Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues

Marek Bartosovic, Mukund Kabbe and Gonçalo Castelo-Branco

In contrast to single-cell approaches for measuring gene expression and DNA accessibility, single-cell methods for analyzing histone modifications are limited by low sensitivity and throughput. Here, we combine the CUT&Tag technology, developed to measure bulk histone modifications, with droplet-based single-cell library preparation to produce high-quality single-cell data on chromatin modifications. We apply single-cell CUT&Tag (scCUT&Tag) to tens of thousands of cells of the mouse central nervous system and probe histone modifications characteristic of active promoters, enhancers and gene bodies (H3K4me3, H3K27ac and H3K36me3) and inactive regions (H3K27me3). These scCUT&Tag profiles were sufficient to determine cell identity and deconvolute regulatory principles such as promoter bivalency, spreading of H3K4me3 and promoter-enhancer connectivity. We also used scCUT&Tag to investigate the single-cell chromatin occupancy of transcription factor OLG2 and the cohesin complex component RAD21. Our results indicate that analysis of histone modifications and transcription factor occupancy at single-cell resolution provides unique insights into epigenomic landscapes in the central nervous system.

The advent of single-cell sequencing technologies has inaugurated a new era in developmental biology, allowing exploratory analysis of tissue complexity and cell heterogeneity1, in-depth analysis of gene regulatory networks, imputation of developmental trajectories and prediction of future cell states (for example RNA velocity)2. Single-cell technologies for probing the epigenetic landscape, such as single-cell ATAC–seq or single-cell DNA methylation–sequencing, have shed light on the epigenetic heterogeneity of tissues. While chromatin accessibility and DNA methylation can provide genome-wide snapshots of active and repressive states, the study of diverse chromatin modifications might provide further insights into epigenomic and cellular states. For decades, the gold-standard assay in the field of epigenetics has been chromatin immunoprecipitation (ChiP) coupled with deep sequencing (ChiP–seq). However, ChiP–seq experiments suffer from low signal-to-noise ratio, inconsistencies due to immunoprecipitation and requirements for high quantities of sample. Recent methods based on in situ chromatin cleavage or tagmentation with low input requirements, such as CUT&Run3 and CUT&Tag4, have raised the possibility of investigating chromatin modifications at the single-cell level at large scale5.

Modifications of histone tails represent a unique system of regulation of gene expression. Histone acetyltransferases and methyltransferases deposit post-translational modifications at various genomic elements (for example, promoters and enhancers) to regulate gene expression both positively (for example, H3K27ac, H3K4me3 and H3K4me1) and negatively (for example, H3K27me3 and H3K9me3). Bulk studies of these modifications have succeeded in defining the regulatory elements but failed to uncover potential cell heterogeneity within tissue samples. Recent studies have characterized the state of post-translational modifications of histones at a single-cell level in cultured cells and embryos6–12. However, for highly complex adult organs such as the brain, such single-cell characterization has not been achieved.

Here, we developed and applied a single-cell Cut&Tag (scCUT&Tag) protocol by adapting the droplet-based 10x Genomics single-cell ATAC–seq platform to investigate histone modification profiles at the single-cell level in the mouse brain. We focused on the oligodendrocyte lineage (OLG), which we have recently shown to be heterogeneous and to be able to transition to alternative cell states during development and disease13–16. We were able to resolve single cells into discrete populations based exclusively on histone modification data, finding unique cell-type-specific markers and quantitative differences in the levels of histone modifications. We used the data obtained to gain unique insights into promoter mark spreading, bivalency and identification of enhancer–promoter interactions. Finally, we were able to obtain single-cell binding profiles for nonhistone proteins, namely chromatin architecture factor and subunit of cohesin complex RAD21 and the OLG-specific transcription factor (TF) OLG2. We have generated web resources, available at https://ki.se/en/mbb/oligointernode and https://mouse-brain-cutanandtag.cells.ucsf.edu, where these scCUT&Tag datasets can be explored. This study provides a method to study epigenetic regulation in complex tissues in great detail and a deeper understanding of epigenetic heterogeneity in the mouse brain.

Results
CUT&Tag profiling of single cells. To perform scCUT&Tag on thousands of cells, we coupled antibody-directed tagmentation (CUT&Tag)1 in bulk with an existing single-cell ATAC–seq protocol (10x Genomics, Fig. 1a). To reduce clumping of the nuclei during the incubations, we added 1% bovine serum albumin (BSA) to specific buffers (Methods) and performed trial bulk CUT&Tag experiments (Extended Data Fig. 1a). We found that addition of BSA reduced clumping of nuclei (Extended Data Fig. 1b) but did not substantially alter the efficiency of tagmentation or the signal distribution for bulk CUT&Tag (Extended Data Fig. 1c).

To verify whether scCUT&Tag data can be used to deconvolute heterogeneous cell populations, we prepared a mixture of three cell lines: mouse embryonic stem cells (mESC, C57Bl/6J origin), mouse embryonic fibroblasts (NIH-3T3) (ATCC) and a mouse oligoden-
Articles NATURE BIOTECHNOLOGY

OLG differentiation, we sorted GFP + cells on day 15 (P15 timepoint). Additionally, to gain more insight into OLGs in mouse (RCE)13,20,21, which labels primarily OLGs in the mouse central nervous system (CNS) cell populations in the scCUT&Tag assay for the different histone modifications in the mouse brain. Single-cell profiles of several histone modifications (H3K4me3, H3K27ac and H3K27me3) or it coincided with H3K4me3. We identified single cells based on the number of reads per barcode and fraction of reads falling into peak regions, called from merged bulk data (Extended Data Fig. 3b). Altogether, we obtained scCUT&Tag profiles of various histone modifications for 47,340 single cells, with the median ranging between 98 (H3K36me3) and 453 (H3K27ac) unique fragments per cell (Fig. 1c and Supplementary Table 4). Between 39.4% and 85.6% of fragments fell within narrow peak regions (Fig. 1d and Extended Data Fig. 3c), indicating a low level of background. The fragment length distribution was consistent with the capture of subnucleosomal fragments, as well as mono-, di- and tri-nucleosomes for all modifications (Fig. 1e).

We compared performance metrics of H3K27me3 scCUT&Tag with those of previously published technologies targeting the same epitope. Our scCUT&Tag data showed similar or higher specificity (as measured by fraction of fragments in peaks) and equal or lower number of unique fragments per cell but yielded more cells per experiment compared with the iCell8 scCUT&Tag and the latest scChIP–seq method12 (Extended Data Fig. 3d–f). Consistent with this, fingerprint plots showed that scCUT&Tag displayed higher specificity and better signal-to-noise ratio compared with scChIP–seq12 and a level of specificity similar to that of iCell8 scCUT&Tag (Extended Data Fig. 3g).

scCUT&Tag of individual histone modifications allows identification of specific cell populations from the mouse brain. To deconvolute and cluster the cells, we generated cell-feature matrices for all datasets using 5 kb (H3K4me3, H3K27ac and H3K27me3) or 50 kb genomic bins (H3K36me3) (Fig. 1b) and identified all major CNS cell populations in the scCUT&Tag assay for the different cell types.

Fig. 1 | Single-cell profiling of several histone modifications in the mouse brain. a, Schematic of the scCUT&Tag experimental design. Cells were isolated from mouse brain at age P15 or P25 and sorted into GFP+ and GFP− populations; nuclei were isolated and incubated with specific antibodies against chromatin modifications or TFs, tagged using protein A–Tn5 fusion and processed by the 10x Chromium scATAC–seq protocol. SSC-A, side scatter area; FITC-A, fluorescein-5-isothiocyanate mean fluorescence intensity. b, Schematic of the analysis strategy. scCUT&Tag signal was aggregated into a cell × bin matrix, with various bin sizes (5 kb or 50 kb); dimensionality reduction was performed using LSI and UMAP and clustering using SNN. Cell clusters were used to identify marker regions, gene activity scores were calculated per cell and integration with other datasets was performed. c, Comparison of number of unique reads per antibody per cell in scCUT&Tag experiments. Lower and upper bounds of boxplot specify 25th and 75th percentile, respectively, and lower and upper whiskers specify minimum and maximum, respectively, no further than 1.5× interquartile range. Outliers are not displayed. H3K4me3: n = 13,739 cells from four biological replicates; H3K27ac: n = 10,414 cells from two biological replicates; H3K27me3: n = 13,932 from four biological replicates; H3K36me3: n = 4,350 cells from two biological replicates. d, Comparison of percentage of reads falling into peak regions per antibody in scCUT&Tag experiments. Peaks were obtained by peak calling in merged bulk datasets. Lower and upper bounds of box plot specify 25th and 75th percentiles, respectively, and lower and upper whiskers specify minimum and maximum, respectively, no further than 1.5× interquartile range. Outliers are not displayed. H3K4me3: n = 13,739 cells from four biological replicates; H3K27ac: n = 10,414 cells from two biological replicates; H3K27me3: n = 13,932 from four biological replicates; H3K36me3: n = 4,350 cells from two biological replicates. e, Distribution of fragment lengths in scCUT&Tag experiments per antibody. f–i, 2D UMAP representation of the scCUT&Tag data for H3K4me3, n = 4, f; H3K27me3, n = 4, g; H3K27ac, n = 2, h and H3K36me3, n = 2, i. j, Pseudobulk scCUT&Tag profiles of H3K4me3 aggregated by cell type at marker loci. k, Heatmap showing H3K4me3 signal intensity in top 50 most specifically enriched genomic bins per cluster (rows) and single cells (columns). Cells are randomly sampled and 5% of total number of cells are displayed. Color bars in rows specify marker cluster, and in columns cell metadata (Age, GFP).
modifications (Fig. 1f–i and Fig. 2a–f). By identifying specific peaks proximal to promoters of marker genes (Fig. 1j,k and Fig. 2a–f), we manually annotated the populations as mature oligodendrocytes (MOL, Mbp+, Mog+ and Cldn11+), astrocytes (AST, Slc1a2+, Rfx4+ and Aqp4+), olfactory ensheathing cells (OEC, Alx3+, Alx4+ and Frzb+), vascular cells (VAS, Nes+, Tbx18+ and Foxf2+)

\[
\begin{align*}
\text{P15/P25} & \quad \text{Sox10-Cre/GFP} \\
\text{Sox10-Cre/GFP} & \quad \text{10x Chromium scATAC}
\end{align*}
\]
**Fig. 2 | De novo identification of cell types by cell-type-specific scCUT&Tag marker regions.** a, b, Projection of scCUT&Tag gene activity scores of H3K4me3 (a) and H3K27me3 (b) on 2D UMAP embedding. Gene name is depicted in the title and specific population is highlighted in the UMAP plot by labeling the cell type. c, d, Heatmap representation of the scCUT&Tag signal for H3K4me3 (c) and H3K27me3 (d). The x axis represents genomic region, each row in the y axis contains data from one cell. Cell correspondence to clusters is depicted by the color bar on the right side of the heatmap and annotated with cell type. Signal is aggregated per 250 bp windows and binarized. e, f, Aggregated pseudobulk scCUT&Tag profiles for four histone modifications in all identified cell types at the loci of selected marker genes (Slc1a2 representing AST (e) and Mbp oligodendrocytes (f)).
and cells within the spectrum of oligodendrocyte progenitors cells (OPCs), committed OPCs (COPs) and newly formed oligodendrocytes (NFOLs) (Pdgfra+, Neu4+ and Gpr17+) in the GFP+ fraction (Fig. 1j, Fig. 2a–f and Extended Data Fig. 4a). GFP− cells comprised mainly neurons (NEU, Rbfox3+ and Neurod2+), both excitatory (Exc, Slic7a5+) and inhibitory (Inh, Gad1+ and Gad2+), astrocytes (S1a2+, Rjkd4+ and Aapg4+) and microglia (MGL, Ciga+ and CD45/Ptprc+) (Fig. 1j, Fig. 2a–f and Extended Data Fig. 4a). We could identify similar populations in H3K27me3 scCUT&Tag and annotate them using a combination of markers that lacked the repression mark in the vicinity of the marker gene regions (Fig. 1j, Fig. 2b, d–f and Extended Data Fig. 4a).

The clustering was highly reproducible among biological replicates, and cells originating from P15/P25 age were well intermingled within the clusters (Extended Data Fig. 5a–c). The cell states of OLG lineage reflect mouse age, with the majority of OPCs originating from P15 and differentiated OLG coming from P25 (Extended Data Fig. 5b,c). Interestingly, we detected a major, likely transient, population of AST in the GFP+ fraction of the Sox10Cre/RCE mouse (Extended Data Fig. 2a,c), most probably derived from ventral regions.  

To benchmark the scCUT&Tag profiles of brain cell populations, we compared them with previously published bulk brain H3K27me3 ChIP-seq and in-house generated bulk H3K27me3 CUT&Run data. Inspection of genome browser tracks revealed similarity between corresponding populations of bulk and single-cell tracks and lower level of background signal in scCUT&Tag data (Extended Data Fig. 6a). Moreover, the corresponding bulk and scCUT&Tag datasets clustered together on a PCA plot generated from the top 150 most variable marker regions (Extended Data Fig. 6b). We also performed scRNA-seq from Sox10Cre/GFP+ sorted cells and determined that the proportions of cell types obtained by scCUT&Tag and by scRNA-seq were similar (Extended Data Fig. 6c).

Integration of scCUT&Tag data with single-cell gene expression. To validate the manual annotation of clusters, we used the adolescent mouse brain scRNA-seq atlas. We picked the 100 most specifically expressed marker genes for selected populations, generated metagene modules and then calculated a gene activity score (scCUT&Tag signal in gene body and promoter) within the module. We found that the specific cell clusters showed enrichment of the metagene signal for the active modifications and were depleted of signal in the H3K27me3 dataset (Extended Data Fig. 7), supporting the cluster annotations. Furthermore, we integrated the H3K4me3 scCUT&Tag with the scRNA-seq data using canonical correlation analysis (CCA)7 at the single-cell level. We found that the major cell populations cocluster together with the corresponding scRNA-seq population (Fig. 3a). Finally, we used gene ontology (GO) terms analysis to functionally annotate the H3K4me3 scCUT&Tag clusters and found GO terms such as astrocyte differentiation and activation (AST), myelination (OLG), regulation of myelination (OEC), cell migration involved in vasculogenesis (VAS), glial cell development (OPCs), neuron development, neuron maturation and axonogenesis (NEU) and microglial cell activation involved in immune response (MGL) (Extended Data Fig. 8) to be specifically enriched in respective clusters.

scCUT&Tag data were able to resolve the major cell types in a heterogeneous sample. However, a previous study had shown that further cell subtypes can be detected in the population of oligodendrocytes48. Therefore, we asked whether this heterogeneity can be resolved by integration of scCUT&Tag data with an existing OLG scRNA-seq dataset. For this purpose, we co-embedded the H3K4me3 scCUT&Tag and scRNA-seq using CCA, which showed good integration of the techniques, while the scRNA-seq clustering was retained (Fig. 3b). We then used metagene scores of OPC, MFO, MOL1, MOL2 and MOL5 to reveal cell subtype signatures within the H3K4me3 scCUT&Tag data. Interestingly, we found that the population of oligodendrocytes that appeared homogenous could be further deconvoluted into subpopulations enriched in module-specific genes (Fig. 3c), indicating that oligodendrocyte heterogeneity is reflected at an epigenetic level.

Differential global and genome-wide patterns of histone modifications in single-cell populations. Since the scCUT&Tag profiles were generated simultaneously for all present populations, it allowed for quantitative analysis of their global and genome-wide pattern of histone modifications. We used the number of unique reads per cell as a proxy for the absolute amount of histone modification in single cells. We found a substantial variability in this regard (Fig. 3d). This is most prominent for H3K27me3, which is enriched in populations of oligodendrocytes, microglia and a subset of neurons relative to the other populations (Fig. 3d). Enrichment of H3K27me3 in oligodendrocytes is consistent with the recent finding that H3K27me3 drives the oligodendrocyte–astrocyte switch during development. Interestingly, we also observed relatively higher amounts of H3K36me3 in the population of immature oligodendrocytes (OPC/COP–NFOL stages) (Fig. 3d). Although heterogeneity in the global levels of antibody-specific signal in cell types could be caused by differential permeability and/or tagmentation, we did not notice consistent enrichment of signal across all histone modifications in specific cell types. Therefore, it is unlikely that signal heterogeneity is caused by differential tagmentation efficiency between cell types, but rather by differential levels of modifications between cell types.

Next, we asked whether we could assign cell populations across the different active modifications and cross-correlate them. For this purpose, we used the CCA to integrate the data at gene resolution. Strikingly, the two-dimensional (2D) representation of the data obtained recapitulated the original nonsupervised clustering with great precision (Fig. 3e), and the clusters annotated with the same cell type in different datasets co-occupied the same low dimensional space (Fig. 3e). To further examine the interplay between active and repressive marks, we identified all active promoters specific for individual populations marked by H3K4me3, and plotted the signals of H3K4me3 (Fig. 3f) and H3K27me3 (Fig. 3g) per cluster for all populations. As expected, we observed depletion of H3K27me3 when the promoter is enriched in H3K4me3 in the respective population (Fig. 3g). Interestingly, we noticed that astrocyte-specific genes had higher H3K4me3 signal in OLGs than in MOLs (Fig. 3f).
Besides, H3K27me3 signal was depleted from OPC-specific genes in MOLs, but not in astrocytes (Fig. 3g), suggesting that H3K27me3 is not required to repress OPC genes during MOL differentiation. In contrast, H3K27me3 signal was present in AST-specific genes in OPCs and MOLs (Fig. 3g). Consistently, it was reported recently that disruption of H3K27me3 impairs OPC differentiation to MOL, and triggers a switch towards an astrocytic fate29. Moreover, the presented epigenetic profiles suggest that AST are epigenetically (H3K4me3) related to OPCs.

**Increase in breadth of H3K4me3 upon oligodendrocyte differentiation.** The breadth of the H3K4me3 mark has been linked previously to cell identity, gene expression and transcriptional consistency across a variety of cell types30. We noticed in the H3K4me3 pileup analysis that both amplitude and breadth of the H3K4me3 signal were increased at population-specific marker gene promoters when compared with marker gene promoters of other populations (Extended Data Fig. 4e–h). To quantify breadth, we specifically looked at promoters of marker genes identified from scRNA-seq data. We found that the marker genes of the identified populations had, on average, higher H3K4me3 breadth (Fig. 4a). Moreover, the magnitude of the breadth was different for individual cell types, with the broadest H3K4me3 peaks on the promoters of AST and oligodendrocytes, and the narrowest on vascular and leptomin-geal cells (VLMCs) but also OPCs (Fig. 4b), which could suggest an increase of H3K4me3 breadth during transition from progenitor states to fully differentiated states. To further look into the dynamics of H3K4me3 spread, we leveraged the ability of scCUT&Tag to visualize H3K4me3 spreading at single-cell resolution and investigated H3K4me3 breadth during the process of differentiation of MOLs from OPCs. We generated single-cell H3K4me3 metagene profiles around MOL-specific marker genes. We then ordered the cells (OPCs and MOLs) in the matrix according to H3K4me3 signal coming from genes expressed in MOLs (MOL signature) and validated by pseudotime analysis (Fig. 4c,d). Strikingly, we observed a gradual increase in the breadth of the H3K4me3 signal at MOL promoters with single-cell resolution (Fig. 4e), which is consistent with spread of H3K4me3 as the cells progress towards differentiated oligodendrocyte identity.

**scCUT&Tag of TFs.** TF binding is notoriously difficult to profile using ChIP-seq in low input samples. Therefore, we asked whether scCUT&Tag was able to uncover the binding of TFs at single-cell resolution. We chose the TFs OLIG2, as it is specific for glial populations, and RAD21, a general chromatin architecture factor and a subunit of the cohesin complex. We performed scCUT&Tag in GFP+ sorted cells from the brain at postnatal day P25. The number of unique reads per cells was lower for TF scCUT&Tag compared with histone modifications. Nevertheless, we were able to obtain a median 48 and 240 unique reads per cell in OLIG2 and
RAD21 scCUT&Tag, respectively. We reduced the dimensionality of the dataset using LSI and UMAP, and obtained specific clusters for RAD21 and OLIG2 (Fig. 5a–d). The OLIG2 dataset separated into two clusters based on depth (number of unique reads per cell), and we annotated clusters with low numbers of unique reads as ‘low binders’ (Fig. 5a). Since manual annotation of populations based on markers is challenging in TF CUT&Tag (Fig. 5e), we based cluster annotation on the assumption that OLIG2/RAD21 binding in specific cell types is correlated with enhancer/promoter activity. Therefore, we analyzed OLIG2/RAD21 binding in promoter regions of genes that are specifically modified by H3K4me3 in scCUT&Tag data, and identified populations of AST, OLG and OEC in RAD21 scCUT&Tag and OLIG and non-OLGs (low binders) in OLIG2 scCUT&Tag (Extended Data Fig. 9a–d). The OLG population in OLIG2 scCUT&Tag is likely composed of both mature OLGs and OPCs, which appear to form a subpopulation within the OLG cluster (Extended Data Fig. 9a). To further strengthen the cluster annotation, we performed co-embedding of the RAD21/OLIG2 with another histone modification – H3K27ac using CCA – and found that the identified clusters consistently co-embedded with the corresponding H3K27ac clusters (Extended Data Fig. 9e). Interestingly, the low binder non-OLGs cells randomly co-embedded with the populations of OEC, AST and vascular cell clusters that were defined by H3K27ac, whereas OLG clusters specifically co-embedded with the OLG H3K27ac signal (Extended Data Fig. 9e). This finding is consistent with expression of Olig2 throughout cell types, it being highly expressed throughout the OLG lineage, whereas OECs and VLMCs do not express Olig2 and only small portion of ASTs express Olig2 (Fig. 5f).

To validate the specificity of scCUT&Tag, we searched for enriched motifs in binding sites using MEME suite in merged pseudobulk datasets of RAD21 and OLIG2. We found the motif of chromatin architecture factor CTCF as the highest enriched in the RAD21 dataset (Fig. 5g), which is consistent with the cooperativity between CTCF and cohesin. We found several motifs enriched in OLIG2 scCUT&Tag, including the motif CAGMTG, similar to the previously identified CAGMTG/CAGCTG motif specific for OLIG2 (Fig. 5h) in mouse and rat, respectively. Together with the previously identified OLIG2 motif, we found enrichment of multiple general eukaryotic enhancer and promoter features (GC box and CAAT box). Interestingly, we also found a motif similar to the motif of TFs from the SOX family (ACARWR, Extended Data Fig. 9f), which is consistent with physical interaction and cooperativity between OLIG2 and members of the SOX family of TFs (SOX8, SOX10).

**Prediction of enhancer–promoter interactions from scCUT&Tag data.** Perhaps the most intriguing and challenging task of epigenomics is to use epigenetic data to predict gene regulatory networks. Our scCUT&Tag dataset is a rich resource that can be used to tackle this question in specific cell populations. We used the
Fig. 6 | Prediction of gene regulatory networks from scCUT&Tag data. a, Schematic depicting the strategy used to predict and validate promoter-enhancer specific loops. Loops are predicted using the H3K27ac scCUT&Tag and other publicly available datasets (Methods). Presence of loops is validated by H3K27ac HiChIP analysis of purified populations of OLG lineage. b, Pileup analysis of 200,000 loops predicted by the ABC model. Signal of HiChIP performed in either mESC or OLG lineage cells was aggregated and plotted as a heatmap with the center at the intersection of the loop coordinates. c, Pileup analysis of loops predicted by ABC and filtered using H3K4me3 (~61,000 loops) and RAD21 (~5,000 loops) scCUT&Tag data. d, Representative overlay of OLG HiChIP matrix with the loops predicted by the ABC model from scCUT&Tag data and filtered with the H3K4me3 signal (black dots). e, Representative example of loops predicted by Cicero and the ABC model from H3K27ac scCUT&Tag data. Known Sox10 enhancers and putative new candidate enhancers are highlighted with gray bars.
activity-by-contact (ABC) model of enhancer–promoter interactions to predict the gene-enhancer regulatory networks from aggregated scCUT&Tag data (Fig. 6a). We focused on OLG and ran the ABC model using oligodendrocyte H3K27ac scCUT&Tag, oligodendrocyte scATAC-seq data (P50 cortex, 10x Genomics) and Hi-C data of neural progenitor cells (OLG). By using Hi-C data of neural progenitor cells we ensured that we did not provide the model with accurate measurements of DNA looping, but rather used the Hi-C data to estimate long-distance chromosome topology. The ABC model predicted ~200,000 enhancer–promoter loops in the OLG (Supplementary Table 2). The predicted loops were consistent with loops predicted by the ABC model from bulk OLG CUT&Run data (Extended Data Fig. 10a,b). We also downscaled the scCUT&Tag data and could see the robustness of predictions from as few as 100 cells (74.4% overlap with full dataset, Extended Data Fig. 10c).

To examine the contact probability of predicted interactions, we performed HiChIP of H3K27ac from primary mouse OLG cultures and could see the robustness of predictions from as few as 100 cells (Extended Data Fig. 10a,b). We also downscaled the scCUT&Tag data and could see the robustness of predictions from as few as 100 cells (74.4% overlap with full dataset, Extended Data Fig. 10c).

To demonstrate the performance of scCUT&Tag, we generated scCUT&Tag profiles for glial TF OLIG2 and chromatin architecture factor RAD21. While OLIG2 preferentially binds OLG-specific enhancers and promoters and binds only weakly to AST-specific gene regulatory regions, RAD21, a ubiquitously expressed genome architecture factor, binds cell-type-specific promoters and enhancers. Although TF scCUT&Tag data is sparser than scCUT&Tag data of histones, we were still able to distinguish cell identity based only on RAD21 binding for all major cell types, except for glial cells and OPCs, which are the smallest cell populations and share features with larger populations. RAD21 binding at single-cell resolution provides valuable insights into enhancer–promoter connectivity and we show that it can be incorporated into models that aim to uncover enhancer–promoter connections.

Although the datasets obtained provide unique insights into epigenetic regulation, improvements in the technology, especially the number of unique fragments per cell, can further enhance possible applications. In its current state, scCUT&Tag is able to distinguish the major cell types but fails to uncover heterogeneity of subpopulations in an unsupervised manner. We could reveal the oligodendrocyte heterogeneity at H3K4me3 level only upon integration with more complex scRNA-seq data. Potential introduction of multomic approaches, such as simultaneous measurement of scCUT&Tag and RNA-seq signal from the same cell, will help us to further understand causal relationships between epigenetic modifications and gene expression.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of
Received: 1 September 2020; Accepted: 18 February 2021; Published online: 12 April 2021

References

1. Zeisel, A. et al. Molecular architecture of the mouse nervous system. Cell 174, 999–1014.e22 (2018).
2. La Manno, G. et al. RNA velocity of single cells. Nature 560, 494–498 (2018).
3. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523, 486–490 (2015).
4. Lorthongpanich, C. et al. Single-cell DNA methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science 341, 1110–1112 (2013).
5. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Cell* 6, e21856 (2017).
6. Kaya-Otkur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* 10, 1930 (2019).
7. Rotem, A. et al. Single-cell ChIP–seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* 33, 1165–1172 (2015).
8. Ai, S. et al. Profiling chromatin states using single-cell iChIP–seq. *Nat. Cell Biol.* 21, 1164–1172 (2019).
9. Wang, Q. et al. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell* 76, 206–216.e7 (2019).
10. Xu, W. L. et al. Single-cell chromatin immunocleavage sequencing (scChIC–seq) to profile histone modification. *Nat. Methods* 16, 323–325 (2019).
11. Carter, B. et al. Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT–seq). *Nat. Commun.* 10, 3747 (2019).
12. Grosselin, K. et al. High-throughput single-cell ChIP–seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* 51, 1060–1066 (2019).
13. Falcão, A. M. et al. Disease-specific oligodendrocyte lineage cells arise in multiple sclerosis. *Nat. Med.* 24, 1837–1844 (2018).
14. Marques, S. et al. Transcriptional convergence of oligodendrocyte lineage progenitors during development. *Dev. Cell* 46, 504–517.e7 (2018).
15. Jäkel, S. et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. *Nature* 566, 543–547 (2019).
16. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* 352, 1326–1329 (2016).
17. Jung, M. et al. Lines of murine oligodendroglial precursor cells immortalized by an activated neu tyrosine kinase show distinct degrees of interaction with axons in vitro and in vivo. *Eur. J. Neurosci.* 7, 1245–1265 (1995).
18. Davis, C. A. et al. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* 46, D794–D801 (2018).
19. Xiao, Y. et al. Hippo pathway deletion in adult resting cardiac fibroblasts initiates a cell state transition with spontaneous and self-sustaining fibrosis. *Genes Dev.* 33, 1491–1505 (2019).
20. Sousa, V. H., Miyoshi, G., Hjerling-Leffler, J., Karayannis, T. & Fishell, G. Characterization of Nkx6-2–derived neocortical interneuron lineages. *Cereb. Cortex* 19, 11–110 (2009).
21. Matsuoka, T. et al. Neural crest origins of the neck and shoulder. *Nature* 436, 347–355 (2005).
22. Huang, W. et al. Novel NG2-CreERT2 knock-in mice demonstrate heterogeneous differentiation potential of NG2 glia during development. *Glia* 62, 896–913 (2014).
23. Zhu, X. et al. Age-dependent fate and lineage restriction of single NG2 cells. *Development* 138, 745–753 (2011).
24. Zhu, X., Bergles, D. E. & Nishiyama, A. NG2 cells generate both oligodendrocytes and gray matter astrocytes. *Development* 135, 145–157 (2008).
25. Matsuda, T. et al. Pioneer factor NeuroD1 rearranges transcriptional and epigenetic profiles to execute microglia–neuron conversion. *Neuron* 101, 472–485.e7 (2019).
26. Boney, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* 171, 557–572.e24 (2017).
27. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21 (2019).
28. Wang, W. et al. PR2C acts as a critical timer that drives oligodendrocyte fate over astrocyte identity by repressing the notch pathway. *Cell Rep* 32, 108147 (2020).
29. Wang, J. et al. EED-mediated histone methylation is critical for CNS myelination and remyelination by inhibiting WNT, BMP, and senescence pathways. *Sci. Adv.* 6, eaae6477 (2020).
30. Benyoun, B. A. et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* 158, 673–688 (2014).
31. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208 (2009).
32. Li, Y. et al. The structural basis for cohesin–CTCF–anchored loops. *Nature* 578, 472–476 (2020).
33. Yu, Y. et al. Olig2 targets chromatin remodelers to enhancers to initiate oligodendrocyte differentiation. *Cell* 152, 248–261 (2013).
34. Darr, A. J. et al. Identification of genome-wide targets of Olig2 in the adult mouse spinal cord using ChIP-Seq. *PloS One* 12, e0186091 (2017).
35. Wißmüller, S., Kosian, T., Wolf, M., Finzsch, M. & Wegner, M. The high-mobility-group domain of Sox proteins interacts with DNA-binding domains of many transcription factors. *Nucleic Acids Res.* 34, 1735–1744 (2006).
36. Fulco, C. P. et al. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* 51, 1664–1669 (2019).
37. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* 49, 1602–1612 (2017).
38. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* 171, 305–320.e24 (2017).
39. Pliner, H. A. et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* 71, 858–871.e8 (2018).
40. Werner, T., Hammer, A., Wahlbuhl, M., Bösl, M. R. & Wegner, M. Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. *Nucleic Acids Res.* 35, 6526–6538 (2007).
41. Moore, J. E. et al. Expanded encyclopedia of DNA elements in the human and mouse genomes. *Nature* 583, 699–710 (2020).

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. 2021
Methods

Animals. The mouse line used in this study was generated by crossing Sox10::Cre (The Jackson Laboratory mouse stock number 025807) on a C57BL/6j genetic background with RCEloxP (enhanced green fluorescent protein (EGFP)) animals (The Jackson Laboratory mouse stock number 32307-JAX on a C57BL/6xCD1 mixed genetic background. Females with a homozygous Cre allele were mated with males lacking the Cre allele, whereas the reporter allele was kept in hemizygosity or homozygosity in both females and males. In the resulting Sox10::Cre–RCEloxP(EGFP) animals the entire OLG was labeled with EGFP.

Breeding with males containing a hemizygous Cre allele in combination with the reporter allele to non-Cre carrier females resulted in early Cre activation in the reporter allele. Sox10::Cre/RCE animals on P15 or P21–P25 using fluorescent-assisted cell sorting (FACS) were used for reporter allele to non-Cre carrier females resulted in early Cre activation in the reporter allele.

All cage changes were done in a laminar airflow cabinet. Facility personnel wore dedicated scrubs, socks and shoes. Respiratory masks were used when working outside of the laminar airflow cabinet. Littermate animals were sacrificed either outside of the laminar airflow cabinet. Littermate animals were sacrificed either.

Food and water were provided as standard chow and water.

General housing parameters such as relative humidity, temperature and ventilation follow the European convention for the protection of vertebrate animals used for experimental and other scientific purposes treat ETS 123. Briefly, consistent relative air humidity of 50%, 22°C is controlled. The use of stand-alone housing units supplemented with high efficiency particulate air-filtered air.

The processing pipeline was built using the Snakemake platform. Preprocessing pipeline and R notebooks used to perform analysis and generate the figures were shared at https://github.com/Castelo-Branco-lab/scCut-Tag. 2020. The Integrative Genomics Viewer server with bigwig tracks merged per replicate is available at https://raw.githubusercontent.com/marzix/IGV_track_server/master/registry/IGV_registry.txt. Additional information is available in the Nature Research Reporting Summary linked to this article.

Bulk CUT&Run and CUT&Tag. CUT&Run was performed on Oli-neu (H3K27me3) or MACS-sorted OPCs (H3K27me3) as described in Skene and Henikoff. The sequencing library was prepared using a Kapa HyperPrep library prep kit (Kapa Biosystems, 0796236001). CUT&Tag was performed on CD140 MACSed and cultured primary mouse OPCs (C5BL/6j WT) with anti-H3K27ac antibodies as described in ref. 48.

Protein–A–Tn5 production. The 3xFLAG protein A (PA)–fusion sequence was acquired from Addgene plasmid #124601 and inserted into plasmid pTSn5 (Addgene #79107) to generate a pa–Tn5 fusion construct (6xHis–TEV–3xFLAG–pa–Tn5). The production was purified from 31 of Escherichia coli culture grown in the LEX system, with a protein expression temperature of 18°C and a cultivation temperature of 30°C. The temperature was switched to 18°C at optical density (OD) = 2 and expression induced at OD = 3 with 0.5 mM isopropyl-β-D-thiogalactoside. Bacteria were disrupted by sonication (4× on/off cycle, 3 min, 80% amplitude), centrifuged for 20 min at 40000 g, filtered through a 0.45μm filter, loaded on an AKTA express column and purified overnight. His-TEV–3xFLAG–pa–Tn5 was dialyzed for affinity purification. Pull-down purification was performed on HiLoad 16/60 Superdex 200 (GE Healthcare). Fractions were examined on SDS–PAGE gel before pooling, then desired fractions were combined, the sample was diluted 1:5 to contain final 30% glycerol, divided into 200 μl aliquots and snap frozen using liquid nitrogen. Enzyme was stored at −80°C until loading.

Tn5 loading. To begin complex formation, Mosaic-end–adapter A (TsN5ME-A, TCGTTGCCAGCCGTCAGATGTGTATAGAGAGACAG) and Mosaic-end–adapter B (TsN5ME-B, GTCCTCGGGCCTCGAGATGTGTATAGAGAGACAG) oligonucleotides were each annealed with Mosaic end–reverse oligonucleotides (TsN5MErev, 5′-phos)GCTGTCCTTATACACATCT-3′). To anneal, the oligonucleotides were diluted to 100μM. Each pair of oligos, TsN5MErev/TsN5ME-A and TsN5MErev/TsN5ME-B, was mixed and annealed separately. The mixed oligonucleotides were denatured in a thermocycler for 5 min at 95°C and then cooled slowly at ramp rate 0.1°C s⁻¹. The loading of pa–Tn5 with the
oligonucleotides was carried out by incubating 2 μl 50μM TsN5Merev/TsN5Merev-A, 2 μl 50μM TsN5Merev/TsN5Merev-B, 21.56 μl glycerol, 2.31 μl 2x dialysis buffer (100 mM HEPES–KOH pH 7.2, 0.2 M NaCl (Thermo Fisher Scientific, AM9760G), 0.2 mM EDTA (Thermo Fisher Scientific, AM9260G), 2 mM DTT, 0.2% Triton X-100 (Thermo Fisher Scientific, 8511) and 20% glycerol) and 3.14 μl PA-Tn5 (63.64 μM (3.5 mg/mL)), for 1 h at room temperature (the final concentration is 2 μM). The enzyme was stored at −20 °C until further use, or at −80°C for long-term storage.

Tissue dissociation. Mice were sacrificed, perfused with 1× PBS, pH 7.4 and the brain was removed. The brain was dissociated into single-cell suspension using Neural Tissue Dissociation Kit P (Millenyi Biotec, 130-092-628) according to the manufacturer’s protocol. For mice older than P7, myelin was removed using debris removal solution (Millenyi Biotec, 130-109-398) according to the manufacturer’s instructions. Single-cell suspension was filtered through a 50 μm cell strainer and briefly stained with 1:5,000 diluted DAPI (1 μg/mL) to assess cell viability. For FACs, cells were resuspended in 1× PBS supplemented with 1% BSA and 2 mM EDTA and kept at 4°C until sorted.

Primary OPC culture. Mice brains from pups at P4–P6 stage were dissected and dissociated (see above), and the single-cell suspension was used to enrich for OPCs using CD140a microbeads (Millenyi Biotec, 130-101-502) according to the manufacturer’s instructions. Brains from four to five mice were pooled for each batch of OPC culture. Cells were seeded on Petri dishes or multwell plates precoated with poly-l-lysine (Sigma, P4707) for >1 h (Sigma, F1141, 1 mg ml−1 stock, 1:1,000 diluted in 1× PBS). Cells were cultured in OPC proliferation medium Dulbecco’s Modified Eagle Medium/ GlutaMAX (DMEM/Gmax; Thermo Fisher Scientific, 10565018), 1× N2 supplement (Thermo Fisher Scientific, 17520048), 1× penicillin-streptomycin (Thermo Fisher Scientific, 15102122), 1× Neurobrew (ThermoFischer Scientific, 130-093-566), basic fibroblast growth factor (bFGF) 20 ng/ml (Peprotech, 100-188) and platelet-derived growth factor (PDGF)-AA 10 ng/ml (Peprotech, 100-13) until confluency (5–6 days), passed once and collected 72 h later.

Cell-line culture. NIH-3T3 cells were cultured in DMEM (Gibco, 41966029) supplemented with 10% fetal bovine serum (FBS; Gibco, 10099182) and 1× Earle’s supplemented with 10% fetal bovine serum (Gibco, 10500064) and 1× penicillin-streptomycin (Gibco, 151022). N2, and mESC, Thermo Fisher Scientific, A1110501) when they reached 80% confluency. 

scRNA-seq. GFP+ cells were sorted using the same protocol as for scCUT&Tag but collected into 1× PBS + 0.5% BSA instead of antibody buffer. Then, cells were counted and ~7,000 cells processed using chromium single cell 3′ reagent kit (10x Genomics) according to the manufacturer’s instructions in two biological replicates.

HiChIP. HiChIP was performed in three biological replicates with at least 1 million cells used as input. Briefly, cultured cells were collected using TrypLE Express (3T3, Thermo Fisher Scientific, 12604013) or accutase (Öli-neu and mESC, Thermo Fisher Scientific, A1110501) when they reached 80% confluency.

scRNA-seq. GFP+ cells were sorted using the same protocol as for scCUT&Tag but collected into 1× PBS + 0.5% BSA instead of antibody buffer. Then, cells were counted and ~7,000 cells processed using chromium single cell 3′ reagent kit (10x Genomics) according to the manufacturer’s instructions in two biological replicates.

The pipeline’s hicpro2juicebox.sh script was used to generate .hic files, which were loaded into Juicebox for viewing contact maps. The hic2cool tool (https://github.com/4dn-dcc/hic2cool) was used to generate 5-kb resolution .cool files.

ABC model. ATAC–seq data were downloaded from 10x Genomics online resources (https://support.10xgenomics.com/single-cell-atac/datasets/1.2.0/atac_v1_adult_brain_fresh_5k) and cell-type-specific peaks were called using MACS2. Cell-type-specific H3K27ac bam files were generated from the cell-ranger ATAC output files. Gene expression data from the scRNA-seq data were processed using scChIPseq_DataEngineering pipeline (https://github.com/vallotlab/scChIPseq_DataEngineering)12. We used deepTools plotFingerprint to generate fingerprint plots.

Comparison with other datasets. We used the deepTools multiBigWig summary to generate summary matrices for bulk and single-cell bigwig tracks that were used as predefined candidate regions to look for enrichment. Peaks were performed on OPC-derived H3K27ac HiChIP data and mESC data. All HiChIP data were balanced using the cooler library before performing the pileups. To find overlapping loops called by the ABC model and Cicero, we first binned the loops into genomic bins with 10-kb resolution and flattened the overlapping loops within one sample. Then we searched for overlaid loops between the samples.

Data availability
Raw data are deposited in GEO under accession GSE163532. The scCUT&Tag dataset can be explored at web resources available at https://ki.se/en/mmb/oligointernode and https://mouse-brain-cutandtag.cells.ucsc.edu/, The following publicly available datasets were used in this study: GSE96107 (HiC of mESC and H3K27me3 ChIP–seq of cortical neurons), GSE124857 (scCUT&Tag in icells8 platform), GSE17389 (scCUT&seq), GSE194435 (H3K27me3 ChIP–seq of microglia), SRP135960 (scRNA-seq of the mouse brain mouse), GSE75330 (scRNA-seq of oligodendrocyte lineage) and GSE135296 (H3K27me3 CUT&Run of mouse NIH/3T3 cells). Data were also accessed for the Mouse Brain Atlas (https://storage.googleapis.com/innarsson-lab-loom/l5_allloom), 10x Genomics single-cell ATAC–seq of P50 mouse cortex (https://support.10xgenomics.com/single-cell-atac/datasets/1.2.0/atac_v1_adult_brain_fresh_5k) and ENCODE H3K27me3 ChIP–seq of Bruese mEsCs (https://www.encodeproject.org/experiments/ENCST000CFJ/).

Code availability
Code is available at https://github.com.Castelo-Branco-lab/scCUT-Tag_2020. References
42. Zhang, Y. et al. Model-based analysis of ChIP–Seq (MACS). Genome Biol. 9, R137 (2008).
43. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 19, 477 (2018).
44. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 46, D260–D266 (2018).
45. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016).
46. Wickham, H. ggplot2: Elegant Graphics for Data Analysis (Springer, 2009); https://doi.org/10.1007/978-0-387-99910-6_3
47. Körner, J. & Rahmann, S. Snap-seqmate: a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522 (2012).
48. Kaya-Okur, H. Bench top CUT&Tag. https://doi.org/10.17504/protocols.io.bcluvh6 (2020).
49. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259 (2015).
50. Durand, N. C. et al. Juicbox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101 (2016).
51. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile engine.


Acknowledgements

We would like to thank T. Jimenez-Beristain for writing laboratory animal ethics permit 1995_2019 and assistance with animal experiments; I. Kerby, M. Mejer and P. Kukanja for assistance with animal experiments; F. Gabriel for setting up CUT&Bun technology; S. Henikoff for providing protein A–Tn5 and protein A–MNase; B. Hervé for setting up the Shiny web resource; M. Speir and M. Haeusler at the University of California Santa Cruz Genomics Institute for setting up the UCSC Cell Browser and UCSC Genome Browser resources; S. Elsäser and R. Sandberg for proofreading and providing critical comments on the manuscript; T. Johansson and E. Dunevall and the Karolinska Institute Protein Science facility for cloning the pA–Tn5 construct and the protein purification; M. Eriksson for performing the 10x ATAC–seq protocol, the Single Cell Genomics Facility and S. Elsäser, Karolinska Institutet, for providing mESCs; O. Fernandez Capetillo, Karolinska Institutet, for providing NIH-3T3 cells and the staff at Comparative Medicine-Biomedicum and Biomedicum flow cytometry core facility (BFC). Figure 1a was generated using BioRender. We acknowledge support from the National Genomics Infrastructure in Stockholm funded by the Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish National Infrastructure for Computing/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. M.B. is funded by the Vinnova Seal of Excellence Marie-Skłodowska Curse Actions grant RNA-centric view on Oligodendrocyte lineage development (RODent). Work in G.C.-B’s research group was supported by the Swedish Research Council (grant nos. 2015-03558 and 2019-01360), the European Union (Horizon 2020 Research and Innovation Programme/European Research Council Consolidator Grant EPiSCOPE, grant agreement no. 681893), the Swedish Brain Foundation (FO2017-0075 and FO2018-0162), the Swedish Cancer Society (Cancerfonden; 190394 P), the Knut and Alice Wallenberg Foundation (grants no. 2019-0107 and 2019-0089), The Swedish Society for Medical Research (SSMF, grant no. JUB2019), the Ming Wai Lau Center for Reparative Medicine and the Karolinska Institutet.

Author contributions

M.B. and G.C.-B. conceived the study, designed the experiments and analysis and wrote the manuscript. M.B. optimized and performed the scCUT&Tag experiments and analyzed the scCUT&Tag data. M.K. and M.B. performed the HiChIP experiment. M.K. analyzed the HiChIP data and helped with generation of related figures. All authors contributed and approved the manuscript.

Competing interests

M.B. and M.K. have performed paid consultation for the company Abcam.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-021-00869-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-00869-9.

Correspondence and requests for materials should be addressed to M.B. or G.C.-B.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Optimization of tagmentation step in the scCUT&Tag protocol by addition of BSA. 

**a**, Table depicting scCUT&Tag protocol steps (Primary Antibody, Secondary Antibody, Tn5 binding, Tagmentation) and whether BSA was included in the scCUT&Tag buffers during these steps. 

**b**, DAPI counterstaining of the nuclei after the scCUT&Tag procedure. Inclusion of BSA in the procedure substantially reduces clumping of the nuclei. 

**c**, Genome browser profiles of bulk CUT&tag experiment from the different BSA conditions (described in **a**). 

**d**, Gating strategy depicting sorting of GFP cells. P1 Depicts general gate for selection of cells, P2 gate selects for singlets, DAPI- gate selects only live cells and GFP+ gate selects cells that possess GFP signal (Sox10-Cre/GFP).
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | scCUT&Tag of mixture of mouse cell lines. a, UMAP projection of H3K27me3 scCUT&Tag in two dimensions. Points are colored by assigned cell identity. N = 2 technical replicates b, Genome browser view of merged pseudobulk profiles of 5 scCUT&Tag clusters and bulk ChIP-seq or CUT&Run profiles of the respective cell lines. c, PCA analysis and Pearson correlation matrix of 5 scCUT&Tag clusters and bulk ChIP-seq or CUT&Run data. The PCA was performed on the top 150 most variable marker regions selected from the scCUT&Tag data. Heatmap shows Pearson's correlation coefficient of signal in the same features. d, Scatterplot showing correlation of scCUT&Tag signal among clusters and technical replicates. e, Stacked barplot showing relative proportions of cell types identified using scCUT&Tag data. f,g, Metagene plot showing the distribution of ChIP-seq for mESCs (f) and 3T3 cells (g) and downscaled scCUT&Tag signal around peaks that were called from the bulk dataset.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Quality control of the scCUT&Tag data. a, Merged pseudobulk profiles of scCUT&Tag with four antibodies against modified histones. b, Scatterplot of number of reads per cells (x axis) and fraction of reads originating from peak regions (y axis) that was used to set cutoffs for cells identification. Cutoffs were set after manual inspection of the plots and are depicted as horizontal and vertical lines overlaid over the plot. c, Histogram of number of features identified per cell for each antibody used in scCUT&Tag. d, Violin plots showing fraction of reads per cell that overlap peak regions that were called on merged bulk profiles using the same parameters for all compared samples. scCUT&Tag peak calling parameters are different from parameters used in Figure 1d. Point specifies mean of the distribution and lines standard error of mean. Number of cells per sample – H3K27me3_N1 3304, H3K27me3_N2 3090, H3K27me3_N3 5145, H3K27me3_N4 2393, H3K27me3_cell_lines_1 4872, H3K27me3_cell_lines_2 3873, K562_H3K4me2_iCell8 807, K562_H3K27me3_iCell8 1387, H1_H3K27me3_iCell8 486, Grosselin_1 2005, Grosselin_2 4122, Grosselin_3 960. e, Violin plot showing number of unique reads per cell. Point specifies mean of the distribution and lines standard error of mean. Number of cells per sample – H3K27me3_N1 n = 3304, H3K27me3_N2 n = 3090, H3K27me3_N3 n = 5145, H3K27me3_N4 n = 2393, H3K27me3_cell_lines_1 n = 4872, H3K27me3_cell_lines_2 n = 3873, K562_H3K4me2_iCell8 n = 807, K562_H3K27me3_iCell8 n = 1387, H1_H3K27me3_iCell8 n = 486, Grosselin_1 n = 2005, Grosselin_2 n = 4122, Grosselin_3 n = 960. f, Barplot depicting number of analyzed cells per experiment. g, Fingerprint plot representing relationship between cumulative signal of scCUT&Tag and scChIP–seq relative to fraction of genomic bins analyzed.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | De novo identification of cell types by cell type specific marker regions. Projection of gene activity scores of (a) H3K27ac and (b) H3K36me3 scCUT&Tag on the two-dimensional UMAP embedding. Gene name is depicted in the title and specific population is highlighted in the UMAP plot by labeling the cell type. c-d, Heatmap representation of the scCUT&Tag signal for (c) H3K27ac and (d) H3K36me3. X axis represents genomic region, each row in Y axis contains data from one cell. Cell correspondence to clusters is depicted by color bar on the right side of the heatmap and annotated with cell type. Signal is aggregated per 250 bp windows and binarized. e-h, Aggregated pseudobulk scCUT&Tag profiles for four histone modifications in all identified cell types around selected marker genes.
Extended Data Fig. 5 | Summary of meta features of cells analyzed by scCUT&Tag. 

**a.** Two-dimensional UMAP embedding of the scCUT&Tag data. Cells are colored by correspondence to GFP population, developmental age and biological replicate. **b-c.** Bar plot summary of the correspondence to the (b) GFP population and (c) developmental age per cell type identified from the H3K4me3 scCUT&Tag data.
Extended Data Fig. 6 | Comparison of merged profiles of populations identified from scCUT&Tag data and corresponding bulk ChIP–seq or bulk Cut&Run data. 

**a**. Genome browser view of a representative region harboring microglia-specific and neuron-specific H3K27me3 peak regions (highlighted in gray).

**b**. PCA analysis and Pearson correlation matrix of merged scCUT&Tag profiles per cluster and bulk ChIP–seq and bulk CUT&Run data. PCA was performed on top 150 most variable marker regions selected from scCUT&Tag data. Heatmap shows Pearson’s correlation coefficient of signal in the same features.

**c**. Relative cell type proportions identified from scCUT&Tag data and scRNA-seq data from biological replicates.
Extended Data Fig. 7 | Metagene analysis of gene activity scores. a-d. Metagene activity projection of scCUT&Tag data on the UMAP embeddings of four histone modification scCUT&Tag datasets. Metagenes are selected as top 100 most specifically expressed in the scRNA-seq data.
Extended Data Fig. 8 | Gene Ontology analysis of H3K4me3 scCUT&Tag marker genes. a. Gene ontology analysis of the marker genes determined by gene activity scores from the H3K4me3 scCUT&Tag data. GO terms were manually selected from the list of all enriched GO terms in all populations.
Extended Data Fig. 9 | scCUT&Tag of transcription factors. a, Meta-region activity scores of marker regions determined from H3K4me3 scCUT&Tag data and specific for the respective population for (a) OLIG2 and (b) RAD21 scCUT&Tag data. c-d, Boxplot representation of a and b, single cell meta-region profiles aggregated per cell type. Lower and upper bound of boxplot specify 25th and 75th percentile and lower and upper whisker specifies minimum and maximum no further than 1.5 times of interquartile range. Outliers are not displayed. non_oligo cells n=2877, oligo cells n=1667. e-f, Co-embedding of (e) H3K27ac and OLIG2 and (f) H3K27ac and RAD21 in single two-dimensional UMAP space. g, Additional motifs identified using MEME from the merged pseudobulk profile of OLIG2 scCUT&Tag.
Extended Data Fig. 10 | Benchmarking of loops predicted by the ABC model with scCUT&Tag data. a, Bar plot depicting fraction of loops predicted by the ABC model with scCUT&Tag data that overlap with loops predicted by ABC model with bulk CUT&Tag data. b, Venn diagram showing the overlap of loops predicted with scCUT&Tag data with loops predicted with ABC model with bulk CUT&Run data and Cicero. c, Scatterplot showing consistency of predictions of ABC model run with downscaled scCUT&Tag data. d, Boxplot representation of lengths of the loops predicted by various methods. Lower and upper bound of boxplot specify 25th and 75th percentile and lower and upper whisker specifies minimum and maximum no further than 1.5 times of interquartile range. Outliers are not displayed. mOro n = 1796, Astrocytes n = 1506, OEC n = 913, Unknown n = 160.
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 our Editorial Policies and the Editorial Policy Checklist.

Statistics

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

- 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

For further information and guidance on statistical practices in Nature Research journals, please see our statistical guidance pages.

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**

Data was collected using standard Illumina sequencing with standard software. BD FACSD Software version 1.2.0.142 and BD FACSDiva 8.0.2 were used for FACS sorting selection of GFP+ and GFP- populations.

**Data analysis**

Data was analyzed using combination of published tools and custom scripts. All code needed to generate the figures is published at [https://github.com/Castelo-Branco-lab/scCut-Tag_2020](https://github.com/Castelo-Branco-lab/scCut-Tag_2020)

List of Software used in the analysis:

- snakemake v5.31.1
- maccs2 v2.2.7.1
- meme v5.0.2
- samtools v1.9
- picard v2.22.0
- de stoools v3.4.4.2
- bedtools v2.29.2
- pygenometracks v3.6
- ABC model v0.2.1

R packages:

- Seurat v3.2.3
- Signac v1.1.0
- ggplot2 v3.3.3
For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature 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

The data generated in this publication are available at GEO accession GSE163532: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163532

Publicly available datasets used in this study:
- GSE75330 - scRNAseq of oligodendrocyte heterogeneity
- SRP135960, https://storage.googleapis.com/linnarsson-lab-loom/i5_allloom - mouse brain atlas scRNA-seq
- https://support.10xgenomics.com/single-cell-atac/datasets/1.2.0/atac_v1_adult_brain_fresh_5k - Fresh P50 mouse brain scATAC-seq
- GSE96107 – HiC of mESC and H3K27me3 ChIP-seq of cortical neurons. GSE124557 – scCUT&Tag in iCell8 platform. GSE117309 – scChIP-seq. GSE104435 – H3K27me3 ChIP-seq of microglia, SRP135960 – mouse brain atlas (https://storage.googleapis.com/linnarsson-lab-loom/i5_allloom), GSE75330 – scRNA-seq of oligodendrocyte lineage, 10x genomics single-cell ATAC-seq of P50 mouse cortex - https://support.10xgenomics.com/single-cell-atac/datasets/1.2.0/atac_v1_adult_brain_fresh_5k, ENCODE H3K27me3 ChIP-seq of mouse embryonic stem cells Bruce4 https://www.encodeproject.org/experiments/ENC00000CFN/, GSM1246686 - H3K27me3 ChIP-seq of mouse NIH/3T3 cells

**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

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

**Life sciences study design**

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

**Sample size**

No sample size calculation was performed to determine sample sizes. The sample sizes were in the same range of other studies in the literature. The scC&T experiments were performed in 2-4 replicates (exception is Rad21 which was performed in one replicate), and integrated in a single analysis. HiChIP was performed in three biological replicates.

**Data exclusions**

For scC&T individual thresholds were set for number of reads, number of unique reads, percentage of reads in peaks, and number of reads in blacklist regions per cell.

**Replication**

We assessed the replication by examination of intermingling of cells originating from different experiments in clusters and in low-dimensional UMAP space. HiChIP replication was assessed by manual inspection of matrices in Juicebox and subsequently the matrices were merged and used in the downstream analysis.

**Randomization**

Randomization was not applied. Animals from multiple litters over a course of approximately 6 months were used for the experiments.

**Blinding**

Blinding was not performed in this study. The clustering of the scC&T data was unsupervised.

**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.
### Antibodies

**Antibodies used**

- H3K4me3 primary Diagenode C15410030 (dilution 1:50)
- H3K27ac Primary Abcam Ab177178 (dilution 1:50)
- H3K27me3 Primary Cell signaling 9733T (dilution 1:50)
- H3K36me3 Primary Abcam Ab9050 (dilution 1:50)
- Rad21 Primary GeneTex GTX106012 (dilution 1:50)
- Olig2 primary Novus Biologicals NBP1-28667 (dilution 1:50)
- Guinea pig anti-rabbit Secondary Novus Biologicals NBP1-72763 (dilution 1:50)

**Validation**

All antibodies used in this study have been validated and tested by the provider company and/or have been cited by other authors. References are available on the web page of the provider company. In addition, all Cut&Run antibodies have been validated by the provider company to work for ChIP-seq applications and the signal of the peaks were consistent with expected enrichment (promoters/enhancers/polycomb repressed regions) and with previous experiments.

- H3K4me3 Diagenode C15410030: https://www.diagenode.com/en/p/h3k4me3-polyclonal-antibody-classic-50-ug-25-ul

Species reactivity: Human, mouse, Arabidopsis: positive. Other species: not tested.

Applications

- Suggested dilution: References
  - ChIP/ChIP-seq: *1 μg/ChIP Fig 1, 2
  - Dot Blotting 1:2,000 Fig 3
  - Western Blotting 1:500 Fig 4
  - Immunofluorescence 1:100 Fig 5

**H3K27ac Abcam Ab177178**

https://www.abcam.com/histone-h3-acetyl-k27-antibody-ep16602-chip-grade-ab177178.html

Reacts with: Mouse, Rat, Human

Suitable for: Flow Cyt, ICC, PepArr, IHC-P, WB, ChIP, ChIP-sequencing

Flow Cyt: 1/1500.

ICC: 1/2000.

PepArr: Use at an assay dependent concentration.

IHC-P: 1/1500 – 1/10000. Perform heat mediated antigen retrieval with Tris/EDTA buffer pH 9.0 before commencing with IHC staining protocol.

WB: 1/10000 – 1/100000. Detects a band of approximately 15 kDa (predicted molecular weight: 15 kDa).

ChIP: Use 2 μg for 25 μg of chromatin.

Use GAPDH ChIP primer pair ab267832 as positive control.

ChIP-sequencing: Use 4 μg for 107 cells.

- **H3K27me3 Cell Signaling 9733T**
  https://www.cellsignal.com/products/primary-antibodies/tri-methyl-histone-h3-lys27-c36b11-rabbit-mab/9733

**REACTIVITY** Human Mouse Rat Monkey

Application

- Dilution: Western Blotting 1:1000
- Immunohistochemistry (Paraffin) 1:200
- Immunofluorescence (Immunochemistry) 1:1600

**H3K36me3 Abcam Ab9050**

Reacts with: Cow, Human

Predicted to work with: Mouse, Rat, Saccharomyces cerevisiae, Xenopus laevis, Arabidopsis thaliana, Caenorhabditis elegans,
Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)
mESCs [Simon Elsässer, Karolinska Institutet], NIH-3T3 cells [Oscar Fernandez Capetillo, Karolinska Institutet]

Authentication

Cell lines were not specifically authenticated but were used in genomic studies that were consistent with their identity. This study also compares H3K27me3 data with previously published and ENCODE datasets and found them to be consistent.

Mycoplasma contamination

Cell lines were regularly checked for mycoplasma contaminations and were found negative.

Commonly misidentified lines

(See ICCLABS register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Animals and other organisms

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

Laboratory animals

The mouse line used in this study was generated by crossing Sox10-Cre animals? (The Jackson Laboratory mouse strain D25807) on a C57BL/6J genetic background with RCE-loxp [EGFP] animals? (The Jackson Laboratory mouse strain 32037-JAk) on a C57BL/6xC3H mixed genetic background. Females with a hemizygous Cre allele were mated with males lacking the Cre allele, while the reporter allele was kept in hemizygosity or homozygosity in both females and males. In the resulting Sox10-Cre-Rce-LoxP [EGFP] animals the entire OL lineage was labeled with EGFP. Breeding with males containing a hemizygous Cre allele in combination with the reporter allele to non-Cre carrier females resulted in offspring where all cells were labeled with EGFP and was therefore avoided. All animals were free from the most common mouse viral pathogens, ectoparasites, endoparasites, and mouse bacteria pathogens harbored in research animals. The battery of screened infective agents met the standard health profile established in Karolinska Institutet animal housing facilities. Mice were kept with the following light/dark cycle: dawn 6:00-7:00, daylight 7:00-18:00, dusk 18:00-19:00, night 19:00-6:00 and housed to a maximum number of 5 per cage in individually ventilated cages (IVC sealsafe GMS500, Tecniplast). Cages contained hardwood bedding [Tapvei, Estonia], nesting material, shredded paper, gnawing sticks and card box shelter [Scangold]. The mice received regular chew diet (either R70 diet or R34, Lantmännens Lantbruk, Sweden). General housing parameters such as relative humidity, temperature, and ventilation follow the European convention for the protection of vertebrate animals used for experimental and other scientific purposes treaty ETS 123. Briefly, consistent relative air humidity of 50%, 22 °C and the air quality is controlled with the use of stand-alone air handling units supplemented with HEPA filtrated air. Monitoring of husbandry parameters is done using Scanclime (Scanbur) units. Water was provided by using a water bottle, which was changed weekly. Cages were changed every other week. All cage changes were done in a laminar air-flow cabinet. Facility personnel wore dedicated scrubs, socks...
and shoes. Respiratory masks were used when working outside of the laminar air-flow cabinet. For primary cell culture, animals of both sexes were sacrificed at P4-P6. scC&T experiments were performed on P15 and P21-P25 animals of both sexes.

Wild animals
No wild animals were used in this study.

Field-collected samples
No field-collected samples were used in this study.

Ethics oversight
All experimental procedures on animals were performed following the European directive 2010/63/EU, local Swedish directive L150/SJFvS/2019, Sankt L150 and Karolinska Institutet complementary guidelines for procurement and use of laboratory animals, Dnr. 1937/03-640. The procedures described were approved by the local committee for ethical experiments on laboratory animals in Sweden (Stockholms Norra Djurförsöksnämnd), lic. nr. 131/15 and 144/16.

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

ChIP-seq

Data deposition
- Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.
IGV server URL: https://raw.githubusercontent.com/mardzix/IGV_track_server/master/registry/IGV_registry.txt
Paste the URL into IGV -> View -> Preferences -> Advanced -> Data registry URL
Data can be browsed at https://ki.se/en/mmb/oligointernode and and https://mouse-brain-cutandtag.cells.ucsc.edu

Files in database submission
The data generated in this publication are available at GEO accession GSE163532:
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163532

Genome browser session
(eg UCSC)
IGV data server link is provided. Data can also be browsed at https://ki.se/en/mmb/oligointernode and and https://mouse-brain-cutandtag.cells.ucsc.edu

Methodology

Replicates
All scC&T experiments were performed in 2-4 replicates except for Rad21 scC&T that was performed in one replicate.

Sequencing depth
Libraries were sequenced for target depth at least 20,000 reads per cell.

Antibodies
H3K4me3 primary Diagenode C15410030
H3K27ac Primary Abcam Ab177178
H3K27me3 Primary Cell signaling 9733T
H3K36me3 Primary Abcam Ab9050
Rad21 Primary GeneTex GTX106012
Olig2 primary Novus Biologicals NBP1-28667
Guinea pig anti-rabbit Secondary Novus Biologicals NBP1-72763
(dilution 1:50 to all antibodies)

Peak calling parameters
Peaks specific for all subpopulations were not called. Merged data peaks were used to construct peaks x cell matrix that is provided in the GEO database.

Data quality
The data quality was checked manually by inspection of histone modifications / TF profiles in genome browser. Briefly, H3K4me3 and H3K27me3 were mostly enriched at promotor regions. H3K27me3 was enriched at ubiquitously repressed hox genes. H3K27ac was enriched at promoters and intergenic regions (likely enhancers) and H3K36me3 was enriched in gene body regions with higher enrichment in exons and towards the 3’. Cut&TG has generally low signal/noise ratio than ChIP-seq, so the majority of peaks had fold enrichment >5.

Software
https://github.com/Castello-Branco-lab/scCut-Tag_2020
Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- 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).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation: Brain was extracted from P15 and P21-P25 mouse subjects, dissociated into single cell suspension using Neural Tissue Dissociation Kit (P) (Miltenyi Biotec). Cells were stained for viability by 1:5000 DAPI diluted in 1xPBS before filtering. Cells were filtered through 50 um strainer immediately before FACS and resuspended in 1xPBS supplemented with 1% BSA and 2mM EDTA and kept at 4 degrees celsius.

Instrument: Facs Aria III, or Facs Aria Fusion (BD Biosciences)

Software: BD FACSDiva 8.0.2 and BD FACS Software version 1.2.0.142

Cell population abundance: GFP is the reporter for Sox30 which is a transcription factor expressed in the oligodendrocyte lineage in the CNS. scC&T was performed on the sorted populations that were GFP+ and/or GFP-. Cca 5-8% cells were GFP+, out of parent population.

Gating strategy: The first gate is for selecting for cells in a non-restrictive manner. The second gate is selecting for singlets. The DAPI+ gate is selecting the live cells, there is a clear separation between DAPI+ and DAPI-. The GFP+ gate is selecting for the GFP+ cells, with again a clear separation between the GFP+ gate and GFP+ gate.

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