

scChIX-seq infers dynamic relationships between histone modifications in single cells

Received: 27 April 2021

Accepted: 12 October 2022

Published online: 02 January 2023

 Check for updates

Jake Yeung^{1,2,3}✉, Maria Florescu^{1,3}, Peter Zeller^{1,3},
Buys Anton de Barbanson¹, Max D. Wellenstein¹ &
Alexander van Oudenaarden¹✉

Regulation of chromatin states involves the dynamic interplay between different histone modifications to control gene expression. Recent advances have enabled mapping of histone marks in single cells, but most methods are constrained to profile only one histone mark per cell. Here, we present an integrated experimental and computational framework, scChIX-seq (single-cell chromatin immunocleavage and unmixing sequencing), to map several histone marks in single cells. scChIX-seq multiplexes two histone marks together in single cells, then computationally deconvolves the signal using training data from respective histone mark profiles. This framework learns the cell-type-specific correlation structure between histone marks, and therefore does not require a priori assumptions of their genomic distributions. Using scChIX-seq, we demonstrate multimodal analysis of histone marks in single cells across a range of mark combinations. Modeling dynamics of *in vitro* macrophage differentiation enables integrated analysis of chromatin velocity. Overall, scChIX-seq unlocks systematic interrogation of the interplay between histone modifications in single cells.

Gene expression in animals relies on epigenetic marks such as histone modifications to regulate the accessibility and function of the genome in different cell types¹. Large-scale efforts characterizing different histone modifications in a variety of cell populations commonly use chromatin immunoprecipitation followed by sequencing (ChIP-seq)^{2–8}. Alternative strategies to ChIP-seq based on enzyme tethering (chromatin immunocleavage, ChIC) have reduced the background signal in profiling the epigenome⁹, and have enabled single-cell profiling of histone modifications^{8,10–19}. Tethering strategies involve incubating cells with an antibody against a histone modification of interest, which then tethers either protein A-MNase^{10,12,18,19} or protein A-Tn5^{11,13–17} fusion protein to generate targeted DNA fragments in single cells. However, most experimental techniques to map single-cell histone modifications are limited to only one histone modification per single cell.

We present an integrated experimental and computational framework for multiplexing histone modifications in single cells. To profile two histone modifications in single cells (Fig. 1a), we first generate

three genome-wide sortChIC¹⁸ datasets: two datasets by incubating cells with one of the two histone modification antibodies separately (single-incubated; Fig. 1b), and the third by incubating cells with both histone modification antibodies together (double-incubated; Fig. 1b). We then use our two single-incubated datasets as training data to generate the possible pairs of genome-wide histone modification profiles that, when added together, fit to a single-cell profile from the double-incubated dataset (Fig. 1c). For each double-incubated cell, we then deconvolve the multiplexed data by probabilistically assigning each fragment back to their respective histone modification.

scChIX-seq links single-cell maps of different histone modifications, revealing relationships between histone modifications in single cells. In these linked maps, information derived from one chromatin state, such as cell types, histone mark levels and pseudotimes, can transfer to another chromatin state (Fig. 1d), unlocking joint analysis of several histone modifications in single cells. We first validated scChIX-seq using simulation, purified blood cell types and whole bone

¹Onco Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Utrecht, the Netherlands. ²Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria. ³These authors contributed equally: Jake Yeung, Maria Florescu, Peter Zeller. ✉e-mail: jake.yeung@ist.ac.at; a.vanoudenaarden@hubrecht.eu

marrow. We then applied scChIX-seq to two complex biological systems, one in mouse organogenesis to uncover orthogonal dynamics in H3K36me3 and H3K9me3, and the other in macrophage in vitro differentiation to reveal coordinated dynamics between H3K4me1 and H3K36me3.

Results

Benchmarking across histone modification relationships

To test whether scChIX-seq is accurate for histone modification patterns that are mutually exclusive as well as highly overlapping, we apply scChIX-seq to simulated single-cell data with known amounts of overlap to benchmark our method across different overlapping patterns between histone modifications. We simulate single-cell histone modification data by modifying simATAC²⁰ to generate sparse count data from different overlapping patterns from the same cell (Fig. 1e and Extended Data Fig. 1a,b; Methods). Our simulations span three scenarios to cover varying degrees of overlapping patterns (Extended Data Fig. 1c). (1) Mutually exclusive scenario with only 1% of loci overlapping. (2) Intermediate scenario with 50% of loci overlapping. (3) Correlated scenario with 99% of loci overlapping. In these simulations, we provide a ground truth parameter p for each genomic locus and then estimate this parameter using our statistical framework to assess the uncertainty in our inferences. Here, p is the expected fraction of double-incubated reads in a locus that belongs to a reference histone modification (that is, $p = 0.5$ if locus is exactly overlapping, $p = 1$ or 0 if locus is exactly mutually exclusive). Applying scChIX-seq to each scenario, we find that the distribution of our estimates \hat{p} across all loci are comparable with the ground truth distribution of p (Extended Data Fig. 1c,d). Furthermore, scChIX-seq accurately recovers the different cell types underlying the simulated data, and links the two histone modification landscapes into a joint uniform manifold approximation and projection (UMAP) (Extended Data Fig. 1e). Summarizing the three scenarios, scChIX-seq can estimate p accurately for all degrees of overlap, with confidence intervals (CI) better than $\hat{p} \pm 0.05$ (Fig. 1e (right) and Extended Data Fig. 1f). Our simulation study confirms that scChIX-seq is accurate in inferring several histone modifications in single cells in both mutually exclusive as well as overlapping histone modification patterns.

Validating with ground truth data from purified cell types

To validate our method experimentally, we generate a ground truth sortChIC dataset by purifying three known cell types from mouse bone marrow: B cells, granulocytes and natural killer (NK) cells, using fluorescence-activated cell sorting (FACS) and applying scChIX-seq (Methods). Of note, the sortChIC method is designed to integrate FACS with histone modification mapping¹⁸, so we can enrich for a cell type and map histone modifications in one workflow. We split bone marrow cells into three technical batches: one batch incubated with anti-H3K27me3 antibody alone (single-incubated), one with anti-H3K9me3 alone (single-incubated) and the third with both anti-H3K27me3 and anti-H3K9me3 antibodies together (double-incubated, H3K27me3+H3K9me3). We then sorted cells into 384-well plates, each plate containing all three cell types, and generate targeted cut fragments (Extended Data Fig. 2a,b). We chose H3K27me3 and H3K9me3 because they have been shown to have a mutually exclusive relationship²¹, allowing us to verify whether we can infer the correct cell type as well as the generally mutually exclusive relationship. Of note, although H3K27me3 and H3K9me3 are known to be nonoverlapping, it is unclear how this relationship precisely changes to make cell-type-specific patterns at different loci, and therefore modeling the two relationships is still needed to accurately infer the two chromatin profiles in individual cells.

From the double-incubated data alone, we would not know which cut fragments correspond to H3K27me3 and which to H3K9me3, but would observe only a superposition of the two profiles. We therefore

used the single-incubated sortChIC data to train a statistical model of how cells from the same cell type combine their H3K27me3 and H3K9me3 profiles to generate double-incubated cut fragments. This model was then used to deconvolve the single-cell multiplexed signal into their respective histone modifications (Methods).

To learn an interpretable latent space for H3K27me3 and H3K9me3, we applied latent Dirichlet allocation (LDA)^{22,23} to the single-incubated H3K27me3 and H3K9me3 datasets, which factorizes count matrices based on a multinomial model (Methods). (Extended Data Fig. 2c,d). LDA learns cell-type-specific vectors of probabilities. These parameters model the probability that a cut fragment would fall into a specific genomic region. These probabilities can therefore be interpreted as genome-wide histone modification distributions that depend on cell type, and each cell generates a high-dimensional sparse count vector with n total fragments by drawing n independent trials from these multinomial distributions.

Demultiplexing the double-incubated data involves two steps. First, we used the training data to infer which genome-wide H3K27me3 distribution was added to which H3K9me3 distribution to generate a linear combination of two distributions (H3K27me3+H3K9me3). Second, we probabilistically assigned each double-incubated cut fragment to either H3K27me3 or H3K9me3, given that we know the underlying linear combination of the two profiles.

The deconvolved H3K27me3+H3K9me3 data generated two sets of cuts for each cell: one set coming from H3K27me3 and the other from H3K9me3. We projected the two sets of cuts onto the H3K27me3 or H3K9me3 latent space (learned from LDA), respectively (Fig. 2a). Since each deconvolved cell has a set of cuts in H3K27me3 and H3K9me3 simultaneously, we can link the UMAPs together, creating a joint chromatin regulation space (Fig. 2a).

The double- and single-incubated cells in the H3K27me3 and H3K9me3 UMAPs intermingle, suggesting that the model accurately assigns cut fragments to their respective histone modification (Extended Data Fig. 2e,f). Comparing the H3K27me3 deconvolved pseudobulk signal with our ground truth single-incubated pseudobulk shows high correlation for the expected cell type, and lower for the other two cell types (Extended Data Fig. 2g). The H3K9me3 deconvolved pseudobulk signal also shows highest correlation with the expected cell type, with lower correlation from other cell types (Extended Data Fig. 2h). Finally, we compared the fragments per cell obtained from scChIX-seq versus multi-CUT&TAG²⁴, and found that scChIX-seq achieves higher sensitivity than multi-CUT&TAG (Extended Data Fig. 2i). Overall, our ground truth dataset demonstrates that scChIX-seq is accurate and sensitive in assigning cut fragments to their respective histone modification.

To quantify the accuracy of scChIX-seq in selecting the correct H3K27me3-H3K9me3 cluster pair to mix together, we color each cell by its ground truth label and plot its inferred H3K27me3-H3K9me3 pair on a two-dimensional (2D) grid (Fig. 2b, left). The false discovery rates (FDRs) of scChIX-seq predicting B cells, granulocytes or NK cells are 10%, 3% and 1%, respectively (Fig. 2b, right). Similarly, scChIX-seq has high specificity and sensitivity in inferring the correct cluster pairs (Fig. 2b, right).

Next, scChIX-seq assigns each double-incubated cut fragment to either H3K27me3 or H3K9me3 (Fig. 2c; Methods). The deconvolved B cell repressive landscapes correspond with their respective ground truth, exemplified in the *Bcl2* (Fig. 2d) and *Crim1* (Extended Data Fig. 3a) locus. We also find cell-type-specific signal in H3K27me3 (Extended Data Fig. 3b) and H3K9me3 signal (Extended Data Fig. 3c).

Our model infers p , the expected fraction of double-incubated fragments at a locus that belongs to H3K27me3. That is, $p = 0$ if all fragments belong to H3K9me3 and $p = 1$ if they all belong to H3K27me3. Plotting these probabilities across all loci reveals a bimodal distribution with peaks near 0 and 1 (Extended Data Fig. 3d). Classifying these loci as H3K9me3-specific ($P < 0.5$) or H3K27me3-specific ($P \geq 0.5$),

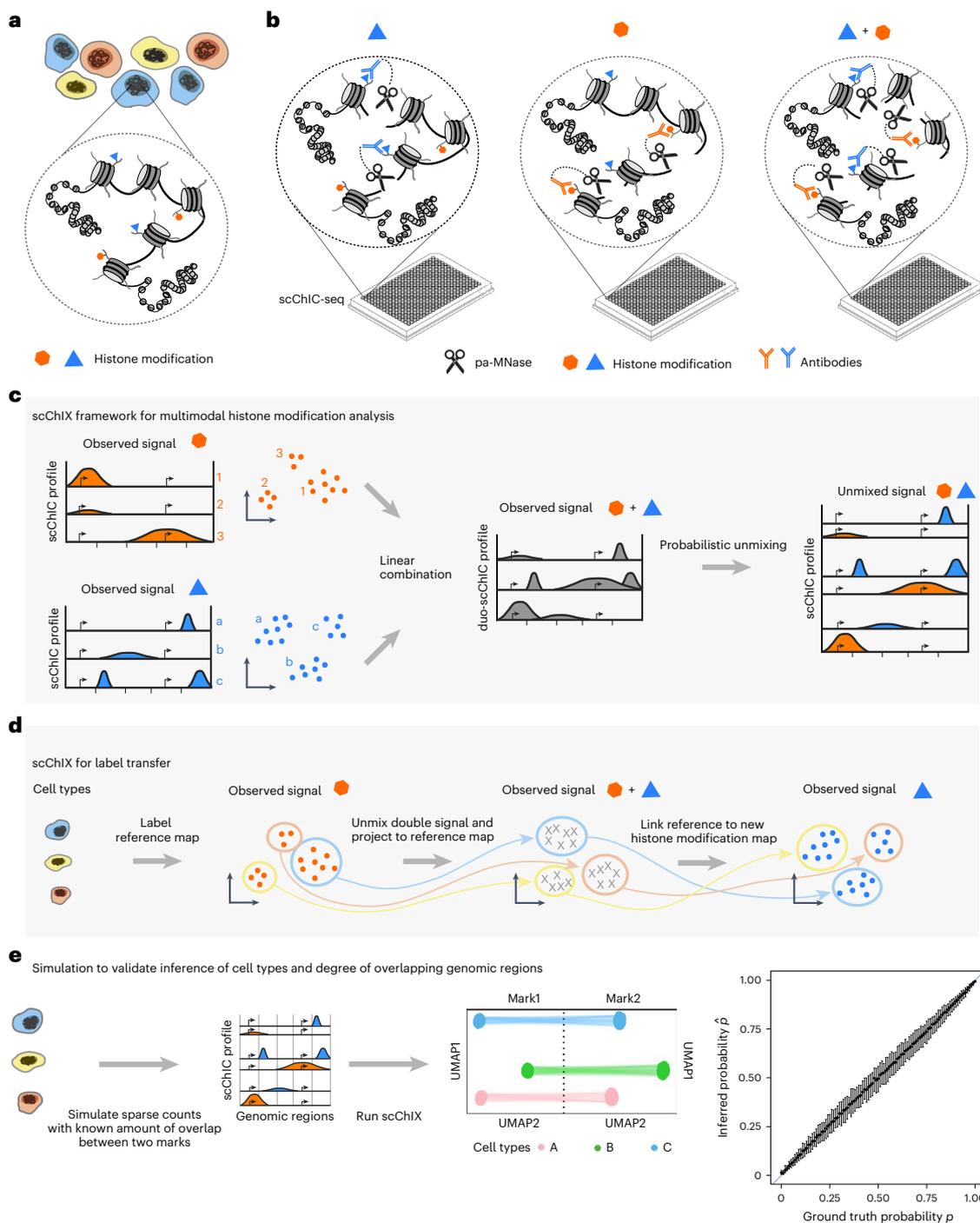


Fig. 1 | Overview of the scChIX-seq method. **a**, Chromatin regulation of different cell types (different colored cells) is regulated in part through several histone modifications (two histone modifications shown as an example). **b**, scChIX-seq uses three sortChIC antibody incubation conditions: two conditions each target a single histone modification (single-incubated) only and the third condition targets both histone modifications simultaneously (double-incubated). **c**, Schematic of scChIX-seq for deconvolving multiplexed histone modifications. The two single-incubated sortChIC datasets (one targeting an orange histone modification, the other a blue modification, each modification reveals three clusters) are training data to define the possible pairs of histone modification distributions that can be combined to generate a hypothetical double-incubated cell. For each observed double-incubated cell, we then assign the cell to the most probable pair of cell states, one from each histone modification. We then probabilistically assign each pa-MNase cut into their respective histone modification. Cartoons represent genome-wide distribution of histone

modification signals in different modifications and cell types; axes represent genomic distance, and vertical ticks are arbitrary distance markers. **d**, Label transfer allows joint analysis of two single-incubated sortChIC datasets targeting functionally distinct histone modifications. Information derived from one histone modification, such as cell types, histone mark levels and pseudotime, can be transferred to another histone modification using the double-incubated cells as a link. **e**, Simulation study shows that scChIX-seq can unbiasedly assign reads to each mark regardless of the amount of overlap there is between the two marks across the genome. *x* axis of cartoon genome-wide distributions (middle-left) is genomic distance. Right: ground truth probabilities versus inferred probabilities from scChIX. p is the expected fraction of double-incubated reads in a genomic locus that belongs to mark 1. \hat{p} is the estimate of the probability; $n = 101$ simulation datapoints spread evenly between 0 and 1 inclusive. Error bars are 95% CI, centers are the mean.

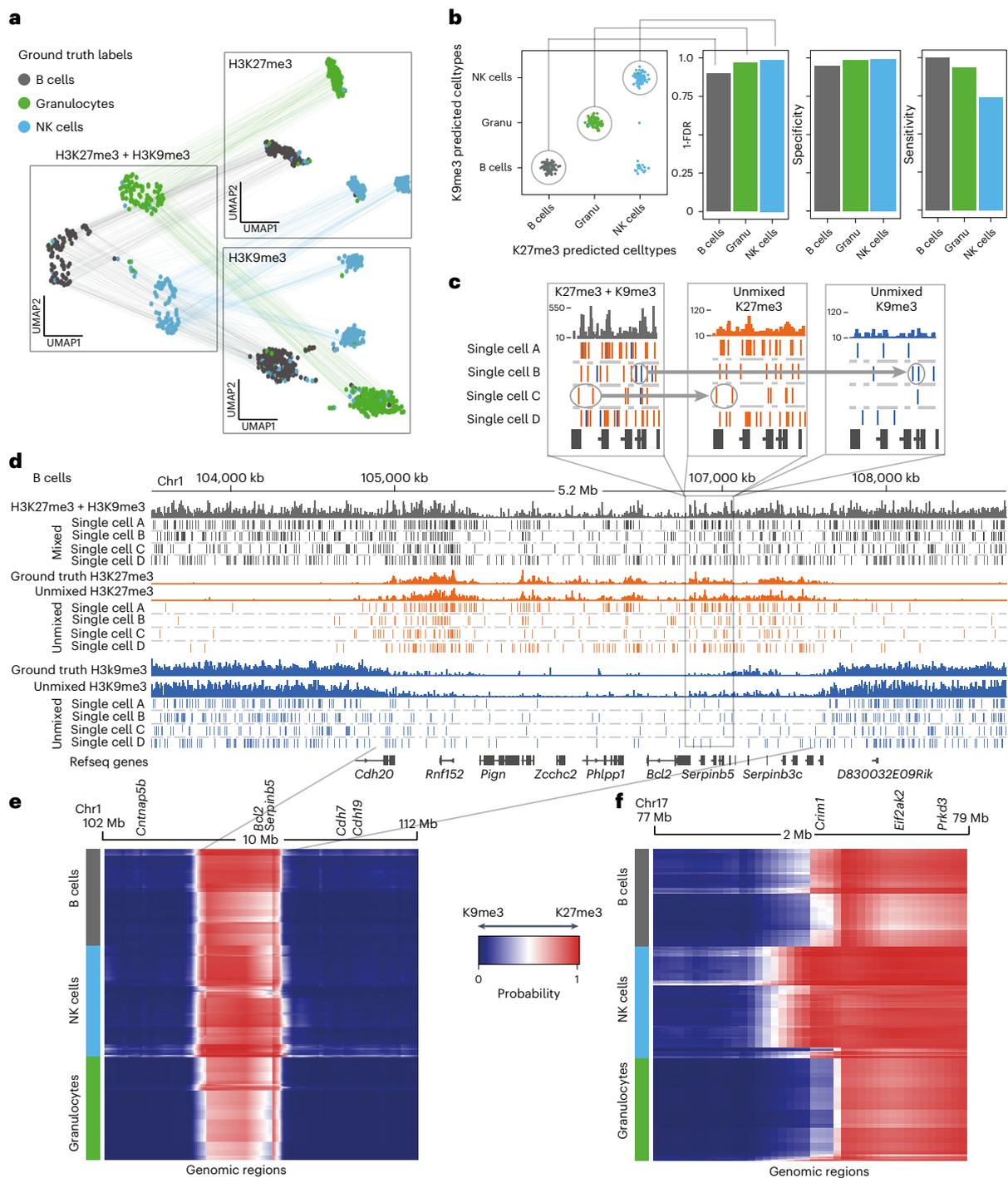


Fig. 2 | scChIX-seq accurately deconvolves multiplexed histone modifications in single cells. **a**, UMAP representation of the H3K27me3 ($n = 367$) and H3K9me3 ($n = 376$) histone modification space derived from the two single-incubated datasets (right two panels), and the H3K27me3+H3K9me3 space (left panel, $n = 290$) derived from the double-incubated data. Cells are colored by their ground truth cell-type labels. The cells in the H3K27me3- and H3K9me3-only space have unmixed double-incubated cells whose deconvolved signal has been projected onto their respective UMAPs. Lines connecting across datasets connect where each double-incubated cell is located in each of the three histone modification space. **b**, Matrix summarizing the cluster pair that scChIX-seq selected for each double-incubated cell. Cells along the diagonal are predicted to be B cells, granulocytes and NK cells, respectively. Cells in the off-diagonal are false negatives. Barplots summarizing FDR, sensitivity and specificity of assigning each cell type (right). **c**, Zoom-in coverage plot and single-cell cut fragments in B cells of mixed (H3K27me3+H3K9me3, gray bars), unmixed

(H3K27me3 and H3K9me3, orange and blue bars). Positions of cut fragments are shown for four single cells (single cells A, B, C and D) for H3K27me3+H3K9me3 signal (gray ticks) as well as their unmixed outputs (orange and blue ticks). Circled reads and arrow highlight examples of cut fragments being assigned to either H3K27me3 (orange) or H3K9me3 (blue). **d**, Zoom-out of the *Serpinb5* locus. Cut fragments from H3K27me3+H3K9me3 are colored based on whether they have been assigned to H3K27me3 (orange) or H3K9me3 (blue). Ground truth coverage are single-incubated sortChIC data targeting H3K27me3 (orange) and H3K9me3 (blue). **e**, Heatmap of probabilities p of assigning reads to H3K27me3 ($p = 1$, red) or H3K9me3 ($p = 0$, blue) around the *Bcl2* locus. Rows are single cells (ordered by predicted cell type), columns are genomic regions (50 kb bins). Transitions between H3K9me3- and H3K27me3-marked chromatin states are independent of cell type. **f**, Same as **e** but at the *Crim1* locus, where transitions from H3K9me3 to H3K27me3 (blue to red) are cell-type specific.

we compare the GC content and distance to transcription start site (TSS) of the two classes of loci (Extended Data Fig. 3e,f). We find H3K9me3-specific regions to have lower GC content and increased distance from TSSs compared with H3K27me3-specific regions. Of note, we observe this difference across all three cell types, suggesting that GC-poor and gene-poor regions of the genome is a general feature of H3K9me3-specific regions²¹.

Summarizing these probabilities in single cells along the genome as a heatmap, the *Bcl2* locus reveals the mutual exclusive relationship between H3K27me3 and H3K9me3, where the chromatin state is predominantly H3K9me3, then switches to H3K27me3, and then switches back to H3K9me3 (Fig. 2e). For *Bcl2*, these transitions occur at the same location independent of the cell type. However, we also find that these transitions can be cell-type specific, as exemplified by the *Crim1* locus (Fig. 2f), where the H3K27me3 region extends further upstream of *Crim1* in NK cells compared with B cells and granulocytes. Our ground truth experiment demonstrates that scChIX-seq can accurately map two histone modifications in single cells, and the inferred probabilities can be biologically interpreted as relationships between the two histone modifications in single cells.

scChIX-seq reveals H3K4me1/H3K27me3 relationships in bone marrow

We next apply scChIX-seq to integrate active (H3K4me1) and repressive (H3K27me3) chromatin states in a complex mixture of cells by sampling mouse bone marrow (Extended Data Fig. 4a,b). We use scChIX-seq to transfer labels and link UMAPs between active and repressive histone modifications (Fig. 3a,b) to perform a joint analysis of the two marks.

To define cell types from the H3K4me1 sortChIC data, we ranked the top 150 genes associated with different clusters from sortChIC and used a publicly available scRNA-seq dataset to compare mRNA abundances of cluster-specific genes across different blood cell types²⁵ (Extended Data Fig. 4c). scChIX-seq takes each H3K4me1+H3K27me3 cell and infers the most probable cluster pair (one from H3K4me1, the other from H3K27me3), which systematically transfers cell-type labels defined from H3K4me1 onto the H3K27me3 data (Extended Data Fig. 4d). We find that a small minority of double-incubated cells have low-confidence cluster pair predictions. Plotting the cluster pairs onto the H3K4me1+H3K27me3 UMAP confirms that the single-cell assignment produces precise clusters where neighboring cells are probably assigned to the same pair. Low-confidence predictions arise from cells that border between clusters (Extended Data Fig. 4e), which we remove from further analysis. Overall, scChIX-seq allows systematic transfer of cell-type labels from one histone modification to another.

We next deconvolve the double-incubated cells into their respective histone modification. The UMAPs from H3K4me1 and H3K27me3 show that single-incubated and deconvolved single cells intermingle, suggesting that deconvolution does not produce batch effects (Extended Data Fig. 4f,g). The deconvolved single cells provide anchors to systematically link one histone modification with another (Fig. 3c). To validate the predicted cell types in both the single and deconvolved datasets, we compared with data from cell types purified by FACS. For H3K4me1 clusters, we compared with publicly available ChIP-seq⁵. Pearson correlation between ChIP-seq of B cells, erythroids, granulocytes and NK cells versus sortChIC from single- and double-incubated cells is highest for the predicted cell type (Extended Data Fig. 5a–d). Although single-incubated cells have higher correlation with ChIP-seq reference data than deconvolved cells for the matched cell type, the deconvolved cells of the matched cell type consistently had higher correlation with ChIP-seq than unmatched cell types. For H3K27me3 clusters, we used our ground truth sortChIC data purified from FACS. Pearson correlation of sortChIC signal between FACS-sorted B cells, granulocytes and NK cells versus pseudobulks derived from whole bone marrow is highest for the predicted cell type (Extended Data Fig. 5e–g).

Classifying these loci as H3K27me3-specific or H3K4me1-specific using a cluster-specific cutoff for *p* (Extended Data Fig. 5h), we again compare the GC content and distance to TSS of the two classes of loci. We find that H3K4me1-marked regions tend to be closer to TSSs compared with H3K27me3 (Extended Data Fig. 5i), and that GC content is higher in H3K27me3-specific compared with H3K4me1-specific regions (Extended Data Fig. 5j). The increase in GC content for H3K27me3-marked regions is consistent with previous studies showing that GC-rich elements in transcriptionally inactive regions can recruit PRC2 (ref. 26).

We use the joint landscape to reveal active and repressive histone modification dynamics within cell types. To find differences in chromatin regulation between pro-B cells versus B cells, we select only pro-B or B cells and recluster the cells in both H3K4me1 and H3K27me3 separately (Extended Data Fig. 6a,b). With multimodal data, we can transfer cell-type-specific H3K4me1 signal onto the H3K27me3 UMAP to distinguish pro-B and B cells with more confidence. Using pro-B cell-specific genes, *Pax5* (ref. 27) and *Pten*²⁸, we project the H3K4me1 signal at loci overlapping these genes onto both H3K4me1 and H3K27me3 landscapes, confirming a subset of pro-B cells within the B cell population (Extended Data Fig. 6c). Similarly, we use marker genes associated with more differentiated B cells, such as *Irf4* (ref. 27), *Igkv3-2* locus²⁹ and *Cd72* (ref. 30) to confirm a more differentiated B cell population (Extended Data Fig. 6d). Plotting the heatmap of H3K4me1-H3K27me3 assignment probabilities at the *Igk* locus reveals that the chromatin state is repressed in pro-B cells but becomes activated in B cells (Fig. 3d), consistent with the progressive activation of the chromatin state during B cell development²⁹.

Next, we recluster neutrophils to analyze differences in chromatin regulation along pseudotime (Extended Data Fig. 7a). Reclustering neutrophils in H3K27me3 reveals a shared pseudotime trajectory that varies smoothly between neutrophils in both the H3K27me3 and H3K4me1 landscapes. H3K4me1 levels at the *Retnlg* locus—a marker gene for mature neutrophils³¹—increases along pseudotime, while H3K27me3 levels decreases (Extended Data Fig. 7b). The H3K27me3 gene loadings associated with pseudotime consists of a module of *Hox* and other developmental genes (Extended Data Fig. 7c–e). Of note, these genes have low levels of mRNA abundances in neutrophils (Extended Data Fig. 7f), suggesting that this module is transcriptionally silent. At a locus overlapping the *Hoxa* locus, we find that H3K27me3 was highly marked while H3K4me1 was lowly marked across all neutrophils. Along pseudotime, H3K27me3 increases further, while H3K4me1 decreases further (Extended Data Fig. 7c). Our pseudotime analysis suggests that dynamics in histone modifications can occur even in regions associated with low-expressed genes.

H3K36me3/H3K9me3 relationships during mouse organogenesis

To demonstrate the method in more complex biological scenarios, we applied scChIX-seq during mouse organogenesis (E9.5 to E11.5) to study H3K36me3 and H3K9me3 dynamics at single-cell resolution (Fig. 4a and Extended Data Fig. 8a,b). We took the top 250 cluster-specific bins from the H3K36me3 data to identify cell types (Methods). These loci associate with gene bodies of cell-type-specific genes. For example, we find H3K36me3 signal around genes enriched in specific cell types, such as erythroids (*Sptb*)³², white blood cells (*Lcp2* (ref. 33)), endothelial cells (*Emcn*)³⁴, neural tube (*Rfx4*)³⁵, neurons (*Elavl4*)³⁶, Schwann precursors (*Cdh6*)³⁷, epithelial cells (*Grhl2*)³⁸, mesenchymal progenitors (*Prx1*)³⁹ and cardiomyocytes (*Gata6*, *Tpm1*)^{40,41} (Extended Data Fig. 8c–l).

To uncover whether distinct H3K36me3 cell types could share common H3K9me3 landscapes, we deconvolved the H3K36me3 + H3K9me3 cells and projected each cell to both landscapes (Fig. 4b). scChIX-seq reveals that erythroid and white blood cells have both distinct active chromatin and heterochromatin, but the other nonblood cell types show similar heterochromatin distribution. Assigning each

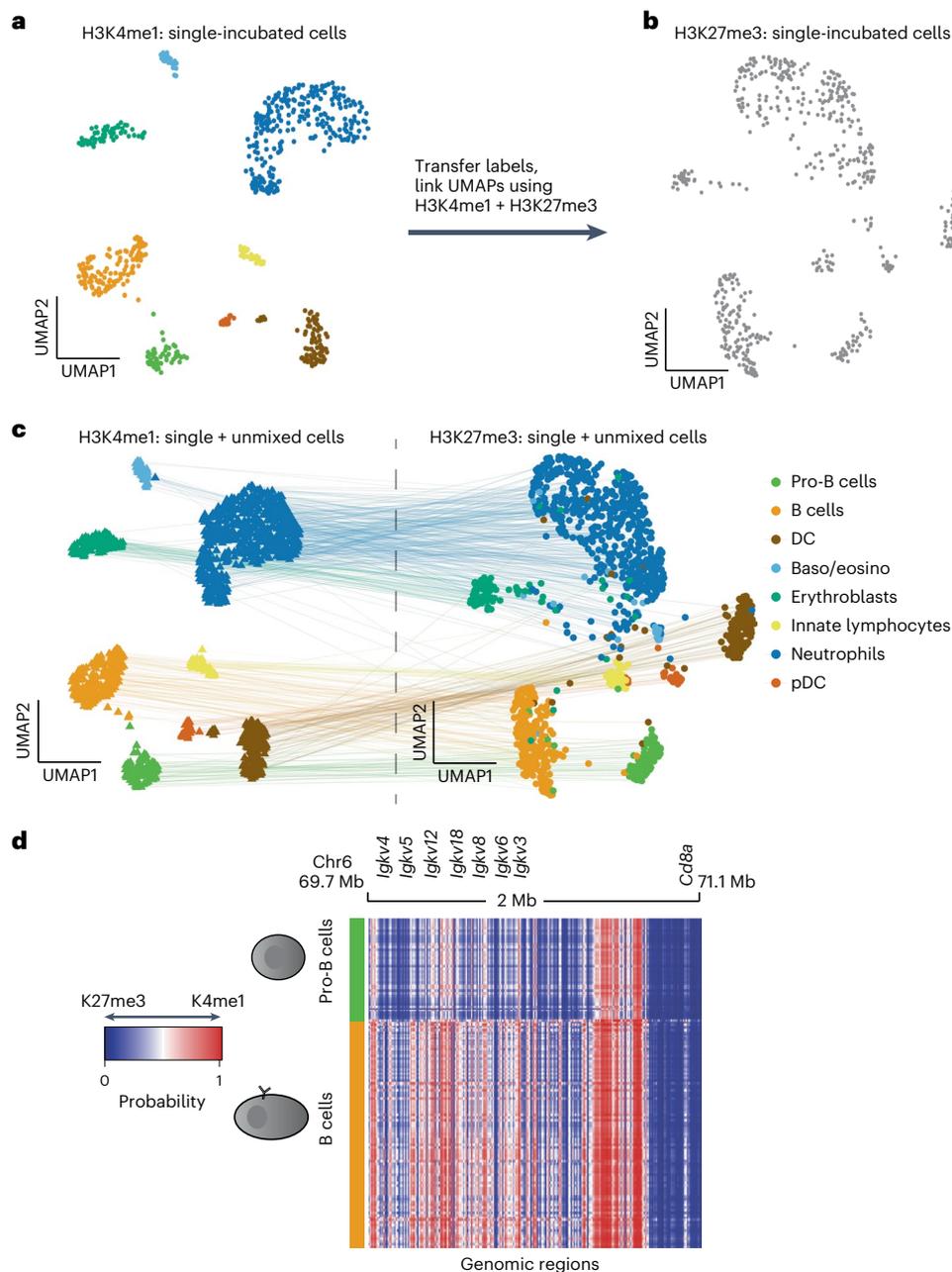


Fig. 3 | scChIX-seq enables joint analysis of distinct histone modifications in single cells. **a**, UMAP of sortChIC signal of H3K4me1 in bone marrow ($n = 639$ cells). Clusters are colored by cell type. Latent space calculated using LDA with 50 kb bins. **b**, UMAP of sortChIC signal of H3K27me3 in whole bone marrow ($n = 517$ cells). Cell types in H3K27me3 are inferred by transferring labels from H3K4me1. **c**, H3K4me1 and H3K27me3 UMAPs linked together by deconvolved double-incubated cells ($n = 1,711$ cells). H3K4me1 and H3K27me3 portions of the double-incubated cells are projected onto their respective UMAPs. Lines connect

where the active signal and the corresponding repressive signal are located for each double-incubated cell. DC, dendritic cells; pDC, plasmacytoid dendritic cells. **d**, Heatmap showing probability of assigning a read in a region to either H3K27me3 or H3K4me1 at 5 kb resolution. Heatmap shows the *Igk* locus for pro-B versus B cells. Rows are single cells, columns are 5 kb genomic regions. Blue represents regions where cut fragments are probably coming from H3K27me3, while red represents regions where cut fragments are probably coming from H3K4me1.

double-incubated cell to a H3K36me3 and H3K9me3 cluster confirms that cells with distinct H3K36me3 can share the same H3K9me3 cluster (Fig. 4c). Of note, the variable genes that show cell-type-specific differences in both active chromatin and publicly available mRNA abundances⁴² (Extended Data Fig. 9a,b) have low signal across cell types in H3K9me3 (Extended Data Fig. 9c), suggesting that using conventional marker genes from RNA-seq would not reveal cell-type differences in H3K9me3.

Differential expression across the three H3K9me3 clusters reveals cluster-specific repressed loci (Extended Data Fig. 9d), with

the largest effect coming from erythroid-specific regions. These erythroid-repressed regions are associated with decreased mRNA abundances (Extended Data Fig. 9e–g). Subsetting the data and running LDA on only nonblood cells in H3K9me3, we find that H3K9me3 varies over organogenesis stages (Fig. 4d), suggesting that heterochromatin differences are stronger across organogenesis stages than between cell types.

Because the double-incubated cells have cut fragments associated with both histone modifications, we hypothesized that the deconvolved data could precisely quantify the ratio between the two histone modifications, and how this ratio changes across cell types. Counting

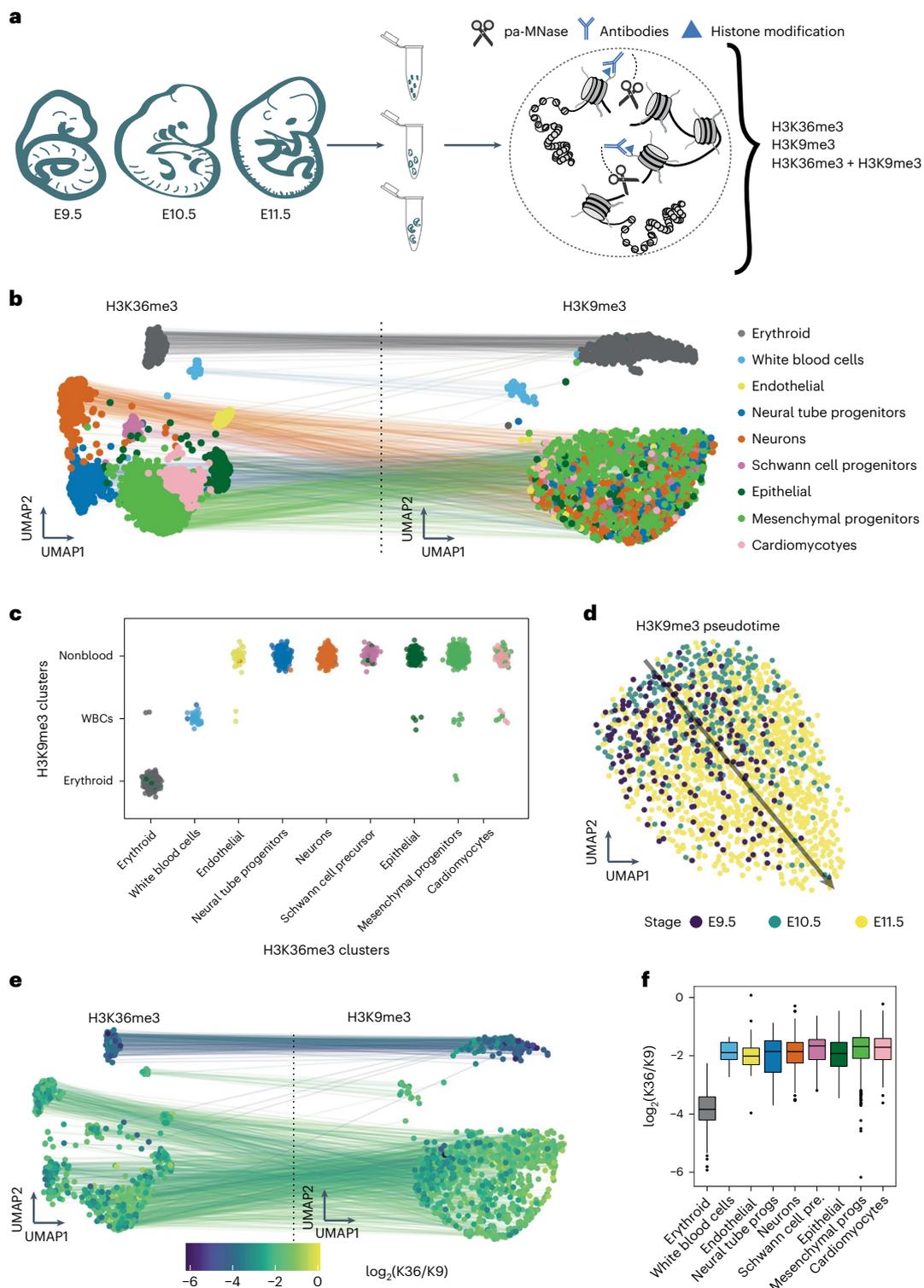


Fig. 4 | Applying scChIX-seq to mouse organogenesis reveals shared heterchromatin landscapes and cell-type-specific differences in H3K36me3:H3K9me3 ratios. a, Schematic of mouse organogenesis experiment to study H3K36me3 and H3K9me3 in single cells. **b**, Joint UMAP of mouse organogenesis after deconvolution from scChIX-seq ($n = 2,911$ H3K36me3 cells, $n = 2,166$ H3K9me3 cells). **c**, Assignment of several H3K36me3 cell types to one H3K9me3 cluster. The H3K36me3 (columns) and H3K9me3 (rows) label for each double-incubated cells ($n = 1,197$ cells) are plotted onto a matrix to H3K36me3 cell types to H3K9me3 clusters. Cells are colored by their cell-type label as in **b**. **d**, Subclustering of nonblood cells for H3K9me3, represented as a UMAP.

Arrow denotes a pseudotime axis. Pseudotime defined as the first PC of the 2D UMAP. **e**, Joint UMAP of deconvolved double-incubated cells ($n = 1,197$ cells), colored by the log ratio of number of H3K36me3 cuts versus number of H3K9me3 cuts. **f**, Boxplot of H3K36me3:H3K9me3 ratio across cell types. Number of double-incubated cells for each cell type: $n = 163$ erythroid, $n = 17$ white blood cells, $n = 24$ endothelial, $n = 136$ neural tube progenitors, $n = 197$ neurons, $n = 46$ Schwann cell precursors, $n = 73$ epithelial, $n = 458$ mesenchymal progenitors and $n = 83$ cardiomyocytes. Boxplots show 25th percentile, median and 75th percentile, with the whiskers spanning 97% of the data.

total reads from single-incubated data would lead to large cell-to-cell technical variability because counts per cell can span several orders of magnitude. However, comparing the counts of the two histone modifications in the same cell could overcome this technical variability. We therefore asked whether the global ratio of H3K36me3 and H3K9me3 in individual cells varies. Plotting the ratio of H3K36me3 and H3K9me3 reveals that most cells have comparable ratios, but that erythroid cells have lower ratios than other cell types (Fig. 4e,f). This lower ratio is consistent with mass spectrometry studies showing a global decrease in H3K36me3 but no change in H3K9me3 during erythroid maturation⁴³. Of note, inferring this global change without scChIX-seq, such as by counting total unique fragments from single-incubation data, is challenging due to the large variability in total counts across cells and the inability to distinguish cell types in certain H3K9me3 clusters (Extended Data Fig. 9h,i).

In sum, applying scChIX-seq to H3K36me3 and H3K9me3 during organogenesis reveals unique insights from multimodal analysis. The complex relationships between the two histone modifications as well as their global changes would not have been elucidated by analyzing single-incubated data alone.

Mark-specific pseudotimes and chromatin velocity

Finally, we applied scChIX-seq to study the dynamic relationships between two active histone modifications, H3K4me1 and H3K36me3, over an in vitro differentiation timecourse. We sorted blood progenitors from mouse bone marrow, added macrophage colony-stimulating factor (MCSF) and collected cells over 7 days (Fig. 5a and Extended Data Fig. 10a,b; Methods). We incubated cells with either H3K4me1, H3K36me3 or both H3K4me1 and H3K36me3, then performed scChIX-seq.

Genome tracks of H3K4me1 and H3K36me3 signal for each day shows upregulation of macrophage-specific genes, such as *Mertk*⁴⁴ (Extended Data Fig. 10c). Heatmap of H3K4me1 and H3K36me3 dynamics at gene bodies along pseudotime reveals that the two histone modifications up- and downregulate genes with different dynamics. H3K36me3 shows a gradual up- or downregulation of signal while H3K4me1 reaches a new steady state earlier along pseudotime (Fig. 5b). Summarizing log₂ fold change of the two histone modifications genome-wide, we find that dynamics in H3K36me3 are often larger than in H3K4me1 (Extended Data Fig. 10d). Comparing pseudotime progression with day of sample collection shows that changes in H3K4me1 peak at day 2 and then increases progressively over the day while H3K36me3 dynamics peak around day 3 and 4 before relaxing towards steady state (Fig. 5c). The time of the largest change in H3K4me1 dynamics occurs 1 day before H3K36me3 (Fig. 5d), suggesting that global changes in H3K4me1 precede changes in H3K36me3. Summarizing at the genome-wide level, UMAPs of H3K4me1 and H3K36me3 of single-incubated cells show that both active marks move progressively towards a macrophage state during the timecourse (Fig. 5e).

Using continuous pseudotime of H3K4me1 and H3K36me3 as our training data (Methods), for both H3K4me1 and H3K36me3 we infer where along pseudotime each double-incubated cell came from. Plotting the inferred pseudotimes of each mark for each cell uncovers the dynamic relationships between the two marks (Fig. 5e). H3K4me1 pseudotime initially progresses while H3K36me3 remains relatively unchanged. As H3K4me1 pseudotime approaches 0.5, H3K36me3 then progresses rapidly towards 1. This sigmoidal-like relationship between H3K4me1 versus H3K36me3 pseudotime progression is consistent with H3K4me1 dynamics occurring before H3K36me3. Finally, we used this inferred pseudotime information to project the deconvolved cells onto the H3K4me1 and H3K36me3 UMAPs. Both UMAPs showed that the single-incubated and deconvolved cells intermingle with each other, suggesting that deconvolution was successful (Extended Data Fig. 10e,f). Using the deconvolved cells as anchors, we then linked the two histone modification maps together (Fig. 5f).

Since we observed that H3K4me1 dynamics occur before H3K36me3, we asked whether we could model the H3K36me3 dynamics as a first-order differential equation analogous to RNA velocity⁴⁵ (Fig. 5g, top; Methods). Since our data come from a timecourse, we directly fitted the exponential curves for dynamic genes along pseudotime for H3K36me3 (Extended Data Fig. 10g), which avoids making steady-state assumptions and leverages information from both single-incubated and deconvolved cells across histone modifications. The distribution of inferred rate constants from the exponential fit show a median of approximately 2.3 per pseudotime (Extended Data Fig. 10h). These rate constants were then used to predict the H3K36me3 levels for each cell over small pseudotime steps ($\Delta t = 0.02$; Fig. 5g). Finally, summarizing the predictions of dynamic genes, we projected the high-dimensional velocity vectors onto the first two principal components (PCs). From the chromatin velocity summary, we found that differentiation starts with large changes in H3K36me3 dynamics, and then relaxes towards the macrophage state.

In summary, we applied scChIX-seq to two active histone modifications to find dynamic relationships between activation states. We then model these dynamics to infer chromatin velocity during macrophage differentiation.

Discussion

Here, we demonstrate that scChIX-seq can deconvolve multiplexed histone modifications, expanding the number of histone marks that can be profiled in single cells. Using simulations, purified cell types and whole bone marrow, we demonstrate that scChIX-seq can accurately map several histone marks. To show how scChIX-seq can reveal unique biological insights in more challenging systems, we applied scChIX-seq to study H3K36me3 and H3K9me3 dynamics during mouse organogenesis to reveal the joint transcriptional and heterochromatin relationships in single cells. scChIX-seq can identify complex cell-type relationships between histone modifications, such as when several cell types can share a similar heterochromatin landscape. Finally, we applied scChIX-seq to two active marks during macrophage in vitro differentiation to quantify the relationship between two correlating marks. Importantly, scChIX-seq is flexible in which histone modifications can be used. The correlation structure between modifications is inferred from the model and therefore does not require a priori assumptions of specific features of the two modifications. Thus, scChIX-seq complements a recent method that focuses on differences in fragment lengths between Pol2 serine-5 phosphate and H3K27me3 to assign reads to their respective mark⁴⁶.

Recently, there have been new experimental innovations to CUT&TAG that modify the pA-Tn5 complex to map several histone modifications in single cells^{24,47–49}. One drawback of Tn5-based approaches (for example, CUT&TAG) compared with MNase-based (for example, sortChIC and CUT&RUN) used in this study is that Tn5 can have biases to open chromatin⁵⁰. Current CUT&TAG methods suppress this bias by using more stringent washing conditions⁵¹, but exceedingly high salt conditions reduce the sensitivity and could wash away weakly bound factors such as transcription factors^{50,51}. On the flip side, MNase-based approaches involve more experimental effort than Tn5-based approaches, reducing the number of single cells that can be processed per round. Although we demonstrate our scChIX-seq method using an MNase-based approach (sortChIC), our computational and experimental framework can also be applied to Tn5-based strategies. Furthermore, our scChIX-seq method may have synergies with recent nanobody-based methods^{47,48}. For example, using two nanobodies, each specific to a different species of immunoglobulin G, one can profile four histone modifications by generating two sets of scChIX-seq simultaneously: two antibodies raised from one species and the other two antibodies raised from the second species.

A limitation in scChIX-seq is that the maximum number of cuts at a specific base pair location is fundamentally limited by the copy

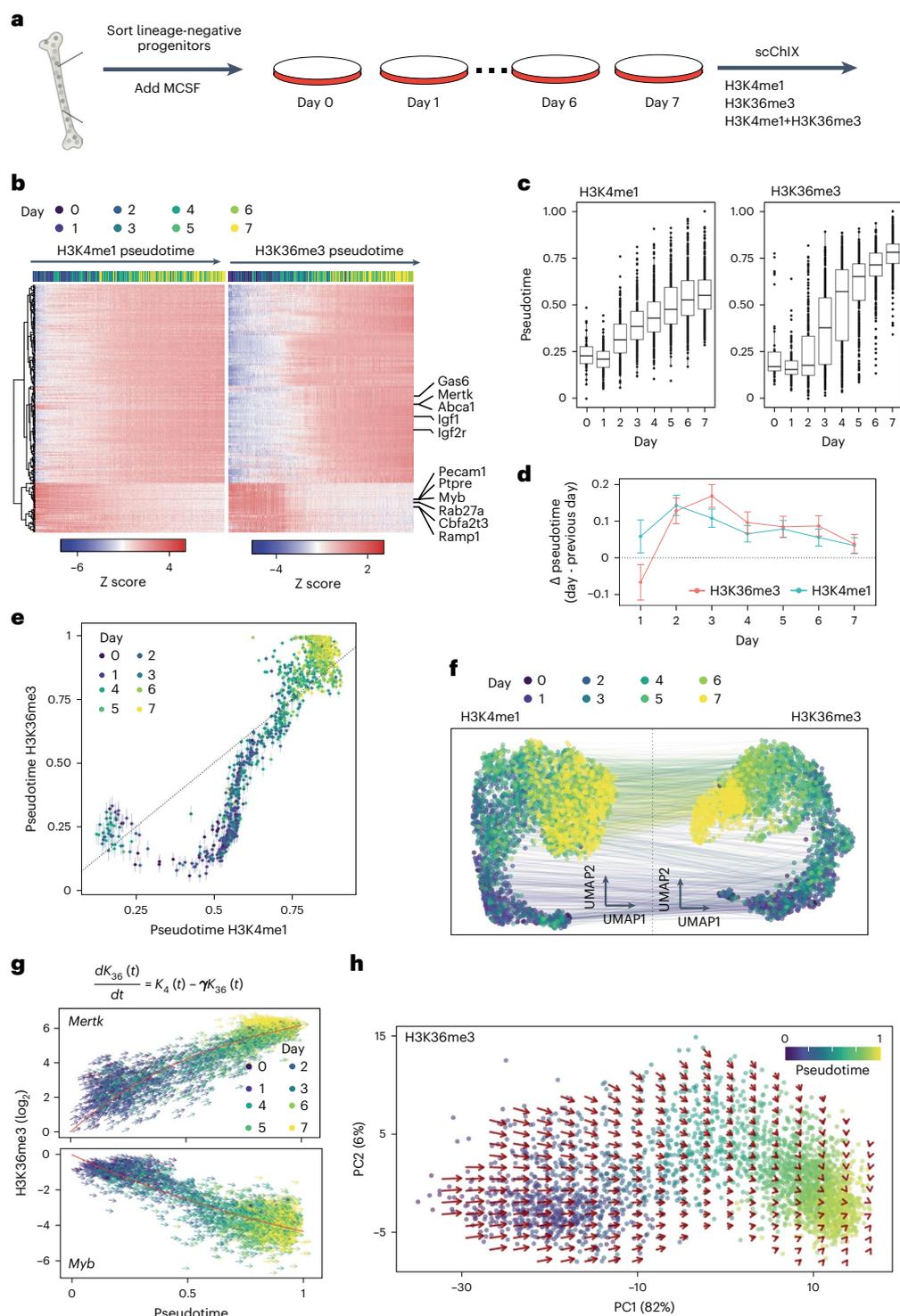


Fig. 5 | Applying scChIX-seq to two active marks reveals chromatin velocity during in vitro macrophage differentiation. **a**, Schematic of mouse macrophage in vitro differentiation timecourse experiment to study H3K4me1 and H3K36me3 in single cells. **b**, Heatmap of histone modification signal on the bodies of dynamic genes over pseudotime. Rows are gene bodies and columns are single-incubated cells ordered along pseudotime. Color labels of columns are days from which the cells were recovered during the timecourse. **c**, Boxplots of pseudotime estimates of single-incubated cells along the timecourse. Number of cells per day for H3K4me1: $n = 58$ day 0, $n = 148$ day 1, $n = 249$ day 2, $n = 350$ day 3, $n = 369$ day 4, $n = 383$ day 5, $n = 488$ day 6, $n = 519$ day 7. For H3K36me3: $n = 42$ day 0, $n = 125$ day 1, $n = 176$ day 2, $n = 301$ day 3, $n = 384$ day 4, $n = 366$ day 5, $n = 522$ day 6, $n = 567$ day 7. Boxplots show 25th percentile, median and 75th percentile,

with the whiskers spanning 97% of the data. **d**, Estimate of the average difference of pseudotime from one day to the next. Error bars indicate 95% CI, calculated by a linear model of the pseudotime differences between days. Statistics derived from number of cells indicated in **c**. **e**, Estimates of two different pseudotimes from a single cell. Error bars are 95% CI of the estimates. Each point is a double-incubated cell. **f**, Joint UMAP of H3K4me1 and H3K36me3 from scChIX-seq, lines connect single cells with multimodal information. **g**, Chromatin velocity estimates of an upregulated gene (above) and a downregulated gene (below). Red curve is the exponential relaxation fit according to the solution of the first-order differentiation equation. **h**, High-dimensional chromatin velocities of dynamic genes projected onto PCs 1 and 2. Vector field estimated by smoothing across nearest neighbors of cells (Methods).

number in that cell. Therefore, a nucleosome that has several modifications in their histone tails would still be cut only once. Currently, our binning strategy (5 kilobase (kb), 50 kb or gene bodies, depending on the biological question) and multinomial model assumes that there is no limit to the number of fragments that can be generated in one bin, which is an approximation that is valid when the bins are large and the number of cuts within the bins are small (for example, due to dropouts).

We demonstrate that scChIX-seq can reveal biological insights by multimodal analysis that would otherwise be obscured by analyzing each modality separately. Overall, scChIX-seq unlocks multimodal analysis in antibody-based chromatin profiling and enables joint analysis of distinct histone modifications in single cells.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-022-01560-3>.

References

- Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* **1839**, 627–643 (2014).
- Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
- Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
- Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
- Lara-Astiaso, D. et al. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
- Rotem, A. et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
- Grosselin, K. et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
- Ai, S. et al. Profiling chromatin states using single-cell itChIP-seq. *Nat. Cell Biol.* **21**, 1164–1172 (2019).
- Schmid, M., Durussel, T. & Laemmli, U. K. ChIC and ChEC: genomic mapping of chromatin proteins. *Mol. Cell* **16**, 147–157 (2004).
- Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
- Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1–10 (2019).
- Ku, W. L. et al. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).
- Wang, Q. et al. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell* **76**, 206–216.e7 (2019).
- Harada, A. et al. A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296 (2019).
- Wu, S. J. et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat. Biotechnol.* **39**, 819–824 (2021).
- Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* **39**, 825–835 (2021).
- Janssens, D. H. et al. Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nat. Genet.* **53**, 1586–1596 (2021).
- Zeller, P. et al. Hierarchical chromatin regulation during blood formation uncovered by single-cell sortChIC. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.26.440606> (2021).
- Ku, W. L., Pan, L., Cao, Y., Gao, W. & Zhao, K. Profiling single-cell histone modifications using indexing chromatin immunocleavage sequencing. *Genome Res.* **31**, 1831–1842 (2021).
- Navidi, Z., Zhang, L. & Wang, B. simATAC: a single-cell ATAC-seq simulation framework. *Genome Biol.* **22**, 1–16 (2021).
- Pauler, F. M. et al. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res.* **19**, 221–233 (2009).
- Blei, D. M., Ng, A. Y. & Jordan, M. I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
- Grün, B. & Hornik, K. topicmodels: an R package for fitting topic models. *J. Stat. Softw.* **40**, 1–30 (2011).
- Gopalan, S., Wang, Y., Harper, N. W., Garber, M. & Fazio, T. G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol. Cell* **81**, 4736–4746.e5 (2021).
- Giladi, A. et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat. Cell Biol.* **20**, 836–846 (2018).
- Mendenhall, E. M. et al. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* **6**, e1001244 (2010).
- Zou, F. et al. Expression and function of tetraspanins and their interacting partners in B cells. *Front. Immunol.* **9**, 1606 (2018).
- Benhamou, D. et al. The c-Myc/miR17-92/PTEN axis tunes PI3K activity to control expression of recombination activating genes in early B cell development. *Front. Immunol.* **9**, 2715 (2018).
- Goldmit, M. et al. Epigenetic ontogeny of the Igk locus during B cell development. *Nat. Immunol.* **6**, 198–203 (2005).
- Pan, C., Baumgarth, N. & Parnes, J. R. CD72-deficient mice reveal nonredundant roles of CD72 in B cell development and activation. *Immunity* **11**, 495–506 (1999).
- Grün, D. et al. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016).
- Pishesha, N. et al. Transcriptional divergence and conservation of human and mouse erythropoiesis. *Proc. Natl Acad. Sci. USA* **111**, 4103–4108 (2014).
- Koretzky, G. A., Abtahian, F. & Silverman, M. A. SLP76 and SLP65: complex regulation of signalling in lymphocytes and beyond. *Nat. Rev. Immunol.* **6**, 67–78 (2006).
- Brachtendorf, G. et al. Early expression of endomucin on endothelium of the mouse embryo and on putative hematopoietic clusters in the dorsal aorta. *Dev. Dyn.* **222**, 410–419 (2001).
- Sedykh, I. et al. Zebrafish Rfx4 controls dorsal and ventral midline formation in the neural tube. *Dev. Dyn.* **247**, 650–659 (2018).
- DeBoer, E. M. et al. Prenatal deletion of the RNA-binding protein HuD disrupts postnatal cortical circuit maturation and behavior. *J. Neurosci.* **34**, 3674–3686 (2014).
- Inoue, T. et al. Analysis of mouse Cdh6 gene regulation by transgenesis of modified bacterial artificial chromosomes. *Dev. Biol.* **315**, 506–520 (2008).
- Chen, A. F. et al. GRHL2-dependent enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit from naive pluripotency. *Cell Stem Cell* **23**, 226–238.e4 (2018).
- Logan, M. et al. Expression of Cre recombinase in the developing mouse limb bud driven by aPrlx enhancer. *Genesis* **33**, 77–80 (2002).
- Takeuchi, J. K. & Bruneau, B. G. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459**, 708–711 (2009).

41. Zhao, R. et al. Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev. Biol.* **317**, 614–619 (2008).
42. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
43. Murphy, Z. C. et al. Regulation of RNA polymerase II activity is essential for terminal erythroid maturation. *Blood* **138**, 1740–1756 (2021).
44. Gautier, E. L. et al. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118–1128 (2012).
45. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
46. Janssens, D. H. et al. CUT&Tag2for1: a modified method for simultaneous profiling of the accessible and silenced regulome in single cells. *Genome Biol.* **23**, 81 (2022).
47. Stuart, T. et al. Nanobody-tethered transposition allows for multifactorial chromatin profiling at single-cell resolution. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.03.08.483436> (2022).
48. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.03.08.483459> (2022).
49. Meers, M. P., Llagas, G., Janssens, D. H., Codomo, C. A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using Multi-Tag. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-022-01522-9> (2022).
50. Wang, M. & Zhang, Y. Tn5 transposase-based epigenomic profiling methods are prone to open chromatin bias. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.07.09.451758> (2021).
51. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283 (2020).

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Methods

Animal experiments

All mice used in this study were Cast-Eij/Bl6 mice and were bred and maintained in the Hubrecht Institute Animal Facility. All mouse experimentation was approved by the Animal Experimentation Committee (DEC) from the Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) and complied with existing European Union legislation and local standards.

Mouse bone marrow. Male 13-week-old C57BL/6 mice were used to extract bone marrow cells. Femurs and tibia were extracted, the bone ends were cut away to access the bone marrow, which was flushed out using a 22G syringe with HBSS (– calcium, – magnesium, – phenol red, Gibco, catalog no. 14175053) supplemented with Pen-Strep and 1% fetal calf serum. The bone marrow was dissociated and debris removed by passing through a 70 μ m cell strainer (Corning, catalog no. 431751). Cells were washed with 25 ml supplemented HBSS before depleting the sample of un-nucleated cells using IOtest 3 Lysing solution (Beckman Coulter) following the provider's instructions. Cells were washed an additional two times with PBS before processing them by the sortChIC protocol for histone modifications. For whole bone marrow experiments (that is, not enriched for specific cell types), we processed cells using the sortChIC protocol for unfixed cells (without ethanol fixation). For the ground truth experiment with sorted cell types, we processed cells using the sortChIC protocol for ethanol-fixed cells. For ethanol fixation, cells were resuspended in 70% ethanol and fixed for 1 h at -20°C . Afterwards cells were resuspended in Storage buffer (42.5 ml H_2O RNase free, 1 ml 1 M HEPES pH 7.5 (Invitrogen), 1.5 ml 5 M NaCl, 3.6 μ l spermidine (Sigma Aldrich, catalog no. S2626-5G), protease inhibitor (Sigma Aldrich, catalog no. 5056489001), 200 μ l 0.5 M EDTA, 5 μ l dimethylsulfoxide) and frozen at -80°C , before processing by the sortChIC protocol.

Mouse organogenesis. No randomization or blinding was performed. Sex of embryos was not known at the time of collection. Four to five embryos were pooled for each reported timepoint (E9.5, E10.5, E11.5) before single-cell isolation. Pooled embryos were dissociated in TrypleE for 10 min at room temperature. Undigested portions were physically removed and the remainder filtered through a 30 μ m filter before the single-cell suspension was split into three samples for each timepoint and each scChIX-seq experiment. Per timepoint, two single-cell samples were used each for a single antibody incubation (H3K36me3 or H3K9me3) and one sample for the double antibody incubation (H3K36me3 + H3K9me3). Antibody incubation was performed as described in the scChIX-seq protocol before single-cell capture using flow cytometry. A DNA library was prepared for each sample using the sortChIC protocol for unfixed cells.

In vitro macrophage differentiation. For in vitro differentiation of bone marrow-derived macrophages, bone marrow was collected aseptically by flushing tibia and femurs from euthanized wild-type male C57BL/6 mice with sterile RPMI and 10% FCS through a 70 μ m cell strainer (Corning). To enrich for stem and progenitor cells, lineage marker-positive (Lin^+) cells were depleted by magnetic-activated cell sorting using a mouse Lineage Cell Depletion kit (Miltenyi Biotec). Lin^- cells were cultured on nontissue-culture-treated plates (Corning) for 7 days in RPMI medium supplemented with 10% FCS, 100 IU ml^{-1} penicillin, 100 mg ml^{-1} streptomycin and 10 ng ml^{-1} recombinant murine MCSF (Peprotech). Medium was refreshed after 3 days. Every 24 h, suspension cells were collected and adherent cells were harvested by incubating 10 min in 2 mM EDTA/0.5% BSA in PBS. Suspension and adherent cells were combined and stained with CellTrace

fluorescent labels (Thermo Fisher), according to manufacturer's instructions. Briefly, cells were pelleted and resuspended in 37 $^{\circ}\text{C}$ PBS containing fluorescent dyes (working concentrations CellTrace CFSE (CTC): 2.5 μM ; CellTrace Yellow (CTY): 2.5 μM ; CellTrace Far Red (CTFR): 0.5 μM) at a concentration of 1,000,000 cells ml^{-1} . Cells were incubated at 37 $^{\circ}\text{C}$ protected from light for 20 min. Staining reactions were stopped by adding two volumes of RPMI/10% FCS and incubating for 5 min at room temperature, protected from light, after which cells were washed twice in PBS. The following combinations of labels were used: unstained (day 0), CTC (day 1), CTY (day 2), CTFR (day 3), CTC + CTY (day 4), CTC + CTFR (day 5), CTY + CTFR (day 6) and CTC + CTY + CTFR (day 7). After harvesting and staining, cells were fixed in 70% ethanol for 1 h and stored for later by the sortChIC protocol for fixed cells.

Cell preparation without ethanol fixation for sortChIC experiments

Cells from whole bone marrow (H3K4me1+H3K27me3) and mouse embryos (H3K36me3+H3K9me3) were processed using the sortChIC method without ethanol fixation. Cells were processed in 0.5 ml protein low-binding tubes. Following steps were performed on ice. Cells were resuspended in 500 μ l wash buffer (47.5 ml H_2O RNase free, 1 ml 1 M HEPES pH 7.5 (Invitrogen), 1.5 ml 5 M NaCl, 3.6 μ l pure spermidine solution (Sigma Aldrich)). Cells were pelleted at 600g for 3 min and resuspended in 400 μ l wash buffer 1 (wash buffer with 0.05% saponin (Sigma Aldrich), protease inhibitor cocktail (Sigma Aldrich), 4 μ l 0.5 M EDTA) containing the primary antibody (1:100 dilution for the antibody, saponin has to be prepared fresh every time as a 10% solution in PBS). Cells were incubated overnight at 4 $^{\circ}\text{C}$ on a roller, before being washed once with 500 μ l wash buffer 2 (wash buffer with 0.05% saponin, protease inhibitor). Afterwards cells were resuspended in 500 μ l wash buffer 2 containing Protein A-Micrococcal Nuclease (pA-MNase) (3 ng ml^{-1}) and incubated for 1 h at 4 $^{\circ}\text{C}$ on a roller.

Finally, cells were washed an additional two times with 500 μ l wash buffer 2 before passing it through a 70 μ m cell strainer (Corning, catalog no. 431751) and sorting G1 cells based on Hoechst staining on a BD Influx FACS machine into 384-well plates containing 50 nl wash buffer 3 (wash buffer containing 0.05% saponin) and 5 μ l sterile filtered mineral oil (Sigma Aldrich) per well. Small volumes were distributed using a Nanodrop II system (Innovadyme).

Cell preparation with ethanol fixation and surface antibody incubation for sortChIC experiments

Cells from sorted bone marrow (H3K27me3+H3K9me3) and macrophage in vitro differentiation (H3K4me1+H3K36me3) were processed using the ethanol fixation protocol. Sorted bone marrow cells were also incubated with surface antibody to enrich for known cell types. For the ethanol-fixed cells the above described sortChIC protocol was adapted. Wash buffers were used as described above, except that 0.05% saponin was exchanged for 0.05% Tween. Ethanol-fixed cells were thawed on ice. Cells were spun at 400g for 5 min and washed once with 400 μ l wash buffer 1. Cells were spun again at 400g and resuspended in 400 μ l wash buffer 1. Cell suspension was split into three samples each having a volume of 400 μ l and incubated with one or two antibodies (1:100 dilution for the antibody) overnight on a roller at 4 $^{\circ}\text{C}$. The next day, cells were spun at 400g, washed once with 400 μ l wash buffer 2 and resuspended in 500 μ l wash buffer 2 containing pA-MNase (3 ng ml^{-1}) and incubated for 1 h on a rotator at 4 $^{\circ}\text{C}$. Next, cells were spun at 400g and resuspended in 400 μ l wash buffer 2 (with addition of 5% blocking rat serum). To sort for defined cell types in the ground truth bone marrow experiment, surface antibodies were added according to these concentrations and were incubated for 30 min on ice:

antibody	info
GR1	A647, anti-mouse Ly-6G/Ly-6C (Gr-1) Antibody, clone: RB6-8C5
NK1	A488, anti-mouse NK-1.1 Antibody, clone: PK136
CD19	BV421, anti-mouse CD19 Antibody, clone: 6D5
working concentration	
1	: 8,000
1	: 400
1	: 200

BD FAC software v.1.2.0.142 was used to collect data from the FACS machine during cell sorting; see Supplemental Fig. 1 for the gating strategy.

Finally, samples were washed once with 500 μ l wash buffer 2 before passing them through a 70 μ m cell strainer (Corning, catalog no. 431751) and sorting on a BD Influx FACS machine, with surface antibody specific gating, into 384-well plates containing 50 nl wash buffer 3 (wash buffer containing 0.05% Tween) and 5 μ l sterile filtered mineral oil (Sigma Aldrich) per well. Small volumes were distributed using a Nanodrop II system (Innovadyme).

MNase activation for sortChIC experiments

Targeted fragmentation was started by the addition of 5 μ l wash buffer 2 containing 4 mM CaCl_2 . For digestion, plates were incubated for 30 min in a PCR machine set at 4 $^\circ\text{C}$. Afterwards the reaction was stopped by adding 100 nl of a stop solution containing 40 mM EGTA, 1.5% NP40, and 10 nl 2 mg ml^{-1} proteinase K. Plates were incubated in a PCR machine for further 20 min at 4 $^\circ\text{C}$, before chromatin was released and pA-MNase permanently destroyed by proteinase K digestion at 65 $^\circ\text{C}$ for 6 h followed by 80 $^\circ\text{C}$ for 20 min to heat inactivate proteinase K. Afterwards plates were stored at -80°C until further processing.

Library preparation for sortChIC experiments

DNA fragments were blunt ended by adding 150 nl end repair mix per well and incubating for 30 min at 37 $^\circ\text{C}$ followed by 20 min at 75 $^\circ\text{C}$ for enzyme inactivation. End repair mix per well: Klenow large (NEB, catalog no. M0210L) 2.5 nl, T4 PNK (NEB, catalog no. M0201L) 2.5 nl, dNTPs 10 mM 6 nl, ATP 100 mM 3.5 nl, MgCl_2 25 mM 10 nl, PEG8000 50% 7.5 nl, PNK buffer 10 \times (NEB, catalog no. B0201S) 35 nl, BSA 20 ng 1.8 nl, nuclease-free water 81.3 nl.

Blunt fragments were subsequently A-tailed by adding 150 nl per well of A-tailing mix and incubated for 15 min at 72 $^\circ\text{C}$. Through the strong preference of AmpliTaq 360 to incorporate dATP as a single base overhang even in the presence of other nucleotides, a general dNTP removal was not necessary. A-tailing mix per well: AmpliTaq 360 (Thermo Fisher Scientific, catalog no. 4398828) 1 nl, dATPs 100 mM 1 nl, KCl 1 M 25 nl, PEG8000 50% 7.5 nl, BSA 20 ng 0.8 nl, nuclease-free water 114.8 nl.

Fragments were ligated to T-tail containing forked adapters containing a T7 polymerase binding site for in vitro transcription (IVT)-based amplification.

Top strand: 5'-GGTGATGCCGTAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATCNNNACACTAT-3'

Bottom strand: 5'-TAGTGTGTTNNGATCGTCGGACTGTAGAACTCCCTATAGTGAGTCGTATTACCGGCGAGCTT-3'

The three random nucleotides (NNN) were the unique molecular identifier used for read deduplication and the eight bases afterwards represent the cell barcodes, which were different for each of the 384 wells. For a full list of adapters and the cell barcodes for each well, see the excel sheet in Supplemental Table 1. Cell barcodes for each 384-well plates are also found as a text file in the scChIX-seq Github repository:

(https://github.com/jakeyeung/scChIX/blob/main/inst/extdata/cellbarcodes_384_NLA_annotated.bc).

For ligation, 50 nl of 5 μ M adapter in 50 mM Tris pH 7 was added to each well with a Mosquito HTS (ttp labtech). After centrifugation, 150 nl of ligation mix was added before incubating plates for 20 min at 4 $^\circ\text{C}$, followed by 16 h at 16 $^\circ\text{C}$ for ligation and 10 min at 65 $^\circ\text{C}$ to inactivate ligase. Adapter ligation mix per well: T4 ligase (400,000 U ml^{-1} , NEB, catalog no. M0202L) 25 nl, MgCl_2 1 M 3.5 nl, Tris 1 M pH 7.5 10.5 nl, DTT 0.1 M 52.5 nl, ATP 100 mM 3.5 nl, PEG8000 50% 10 nl, BSA 20 ng 1 nl, nuclease-free water 44 nl.

Before pooling, 1 μ l nuclease-free water was added to each well to minimize material loss. Ligation products were pooled by centrifugation into oil coated VBLOK200 Reservoir (ClickBio) at 500g for 2 min and the liquid face was transferred into 1.5 ml Eppendorf tubes and then purified by centrifugation at 13,000g for 1 min and transferred into a fresh tube twice. DNA fragments were purified using Ampure XP beads (Beckman Coulter, prediluted one in eight in bead binding buffer: 1 M NaCl, 20% PEG8000, 20 mM Tris pH 8, 1 mM EDTA) at a bead to sample ratio of 0.8. After 15 min incubation at room temperature, beads were washed twice with 1 ml 80% ethanol resuspending the beads during the first wash and resuspended in 20 μ l nuclease-free water. After 2 min elution, the supernatant was transferred into a fresh 0.5 ml tube. A second cleanup was performed adding 26 μ l undiluted Ampure XP beads and the beads were resuspended in 8 μ l nuclease-free water. The cleaned DNA was then linear amplified by IVT by adding 12 μ l of MEGAScript T7 Transcription Kit (Fisher Scientific, catalog no. AMB13345) for 12 h at 37 $^\circ\text{C}$. Template DNA was removed by addition of 2 μ l $^{-1}$ TurboDNase (IVT kit) and incubation for 15 min at 37 $^\circ\text{C}$. The RNA produced was further purified using RNA Clean XP beads (Beckman Coulter) at a beads to sample ratio of 0.8 and samples were resuspended in 22 μ l of nuclease-free water. RNA was fragmented by mixing in 4.4 μ l fragmentation buffer (200 mM Tris-acetate pH 8.1, 500 mM KOAc, 150 mM MgOAc) and incubation for 2 min at 94 $^\circ\text{C}$. Fragmentation was stopped by transferring samples to ice, adding 2.64 μ l 0.5 M EDTA and another bead cleanup; samples were resuspended in 12 μ l nuclease-free water.

RNA (5 μ l) was primed for reverse transcription by adding 0.5 μ l 10 mM dNTPs and 1 μ l 20 mM randomhexamerRT primer (5'-GCCTTGGCACCCGAGAATTCANNNNNN-3') and hybridizing it by incubation at 65 $^\circ\text{C}$ for 5 min followed by direct cool down on ice. Reverse transcription was performed by further addition of 2 μ l first strand buffer (part of Invitrogen kit, catalog no. 18064014), 1 μ l 0.1 M DTT, 0.5 μ l RNAseOUT (Invitrogen, catalog no. LS10777019) and 0.5 μ l SuperscriptII (Invitrogen, catalog no. 18064014) and incubating the mixture at 25 $^\circ\text{C}$ for 10 min followed by 1 h at 42 $^\circ\text{C}$. Single-stranded DNA was purified through incubation with 0.5 μ l RNAseA (Thermo Fisher, catalog no. EN0531) and incubation for 30 min at 37 $^\circ\text{C}$.

A final PCR amplification to add the Illumina small RNA barcodes and handles was performed by adding 25 μ l of NEBNext Ultra II Q5 Master Mix (NEB, catalog no. M0492L), 11 μ l nuclease-free water and 2 μ l of 10 μ M RPI and RPIx primers.

PCR protocol for sortChIC experiments

Activation for 30 s at 98 $^\circ\text{C}$, 8–12 cycles (depending on starting material), 10 s at 98 $^\circ\text{C}$, 30 s at 60 $^\circ\text{C}$, 30 s at 72 $^\circ\text{C}$, final amplification 10 min at 72 $^\circ\text{C}$.

PCR products were cleaned by two consecutive DNA bead clean-ups with a bead to sample ratio of 0.8. Final product was eluted in 7 μ l nuclease-free water. The abundance and quality of the final library were assessed by QUBIT and bioanalyzer.

Data processing

All DNA libraries were sequenced on an Illumina NextSeq500 with 2 \times 75 bp. We ran the raw fastq files through the Single-Cell MultiOmics (SCMO) workflow (github.com/BuysDB/SingleCellMultiOmics⁵²). The workflow comprises of six steps.

(1) Demultiplex raw fastq files using demux.py (SCMO). (2) Trim fastq files by removing adapters using cutadapt (v.3.5). (3) Map trimmed fastq files using bwa (v.0.7.17-r1188). (4) Tag bam files with cell barcode information, using bamtagmultiome.py (SCMO). (5) Generate count tables using bamToCountTable.py (SCMO). (6) Run dimensionality reduction of count matrices using run_LDA_model.R. See an example of the pipeline in the scChIX-seq Github repository⁵³.

Unmixing scChIX-seq signal

Single-cell epigenomics techniques (for example, sortChIC, CUT&RUN and CUT&TAG) generate a vector of counts indicating the number of cut fragments that map in each genomic region for each cell. We model the vector of counts from a double-incubated cell \vec{y} as a linear combination of two multinomial distributions: one coming from a cluster c of histone modification 1, parameterized by \vec{p}_c , the other from another cluster d of histone modification 2 \vec{q}_d . The log-likelihood for a linear combination of two multinomials is:

$$L_{(c,d)} = \log(P(\vec{y}|\vec{p}_c, \vec{q}_d, w)) \propto \sum_{g=1}^G y_g \log(wp_{c,g} + (1-w)q_{d,g}). \quad (1)$$

\vec{y} is the number of cuts across the genome for a double-incubated cell. $p_{c,g}$ and $q_{d,g}$ are cluster-specific probabilities indicating the likelihood that a cut fragment maps to region g in histone modifications 1 and 2, respectively. w is the mixing fraction of histone modification 1 in the double-incubated cell, which we estimate by maximizing the log-likelihood given \vec{y} , \vec{p}_c and \vec{q}_d .

Applying single-cell techniques to complex tissues generates data with many clusters. Therefore, given a double-incubated cell, we do not know which pair of clusters (c, d) were combined to generate the observed counts. We therefore calculate the log-likelihood for all possible pairs of clusters learned from the training data and then select the cluster pair with the highest probability for each cell.

Cluster-specific probabilities \vec{p}_c and \vec{q}_d are learned by applying LDA (with $k = 30$ topics) using the topicmodels R package⁵⁴ to the training data (that is, single-incubated cells), which are count matrices.

After assigning each cell to the most probable cluster pair (\hat{c}, \hat{d}), we assign $y_{i,j}$, the j th read mapped to region g in cell i , to histone mark 1 with probability $P_{i,j}$:

$$P_{i,j} = \frac{wp_{\hat{c},g}}{wp_{\hat{c},g} + (1-w)q_{\hat{d},g}}. \quad (2)$$

This assignment generates a pair of vectors $\vec{y}_{1,i}$ and $\vec{y}_{2,i}$ that are linked because they both come from cell i . Unmixed counts $\vec{y}_{1,i}$ and $\vec{y}_{2,i}$ are then projected back onto the space inferred from training data of histone modification 1 and 2, respectively. The links between histone modification 1 and 2 are used to transfer labels and create linked UMAPs between the two histone modifications.

Latent Dirichlet allocation

LDA is a probabilistic matrix decomposition model that is useful when the input data is a matrix of counts. LDA uses hierarchical multinomial models to estimate the relative frequencies of cuts in each genomic region in single cells.

To generate the genomic location of the j th read for cell i :

Choose a topic $z_{i,j}$ by sampling from the cell-specific distribution of topics:

$$\begin{aligned} \vec{U}_i &\sim \text{Dirichlet}(\alpha) \\ z_{i,j} &\sim \text{Multinomial}(\vec{U}_i, 1) \end{aligned}$$

Choose genomic region $w_{i,j}$ by sampling from the topic-specific distribution of genomic regions:

$$\begin{aligned} \vec{V}_k &\sim \text{Dirichlet}(\delta) \\ w_{i,j} &\sim \text{Multinomial}(\vec{V}_{z_{i,j}}, 1) \end{aligned}$$

The Dirichlet distributions are priors to prevent overfitting when there are few cuts in the region. We used the LDA model implemented by the topicmodels R package, using the Gibbs sampling implementation with hyperparameters $\alpha = 1.67$, $\delta = 0.1$, where K is the number of topics²³.

We estimate \vec{p}_c and \vec{q}_d for each cluster in histone modification 1 $\{\vec{p}_1, \vec{p}_2, \dots, \vec{p}_c\}$ and modification 2 $\{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_D\}$ by averaging the estimated probabilities across cells assigned to each cluster for each gene g :

$$p_{g,c} = \frac{1}{|C|} \sum_{i \in C} \sum_{k=1}^K V_{g,k} U_{k,i}$$

where C is the set of cells that belong to cluster c .

Simulation of single- and double-incubated histone modification data

To simulate multimodal single-cell histone modification data with varying degrees of overlap, we extended simATAC⁵⁵ to allow generating cell-type profiles from histone modifications of varying mutually exclusive relationships.

For each cell type, we first run simATAC to generate sparse count data of 10,000 loci across 750 cells partitioned into three technical replicates of 250 cells each. The high-dimensional count data are sparse. Counts from each locus are generated according to a Poisson likelihood with locus-specific means (λ) matching real single-cell ATAC-seq from K562 cells (GSE99172).

In our 750 cells, cells 1–250 represent single-incubated cells from mark 1; cells 251–500 from mark 2; cells 501–750 from double-incubated cells. Cells from mark 1 have counts generated from locus-specific means λ . Cells from mark 2 also have counts generated from λ , but we swap the top $x\%$ of bins with highest λ with bins with lowest λ , allowing precisely defined sets of mutually exclusive and overlapping bins. We use $x = 1\%$, 50% and 99% to benchmark our method from mostly overlapping (that is $x = 1\%$) to mostly mutually exclusive (that is $x = 99\%$) Cells from mark 3 are generated by adding counts generated from mark 1 and mark 2 to simulate double-incubated cells.

To generate cell-type-specific profiles, we repeat the above with a cell-type-specific random seed and shuffle the order of the bins. This generates count data where λ is cell-type specific, but the distribution of λ are preserved genome-wide.

Estimating the top cluster-specific bins

We use the LDA matrix factorization to identify the top cluster-specific bins in the data. We rank the bin loadings for each cell type and take the top 150 (whole bone marrow) or 250 (mouse organogenesis) bins with the largest loadings.

Inferring pseudotime in differentiation data

To analyze the macrophage differentiation data, we first removed erythroblasts, plasmacytoid dendritic cells, and innate lymphocyte cells from the data, which were concentrated at day 0 and not considered to be part of the macrophage differentiation trajectory. We then ran LDA ($k = 30$ topics) and performed principal component analysis (PCA) on the LDA outputs, which retrieves the principal components that explain the largest amount of variance after denoising the data. We used the first principal component for H3K4me1 and H3K36me3 to define pseudotime, which we found correlates with the day along the timecourse.

Unmixing scChIX-seq signal from continuous pseudotime

To apply scChIX-seq on continuous pseudotime, we modify the log-likelihood (equation (1)) to account for a continuous variable:

$$L(t_1, t_2) = \log(P(\vec{y}|\vec{p}(t_1), \vec{q}(t_2), w)) \propto \sum_{g=1}^G y_g \log(wp_g(t_1) + (1-w)q_g(t_2)) \quad (3)$$

where $t_1 \in [0, 1]$ is pseudotime from histone modification 1 and $t_2 \in [0, 1]$ is pseudotime from modification 2.

To estimate pseudotime, we ran LDA to denoise the count matrix, and then ran PCA to estimate largest principal components explaining the variance in the data. We took the first principal component as our pseudotime estimate for both marks, which captured the epigenomic changes over the 7-day timecourse.

$p_g(t)$ is estimated by fitting the signal from histone modification 1 at genomic region g with a lowess curve along pseudotime. We estimate q_g analogously but using signal from histone modification 2.

To infer the pseudotime of histone modifications 1 and 2 simultaneously given a vector of counts from a double-incubated cell, we estimate t_1 and t_2 that minimizes the log-likelihood L from equation (3). We estimate the variance-covariance matrix of t_1 and t_2 by the square root of the inverse of the Hessian matrix, which we use to calculate the standard errors.

Since the t_1 and t_2 are constrained between 0 and 1, we use the L-BFGS-B optimization algorithm implemented in R. Since estimates from a single cell can sometimes be noisy due to low counts, we sum the counts across the 25-nearest neighbors (estimated from the latent space inferred by LDA) for each double-incubated cell.

Chromatin velocity during macrophage differentiation

We assume that dynamic genomic regions in H3K36me3 can be modeled using a first-order differential equation

$$\frac{dK_{36}(t)}{dt} = K_4(t) - \gamma K_{36}(t). \quad (4)$$

We estimate the time constant γ for each genomic region by fitting an exponential relaxation function across pseudotime

$$K_{36}(t) = y_0 + A(1 - e^{-\gamma t}), \quad (5)$$

where y_0 is the signal at $t = 0$ and A is the predicted H3K36me3 levels at steady state. Fitting the γ directly from the pseudotime allows us to leverage signal from both single- and deconvolved cells.

To predict future values of H3K36me3 levels for each cell at each genomic region, we use the Euler method and plug in the estimated γ , H3K4me1 levels at time t and time step h of 0.02 pseudotime units:

$$K_{36}(t + 1) = K_{36}(t) + h(K_4(t) - \gamma K_{36}(t)). \quad (6)$$

Finally, we project the single- and double-incubated H3K36me3 signal onto the first two principal components and project the predicted future values onto the PCA. We use the velocity grid flow visualization as implemented in *velocity*⁵⁶ to visualize the velocity vectors on the PCA space.

Comparison with multi-CUT&TAG

Raw fastq files (R1, R2 and R3) from the single-cell experiments were downloaded from Gene Expression Omnibus accession number GSE171554. The first 42 bases of the reads in R1 and R2 were trimmed to remove the barcodes and the bases common to all Tn5 adapter sequences. The 16-base cell barcodes in R3 were added to the fastq headers of R1 and R2. The trimmed and cell-barcoded R1 and R2 reads were then aligned to the mm10 mouse genome using Burrows-Wheeler aligner (*bwa* v.0.7.17-r1188). Fragments that start at same location and have the same cell barcode were considered

duplicates and discarded. Cells with more than 100 fragments with MAPQ scores in R1 greater than or equal to 40 were kept for comparison with scChIX-seq.

Reporting summary

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

Data availability

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through Gene Expression Omnibus Series accession number [GSE155280](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155280) (ref.⁵⁷).

Code availability

We developed the SingleCellMultiOmics package, in which there are modules used for processing sortChIC data (<https://github.com/BuysDB/SingleCellMultiOmics>)⁵², and an R package that implements scChIX-seq and contains snakemake workflows for processing data and example notebooks for downstream analyses (<https://github.com/jakeyeung/scChIX>)⁵³.

References

- de Barbanson, B. A. et al. BuysDB/SingleCellMultiOmics: 0.1.30 (v.0.1.30). Zenodo. <https://doi.org/10.5281/zenodo.7074511> (2022).
- Yeung, J. *jakeyeung/scChIX*: v.1.0.1 (v.1.0.1). Zenodo. <https://doi.org/10.5281/zenodo.7152037> (2022).
- Grün, B. & Hornik, K. *topicmodels*: an R package for fitting topic models. *J. Stat. Softw.* **40**, 1–30 (2011).
- Navidi, Z., Zhang, L. & Wang, B. *simATAC*: a single-cell ATAC-seq simulation framework. *Genome Biol.* **22**, 74 (2021).
- La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
- Yeung, J., Florescu, M., Zeller, P., de Barbanson, B. A., Wellenstein, M. D. & van Oudenaarden, A. scChIX-seq infers relationships between histone modifications in single cells. Datasets. Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155280> (2022).

Acknowledgements

We thank M. van Loenhout for experimental advice on purifying cell types from the bone marrow, R. van der Linden for expertise with FACS and M. Blotenburg for help with cell typing the mouse organogenesis dataset. We thank M. Saraswat and O. Stegle for discussions on multinomial distributions. This work was supported by a European Research Council Advanced grant (ERC-AdG 742225-IntScOmics); Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP grant (NWO CW 714.016.001) and NWO grant (OCENW.GROOT.2019.017); the Swiss National Science Foundation Early Postdoc Mobility (P2ELP3-184488 to P.Z. and P2BSP3-174991 to J.Y.); Marie Skłodowska-Curie Actions Postdoc (798573 to P.Z.) and the Human Frontier for Science Program Long-Term Fellowships (LT000209-2018-L to P.Z. and LT000097-2019-L to J.Y.). This work is part of the OncoCode Institute which is financed partly by the Dutch Cancer Society.

Author contributions

J.Y., M.F., B.A.d.B. and A.v.O. conceived the project. M.F. developed double-incubation techniques and performed mouse bone marrow and organogenesis experiments with help from P.Z. P.Z. developed single-incubation techniques. P.Z. and M.D.W. designed and performed macrophage in vitro differentiation experiments. J.Y., M.F. and A.v.O. analyzed the data. J.Y. developed and applied statistical methods with help from M.F. and B.A.d.B. B.A.d.B. wrote the sortChIC preprocessing pipeline, with help from M.F. and J.Y. J.Y.,

M.F. and A.v.O. wrote the manuscript, with input from P.Z., M.D.W. and B.A.d.B.

Competing interests

The authors declare no competing interests.

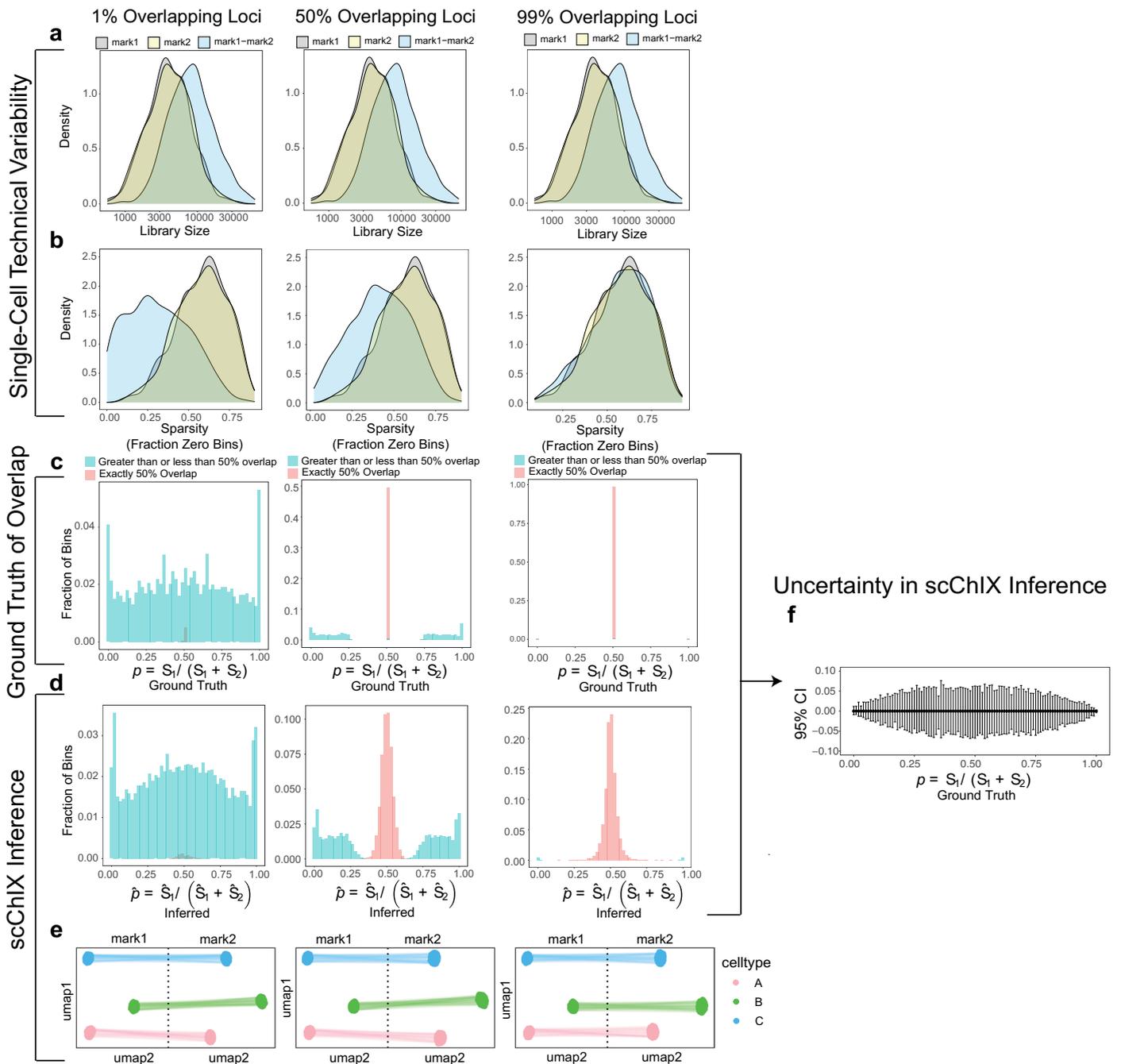
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41587-022-01560-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-022-01560-3>.

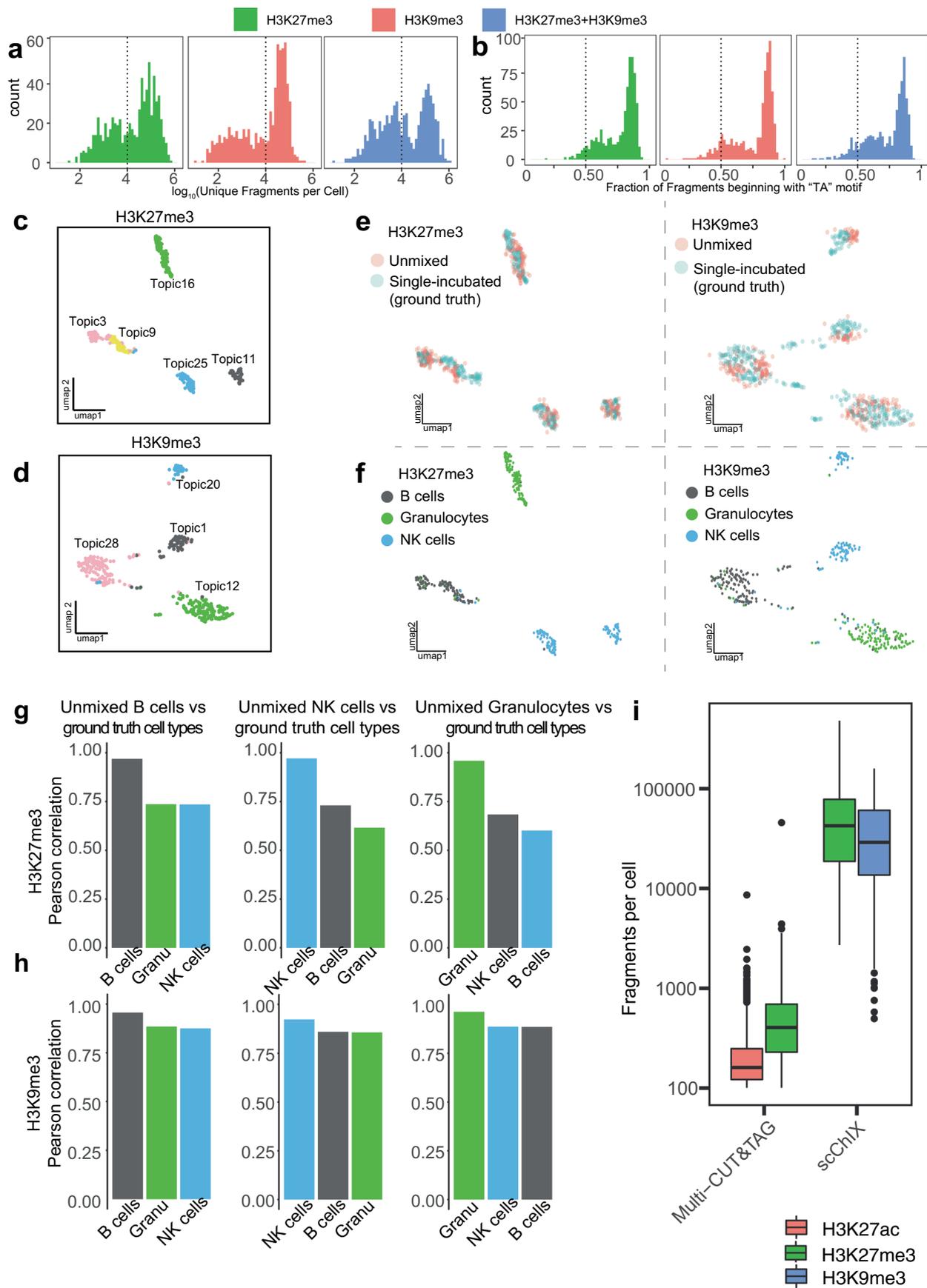
Correspondence and requests for materials should be addressed to Jake Yeung or Alexander van Oudenaarden.

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



Extended Data Fig. 1 | Benchmarking scChIX-seq across a range of overlapping patterns. Left column: simulation results in a mutually exclusive scenario (that is 1% of loci are overlapping). Middle column: results for an intermediate amount of overlap (that is 50% of loci are overlapping). Right column: results for highly correlated scenario (that is 99% of loci are overlapping). **(a)** Distribution of unique fragment cuts per cell in simulation. **(b)** Sparsity of the input matrix. Note that in the mutually exclusive scenario, the double-incubated marks is less sparse than single-incubated marks because loci with zero reads in one mark often have non-zero reads in another mark. **(c)** Distribution of the degree of overlap (defined as fraction of double-incubated

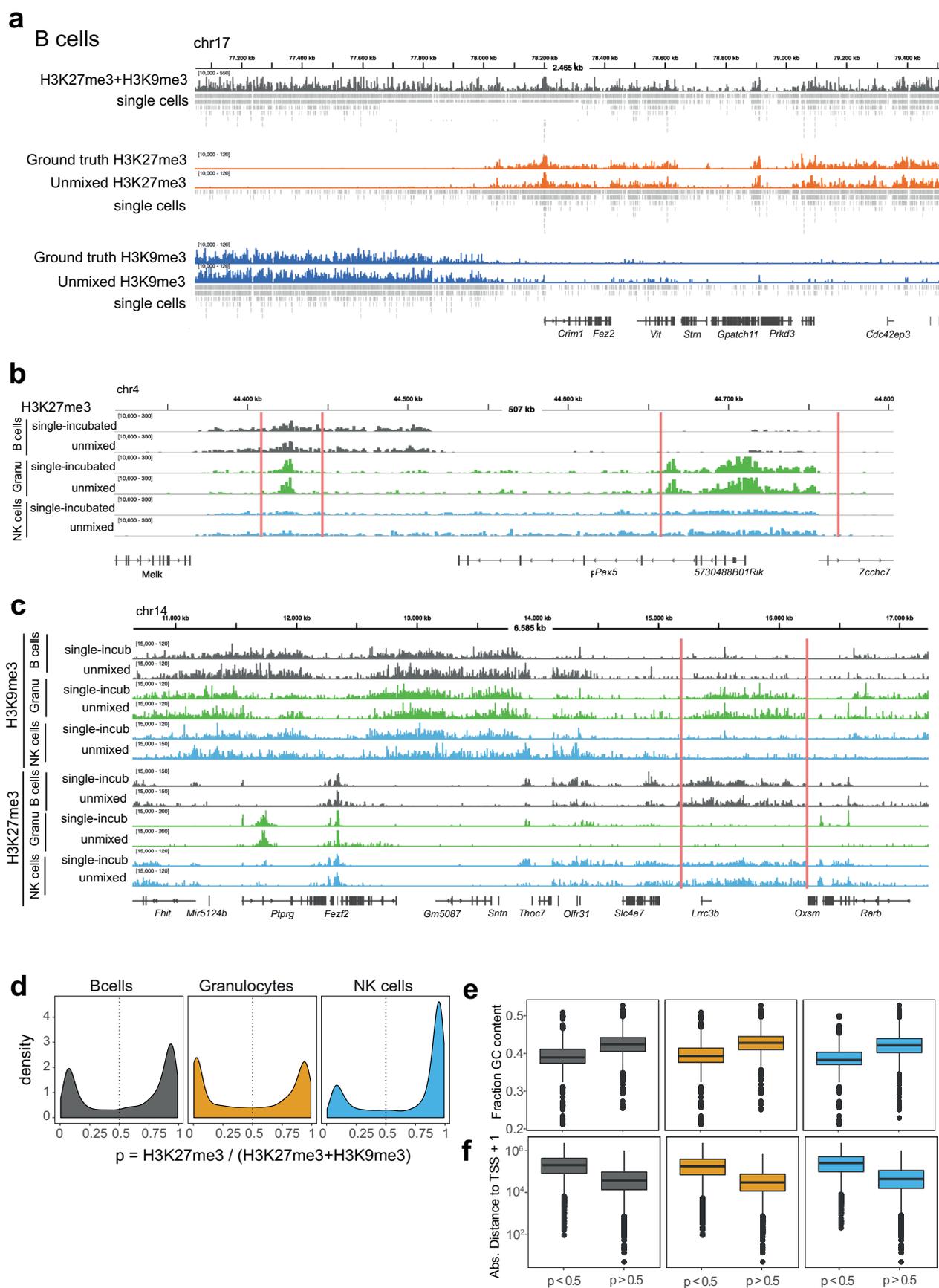
signal belonging to mark1: $p = \frac{S_1}{S_1 + S_2}$) for each locus genome-wide. **(d)** Estimated degree of overlap from scChIX-seq. **(e)** UMAP representation of the three cell types underlying simulation. UMAPs from the two marks are linked by double-incubated cells that are deconvolved by scChIX-seq. **(f)** Empirical 95% confidence interval across the range of $\hat{p} = \frac{S_1}{S_1 + S_2}$ (from 0 to 1). Range obtained by aggregating results from the three overlapping patterns. n=101 simulation datapoints spread evenly between 0 and 1 inclusive. Error bars are empirical 95% confidence intervals, centers are the mean.



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

Extended Data Fig. 2 | scChIX-seq accurately deconvolves double-incubated signal into their respective histone modifications. (a) Histogram of unique fragment cuts per cell. (b) Histogram of fraction of unique fragments starting with a “TA” motif. (c, d) UMAP of latent Dirichlet allocation (LDA) embedding using $k=30$ topics for H3K27me3 (c) and H3K9me3 (d). (e, f) UMAP representation of H3K27me3 (left) and H3K9me3 (right) data colored by unmixed or single-incubated cells (e) or ground truth cell type labels defined

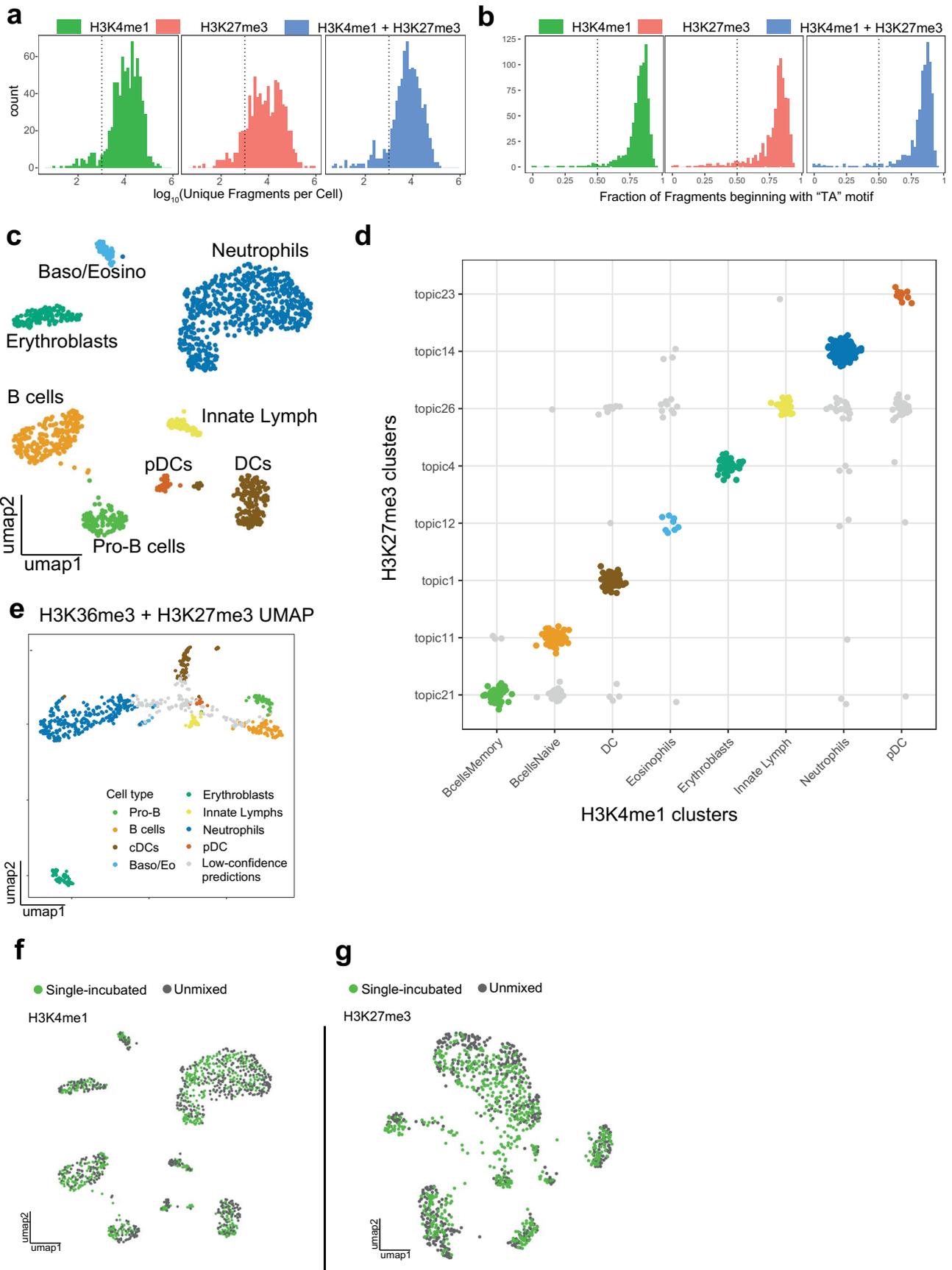
by FACS (f). (g, h) Genome-wide Pearson correlation between deconvolved H3K27me3 (g) and H3K9me3 (h) signal versus ground truth sortChIC purified by FACS. Shared genomic regions were calculated by using 1 kb bins across the genome. (i) Comparison of fragments per cell obtained from Multi-CUT&TAG versus scChIX-seq. Multi-CUT&TAG data came from a mixture of embryonic and trophoblast stem cells *in vitro*, while scChIX-seq came from sorted bone marrow cells *in vivo*. $n=1806$ cells for Multi-CUT&TAG, $n=290$ for scChIX-seq.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Coverage tracks of deconvolved cells and genome statistics. (a) Coverage tracks for B cells visualizing the H3K27me3+H3K9me3, deconvolved H3K27me3 or H3K9me3, and ground truth H3K27me3 or H3K9me3 histone modification levels for three different genomic regions. Double-incubated signal in grey, H3K27me3 single, and unmixed signal in orange, and H3K9me3 single and unmixed signal in blue. Under each coverage track are cut fragments of single cells. Each row of the single cells track are cuts from an individual cell. Shown are a subset of cells, which were chosen for their high number of cuts in the region. Rows are in decreasing order of total number of cuts. **(b)** H3K27me3 coverage tracks showing the region around *Pax5* for the ground truth H3K27me3 pseudobulk signal from single-incubated cells and for the deconvolved H3K27me3 pseudobulk signal from double-incubated cells for three cell types: B cells (grey), granulocytes (green), and NK cells (blue).

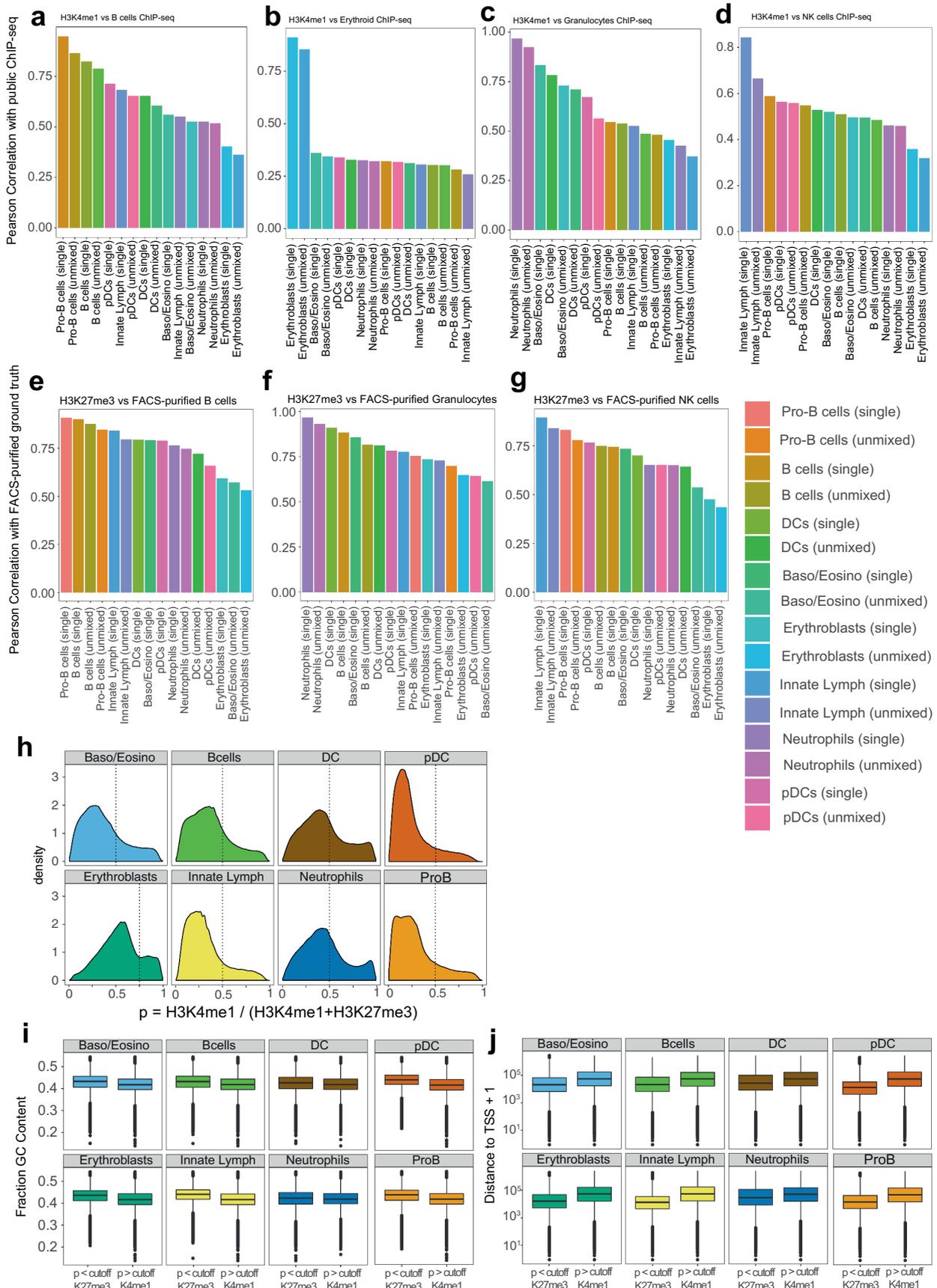
(c) H3K9me3 (top) and H3K27me3 (bottom) coverage tracks showing the region around *Auts2* for ground truth (single-incubated) and for the unmixed (unmixed) for B cells (grey), granulocytes (green) and NK cells (blue), respectively. **d** Distribution of assignment probability estimates in the genome for the three cell types. Vertical dotted lines represent cutoffs to define H3K9me3-specific (that is $p < 0.5$) or H3K27me3-specific regions (that is $p \geq 0.5$). **e** Boxplot distributions of GC content in H3K27me3-marked and H3K9me3-marked regions. **f** Boxplot distributions of distance to TSS in the two classes of regions. Distances are measured from the center of the 50 kb locus to the nearest TSS. Number of bins in each boxplot: $n=9962$ for B cells $p < 0.5$, $n=15877$ for B cells $p \geq 0.5$, $n=12483$ for granulocytes $p < 0.5$, $n=13345$ for granulocytes $p \geq 0.5$, $n=7337$ for NK cells $p < 0.5$, $n=18491$ for NK cells $p \geq 0.5$. Boxplots show 25th percentile, median and 75th percentile, with the whiskers spanning 97% of the data.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Inferring cluster pairs from H3K4me1+H3K27me3 transfers cell type labels. (a) Histogram of unique fragment cuts per cell. (b) Histogram of fraction of unique fragments starting with a “TA” motif. (c) UMAP of H3K4me1 sortChIC data, cells colored by cell type. (d) Assignment plot showing individual H3K4me1+H3K27me3 cells (represented as dots) assigned to a pair of topics (x-axis labels are H3K4me1 clusters, named by their associated cell type, while y-axis are H3K27me3 clusters). Cells along the

diagonal are high-confidence predictions that match a H3K4me1 cluster with a H3K27me3 topics, and are colored by the H3K4me1-derived cell type labels. (e) UMAP of H3K4me1+H3K27me3 sortChIC. Cells are colored by their cell type inferred from cluster pairs. Low-confidence predictions are colored in grey. (f, g) UMAP representation of H3K4me1 (f) and H3K27me3 (g). Cells are colored by whether the epigenome was generated by single-incubation or by unmixing by scChIX-seq.

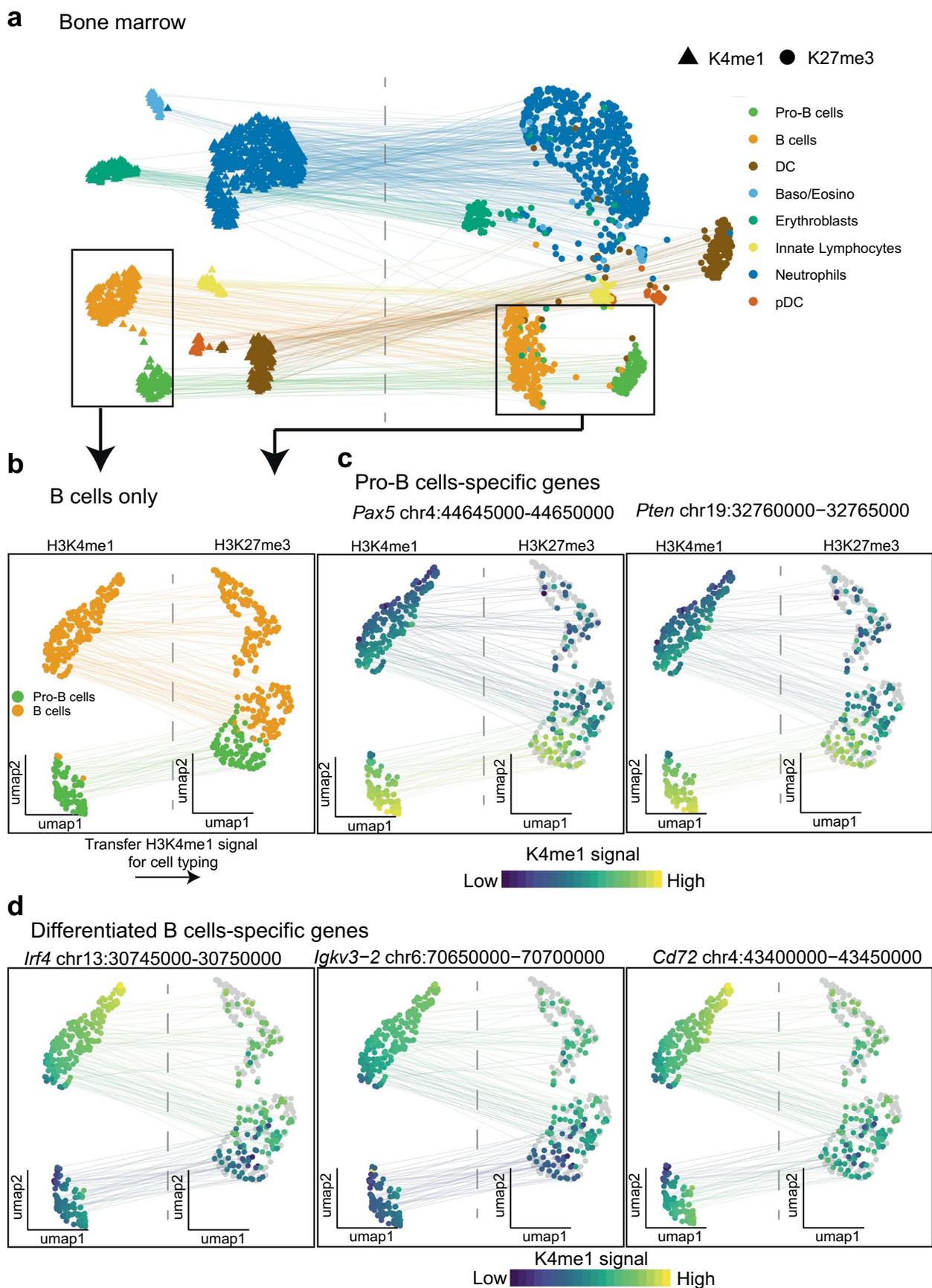


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Histone modification signal of deconvolved cell types correlates with public H3K4me1 ChIP-seq and H3K27me3 sortChIC ground truth data. (a-d)

Pearson correlation between publicly available H3K4me1 ChIP-seq⁵ data of purified B cells (a), erythroid (b), granulocytes (c), or NK cells (d) versus H3K4me1 profiles of different cell types derived from scChIX-seq. Single: pseudobulk profiles generated by single incubation, unmixed; pseudobulk profiles deconvolved by scChIX-seq. (e-g) Pearson correlation between H3K27me3 sortChIC from FACS-purified B cells (e), granulocytes (f), NK cells (g) versus H3K27me3 sortChIC derived from pseudobulks of whole bone marrow without FACS purification. Single: pseudobulk profiles generated by

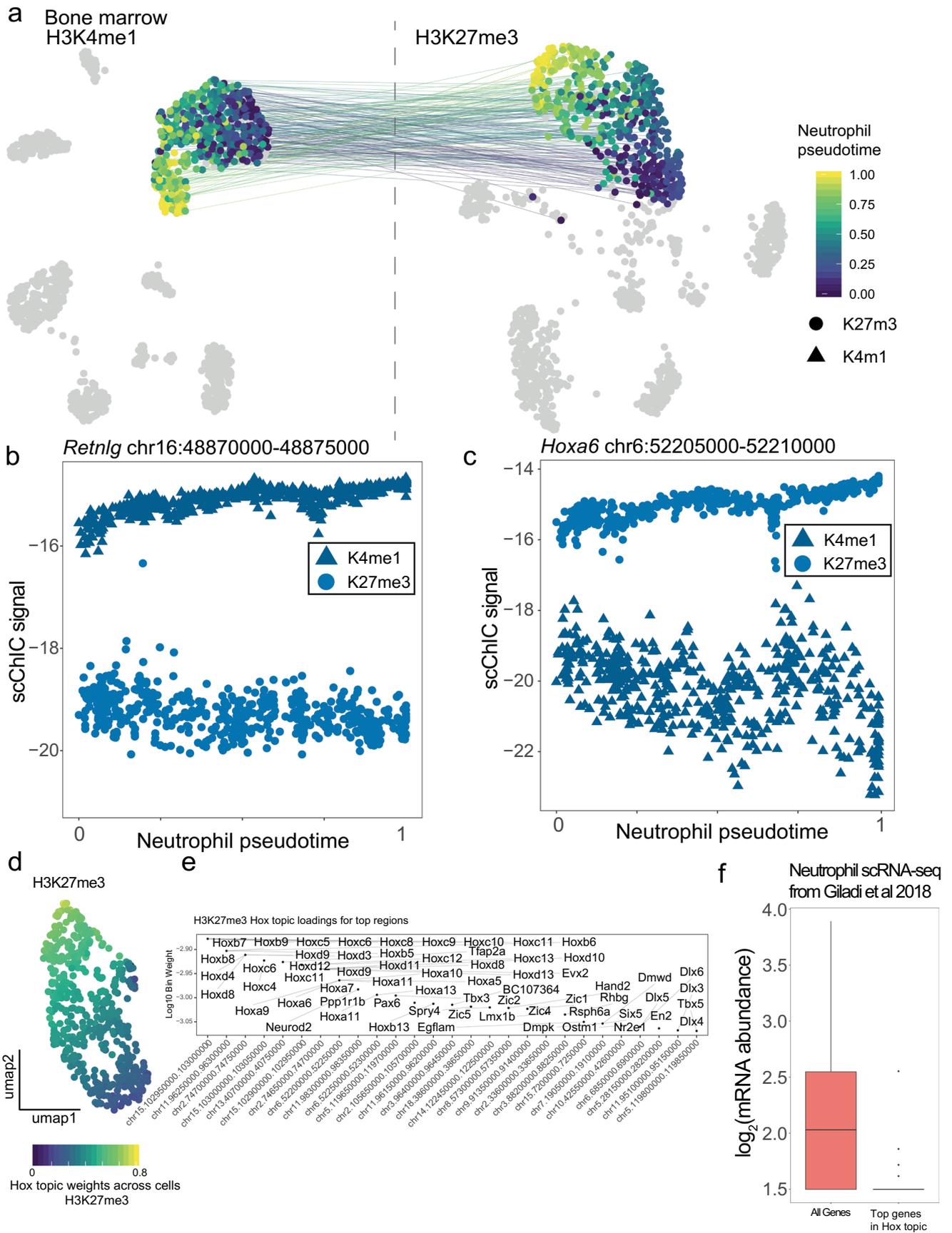
single incubation, unmixed; pseudobulk profiles deconvolved by scChIX-seq. (h) Distribution of assignment probability estimates p in the genome for the three cell types. Vertical dotted lines represent cutoffs for p to define H3K27me3-specific and H3K4me1-specific regions. p is the expected fraction of reads that belong to H3K4me1 in a specific genomic locus. (i) Boxplot distributions of GC content for the two classes of regions. (j) Boxplot distributions of distance to TSS in the two classes of regions. Distances are measured from the center of the 5 kb locus to the nearest TSS. Boxplots show 25th percentile, median and 75th percentile, with the whiskers spanning 97% of the data.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Re-clustering on B cells reveals heterogeneity within B cells. (a) UMAP visualization of H3K4me1 and H3K27me3 (single signal and unmixed signal), colored by cell types derived from H3K4me1 and transferred to H3K27me3. Black rectangle indicates the B cell population used to re-cluster in (b,c,d). (b) UMAP of pro-B and B cells only. (c,d) Projection of H3K4me1 signal

of marker genes for pro-B (c) or for differentiated B cells (d). H3K4me1 signal is measured in all cells of the H3K4me1 UMAP (that is both single- and double-incubated have H3K4me1 signal in the H3K4me1 UMAP). Double- (colored) but not single-incubated (grey) cells have H3K4me1 signal in the H3K27me3 UMAP.

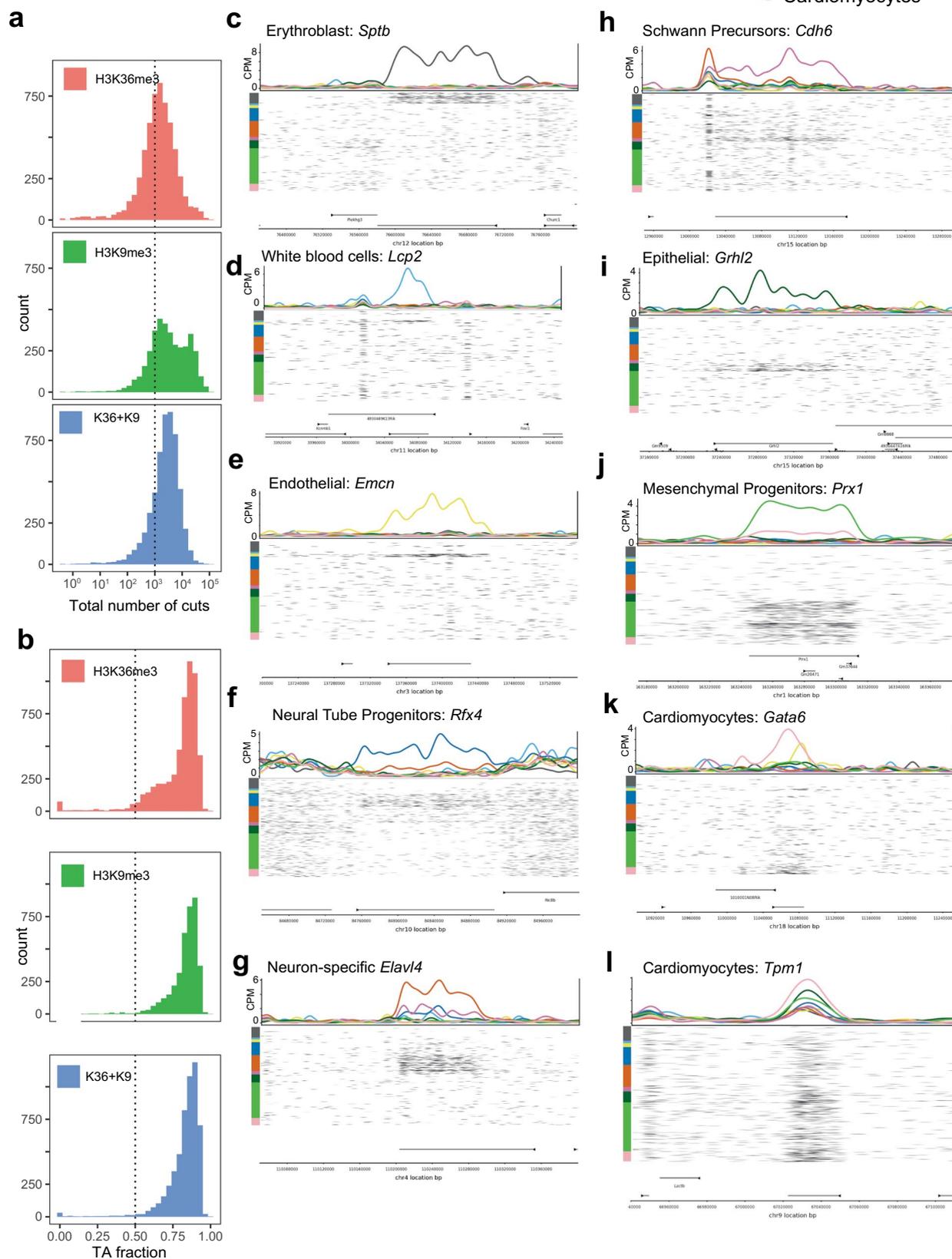


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | H3K4me1 and H3K27me3 signal during neutrophil maturation. (a) UMAP visualization of H3K4me1 and H3K27me3, lines join H3K4me1 and H3K27me3 UMAPs of double-incubated neutrophils. Heterogeneity within neutrophils are colored as neutrophil pseudotime. (b) H3K4me1 and H3K27me3 modification levels at the *Retnlg* (a mature neutrophil marker gene) locus along neutrophil pseudotime. (c) H3K4me1 and H3K27me3 modification levels at the *Hoxa* along neutrophil pseudotime. (d) UMAP of H3K27me3 signal across single cells colored by weights of a topic

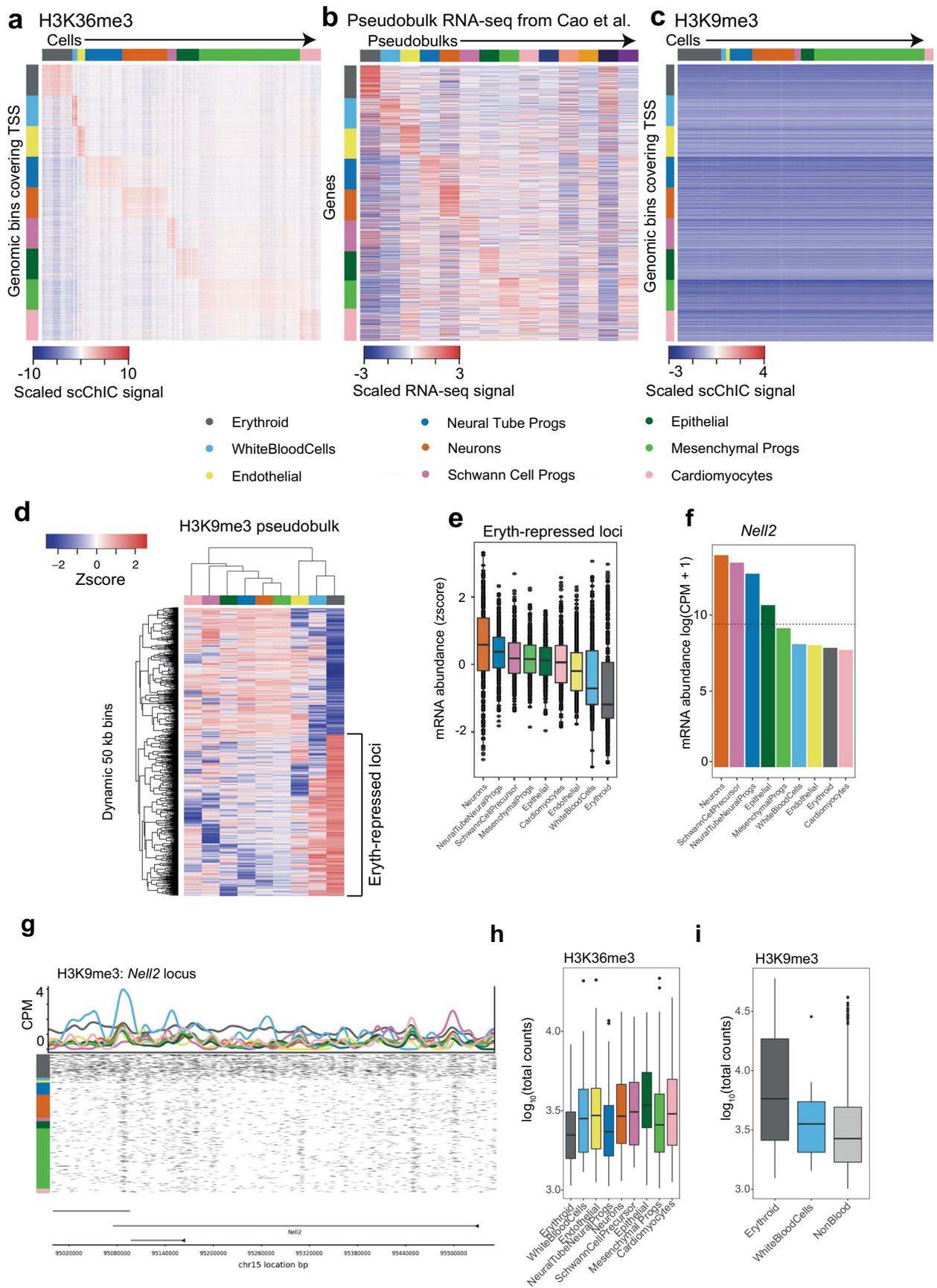
containing high H3K27me3 levels at many *Hox* and developmental gene loci (*Hox* topic). (e) Topic weights of the top 150 genes associated with loci in the *Hox* topic for H3K27me3. (f) Neutrophil mRNA abundance of genes in the *Hox* topic compared to other genes derived from publicly available scRNA-seq data²⁵. Number of genes per boxplot: n=17986 for All Genes, n=127 for genes in the *Hox* topic. Boxplots show 25th percentile, median and 75th percentile, with the whiskers spanning 97% of the data.

- Erythroid
- Endothelial
- Neurons
- Epithelial
- White Blood Cells
- Neural Tube Progs
- Schwann Precursors
- Mesenchymal Progs
- Cardiomyocytes



Extended Data Fig. 8 | Cell typing mouse organogenesis dataset using H3K36me3 using marker genes. (a) Histogram of unique fragment cuts per cell. (b) Histogram of fraction of unique fragments starting with a “TA” motif. (c-l)

Genome browser plots of cell type-specific H3K36me3 loci showing pseudobulk CPM signals (colored lines, top) and cut locations of individual cells (bottom, black marks). Cells are ordered by cell type (color-coded on the left).

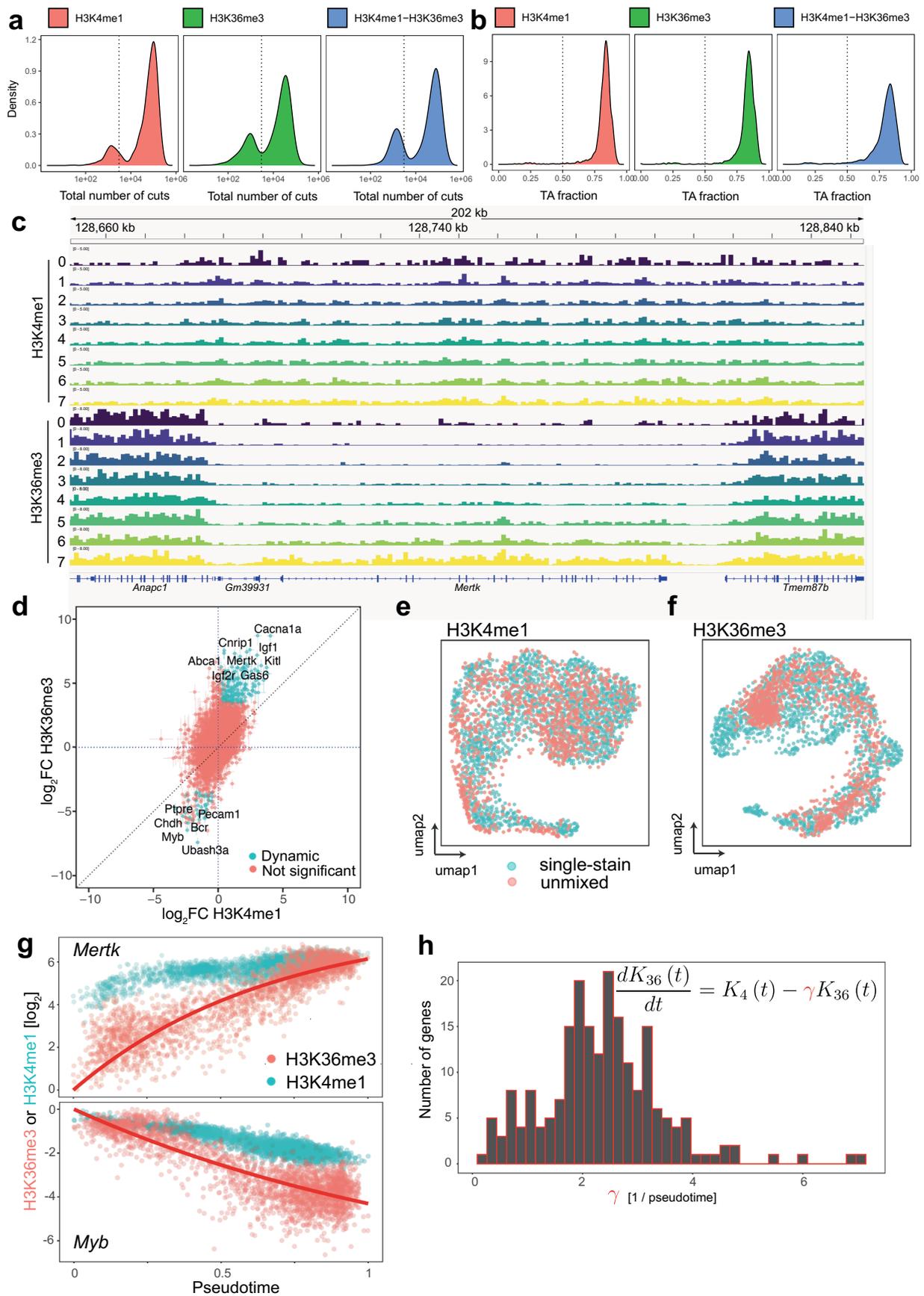


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | H3K9me3-specific regions across cell types.

(a) Heatmap of H3K36me3 signal for the top 250 H3K36me3-specific loci (rows) across cell types (columns). **(b)** Heatmap of mRNA abundances for the genes associated with the H3K36me3-specific loci in (a) across pseudobulks. Data processed from publicly available scRNA-seq data from Cao et al.⁴². **(c)** Heatmap of H3K9me3 signal for the same top 250 H3K36me3-specific loci as in (a). The H3K36me3 and H3K9me3 heatmaps are mean-centered and scaled using a common mean and standard deviation calculated across both marks. **(d)** Heatmap of H3K9me3 signal across pseudobulks at H3K9me3-variable loci. **(e)** Relative mRNA abundances⁴² at n=364 genes associated with erythroblast-repressed loci across nine cell types. **(f)** mRNA abundance of an erythroblast-repressed gene, *Nell2*, across pseudobulks. **(g)** Genome browser plot of around

the *Nell2* locus, an erythroblast-specific region for H3K9me3. Top of plot is pseudobulk H3K9me3 CPM signals, below are cut locations of individual cells (black marks). Cells are ordered by cell type (color-coded as in heatmaps). **(h, i)** Total unique fragments across cell types for single-incubated cells for H3K36me3 (h) and H3K9me3 (i), showing that the variability of the number of cuts across cells can span orders of magnitude. Number of single-incubated H3K36me3 cells for each boxplot: n=154 erythroid, n=36 white blood cells, n=60 endothelial, n=250 neural tube progenitors, n=272 neurons, n=58 Schwann cell precursors, n=154 epithelial, n=570 mesenchymal progenitors, n=160 cardiomyocytes. For H3K9me3: n=207 erythroid, n=26 white blood cells, n=736 non-blood cell types. Boxplots in (e), (h), (i) show 25th percentile, median and 75th percentile, with the whiskers spanning 97% of the data.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Distinct dynamics of H3K4me1 and H3K36me3 during macrophage in vitro differentiation. (a) Density plots of total number of cuts across cells for H3K4me1, H3K36me3, and H3K4me1+H3K36me3 labeled cells. (b) Density plots of fraction of cuts starting with a TA motif across cells for H3K4me1, H3K36me3, and H3K4me1+H3K36me3 labeled cells. (c) Genome-browser plot around gene body of *Mertk*, a macrophage-specific gene. Tracks are bigwigs from pseudobulks averaged across the time course. (d) Log₂ fold change estimates along pseudotime on gene bodies in the genome. Colored dots

are considered significant (\log_2 fold change in H3K36me3 > 3.5, zscore in both H3K36me3 and H3K4me1 > 2) and used for chromatin velocity estimates. (e, f) UMAP of H3K4me1 (e) and H3K36me3 (f) of single-incubated and deconvolved cells showing intermingling of the two types of cells. (g) Examples of H3K4me1 and H3K36me3 for an upregulated (above) and downregulated (below) gene along pseudotime. (h) Histogram of estimates of the rate constant γ for the 209 dynamic genes highlighted in (d).

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.

n/a Confirmed

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection BD FACS™ Software 1.2.0.142 was used to collect data from the FACS machine during cell sorting.

Data analysis Python package for processing fastq files into bams: <https://github.com/BuysDB/SingleCellMultiOmics>
R package associated with deconvolving the signal: <https://github.com/jakeyeung/scChX>

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

Data has been uploaded to Gene Expression Omnibus (GEO) accession number: GSE155280.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.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. The number of plates we used was based on our estimates of the number of different cell types we expected to see. For the ground truth experiment, we expected only a few cell types, and therefore used 9 plates (i.e. 3 plates per antibody condition). For organogenesis where we expected more cell types, we used 33 plates (13 plates for H3K36me3, 10 plates for H3K9me3, and 10 plates for dual-incubation). For macrophage in vitro differentiation, we used 8 plates per antibody condition.
Data exclusions	Cells that did not pass quality controls were excluded from the analysis. We removed all cells that had fewer than 50 percent of reads starting with a TA sequence (removed for low MNase specificity). In bone marrow H3K27me3 and H3K9me3 samples, we further removed cells that had fewer than 1000 unique cuts. In bone marrow H3K4me1, we removed cells that had fewer than 500 unique cuts. For mouse organogenesis, we removed cells that had fewer than 1000 unique cuts. For macrophage differentiation, we removed cells that had fewer than 3000 unique cuts.
Replication	We performed experiments across multiple plates and found the results across these technical replicates to be reproducible. When projecting cells across technical replicates onto a low-dimensional manifold, we did not observe effects coming from differences in technical replicates.
Randomization	We used blocking in the experimental plate design to reduce unexplained variability, within each block the cells were randomly assigned onto the plates. In the ground truth experiment, we minimized effects across plates by sorting different cell types onto the same plate. The location of each cell was not randomly assigned to the well on the plate. For macrophage in vitro differentiation, samples were collected over 7 days, but FACS sorting was done onto plates to pool 7 days of samples evenly onto the plates to reduce batch effect. For organogenesis experiment, different plates corresponded to different stages of development.
Blinding	No blinding was done because the experiments did not involve conditions that would induce a bias from the experimentalist.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	H3K4me1: rabbit anti-mouse H3K4me1, polyclonal, Ab8895, Lot: GR3206285-1, Abcam H3K27me3: rabbit anti-mouse H3K27me3, monoclonal, Identifier: 9733S, NEB H3K9me3: rabbit anti-mouse H3K9me3, polyclonal, Ab8898, Lot: GR3217826-1, Abcam H3K36me3: rabbit anti-mouse H3K36me3, monoclonal, clone: RM155, Merck GR1: A647, rat anti-mouse Ly-6G/Ly-6C, monoclonal, clone: RB6-8C5, Lot: 108420, Biolegend NK1: A488, rat anti-mouse anti NK-1.1, clone: PK136, Lot: 108717, Biolegend CD19: BC421, rat anti-mouse CD19, clone: 6D5, Lot: 11537, Biolegend Haematopoietic stem and progenitor enrichment pool: mix of biotinylated antibodies against CD5, CD11b, CD19, CD45R/B220, Ly6G/C(Gr-1), TER119, 7-4, part of #19856, Stemcell
Validation	We validated antibodies by performing sortChIC on K562 cells and confirmed that we reproduced the publicly available ChIP-seq

Validation

signals from the ENCODE project.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Male 13-week-old C57BL/6 mice were used to extract bone marrow cells. Embryos from E9.5, E10.5, and E11.5 were used to extract cells for mouse organogenesis study. Mice were kept in 12h:12h light:dark cycles in controlled ambient temperature and humidity, food and water ad libitum.

Wild animals

No wild animals were used in this study.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

Experimental procedures were approved by the Dier Experimenten Commissie of the Royal Netherlands Academy of Arts and Sciences and performed according to the guidelines.

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.

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155280>

GEO Accession number: GSE155280

Files in database submission

bam
bigwigs
processed count tables

Genome browser session

(e.g. [UCSC](#))

Bigwig files can be downloaded and directly viewed using IGV viewer

Methodology

Replicates

Validation experiments using ground truth cell types were performed across three technical plates. Experiments from whole bone marrow were performed across two technical plates.

Sequencing depth

In validation experiments, we sequenced to a mean depth of 48343 unique cut fragments per cell. In experiments from whole bone marrow, we sequenced to a mean depth of 9041 unique cut fragments per cell.

Antibodies

H3K4me1: rabbit anti-mouse H3K4me1, polyclonal, Ab8895, Lot: GR3206285-1, Abcam
H3K27me3: rabbit anti-mouse H3K27me3, monoclonal, Identifier: 9733S, NEB
H3K9me3: rabbit anti-mouse H3K9me3, polyclonal, Ab8898, Lot: GR3217826-1, Abcam
H3K36me3: rabbit anti-mouse H3K36me3, monoclonal, clone: RM155, Merck

Peak calling parameters

No peak calling was performed in this study.

Data quality

We removed all cells that had fewer than 50 percent of reads starting with a TA sequence (removed for low MNase specificity). For ground truth bone marrow study: H3K27me3 and H3K9me3 samples, we further removed cells that had fewer than 1000 unique cuts.
For whole bone marrow study: H3K27me3 and H3K4me1 used cut off of 1000 and 500 unique cuts, respectively.
For mouse organogenesis, we removed cells that had fewer than 1000 unique cuts.
For macrophage differentiation, we removed cells that had fewer than 3000 unique cuts.

Software

Python package for processing fastq files into bams: <https://github.com/BuysDB/SingleCellMultiOmics>
R package associated with deconvolving the signal: <https://github.com/jakeyeung/scChIX>

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

Ethanol fixed cells were thawed on ice. Cells were spun at 400 g for 5 minutes and washed once with 400 microlitre Wash Buffer 1 (47.5 ml H₂O RNAse free, 1 ml 1 M HEPES pH 7.5 (Invitrogen), 1.5 ml 5M NaCl, 3:6 µl pure spermidine solution (Sigma Aldrich), 0:05% saponin). Cells were spun again at 400 g and resuspended in 400 microlitre Wash Buffer 1. Cell suspension was split into 3 samples each having a volume of 400 microlitre and incubated with one or two antibodies (1:100 dilution for H3K27me3, H3K9me3 and H3K27me3+H3K9me3) overnight on a roller at 4 degrees Celsius. The next day cells were spun at 400 g, washed once with 400 microlitre Wash Buffer 2 and resuspended in 500 microlitre Wash Buffer 2 containing pA-MNase (3 ng/mL) and incubated for 1 hour on a rotator at 4 degrees Celsius. Next, cells were spun at 400 g and resuspended in 400 microliter Wash Buffer 2 (with addition of 5% blocking rat serum). Surface antibodies were added according to these concentrations and were incubated for 30 minutes on ice:

GR1 & A647, anti-mouse Ly-6G/Ly-6C (Gr-1) Antibody, clone: RB6-8C5 & 1:8000
 NK1 & A488, anti-mouse NK-1.1 Antibody, clone: PK136 & 1:400
 CD19 & BV421, anti-mouse CD19 Antibody, clone: 6D5 & 1:200

Finally, samples were washed once with 500 microlitre Wash Buffer 2 before passing them through a 70 micron cell strainer (Corning, 431751) and sorting on a BD Influx FACS machine, with surface antibody specific gating, into 384 well plates containing 50 nanoliter Wash buffer 3 (Wash buffer containing 0.05 % Tween) and 5 microlitre sterile filtered mineral oil (Sigma Aldrich) per well. Small volumes were distributed using a Nanodrop II system (Innovadyme).

Instrument

BD Influx System

Software

BD FACSTM Software 1.2.0.142

Cell population abundance

Purity of the sorted cell populations was assessed by performing scChIX and doing dimensionality reduction of chromatin levels across the genome in single cells.

Gating strategy

We used a forward scatter gate to remove debris (low FSC) and trigger pulse width to remove doublets (high trigger pulse width). We selected GR1+ cells as granulocytes, NK1.1+, GR1- cells as NK cells, and NK1.1-, GR1-, CD19+ cells are B cells.

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