High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell

Song Chen, Blue B. Lake and Kun Zhang

Single-cell RNA sequencing can reveal the transcriptional state of cells, yet provides little insight into the upstream regulatory landscape associated with open or accessible chromatin regions. Joint profiling of accessible chromatin and RNA within the same cells would permit direct matching of transcriptional regulation to its outputs. Here, we describe droplet-based single-nucleus chromatin accessibility and mRNA expression sequencing (SNARE-seq), a method that can link a cell’s transcriptome with its accessible chromatin for sequencing at scale. Specifically, accessible sites are captured by Tn5 transposase in permeabilized nuclei to permit, within many droplets in parallel, DNA barcode tagging together with the mRNA molecules from the same cells. To demonstrate the utility of SNARE-seq, we generated joint profiles of 5,081 and 10,309 cells from neonatal and adult mouse cerebral cortices, respectively. We reconstructed the transcriptome and epigenetic landscapes of major and rare cell types, uncovered lineage-specific accessible sites, especially for low-abundance cells, and connected the dynamics of promoter accessibility with transcription level during neurogenesis.

RNA-seq of single cells or nuclei reveals their transcription state, whereas chromatin accessibility sequencing (ATAC-seq) uncovers the associated regulatory landscape. Current strategies 1,2, which involve profiling these modalities separately followed by computational integration, may not fully recapitulate the true biological state. Joint profiling of two layers of omics information within the same cells would enable direct matching of transcriptional regulation to its output, allowing for more accurate reconstruction of the molecular processes underlying a cell’s physiology.

To enable highly parallel profiling of chromatin accessibility and mRNA from individual nuclei, we developed SNARE-seq, implemented on a micro-droplet platform3. In this method, the accessible chromatin in permeabilized nuclei is captured by Tn5 transposase, before droplet generation. We reason that, without heating or detergent treatment, binding of transposase to its DNA substrate after transposition could maintain the contiguity of the original DNA strands4, allowing for the co-packaging of accessible genomic sites and mRNA from individual nuclei in the same droplets. As such, we designed a splint oligonucleotide with sequence complementary to the adaptor sequence inserted by transposition (5’ end) and the poly(A) bases (3’ end) allowing capture by oligo(dT)-bearing barcoded beads. After encapsulation of nuclei, their mRNAs and fragmented chromatin can be released by heating the droplets, allowing access to splint oligonucleotides and adaptor-coated beads with a shared cellular barcode (Fig. 1a). A pair of RNA-seq and chromatin accessibility libraries can then be generated for sequencing (see Methods). The resulting data can be connected by their shared cellular barcodes, without the need for probabilistic mapping of single-cell clusters from separate analyses. While SNARE-seq shows similarities to sci-CAR2 conceptually, our method was implemented on a widely accessible Drop-seq platform and provides denser chromatin information owing to a design that captures chromatin information first and then links it to the transcriptome.

To evaluate the ability of SNARE-seq to capture accessible chromatin, we first performed a proof-of-concept experiment on lymphoblastoid cell line GM12878, which has extensively characterized chromatin landscapes. Ensemble profiles of SNARE-seq accessibility data showed a signal-to-noise ratio similar to that in ATAC-seq and Omni-ATAC1 (Fig. 1b). The aggregated SNARE-seq data also showed the expected periodical nucleosome pattern and a strong enrichment of fragments within canonical promoter regions (Supplementary Fig. 1a,c), which are typical characteristics of bulk ATAC-seq data. We validated the peaks from the SNARE-seq data by overlapping them with those of published bulk ATAC-seq and Omni-ATAC data (Supplementary Fig. 1b) and found that 85.9% of ATAC-seq peaks were shared among all three assays and that 87.6% of Omni-ATAC peaks were shared by Omni-ATAC and SNARE-seq. After filtering out low-quality data, we obtained a median of 2,720 accessible sites per nuclei, which is comparable to the number for several published single-cell or single-nucleus ATAC-seq methods and roughly four- to fivefold less dense than for the 10x Genomics sc-ATAC-seq method (Supplementary Figs. 1d and 2a,c).

To assess the accuracy of SNARE-seq in identifying cell types, we performed SNARE-seq on mixtures of cultured human BJ, H1, K562 and GM12878 cells and collected 1,047 profiled pairs (median of 500 unique molecular identifiers (UMIs); median of 805 accessible sites; Supplementary Fig. 2a,d). Separate clustering of expression and chromatin accessibility data showed clear separation into four distinct clusters (Fig. 1c). Differential expression of marker genes (Supplementary Fig. 3a) validated these cluster identities. The classification results from the two profiles were in good agreement (kappa coefficient of 0.92; Fig. 1d). Notably, we found that transcription factors JUN, IRF8, POU5F1 and GATA1, which showed enriched expression in BJ, GM12878, H1 and K562 cells, respectively (Supplementary Fig. 3c), also exhibited a similar pattern of preferential binding to chromatin sequences captured by SNARE-seq accessibility assay (Supplementary Fig. 3b,d), consistent with previous observations5. We improved the detection sensitivity on chromatin further by using NP-40-based Nuclei EZ buffer to boost tagmentation efficiency and by adding an RNase inhibitor combination6 to protect RNA from degradation. From the mixed cell lines we acquired 1,043 paired profiles with a median number of 1,159 UMIs and 2,254 accessible sites captured (Supplementary Figs. 2a,d and 4c). We then compared SNARE-seq expression and chromatin data to those generated from snDrop-seq or SNARE-seq chromatin-only experiments and observed consistent clustering (Supplementary Fig. 4a,b) and a high level of correlation of raw reads (Supplementary Fig. 4d), as well as an efficient recovery rate
Furthermore, a species-mixing experiment indicated high purity and a low doublet rate (6%; Supplementary Fig. 4f) of SNARE-seq data. Therefore, in a simple cell mixture, SNARE-seq can effectively separate cell types on the basis of both their chromatin signatures and transcriptomes, with a high level of concordance.

We next applied SNARE-seq to mouse neonatal cerebral cortex (postnatal day (P) 0, n = 5) and recovered 5,081 nuclei that had linked transcriptome (median of 357 UMIs) and chromatin accessibility (median of 2,583 accessibility sites) data after quality filtering (Supplementary Fig. 2a,d). Correlation analyses of the expression profiles and the chromatin profiles demonstrated high reproducibility between independent SNARE-seq experiments (Supplementary Fig. 5a,b). Among all RNA reads, 94% aligned to the genome, with 37% of these mapped to exons and 42% mapped to introns (Supplementary Fig. 5c), reflecting the enrichment of nascent transcripts in the nucleus. In comparison, despite a similar mapping rate (>91%), the chromatin accessibility data showed a larger fraction of reads (34%) mapped to intergenic regions. There was also enrichment of accessibility reads in close proximity to the transcription start site (10%) and low coverage in exons, suggestive of enhancer and promoter sequences present in those non-coding regions. Therefore, both RNA and chromatin reads showed the expected genome distributions, comparable to those seen in snDrop-seq and snATAC data.

Unsupervised clustering of cerebral cortex transcriptomes identified 19 cell clusters, including astrocytes/radial glia (Ast/RG); intermediate progenitor cells (IP); excitatory neurons (Ex); migrating inhibitory neurons (In); and Cajal–Retzius cells. We further detected several non-neuronal cell types, including oligodendrocyte progenitor cells (OPCs); endothelial cells (End); pericytes (Peri); and microglia (Mic). These cell clusters ranged in size from 37 (0.7%) to 542 (10.7%) cells (Supplementary Fig. 6a) and were independent of batch or sequencing depth (Supplementary Fig. 6b–e). Uniform manifold approximation and projection (UMAP) revealed a trajectory extending from the progenitor states, reflective of the developmental progression of the cerebral cortex.
sequential development of cortical cell fates. Consistent with this trajectory, nuclei occurring adjacent to intermediate progenitors represented those of the late-born neurons of the superficial layers (proceeding deep-layer neurons) and glial cell types associated with the onset of gliogenesis that is expected at this time point (Fig. 2a). We compared SNARE-seq transcriptome data to a recently published single-cell RNA-seq dataset of the mouse cortex at a similar developmental time point that was generated by SPLIT-seq5. Despite a lower number of detected UMIs in SNARE-seq, the cell types and their signatures were reasonably well correlated (Supplementary Fig. 7a–c). Notably, we captured finer distinctions between closely related cellular states and identified three subclusters of intermediate progenitor cells: cluster IP–H1mg2, expressing Mki67, Top2a and Klf23 (Fig. 2b, Supplementary Fig. 7d and Supplementary Table 1), representing cycling progenitors; cluster IP–Gadd45g that was enriched for Gadd45g expression, representing apical progenitors that exited from the cell cycle1; and cluster IP–Eomes, representing basal progenitors that show early commitment to the neuronal lineage. The cell type and layer identities of our clusters were further validated by expected expression of known marker genes and in situ staining of markers discovered here (Fig. 2b, Supplementary Fig. 8 and Supplementary Table 1).

We compared aggregated SNARE-seq chromatin accessibility profiles with published bulk ATAC-seq ENCODE data on neonatal mouse brain cortex and found a strong concordance between these two methods (Supplementary Fig. 7c,f). To cluster co-assayed cells on the basis of their chromatin accessibility profiles, we used the corresponding transcriptional profiles to aggregate chromatin accessibility signals for each cluster separately, followed by peak calling and the probabilistic topic modeling method implemented in cisTopic17. After projecting onto lower dimensions using UMAP, most single-nucleus chromatin accessibility clusters (Fig. 2c) corresponded to the same cell types resolved from the corresponding expression data (Fig. 2a). Notably, the chromatin accessibility of deep-layer excitatory neurons and migrating inhibitory neurons, which differentiated earlier in the cerebral cortex and ganglionic eminences, respectively, showed well-separated clusters, whereas those of late-generated superficial-layer excitatory neurons were less distinct. The diffuse boundaries identified by expression profiles were also clustered as subtypes on the basis of chromatin information (Supplementary Fig. 9a,b). These subtypes may represent datasets with insufficient clustering power owing to the sparsity of chromatin data and/or dynamic epigenetic states that are still undergoing maturation. Cell-type identities of the major clusters were further supported by specific accessibility in the promoter region for marker gene loci Hes5 (Ast/RG), Gadd45g (IP), Meg3 (neuron), Pdgfra (OPC), Vtn (Peri) and Apbbi1ip (Mic) (Supplementary Fig. 9d). We found that the promoter accessibility of lineage markers Vtn and Cd45 (for pericyte and microglia, representing 1% and 0.7% of total cells, respectively) was present only in cell-type-aggregated chromatin profiles that were identified de novo with transcriptome data. In contrast, chromatin data analyzed on the basis of the accessible peaks called from batch-aggregated profiles, the current default strategy for analyzing sc-ATAC-seq data, failed to recover these accessible peaks unique to rare cells in the presence of background noise from other more abundant cells (Fig. 2e).

Fig. 2 | Dual-omics profiling of neonatal mouse cerebral cortex with SNARE-seq (n = 5 replicates). a, UMAP projection of 5,081 SNARE-seq expression data of mouse cortical nucleus. Cells were assigned on the basis of known markers. b, Heat map showing the normalized expression of cell-type-specific genes relative to the maximum expression level across all cell types. CR, Cajal–Retzius cells. c, UMAP projection of SNARE-seq chromatin accessibility data of mouse cerebral cortex nuclei. Cells are labeled with the same color codes for cell types identified by the linked expression data. d, Heat map showing the normalized chromatin accessibility of cell-type-specific accessible sites, relative to the maximum accessibility across all cell types. e, Chromatin accessibility tracks generated from cell-type-specific or batch-aggregated (batch codes 12 A, 12 B and 12 C) chromatin accessibility data at pericyte (left) and microglia (right) marker gene loci (Vtn and Cd45, respectively). For better visualization, the promoter regions are shaded in gray. f, Pseudotime trajectories constructed with SNARE-seq expression (top) and chromatin accessibility (bottom) profiles for 1,469 nuclei (214 IP-H1mg2, 99 IP–Gadd45g, 437 IP–Eomes, 177 Ex-L2/3–Ctn2n and 542 Ex-L2/3–Cux1) from the mouse cerebral cortex. Cells are colored according to pseudotime score (left) or cellular identity (right). g, Promoter accessibility (yellow) and gene expression (red) changes of Sox6, Gpm6b, Nrxi1 and Kdhrbs2 across pseudotime during early neurogenesis. Misc, cells of miscellaneous clusters.
genes were also differentially expressed by pseudotime. Most of these genes presented similar directional changes in promoter accessibility and expression level (Supplementary Fig. 12b,c). For example, Sox6, encoding a transcription factor required for maintenance of neural precursor cells\textsuperscript{18}, and membrane-protein-encoding Mlc1 showed a decline along neuronal differentiation, whereas Khdrbs2 (SLM1), encoding an RNA-binding protein participating in alternative splicing, and its regulating target Nrxn1 (ref. \textsuperscript{19}) showed a similar directional rise along neurogenesis (Fig. 2g and Supplementary Fig. 12c,d). Thus, SNARE-seq provided linked expression and chromatin accessibility profiles that enabled construction of regulatory dynamics during developmental programs,
as well as detailed characterization of epigenetic state for the cell clusters (Supplementary Fig. 13).

We finally applied SNARE-seq to adult mouse cerebral cortex and obtained 10,309 paired profiles (median of 1,332 RNA UMIs and median of 2,000 chromatin accessibility per nucleus) after quality-control filtering (Supplementary Fig. 2a,d). Unsupervised clustering of the 10,309 transcriptomes revealed 22 cell clusters, including 10 excitatory neuron types, 4 inhibitory neuron types (expressing Pvalb, Sst, Npy and Vip) and OPCs, newly formed Itpr2-expressing oligodendrocytes (Oli–Itpr2) and mature oligodendrocytes (Oli–Mal), as well as other non-neuronal cells (Fig. 3a). Most of the clusters could be identified with existing lineage or cortical layer markers. These marker genes were expressed in a similar but more specific pattern (Supplementary Figs. 8 and 14a–d, and Supplementary Table 1) in the cell clusters. To investigate the epigenetic patterns of each cell cluster, we aggregated SNARE-seq chromatin data of adult cerebral cortex, which showed high similarity to bulk ATAC-seq data (Supplementary Fig. 14e), on the basis of cell-type identities defined by de novo clustering of linked transcriptome data. We then performed peak calling and clustering using the topic modeling method\(^1\). The cell clusters (Fig. 3b) were more cleanly and distinctly separated, compared to the chromatin profile of the neonatal cortex, probably owing to the more discrete cell states of the adult brain. We next performed gene ontology and motif enrichment analysis on the differential accessible peaks identified across all cell clusters (Supplementary Table 3). Although some clusters, such as astrocytes (Ast) and microglia (Mic), showed similar enrichment of biological processes and transcription factor binding in comparison to corresponding cells in the developing mouse cortex (Supplementary Fig. 15), most other clusters exhibited different features, which might reflect the postnatal maturation within the brain.

Overall, SNARE-seq is a robust method allowing the joint measurement of the transcriptome and chromatin accessibility in single cells or nuclei. Owing to a simple design that does not rely on proprietary reagents, SNARE-seq can be widely implemented. Compared to the recently reported sci-CAR\(^1\), SNARE-seq detects RNA molecules at a sensitivity comparable to that of other single-nucleus RNA-seq methods (Supplementary Figs. 2d,e and 14f) and captures 4–5 times more accessible sites (Supplementary Fig. 2a,b), which improved the discovery of differentially accessible sites by approximately twofold and provided a better separation of cell clusters (Supplementary Fig. 13). Finally, the throughput of this assay can be further improved through integration with a cellular combinatorial indexing strategy\(^1\). SNARE-seq represents a valuable tool for characterizing tissue complexity in both the inputs and outcomes of transcriptional regulation units and would be especially useful for creating cell atlases of human tissues and clinical samples.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41587-019-0290-0.

Received: 21 February 2019; Accepted: 12 September 2019; Published online: 14 October 2019.

References
1. Lake, B. B. et al. Nat. Biotechnol. 36, 70–80 (2018).
2. Duren, Z. et al. Proc. Natl Acad. Sci. USA 115, 7723–7728 (2018).
3. Macosko, E. Z. et al. Cell 161, 1202–1214 (2015).
4. Amini, S. et al. Nat. Genet. 46, 1343–1349 (2014).
5. Cao, J. et al. Science 361, 1380–1385 (2018).
6. Buenrostro, J. D. et al. Nature 556, 3970–3986 (2017).
7. Bravo González-Blas, C. et al. Nat. Methods 16, 397–400 (2019).
8. Subramanian, L. et al. Proc. Natl Acad. Sci. USA 108, E265–E274 (2011).
9. Zhang, S. et al. Mol. Neurobiol. 55, 9001–9015 (2018).
10. Harrison-Uy, S. J. Cold Spring Harb. Perspect. Biol. 4, a00894 (2012).
11. Artegiani, B. et al. Cell Rep. 21, 3271–3284 (2017).
12. La Manno, G. et al. Nature 560, 494–498 (2018).
13. Lee, K. E. et al. Proc. Natl Acad. Sci. USA 111, 2794–2799 (2014).
14. Hiyama, T. et al. Cell 147, 1601–1614 (2011).

Acknowledgements
This project was supported by National Institutes of Health grants U01MH098977, R01HL123755 and U54HL145668 to K.Z.

Author contributions
S.C. and K.Z. conceived the study. S.C. designed and conducted the experiments. S.C. and B.B.L. analyzed the data. S.C. and K.Z. wrote the manuscript.
Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0290-0.

Correspondence and requests for materials should be addressed to K.Z.
Reprints and permissions information is available at www.nature.com/reprints.
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Ethics. The human embryonic stem cell line H1 was purchased from WiCell and the related study was approved by the University of California San Diego Embryonic Stem Cell Research Oversight Committee.

Cell culture. BJ and K562 cells were maintained in DMEM supplemented with 10% FBS. GM12878 cells were maintained in 1640 medium supplemented with 15% FBS. The H1 human embryonic stem cell line was maintained in feeder-free mTeSR medium and passaged with ReLeSR according to the manufacturer’s instruction.

Nuclei preparation. GM12878 nuclei were extracted with ATAC resuspension buffer containing 0.1% NP-40, 0.1% Tween-20 and 0.01% digitonin as described previously. Nuclei from the human cell line mixture were extracted with either nuclear extraction buffer (0.32 M sucrose, 5 mM CaCl2, 3 mM magnesium acetate, 0.1 mM EDTA, 20 mM Tris-Cl [pH 8], and 0.1% Triton X-100) or ice-cold Nuclei EZ Prep buffer (cat. no. NUC101). To extract nuclei from mouse cerebral cortex (C57Bl/6 mouse cortex at P0 and 2 months, purchased from BrainBits (cat. no. C57PCX)), the tissue samples were chopped into small pieces with a razor blade and were homogenized using a glass Dounce tissue grinder (10 times with pastel A and 20 times with pastel B) in 2 ml of ice-cold Nuclei EZ Prep buffer. Nuclei were then passed through a 30-μm filter (Simsx Partec), spun down for 10 min at 900g, and washed and resuspended in PBS supplemented with 1% fatty-acid-free BSA.

Tn5 transposon insertion. Nuclei were counted with an automated cell counter, and approximately 10,000 nuclei were used for transposon insertion. Nuclei pellets were resuspended in a total of 50 μl of reaction mix containing 25 μl of 2× Nextera Targent DNA Buffer, 8 μl of TDE1 Tagment DNA Enzyme and 1 μl of NaGen RNase inhibitor (or 1 μl of Enzymatics RNase-In and 1 μl of Superase-In for the human cell line mix and adult mouse cerebral cortex experiments) and incubated at 37°C for 30 min with shaking at 300 r.p.m. After transposon insertion, nuclei were resuspended and washed with PBS containing 1% BSA and kept at 4°C until droplet generation.

Nuclei barcoding. Droplet generation was performed as described previously, with a few modifications. Briefly, tubing and syringes were coated with 1% BSA to prevent nonspecific binding and then rinsed with PBS before the experiment. Picoll PM-400 was added in nuclei suspension buffer instead of lysis buffer to mitigate nuclei settling. To capture released chromatin fragments with barcoded beads, 1 μl of splint oligonucleotides (Nextera-R1-rc-polyA, 10μM; Supplementary Table 3) was added into Drop-seq lysis buffer. Nuclei suspension at a concentration of 100 nuclei per microliter was co-encapsulated with barcoded beads (ChemGenes, cat. no. Maco6021110) in droplets. When encapsulation was complete, microfluidic emulsion collected in Falcon tubes was overlaid with a layer of mineral oil and then was incubated at 37°C for 30 min with shaking at 300 r.p.m. After transposon insertion, nuclei were resuspended and washed with PBS containing 1% BSA and kept at 4°C until droplet generation.

Sequence data preprocessing. For the human cell line mixture, barcodes with fewer than 200 UMIs or more than 2,000 UMIs (Triton X-100 lysed) or 5,000 UMIs (Nuclei EZ lysed) were omitted, and barcodes with both transcriptome and chromatin accessibility profiles were selected. The expression count matrix was then normalized in the PAGODA2 package (https://github.com/hms-dbmi/pagoda2). A Winsorization procedure was employed to cap the magnitude of the top 0.1% extreme values for each gene. Variance of each gene was modeled as dependency on the exposure magnitude (log scale) as a smoothed generalized additive model with smoothing term k=10 (mgcv package in R). The observed-to-expected variance ratio for each gene was modeled by F distribution using the degrees of freedom corresponding to the number of successful gene observations. To normalize the contribution of each gene in the subsequent principal-component analysis, we rescaled the variance of each gene to match the tail probability obtained from the F distribution under a standard normal sampling process. Cell clusters were determined from an approximate k-nearest-neighbors graph on the basis of the cosine distance of the top 10 principal components derived from the top 1,000 most variable genes from the variance-adjusted expression matrix, using the Infomap community detection algorithm (as implemented in the igraph R package). Cell clusters were visualized by t-SNE. For P0 mouse cerebral cortex experiments, 6,663 barcodes with more than 200 UMIs and fewer than 1,200 UMIs were retained, and 5,488 (82.4%) barcodes were left after a second round of filtration to remove those with fewer than 250 accessible sites and fraction of reads in peak lower than 0.4. The expression count matrices were combined across independent experiments and were batch corrected and normalized in the PAGODA2 package. Expression variance was adjusted as mentioned previously. Then the top 2,000 most variable genes were used to derive the top 50 prinicipal components, and cell clusters were determined from the k-nearest-neighbors graph. Cell clusters with fewer than 25 cells were omitted from further analysis, dissecting out 5,081(76.3%) cells were re-classified and the projection on the top 20 principal components. Genes that were differentially expressed between cell types were identified by Wilcoxon rank-sum test in Seurat (v.2.3.4: https://satijalab.org/seurat/). Cell clusters were annotated manually on the basis of known markers for the cerebral cortex and gene expression pattern from DropViz (http://dropviz.gen). Overall, 10,369 adult mouse cerebral cortex expression datasets were recovered and clustered in a similar way but using different cutoffs (minimum of 200 genes, maximum of 2,500 genes and fraction of reads in peak higher than 0.5).

Comparison of SNARE-seq expression data with SPLIT-seq and DroNc-seq data. The top 20 genes from the statistically significant principal components differentiating cell types, as well as the top 50 differentially expressed genes associated with each cell type, were identified by Seurat, and cluster-averaged expression values were used for correlation analysis between the SNARE-seq P0 and SPLIT-seq P2 mouse cerebral cortex expression datasets and between the SNARE-seq and DroNc-seq adult mouse cerebral cortex datasets.

Cell-cycle phase assignments. Each cell was scored using the CellCycleScoring function in Seurat on the basis of its expression of G2/M and S-phase marker genes. Cells with high G2/M or S scores were assigned as G2/M phase or S phase respectively, while cells expressing neither were assigned as G0/G1 phase.

Clustering of chromatin accessibility data. To cluster chromatin accessibility data from the human cell line mixture, the count matrix was first binarized and peaks with fewer than five counts overall or expressing in more than 10% of cells were removed. The probabilities of a region-topic distribution and topic-cell distribution

NATURE BIOTECHNOLOGY | www.nature.com/naturebiotechnology
were calculated using a latent Dirichlet allocation model with a collapsed Gibbs sampler in cisTopic (v.0.1; https://github.com/aertslab/cisTopic). The number of topics with the highest likelihood were picked, and principal-component analysis was performed for all topics. Clustering was visualized by UMAP projection of principal component analysis scores. For mouse cerebral cortex accessibility datasets, cell clusters identified by expression data were used with raw chromatin reads associated with barcodes from the same cell types aggregated together and cluster-specific peaks called with a bulk ATAC-seq pipeline for each identified cluster. Peak lists were then merged, and the accessibility count matrices were generated by overlapping reads with the merged list. The accessibility count matrices were combined across experiments and clustering was performed in the same way in cisTopic as mentioned previously. Cell clusters were visualized by UMAP projection of the principal-component scores of the top 25 topics.

**Identification of differential accessible sites.** To identify cluster-specific accessible sites, differential accessible probabilities ($P$ values) for each peak in each cluster were calculated using Fisher’s exact test. $P$ values were then converted to $Q$ values by the Benjamini–Hochberg procedure, and peaks with $P$ values lower than 0.05 in each cluster were kept. The cluster-specific peak counts per cell were then aggregated and normalized by cell-specific library size factors computed separately by estimateSizeFactorsForMatrix in Monocle (v.2.10; http://cole-trapnell-lab.github.io/monocle-release/) and were visualized using a heat map.

**Developmental ordering of the early neurogenesis subset.** To order cells according to their developmental trajectory of early neurogenesis on the basis of expression data, we selected 1,498 expression datasets for cells from the mouse cerebral cortex identified as IP–Hmgn2, IP–Gadd45g, IP–Eomes, Ex-L2/3–Cntn2 and Ex-L2/3–Cux1 by the previous PAGODA2 clustering-based approaches. Differentially expressed genes across cell types were identified with the differentialGeneTest function of Monocle, and the 503 most significant genes ($qval < 0.001$) were retained to construct the pseudotime trajectory. Cells were ordered according to their value along the trajectory tree. The gene expression along pseudotime was calculated in the same way, and genes passing the significance test ($qval < 0.05$) and gene expression kinetics were visualized using the plot_genes_in_pseudotime function in Monocle. Chromatin accessibility dynamics along pseudotime were calculated similarly to gene expression. Briefly, peaks within 10kb of each other were merged in Cicero and differential accessible sites across cell types were tested. After ranking accessible sites by significance (as reported by differentialGeneTest), the top 1,300 most significant sites ($qval < 0.1$) were used to construct the pseudotime trajectory. To select the differentially accessible promoters along pseudotime, we first selected the differential accessible sites within 2 kb of a gene’s transcription start site and intersected these with the list of differential expressed genes obtained from the step above. Promoter accessibilities were then visualized with the plot_accessibility_in_pseudotime function in Monocle, and a natural spline was used to fit the promoter accessibilities along pseudotime with the percentage of accessible cells as a covariate.

**Annotation of genomic elements.** The GREAT algorithm (http://great.stanford.edu/public.html/) was used to annotate differential accessible sites using the following settings: 1 kb upstream and 1 kb downstream, up to a 500-kb maximum extension. The HOMER package (v.4.10; http://homer.ucsd.edu/homer/) was used to determine motif enrichment using default settings.

**External data.** Published Omni-ATAC (SRP103230), sc-ATAC-seq (GSE65360), snATAC (GSE100033), SPLiT-seq (GSE110823), sci-ATAC (GSE668103), sci-CAR (GSE117089), DroNc-seq (https://portals.broadinstitute.org/single_cell) and ATAC-seq (ENCODE; https://www.encodeproject.org/experiments/ENCSR310MLB/) data were reprocessed. RNA in situ hybridization images for marker genes were taken from the Allen Institute Brain Atlas.

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

**Data availability**
Raw and processed data are available at the Gene Expression Omnibus database under accession number GSE126074.

**Code availability**
The custom script for processing single-nucleus chromatin accessibility reads is available at https://github.com/chensong611/SNARE_prep.
Reporting Summary

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

Statistics

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

- [ ] n/a
- [ ] Confirmed

- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] The statistical test(s) used AND whether they are one- or two-sided

*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*

- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

*Give P values as exact values whenever suitable.*

- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [ ] Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

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

Software and code

Policy information about availability of computer code

| Data collection | bcl2fastq v2.17, Drop-seq_tools-2.1.0 |

Data analysis

| STAR v2.5, atac_dnaise_pipelines v0.3.0, PAGODA2, Seurat v2.3.4, Monocle 2.10, HOMER V4.1, GREAT v3.0, deepTools v2.0, R v3.5.1, cisTopic v0.1, RSeQC v2.6.4 |

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

Data

Policy information about availability of data

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

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

Raw and processed data is available at Gene Expression Omnibus database under the accession number GSE126074.

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.

- [x] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences
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

Data exclusions

Low quality single nuclei data sets (barcodes with fewer than 200 UMIs or more than 2,000 UMIs (for Triton-X lysed nuclei) or 5,000 UMIs (for Nuclei EZ lysed nuclei) in expression data, and barcodes with fewer than 250 accessible sites and fraction of reads in peaks lower than 0.4 in chromatin accessibility data) and nuclei clusters (cluster size smaller than 25) were excluded from downstream analysis as outlined in the Methods section.

Replication

We performed 5 independent experiments for postnatal day 0 mouse cerebral cortex and 12 independent experiments for adult mouse cerebral cortexand verified reproducibility by correlation of gene expression and chromatin accessibility of aggregate sSNARE-seq datasets.

Randomization

The experiments were not randomized

Blinding

Clustering of single nuclei for each dataset were performed in an unbiased, unsupervised manner.

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 | Methods |
|---------------------------------|---------|
| n/a Involved in the study       | n/a Involved in the study |
| ☒ Antibodies                   | ☒ ChIP-seq |
| ☒ Eukaryotic cell lines        | ☒ Flow cytometry |
| ☒ Palaeontology                | ☒ MRI-based neuroimaging |
| ☒ Animals and other organisms  |         |
| ☒ Human research participants  |         |
| ☒ Clinical data                |         |

Eukaryotic cell lines

Policy information about cell lines

Cell line BI, GM12878 and K562 were bought from ATCC and human embryonic stem cell line H1 was purchased from Wicell.

Authentication

Cell lines were authenticated at the respective repositories (ATCC or Wicell) prior to shipping.

Mycoplasma contamination

Cell lines were tested negative for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

No commonly misidentified cell lines were used.

Animals and other organisms

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

Laboratory animals

CS7BL/6 male mouse cortex at postnatal day 0 and 2 months were purchased from BrainBits (Cat # CS7PCX, CS7ACK).

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

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