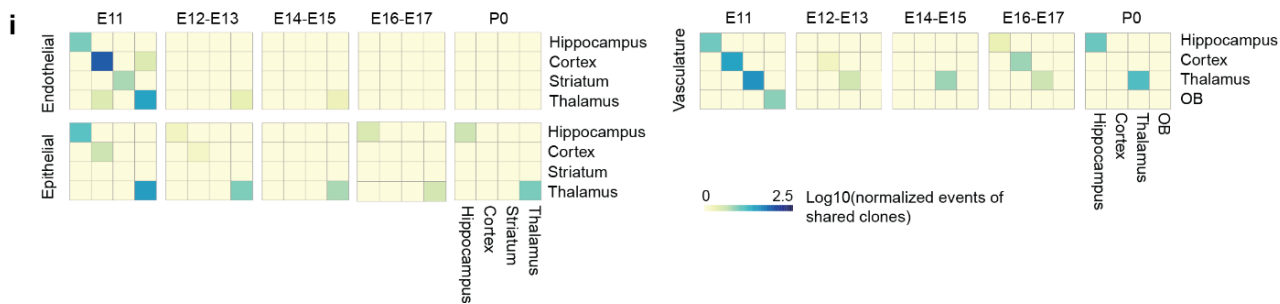
4. The manuscript reports early generation and broad early dispersion of microglia, with late-labeled microglial clones no longer dispersed across regions. Could the authors clarify whether endothelial, vascular, and epithelial lineages were similarly evaluated for clonal dispersion, and whether comparable patterns were observed?

Response:

To address this interesting question, we performed similar clonal dispersion analyses for endothelial, vascular, and epithelial lineages. Interestingly, we did not detect early broad dispersion seen in microglial clones (**Extended Data Fig. 7i**), indicating that this feature is specific to microglia rather than broadly shared across these other lineages. We have incorporated the corresponding analyses and figures into the revised manuscript.

“However, no clonal barcodes were shared between microglia in the hippocampus and the OB. Most microglia, particularly those forming regionally dispersed clones, were labeled early, around E11. And

this dispersed pattern was not found in other non-neuroepithelial lineages such as vascular or endothelial cells (Extended Data Fig. 7i)."



Extended Data Figure 7i. Heatmap showing the number of shared clones between endothelial cells, epithelial cells, and vasculature cells across different injection timepoints. Clone counts are normalized by the total number of clones detected at each injection age.

Additionally, did you detect transient-amplifying microglia or intermediate proliferative states, and if so, how were these distinguished (markers, spatial context, clone size distributions)?

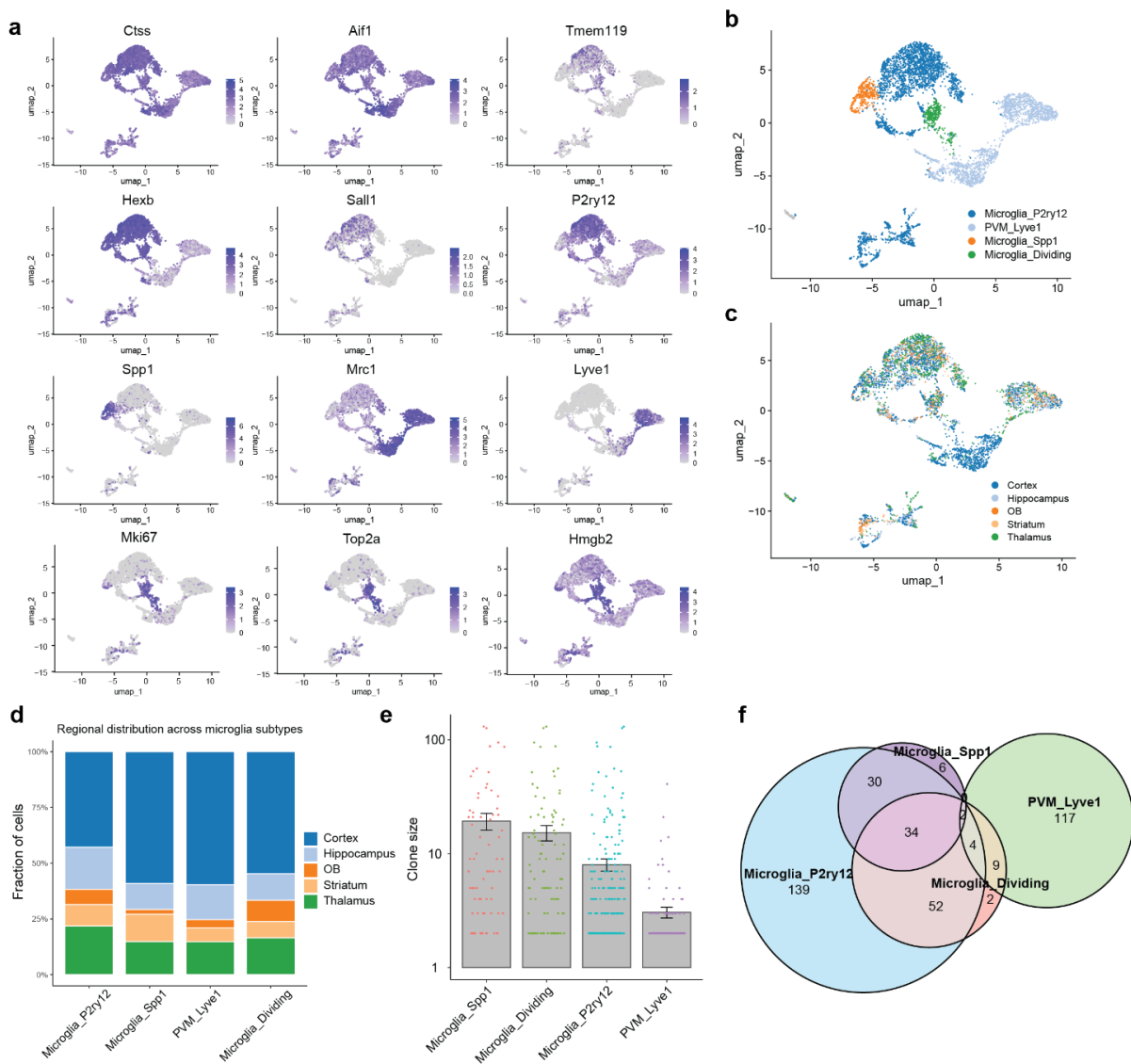
Response:

We thank the reviewer for this insightful question. We re-examined our dataset for evidence of transient amplifying or intermediate proliferative microglial states and identified a population consistent with this identity (**Extended Data Fig. 7a,b**). This population was defined based on the expression of canonical cell-cycle markers, including *Mki67* and *Top2a*, and has been reannotated accordingly in the revised manuscript (**Extended Data Fig. 7a,b**). We further analyzed the spatial distribution and clonal properties of this proliferative microglial population. These cells were found across all regions with a higher ratio in OB (**Extended Data Fig. 7c,d**). In addition, we found the proliferative microglia are associated with larger clone sizes, consistent with their proliferative expansion (**Extended Data Fig. 7e**). Clonal relationship analysis indicates that these proliferative microglia are most frequently related to *P2ry12*⁺ and *Spp1*⁺ microglia rather than *Lyve1*⁺ cells (**Extended Data Fig. 7f**).

Corresponding analyses and figures have been added to the revised manuscript, as follows.

*"further subdivided into *P2ry12*⁺ homeostatic, *Spp1*⁺ axon track-associated subtypes, and transiently dividing microglia, in addition to *Lyve1*⁺ perivascular macrophages (Extended Data Fig. 7a-d)"*

*"By contrast, *P2ry12*⁺ and *Spp1*⁺ frequently shared clonal barcodes, indicating these microglial subpopulations can share a common origin. Transiently dividing microglia were most frequently related to *P2ry12*⁺ and *Spp1*⁺ microglia rather than *Lyve1*⁺ cells, suggesting their higher proliferation at P4."*



Extended Data Figure 7: Lineage analysis of microglia subtypes

- a**, Expression of marker genes across microglia subtypes. Each dot represents a single cell.
- b**, UMAP visualization of microglia in clones from five brain regions. Cells are colored by subtypes of microglia.
- c**, UMAP visualization of microglia in clones from five brain regions. Cells are colored by brain regions.
- d**, Stacked bar plot showing the distribution of different microglial subtypes across regions.
- e**, Bar plot showing the average clone size for microglia subtypes. Error bars indicate the standard error, and each dot represents an individual clone.
- f**, Venn diagram showing shared clones of microglia between different subtypes.

5. Regarding the clonal analysis of microglia, could the authors clarify how these cells are labeled in their system? Since microglia originate from yolk-sac progenitors rather than ventricular neuroepithelium, their labeling likely reflects infection of already colonizing cells that subsequently amplify locally within the brain. It would be helpful to make this point explicit, and to comment on whether similar considerations apply to other non-neuroepithelial lineages such as vascular or endothelial cells.

Response:

We thank the reviewer for this question. During in utero injection, viral particles are delivered into the ventricular space and can infect cells that are in direct contact with the ventricular surface at the time of injection. Notably, microglia have been reported to be present within or immediately adjacent to the ventricular

zone during embryonic development¹². In addition, a small fraction of cells may be labeled along the injection tract. Accordingly, the microglial clones observed in our analysis are most likely labeled through infection of microglial cells that are already present in or adjacent to the ventricular zone and subsequently expand locally within the brain. Similar considerations likely apply to other non-neuroepithelial cell types identified in our dataset, such as vascular- or endothelial-associated cells, which may also be infected via exposure to virus in the ventricular space or along the injection site. The epithelial cells detected in our dataset are mostly choroid plexus epithelial cells which are located and can be infected within the ventricles.

This point is now explicitly noted in the revised manuscript.

“Labeling of microglia likely reflects infection of already colonizing cells that subsequently amplify locally within the brain¹², and similar considerations may apply to other non-neuroepithelial lineages such as vascular or endothelial cells.”

6. The section describing thalamic clonal relationships includes habenular neurons alongside ventral/dorsal thalamus. As the habenula is part of the epithalamus (within the diencephalon) rather than thalamus as strictly defined, can the authors please clarify anatomical labels and ensure consistency across text and figures (a brief note in Extended Data Fig. 6 or Methods would help).

Response:

We thank the reviewer for this helpful clarification regarding anatomical terminology. In our study, the term “thalamus” was used to describe the intended dissection target region based on the reference atlas from Allen Institute. However, due to the nature of the dissection, adjacent diencephalic regions, including the habenula (epithalamus), were also collected and included in the analysis. Similar considerations apply to the other dissected brain regions.

To avoid ambiguity, we have revised the text and figure legend to clarify this and ensure consistent anatomical labeling throughout the manuscript. Specifically, we have now clearly clarified in the main text and the figure legends for **Fig. 1b**, revised **Extended Data Figs. 10, 14, and 18** that “thalamus” includes the epithalamus and may also contain tissue from neighboring regions. These clarifications improve anatomical precision without affecting the interpretation of the clonal analyses.

As in main text:

“We harvested the brains at postnatal day 4 (P4), and dissected the tissue to enrich cells in the cortex (including corpus callosum; rostral and caudal separated), striatum, thalamus (including epithalamus), hippocampus, and the olfactory bulb (OB)”

“thalamic (including epithalamus) glutamatergic neurons emerge from progenitor cells located in the diencephalon surrounding the third ventricle”

As in Figure legend of Fig. 1b, Extended Data Figs. 10, 14, and 18:

“Dissection was performed with reference to the Allen Institute atlas. The thalamic sample included both thalamic and epithalamic tissue. The dissections may have included a small amount of tissue from adjacent regions.”

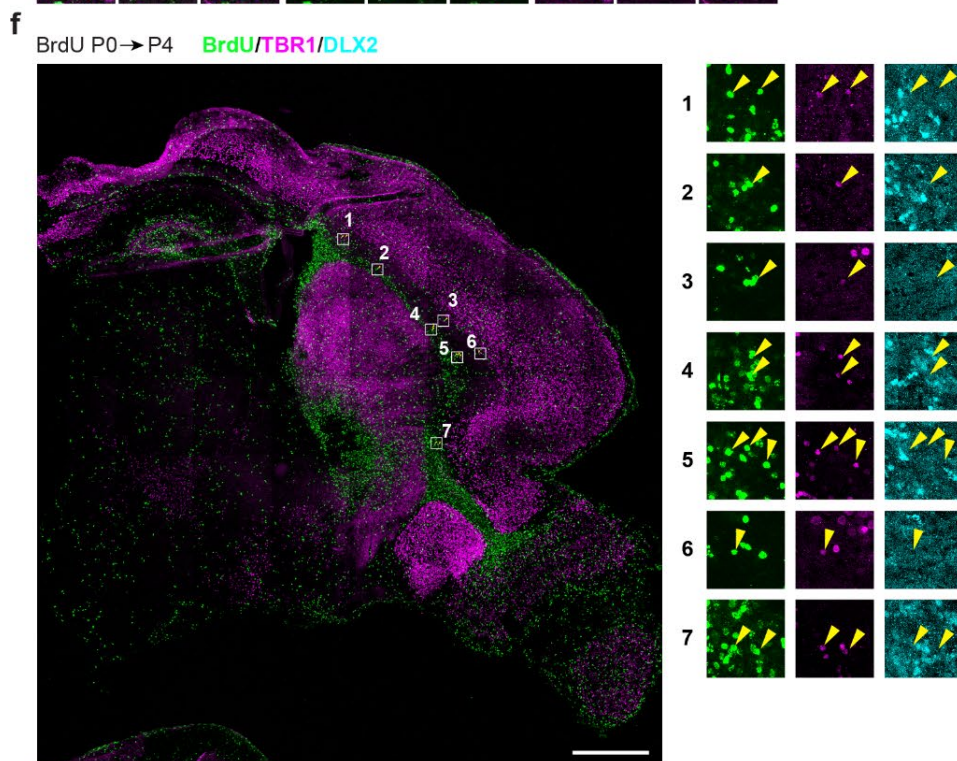
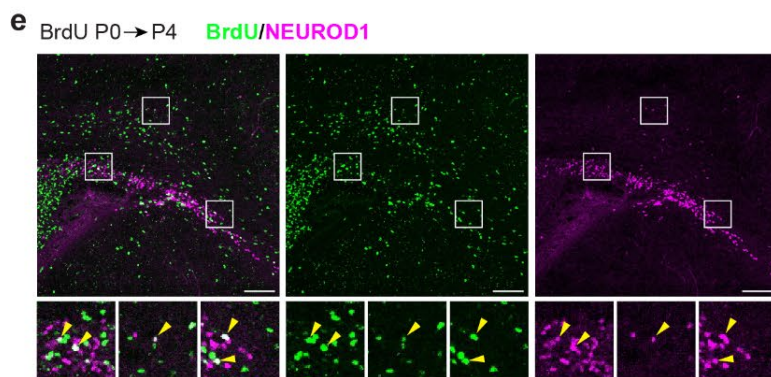
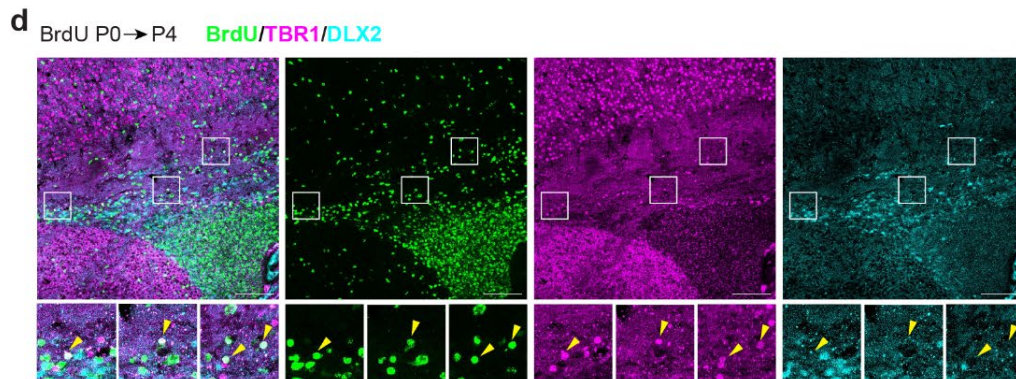
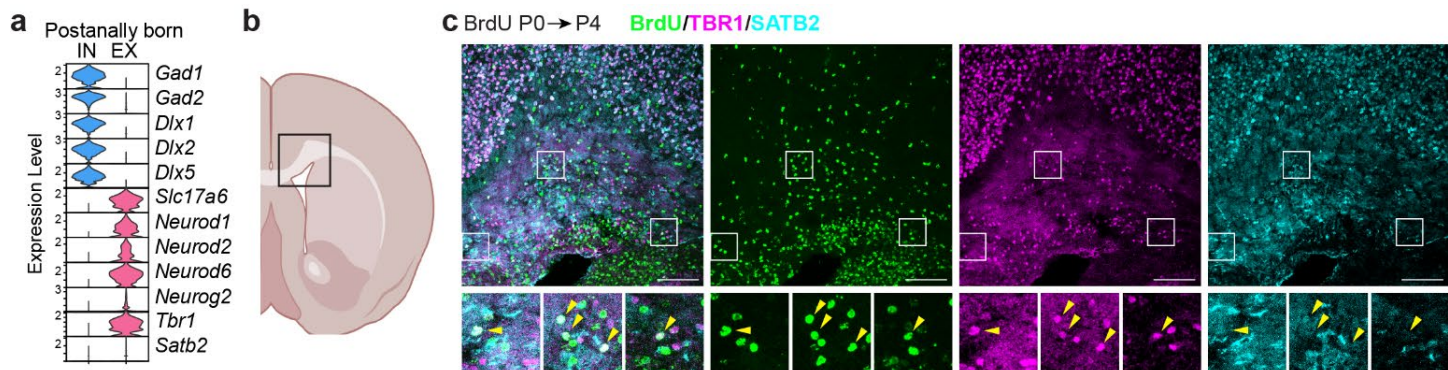
7. The finding that postnatally born glutamatergic neurons share clones with prenatally generated glutamatergic neurons and glia suggests continuity across the neurogenesis–gliogenesis transition. While postnatal neurogenesis is well established in the hippocampus, the evidence presented here for cortical postnatal glutamatergic neurogenesis appears novel and important, as cortical development is generally thought to shift entirely from neurogenesis to gliogenesis after birth. This seems novel and important; consider emphasizing it more explicitly in the Results and discuss it.

Response:

We thank the reviewer for highlighting this important and potentially novel observation. In response, we performed additional analyses and experiments to further define the identity of this population and discussed its possible fate. Specifically, we examined the transcriptomic identity of these postnatally born glutamatergic neurons using established molecular markers, in comparison with postnatally born GABAergic neurons (**Extended Data Fig. 11a**). These cells express canonical glutamatergic neuronal markers, including *Neurod1* and *Tbr1*, and lack inhibitory neuronal markers such as *Dlx2*, *Gad1*, and *Gad2*, supporting their classification as glutamatergic neurons. We further validated this population by immunostaining with BrdU applied at P0 and collected at P4, which confirmed the presence of BrdU⁺/TBR1⁺/DLX2⁻; BrdU⁺/NEUROD1⁺; BrdU⁺/TBR1⁺/SATB2⁻ cells (**Fig. 2i; Extended Data Fig. 11b-e**). Because our dataset was collected at P4, these neurons are mostly observed in the corpus callosum and are likely still migrating (**Extended Data Fig. 11f**), such that their final regional positions cannot yet be determined from the current data. Nevertheless, their clonal relationship with prenatally generated glutamatergic neurons and glia suggests continuity of lineage (**Figs. 2j,k**). We also note that this population may partially overlap with previously reported postnatally generated glutamatergic neurons derived from *Neurog2*⁺ progenitors that migrate to the olfactory bulb¹³ or cortex¹⁴. We have revised the manuscript to highlight this observation more explicitly and to discuss its potential developmental significance.

As in the main text:

*“In the cortex, postnatally generated glutamatergic neurons mapped predominantly to the white matter region above the dorsolateral ventricles (Fig. 2h; Extended Data Fig. 8). Gene expression analysis was performed to define the molecular features of this population, showing that these cells expressed glutamatergic neuronal markers, including *Slc17a6*, *Neurod1* and *Tbr1*, lacked *Satb2*, and were negative for inhibitory neuronal markers, including *Dlx2*, *Gad1* and *Gad2* (Extended Data Fig. 11a). We further validated this population by P0 BrdU labeling followed by P4 immunostaining, which confirmed the presence of BrdU⁺/TBR1⁺/DLX2⁻ cells, mainly in the white matter above the dorsolateral ventricles (Fig. 2i; Extended Data Fig. 11b-e). Analysis of sagittal sections showed that this population was distributed along the rostrocaudal axis. Because these neurons are still migrating, it remains unclear whether they subsequently migrate into the cortex or toward the OB (Extended Data Fig. 11f). They may partially overlap with previously described postnatally generated glutamatergic neurons^{13–15}. Clonal lineage analysis showed that, at P0 labeling, these cells shared clones with mitotic glial cells and GABAergic neurons migrating toward the OB (Fig. 2j,k). Notably, this population of postnatally glutamatergic neurons shared clonal relationships with prenatally generated glutamatergic neurons and glial cells in labeling of E11-E15, suggesting a continued lineage (Fig. 2j,k).”*



Extended Data Figure 11: Molecular identities of postnatally born glutamatergic neurons

a, Violin plots showing expression of canonical excitatory and inhibitory neuronal marker genes in postnatally born excitatory and inhibitory neurons isolated from dissected cortical regions.

b, Diagram highlighting regions enriched for postnatally generated glutamatergic neurons in the validation dataset.

c-e, Immunostaining validation of postnatally born glutamatergic neurons in representative coronal sections. BrdU was administered intraperitoneally at P0, and whole-mount analysis was performed at P4. Double-positive cells are indicated by arrows. Scale bar = 100 μ m.

f, Immunostaining validation of postnatally born glutamatergic neurons in representative sagittal sections. BrdU was administered intraperitoneally at P0, and whole-mount analysis was performed at P4. Double-positive cells are indicated by arrows. Scale bar = 1 mm.

8. The authors report CGE-derived clusters mostly generated at E12–E14, with MGE-derived clusters also within this window. Prior literature generally places MGE onset around E12 and CGE slightly later (~E14 onward). Could the authors clarify whether your data support an earlier onset of CGE-derived neurogenesis, or whether this may reflect clustering or labeling differences?

Response:

We appreciate the reviewer's careful reading and helpful suggestions. In our study, progenitors were labeled by lentiviral delivery at discrete embryonic time points, and lineage origin was inferred retrospectively from transcriptional identity. We do not directly observe the anatomical site of origin of individual neurons, and our assignments of MGE- versus CGE-derived lineages are based on established molecular markers and transcriptomic similarity. In addition, because lentivirus can integrate into the genome in the absence of cell division, it may label a progenitor that divides several days later, distinguishing this approach from classical BrdU birthdating methods.

Within this framework, our data do not necessarily support an earlier-than-expected onset of CGE-derived neurogenesis. The apparent overlap in labeling windows likely reflects the temporal resolution of the lentiviral labeling strategy. Consistent with the established model, MGE-associated lineages are more prominent at earlier labeling time points, whereas CGE-associated clusters are enriched at later stages, in line with the general view that MGE neurogenesis precedes CGE neurogenesis. We have revised the text to change the interpretation of "generation" into "labeling" and note the labeling of CGE neurons at later stages.

Regarding Comment 10 of reviewer #2, we also revised the order of description, as in main text:

"In the cortex, MGE-derived clusters were fate-biased (160 multicellular clones), transcriptomically distinct (e.g., Npy^+ , $Prox1^+$, Sst^+), and labeled mainly in E12-E14. Notably, Npy^+ neurons were not found to share clones with Sst^+ neurons. While CGE-derived clusters were mostly labeled between E12 and E14, one cluster extended into later stages."

9. Please clarify the x-axis in Fig. 3e (e.g., "clone size" or "number of cells per clone") for the cumulative proportion plot to aid interpretation.

Response: We thank the reviewer for this suggestion. We have updated the revised **Fig.3j** to clearly label the x-axis as "*number of cells per clone*" to aid interpretation.

10. Across Extended Data Figs. 6, 8, and 10, references to clusters and clonal groups are difficult to reconcile with the figures due to dense text annotations and lack of a clear color legend. Consider adding a color annotation with legend or other graphical markers (e.g., explicit labels for triangles in half-heatmaps) to map the described groups directly onto the plots.

Response:

We thank the reviewer for this helpful suggestion. To improve clarity and facilitate direct mapping between the text and the figures, (i) we removed the redundant text in the labeling in the figures. For example, we remove the region in the cluster names of revised **Extended Data Figs. 10, 14, and 18**, since there is already a regional annotation. In addition, we optimized the labeling of cluster names in **Fig. 2a-b, Fig. 3a-b, and Fig. 4a-b** to change the region names into abbreviations. (ii) We have added explicit color annotations with corresponding legends to revised **Extended Data Figs. 10, 14, and 18**. Similarly, we added color annotations of cell types in **Fig. 1d, Fig. 4h, Extended Data Fig. 17** to avoid dense text annotations and aid interpretation. (iii) In addition, we added clear labels to the triangles in revised **Extended Data Figs. 10, 14, and 18**. Moreover, we revised the order of descriptions in the main text to match the label order, making it easier to link each group in the text to the corresponding plot.

One instance is:

“Striatal GABAergic neurons were divided into two major clonal groups. The first group was enriched at E14 and comprised LGE-derived neurons in the striatal core. The second arose between E12 and E16, persisted until P0, and was transcriptomically similar to migrating OB neurons (Extended Data Fig. 14, Extended Data Fig. 8).”

These changes reduce reliance on dense text annotations and make the correspondence between described groups and plotted data more immediately interpretable.

11. The identification of $Meis2^+Sp8^+$ dispersed interneurons is striking — could their clonal analysis be linked to embryonic single-cell datasets to infer developmental origin more precisely?

Response:

We thank the reviewer for this insightful suggestion. Because available embryonic reference dataset¹⁶ with clearly dissected CGE, MGE, and LGE annotations were generated using different chemistries and from early embryo stages, direct integration with our dataset would likely be confounded by both batch effects and dominant differentiation-state differences. Instead, we integrated marker sets from embryonic CGE-, MGE-, and LGE-dissected reference dataset¹⁶ together with well-established regional markers to infer the developmental origin. We visualized these canonical regional markers (**Extended Data Figs. 12a**) and performed module-scoring analyses (**Extended Data Figs. 12b,c**) using marker sets to assess correspondence between interneuron clusters and developmental origins.

These analyses suggest that the $Meis2^+Sp8^+$ population is derived predominantly from LGE lineages (**Extended Data Figs. 12c**), providing additional support for its inferred developmental origin. At the same time, some clusters could not be clearly separated between CGE- and LGE-like identities (**Extended Data Figs. 12c**). This ambiguity may partly reflect the inclusion of P0-born immature neurons in our dataset, as SVZ-derived immature neurons can retain transcriptional features that overlap with CGE/LGE developmental programs. The analysis also helped refine the annotation of several previously ambiguous clusters. For

conceptual advances are important. Our comments above are intended to help strengthen the clarity and impact of the presentation — the overall quality of the work is excellent.

Response:

We sincerely thank the reviewer for the positive overall assessment of our work. We are grateful for the recognition of both the conceptual advances and the value of the dataset as a community resource. We also appreciate the thoughtful suggestions provided above, which have helped us improve the clarity and impact of the presentation.

Referee #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Response:

We thank the reviewer for co-reviewing this manuscript and greatly appreciate their time and effort.

Referee #4 (Remarks to the Author):

This study uses frontier and state of the art technologies aimed at an interesting topic. However from the data shown it remain unclear whether this study goes beyond what is already known using more classical tools. The results shown appear to be mostly confirmatory, not only in the cortex but also in striatum and thalamus. In particular it is unclear whether the new data bring more insights than previous data reported for thalamus and striatum using more conventional technologies (see for instance Kelly Neuron 2018; Wong PLOS biol 2028)?

What is particularly missing here is a more in depth analysis with much higher spatial resolution, which would be required to disentangle clonal relationships in subregions of interest in the described regions, in particular thalamus (dorsal / ventral - all thalamic nuclei) and basal forebrain (including subregions of striatum, GP, CP ..). Indeed these important aspects of lineage relationships remain much less known (although for thalamus see recent report of Puche aroca 2025).

Response:

We appreciate the reviewer's concern that several broad lineage principles are consistent with prior work. We agree that concordance with previous studies is an important benchmark for a resource dataset. In revision, we substantially expanded the spatially resolved analysis to determine whether our integrated clonal, transcriptomic, temporal, and MERFISH framework could reveal lineage relationships not accessible from broader regional annotations. This analysis mapped 416 clusters to anatomical subregions and uncovered several subtype- and subregion-specific findings, including prolonged neurogenesis of ventral D2 MSNs, late clonal decoupling of striatal MSNs and astrocytes, distinct OB lineage relationships of late-born striatal neurons versus astrocytes, region-specific temporal differences in astrocyte–OPC coupling, and hippocampal subregion-specific neuron–glia lineage relationships.

In the striatum, the higher-resolution annotation resolved MSN and interneuron populations with clear subregional specificity, including clusters enriched in distinct dorsal and ventral compartments (**Fig. 3d**,

Extended Data Fig. 15a). Re-analysis at this level showed that MGE-derived interneuron populations and LGE-derived MSN populations were clonally distinct, providing additional support for both our lineage analysis and cell-type annotation, while MSN populations across striatal subregions remained clonally related overall but differed in temporal production dynamics (**Fig. 3e**). For example, D3 MSNs enriched in the olfactory tubercle were recovered predominantly from early labeling stages and were not appreciably generated later, whereas D1 MSNs were primarily recovered from E11–E15 labeling, consistent with previous studies, and ventral D2 MSNs showed a more prolonged neurogenic window than dorsal D2 MSNs (**Fig. 3e**). Taking advantage of the temporal coverage of the dataset, we further asked whether this prolonged neurogenesis coincided with astrogenesis by examining clonal relationships between MSN subtypes and striatal astrocytes labeled either at E11–E15 or at E16–P0 (**Fig. 3f**). Notably, late-born neurons and late-generated astrocytes after E16 were largely clonally separated (**Fig. 3f**). To further characterize these populations, we performed differential gene expression analysis and identified genes preferentially enriched in early- versus late-generated cells, including *Sox11* and *Nnat* in late-born MSNs (**Extended Data Fig. 15b**). Functional analyses suggested that these populations occupy distinct cellular states (**Extended Data Fig. 15c**), and overlap analysis with neurodevelopmental disorder gene sets showed that late-born MSNs were more enriched for NDD/ASD-associated genes¹⁷ (**Extended Data Fig. 15d**). These analyses provide additional temporal, molecular, and subregional resolution beyond what was accessible in previous targeted studies.

In the olfactory bulb, mapping interneuron clusters to subregional locations resolved three broad spatial groups: GL, superficial GCL, and deep GCL (**Extended Data Fig. 16a**). This analysis revealed a broad temporal gradient, with more superficial populations generated earlier than deeper populations, while also uncovering substantial heterogeneity within each subregion. GL populations were labeled predominantly during early embryonic development, superficial GC populations were generated mainly during embryonic stages, and deep GC populations included subtypes produced primarily in late embryonic stages as well as others generated largely postnatally (**Extended Data Fig. 16b**). At the same time, substantial temporal heterogeneity was evident within subregions (**Extended Data Fig. 16c**). For example, superficial GC contained clusters generated predominantly before E16 as well as others enriched after E16, whereas deep GC included one cluster labeled mainly before E16, five clusters labeled predominantly after E16, and four clusters generated mainly postnatally (**Extended Data Fig. 16c**). Finer subdivision also exposed heterogeneous clonal relationships within subregions (**Extended Data Fig. 16d-e**), as illustrated by the subtype-biased clonal patterns associated with OB:IN_Meis2_Sp8_9 and OB:IN_Meis2_Sp8_19. We further examined clonal relationships between OB interneuron clusters and striatal cell types (**Extended Data Fig. 16f**). GL clusters showed less clonal sharing with striatal interneurons than GCL. Heterogeneous patterns across specific OB and striatal subtypes were detected (**Fig. 3f, Extended Data Fig. 16f**). D1 MSNs showed a lower proportion of shared clones with OB inhibitory neurons, whereas both D1 and D2 MSNs shared more clones with deep GCL neurons than with superficial GCL neurons, consistent with previous findings⁶. Notably, an *Oprm1*-positive MSN cluster showed the highest proportion of shared clones with OB inhibitory neurons (**Fig. 3f**). These findings show that higher-resolution spatial and transcriptomic subdivision reveals lineage structure in the OB that is not apparent at broader levels of analysis.

In the thalamic regions, we mapped astrocyte and OPC clusters across thalamic and epithalamic territories and found that clonal coupling between epithalamic and thalamic astrocytes was strongest at early labeling stages and markedly reduced after E16 (**Extended Data Fig. 19a,b**). Likewise, thalamic astrocyte–OPC clonal coupling was primarily associated with early embryonic labeling, with minimal sharing after E16 (**Extended Data Fig. 19a,b**). Notably, this contrasts with the cortex, where astrocyte–OPC clonal coupling persisted later, including after birth (**Extended Data Fig. 19c,d**). We further found that cortical OPC–astrocyte clone sharing remained detectable after P0 (**Extended Data Fig. 19d**). However, the OPCs involved in this postnatal clonal coupling were located predominantly in white matter regions (**Extended Data Fig. 19c**), making it unclear whether they represent cells migrating into the cortex or a distinct white matter-associated population.

Together, these analyses reveal previously unresolved regional differences in the temporal organization of gliogenic lineage relationships.

In the hippocampus, assigning astrocyte, OPC, and neuronal clusters to specific subregions and layers of CA1, CA3, and DG revealed subtype-specific lineage patterns (**Extended Data Fig. 20**). Notably, only one OPC cluster shared clones with neurons from all three hippocampal regions, suggesting transcriptomic convergence of OPCs arising from distinct lineage origins (**Fig. 4g**), whereas most OPC clusters shared clones primarily with CA1 neurons, indicating that the majority of hippocampal OPCs are linked to the CA1 lineage (**Fig. 4g**). After E16, most OPCs no longer shared clones with neurons, consistent with lineage divergence between neurogenesis and gliogenesis (**Fig. 4g**). In DG, this likely reflects its prolonged neurogenic window, whereas in CA1 it is consistent with a developmental switch toward gliogenesis. Interestingly, astrocyte–neuron clonal coupling persisted longer, particularly in DG, where DG-localized astrocyte clusters continued to share clones with DG neurons (**Fig. 4g**). In addition, We found that hippocampal CR cells shared clones with one ependymal-like astrocyte cluster and with hippocampal ependymal cells, whereas no comparable clonal coupling was detected in the cortex (**Extended Data Figs. 6d-f**). Together, these analyses reveal additional subregional lineage relationships in the hippocampus that were not evident in the broader analysis.

Overall, we therefore believe that this study extends previous work by providing a unified lineage resource with molecular, temporal, and spatial resolution sufficient to resolve subregional and subtype-specific developmental relationships across the forebrain. We are grateful to the reviewer for this critical suggestion, which significantly strengthened the manuscript.

We added these new results into our revised main text, figures, and supplementary figures as follows:

“Spatial mapping resolved striatal clusters into medium spiny neurons (MSN) and interneurons with dorsal–ventral subregional specificity (Fig. 3d, Extended Data Fig. 15a). MSN populations across striatal subregions remained broadly clonally related, but differed in their temporal production dynamics (Fig. 3e). For example, D1 MSNs were mainly recovered from E11–E15 labeling, whereas D2 MSNs were continuously labeled, consistent with previous studies⁶. Notably, ventral D2 MSNs showed more prolonged neurogenesis than dorsal D2 MSNs (Fig. 3e), while olfactory tubercle-enriched D3 MSNs were recovered predominantly from early labeling stages. We also identified an Oprm1 MSN population⁹, marked by the mu-opioid receptor associated with opioid-related reward behaviors, which was labeled at both early (E11–E15) and late (E16–P0) stages. Leveraging the temporal coverage of our dataset, we next asked whether the prolonged MSN neurogenesis coincided with astrogenesis by examining clonal relationships between MSN subtypes and striatal astrocytes labeled at E11–E15 or E16–P0. Interestingly, late-born MSNs and striatal astrocytes labeled after E16 were largely clonally separated (Fig. 3f), suggesting that the MSN and astrocyte lineages in the striatum diverge after E16.”

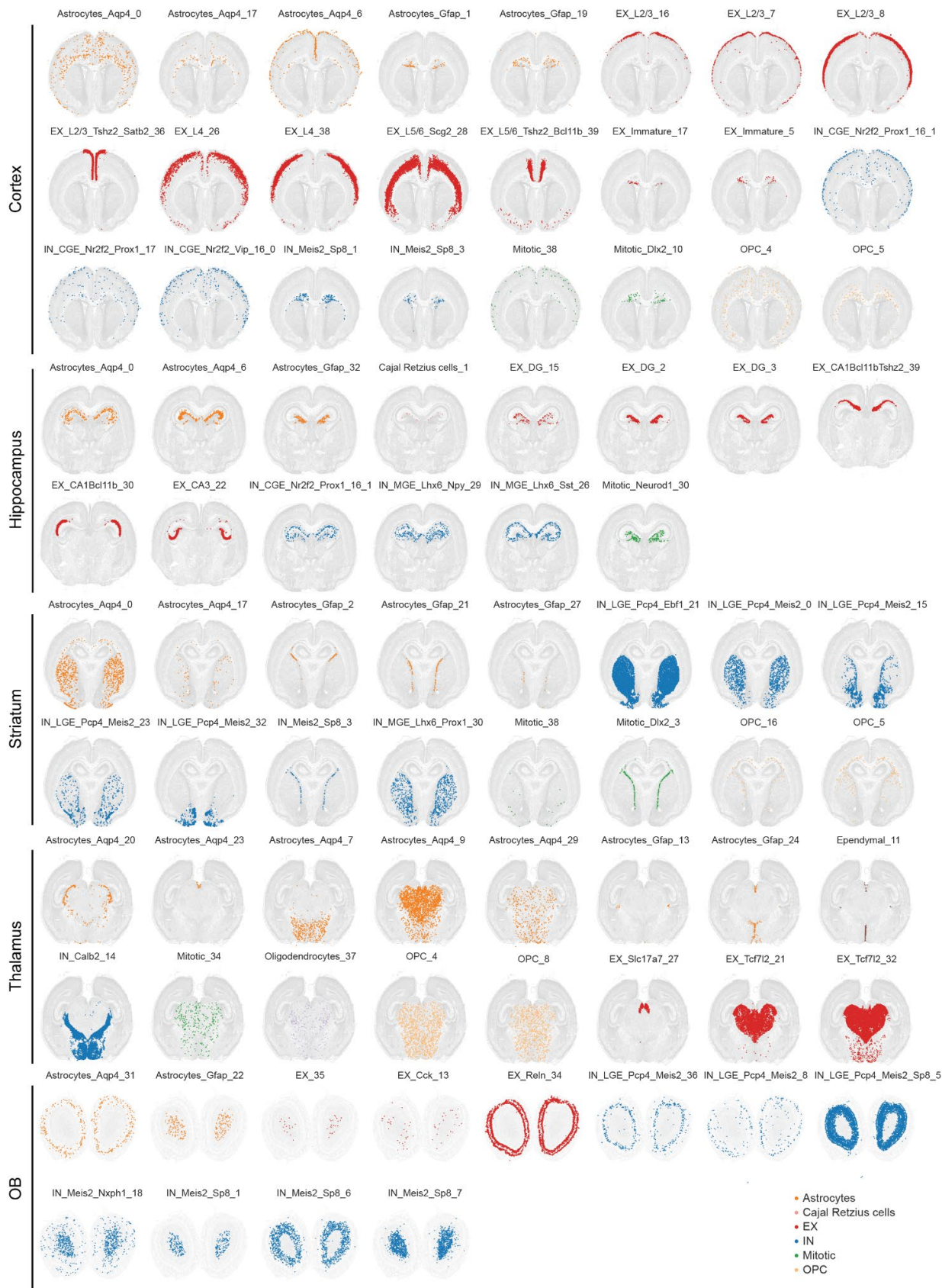
“Mapping OB interneuron clusters to subregional locations resolved three broad spatial groups: glomerular layer (GL), superficial granule cell layer (GCL), and deep GCL (Extended Data Fig. 16a). These results suggest that OB cell populations at P4 emerge according to a temporal order: superficial populations are generated first, whereas deeper populations emerge subsequently (Extended Data Fig. 16b). Within each anatomical domain, multiple transcriptomic subpopulations of OB interneurons with heterogeneous clonal patterns can be identified (Extended Data Fig. 16c-e). Extending the analysis to striatal populations, GL clusters showed less clonal sharing with striatal interneurons than GCL clusters. In contrast, most OB GCL clusters shared clones with MSNs, especially at E11–E15 labeling, with marked heterogeneity across specific OB and striatal subtypes (Extended Data Fig. 16f).”

“Analyzing the clones shared between striatal and OB GABAergic neurons, notably, we found D1 MSNs showed a lower proportion of shared clones with OB inhibitory neurons than D2 MSNs, whereas

both D1 and D2 MSNs shared more clones with deep GCL than with superficial GCL neurons, consistent with previous findings¹⁰. Notably, the Oprm1 MSN cluster showed a significantly higher proportion of clone sharing with OB inhibitory neurons than other MSN clusters (~ 3 fold) (Fig. 3g). Unlike D1 and D2 MSNs, Oprm1 MSNs shared clones with superficial and deep GCL neurons at comparable proportions at early labeling (E11–E15), but biased towards deep GCL neurons at late labeling (E16–P0). Given the clonal separation between MSNs and striatal astrocytes after E16, we next asked whether these striatal populations were linked to the same OB populations. Both late-labeled MSNs and astrocytes shared clones with deep GCL neurons; however, only MSNs shared clones with GL neurons, whereas only astrocytes shared clones with superficial GCL neurons. These results suggest that late striatal neurogenesis and astrogenesis are associated with distinct, although partially overlapping, OB lineage relationships.”

“Higher-resolution spatial mapping further revealed subregion-specific lineage relationships. For example, astrocytes in the thalamus and epithalamus shared clones at early stages, but these shared clones were markedly reduced after E16 (Extended Data Fig. 19a,b). Thalamic astrocytes and OPCs shared clonal barcodes at early embryonic stages, with minimal clonal coupling observed after E16 (Extended Data Fig. 19b). In contrast, cortical astrocyte–OPC clonal coupling persisted later, including after birth, and cortical OPC–astrocyte clone sharing remained detectable after P0, although the OPCs in these shared clones were predominantly located in white matter, making it unclear whether they represent cells migrating into the cortex or a distinct white matter-associated population (Extended Data Fig. 19c,d). In the hippocampus, assigning astrocyte, OPC, and neuronal clusters to specific subregions and layers of CA1, CA3, and DG revealed subtype-specific lineage patterns (Extended Data Fig. 20). Interestingly, most OPC clusters shared clones primarily with CA1 neurons, indicating that the majority of hippocampal OPCs are linked to the CA1 lineage, whereas one OPC cluster shared clones with neurons across CA1, CA3, and DG, suggesting transcriptomic convergence from distinct lineage origins (Fig. 4g). After E16, OPC–neuron clonal sharing was largely lost, consistent with divergence of neurogenesis and gliogenesis. By contrast, astrocyte–neuron clonal coupling persisted longer, particularly in DG, where DG-localized astrocyte clusters continued to share clones with DG neurons (Fig. 4g).”

“Notably, hippocampal CR cells were found sharing clonal lineage relationships with astrocytes and ependymal cells (Fig. 1d, Extended Data Fig. 6d-e), which were not detected in cortical CR cells (Extended Data Fig. 6f).”



Extended Data Figure 8: MERFISH spatial maps showing the subregional distribution of representative clusters. This figure provides anatomical context for cluster annotations and supports interpretation of region- and subregion-resolved clonal analyses in the main figures.

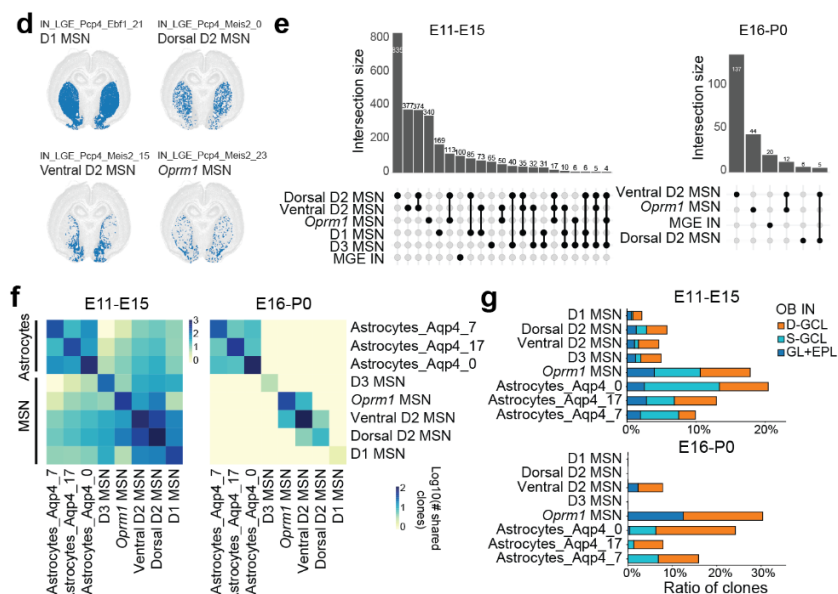


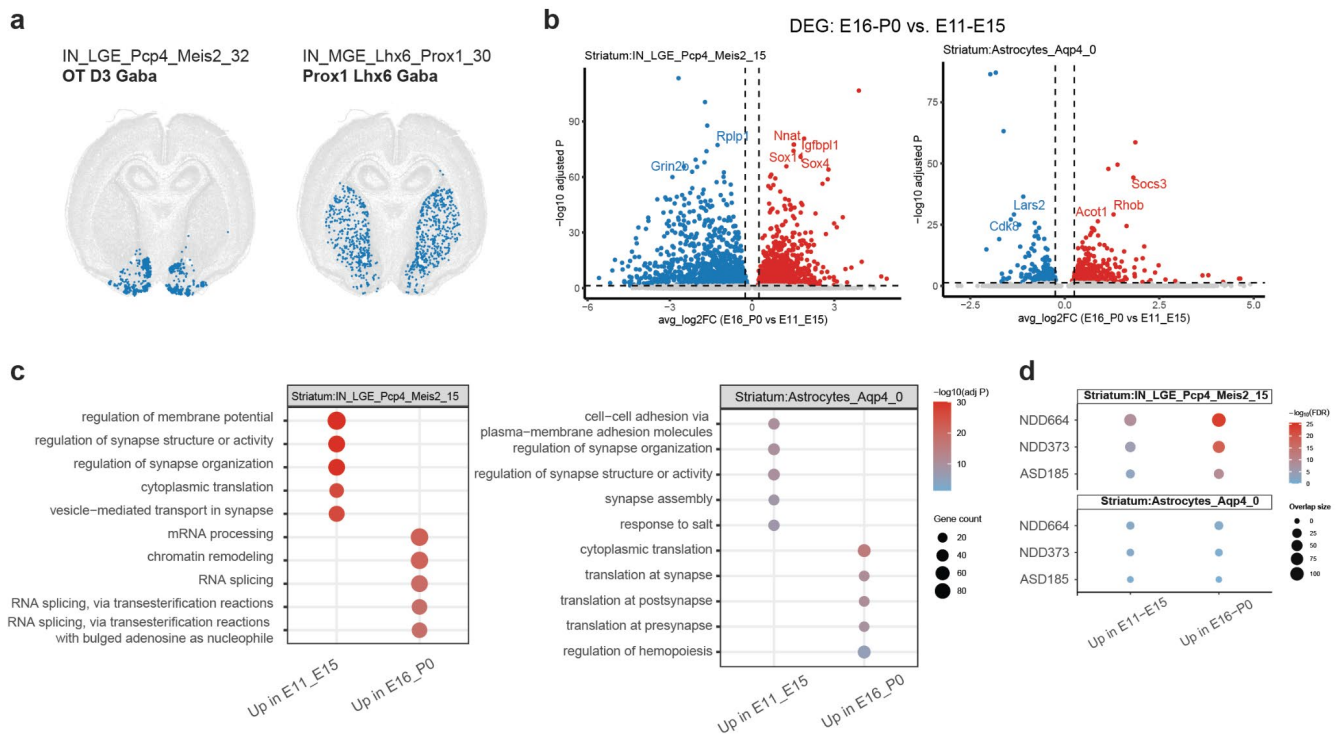
Figure 3: Regional specific and regional dispersed clonal specification of GABAergic neurons

d, MERFISH spatial maps showing the distribution of GABAergic neuron subtypes across striatal subregions. This spatial analysis resolves MSN and interneuron populations with dorsal, ventral, and olfactory-tubercle-associated distributions.

e, UpSet plots showing shared-clone counts among striatal GABAergic neuron subtypes. Counts were calculated from pooled E11–E15 injections (left) or E16–P0 injections (right). Comparing early and late labeling windows tests whether clonal relationships among striatal GABAergic subtypes change over developmental time.

f, Heatmap showing clonal coupling between striatal MSN and astrocyte clusters. Values represent shared-clone counts. Diagonal values indicate the number of clones containing each cell type. Cell-type pairs with at least two shared clones are shown. This panel highlights late-generated striatal neuronal populations becoming clonally decoupled with astrocyte lineages.

g, Proportion of striatal GABAergic-neuron and astrocyte clusters sharing clones with OB GABAergic neurons across the categories defined in Extended Data Fig. 16a.



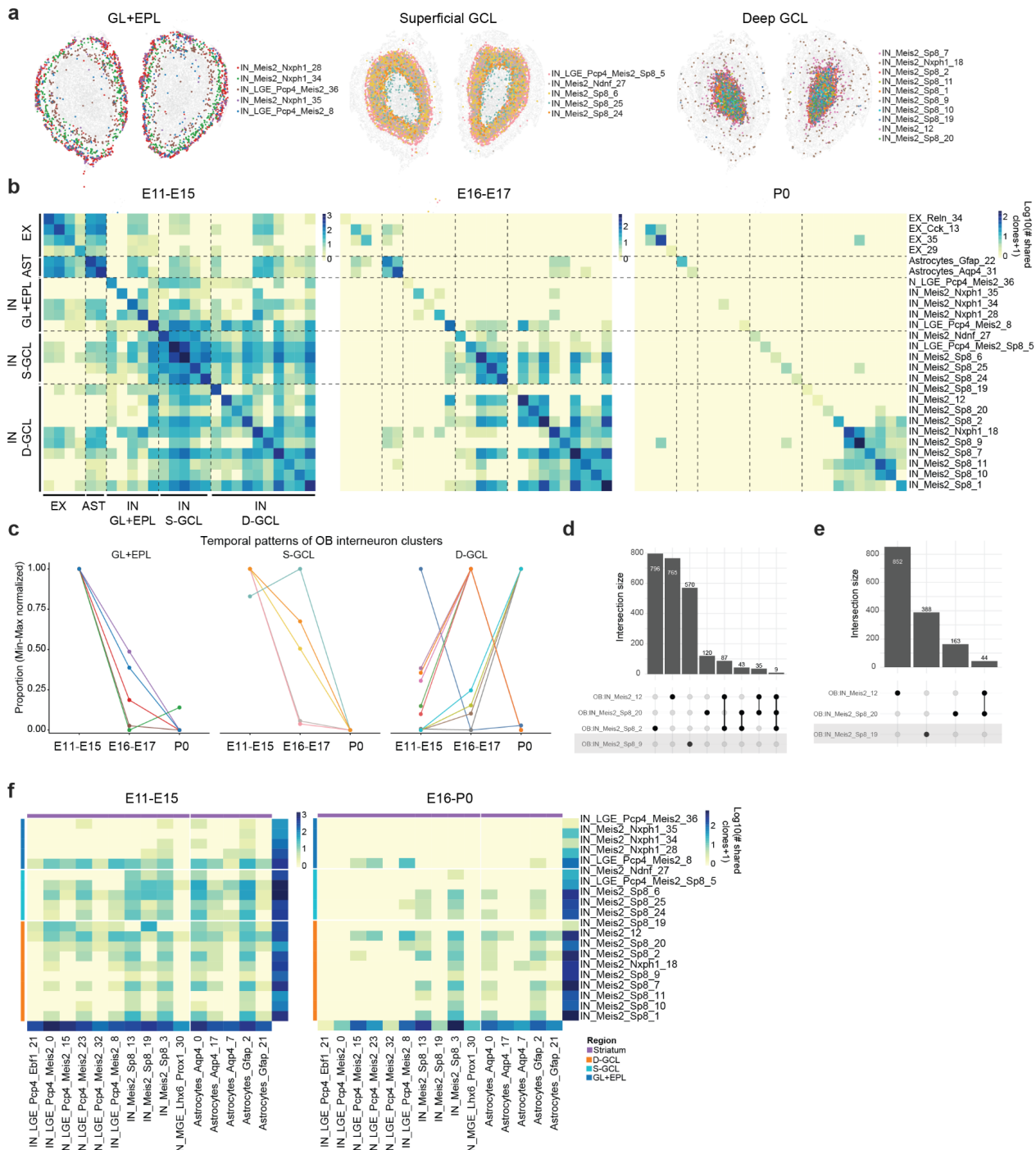
Extended Data Figure 15: Subregion-resolved analysis highlights late clonal decoupling between striatal neurons and astrocytes

a, MERFISH spatial maps showing the distribution of different GABAergic neuron subtypes across distinct striatal subregions.

b, Volcano plots showing differentially expressed genes between cells labeled at E16–P0 and E11–E15 within the IN_LGE_Pcp4_Meis2_15 and Astrocytes_Aqp4_0 clusters.

c, Gene Ontology biological process enrichment analysis of genes enriched in E16–P0- or E11–E15-labeled cells within the IN_LGE_Pcp4_Meis2_15 and Astrocytes_Aqp4_0 clusters.

d, Enrichment analysis of differentially expressed genes between E16–P0- and E11–E15-labeled cells within the IN_LGE_Pcp4_Meis2_15 and Astrocytes_Aqp4_0 clusters against ASD- and NDD-associated gene sets¹⁷. Enrichment was assessed using Fisher's exact test, and P values were adjusted by the Benjamini–Hochberg method.



Extended Data Figure 16: Subregion-resolved analysis in OB

a, MERFISH spatial maps showing the distribution of different GABAergic neuron subtypes across distinct OB subregions.

b, Heatmap showing clonal coupling between OB GABAergic neuron clusters. Color indicates shared-clone counts. Diagonal values indicate the number of clones containing each cell type. Cell-type pairs with at least two shared clones are shown. OB GABAergic neurons collected within the OB from the P0 labelling showed very limited clone sharing with other regions, including the striatum and cortex (Fig. 1d). This suggests that many OB GABAergic neurons captured at P4 (showing here) were born embryonically or locally generated, whereas postnatally generated neurons migrating toward the OB through the RMS had largely not yet reached the OB at P4.

c, Proportion of cell clusters across temporal windows within each OB subregion. Proportions were calculated relative to all OB GABAergic neurons and min-max normalized for visualization. This panel summarizes temporal differences in subtype abundance across OB subregions.

d, Heatmap showing clonal coupling between OPC and astrocyte clusters in the cortex. Color indicates shared-clone counts. Diagonal values indicate the number of clones containing each cell type. Cell-type pairs with at least two shared clones are shown.

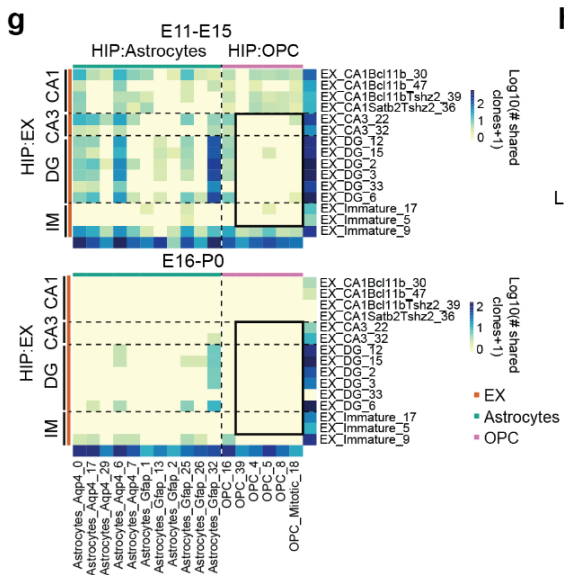
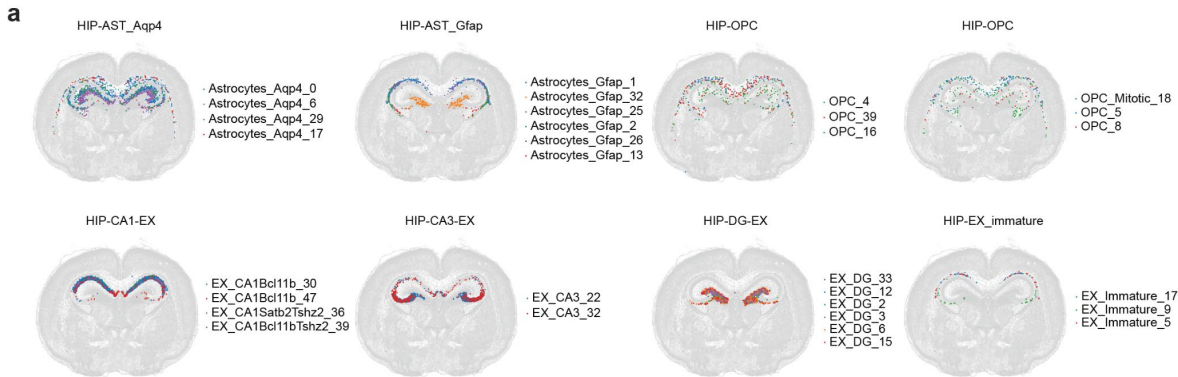
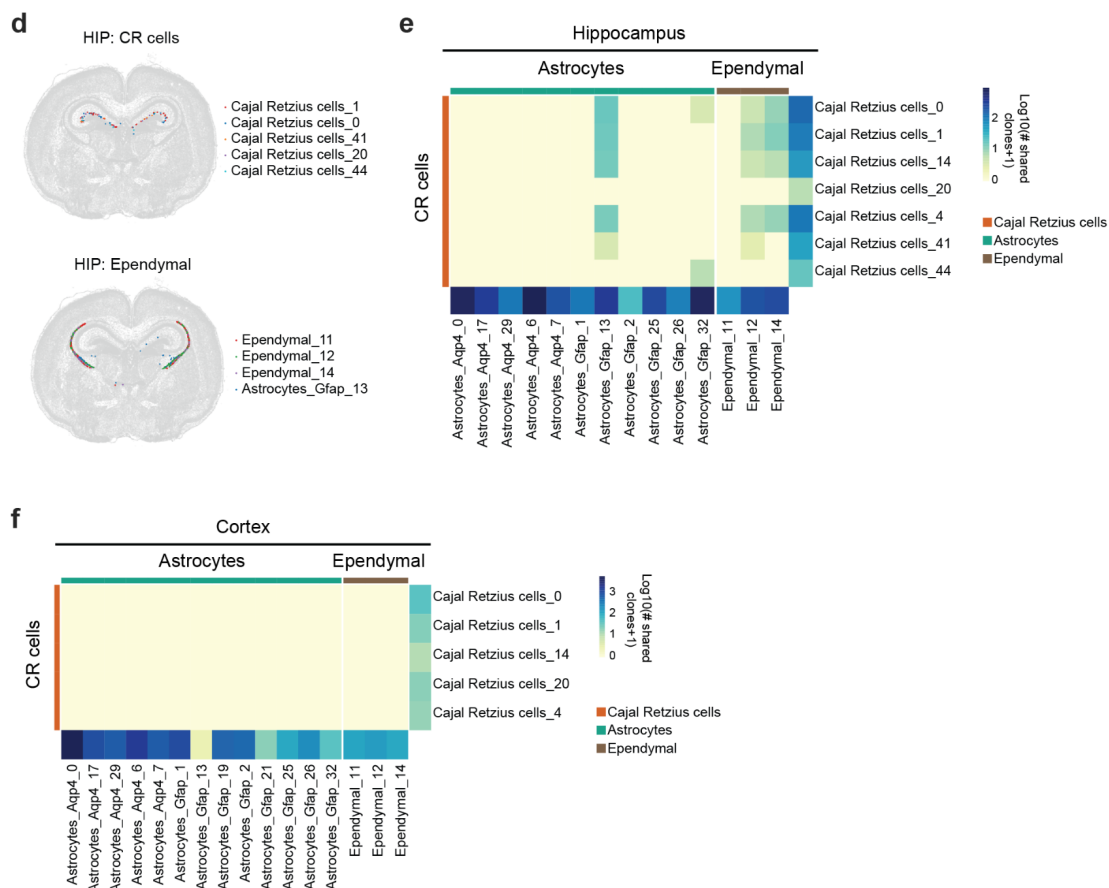


Figure 4g, Heatmap showing clonal coupling between hippocampal clusters. Color indicates shared-clone counts. Right and bottom color bars indicate the number of multicellular clones containing each cell type. Cell-type pairs with at least two shared clones are shown. Temporal windows indicate injection timing.



Extended Data Figure 6: Lineage analysis of CR cells

d, MERFISH spatial maps of representative CR cell clusters (top) and astrocyte and ependymal cell clusters (bottom). Cells are colored by cluster identity.

e, Heatmap showing shared-clone counts between hippocampal CR cell clusters and astrocyte or ependymal cell clusters. Color indicates shared-clone counts. Right and bottom color bars indicate the number of multicellular clones containing each cell type. Cell-type pairs with at least two shared clones are shown. Data is pooled across all injection time points.

f, Heatmap showing shared-clone counts between cortical CR cell clusters and astrocyte or ependymal cell clusters. Color indicates shared-clone counts. Right and bottom color bars indicate the number of multicellular clones containing each cell type. Cell-type pairs with at least two shared clones are shown. Data is pooled across all injection time points.

The lack of description of the data for other structures of the forebrain such as hypothalamus and amygdala is also crucially missing here.

Response:

We thank the reviewer for bringing up this important point. In this study, we focused our analyses on five forebrain regions (cortex, thalamus, hippocampus, striatum, and olfactory bulb) and therefore did not collect hypothalamus or amygdala in this dataset. We have clarified this anatomical scope in the Abstract and Discussion and now explicitly note that future extensions incorporating additional forebrain regions, including hypothalamus and amygdala, will further expand the value.

“This atlas covers five regions collected in our dataset, and additional forebrain structures not sampled here may be incorporated in future work to further expand its anatomical scope.”

Moreover the description and presentation of the data is also often too succinct (very heavy figures and very limited legend explanations), making it difficult to grasp the data and understand the logic behind the proposed interpretations.

Response:

We thank the reviewer for this helpful feedback, which prompted us to improve both the visual presentation and explanatory legends throughout the manuscript. Specifically, we reduced visual density in several figures by simplifying annotations and improving layout. As in our response to Comment 10 of Reviewer #2, (i) we removed the redundant text in the labeling in the figures. For example, we remove the region in the cluster names of revised **Extended Data Figs. 10, 14, and 18**, since there is already a regional annotation. In addition, we optimized the labeling of cluster names in **Figs. 2a-b, Figs. 3a-b, and Figs. 4a-b** to change the region names into abbreviations. (ii) We have added explicit color annotations to revised **Extended Data Figs. 10, 14, and 18**. Similarly, we added color annotations of cell types in **Fig. 1d, Fig. 4h, Extended Data Fig. 17** to avoid dense text annotations and aid interpretation. (iii) In addition, we added clear labels to the triangles in revised **Extended Data Figs. 10, 14, and 18**. We also revised the order of descriptions in the main text to match the label order, making it easier to link each group in the text to the corresponding plot.

In addition, we added new main figure panels (**Figs. 3d–i and 4g**) and Extended Data figures (**Extended Data Figs. 6d–f, 15, 16, 19, and 20**) to better illustrate the regional lineage analyses underlying our major conclusions.

Finally, we made our P4 MERFISH dataset available through the UCSC Cell Browser, enabling readers to explore the spatial distribution of cell types and clusters and to interpret these spatial patterns together with the single-cell lineage dataset. Together, these revisions make the data presentation clearer and the logic underlying our interpretations more transparent.

On the other hand, the study has the ambition to become a landmark resource, it would be important to make the data more readily available online using a user-friendly interface, as usually done remarkably by the Allen Institute.

Response:

We agree that a landmark resource is most valuable when it can be readily explored by the community. To facilitate this, we made the dataset available through a user-friendly, web-based interface built on the UCSC Cell Browser, allowing interactive query and visualization without the need for custom code. Specifically, (i) we have organized our single-cell clonal dataset in a UCSC Cell Browser portal that supports searchable cell type and region annotations on UMAP, (ii) gene-level visualization and marker inspection, and (iii) direct download of processed matrices and metadata, including single cell identity and clonal information. In addition, (iv) We have also made our P4 MERFISH dataset available through the UCSC Cell Browser, enabling readers to explore the spatial distribution of cell types and clusters and to interpret these spatial patterns together with the single-cell dataset.

“An interactive browser of single-cell data and MERFISH data can be found at the University of California, Santa Cruz (UCSC) cell browser <https://p4-brain-lineage.cells.ucsc.edu>.”

Below are more specific comments related to technical issues that have to be solved and/or clarified.

- From Fig.1 and associated suppl data it seems like there are relatively few glutamatergic neurons in the dataset, compared with GABAergic cells (there seem to be more GABA-ergic cells in thalamus and cortex for instance). This points to a bias in the cell or transcriptome collection that could mislead the interpretation of the data.

Response:

We thank the reviewer for this comment and agree that glutamatergic neurons appear relatively underrepresented in some regions. As our response to the Comment 2 of Reviewer #2. The relatively lower apparent proportion of glutamatergic neurons likely reflects a combination of biological and technical factors. First, the cells shown are virus-labeled EGFP+ cells rather than all cells present in the corresponding anatomical region at P4, and our labeling scheme may not fully capture cells generated at the earliest developmental stages. Second, excitatory neurons are generally more mature and possess longer projections, which likely makes them more vulnerable to loss during dissociation and downstream processing, leading to their relative underrepresentation among recovered cells. We have included a brief discussion of this point in the figure legend of **Extended Data Fig. 4b** and clarified that the values should not be used to derive values of absolute excitatory-to-inhibitory ratios.

“The proportions were calculated by normalizing the count of each cell type to the total number of cells across all cell types within each region. These proportions were derived from the GFP+ cells we collected and not the absolute cell-type ratios in the brain.”

- The authors describe fate restricted clusters of upper layer neurons in the cortex that could relate to previously described clonal analyses using Cux2-Cre lineage that had remained controversial. This is potentially interesting but the data shown should be explicated more clearly: to which time-points of viral injection do they correspond? If they consist of late generated clones the interpretation would be quite different.

Response:

We thank the reviewer for raising this important point. We agree that the viral labeling time points are important for interpreting these upper-layer neuronal clones. To address this, we analyzed the number of multicellular clones contributing to these clusters across different viral injection time points. The upper-layer neuronal clusters described in the manuscript contained clones from both early and late labeling, but the majority were derived from later injections. Specifically, Cortex:EX_L2/3_23 and Cortex:EX_L2/3_26 contained 207 multicellular clones in total, of which 17 clones were recovered from early E11–E13 labeling.

We further compared these clusters with the all major early-labeled (E11–E13) deep- and upper-layer excitatory neuron populations (excluded cingulate cortex and other fate biased clusters). Among E11–E13 labeled clones, L5/6 clusters, including Cortex:EX_L5/6_11, Cortex:EX_L5/6_42, and Cortex:EX_L5/6_Scgl2_28, contained 368 multicellular clones, whereas L2/3 clusters, including Cortex:EX_L2/3_16, Cortex:EX_L2/3_7, Cortex:EX_L2/3_8, Cortex:EX_L2/3_23, and Cortex:EX_L2/3_26, contained 230 multicellular clones. These results indicate that the 17 clones of upper-layer clusters described in the manuscript represent a relatively small subset of early-labeled L2/3 populations.

We have now explicitly clarified the number of multicellular clones and their labeling time points in the main text. Although the presence of early-labeled upper-layer fate-biased clones could be consistent with a small population of fate-biased upper-layer progenitors, these clones are not sufficiently represented in our dataset

to support a strong conclusion regarding fate-restricted upper-layer neurogenesis. We therefore revised the text to remove the overinterpretation in relation to prior Cux2-Cre lineage studies.

The revised main text now states:

“In addition, two fate-biased clusters of upper-layer glutamatergic neurons (cluster 23 and 26; Prdm16, Tsix; 17 multicellular clones in E11-E13) rarely shared barcodes with deep-layer neurons (Extended Data Fig. 9c, Extended Data Fig. 10; 368 multicellular clones in E11-E13).”

- The data claiming neurogenesis after birth should be substantiated further. Since the library used is lentiviral, how can the authors conclude that the cells shown are born at P0, why not using retroviruses instead? Also, what is the identity of the cortical neurons generated?

Response:

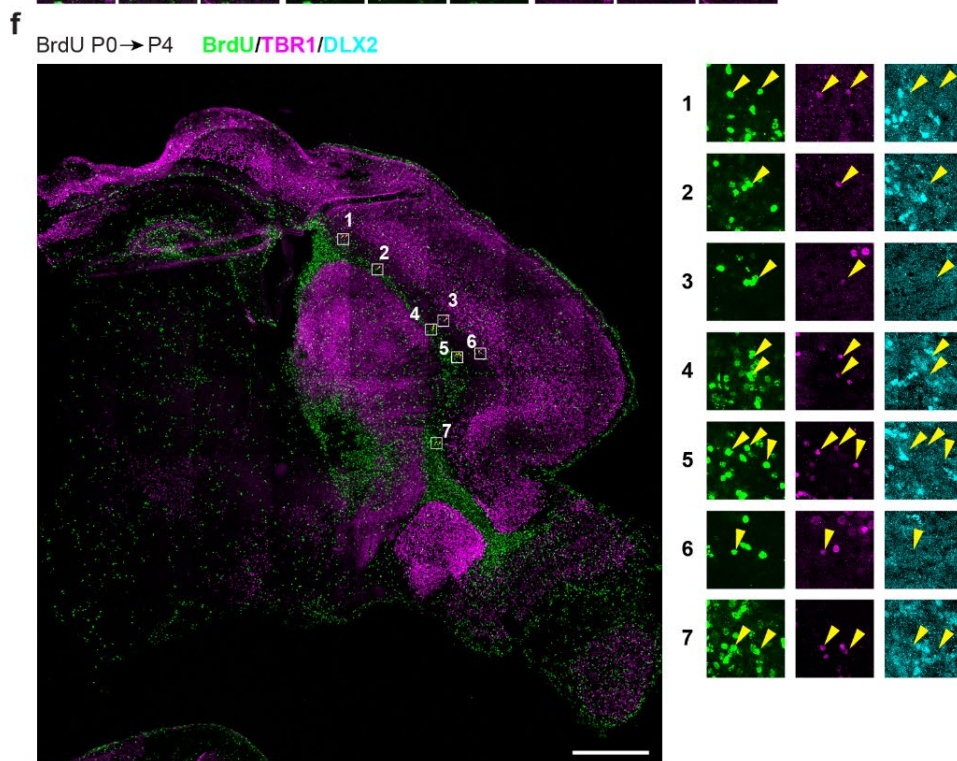
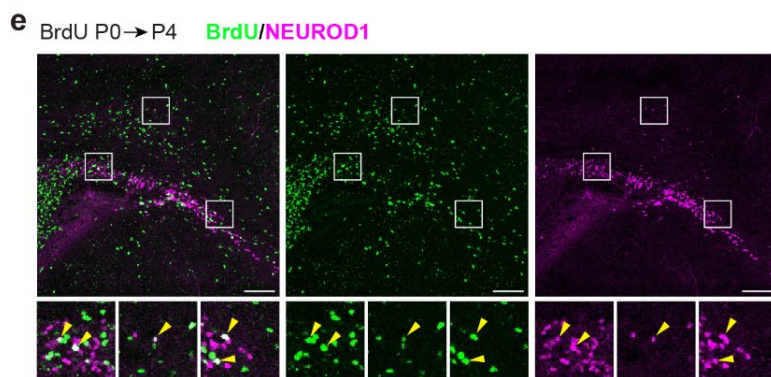
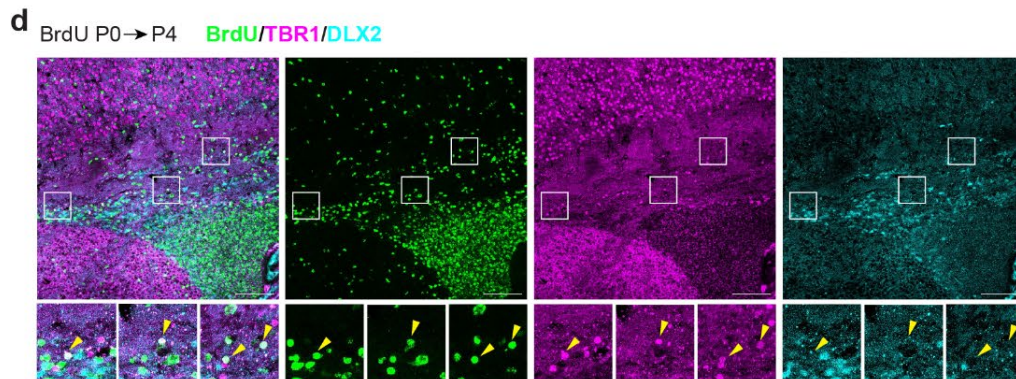
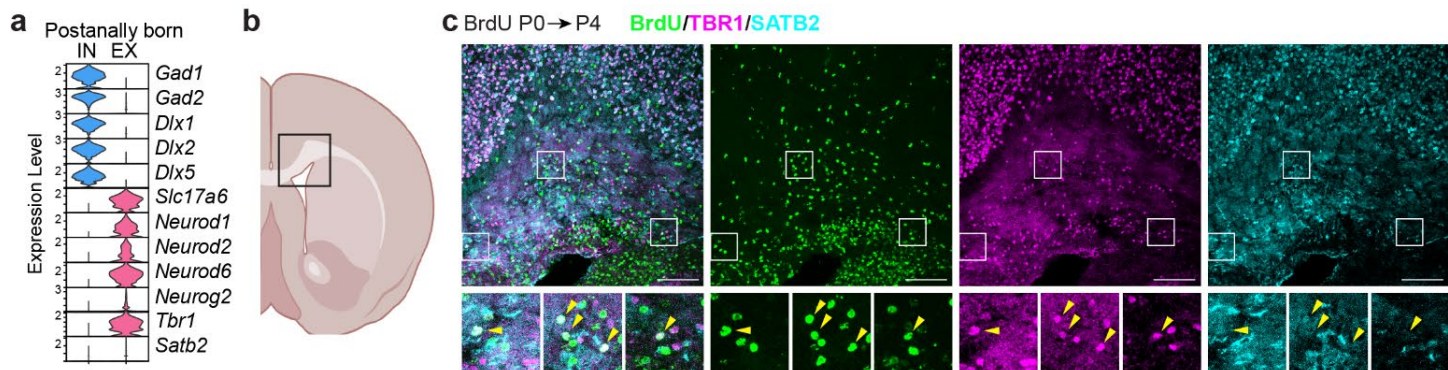
Thanks for the reviewer’s suggestion and comments. We agree that lentiviral labeling alone does not provide the temporal specificity of a retrovirus. For this reason, our analysis is based only on multicellular clones, for which the labeled progenitor needs to have undergone division between P0 and P4. We used lentivirus because it labels both cycling and quiescent progenitors, provides broader progenitor coverage, and offers higher titer needed for this study.

As in the response to Comment 7 of reviewer #2, to further define the identity of the postnatally generated cortical glutamatergic neurons, we performed additional experiments and analyses. Specifically, we first examined their transcriptomic identity using established markers, in comparison with postnatally generated GABAergic neurons (**Extended Data Fig. 11a**). These cells express canonical excitatory neuronal markers (as *Neurod1* and *Slc17a6*) and lack inhibitory markers (as *Dlx2* and *Gad2*), supporting their classification as glutamatergic neurons. Moreover, they are positive for *Tbr1* but negative for *Satb2*. We further validated this population by immunostaining with BrdU applied at P0 and collected at P4, which confirmed the presence of BrdU⁺/TBR1⁺/DLX2⁻; BrdU⁺/NEUROD1⁺; BrdU⁺/TBR1⁺/SATB2⁻ cells (**Fig. 2i; Extended Data Fig. 11b-e**). Because our dataset was collected at P4, these neurons are mostly observed in the corpus callosum and are likely still migrating (**Extended Data Fig. 11f**), such that their final regional positions cannot yet be determined from the current data. Nevertheless, their clonal relationship with prenatally generated glutamatergic neurons and glia suggests continuity of lineage (**Figs. 2j,k**). We also note that this population may partially overlap with previously reported postnatally generated glutamatergic neurons derived from *Neurog2*⁺ progenitors that migrate to the OB¹³ or cortex¹⁴. We have revised the manuscript to clarify these points, more explicitly justify the interpretation of postnatal neurogenesis, and further discuss the identity of this population.

As in the main text:

*“In the cortex, postnatally generated glutamatergic neurons mapped predominantly to the white matter region above the dorsolateral ventricles (Fig. 2h; Extended Data Fig. 8). Gene expression analysis was performed to define the molecular features of this population, showing that these cells expressed glutamatergic neuronal markers, including *Slc17a6*, *Neurod1* and *Tbr1*, lacked *Satb2*, and were negative for inhibitory neuronal markers, including *Dlx2*, *Gad1* and *Gad2* (Extended Data Fig. 11a). We further validated this population by P0 BrdU labeling followed by P4 immunostaining, which confirmed the presence of BrdU⁺/TBR1⁺/DLX2⁻ cells, mainly in the white matter above the dorsolateral ventricles (Fig. 2i; Extended Data Fig. 11b-e). Analysis of sagittal sections showed that this population was distributed along the rostrocaudal axis. Because these neurons are still migrating, it remains unclear whether they subsequently migrate into the cortex or toward the OB (Extended Data Fig. 11f). They*

may partially overlap with previously described postnatally generated glutamatergic neurons^{13–15}. Clonal lineage analysis showed that, at P0 labeling, these cells shared clones with mitotic glial cells and GABAergic neurons migrating toward the OB (Fig. 2j,k). Notably, this population of postnatally glutamatergic neurons shared clonal relationships with prenatally generated glutamatergic neurons and glial cells in labeling of E11-E15, suggesting a continued lineage (Fig. 2j,k)."



Extended Data Figure 11: Molecular identities of postnatally born glutamatergic neurons

a, Violin plots showing expression of canonical excitatory and inhibitory neuronal marker genes in postnatally born excitatory and inhibitory neurons isolated from dissected cortical regions.

b, Diagram highlighting regions enriched for postnatally generated glutamatergic neurons in the validation dataset.

c-e, Immunostaining validation of postnatally born glutamatergic neurons in representative coronal sections. BrdU was administered intraperitoneally at P0, and whole-mount analysis was performed at P4. Double-positive cells are indicated by arrows. Scale bar = 100 μ m.

f, Immunostaining validation of postnatally born glutamatergic neurons in representative sagittal sections. BrdU was administered intraperitoneally at P0, and whole-mount analysis was performed at P4. Double-positive cells are indicated by arrows. Scale bar = 1 mm.

Minor comment :

- the study is focused on forebrain, so this should be clearly explicated in the abstract and introduction, in which 'brain' is often used instead of 'forebrain'. Unless the authors add their data on midbrain, hindbrain, cerebellum, ..

Response:

We thank the reviewer for this suggestion. We agree that the terminology should accurately reflect the scope of the study. Accordingly, we have revised the title, abstract, introduction, and discussion to consistently use “forebrain” in place of the broader term “brain” where appropriate.

REFERENCE

1. Gao, P. *et al.* Deterministic progenitor behavior and unitary production of neurons in the neocortex. *Cell* **159**, 775–788 (2014).
2. Llorca, A. *et al.* A stochastic framework of neurogenesis underlies the assembly of neocortical cytoarchitecture. *Elife* **8**, e51381 (2019).
3. Xu, H.-T. *et al.* Distinct lineage-dependent structural and functional organization of the hippocampus. *Cell* **157**, 1552–1564 (2014).
4. Shi, W. *et al.* Ontogenetic establishment of order-specific nuclear organization in the mammalian thalamus. *Nat. Neurosci.* **20**, 516–528 (2017).
5. Wong, S. Z. H. *et al.* In vivo clonal analysis reveals spatiotemporal regulation of thalamic nucleogenesis. *PLoS Biol.* **16**, e2005211 (2018).
6. Kelly, S. M. *et al.* Radial glial lineage progression and differential intermediate progenitor amplification underlie striatal compartments and circuit organization. *Neuron* **99**, 345–361.e4 (2018).
7. Shen, Z. *et al.* Distinct progenitor behavior underlying neocortical gliogenesis related to tumorigenesis.

- Cell Rep.* **34**, 108853 (2021).
8. Ratz, M. *et al.* Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat. Neurosci.* **25**, 285–294 (2022).
 9. Andraka, E., Phillips, R. A., 3rd, Brida, K. L. & Day, J. J. Chst9 marks a spatially and transcriptionally unique population of Oprm1-expressing neurons in the nucleus accumbens. *Addict. Neurosci.* **11**, 100153 (2024).
 10. Fuentealba, L. C. *et al.* Embryonic origin of postnatal neural stem cells. *Cell* **161**, 1644–1655 (2015).
 11. Bandler, R. C. *et al.* Single-cell delineation of lineage and genetic identity in the mouse brain. *Nature* **601**, 404–409 (2022).
 12. Squarzoni, P. *et al.* Microglia modulate wiring of the embryonic forebrain. *Cell Rep.* **8**, 1271–1279 (2014).
 13. Winpenny, E. *et al.* Sequential generation of olfactory bulb glutamatergic neurons by Neurog2-expressing precursor cells. *Neural Dev.* **6**, 12 (2011).
 14. Donega, V. *et al.* Transcriptional dysregulation in postnatal glutamatergic progenitors contributes to closure of the cortical neurogenic period. *Cell Rep.* **22**, 2567–2574 (2018).
 15. Nowakowski, T. J., Mysiak, K. S., Pratt, T. & Price, D. J. Functional dicer is necessary for appropriate specification of radial glia during early development of mouse telencephalon. *PLoS One* **6**, e23013 (2011).
 16. Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
 17. Fu, J. M. *et al.* Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).