Title
Author Correction: Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation

Permalink
https://escholarship.org/uc/item/9qs064xj

Authors
Preissl, S
Fang, R
Huang, H
et al.

Publication Date
2018-03-01

DOI
10.1038/s41593-018-0114-4

Peer reviewed
Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation

Sebastian Preissl, Rongxin Fang, Hui Huang, Yuan Zhao, Ramya Raviram, David U. Gorkin, Yanxiao Zhang, Brandon C. Sos, Veena Afzal, Diane E. Dickel, Samantha Kuan, Axel Visel, Len A. Pennacchio, Kun Zhang and Bing Ren

Analysis of chromatin accessibility can reveal transcriptional regulatory sequences, but heterogeneity of primary tissues poses a significant challenge in mapping the precise chromatin landscape in specific cell types. Here we report single-nucleus ATAC-seq, a combinatorial barcoding-assisted single-cell assay for transposase-accessible chromatin that is optimized for use on flash-frozen primary tissue samples. We apply this technique to the mouse forebrain through eight developmental stages. Through analysis of more than 15,000 nuclei, we identify 20 distinct cell populations corresponding to major neuronal and non-neuronal cell types. We further define cell-type-specific transcriptional regulatory sequences, infer potential master transcriptional regulators and delineate developmental changes in forebrain cellular composition. Our results provide insight into the molecular and cellular dynamics that underlie forebrain development in the mouse and establish technical and analytical frameworks that are broadly applicable to other heterogeneous tissues.

Results

Method optimization and computational analysis framework. We adopted a combinatorial barcoding assisted single cell ATAC-seq strategy and optimized it for frozen tissue samples (see Methods). Compared to previous reports, key modifications were made to maximally preserve nuclei integrity during sample processing and optimize transposase-mediated fragmentation of chromatin in individual nuclei (Supplementary Figs. 1–3). We applied this modified protocol, hereafter referred to as snATAC-seq (single-nucleus ATAC-seq), to forebrain tissue from 8-week-old adult mice (postnatal day (P) 56) and from mouse embryos at seven developmental stages from embryonic day (E) 11.5 to birth (P0; Fig. 1a,b). DNA libraries were sequenced to near-saturation as indicated by a read duplication rate of 36–73% per sample (Supplementary Table 1). The barcode collision rate, which assesses the probability of two nuclei sharing the same barcode combination, was ~16%, slightly higher than expected and reported before (Supplementary Fig. 3c). We filtered out low-quality datasets using three stringent quality-control criteria including read depth (Supplementary Fig. 3d), recovery rate of constitutively accessible promoters in each nucleus (Supplementary Fig. 3e) and signal-over-noise ratio estimated by fraction of reads in peak regions (Supplementary Fig. 3f and Methods). In total, 15,767 high-quality snATAC-seq datasets were obtained. The median read depth per nucleus ranged from 9,275 to 18,397, with the median promoter coverage at 11.6% and the median fraction of reads in peak regions...
at 22% (Supplementary Tables 2 and 3). Our protocol maintains the extraordinary scalability of combinatorial indexing, while featuring a ~6-fold increase in read depth per nucleus compared to previous reports (Supplementary Table 3). The high quality of the single-nucleus chromatin accessibility maps was supported by strong concordance between the aggregate snATAC-seq data and bulk ATAC-seq data (R \geq 0.91; Fig. 1c) and excellent reproducibility between independent snATAC-seq experiments (R > 0.9; Fig. 1c and Supplementary Fig. 4).

The snATAC-seq profiles from each forebrain tissue arise from a mixture of distinct cell types. Enhancer regions are well known to display cell-type-dependent chromatin accessibility, and are more effective at classifying cell types than promoters or transcriptomic elements (Fig. 1d, Methods and Supplementary Table 4). Next, we constructed a binary accessibility matrix of open chromatin regions, using 0 or 1 to indicate the absence or presence of a read at each open chromatin region in each nucleus (Fig. 1d). We then calculated the pairwise similarity between cells using a Jaccard index and applied a nonlinear dimensionality reduction method, t-SNE, to project the Jaccard index matrix to a low-dimension space (Fig. 1d). The final t-SNE plot depicts cell types as distinct clusters in a three-dimensional space (Fig. 1d).

Identification of forebrain cell types from snATAC-seq profiles. We applied this computational framework first to 3,033 high-quality snATAC-seq profiles obtained from the adult forebrain (Fig. 2a and Supplementary Table 2). As a negative control, we included 200 'shuffled' nuclear profiles (Supplementary Fig. 5c,d and Online Methods). This analysis revealed 10 total clusters. As expected, the shuffled nuclei formed a distinct cluster with low intracluster similarity. In addition, one other cluster showing low intracluster similarity likely represented low-quality nuclei or accessibility profiles resulting from barcode collision events (Supplementary Fig. 3c). After eliminating these nuclei, we determined eight distinct cell-type clusters from the adult forebrain tissue (Fig. 2a) and seven fetal development timepoints and at maturity, resulting in a total of 140,103 TSS-distal accessibility profiles into distinct cell types. We developed a computational framework to uncover distinct cell types from the snATAC-seq profiles of mouse forebrain tissue at seven fetal development timepoints and at maturity, resulting in a total of 140,103 TSS-distal

Fig. 1 | Overview of the experimental and computational procedures of snATAC-seq. a, Following nuclei isolation from frozen forebrain tissue biopsies, tagmentation of 4,500 permeabilized nuclei was carried out using barcoded Tn5 in 96-wells plates. After pooling, 25 nuclei were sorted into each well of a 384-well plate, and PCR was carried out to introduce the second set of barcodes. FANS, fluorescence-assisted nuclei sorting. b, Overview of the experimental and computational procedures of snATAC-seq. a, Following nuclei isolation from frozen forebrain tissue biopsies, tagmentation of 4,500 permeabilized nuclei was carried out using barcoded Tn5 in 96-wells plates. After pooling, 25 nuclei were sorted into each well of a 384-well plate, and PCR was carried out to introduce the second set of barcodes. FANS, fluorescence-assisted nuclei sorting. b, Overview of the experimental and computational procedures of snATAC-seq. c, Chromatin accessibility profiles of aggregate snATAC-seq (black tracks) agree with bulk ATAC-seq (gray, top track) and are consistent between independent experiments. R1, replicate 1; R2, replicate 2.

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.
results were highly reproducible for two independent experiments (Supplementary Fig. 5e,f).

To categorize each cluster, we generated aggregate chromatin accessibility maps for each cluster and examined the patterns of chromatin accessibility at known cell type marker genes. We found three clusters with chromatin accessibility at Neurod6 and other excitatory neuron-specific genes (clusters EX1–3; Fig. 2b and Supplementary Fig. 6a); two clusters with accessibility at the gene locus of Gad1 likely representing inhibitory neurons (clusters IN1 and 2; Fig. 2b and Supplementary Fig. 6a); one cluster with...
accessibility at the Apoe locus and other known astroglia markers (cluster AC; Fig. 2b); one cluster with accessibility at the Mog gene locus and other oligodendrocyte marker genes (cluster OC; Fig. 2b); and one microglia cluster with accessibility at genes encoding complement factors including the gene C1qb (cluster MG; Fig. 2b and Supplementary Fig. 7a–c). We also compared the aggregate chromatin accessibility maps for each cluster to previously published maps from sorted excitatory neurons, GABAergic neurons, microglia and NeuN-negative nuclei (which mostly comprise non-neuronal cells including astrocytes and oligodendrocytes; Fig. 2b and Supplementary Fig. 7a–c). Consistent with the accessibility patterns at marker gene loci, we observed that clusters EX1–3 were highly similar to sorted excitatory neurons. To further characterize the distinct excitatory neuron clusters, we compared EX1–3 with published bulk ATAC-seq data from different cortical layers and from dentate gyrus23. Notably, we found that EX1 and EX3 were more similar to upper and lower cortical layers, respectively, whereas EX2 showed properties of dentate gyrus neurons (Supplementary Fig. 8a). Cluster IN1 was highly similar to sorted cortical GABAergic neurons. Unexpectedly, IN2 was more similar to sorted excitatory neurons than cortical GABAergic neurons. IN2 showed properties of dentate gyrus neurons23. Notably, we found that EX1 and EX3 were more similar to upper and lower cortical layers, respectively, whereas EX2 showed properties of dentate gyrus neurons (Supplementary Fig. 8a). Cluster IN1 was highly similar to sorted cortical GABAergic neurons. Unexpectedly, IN2 was more similar to sorted excitatory neurons than cortical GABAergic neurons. Distinctions between the inhibitory neuron clusters (IN1 and IN2) were not clear at this stage, but came into focus later when we analyzed transcription factor (TF) motifs enriched in the accessible chromatin regions (described below). Clusters OC and AC resembled sorted NeuN-negative cells, and cluster MG is similar to sorted microglia (Fig. 2b,c).

According to our snATAC-seq data, the adult mouse forebrain consists of 52% excitatory neurons, 24% inhibitory neurons, 12% oligodendrocytes and 6% astrocytes and microglia, respectively (Fig. 2d). Since cell type proportions vary between different forebrain regions, for example, cortex and hippocampus17, the percentages derived from snATAC-seq represent an average of all forebrain regions (Fig. 2e and Supplementary Fig. 7d,e). The predominance of neuronal nuclei derived from adult forebrain tissue was confirmed by flow cytometry analysis stained against the postmitotic neuron marker NeuN22 (Fig. 2e and Supplementary Figs. 6b and 7b,e).

Delineation of the cis-regulatory landscape of specific cell types in the adult forebrain. The power of snATAC-seq is not simply to delineate cell types but also to reveal the cis-regulatory landscape within each cell type. To this end, we calculated the cell-type-specificity of each putative cis-regulatory element (i.e., chromatin accessibility region) using a Shannon entropy index (Supplementary Fig. 9). As expected, proximal promoter elements were accessible in more cell types, while the distal enhancer elements showed significantly higher cell-type-specificity (median value of 4.2% for proximal elements vs. 0.4% for distal elements; Supplementary Fig. 9a,b,d). We next developed a feature selection method (see Methods) to identify the subset of elements that could best distinguish the eight cell type clusters from each other. This approach identified 4,980 elements showing clear cell-type-dependent accessibility (Fig. 2e). To gain insight into the key transcriptional regulators and pathways active in each cell type, we performed k-means
clustering followed by motif enrichment analysis for these genomic elements (Fig. 2e,f, Supplementary Fig. 9d and Supplementary Table 4). For each cell type, we observed an enrichment of binding motifs corresponding to key TFs (Fig. 2f). For example, the binding motif for ETS-factor PU.1 was enriched in MG elements, motifs for SOX proteins were enriched in OC elements, bHLH motifs were enriched in EX1–3 elements and DLX homeodomain factor motifs were enriched in IN elements (Fig. 2f). Moreover, this analysis revealed an important difference between the inhibitory neuron clusters IN1 and IN2. We found that a binding motif for MEIS factors was enriched in a subset of elements specific to IN2. Previous reports showed that MEIS2 plays a major role in generating medium spiny neurons, the main GABAergic neurons in the striatum. Accordingly, we identified gene loci of Ppp1r1b and Drd1, which encode markers of medium spiny neurons, as highly accessible in IN2 but not IN1 (Supplementary Fig. 10). These data suggest that IN2 may represent medium spiny neurons, while IN1 could represent a distinct class of GABAergic neurons. We also identified motifs that were differentially enriched between EX1, EX2 and EX3. Notably, regions specific for EX1 and 3 were enriched for motifs from the Forkhead family, and EX2 was enriched for motifs recognized by MEF2C (Supplementary Fig. 8c and Supplementary Table 4), which has been shown to play an important role in hippocampus mediated memory. A comparison with data from cell-type-specific differentially methylated regions identified by single cell DNA-methylation analysis of neurons showed that both methods were able to identify inhibitory and excitatory neuron specific elements (Supplementary Fig. 11).

**Fig. 4** | snATAC-seq analysis uncovers cis-regulatory elements and transcriptional regulators of lineage specification in the developing forebrain. a. A heat map shows the results of k-means clustering of 16,364 candidate cis-regulatory elements based on chromatin accessibility in different cell types. b. Gene ontology analysis of each cell type using GREAT. c. Transcription factor motifs enriched in each group. d. Enrichment of enhancers that were functionally validated as part of the VISTA database. e. Representative images of transgenic mouse embryos showing LacZ reporter gene expression under control of the indicated subpallial enhancers. Pictures were downloaded from the VISTA database.

Profiling embryonic forebrain development using snATAC-seq. We next extended our framework by analyzing the snATAC-seq profiles derived from fetal mouse forebrains at seven developmental stages (Fig. 1b), seeking to reveal developmental dynamics of transcriptional regulation at the cellular level. The developmental stages examined cover key events from the onset of neurogenesis to gliogenesis. From 12,733 high-quality snATAC-seq profiles, we identified 12 distinct subpopulations (Fig. 3a) that exhibited changes in abundance through development (Fig. 3a–c). This broad cell-type classification allowed us to profile the dynamic cis-regulatory landscape of forebrain development. Based on chromatin accessibility profiles at gene loci of known marker genes, we assigned these cell populations to radial glia, excitatory neurons, inhibitory neurons, astrocytes and erythromyeloid progenitors (EMP; Fig. 3b). Notably, the EMP cluster was restricted to E11.5, whereas the astrocyte cluster was present after E16.5 and expanded dramatically around birth (Fig. 3b).
Identification of lineage-specific transcriptional regulators during embryonic forebrain development. To identify the transcriptional regulatory sequences in each subpopulation, we identified 16,364 genomic elements that showed cell-population-specific chromatin accessibility and best separated the cell subpopulations (Fig. 4a and Supplementary Table 4). To further characterize these elements, we performed motif enrichment analysis and gene ontology analysis of each cluster using GREAT42. Our analysis showed that genomic elements that were mostly associated with radial glia like cell groups (RG1–4; Fig. 4a) fell into regulatory regions of genes involved in early forebrain developmental processes, including forebrain regionalization (Fig. 4b; K1), central nervous system development (Fig. 4b; K3) or forebrain development (Fig. 4b; K5). These elements were enriched for homeobox motifs corresponding to LHX-transcription factors including LHX2 (Fig. 4c; K1, K3 and K5), which is critical for generating correct neuron numbers by regulating proliferation of neural progenitors43 and for temporally promoting neurogenesis over astroglialgenesis44. Notably, one of these clusters was also enriched for both the proneural bHLH transcription factor ASCL1 (Mash1) and its co-regulator POU3F3 (Brn1; Fig. 4c; K5)45. ASCL1 is required for normal proliferation of neural progenitor cells46 and implicated in a DLX1/2-associated network that promotes GABAergic neurogenesis47. In line with this, associated genomic elements were also accessible in one inhibitory neuron cluster (eIN2; Fig. 4c; K3).

We also identified transcriptional regulators that were specifically associated either with neurogenesis or gliogenesis during forebrain development. For example, the early astrocyte (eAC)-specific elements were located in open chromatin regions near genes involved in glia cell fate commitment, and the top enriched transcription factor motif was NF1-halfsite (Fig. 4a–c; K2). Previous studies show that NF1 transcription factor NF1A alone is capable of specifying glia cells to the astrocyte lineage48. NFIX is another NF1 family member with proneural function49. This motif was enriched together with the bHLH transcription factor NEUROD1, binding sites mainly in open chromatin regions found in the excitatory neuron cell population (Fig. 4c; K4, K12, K13)41. Based on chromatin accessibility profiles at marker gene loci, we had previously assigned two cell clusters to the excitatory neuron lineage (eEX1 and eEX2; Fig. 3b). Compared to cluster eEX2, eEX1 showed increased accessibility at both radial glia associated open chromatin (Figs. 3b and 4a; K4) and chromatin regions associated with CNS neuron differentiation (Fig. 4a; K12). In addition, eEX1 nuclei preceded the emergence of eEX2 nuclei during development (Fig. 3c). These findings indicated that eEX1 might represent a transitional state during excitatory neuron differentiation.

The bHLH transcription factor family consists of several subfamilies that recognize different DNA motifs50. NEUROD1 belongs to a subfamily of transcription factors that bind to a central CAT motif, whereas other transcription factors, such as TCF12, preferentially bind to a CAG motif51. Our snATAC-seq profiles revealed an enrichment of the TCF12-binding motif in regions associated with cortex GABAergic interneuron differentiation, in contrast to the excitatory neuron associated enrichment for NEUROD1 (Fig. 4a–c; K4, K11–K13)36,37,41. Analysis of specific genomic elements of the inhibitory neuron cluster eIN3 showed a notable bias in proximity to genes associated with ‘skeletal muscle and organ development’ (Fig. 4a,b; K8). More detailed analysis revealed that the underlying genes Mef2c/d and Foxp1/2, as well as Drd2/3, encode transcription factors and dopamine receptors, indicating differentiating striatal medium spiny neurons46,49. This finding was consistent with the enrichment for MEIS-homeodomain factors in these regions (Fig. 4c; K8) comparable to the medium spiny neuron cluster in adult forebrain (Fig. 2e,f (K8) and Supplementary Fig. 10). Further, genomic elements specific to the EMP cluster were associated with genes involved in myeloid cell development (Fig. 4a–c; K14) and enriched for motifs of the ubiquitous AP-1 transcription factor complexes that have been described as playing a role in shaping the enhancer landscape of macrophages41.

Finally, we attempted to identify developmental dynamics of elements within each cell cluster (Supplementary Fig. 11). Our analysis revealed between 41 and 2,114 dynamic genomic elements for each cell type (Supplementary Fig. 12c–g). Regions that are more accessible after birth (P0) compared to early timepoints were enriched for the RFX1 motif in GABAergic neurons, including the cluster eIN1 as well as the excitatory neuron cluster eEX2 (Supplementary Fig. 12d,e), indicating a general role of the evolutionarily conserved RFX factors in perinatal adaptation of brain cells. Several family members, including RFX1, are expressed in the brain and have been implicated in regulating cilia, for example, in sensory neurons46.

Functional and anatomical annotation of identified candidate cis-regulatory elements. While assessment of open chromatin plays an important role in predicting regulatory elements in the genome52, it does not provide direct information of functional activity. To address this point, we asked whether cluster-specific transposable-accessible chromatin in the embryonic forebrain overlaps with genomic elements tested in reporter assays to validate enhancer activity in mouse embryonic forebrain in vivo53. First, we focused our analysis on all genomic elements with validated functional activity in the forebrain and on a subset shown to be active only in the subpallium54,55. The subpallium is a brain region that gives rise to GABAergic and cholinergic neurons56. In total, 63.1% (275 of 436) of all forebrain enhancers and 64.8% (59 of 91) of subpallial enhancers were represented in our subset of genomic elements, indicating a high degree of sensitivity. Next, we calculated the relative enrichment of subpallial enhancers over total forebrain enhancers for each cluster. Notably, subpallial enhancers were only enriched in clusters K9–11, which were assigned to the GABAergic neuronal lineage (Fig. 4d,e and Supplementary Fig. 13). Unexpectedly, elements of cluster K5 were active in dorsal and lateral pallial regions, as well as in the lateral ganglionic eminence, indicating conserved roles for these genomic elements in a wide variety of regions in the developing forebrain (Fig. 4a and Supplementary Fig. 13). Integration of genomic elements identified by snATAC-seq in specific cell clusters with transgenic enhancer assays confirmed the high specificity and sensitivity of snATAC-seq in identifying cell populations and their underlying regulatory elements.

Discussion

Tissue heterogeneity has been a significant hurdle in the dissection of gene regulatory programs driving mammalian development. While single-cell-based analysis of chromatin accessibility has been reported, a major challenge lies in the published methods’ requirement for fresh cell populations, whereas most biological samples in tissue banks are either frozen or in formalin fixed paraffin embedded blocks. We report here a general approach (snATAC-seq) and a computational framework that can be used to dissect cellular heterogeneity and delineate cell-type-specific gene regulatory sequences in snap-frozen primary tissues. We applied snATAC-seq to heterogeneous forebrain tissue from adult and embryonic...
mice and resolved specific cell types in these samples. Similarly to other approaches, such as single-cell RNA-seq \(^7,8\) and single-cell DNA-methylation analysis \(^9,10\), snATAC-seq can be used to identify cell types de novo in heterogeneous tissue, facilitating generation of cell atlases in the brain and other tissues. In addition, snATAC-seq catalogs the candidate enhancers for each cell type, enabling the dissection of gene regulatory programs without the need to purify specific cell types. As such, this method is particularly suitable for studying cell populations in complex tissues where cellular surface markers are not available. The current framework allows analysis of major cell types with a relative abundance of at least 5%, as shown for microglia in the adult forebrain. It is expected that increasing the number of cells profiled per experiment will linearly increase the sensitivity of cell-type detection. Indeed, the presented combinatorial barcoding protocol can be scaled up to >5,000 high-quality nuclei per experiment simply by working in 384-well plates rather than 96-well plates. Increasing the number of barcodes during tagging will also help to lower the final barcode collision rate without limiting the throughput\(^1\).

Through integrative analysis of single-nuclei chromatin accessibility profiles, we tracked changes in the relative proportions of these cell types during development, identified putative regulatory elements active within each cell type and used those regulatory elements to reveal key TFs in specific forebrain cell types. Therefore, our results provide a unique view of the cell-type-specific cis-regulatory landscape in the forebrain. We expect that with larger cell numbers in the future it will be possible to uncover previously unknown regulatory elements in rare cell types. Moreover, applying snATAC-seq to human tissue samples and integration with genomic variants may reveal relative contributions of distinct cell types to diseases like schizophrenia or Alzheimer’s. We anticipate that our snATAC-seq approach will be a valuable tool for analysis of other brain regions than 96-well plates. Increasing the number of barcodes during tag-

**Acknowledgements**

We thank B. Li for bioinformatic support. We thank M. He and T. Osothprarop for providing the Tn5 enzyme. We thank D. Gao for sequencing on the MiSeq. This study was funded in part by the National Human Genome Research Institute (U54HG0086997 to B.R.), National Institute Mental Health (1U19MH114831 to B.R., U01MH098977 to K.Z.), NHG (2P50 GM085764 to B.R.), and the Ludwig Institute for Cancer Research (to B.R.). S.P. was supported by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (DFG, PR 1668/1-1). R.R. was supported by a Ruth L. Kirschstein National Research Service Award NIH/NCI T32 CA009523. Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under US Department of Energy Contract DE-AC02-05CH11231, University of California.

**Author contributions**
The study was conceived and designed by B.R., S.P., R.F. and K.Z.; it was overseen by B.R. and K.Z. Experiments were performed by S.P., B.C.S. and H.H.; tissue collection by V.A., D.E.D., A.V., L.A.P. and S.P.; sequencing performed by S.K.; computational strategy developed by R.F. and S.P.; data analysis performed by S.P., R.F., Yu. Z., R.R., Ya. Z. and D.U.G. The manuscript was written by S.P., R.F. and B.R. A.V., L.A.P. and K.Z. provided input and edited the manuscript. All authors discussed results and commented on the manuscript.

**Competing interests**
The authors declare no competing financial interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0079-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to B.R.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Mouse tissues. All animal experiments were approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee or the University of California, San Diego, Institutional Animal Care and Use Committee. Forebrains from embryonic mice (E11.5–E16.5) and early postnatal mice (P0) were dissected from one pregnant female or one litter at a time and combined for breeding. Forbred animals were purchased from Charles River Laboratories (C57BL/6NCrl strain) or Taconic Biosciences (C57BL/6N Tac strain) for E14.5 and P0. Breeding animals for other timepoints were received from Charles River Laboratories (C57BL/6NCrl). Dissected tissues were stored in a dry ice ethanol bath. For the adult time point (P56), forebrains from 8-week-old male C57BL/6 NCrl mice (Charles River Laboratories) were dissected and flash frozen in liquid nitrogen separately. Tissues were pulverized in liquid nitrogen using pestle and mortar. For each timepoint, two replicates were processed (n = 2 per timepoint).

Transposome generation. To generate A/B transposomes, A and B oligos were annealed to common pMEN1 oligos (95°C for 2 min, cooled to 14 °C at a cooling rate of 0.1 °C/s) separately (Supplementary Table 5). Next, barcoded transposomes were mixed in a 1:1 molar ratio with unloaded transposase Tn5, which was annealed to common pMENTs oligos (95 °C for 2 min, cooled to 14 °C at a cooling rate of 0.1 °C/s) separately (Supplementary Table 5). Next, barcoded transposomes were mixed in a 1:1 molar ratio with unloaded transposase Tn5, which was annealed to common pMENTs oligos (95 °C for 2 min, cooled to 14 °C at a cooling rate of 0.1 °C/s) separately (Supplementary Table 5). Next, barcoded transposomes were mixed in a 1:1 molar ratio with unloaded transposase Tn5, which was annealed to common pMENTs oligos (95 °C for 2 min, cooled to 14 °C at a cooling rate of 0.1 °C/s) separately (Supplementary Table 5).

Combinatorial barcoding assisted single-nuclei ATAC-seq. For combinatorial barcoding, we used eight different A transposons and 12 generated at Illumina. Mixture was incubated for 30 min at room temperature (1 °C/min) separately (Supplementary Table 5). Next, barcoded transposons were annealed to common pMENTs oligos (95 °C for 2 min, cooled to 14 °C at a cooling rate of 0.1 °C/s) separately (Supplementary Table 5).

Nuclear isolation. Dissected forebrains were pulverized in liquid nitrogen using pestle and mortar. For each timepoint, two replicates were processed (n = 2 per timepoint).

Nuclear isolation. Dissected forebrains were pulverized in liquid nitrogen using pestle and mortar. For each timepoint, two replicates were processed (n = 2 per timepoint).

Adult cortical tissue. We used adult cortex from embryonic mice (E11.5–E16.5) and early postnatal mice (P0) to harvest NeuN-negative sorting.

Adult cortical tissue. We used adult cortex from embryonic mice (E11.5–E16.5) and early postnatal mice (P0) to harvest NeuN-negative sorting.

ATAC-seq. ATAC-seq was performed on 20,000 sorted nuclei, as described previously, with minor modifications11. After adding IGEPA/C-A630 (Sigma) in a final concentration of 0.1%, nuclei were pelleted for 15 min at 1,000 g. The pellet was resuspended in 19μL 1x DMF buffer (36.3 mM Tris-acetate (pH 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and nuclei were counted using a hemocytometer. Concentration was adjusted to 500 μL/mg nuclei.

ATAC-seq. ATAC-seq was performed on 20,000 sorted nuclei, as described previously, with minor modifications11. After adding IGEPA/C-A630 (Sigma) in a final concentration of 0.1%, nuclei were pelleted for 15 min at 1,000 g. The pellet was resuspended in 19μL 1x DMF buffer (36.3 mM Tris-acetate (pH 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and nuclei were counted using a hemocytometer. Concentration was adjusted to 500 μL/mg nuclei.

ATAC-seq. ATAC-seq was performed on 20,000 sorted nuclei, as described previously, with minor modifications11. After adding IGEPA/C-A630 (Sigma) in a final concentration of 0.1%, nuclei were pelleted for 15 min at 1,000 g. The pellet was resuspended in 19μL 1x DMF buffer (36.3 mM Tris-acetate (pH 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and nuclei were counted using a hemocytometer. Concentration was adjusted to 500 μL/mg nuclei.

ATAC-seq. ATAC-seq was performed on 20,000 sorted nuclei, as described previously, with minor modifications11. After adding IGEPA/C-A630 (Sigma) in a final concentration of 0.1%, nuclei were pelleted for 15 min at 1,000 g. The pellet was resuspended in 19μL 1x DMF buffer (36.3 mM Tris-acetate (pH 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and nuclei were counted using a hemocytometer. Concentration was adjusted to 500 μL/mg nuclei.

ATAC-seq. ATAC-seq was performed on 20,000 sorted nuclei, as described previously, with minor modifications11. After adding IGEPA/C-A630 (Sigma) in a final concentration of 0.1%, nuclei were pelleted for 15 min at 1,000 g. The pellet was resuspended in 19μL 1x DMF buffer (36.3 mM Tris-acetate (pH 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and nuclei were counted using a hemocytometer. Concentration was adjusted to 500 μL/mg nuclei.
Single-nuclei ATAC-seq cluster analysis. Cluster analysis partitions cells into groups such that cells from the same group have higher similarity than cells from different groups. Here, we developed a pipeline to obtain cell clusters (https://github.com/r3fang/snATAC). We first generated a catalog of accessible chromatin regions using bulk ATAC-seq data and created a binary accessible matrix. Chromatin sites were assigned a value of 1 for a given cell if there was a read detected within the peak region. Next, we calculated the pairwise Jaccard index between every two cells on the basis of overlapping open chromatin regions. Next, we applied a nonlinear dimensionality reduction method (t-SNE) to map the high-dimensional structure to a 3-D space. This transforms high-dimensional structures to dense data clouds in a low-dimensional space, allowing partitioning of cells using a density-based clustering method. We then identified the optimal number of clusters using the Dunn index. Finally, we compared our cluster results to those of the shuffled set to further verify that our cluster result was not driven by library complexity or other confounding factors.

Step 1. Determining accessible chromatin sites in single cells. To catalog accessible chromatin sites in individual cells, we first created a reference map of open chromatin sites determined by bulk ATAC-seq. The chromatin accessibility maps from different timepoints (from E11.5 to PS6) were merged into a single reference file using REDTools. For clustering of single cells, we tested clustering performance using accessible promoters (2 kb upstream of TSS) and distal elements, respectively, and found that clusters by distal elements outperformed promoters with lower Kickball-Leibler divergence (Supplementary Fig. 5). Therefore, we decided to focus on distal genomic elements as features to perform clustering. Reads in individual cells overlapping with accessible sites were identified. We generated an accessible matrix of the read counts overlapping each individual accessible sites (columns) in each cell (row).

Step 2. Binary accessible matrix. We next converted the chromatin accessibility matrix to a binary matrix $M_{i,j}$, in which $M_{i,j} = 1$ if any read in cell $i$ is mapped to region $j$.

Step 3. Jaccard index matrix. Jaccard index matrices $I_{i,j}$ were calculated between every two cells in which $I_{i,j}$ measures the commonly shared open chromatin regions between cell $C_i$ and $C_j$ as follows:

$$I_{i,j} = \frac{|M_i \cap M_j|}{|M_i \cup M_j|}$$

Diagonal elements of $I_{i,j}$ are set to 0 as required by t-SNE analysis.

Step 4. Dimensionality reduction using t-SNE. Using Jaccard index matrix $I_{i,j}$ as input, we next applied t-SNE to map the N-dimensional data to a 3-D space. Since t-SNE has a nonconvex objective function, it is possible for different runs to yield different solutions. Thus, we ran t-SNE several times with different initializations and used the result with the lowest Kickball-Leibler divergence and best visualization. In a previous study, sequencing depth was considered as a confounding factor and highly used the result with the lowest Kullback-Leibler divergence and best visualization. This transforms high-dimensional structures to dense data clouds in the 3-D space.

Step 5. Density-based clustering. We applied a density-based clustering method to partition cells into groups in the embedded 3-D space. Since t-SNE has a nonconvex objective function, it is possible for different runs to yield different clusters. We expected that the coherent structure of the open chromatin landscape of cells with high similarity would rely on a continuous and smooth 3-D structure and cells for different groups would locate to distinct parts of the plot. We used t-SNE to transform the high-dimensional structures to dense data clouds in the 3-D space. Finally, we applied a density-based clustering method to identify different cell populations within the embedded 3-D space.

Step 6. Number of clusters. In detail, Habibi method applied the Dunn index to quantify the quality of cluster result as following:

$$DB = \min_{C_i,C_j} \frac{\max_{i,j} \Delta(C_i,C_j)}{\min_{i,j} \Delta(C_i,C_j)}$$

in which $\Delta(C_i,C_j)$ represents the intercluster distance between cluster $C_i$ and $C_j$, and $\Delta(C_i)$ represents the intracluster distance of cluster $C_i$. We used the ‘MaxStep’ distance developed by Habibi et al. to calculate the distance for the Dunn index. Finally, we iterated all possible $(\rho_{i,j}, \delta_j)$ combinations that yield different clusters and calculated their Dunn index. The clustering result with the highest Dunn index was chosen as final cluster (Algorithm 3).

Algorithm 1: Cluster assignment

**Input:** local density ($\rho$) and local distance ($\delta$) for every cell; pairwise Euclidean distance in embedded 3-D space ($D$).

**Output:** Cluster assignment ($C$)

Let $n$ be the total number of cells

Let $C_{\infty}$ be an empty array of length $n$

Let $D_{\infty} = -INF$

for $p_{i,j}$ from 0 to max($\rho$) do

for $\delta_i$ from 0 to max($\delta$) do

choose cells whose $\rho(i)$ and $\delta(i)$ are greater than $p_{i,j}$, $\delta_j$ as Centers

$\Delta = \text{cluster_assignment}(D,C,\text{Centers})$

if Dunn($D,C,\Delta$) > Dunn($D,\infty$)

$D_{\infty} = \text{Dunn}(D,\Delta)$

$C_{\infty} = C$

end

end

Return $C_{\infty}$

* cluster_assignment($D,C,\text{Centers}$) and Dunn($D,C$) were described in ref. 15.

Step 7. Shuffled cells. Due to the limited genome coverage of each single cell, cells may cluster according to their sequencing depth rather than ‘true’ covariation. To verify that our cluster results were not driven by such artifacts, we compared our results to a simulated dataset. For this dataset, binary accessible sites within each cell were randomly shuffled across all accessible sites. In other words, we shuffled the data and removed the biological significance, but maintained the distribution of sequencing depth across cells. Shuffled cells were uniformly distributed as a ‘ball’ in the embedded 3-D space without clear partition of cells. However, we did observe a small portion of cells that tend to form a cluster but did not pass the cutoff ($\rho_{i,j}$ used for the PS6 forebrain dataset).

Identification of cluster-specific features. We next developed a computational method that combines stability selection with LASSO to identify genomic elements (features) that potentially distinguish cells belonging to different clusters. LASSO regression enables sparse feature selections through the use of an L1 penalty. However, LASSO regression often does not result in a robust set of selected features and is sensitive to data perturbation. This is especially true when features are correlated. To overcome these limitations, we adopted a stable lasso to robustly identify features that distinguish every two cell clusters (Algorithm 2). Finally, we combined all identified features that distinguish different cell types to identify genomic elements (features) that potentially distinguish cells belonging to different clusters.

Algorithm 2: Identification of cluster specific elements

**Input:** $X \in \mathbb{R}^{n \times p}$ (binary matrix), $Y \in \{0, 1\}^n$ (cluster label), $\alpha$ (subsampling rate), $\beta$ (perturbation rate), $T$ (iteration)

**Output:** importance score for each feature

for $t=1$ to $T$ do

Randomly perturb the data:

Draw a subset $(X_s, Y_s)$ of $(X, Y)$

Draw a vector $w \sim U([0, 1]^p)$

Reweight the features: $X_s = X_s \times w$

Compute the LASSO path of length $\alpha \cdot n$

Keep the selection matrix $S_i$ where

$S_i(i,j) = 1$, if the $i$th feature selected at $j$th step

$0$, otherwise

end for loop;

Compute the feature importance

$$f = \frac{1}{n \cdot \alpha \cdot t} \sum_{i=1}^{n/2} \sum_{j=1}^{Y} S_i(i,j)$$

Bulk ATAC-seq. Paired-end sequencing reads were aligned to the mm10 reference genome using Bowtie2 in paired-end mode, with following parameters: bowtie2 -p 5 -t -X2000–no-mixed–no-discordant (ref. 53), and PCR duplicates were removed using samtools. Next, mitochondrial reads were removed and the position of alignments adjusted. For visualization the ‘bamCoverage’ utility from deepTools2 was used.

Hierarchical clustering of ATAC-seq profiles in adult forebrain. DeepTools2 was used for correlation analysis and hierarchical clustering of ATAC-seq profiles from cell clusters and sorted cell types in the adult forebrain. First, we computed read coverage for each dataset against the merged list of genomic elements that separate cell clusters in the adult forebrain using the ‘multiBamSummary’ utility. Next we used ‘plotCorrelation’ to generate hierarchical clustering using Spearman correlation coefficient between two clusters.
Accessibility analysis and clustering of genomic elements. To cluster genomic elements based on their accessibility profile, we used promoter distal elements that were capable of distinguishing two cell clusters. For each feature, we extended the summit identified by MACS2® in both directions by 250bp and generated a union set of elements using the ‘mergeBED’ function of BEDTools v2.17.0®. Next, we intersected cluster-specific bam files with the peak list using the ‘coverageBED’ function of BEDTools®. We discarded elements that had fewer than five reads on average. After adding a pseudocount of one, we calculated cluster-specific RPM (reads per million sequenced reads) values for each genomic element. We divided the RPM value for a given cluster by the average value of all clusters (fold over mean) and finally log2-transformed the data. The generated matrix was used for k-means clustering of the elements using Ward’s method. We performed this analysis for all adult clusters, as well as for the excitatory neuron clusters and the 12 developmental cell clusters. A list of elements for each analysis can be found in Supplementary Table 4. To compare clusters of genomic elements in the adult forebrain with previously described single-cell DNA-methylation data®, we calculated the fraction of cell-type-specific differentially methylated regions (DMR) with each cluster using the ‘intersectBED’ function of BEDTools and normalized it by the total number of elements. Since Luo et al.® focused on frontal cortex and specifically purified neurons, we centered the comparison on clusters associated with excitatory and inhibitory neurons.

Motif enrichment analysis. To identify potential regulators of chromatin accessibility, we performed motif enrichment analysis using the AME utility of the MEME suite®. For enrichment of known motifs, one-tailed Fisher's exact test was used to calculate significance. P values were corrected by the Bonferroni method for multiple testing. A P value cutoff of < 10−4 was chosen for known motifs from the JASPAR database (JASPAR_CORE_2016_vertebrates.meme)®. For identification of de novo motifs, the HOMER tool was used with its default settings®.

Annotation of genomic elements. The GREAT algorithm was used to annotate distal genomic elements using the following settings to define the regulatory region of a gene: basal + extension (constitutive, 1 kb upstream and 0.1 kb downstream, up to 500-kb max extension)®. Gene ontology categories ‘molecular function’ and ‘biological processes’ were used.

Analysis of dynamic chromatin accessibility within a cell cluster. First, the ATAC-seq reads were counted in all peaks for each stage, cell type and replicate. For each cell cluster, only stages with more than 250,000 reads overlapping ATAC-seq peaks and more than 50 nuclei were used for dynamic analysis. Peaks with greater than 1 read per million reads (RPM) in at least two samples were kept. We used edget® to assess the significance of difference between adjacent stages for cell clusters with at least 4 of 7 stages passing filtering criteria. P values were corrected using the Bonferroni method. Peaks with a Bonferroni P value < 0.05 were called dynamic peaks. The total numbers of dynamic peaks in each cell type are listed in Supplementary Fig. 11c. For each cell type, the read counts in each peak were normalized to a unit vector (i.e., values were divided by the square root of the sum of the squares of the values). k-means clustering was used for cell clusters with more than 200 dynamic elements (k = 3). Motif enrichment analysis was performed for each peak cluster using HOMER®.

VISTA analysis. Genomic locations of 484 VISTA validated elements® were downloaded from https://enhancer.blb.gov using the search term ‘forebrain’. Genomic locations were converted from mm9 to mm10 using the ‘liftOver’ tool (minimum rematch ratio of 0.95)®. Of these, 91 showed specific activity in the subpallium®. To identify developmental clusters that were enriched for subpallial enhancers, we first calculated the ratio of elements per k-means cluster overlapping with the total forebrain enhancer list and the subpallial subset separately. Finally, we calculated the relative enrichment using the ratio of subpallial over the complete forebrain regions. For anatomical annotation of distinct clusters, we intersected these regions with enhancers that are active in specific areas in the developing mouse forebrain®. After filtering clusters with fewer than five overlapping regions, we performed a binomial test to identify anatomical regions enriched for each cluster. The enrichment score is defined as –log2(1-binomial P value).

External datasets. Published ATAC-seq data of sorted excitatory neurons (GSM1541964, GSM1541965); GABAergic neurons (GSM2333653, GSM233363), microglia (GSM2104286), neurons of the dentate gyrus (GSM2179990, GSM2179991)® and distinct cortical layers (Layer2/3: GSM2333632, GSM2333633; Layer 4: GSM2333644, GSM2333645; Layer 5: GSM2333641, GSM2333642; Layer 6, GSM2333638, GSM2333639)® were reprocessed. In addition, bulk ATAC-seq data for embryonic time points generated by the ENCODE consortium were analyzed for comparison (https://www.encodeproject.org/search/searchTerm=atac+forebrain)

Statistics. No statistical methods were used to predetermined sample sizes, and we have not formally tested the distribution of the data. There was no randomization of the samples, and investigators were not blinded toward the developmental time point investigated. However, clustering of single nuclei based on chromatin accessibility was performed in an unbiased manner. Cell types were assigned afterwards. Low-quality nuclei were excluded from downstream analysis as outlined above.

Distal genomic elements to separate two cell clusters were identified using a stable LASSO approach®. A negative binomial test was used to identify promoters enriched in a specific cell clusters to enable annotation. To identify differentially accessible sites within a given cell type between developmental stages, a negative binomial test was used and the resulting P value was corrected using the Bonferroni method®. Motif enrichment for known transcription factor motifs in different sets of genomic elements was performed using a one-tailed Fisher's exact test in combination with Bonferroni correction for multiple testing®. For significance testing of enrichment of de novo motifs, a hypergeometric test was used without correction for multiple testing®.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Accession codes. Raw and processed data have been deposited to NCBI Gene Expression Omnibus with the accession number GSE100033. Data analysis pipeline can be downloaded at https://github.com/r3fang/snATAC.

Data availability. Raw and processed data to support the findings of this study have been deposited to NCBI Gene Expression Omnibus with the accession number GSE100033.

Code availability. The scripts and pipeline for the analysis can be found at https://github.com/r3fang/snATAC.

References
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size
   Describe how sample size was determined.
   - No statistical methods were used to predetermine sample size

2. Data exclusions
   Describe any data exclusions.
   - Low quality single nuclei data sets (low number of reads, promoter coverage, read fraction in peaks) and nuclei clusters (low cluster density) were excluded from downstream analysis as outlined in the Methods section.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   - For each developmental time point we performed two independent experiments and verified reproducibility by correlation of chromatin accessibility of aggregate single nuclei ATAC-seq data sets.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   - The experiments were not randomized.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   - Clustering of single nuclei for each data set were performed in an unbiased, blinded manner. Cluster names were assigned afterwards. Investigators were not blinded regarding the analyzed developmental time point.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   - A statement indicating how many times each experiment was replicated
   - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   - The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   - Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Bowtie2, SAMTools, BEDTools, DeepTools2, t-SNE, LASSO regression, Picardtools, MACS2, MEME suite, HOMER, GREAT, edgeR, liftOver
The scripts and pipeline for the snATAC analysis can be found at https://github.com/r3fang/snATAC

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

The hyperactive Tn5 transposase was acquired through a collaboration with Illumina. The mutations of this enzyme have been published. In addition, a protocol for production of a Tn5 transposase version with E54K and L372P mutations and wildtype at M56 has been published (Picelli et al. 2014). All other materials are commercially available.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

anti-NeuN antibody (1:5000, MAB377, Lot 2806074, EMD Millipore). The antibody was validated for mouse brain tissue and flow cytometry according to the manufacturer.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. GM12878 from the Coriell Institute for Medical Research

b. Describe the method of cell line authentication used. Cells were not authenticated

c. Report whether the cell lines were tested for mycoplasma contamination. Cells were not tested for mycoplasma

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. None of the cell lines used are listed in the ICLAC database

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Forebrains from embryonic mice (E11.5–E16.5) and early postnatal mice (P0) were dissected from one pregnant female or one litter at a time and combined. For breeding, animals were purchased from Charles River Laboratories (C57BL/6NClrl strain) or Taconic Biosciences (C57BL/6NTac strain) for E14.5 and P0. Breeding animals for other time points were received from Charles River Laboratories (C57BL/6NClrl). For the adult time point (P56), the forebrain from 8-week old male C57BL/6NClrl mice (Charles River Laboratories) were dissected and flash frozen in liquid nitrogen separately.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

The study did not involve human research participants.
Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

- 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- 3. All plots are contour plots with outliers or pseudocolor plots.
- 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.

Forebrains from 8-week old male C57BL/6NCrl mice (Charles River Laboratories) were dissected and flash frozen in liquid nitrogen separately. Tissues were pulverized in liquid nitrogen using pestle and mortar. 10 mg adult mouse forebrain tissue (P56) were resuspend in 500 μl lysis buffer (0.5% BSA, 0.1% Triton-X, cOmplete (Roche), 1 mM DTT in PBS) and incubated for 10 min at 4°C. After spinning down (5 min, 500 x g) sample was resuspended in 500 μl staining buffer (0.5% BSA in PBS). Nuclei suspension was incubated with anti-NeuN antibody (1:5000, MAB377, Lot 2806074) for 30 min at 4°C. After centrifugation nuclei were resuspend in 500 μl staining buffer (0.5% BSA in PBS) containing anti-mouse Alexa488-antibody (1:1000, Life Technologies). After incubation for 30 min at 4°C, nuclei were pelleted (5 min 500 x g) and resuspended in 700 ul sort buffer (1% BSA, 1mM EDTA in PBS). After filtration into a FACS tube 5 ul DRAQ7 (Cell Signalling Technologies) was added and NeuN- nuclei were sorted using a SH800 sorter (Sony) into 5% BSA (Sigma) in PBS.

6. Identify the instrument used for data collection.

Experiments were conducted on a Sony SH800 sorter

7. Describe the software used to collect and analyze the flow cytometry data.

Built-in SH800 software was used for data analysis

8. Describe the abundance of the relevant cell populations within post-sort fractions.

NeuN- nuclei were reanalysed after sorting on a Sony SH800 sorter in analysis mode. The purity was >98 %.

9. Describe the gating strategy used.

First big particles were identified according to FSC/BSC area blot to get rid of small debris. Next duplicates were removed according to signal width in FSC and BSC channels. Finally, nuclei were identified in the gated events according to high DRAQ7 signal that stains DNA.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.