nature methods

Article

https://doi.org/10.1038/s41592-026-03151-5

Spatio-DARLIN enables robust and efficient in situ lineage tracing in mice at single-cell resolution

Received: 29 August 2025

Accepted: 5 June 2026

Published online: 29 June 2026

Check for updates

Jianing Gao (高佳宁)[1,2,5], Zhanhao Zhang (张展浩)[1,2,5], Daolong Chen (陈道龙)[1,2], Sijie Diao (刁思捷)[1,2], Simin Liu (刘思敏)[1,2], Shou-Wen Wang (王寿文)[1,2,3,4] ✉ & Li Li (李莉)[1,2] ✉

Spatially resolved lineage tracing is essential for understanding how clonal relationships shape tissue architecture. However, such an approach has not been established in mice across different tissues. Here we present Spatio-DARLIN, a versatile method that integrates the high-diversity DARLIN lineage-tracing mouse with sequencing-based spatial transcriptomics. Through a dedicated computational pipeline, Spatio-DARLIN achieves accurate clonal mapping at single-cell resolution and recovers reliable lineage information from ~25–50% of cells in the intestine and brain. Spatio-DARLIN identified stereotyped clonal patterns in the intestinal epithelium and revealed clonal dynamics that were consistent with stem-cell neutral drift. In the brain, we uncovered greater clonal expansion of radial glial cells in the cortex and hippocampus during development than in other regions. Moreover, our data strongly suggested that neuronal progenitors across different nuclei in the hypothalamus were already spatially prepatterned by embryonic day E10. Spatio-DARLIN enables high-resolution study of clonal architecture, expansion and migration across diverse tissues in situ.

The architecture and function of multicellular organisms are governed by two fundamental determinants: the lineage of a cell, which encapsulates its developmental history and inherited potential, and its location, which shapes its microenvironmental niche and intercellular interactions. The dynamic interplay between this cellular 'history' and 'geography' orchestrates development, tissue homeostasis and disease pathogenesis. Therefore, a complete understanding of these processes requires the simultaneous resolution of the clonal origin of a cell, its present molecular state and its precise spatial coordinates within an intact tissue. Achieving this integrated, multimodal perspective is essential for deciphering the fundamental rules of tissue formation, maintenance and repair.

Historically, lineage-tracing methodologies have been constrained by a fundamental tradeoff between spatial context and informational depth. Classical approaches using heritable fluorescent reporters preserve excellent spatial fidelity but suffer from limited barcoding capacity and an inability to report cell state[1]. The advent of high-diversity genetic barcoding coupled with single-cell RNA sequencing overcame these limitations, enabling large-scale reconstruction of lineage relationships with deep molecular profiling[2–6]. This powerful synergy has yielded critical insights into cell fate decisions[7–11], migration, cancer evolution[12–14] and epigenetic memory[8,9,15,16]. However, these approaches necessitate tissue dissociation, thus destroying the native spatial architecture of cells. Despite a few specialized studies[17–20], a generic, high-resolution spatial lineage-tracing technology applicable across diverse murine tissues has remained an unmet need.

A naive combination of existing spatial and lineage-tracing technologies is inadequate for three primary reasons. First, current sequencing-based spatial transcriptomic platforms exhibit low mRNA capture efficiency, which is particularly detrimental for recovering

[1]Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China. [2]School of Life Sciences, Westlake University, Hangzhou, China. [3]School of Science, Westlake University, Hangzhou, China. [4]Center for Interdisciplinary Studies, Westlake University, Hangzhou, China. [5]These authors contributed equally: Jianing Gao, Zhanhao Zhang. ✉e-mail: wangshouwen@westlake.edu.cn; lili@westlake.edu.cn

expressed lineage barcodes and results in sparse, uninformative clonal maps. Second, the immense scale of these datasets (profiling hundreds of thousands of cells per section) dramatically increases the probability of barcode homoplasy, where independent editing events by chance generate identical barcodes in unrelated cells. Finally, technical errors such as barcode diffusion, library preparation and sequencing errors and computational cell segmentation errors can lead to erroneous clone assignments that obscure true clonal boundaries. Overcoming these challenges demands coordinated experimental and computational innovation.

We recently developed DARLIN, an inducible CRISPR-based mouse model that supports high-resolution lineage tracing across diverse tissues[9,21]. DARLIN generates a theoretical barcode diversity of ~10^{18} and achieves robust barcode detection in nearly 70% of cells in dissociative single-cell RNA sequencing assays, establishing it as a premier system for in vivo lineage analysis. Here, we introduce Spatio-DARLIN, a versatile method that integrates the DARLIN mouse with sequencing-based spatial transcriptomics to concurrently map clonal history, transcriptional state and spatial location for hundreds of thousands of cells within their native tissue context (Fig. 1a). We rigorously benchmark the performance of Spatio-DARLIN and demonstrate its utility by delineating the spatiotemporal clonal architecture of both the intestine and brain.

Results

Spatio-DARLIN jointly profiles gene expression and clonal barcodes in situ at single-cell resolution

The DARLIN system uses three polyadenylated barcoding arrays (Col1a1 array (CA), TIGRE array (TA) and Rosa26 array (RA)) integrated into distinct genomic loci, each containing ten tandem CRISPR target sites (Fig. 1a). Transient doxycycline (Dox) administration induces editing at these sites, generating diverse and expressible lineage barcodes in each founder cell. To capture these barcodes in situ, we used a commercial spatial transcriptomics chip (BMKMANU S3000) featuring a dense array of 2.5-µm poly(dT) microbeads (each with a unique spot barcode) spaced 3.5 µm apart across a capture area of 6.8 × 6.8 mm^2. Given that mouse cells are ~10 µm in diameter, this bead density enables spatial lineage tracing at single-cell resolution. We enhanced the capture efficiency of DARLIN barcodes through targeted amplification. We refer to this integrated method as Spatio-DARLIN.

We first applied Spatio-DARLIN to map the spatial clonal architecture of the mouse small intestine, an ideal model system due to its rapid cellular turnover (3–5 days)[22] and well-defined spatial organization into crypt–villus units. To induce clonal labeling, we treated an adult DARLIN mouse (Y007) with Dox for 1 week, followed by 1-week chase period without Dox to allow clonal expansion and differentiation (Fig. 1b). The proximal small intestine (PSI) was then collected, coiled into a 'Swiss-roll' configuration and cryosectioned. The tissue section was profiled on the spatial transcriptomics chip to generate spatially barcoded cDNAs, which were sequenced to simultaneously obtain single-cell transcriptomes and DARLIN barcodes. Matched hematoxylin and eosin (H&E) staining was conducted to facilitate precise cell segmentation.

Following cell segmentation and quality control (QC), we recovered 146,584 high-quality cells from a single tissue section (Extended Data Fig. 1a,b). Unsupervised clustering revealed seven major populations in the transcriptome: crypt, villus, lamina, muscle, mesothelium, adipose and isolated lymphoid follicle (Fig. 1c,d). These populations occupied characteristic spatial domains (Fig. 1e). Using canonical marker genes, we further resolved specific cell types, including intestinal stem cells and Paneth cells in the crypt; enterocytes and goblet and tuft cells in the villus; immune cells in the lamina and isolated lymphoid follicle and myocytes and neurons in the muscle layer, all localizing to their expected anatomical regions (Extended Data Fig. 1c–g). These results demonstrate high-quality spatial transcriptomic data.

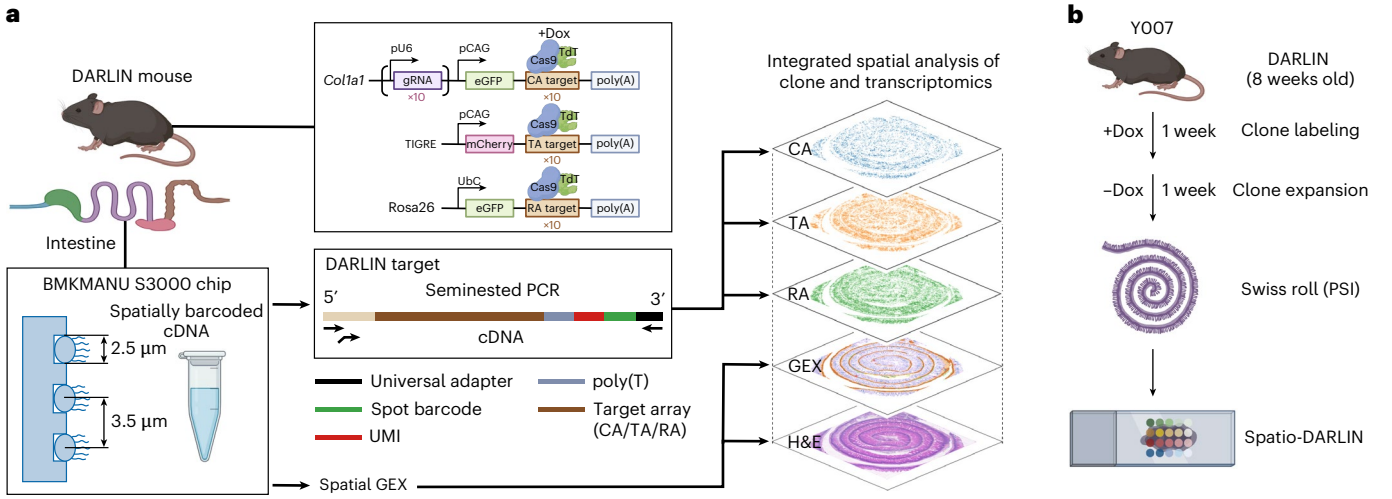
A key challenge for spatial lineage tracing is the accurate assignment of clonal barcodes. Although the transcriptional state of a cell is supported by hundreds of genes, its clonal identity hinges on a few DARLIN barcode sequences, making this identity vulnerable to technical errors such as barcode diffusion, amplification and sequencing errors and cell segmentation errors. Errors in the DARLIN barcode sequence typically lead to clonal fragmentation, where cells from a single clone are erroneously assigned to different barcodes. By contrast, errors in the associated spot barcode are more detrimental, as they misassign the location of a cell, creating spurious, spatially diffuse clonal signals (Fig. 1f), a phenomenon observed in our raw data (Fig. 1g). To ensure accurate recovery of spatially resolved lineage barcodes from amplicon sequencing data, we developed a customized bioinformatic pipeline designed to mitigate specific sources of technical noise (Extended Data Fig. 2 and Supplementary Fig. 1). This pipeline: (1) corrects PCR and sequencing errors, (2) filters amplification artifacts by removing molecules with the same spot barcode and unique molecular identifier (UMI) but that lack a single dominant lineage barcode and (3) eliminates capture-oligonucleotide carryover artifacts by discarding spatial spots characterized by low read:UMI coverage ratios. This stepwise QC strategy effectively reduced sequence-level and spot-level noise, yielding a high-confidence dataset of spatially coherent clonal signals for downstream analysis (Supplementary Fig. 2). Finally, we aggregated lineage barcodes based on cell segmentation to achieve single-cell resolution. Unless otherwise specified, all downstream analyses were conducted using these cellular units. After these preprocessing steps, we observed tightly localized clones in the small intestine, which validated our approach (Fig. 1g,h).

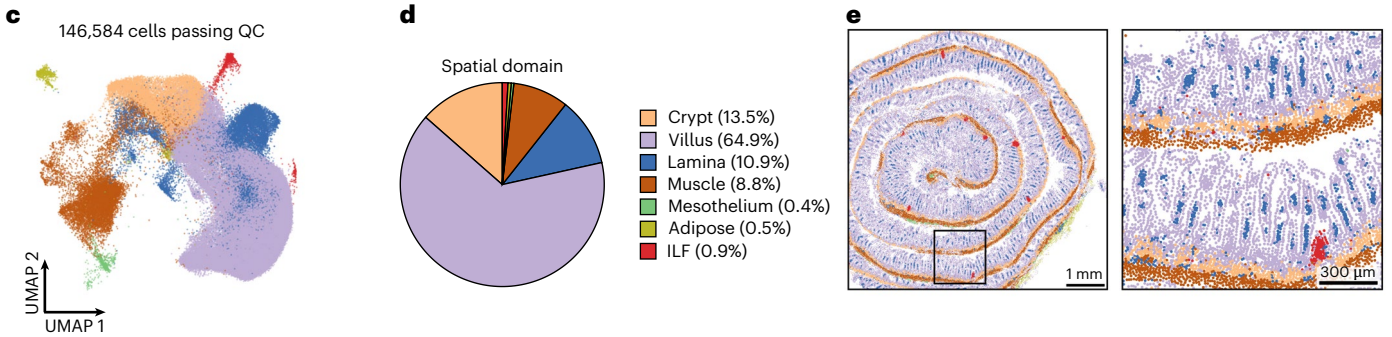
A useful DARLIN barcode must not only be detected, but also carry a unique editing pattern that reliably labels a clone. After transcriptomic QC, we obtained 146,584 cells, in which 51,459 cells (35.1%) contained a detectable DARLIN barcode from at least one target locus

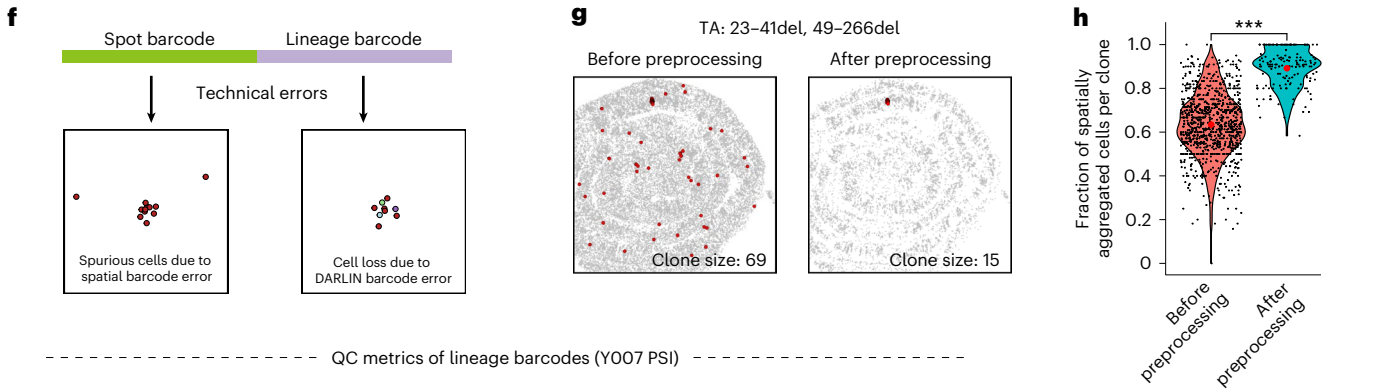
Fig. 1 | Spatio-DARLIN jointly profiles gene expression and clonal barcodes in situ with high resolution. a, Schematic of Spatio-DARLIN. DARLIN mice carry three target arrays (CA, TA and RA) in different genomic loci. After tissue dissection, the intestine was coiled into a Swiss roll, stained with H&E, imaged and subjected to spatial transcriptomics. Clonal barcodes were recovered from spatially barcoded cDNA via seminested PCR and integrated with spatial gene expression profiles. eGFP, enhanced green fluorescent protein; GEX, gene expression; gRNA, guide RNA; TdT, terminal deoxynucleotidyl transferase. **b**, Experimental design for barcode induction, tissue collection and profiling; ILF, isolated lymphoid follicle. **c**, Transcriptome-based uniform manifold approximation and projection (UMAP) of 146,584 cells passing QC. Cells are colored by inferred spatial domains. **d**, Cell proportions of major spatial domains. **e**, Spatial representation of the transcriptomic data (left) and a zoomed-in view (right). Cells are colored by spatial domains as in **c**. **f**, Schematic of potential errors and consequences in Spatio-DARLIN sequencing data. **g**, Representative clone signal before and after preprocessing. **h**, Violin plots showing the fraction of spatially aggregated cells per clone before and after preprocessing; $^{***}P < 2.2 \times 10^{-16}$, two-sided Wilcoxon signed-rank test. **i**, QC of DARLIN barcodes; BCs, barcodes. **j**, Venn diagram showing the number of cells associated with different types of barcode detection. **k**, Editing efficiency across spatial domains for each target array. Frac., fraction. Each dot corresponds to the editing efficiency computed from a spatial domain within a single mouse, colored as in **c**, while the bar plots represent the mean editing efficiency across domains. The sample size (n) corresponds to the number of spatial domains. Spatial domains are treated as independent units of observation within one biological sample; no biological replicates were performed for this analysis. **l**, Relative fraction of cells that pass each QC (that is, detected, edited and rare) or all QCs. Dots indicate statistics from three mouse replicates (Y003, Y004 and Y007), while the bar plots represent the average across these dots. Illustrations in **a** and **b** were created in BioRender; Li, L. https://biorender.com/cl0m82y (2026).

https://doi.org/10.1038/s41592-026-03151-5

(Fig. 1i). Among these, 27% had barcodes detected from two or more arrays. The TA and RA arrays showed higher detection rates than the CA array (Fig. 1j). We detected DARLIN barcodes across all cell types at an efficiency proportional to their total detected mRNA abundance (Extended Data Fig. 3a–c). We observed nearly 100% editing efficiency across all three arrays and major cell populations (Fig. 1k). We previously established a common barcode filtering strategy designed to remove barcode homoplasy. Applying this, we found that ~70% of edited cells were retained using a conservative barcode generation probability cutoff ($\rho_* = 10^{-5}$; Extended Data Fig. 3d–f). Integrating







https://doi.org/10.1038/s41592-026-03151-5

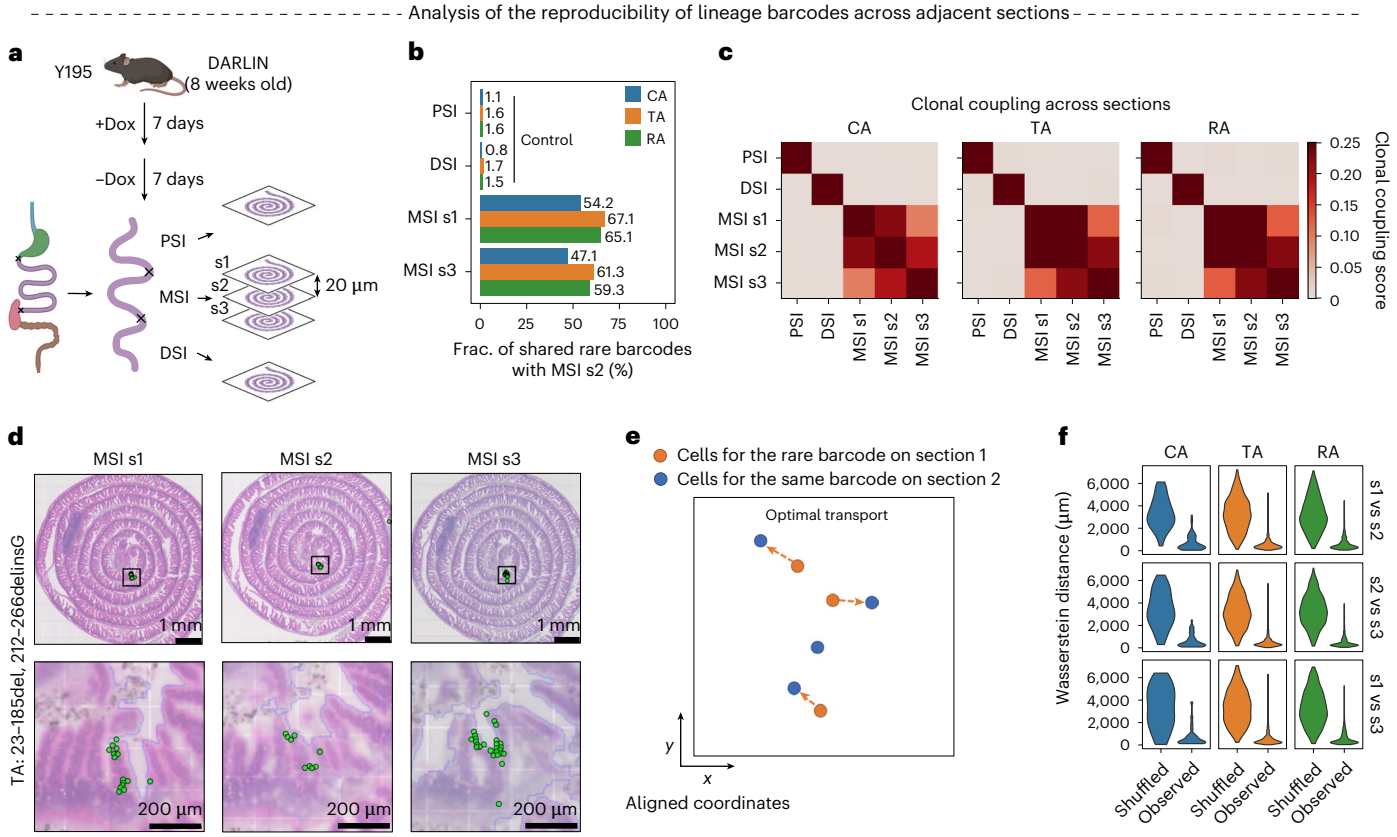


Fig. 2 | Spatio-DARLIN enables negligible barcode homoplasy and reproducible clone detection. a, Experimental design for evaluating spatial clone detection. **b**, Fraction of shared rare barcodes between MSI s2 and other tissue sections (PSI, DSI, MSI s1 and MSI s3) for each target array. **c**, Heat map of clonal coupling scores across tissue sections, computed independently for each array (CA, TA and RA). **d**, Representative examples of clonal localization detected across adjacent sections for TA barcodes. Reproducibility across section comparisons is summarized in **f**. Similar spatial patterns were observed in an independent biological replicate (L126) across three consecutive brain sections (Extended Data Fig. 7). **e**, Schematic of computing the Wasserstein distance between two populations associated with the same rare barcode but from different tissue sections. **f**, Violin plots of the Wasserstein distances between different tissue sections across all rare barcodes. Values for both observed and shuffled barcode labels are shown. Illustration in **a** created in BioRender; Li, L. https://biorender.com/cl0m82y (2026).

barcodes across the three loci to identify clones (Extended Data Fig. 3g), we retained 12,390 reliable clones, including 6,748 clones with more than one cell (Extended Data Fig. 3h). In total, approximately 25% of all profiled cells passed stringent QC for reliable clonal barcodes, providing a robust foundation for downstream analysis. These efficiency metrics were consistent across multiple biological replicates (Fig. 1l), demonstrating the reproducible and efficient barcode recovery of Spatio-DARLIN in situ.

Spatio-DARLIN reproducibly detects spatial clones across adjacent tissue sections

To further validate the spatial detection accuracy of Spatio-DARLIN, we administered Dox to an 8-week-old DARLIN mouse, followed by a 1-week chase before collection. We separately profiled the proximal, middle and distal segments of the small intestine (PSI, MSI and DSI, respectively; Fig. 2a). For the MSI, we analyzed three adjacent cryosections (s1, s2 and s3) spaced 20 µm apart. This design creates a natural negative control, as clones from the non-adjacent PSI and DSI should be genetically unrelated to those in the MSI, allowing us to quantify barcode homoplasy. Conversely, the three adjacent MSI sections serve as technical replicates to assess other experimental noise.

Analysis revealed a minimal overlap of ~1% shared clones between the distinct intestinal segments (PSI, MSI and DSI), confirming a low homoplasy rate (Fig. 2b). By contrast, adjacent sections within the MSI (for example, s1/s2 and s2/s3) shared approximately 50% of DARLIN barcodes. We independently confirmed this using CoSpar[23] to calculate clonal coupling scores, revealing strong connectivity within the MSI series but negligible coupling between the separate segments (Fig. 2c). Furthermore, cells from the same clone exhibited highly consistent spatial localization across sections (Fig. 2d). We quantified this spatial fidelity using the normalized Wasserstein distance, where a smaller distance indicates higher localization consistency. The distances for true clones were substantially smaller than those for shuffled barcode controls (Fig. 2e,f). Additional analyses of barcode consistency and spatial coherence on these data further supported the accuracy of clone assignment by our method (Supplementary Figs. 2 and 3). Collectively, these results demonstrate that Spatio-DARLIN, aided by our optimized preprocessing pipeline, achieves robust and accurate in situ clonal detection with minimal technical artifacts.

Spatio-DARLIN unveils spatiotemporal clonal dynamics of the intestinal epithelium

The small intestine consists of periodic crypt–villus units organized along the proximal–distal axis. Each crypt harbors 12–16 stem cells that continuously divide and differentiate along the crypt–villus axis, with some stem cells being stochastically outcompeted over time[24,25]. Although marker gene-based lineage tracing has provided important insights into this process[24,25], these approaches have been limited in throughput and cell-state resolution. Here, we reanalyzed the spatiotemporal clonal dynamics of intestinal epithelial cells using our high-throughput Spatio-DARLIN platform.

https://doi.org/10.1038/s41592-026-03151-5

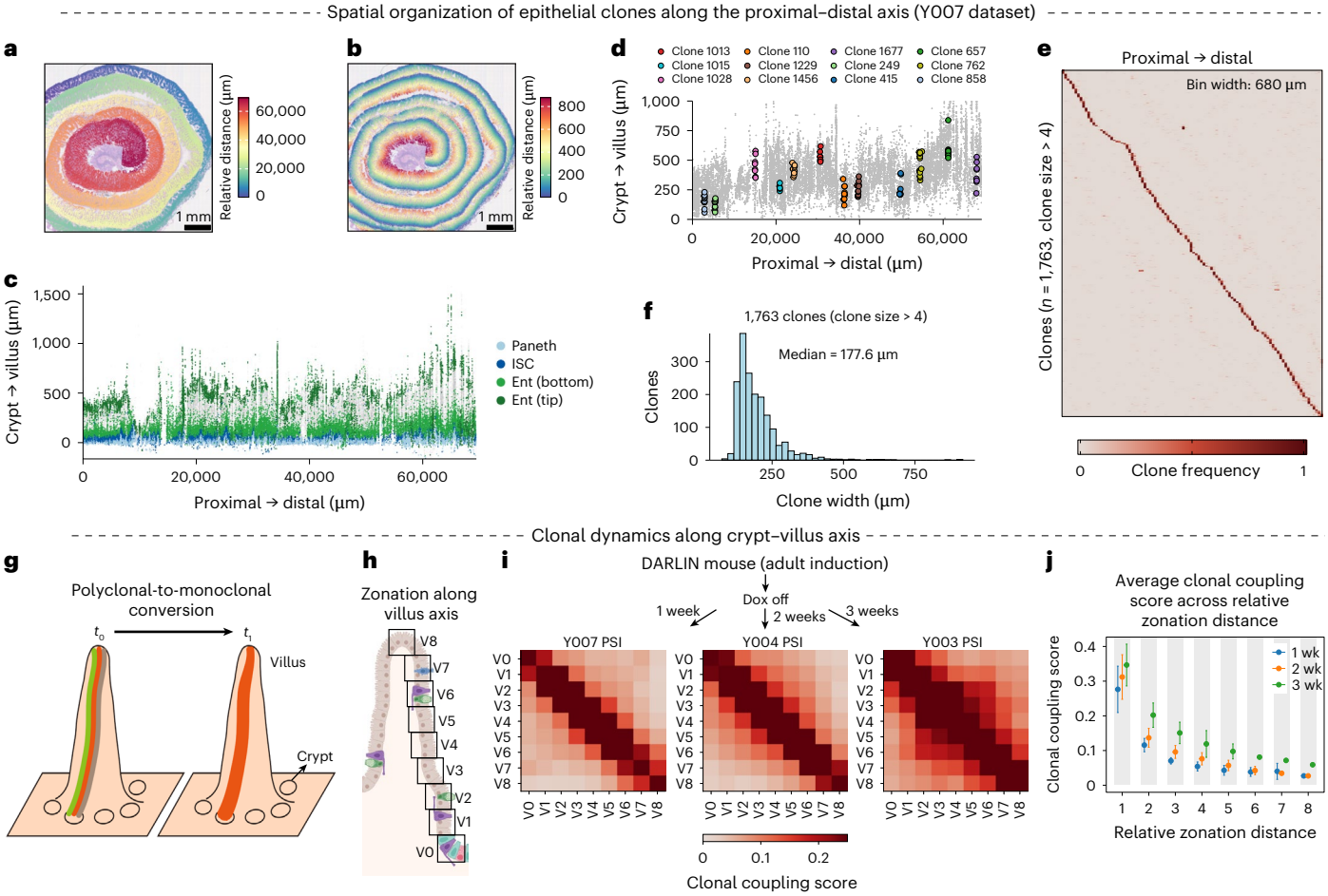


Fig. 3 | Spatio-DARLIN uncovers spatiotemporal clonal architecture in the small intestine. a,**b**, Spatial coordinates of cells in the Y007 PSI section, colored by position along the proximal–distal axis (**a**) and crypt–villus axis (**b**). **c**, Two-dimensional unrolled coordinates of epithelial cells along the proximal–distal axis (*x* axis) and crypt–villus axis (*y* axis). Selected cell types are indicated on this embedding; Ent, enterocytes; ISC, intestinal stem cells. **d**, Spatial localization of representative epithelial clones along the unrolled embedding. **e**, Heat map of the clone-by-cell matrix, visualized with a bin size of 680 μm along the proximal–distal axis. Clones are ordered by their peak density along the proximal–distal axis. Only clones with more than four cells are used. **f**, Histogram of clone width along the proximal–distal axis. **g**, Schematic of polyclonal-to-monoclonal conversion after initial intestinal stem cell labeling at time 0 (t_0). The clonal diversity within a crypt is lost over time. **h**, Schematic of zonation binning along the crypt–villus axis, leading to nine equal-length zones from the crypt base (V0) to the villus tip (V8). **i**, Heat map of clonal coupling scores between crypt–villus zones (V0–V8) at 1 (left), 2 (middle) or 3 (right) weeks after labeling. **j**, Clonal coupling scores as a function of relative zonation distance (for example, V1 and V4 have a relative distance of 3), extracted from the matrix in **i**. Points represent the mean value at a given zonation distance, with error bars indicating ± s.d. wk, weeks. The sample size (*n*) corresponds to the number of pairwise zonation bin comparisons contributing to each distance within each dataset. Data are shown for independent samples collected at different time points (1, 2 and 3 weeks), which represent distinct experimental conditions rather than biological replicates. Illustration in **h** created in BioRender; Li, L. https://biorender.com/cl0m82y (2026).

We first examined the clonal organization along the proximal–distal axis of the small intestine. To enable this analysis, we computationally unrolled the Swiss-roll-like structure, aligning individual crypts along the *x* axis and capturing the crypt–villus progression along the *y* axis. Applying this strategy to the Y007 dataset (Fig. 1), we observed the expected localization of cell types, thereby validating our approach (Fig. 3a–c and Extended Data Fig. 4a–c). In this visualization, detected clones were spatially segregated along the proximal–distal axis (Fig. 3d,e), each spanning a sharp clone width of ~176 μm (Fig. 3f and Extended Data Fig. 4d). This width approximates the distance between two neighboring villus tips, consistent with the idea that epithelial stem cells from a single crypt generate mature epithelial cells distributed along the two adjacent villi.

We next examined clonal dynamics along the crypt–villus axis. After labeling, individual stem cells initiated unique clones that expanded within the crypt through competitive proliferation. Over time, polyclonal crypts transition to a monoclonal state, a phenomenon driven by neutral drift among the stem cell population[24,25] (Fig. 3g). To track this polyclonal-to-monoclonal conversion, we pulse-labeled adult mice and chased for 1–3 weeks, a time window expected to capture substantial loss of clonal diversity. The axis was divided into nine vertical zones (V0–V8), with cells assigned based on their unrolled *y* coordinates (Fig. 3h and Extended Data Fig. 5a). After a 1-week chase, adjacent zones showed notably stronger clonal coupling than distal regions (Fig. 3i, left), a localized pattern that was disrupted in scrambled controls (Extended Data Fig. 5b). This zonal coupling pattern persisted at 2- and 3-week time points and intensified with longer chases (Fig. 3i,j). This progressive increase in zonal coupling is consistent with the loss of clonal diversity; as monoclonal conversion proceeds, dominant clones become wider in three-dimensional (3D) context, increasing the likelihood of being captured as larger, coherent clusters in a two-dimensional (2D) tissue section. We further verified this growing zonal coupling pattern over time within the enterocyte population (Extended Data Fig. 5c–g). Together, our data recapitulate

https://doi.org/10.1038/s41592-026-03151-5

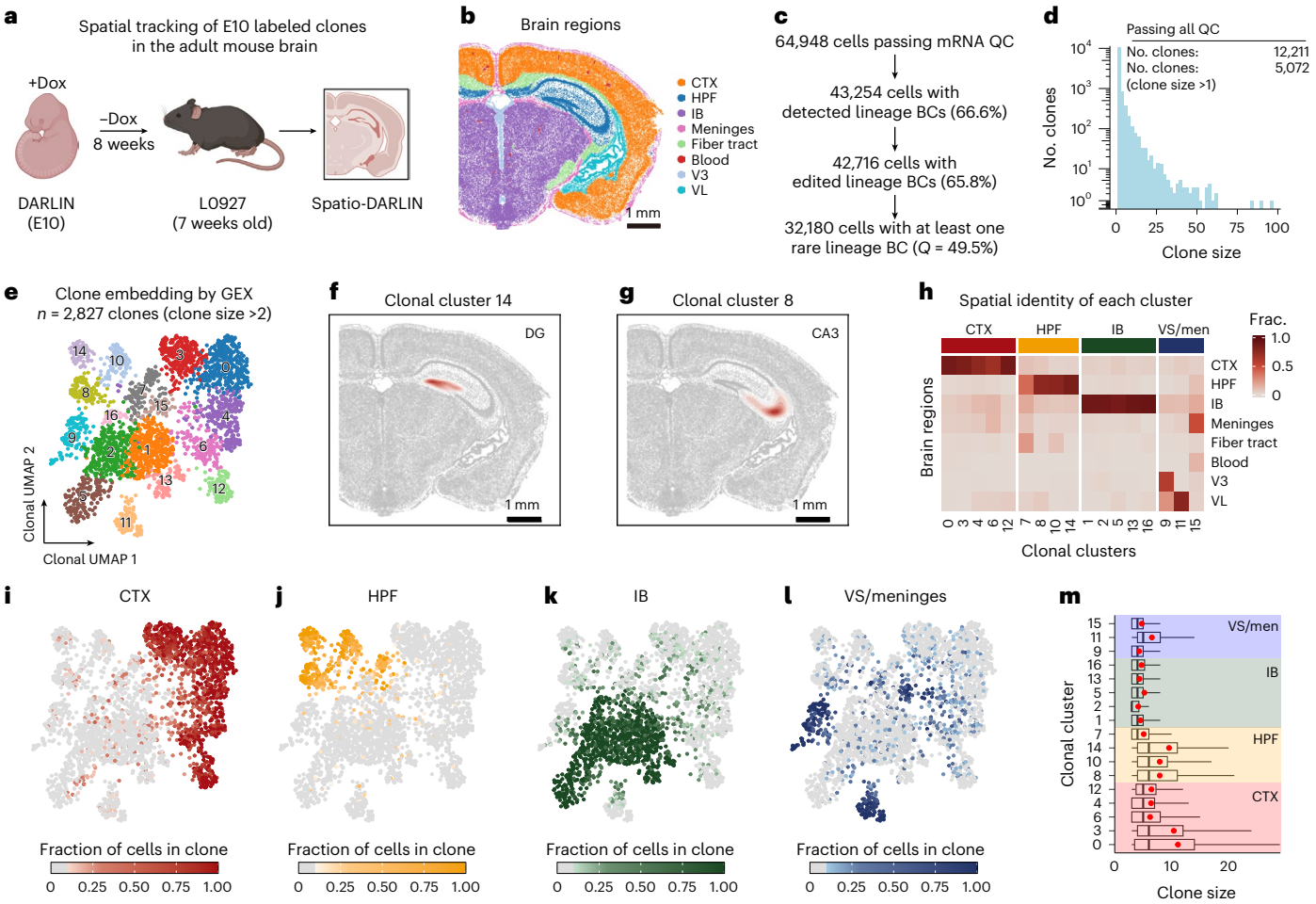


Fig. 4 | Spatio-DARLIN reconstructs the clonal architecture in situ in the brain. a, Experimental design for labeling clones in the brain. **b**, Spatial representation of the transcriptomic data. Cells are colored by spatial domains; CTX, cortex; HPF, hippocampal formation; IB, interbrain; V3, third ventricle; VL, lateral ventricle. **c**, QC of DARLIN barcodes. **d**, Size distribution for all clones passing QC filters, with clone statistics indicated. **e**, Embedding of all 2,827 clones (more than two cells) generated with the clone2vec algorithm. Each dot represents a single clone, which is colored by its cluster identity. **f**,**g**, Density plots for the spatial distribution of selected clonal clusters in **e**. Dark red indicates higher cell density; DG, dentate gyrus; CA3, cornu ammonis 3. **h**, Heat map of cell fractions across spatial domains for each clonal cluster. **i**–**l**, Spatial preference of each clone toward the cortex (**i**), hippocampal formation (**j**), interbrain (**k**) and ventricular system/meninges (VS/men; **l**) on the clonal embedding in **e**. Each clone is colored by its cell fraction at the designated spatial domain. **m**, Box plot of clone size distribution across clonal clusters, as defined in **e**. Box plots show the median (50th percentile), the bounds of the box represent the interquartile range (25th to 75th percentile) and whiskers extend to 1.5× the interquartile range. The sample size (*n*) corresponds to the number of clones within each cluster. Clones are treated as independent units of observation within the analyzed sample. Illustrations in **a** created in BioRender; Li, L. https://biorender.com/cl0m82y (2026).

the spatiotemporal dynamics of clonal competition and monoclonal conversion, yielding spatially restricted clones that migrate upward along the villus.

Spatio-DARLIN resolves clonal architecture of the brain

The intricate anatomy of the brain, with its complex nuclear structures, obscures the developmental principles guiding its assembly. We applied Spatio-DARLIN to investigate its clonal architecture and the timing of cell fate commitment. Previous studies have shown that, around embryonic day 10 (E10), radial glial cells (RGCs) begin to proliferate and undergo neurogenesis and subsequent gliogenesis[26,27]. Thus, we induced DARLIN at E10, resulting in body-wide labeling, including RGCs. We then collected the brains for spatial profiling at 7 weeks (Fig. 4a). We analyzed two biological replicates, collecting a single coronal section from mouse L0927 and three sequential sections from mouse L126 (Extended Data Figs. 6 and 7). For all sections, we generated spatially resolved single-cell transcriptomes alongside DARLIN barcodes. The results from mice L0927 and L126 were highly consistent. Therefore, the subsequent analysis focuses primarily on data from L0927.

We first clustered 64,948 quality-controlled cells based on their transcriptomes and annotated them anatomically (Fig. 4b and Extended Data Fig. 8a–d). Cells were mapped to eight brain regions, including the cortex, hippocampal formation and interbrain, as well as non-neural compartments. As expected, neurons were predominant in cortex, hippocampal formation and interbrain, whereas non-neural cells were enriched in regions such as the meninges and ventricles (Extended Data Fig. 8e,f).

Following stringent QC, we recovered reliable lineage barcodes from 49.5% of cells (Fig. 4c), identifying 12,211 clones. Of these, 5,072 clones contained more than one cell (Fig. 4d). This relatively high recovery rate is consistent with the higher whole-transcriptome capture efficiency observed in brain tissue than in the intestine (Supplementary Fig. 4). Spatio-DARLIN represents a substantial advance over existing spatial lineage-tracing technologies. Compared to 10x Visium V1-based approaches in the brain, it achieves substantially higher lineage barcode recovery when analyzed at a comparable 50-µm resolution (~94% for Spatio-DARLIN versus ~42% for iTracer[28] and ~5% for Space-TREX[29]), while offering superior native resolution (2.5 µm

https://doi.org/10.1038/s41592-026-03151-5

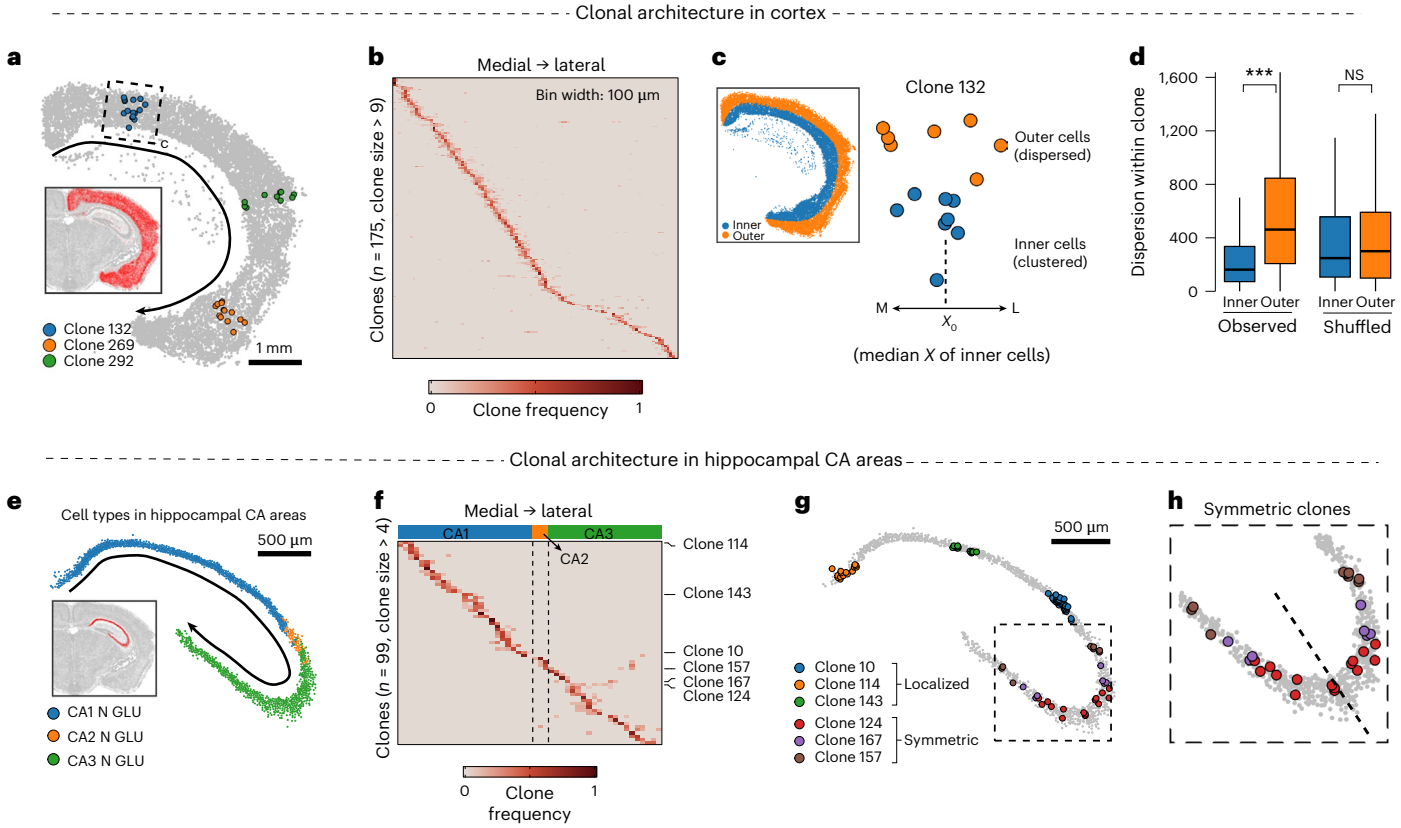


Fig. 5 | Clonal architecture in the cortex and hippocampal CA areas. **a**, Representative clones shown on the cortex. The arrow marks the medial-to-lateral direction. Inset: location of highlighted cortical cells (red dots) on the brain section. The dotted box indicates the region shown in the zoomed-in view in **c**. **b**, Heat map of cell distribution along the medial–lateral axis (100 µm bin) for cortical clones ($n = 175$, clone size > 9). **c**, Left: partition of the cortex into inner and outer layers. Right: schematic of calculating inner and outer dispersion of a clone along the medial–lateral axis. L, lateral; M, medial. **d**, Box plot of inner and outer dispersion of 136 observed clones and their shuffled counterparts (two-sided paired Wilcoxon signed-rank test; $^{***}P = 2.9 \times 10^{-10}$); NS, not significant ($P = 0.50$). Only clones with more than two cells in both inner and outer groups were included. Box plots show the median (50th percentile), the bounds of the box represent the interquartile range (25th to 75th percentile) and whiskers extend to 1.5× the interquartile range. **e**, Cell-type composition in hippocampal CA areas, with CA1, CA2 and CA3 glutamatergic neurons (N GLU) annotated. Inset: location of highlighted hippocampal cells (red dots) on the brain section. **f**, Heat map of cell distribution along the medial–lateral axis for hippocampal clones ($n = 99$; clone size > 4). **g**, Spatial map of representative hippocampal clones. **h**, Zoomed-in view of the CA2–CA3 area in **g**, showing clones with symmetric positioning along the hippocampal fold.

versus 50 µm). Using published data, we provided a detailed comparison against recently reported methods (Extended Data Fig. 10), including KP-Tracer and PEtracer[20,30]. Although this table provides key aspects of different methods, it should be interpreted cautiously because it is not a direct side-by-side comparison of the same biological system.

We computed the transcriptomic similarity among 2,827 clones with more than two cells, and projected them onto a 2D clonal embedding using the clone2vec algorithm[31], where each dot represents a single clone. Leiden clustering revealed 17 distinct clonal clusters; clones within each cluster share similar transcriptional fates (Fig. 4e). These clusters exhibited strong preferential enrichment in specific anatomical regions (Fig. 4f–h). For instance, clusters 0, 3, 4, 6 and 12 were enriched in the cortex (Fig. 4i and Extended Data Fig. 8g), whereas clusters 7, 8, 10 and 14 were specific to the hippocampal formation (Fig. 4f,g,j and Extended Data Fig. 8h). Clusters 1, 2, 5, 13 and 16 were primarily distributed in the interbrain (Fig. 4k and Extended Data Fig. 8i), and clusters 9, 11 and 15 were associated with the ventricular system and meninges (Fig. 4l and Extended Data Fig. 8j,k). Notably, clonal clusters enriched in the same brain region colocalized within the transcriptome-based embedding, suggesting that the RGCs in different brain regions have been spatially prepatterned by E10.

We further identified substantial heterogeneity in clonal expansion between E10 and 7 weeks. Cortical (for example, clusters 0 and 3) and hippocampal clones (for example, clusters 8, 10 and 14) were consistently larger than those from the subcortical regions or the meninges (Fig. 4m). To rule out potential sampling biases inherent to 2D sectioning, we performed orthogonal validation using bulk DARLIN sequencing of the dissected brain regions from E10-induced 5-week-old mice (Extended Data Fig. 9a). By normalizing observed lineage counts using explicitly calculated library and sequencing sampling rates, we generated 3D estimates of absolute clone size. This analysis confirmed that clones in the dorsal cortex and hippocampus are notably larger than those in the thalamus or hypothalamus (Extended Data Fig. 9b–e). This finding suggests that RGCs in the hippocampal formation and cortex undergo substantially more clonal expansion than those in other brain regions.

Spatio-DARLIN reveals distinct clonal organization in the cortex and hippocampus

In the developing mammalian cortex, neurons derived from common RGCs form localized columnar units that span the cortical layers, a pattern resulting from the radial migration of clonally related neurons from the ventricular zone[32]. Consistent with this principle, we observed 175 large clones (size > 9) that were arranged along the radial axis, oriented perpendicular to the medial–lateral axis of the cortex (Fig. 5a,b). Furthermore, many clones exhibited a distinctive cone-shaped morphology, where inner-layer cells were tightly clustered whereas outer-layer cells were more widely dispersed. Quantification

https://doi.org/10.1038/s41592-026-03151-5

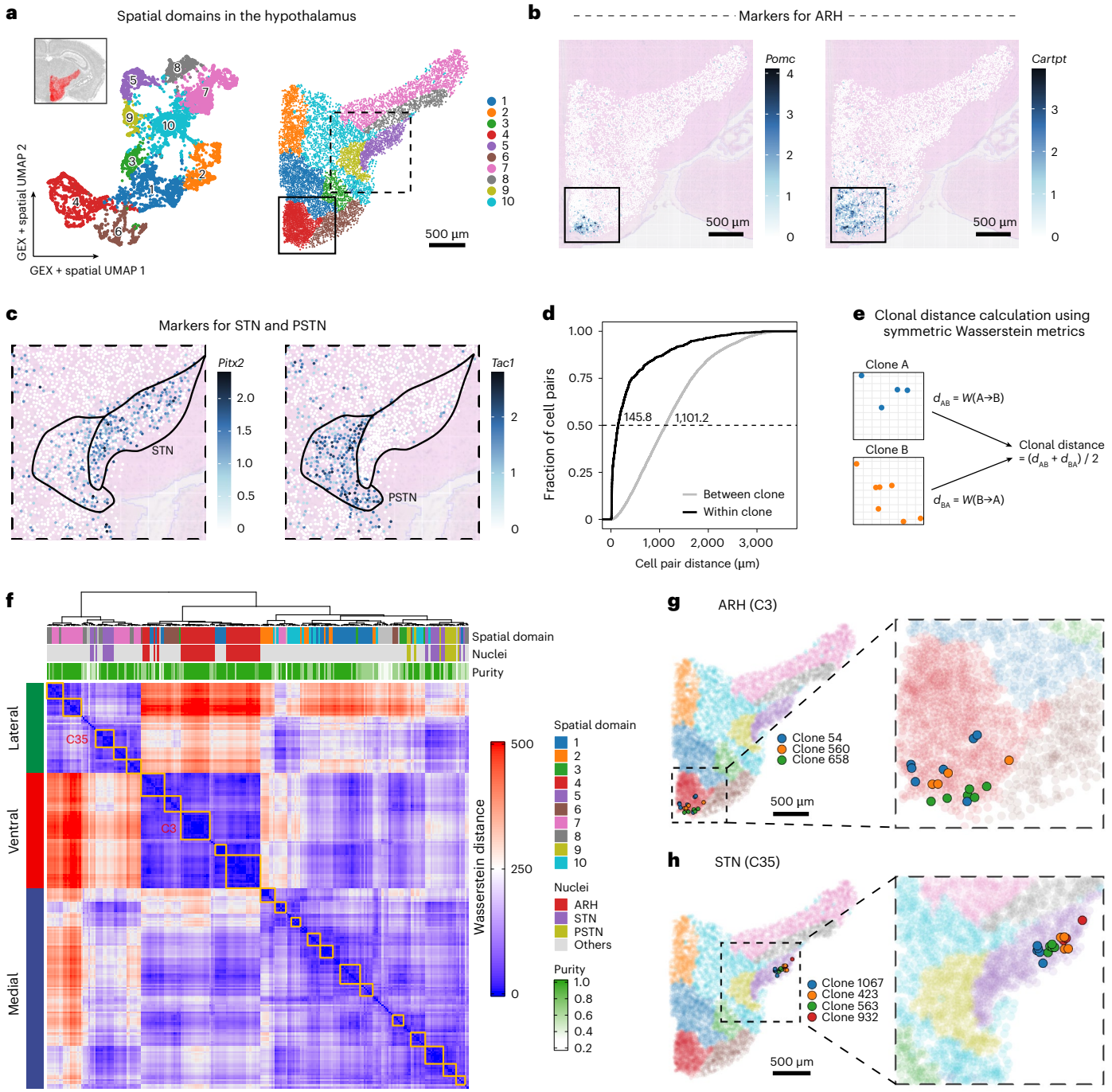


Fig. 6 | Spatio-DARLIN reveals clusters of spatially restricted clones in the hypothalamus. a, Spatial domains in the hypothalamus identified via the BANKSY algorithm. UMAP embeddings of spatially weighted transcriptome (left) or spatial coordinates (right) colored by spatial domains. Inset: location of the hypothalamus region highlighted in red on the brain section. **b**, Expression of ARH marker genes (*Pomc* and *Cartpt*; dotted box in **a**). The ARH region is highlighted in a solid box. **c**, Expression of marker genes of the STN (*Pitx2*) and parasubthalamic nucleus (PSTN; *Tac1*; solid box in **a**). **d**, Cumulative distribution of distances for cells that are within the same clone (black) or between different clones (gray). Median distances are indicated. **e**, Schematic of the clonal distance calculation using the symmetric Wasserstein distance. **f**, Heat map of pairwise clonal distances computed according to **e**, ordered with hierarchical clustering ($n = 233$, clone size > 2). Top color bars show spatial domains (defined in **a**), nuclei assignment and clonal purity (cell fraction in the most populated spatial domain of each clone). **g**,**h**, Spatial map of representative clones from cluster C3 in the ARH (**g**) and cluster C35 in the STN (**h**).

confirmed that outer-layer cells within a clone were significantly more dispersed than inner-layer cells ($P = 2.9 \times 10^{-10}$), a pattern not observed in randomized control clones (Fig. 5c,d). As neurogenesis proceeds from inner layers to outer layers, this increasing radial dispersion likely reflects the cumulative divergence of neuronal migration paths over time[33].

We also identified spatially confined clones in the hippocampal CA regions, organized primarily along the medial–lateral axis (Fig. 5e,f). Although most clones were highly localized (for example, clone 10 and clone 114), a specific subset in the CA2–CA3 region exhibited unusually broad distribution along this axis (for example, clone 124 and clone 157). Notably, these large-spanning clones were positioned with an

apparent symmetry relative to the hippocampal fold (Fig. 5g,h). This phenomenon may be a consequence of the spatial folding that occurs during hippocampal morphogenesis, which can be confirmed with 3D tissue reconstruction.

Spatio-DARLIN uncovers early spatial specification of different nuclei in the hypothalamus

We observed that neuronal clones are spatially prepatterned across different brain regions by E10. Here, we examined whether such spatial specification of clones also occurred in finer structures, such as various nuclei in the hypothalamus[34,35].

To establish a spatial reference, we first applied BANKSY clustering[36] to hypothalamic cells, delineating ten distinct clusters in the transcriptome that corresponded to spatially contiguous domains (Fig. 6a). Several domains aligned with known nuclei; for instance, cluster 4 matched the arcuate hypothalamic nucleus (ARH), and cluster 5 matched the subthalamic nucleus (STN; Fig. 6b,c). We found that cells within individual clones were notably more colocalized than non-clonally related cells (Fig. 6d). Strikingly, 53.6% of the 233 clones with at least three cells were entirely confined to a single spatial domain (clonal purity = 1), indicating that by E10, half of all hypothalamic progenitors are prespecified to populate a single, defined territory, such as a nucleus.

We next quantified spatial similarity between clones using the symmetric Wasserstein distance (Fig. 6e). Hierarchical clustering of these distances revealed a clear spatial structure in the resulting heat map. Clones broadly segregated into three spatial groups corresponding to the lateral, ventral and medial regions of the hypothalamus, mirroring its known anatomical organization (Fig. 6f). Within these groups, we identified highly localized subclusters that correspond to just a single nucleus, such as C3 in the ARH and C35 in the STN (Fig. 6g,h). Together, these results suggest that neighboring hypothalamic progenitors at E10 collectively acquire similar spatial coordinates in the adult hypothalamus, prefiguring the complex nuclear architecture of the mature brain.

Discussion

Here, we present Spatio-DARLIN, a high-resolution, generic platform for spatial lineage tracing in mice. This method simultaneously captures high-quality single-cell transcriptomes and recovers reliable lineage barcodes from ~25–50% of cells in the intestine and brain, yielding thousands of multi-cell clones. To overcome technical noise in the data, we developed a dedicated computational pipeline that ensures negligible barcode homoplasy and enables highly accurate, reproducible spatial clone detection across adjacent tissue sections. Crucially, unlike platforms tailored to specific tissues or cancer models, Spatio-DARLIN is a versatile tool validated for robust applications across diverse organs.

As a proof of concept, we applied Spatio-DARLIN to the adult intestinal epithelium and unveiled compartmentalized clones along the proximal–distal axis, as well as clonal expansion and migration along the crypt–villus axis. These results confirm that Spatio-DARLIN can accurately recapitulate established stem cell kinetics and quantify dynamic tissue turnover in situ. Applying Spatio-DARLIN to study more complex brain development, we not only reconstructed the radially located cone-shaped neuronal clones in the cortex but also revealed substantially more neuronal clone expansion in the cortex and hippocampus. More importantly, our data strongly suggest early spatial specification of neuronal progenitors across various hypothalamic nuclei. Although further validation of this concept awaits time-course barcoding and longitudinal profiling, our observations indicate a primed migration trajectory during neurogenesis. These findings demonstrate that Spatio-DARLIN is a powerful tool that can depict clonal architecture and resolve clonal expansion and migration in situ at single-cell resolution.

Despite the potential of Spatio-DARLIN, our approach is subject to several limitations that warrant consideration and future investigation. First, although the current protocol achieves decent lineage barcode recovery (25–50%), the detection efficiency varies across tissues and replicates, likely due to heterogeneity in transgene expression and capture dynamics. The sensitivity of Spatio-DARLIN could be further enhanced by using customized arrays with capture probes specifically targeting DARLIN barcodes. Second, our approach requires 2D tissue sectioning, which leads to clone subsampling and may generate clonal bias that requires careful interpretation. This issue can be mitigated through an informed choice of section orientation or the use of 3D reconstruction.

Spatio-DARLIN directly addresses the long-standing tradeoff between spatial context and informational depth that has hindered classical lineage-tracing approaches. By generating high-resolution maps that simultaneously capture the clonal origin of a cell, transcriptional state and precise spatial coordinates, we move closer to a full understanding of principles that orchestrate tissue development and homeostasis. DARLIN mice are publicly available from The Jackson Laboratory (038749 and 038750) and support pan-tissue labeling. Apart from working with the BMKMANU S3000 chip demonstrated here, its utility can be extended by porting it to other spatial transcriptomic platforms, such as Stereo-seq[37] and DBiT-seq[38]. We envision that Spatio-DARLIN will enable in situ clonal tracking across diverse tissues and biological processes and finally empower us to systematically decipher the complex dialog between cellular history and geography in health and disease.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of data and code availability are available at https://doi.org/10.1038/s41592-026-03151-5.

References

1. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148**, 33–45 (2012).
2. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
3. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
4. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
5. Pei, W. et al. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).
6. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
7. Pei, W. et al. Resolving fates and single-cell transcriptomes of hematopoietic stem cell clones by PolyloxExpress barcoding. *Cell Stem Cell* **27**, 383–395 (2020).
8. Chen, M., Fu, R., Chen, Y., Li, L. & Wang, S.-W. High-resolution, noninvasive single-cell lineage tracing in mice and humans based on DNA methylation epimutations. *Nat. Methods* **22**, 488–498 (2025).
9. Li, L. et al. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* **186**, 5183–5199 (2023).
10. Weber, T. S. et al. LoxCode in vivo barcoding reveals epiblast clonal fate bias to fetal organs. *Cell* **188**, 3882–3896 (2025).
11. Jindal, K. et al. Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. *Nat. Biotechnol.* **42**, 946–959 (2023).
12. Yang, D. et al. Lineage tracing reveals the phylodynamics, plasticity and paths of tumor evolution. *Cell* **185**, 1905–1923 (2022).

13. Quinn, J. J. et al. Single-cell lineages reveal the rates, routes and drivers of metastasis in cancer xenografts. *Science* **371**, eabc1944 (2021).
14. Lu, Z. et al. Polyclonal-to-monoclonal transition in colorectal precancerous evolution. *Nature* **636**, 233–240 (2024).
15. Shaffer, S. M. et al. Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. *Cell* **182**, 947–959 (2020).
16. Scherer, M. et al. Clonal tracing with somatic epimutations reveals dynamics of blood ageing.*Nature* **643**, 478–487 (2025).
17. Engblom, C. et al. Spatial transcriptomics of B cell and T cell receptors reveals lymphocyte clonal dynamics. *Science* **382**, eadf8486 (2023).
18. Ma, C. et al. Inferring allele-specific copy number aberrations and tumor phylogeography from spatially resolved transcriptomics. *Nat. Methods* **21**, 2239–2247 (2024).
19. Chow, K.-H. K. et al. Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021).
20. Koblan, L. W. et al. High-resolution spatial mapping of cell state and lineage dynamics in vivo with PEtracer. *Science* **390**, eadx3800 (2025).
21. Li, L. et al. DARLIN mouse for in vivo lineage tracing at high efficiency and clonal diversity. *Nat. Protoc.* **20**, 2319–2344 (2025).
22. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
23. Wang, S.-W., Herriges, M. J., Hurley, K., Kotton, D. N. & Klein, A. M. CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nat. Biotechnol.* **40**, 1066–1074 (2022).
24. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822–825 (2010).
25. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing *Lgr5* stem cells. *Cell* **143**, 134–144 (2010).
26. La Manno, G. et al. Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021).
27. Di Bella, D. J., Domínguez-Iturza, N., Brown, J. R. & Arlotta, P. Making Ramón y Cajal proud: development of cell identity and diversity in the cerebral cortex. *Neuron* **112**, 2091–2111 (2024).
28. He, Z. et al. Lineage recording in human cerebral organoids. *Nat. Methods* **19**, 90–99 (2022).
29. Ratz, M. et al. Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat. Neurosci.* **25**, 285–294 (2022).
30. Jones, M. G. et al. Spatiotemporal lineage tracing reveals the dynamic spatial architecture of tumor growth and metastasis. Preprint at *bioRxiv* https://doi.org/10.1101/2024.10.21.619529 (2024).
31. Erickson, A. G. et al. Unbiased profiling of multipotency landscapes reveals spatial modulators of clonal fate biases. Preprint at *bioRxiv* https://doi.org/10.1101/2024.11.15.623687 (2025).
32. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
33. Gao, P. et al. Deterministic progenitor behavior and unitary production of neurons in the neocortex. *Cell* **159**, 775–788 (2014).
34. Zhang, Y.-H. et al. Cascade diversification directs generation of neuronal diversity in the hypothalamus. *Cell Stem Cell* **28**, 1483–1499 (2021).
35. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
36. Singhal, V. et al. BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. *Nat. Genet.* **56**, 431–441 (2024).
37. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
38. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 (2020).

Methods

Dox administration in DARLIN mice

Cas9-TdT-gRNAs-M2 mice (The Jackson Laboratory, 038749) and CA/TA/RA triple target-array mice (The Jackson Laboratory, 038750) were used in this study. The homozygous *Col1a1*$^{tetO-Cas9-TdT-gRNAs}$:*Rosa26*$^{M2-rtTA}$ strain and *Col1a1*:^{CA}TIGRE^{TA}:*Rosa26*^{RA} strain were maintained separately, and DARLIN mice were generated by crossing these two strains. For timed pregnancy experiments, male and female mice were housed together by night, and the vaginas of the female mice were examined the next morning. The presence of a vaginal plug was recorded as E0.5. To induce barcoding, an 8-week-old DARLIN mouse was treated with 2 mg ml^{-1} Dox (ABCONE, D48978) in drinking water (supplemented with 10 mg ml^{-1} sucrose) and received three intraperitoneal injections (50 µg per g (body weight)) for 1 week. For the time-course tracing experiment, 8-week-old DARLIN mice were induced by Dox for 1 week, and the small intestines were collected after another 1, 2 and 3 weeks, respectively. To label embryonic-stage mice, pregnant dams were anesthetized and injected with a single dose of Dox (50 µg per g (body weight)) through the retro-orbital route. Mice were housed in a barrier facility on a 12-h dark/12-h light cycle, a temperature of 20–26 °C and a humidity of 40–70%. All animal procedures were approved by the Institutional Animal Care and Use Committee of Westlake University (AP#24-072-LL).

Small intestine processing

Mice were perfused with 20 ml of precooled PBS to remove blood, and the small intestine was dissected, the intestinal lumen was rinsed with precooled PBS, and adipose tissue was removed. The small intestine was divided into three equal-length segments, namely the PSI, MSI and DSI. Each segment was cut longitudinally, starting from the most distal portion. With the luminal side facing upward, the segment was coiled into a Swiss roll with the distal part in the center and the proximal part in the peripheral region. The fresh tissue was then embedded in a 1 cm × 1 cm container with optimal cutting temperature (OCT) compound (Sakura Tissue-TEK, 4583) on dry ice and stored at −80 °C immediately.

Brain tissue collection

Mice were deeply anesthetized with isoflurane and transcardially perfused with 40 ml of precooled PBS to remove blood. The brain was then carefully dissected from the skull and washed in precooled PBS to remove any residual debris. Sterile gauze was used to gently dry the PBS on the brain surface. The fresh tissue was embedded in a 1 cm × 1 cm container with OCT compound (Sakura Tissue-TEK, 4583) on dry ice and stored at −80 °C immediately. For the biological replicate L126, serial coronal sections were prepared to capture the region of interest. To ensure representative coverage, we used an intermittent sampling strategy: 10-µm sections were initially collected, followed by discarding three consecutive 10-µm sections before collecting the next section. This yielded a total of three sections spanning an axial distance of 90 µm.

Construction and sequencing of the spatial transcriptomic library

The spatial transcriptomic library was constructed with a BMKMANU S3000 Gene Expression kit (BMKMANU, ST03010). OCT-embedded blocks were sectioned at a thickness of 10 µm on tissue slides using a precooled cryostat. Tissue sections were transferred to the 6.8 × 6.8 mm^2 BMKMANU S3000 spatial chip, fixed with methanol, stained with H&E and imaged using a Pannoramic MIDI II (3DHISTECH). The spatial transcriptome library was then constructed according to the manufacturer's instructions and sequenced for approximately 270 Gbp using an Illumina NovaSeq 6000 or SURFSeq 5000 instrument.

Construction and sequencing of the DARLIN barcode amplicon

To perform targeted amplification of DARLIN barcodes, we used the spatially resolved cDNA as starting material and applied a nested PCR approach. We constructed the CA/TA/RA spatial libraries separately. The primers used in this study are included in Supplementary Table 1. KAPA HiFi HotStart ReadyMix (Roche Applied Science, 07958935001) was used to perform targeted amplification. SPRIselect Reagent Beads (0.8×; Beckman Coulter, B23318) were used to purify the PCR product for each step. Specifically, we amplified the CA, TA and RA array libraries for nine cycles using one-fifth of the purified spatial cDNA, the P5_PR1 primer and the NGS_CA_R1, NGS_TA_R1 or NGS_RA_R1 primer. The three PCR products were then purified, and each was subjected to a second PCR with nine cycles, again using the P5_PR1 primer and the NGS_CA_R2, NGS_TA_R2 or NGS_RA_R2 primer. Each of the resulting PCR products was further purified and used to perform an indexing PCR for nine cycles using the primers included in the kit (New England Biosciences, E6609S).

The final indexing DARLIN amplicons were pooled and quantified using a VAHTS Library Quantification kit for Illumina (Vazyme, NQ101) and sequenced on an Illumina NovaSeq 6000 using NovaSeq 6000 SP kits (read 1: 85 cycles; i7 index: 8 cycles; read 2: 350 cycles; Illumina, MS-102-2003) with 10% PhiX sequencing control v3 (Illumina, FC-110-3001).

Bulk DARLIN library preparation for regional clone size analysis

To evaluate clonal expansion across neuroanatomical regions, DARLIN mice induced at E10 were perfused with ice-cold PBS. The dura mater and hypothalamus were isolated, followed by bilateral dissection of the dorsal cortex, ventral cortex, hippocampus and thalamus. Genomic DNA (gDNA) was extracted using a DNeasy Blood and Tissue kit (Qiagen, 69504) and quantified with a Qubit dsDNA HS kit (Invitrogen, Q32854) to estimate cell input.

To control sampling depth, exactly 10% of the total gDNA from each sample was used for library construction. Each sample had two technical replicates. UMIs were incorporated via five cycles of linear amplification using RT_CA primers, followed by purification with 1× AMPure XP beads (Beckman Coulter, A63881). CA amplicon libraries were generated using a standard protocol involving 12 cycles of PCR and subsequent 12 cycles of seminested PCR to enrich lineage barcodes. Final libraries were indexed via a ten-cycle PCR with sample-specific barcodes and purified using 0.8× AMPure XP beads before sequencing.

qPCR for DARLIN array expression

Total RNA was extracted from approximately 20 tissue cryosections (10 µm thickness) using TRIzol reagent (Invitrogen, 15596018CN) following the manufacturer's protocol. RNA concentration was determined using a Qubit RNA HS Assay kit (Invitrogen, Q32852). Subsequently, 1 µg of total RNA was subjected to gDNA removal and reverse transcription using a HiScript IV 1st Strand cDNA Synthesis kit (+gDNA wiper; Vazyme, R412-01). The resulting cDNA was analyzed by qPCR using DARLIN array-specific primers (Supplementary Table 1) and SupRealQ Ultra Hunter SYBR qPCR Master Mix (U+; Vazyme, Q713-02). All reactions were performed on a QuantStudio 6 Pro Real-Time PCR system (Applied Biosystems).

Data preprocessing and QC of spatial transcriptomic data

The count matrix of segmented cells from the BMKMANU S3000 chip was generated using BSTMatrix (v2.4.f.1). Briefly, the spot barcode was extracted from read 1 and assigned to spatial coordinates. The mRNA sequence from read 2 was aligned to the mm10 reference genome using STAR (v2.7.10b). Cellpose (v2.2.2) was used to segment nuclei from single-stranded DNA-stained images (19,954 × 20,000 pixels). To infer whole-cell boundaries, we used flow-field-based mask estimation of Cellpose, specifying a target diameter of 35 pixels (~11.9 µm) for nuclear segmentation and an expansion of 35 pixels to generate cell masks[39]. The gene UMI counts of spots in the same cell mask were summed to generate a cell-level count matrix. The position of the cell was the center of the merged spots. For QC, we filtered the count matrices to remove

cells with <100 detected genes. Below, we use 'dataset' to refer to the data generated from each tissue section using the BMKMANU S3000 chip. Different chips generate different datasets that are considered separate batches.

Analysis of intestinal spatial transcriptomic data

To integrate spatial transcriptomic datasets from different tissue sections (that is, batches), we performed a sketch-based integration using Seurat (v5.1.0)[40]. Briefly, we built a core atlas of the small intestine (PSI core) using the subsampled cells from two proximal intestine datasets from mice Y004 and Y007 mice and projected other cells to the PSI core through reference mapping.

To build the PSI core data, we downsampled both the Y004 and Y007 datasets to 25,000 cells, thus generating a combined dataset of 50,000 cells, with equal contribution from both batches. Here, each cell was weighted by the leverage score (relative contribution to the total variance) evaluated on the top 2,000 highly variable genes (HVGs) of each dataset[40]. With this PSI core matrix, we performed log normalization, recomputed the top 2,000 HVGs, generated a sub-matrix restricting to these HVGs, performed gene normalization and carried out principal component analysis (PCA) to obtain the first 50 principal components. Data integration was performed through the anchor-reciprocal PCA procedure using the IntegrateLayers function in Seurat. We used the top 50 batch-corrected principal components to build a k-nearest neighbor (kNN) graph, with $k = 20$, and performed Louvain clustering with a resolution of 2. Finally, we performed UMAP to generate the 2D embedding for visualization.

To project other cells onto this PSI core atlas, we calculated the anchors between the query and reference, projected them to batch-corrected PCA and UMAP embeddings and transferred the clustering labels from the reference (PSI core) to the query (other cells) through a kNN classifier ($k = 10$). The cell-type annotation of these clusters was performed manually based on marker gene expression.

Computational unrolling of the coiled small intestine

Below, we unrolled the coiled small intestine to facilitate computational analysis. This procedure was inspired by a previous study[41], but with a customized adaptation. A schematic of this procedure is shown in Extended Data Fig. 4a.

First, to build the longitudinal axis (x axis) from the proximal intestine to the distal intestine, we manually drew a continuous green line (reference line) along the serosa layer on the low-resolution H&E image using QuPath (v0.5.1). The pixel coordinates of this line were extracted by thresholding the red–green–blue channels (red < 80%, green > 80%, blue < 80%)[42]. The coordinates of these pixels were used to construct a k-nearest-neighbor graph ($k = 10$) to identify the shortest path from the initial to the final pixel along the reference line using Dijkstra's algorithm[43]. The cumulative distance along this path was then used to define the longitudinal coordinate (x axis).

To infer the coordinate of each cell along the crypt–villus axis (y axis), we projected cells onto the x axis and computed the perpendicular distance. This is tricky due to the irregular folding of the intestinal roll. To address this issue, we implemented an iterative projection strategy. In the first round, we directly projected 'root' cells, defined as cells that were within 40 pixels of the green serosa line, onto the nearest point on the reference line (cell-to-reference projection). We denoted the set of projected cells as S, and the set of remaining cells as R. In subsequent rounds, we identified cells in R that were within 40 pixels from a cell in S and projected these R cells onto certain S cells that have minimum cell-to-cell distances (cell-to-cell projection). Afterward, we updated the set R and S and repeated this procedure until there were no cells left. Effectively, this generated a nearest-neighbor graph for these cells, which enabled us to propagate the coordinates from the initial root cells outward, ensuring robust projection even in irregularly folded regions.

After these steps, each cell was assigned coordinates along both axes. The x axis represents the position along the longitudinal serosa track, and the y axis corresponds to the perpendicular distance from each cell to the serosa line, indicating the coordinate on the crypt–villus axis.

DARLIN barcode recovery

The snakemake_DARLIN pipeline was previously developed for processing single-cell or bulk DARLIN datasets generated from the DARLIN mouse model. To enable allele calling from BMK S3000 spatial transcriptomic data, we developed an optimized pipeline inspired by this original DARLIN workflow, with adaptations to address the following specific challenges: (1) various sequencing and amplification artifacts in this targeted library and (2) the high-throughput nature of spatial transcriptomics, involving millions of spatial spots (~414 million) and resulting in massive total read volumes that exceed the processing capacity of previous workflows.

Here, raw sequencing data from Spatio-DARLIN amplicon libraries were processed as paired-end reads, where read 1 (85 base pairs) contained the spot barcode and unique molecular identifier (UMI; labeled UB in Extended Data Fig. 2), and read 2 (350 base pairs) encompassed the lineage barcode and flanking sequences (Extended Data Fig. 2a). Lineage barcodes were extracted from read 2 using cutadapt (v5.0). Concurrently, spot barcodes and UMIs were extracted from the paired read 1 and corrected for sequencing errors using BSTMatrix (v2.4.f.1), generating revised corrected spot barcodes and corrected UMIs. Subsequently, a customized three-step QC pipeline was implemented to mitigate artifacts inherent to amplicon sequencing (Extended Data Fig. 2b).

First, to correct PCR and sequencing errors within the lineage barcodes, raw lineage barcodes were clustered using umi_tools (v1.1.5). Directional clustering was applied with a length-aware Hamming distance threshold set to a default error rate of 0.02, generating corrected lineage barcodes (Extended Data Fig. 2c). Second, we filtered amplification artifacts in which incomplete products from previous PCR cycles prime subsequent reactions, causing a single molecular identifier (corrected spot barcodes + corrected UMI combination) to erroneously associate with multiple distinct corrected lineage barcodes. We calculated the fraction of reads supporting each specific corrected lineage within a unique molecule. Based on the observed bimodal distribution of these fractions, only lineage barcodes supported by ≥80% of reads within a molecule were retained (Extended Data Fig. 2d). Third, capture-oligonucleotide carryover artifacts were removed at the spot level. These artifacts arise when capture oligos containing spot barcodes dissociate from spatial beads and serve as primers in solution during PCR, incorrectly assigning spatial coordinates to lineage barcodes from unrelated locations. To address this, we calculated k, the average number of supporting reads per UMI for each spatial spot, and applied a stringent filter to retain only spots with $k \geq 10$ (Extended Data Fig. 2e–g). This stepwise pipeline systematically reduced sequence-level and spot-level noise to yield a high-confidence set of spatially resolved lineage barcodes. We have extensively validated the robustness of our pipeline to sequencing depth, sequencing errors and cell segmentation errors (Supplementary Figs. 1–3). This dedicated denoising pipeline is implemented at https://github.com/LLi-Lab/spatio_DARLIN/, with detailed documentation.

After these steps, we determined the mutation patterns by mapping the resulting DARLIN barcodes to their reference sequences. This step requires the MATLAB-based CARLIN pipeline, which is computationally expensive. Calling mutations at the end rather than at the beginning of the preprocessing greatly speeds up the computation, which addresses the second challenge of scalability.

Finally, the UMI counts for each barcode were aggregated across spots belonging to the same cell mask. The cell mask was generated through cell segmentation from the spatial transcriptomics preprocessing step. This provided the cell-by-barcode UMI count matrix for

downstream analyses. As in the spot-level QC, we also excluded spurious cells that did not have a dominant lineage barcode from any recording loci.

Identification of rare barcodes and reliable clones

We followed our previous method to identify rare barcodes from each locus[9,21]. Specifically, rare barcodes were barcodes with an inferred generation probability ρ lower than a cutoff probability ρ_*, that is, $\rho \leq \rho_*$. We adopted the cutoff $\rho_* = 1 \times 10^{-5}$, which should reliably label ~10,000 clones at a false discovery rate of 0.05. Because each tissue section generates ~10,000 clones, this choice should be sufficient for our scenario.

Because each cell may have three barcodes from different loci (CA, TA and RA), we integrated this information to call clones. We followed our previous method for clone identification[9,21]. Briefly, we generated a cell-based graph, where two cells were connected if they shared at least one rare barcode as identified above. Connected components in this graph were identified as clones. To minimize the inclusion of unrelated founder cells, only clones containing no more than three distinct alleles from each of the CA, TA and RA loci were retained for downstream analyses. Furthermore, unusually large clones (that is, > 100 cells) were also excluded, which can be determined from the clone size distribution.

Validation of clone detection accuracy

To assess the spatial coherence of epithelial clones across adjacent tissue sections, we first corrected the horizontal shift between sections by aligning their tissue centroids. Cells from different sections were then projected onto the same space to remove their z axis difference along tissue depth. For cells from the same clone but detected across sections, we computed the Wasserstein distance (a measure from optimal transport theory[44]) to quantify the minimal cumulative shift needed to convert cells detected in one section to cells in another (Fig. 2e). A small distance implies stronger spatial coherence within the same clone across adjacent tissue sections, suggesting a high accuracy of clone detection.

To obtain a negative control without spatial coherence, we generated scrambled datasets by randomly permuting clone labels, separately for each tissue section, which were also used to compute Wasserstein distances as negative controls (Fig. 2f).

Epithelial clone width estimation

We used an autoregression-based strategy to estimate the width of epithelial clones along the proximal–distal axis (Fig. 3f). For each clone, we first calculated the cell density distribution along the proximal–distal axis using a kernel density estimator. A step function was then fitted to the kernel density estimator by applying a cutoff at half the maximal density value. Finally, the autocorrelation function of the step function was computed, and the lag at which the autocorrelation dropped below a threshold of 0.1 was identified as the clone width. A schematic of this procedure is shown in Extended Data Fig. 4d.

Clonal coupling analysis

The clonal coupling score is a normalized correlation coefficient that reflects how often two groups share cells from the same clone. The two groups can be spatial regions, such as tissue slides (Fig. 2c) or villus zones (Fig. 3i). The score was calculated using the cospar.tl.pvalue_for_fate_coupling function in CoSpar (v.0.4.1)[23], using an updated version described in ref. 9.

Analysis of brain spatial transcriptomics data

The gene count matrix was log normalized and scaled based on the top 2,000 HVGs. PCA was conducted, followed by the construction of a shared nearest-neighbor (SNN) graph ($k = 20$) from the top 50 principal components for dimensionality reduction. A UMAP was generated for visualization (Fig. 4b and Extended Data Fig. 8a–d).

Cell clustering was performed using the Louvain algorithm (resolution = 3) applied to the SNN graph, with subsequent refinement through iterative subclustering (resolution = 0.4). For annotation, a supervised deep learning classifier was implemented. This model, architecturally analogous to SuperCT[45], was trained on binarized expression data from two integrated mouse brain single-nucleus RNA-sequencing atlases[35,46] to predict cell identities. Cluster annotations were assigned based on the majority label from classifier predictions, with discrepancies adjudicated through manual curation guided by established marker genes.

Spatial domains were delineated using the BANKSY algorithm[36], which incorporates spatial neighborhood information ($\lambda = 0.5, k = 50$) into the expression matrix of top HVGs. Independent component analysis was applied to this matrix, and the top 20 components were used for UMAP visualization and to construct a spatially informed SNN graph ($k = 50$). Spatial domains were ultimately identified by applying Louvain clustering (resolution = 0.8) to this graph.

Tangential coordinate assignment for cortical and hippocampal CA regions

A principal curve was fitted to the spatial coordinates of cells using the princurve R package (v2.1.6)[47]. For each cell, a tangential coordinate was assigned, defined as the distance along the curve from a manually defined starting point to the projection of the cell onto the curve (Fig. 5a,b,e,f).

Clonal clustering by transcriptome similarity

Clonal heterogeneity was resolved using the clone2vec algorithm (Fig. 4e)[31], which embeds clones into a continuous latent space based on the transcriptional similarity of their progeny. The algorithm adapts the skip-gram architecture from word2vec, treating clones as tokens and cells as context. Embeddings were trained in 10 dimensions. A 2D projection was then generated using UMAP for visualization. To define clonal clusters, an SNN graph ($k = 15$) was constructed from the latent embeddings, and clustering was performed on this graph using the Leiden algorithm at a resolution of 1.0.

Clonal clustering by spatial proximity

To characterize the spatial architecture of hypothalamic clones, we constructed a pairwise distance matrix using a symmetric Wasserstein metric. Cells within each clone were modeled as weighted spatial point sets with uniform mass. We computed the Wasserstein distance $W(\mathrm{A} \to \mathrm{B})$ from clone A to B using the wasserstein() function in the R package transport (v0.15.4). To ensure symmetry, the final distance $d_{\mathrm{A,B}}$ was defined as the mean of the bidirectional transport costs: $d_{\mathrm{A,B}} = (W(\mathrm{A} \to \mathrm{B}) + W(\mathrm{B} \to \mathrm{A}))/2$. Hierarchical clustering was subsequently performed on the resulting distance matrix using Ward's method (ward.D2 in hclust).

Bulk genomic DNA amplicon data analysis

Paired-end FASTQ files were processed using a custom Python pipeline (https://github.com/JarningGau/darlinpy). Reads were assembled using PEAR (v0.9.11), and lineage barcodes and UMIs were extracted based on fixed flanking sequences. Barcodes shorter than 20 nucleotides or supported by fewer than two reads were discarded. Lineage sequences were error corrected using a length-dependent Hamming distance threshold (relative distance of 0.02), and UMIs were deduplicated via directional clustering (umi_tools) to generate an allele-by-sample count matrix. To estimate absolute clone sizes, observed UMI counts were normalized by the effective sampling efficiency ($\alpha \times \beta$), defined as the product of the library sampling rate (α, input / extracted gDNA) and sequencing sampling rate (β, detected UMIs / input cells). Homoplasic alleles were filtered using a precompiled allele bank, and reproducibility was assessed via the Jaccard index. Finally, alleles detected across multiple brain regions were excluded to prevent contamination from circulating cells or microglia.

Statistics and reproducibility

No statistical method was used to predetermine sample size. Pre-established QC criteria were applied as described above, and cells with fewer than 100 expressed genes were excluded from downstream analyses. Key findings were reproduced in independent biological samples, as described in the Results and figure legends. The experiments were not randomized. Investigators were not blinded to group allocation during experiments or data analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The spatial transcriptomic datasets and spatial DARLIN datasets were deposited at the NCBI Gene Expression Omnibus under accession numbers GSE296896 and GSE296897, respectively. Analysis parameters for each dataset used in this study are available in Supplementary Table 2.

Code availability

Scripts for data preprocessing are available at https://github.com/LLi-Lab/spatio_DARLIN.

References

39. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
40. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2023).
41. Parigi, S. M. et al. The spatial transcriptomic landscape of the healing mouse intestine following damage. *Nat. Commun.* **13**, 828 (2022).
42. Mayassi, T. et al. Spatially restricted immune and microbiota-driven adaptation of the gut. *Nature* **636**, 447–456 (2024).
43. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
44. Peyré, G. & Cuturi, M. Computational optimal transport: with applications to data science. *Found. Trends Mach. Learn.* **11**, 355–607 (2019).
45. Xie, P. et al. SuperCT: a supervised-learning framework for enhanced characterization of single-cell transcriptomic profiles. *Nucleic Acids Res.* **47**, e48 (2019).
46. Han, L. et al. Single-cell spatial transcriptomic atlas of the whole mouse brain. *Neuron* **113**, 2141–2160 (2025).
47. Hastie, T. & Stuetzle, W. Principal curves. *J. Am. Stat. Assoc.* **84**, 502–516 (1989).

Acknowledgements

We are grateful to H. Xu and Y. Wang for their insightful discussions. We thank the Westlake University High-Performance Computing Center for computing resources.

Author contributions

L.L. and S.-W.W. conceived the project. Z.Z. performed the experiments with help from D.C., S.D. and S.L. J.G. developed the computational pipeline and conducted data analysis. L.L. and J.G. wrote the first draft of the paper with help from Z.Z. S.-W.W. and L.L. finalized the paper. L.L. and S.-W.W. supervised the project.

Funding

L.L. acknowledges support for the research of this work from the National Key R&D Program of China (2024YFA1306603), the National Natural Science Foundation of China (32595481 and 32570953) and R&D Programs of Zhejiang province (2024SSYS0034). S.-W.W. acknowledges support for the research of this work from the National Natural Science Foundation of China (32470700) and Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM502).

Competing interests

The authors declare no competing interests.

Additional information

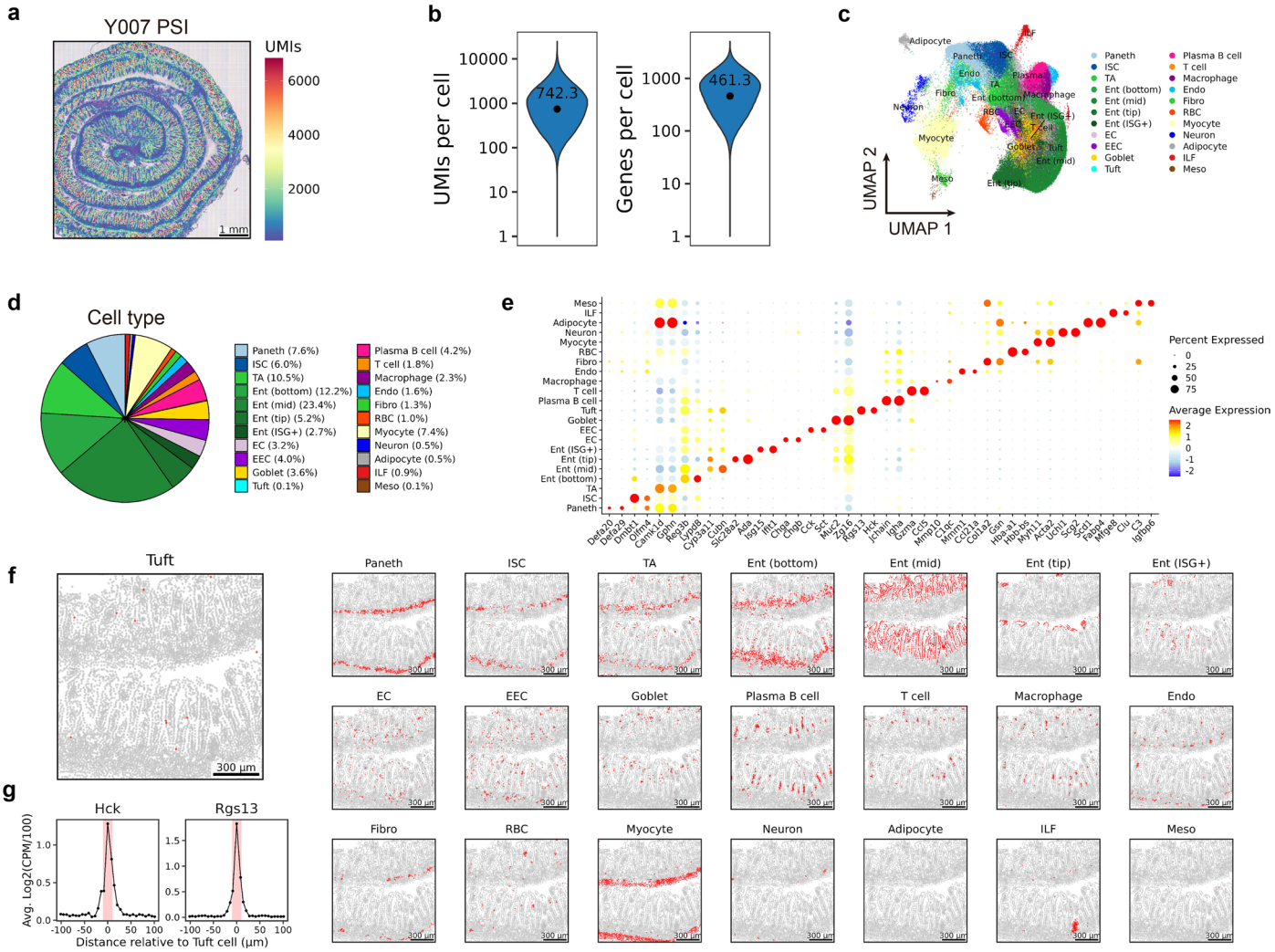
Extended data is available for this paper at https://doi.org/10.1038/s41592-026-03151-5.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-026-03151-5.

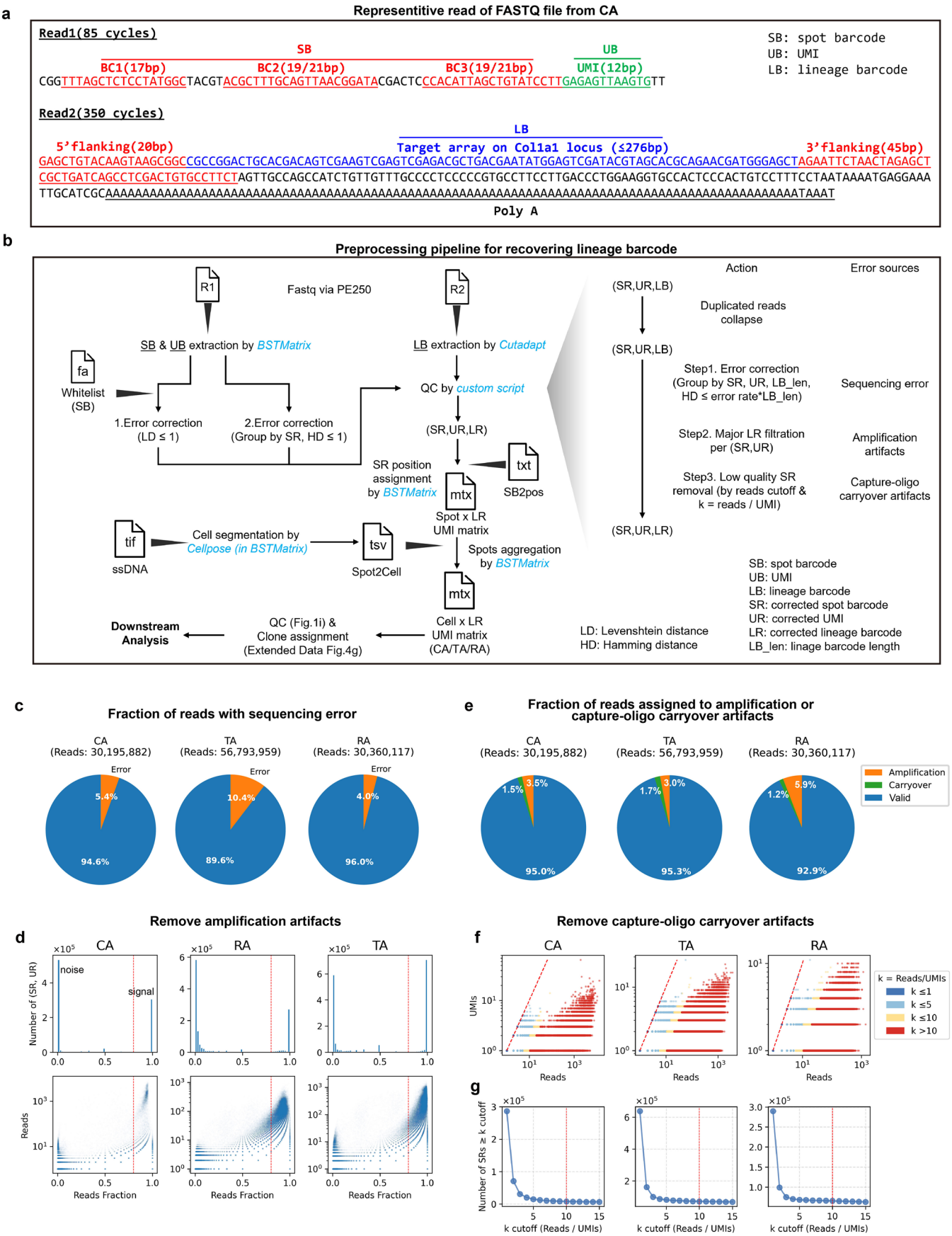
Correspondence and requests for materials should be addressed to Shou-Wen Wang or Li Li.

Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: M. Mukhopadhyay, in collaboration with the *Nature Methods* team.

https://doi.org/10.1038/s41592-026-03151-5



Extended Data Fig. 1 | Overview of Spatio-DARLIN data associated with Y007 PSI. a, Spatial map of UMI counts per cell for Y007 PSI generated in Fig. 1b. **b**, Violin plots showing the distribution of UMI counts (left) and gene counts (right) per cell. Mean values are indicated. **c**, Transcriptomic UMAP embedding colored by assigned cell types. ISC, intestinal stem cell; TA, transit-amplifying cell; Ent (bottom/mid/tip), enterocyte at the bottom, middle, or tip of the villus; Ent (ISG +), interferon-responsive enterocyte; EEC, enteroendocrine cell; EC, enterochromaffin cell; Plasma, plasma B cell; Endo, endothelial cell; Fibro, fibroblast; RBC, red blood cell; ILF, isolated lymphoid follicle-associated cell; Meso, mesothelial cell. **d**, Pie chart summarizing the proportion of each cell type. **e**, Dot plot showing the expression of selected marker genes across annotated cell types. Dot size indicates the proportion of cells expressing the gene; color reflects average log normalized expression. **f**, Spatial maps of all annotated cell types in a representative window, as in Fig. 1e. **g**, Average expression of *Hck* and *Rgs13* as a function of distance from nearest Tuft cells. CPM, count per million.

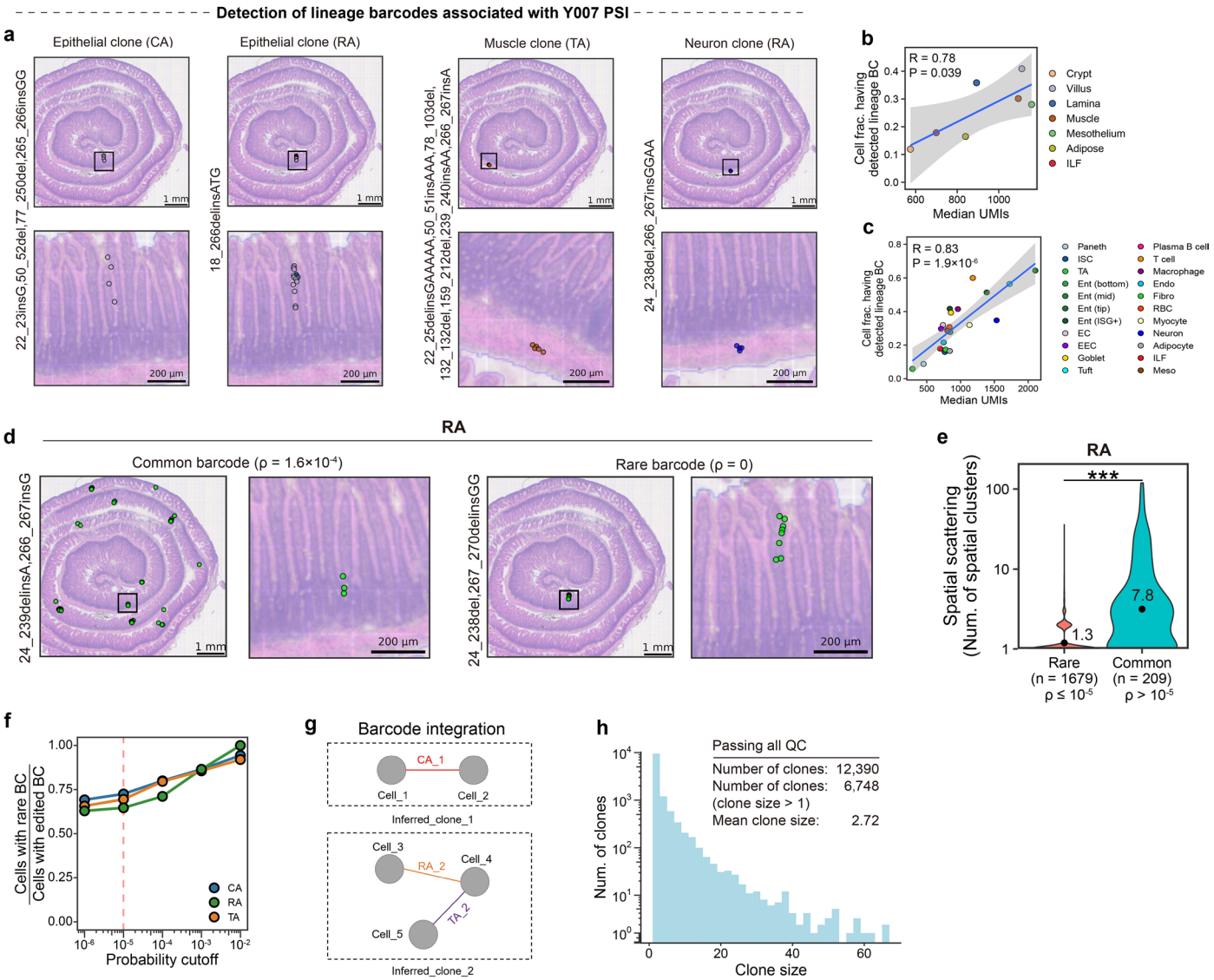


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41592-026-03151-5

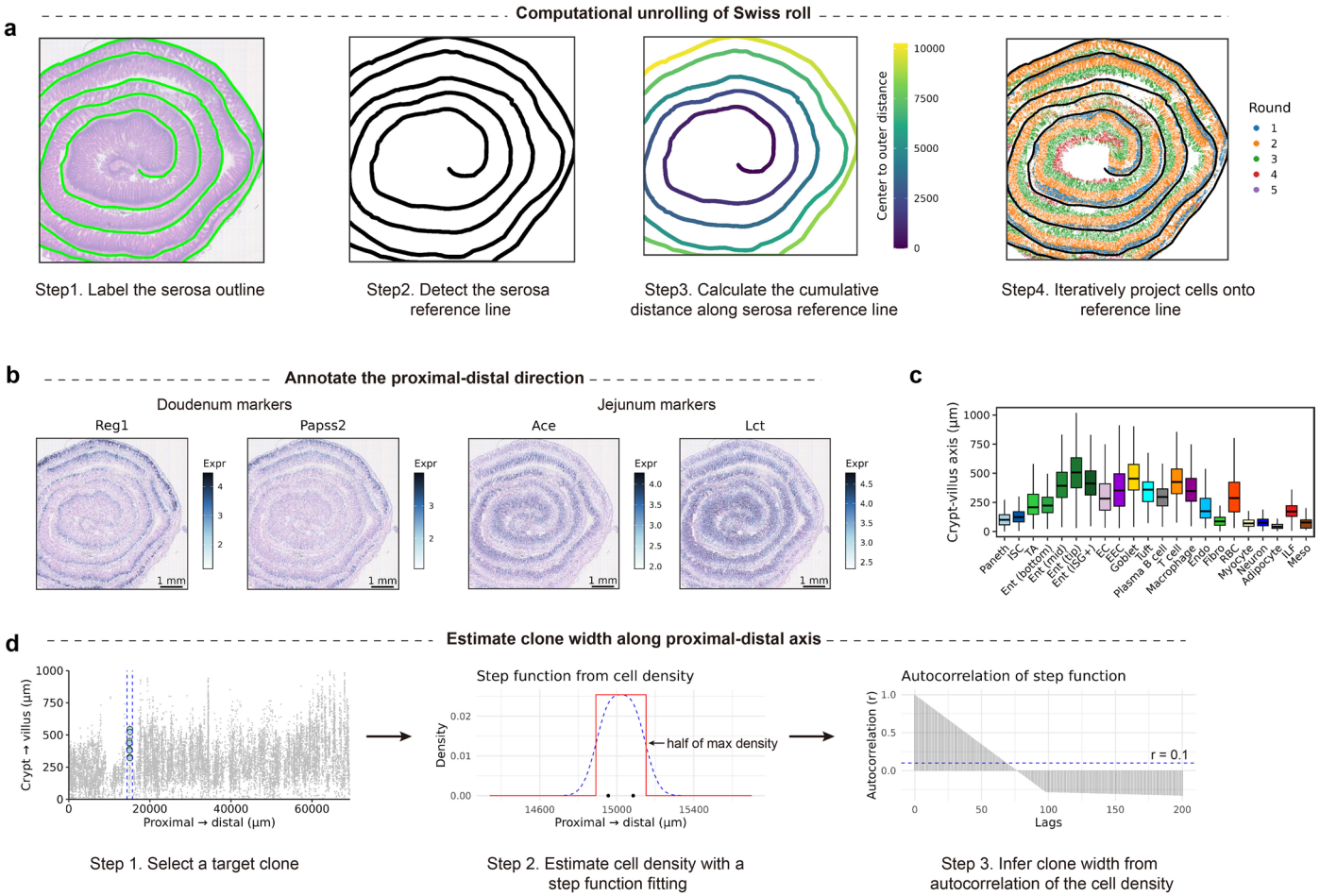
Extended Data Fig. 2 | Read structure, preprocessing pipeline, and quality control for lineage barcode recovery. a, Representative paired-end read structure of CA amplicon library: Read 1 contains the spot barcodes (SB, BC1–BC3) and UMI (UB), and Read 2 spans the lineage barcode located at Colla1 locus (LB), flanked by 5′ and 3′ constant sequences and followed by a poly(A) tail. **b**, Schematic of the spatial and lineage barcode decoder pipeline for Spatio-DARLIN amplicon data. **c**, Fraction of reads containing correctable sequencing errors (marked in orange) in CA, TA, and RA libraries. **d**, Identification and removal of amplification artifacts based on the distribution of read fractions and read counts per (SR, UR) pair. Dashed lines indicate thresholds separating signal from noise. SR, corrected spot barcode. UR, corrected UMI. **e**, Fraction of reads classified as amplification artifacts (orange), capture-oligo carryover artifacts (green), or valid reads in CA, TA, and RA libraries. **f**, Read–UMI relationships for CA, TA, and RA libraries, colored by the read-per-UMI ratio (k). Each dot corresponds to a unique spatial spot. **g**, Effect of varying the k cutoff (reads per UMI) on the number of retained spot barcodes, with dashed lines indicating the cutoff used in this paper ($k \geq 10$).

https://doi.org/10.1038/s41592-026-03151-5



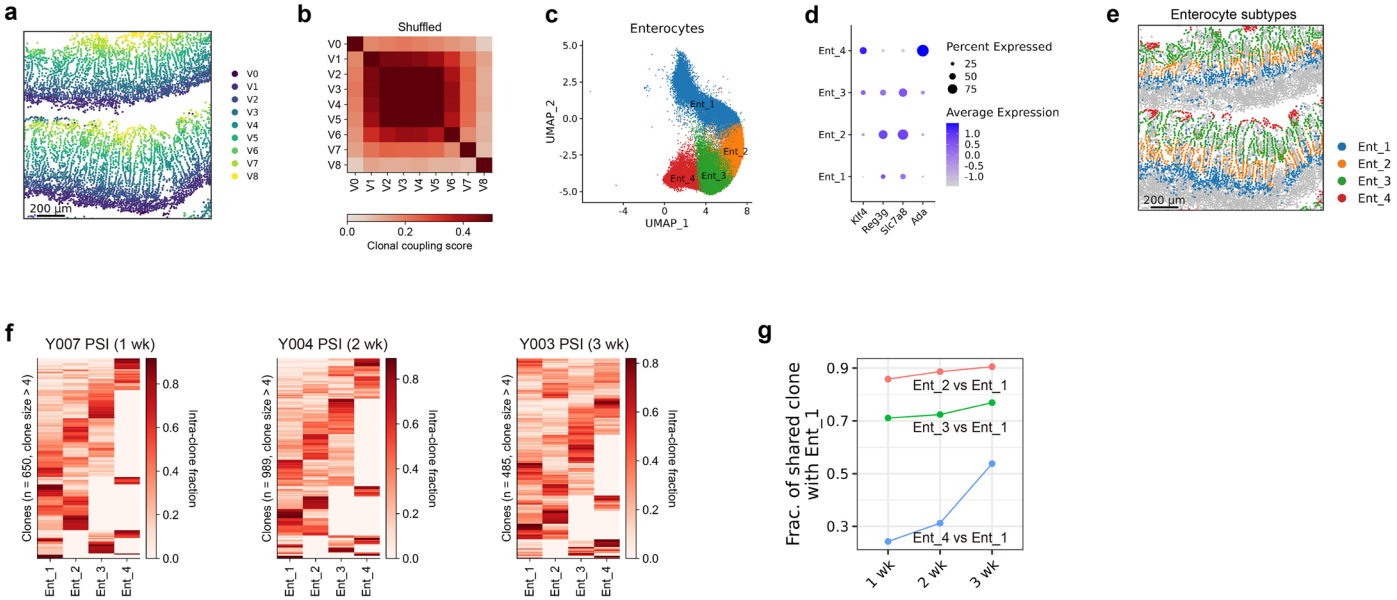
Extended Data Fig. 3 | Detection and quality control of lineage barcodes. **a**, Spatial map of representative clones: two epithelial clones (CA and RA), a muscle clone (TA), and a neuron clone (RA). The upper panels show the full view while the bottom panels give the zoom-in view. **b,c**, Correlation between DARLIN barcode recovery and the median UMI counts detected in each spatial domain (**b**, Pearson R = 0.78, P value = 0.039) or each cell type (**c**, Pearson R = 0.83, P value = 1.9×10^{-6}). Solid lines indicate linear regression fits, with shaded areas representing 95% confidence intervals. **d**, Representative spatial maps of cells associated with common or rare barcodes from RA. The barcode generation probability ρ estimated from the allele bank is indicated for each barcode. **e**, Violin plot of spatial scattering for both the rare barcodes ($\rho \leq 10^{-5}$) and common barcodes ($\rho > 10^{-5}$). Spatial scattering is measured by the number of spatially distinct cell clusters associated with a barcode. The mean cluster number is indicated for each case. ***, $P < 2.2\times10^{-16}$, two-sided Wilcoxon signed-rank test. **f**, Fraction of edited cells that have rare barcodes across varying barcode generation probability cutoffs. The red dashed line indicates the cutoff at 10^{-5}. **g**, Schematic for calling clones by integrating lineage barcodes from three loci. **h**, Size distribution for all clones passing QC filters, with clone statistics indicated.

https://doi.org/10.1038/s41592-026-03151-5



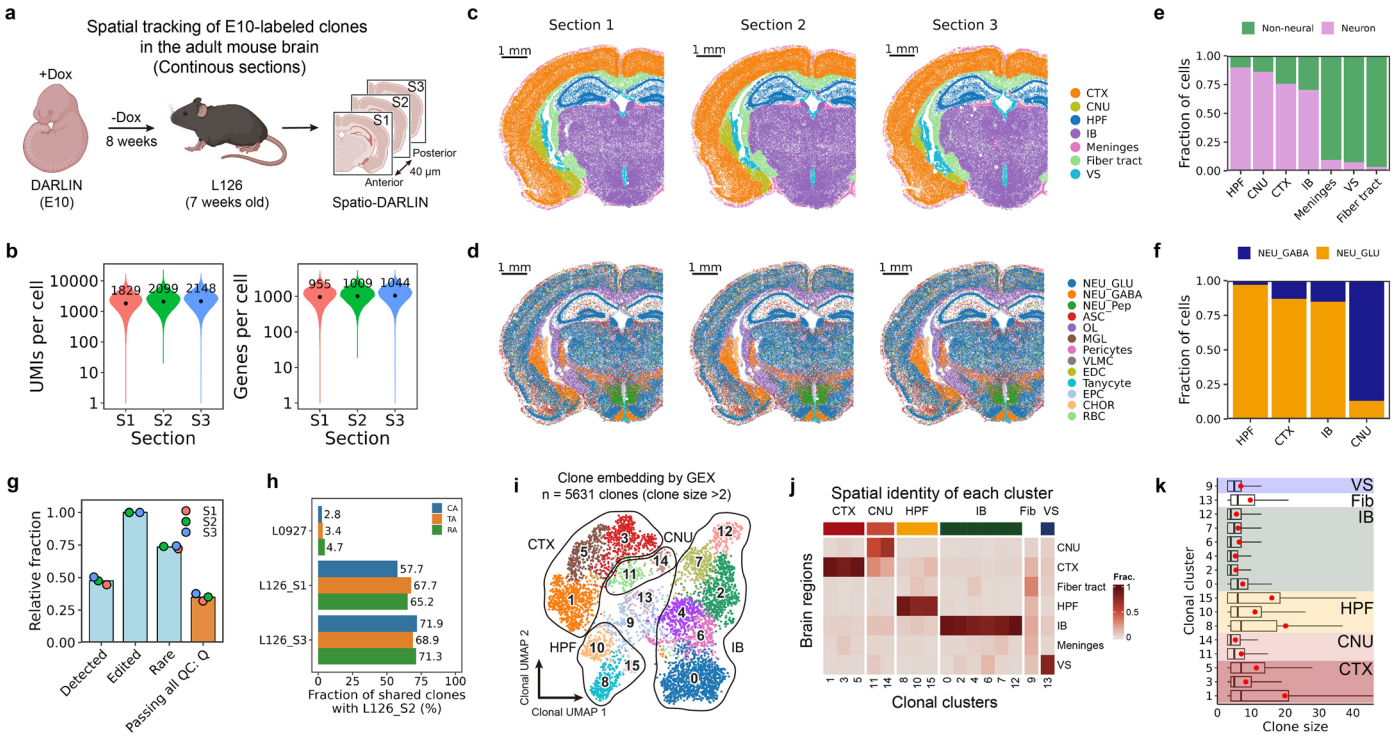
Extended Data Fig. 4 | Computational unrolling and clone width estimation in the intestine. a, Schematic of computational unrolling of the coiled intestine. **b**, Spatial expression of region-specific marker genes. *Reg1* and *Papss2* mark the duodenum, while *Ace* and *Lct* mark the jejunum, confirming spatial fidelity of axis annotation. **c**, Box plot of the inferred position along the crypt–villus axis for each cell type. The sample size (n) corresponds to the number of cells per cell type. Cells are treated as independent units of observation within the analyzed sample (Y007 PSI). Box plots show the median (50th percentile), the bounds of the box represent the interquartile range (25th to 75th percentile) and whiskers extend to 1.5 times the interquartile range. **d**, Schematic of clone width estimation.

https://doi.org/10.1038/s41592-026-03151-5



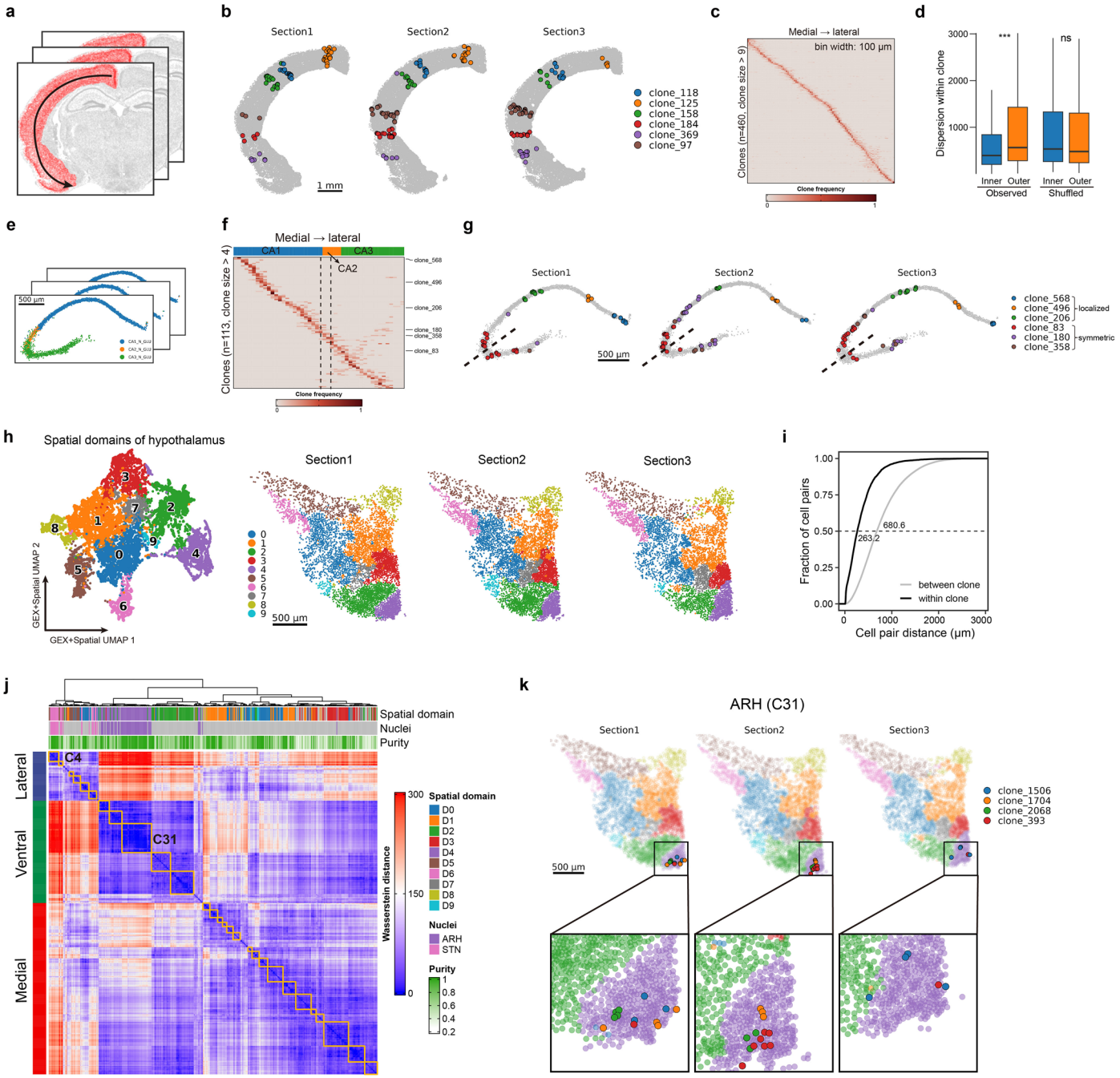
Extended Data Fig. 5 | Additional analyses for clonal dynamics along the crypt–villus axis. a, Spatial map of villus zones (V0-V8) on the Y007 PSI section. **b**, Heatmap of the clonal coupling scores between different zones for shuffled clonal data. **c**, Transcriptomic UMAP embedding of enterocytes colored by subtypes (Ent_1–Ent_4), showing a continuum of enterocyte differentiation. **d**, Dot plot showing expression of villus zonation marker genes across enterocyte subtypes. Otherwise, the same as Extended Data Fig. 1e. **e**, Spatial distribution of enterocyte subtypes (Ent_1–Ent_4) across different tissue sections. **f**, Heatmap of intra-clone fractions from Ent_1 to Ent_4 subtypes across different chasing periods (1-3 weeks). Only clones labeling Ent_1 are shown. **g**, Fraction of clones shared between the Ent_1 and more differentiated enterocyte subtypes (Ent_2 to Ent_4) over time.

https://doi.org/10.1038/s41592-026-03151-5



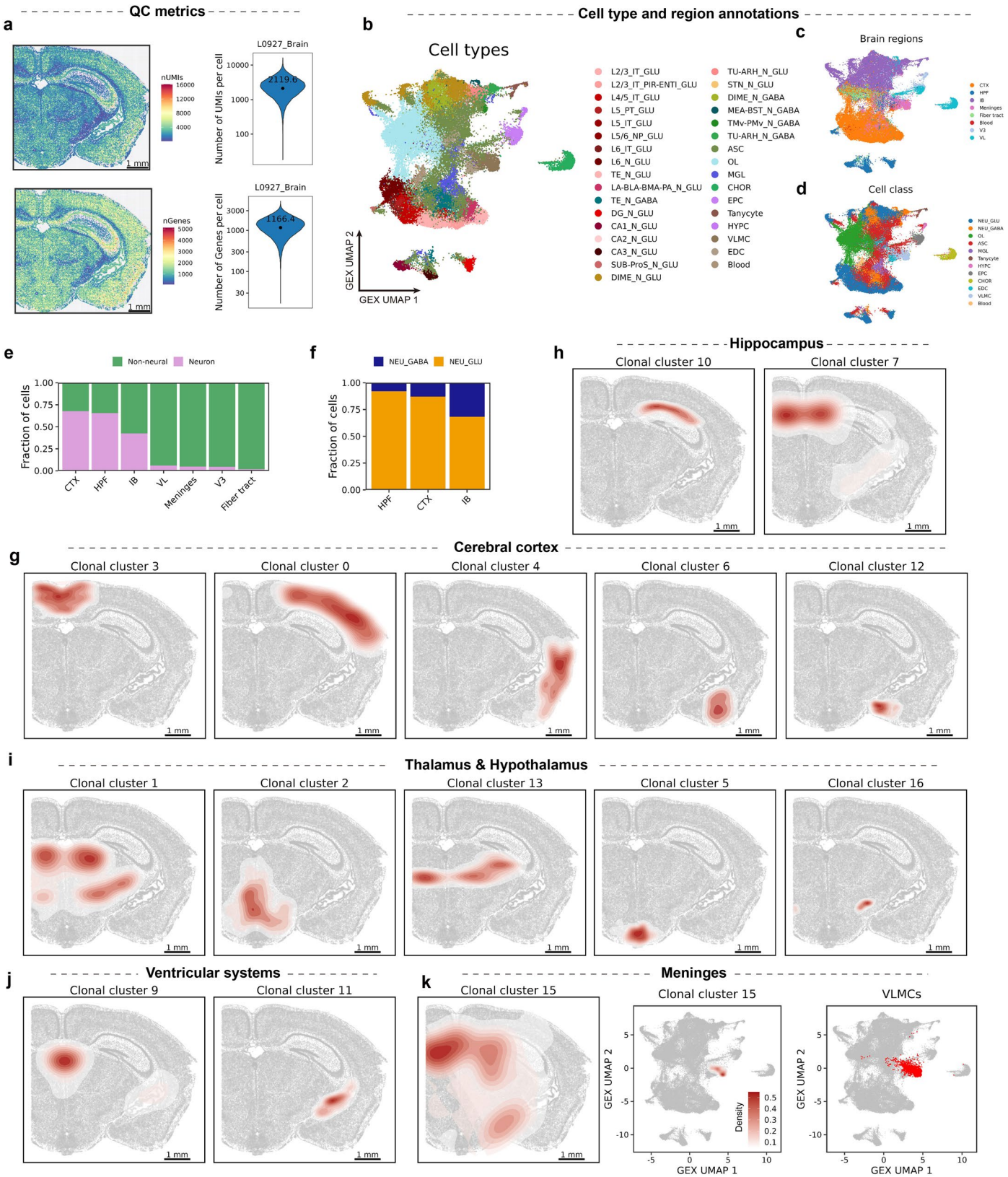
Extended Data Fig. 6 | In situ clonal analyses from the L126 brain dataset. Data recapitulate Fig. 4. **a**, Experimental schematic for tracking E10-labeled clones in biological replicate mouse L126 using three sequential sections. **b**, Quality control metrics showing the distribution of UMIs (left) and genes (right) per cell across sections S1–S3. **c, d**, Spatial maps of anatomical domains (**c**) and major cell types (**d**) across the three sections. **e, f**, Regional composition of neuronal vs. non-neuronal cells (**e**) and excitatory vs. inhibitory neurons (**f**). **g**, Statistics of lineage barcode recovery and QC filtering steps. **h**, Fraction of lineage barcodes shared between adjacent sections (S1–S3) compared to a distinct biological replicate (L0927). **i–k**, Clonal analysis summary: UMAP embedding of 5,631 clones colored by cluster (**i**), heatmap of regional enrichment (**j**), and quantification of clone sizes (**k**) for each clonal cluster. Box plots show the median (50th percentile), the bounds of the box represent the interquartile range (25th to 75th percentile) and whiskers extend to 1.5 times the interquartile range. Illustrations in **a** created in BioRender; Li, L. https://biorender.com/cl0m82y (2026).

https://doi.org/10.1038/s41592-026-03151-5



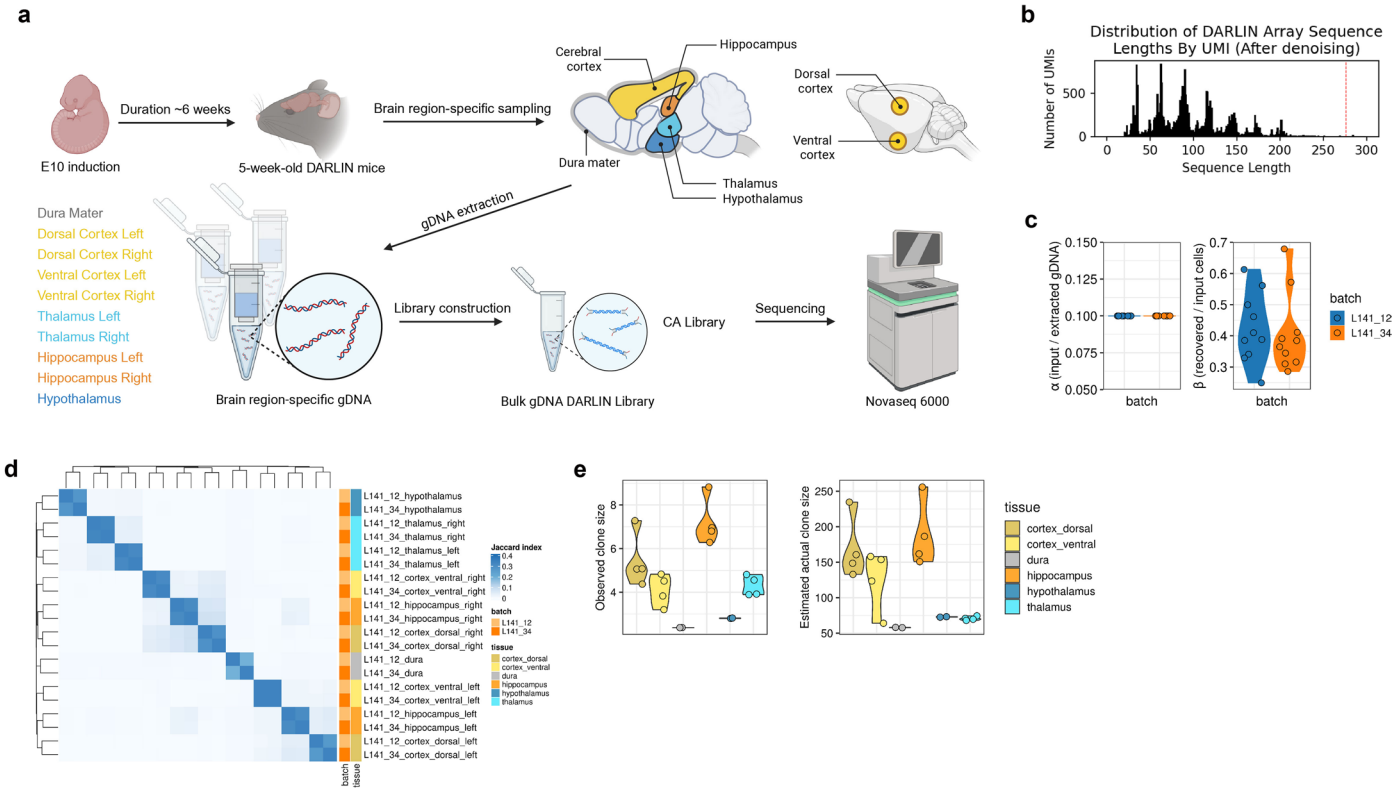
Extended Data Fig. 7 | Additional clonal analyses from the L126 brain dataset. Data recapitulate Figs. 5 and 6. **a**–**c**, Clonal analysis in the cortex: schematic of medial–lateral alignment (**a**), representative clones across three consecutive sections (**b**), and heatmap of clonal distribution along the medial–lateral axis (**c**). **d**, Quantification of inner versus outer dispersion for observed cortical clones (n = 275 clones) compared to shuffled controls (two-sided paired Wilcoxon test, ***, $P = 8.6\times10^{-12}$; ns, not significant, P = 0.20). Only clones with more than 4 cells in both inner and outer groups were included. Box plots show the median (50th percentile), the bounds of the box represent the interquartile range (25th to 75th percentile) and whiskers extend to 1.5 times the interquartile range. **e**–**g**, Clonal analysis in the hippocampus: spatial alignment of CA regions (**e**), clonal distribution heatmap (**f**), and representative clones exhibiting localized or symmetric patterns across sections (**g**). **h**, Identification of hypothalamic spatial domains using BANKSY, visualized via transcriptomic UMAP (left) and spatial maps (right). **i**, Cumulative distribution of pairwise distances for cells within the same clone versus between different clones. **j**, Hierarchical clustering of hypothalamic clones (n = 467, clone size >2) based on pairwise spatial distances. **k**, Representative clones within the arcuate hypothalamic nucleus (ARH) tracked across three sections.

https://doi.org/10.1038/s41592-026-03151-5



Extended Data Fig. 8 | Additional spatial analyses in the brain from mouse L0927. a, Quality control metrics of the L0927 brain dataset. Spatial maps (left) and violin plots (right) of UMI counts (top) and gene counts (bottom) per cell. **b**–**d**, Transcriptome-based UMAP embedding, colored by detailed cell types (**b**), brain regions (**c**), and major cell types (**d**). **e**, Fraction of neuronal and non-neuronal cells across brain regions. **f**, Fraction of glutamatergic (GLU) and GABAergic (GABA) neurons across brain regions. **g**–**j**, Spatial distribution of clonal clusters in the cerebral cortex (**g**), hippocampal regions (**h**), interbrain (**i**), and ventricular system (**j**). Darker red implies a higher cell density for a given clonal cluster. **k**, Clonal clusters associated with the VLMC in meninges (cluster 15). VLMC, vascular and leptomeningeal cell.

https://doi.org/10.1038/s41592-026-03151-5



Extended Data Fig. 9 | Validation of region-specific expansion of E10-labeled RGC clones. a, Schematic overview of the experimental design. Brain regions were sampled bilaterally (left and right hemispheres) from a single mouse (L141). For each anatomical location, two technical replicate libraries were prepared (designated L141_12 and L141_34). Sample IDs follow the format Replicate_Region_Side; for example, L141_12_cortex_dorsal_left denotes the L141_12 technical replicate from the left dorsal cortex. **b**, Length distribution of lineage barcode from CA amplicon associated with the sample L141_12_cortex_dorsal_left. The red dashed line indicates the expected length of the unedited lineage barcode. **c**, Estimation of subsampling rates during gDNA extraction (α) and sequencing (β). **d**, Reproducibility assessment between technical replicates based on the Jaccard index of detected lineage barcodes. **e**, Observed (left) and estimated actual clone sizes (right; observed / ($\alpha\beta$)) across brain regions. Illustrations in **a** created in BioRender; Li, L. https://biorender.com/cl0m82y (2026).

https://doi.org/10.1038/s41592-026-03151-5

Benchmark of the spatial lineage tracing datasets

Methodology	Tissue	ST platform	Spatial resolution	Coarse-grained unit	Clone recovery (Q)	Cost per mm^2 area
iTracer	Cerebral organoid	Visium V1	55 μm	55 μm spots	41.7%	$71.00
Space-TREX	Brain	Visium V1	55 μm	55 μm spots	4.6%	$71.00
KP-Tracer	LUAD	Slide-seq	10 μm	10 μm spots	15–35%[1]	
KP-Tracer	LUAD	Slide-tag	~10 μm	Nuclei	NA	
PEtracer	4T1	MERFISH	0.1–0.3 μm	Segmented cells	36–58%[2]	
SpaceBar	WM989 melanoma	SeqFISH	0.1–0.3 μm	Segmented cells	34%	
Spatio-DARLIN	Brain	BMKMANU	2.5 μm	48 μm spots	75.2–93.9%	$62.00
Spatio-DARLIN	Brain	BMKMANU	2.5 μm	Segmented cells	32.1–49.5%	$62.00

[1] According to Fig.1d (https://doi.org/10.1101/2024.10.21.619529)

[2] According to Fig. S13D (https://doi.org/10.1126/science.adx3800)

Extended Data Fig. 10 | A detailed comparison across different spatial lineage tracing platforms. Comparison of spatial lineage-tracing platforms. Summary table comparing Spatio-DARLIN with recently reported spatial lineage-tracing methods, br /including iTracer, Space-TREX, KP-Tracer and PEtracer. Values were compiled from published datasets and should be interpreted as a literature-based comparison rather than abr /side-by-side experimental benchmark.

nature portfolio

Corresponding author(s): Shou-Wen Wang, Li Li

Last updated by author(s): Apr 14, 2026

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Raw sequencing data were processed using BSTMatrix_v2.4.f.1 to generate gene expression matrices and spot coordinate information. No additional software was used for data collection.
Data analysis	https://github.com/LLi-Lab/spatio_DARLIN

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The accession number for the spatial transcriptomic datasets is NCBI GEO: GSE296896, and for the spatial DARLIN datasets is GSE296897. A table containing analysis parameters for each dataset in this study is available at Supplementary Table 2.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This study did not involve human participants, their data, or biological material.
Reporting on race, ethnicity, or other socially relevant groupings	This study did not involve human participants, their data, or biological material.
Population characteristics	This study did not involve human participants, their data, or biological material.
Recruitment	This study did not involve human participants, their data, or biological material.
Ethics oversight	This study did not involve human participants, their data, or biological material; therefore, human ethics approval was not required.

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

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. Sample sizes are similar to those generally employed in the field and were considered sufficient for robust analysis.
Data exclusions	Cells < 100 expressed genes were excluded based on the general employed quality control criteria.
Replication	All key experiments were successfully replicated with independent biological samples.
Randomization	No randomization was used, as all samples were processed in parallel under identical experimental conditions.
Blinding	Investigators were not blinded to group allocation during experiments or data analysis, as blinding was not relevant to the study design.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Cas9-TdT-gRNAs-M2 mice (The Jackson Laboratory, 038749) and CA/TA/ RA triple target-array mice (The Jackson Laboratory, 038750) were used in this study.
Wild animals	This study did not involve wild animals.
Reporting on sex	Both male and female embryos were used. No sex-based differences were specifically analyzed.
Field-collected samples	This study did not involve field-collected samples.
Ethics oversight	All animal procedures were approved by the Institutional Animal Care and Use Committee of Westlake University (permission number AP#24-072-LL).

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

Plants

Seed stocks	This study did not involve plant seed stocks.
Novel plant genotypes	This study did not involve novel plant genotypes.
Authentication	This study did not involve plants; authentication is not applicable.