HHS Public Access

Author manuscript

Nat Biotechnol. Author manuscript; available in PMC 2023 December 01.

Published in final edited form as:

Nat Biotechnol. 2023 June ; 41(6): 806–812. doi:10.1038/s41587-022-01588-5.

Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution

Tim Stuart[2,3], **Stephanie Hao**[1], **Bingjie Zhang**[2,3], **Levan Mekerishvili**[3,4], **Dan A Landau**[3,4], **Silas Maniatis**[1], **Rahul Satija**[2,3], **Ivan Raimondi**[1,4]

[1]Technology Innovation Lab, New York Genome Center, New York City, NY

[2]Center for Genomics and Systems Biology, New York University, New York City, NY

[3]New York Genome Center, New York City, NY

[4]Weill Cornell Medicine, New York City, NY

Abstract

Chromatin states are functionally defined by a complex combination of histone modifications, transcription factor binding, DNA accessibility, and other factors. Current methods for defining chromatin states cannot measure more than one aspect in a single experiment at single-cell resolution. Here, we introduce nanobody-tethered transposition followed by sequencing (NTT-seq), an assay capable of measuring the genome-wide presence of up to three histone modifications and protein-DNA binding sites at single-cell resolution. NTT-seq utilizes recombinant Tn5 transposase fused to a set of secondary nanobodies (nb). Each nb-Tn5 fusion protein specifically binds to different immunoglobulin-G antibodies, enabling a mixture of primary antibodies binding different epitopes to be used in a single experiment. We apply bulk- and single-cell NTT-seq to generate high-resolution multimodal maps of chromatin states in cell culture and in human immune cells. We also extend NTT-seq to enable simultaneous profiling of cell-surface protein expression and multimodal chromatin states to study cells of the immune system.

Several related methods were recently developed that enable individual aspects of chromatin state to be measured at single-cell resolution via an antibody-guided DNA tagmentation reaction (1–3). However, chromatin states are characterized by combinations of factors at an individual locus (4), including histone posttranslational modifications and the binding of non-histone proteins to the DNA. For example, promoters are commonly marked by

Correspondence to: iraimondi@nygenome.org.

Author contributions

I.R. conceived the study. I.R., S.H., B.Z., and L.M. performed experiments. T.S. and I.R. performed computational analysis. D.A.L., S.M., and R.S. supervised the study. T.S. and I.R. wrote the manuscript with input from all authors.

Competing interests

In the past 3 years, R.S. has worked as a consultant for Bristol-Myers Squibb, Regeneron and Kallyope and served as an SAB member for ImmunAI, Resolve Biosciences, Nanostring, and the NYC Pandemic Response Lab. I.R. and S.M. have filed a patent application based on this work (US Provisional Application No. 63/276,533). The remaining authors declare no competing interests.

Code availability

Signac 1.7.0 (14) and Seurat 4.1.0 (16) were used for all analysis, and are available from CRAN: https://cran.r-project.org/package=Signac; https://cran.r-project.org/package=Seurat. Code to reproduce analyses is available on GitHub: https://github.com/stuart-lab/nanobody. NTT-seq resources can be found at https://ntt-seq.com.

both H3K27ac and H3K4me2, whereas enhancers are marked by H3K27ac but typically lack H3K4me2. Furthermore, active and poised enhancers are both marked by H3K4me1 and can be distinguished by the presence of H3K27ac (5). Therefore, multimodal single-cell chromatin profiling methods are required to fully characterize chromatin states in heterogeneous tissues.

A majority of single-cell chromatin profiling methods employ protein-A/G fused to Tn5 transposase (1–3,6,7). Protein-A/G binds to IgG antibodies, enabling Tn5 to be directed to regions of the genome where an IgG antibody is bound and insert adapters for DNA sequencing. As protein-A/G binds to IgG antibodies from different species with high affinity, such methods are difficult to perform in an antibody-multiplexed design aiming to measure multiple histone modifications in a single experiment. Current approaches for multimodal chromatin profiling using protein-A/G, such as MulTI-Tag, involve complex experimental workflows with multiple wash and incubation steps (7). Such methods have not been demonstrated to work with complex tissues (6,7), thus limiting their broader application. We reasoned that the use of small single polypeptide chain antibodies (nanobodies) that specifically bind IgG from different species or different IgG subtypes in place of protein-A/G may enable the multiplexing of primary antibodies to facilitate a multimodal chromatin assay (8). Nanobodies bind strongly to their target epitope with dissociation constants (Kd) in the high picomolar-scale, whereas protein-A/G has Kd in the low nanomolar scale (9,10). Furthermore, nanobodies are stable under a broad temperature and pH range. We hypothesized that a nanobody-Tn5 (nb-Tn5) fusion would form a stable and specific protein-protein complex with a target primary IgG antibody.

In this study we engineered a set of nanobody-Tn5 fusion proteins and apply these fusion proteins in a multiplexed chromatin-profiling assay, measuring up to three distinct chromatin targets genome-wide simultaneously in single cells. We demonstrate the accuracy of multiplexed chromatin data obtained using our novel assay using cultured cells and human immune cells from the bone marrow and peripheral blood.

Results

We engineered and produced four different recombinant nb-Tn5 fusion proteins, specific for IgG antibodies from different species or IgG subtypes (Fig. 1A, Extended Data Fig. 1A). This included anti-mouse and anti-rabbit IgG nanobodies, as well as isotype-specific nanobodies for mouse IgG1 and IgG2a. Loading nb-Tn5 fusion proteins with barcoded DNA adaptor sequences enables the identity of individual nb-Tn5 fusion proteins that generated the sequenced DNA fragment to be determined through DNA sequencing.

We tested each recombinant nb-Tn5 fusion in a bulk-cell NTT-seq experiment and obtained an NTT-seq library only when the nb-Tn5 matched the target antibody, while the incubation of nb-Tn5 with the unmatched Ab resulted in no library amplification via PCR (Extended Data Fig. 1B). Motivated by this result, we performed multiplexed NTT-seq aiming to profile multiple different chromatin features in a single experiment. In our protocol, extracted nuclei are stained in a single step using primary antibodies for multiple epitopes simultaneously, the excess antibody is washed and nuclei are incubated with a mixture of

adapter-barcoded (11) nb-Tn5s, with each nb-Tn5 recognizing a specific IgG antibody. Subsequently, nb-Tn5s are activated by adding Mg^{2+} resulting in the tagmentation of genomic DNA in proximity of the primary antibody. The released DNA fragments harbor specific barcodes enabling the assignment of sequenced fragments to an individual nb-Tn5 and its associated primary antibody (Fig. 1B).

To test the targeting specificity of our species-specific nb-Tn5 fusion proteins, we used antibodies for H3K27me3 and H3K27ac in bulk human peripheral blood mononuclear cells (PBMCs), as these marks do not co-occur in the genome (12). Multiplexed NTT-seq resulted in libraries with nearly identical genomic distributions for each separate mark to matched NTT-seq performed on the same cells for each histone mark separately (Fig. 1C). The enrichment of sequenced fragments falling in H3K27me3 and H3K27ac peaks was approximately the same across the multiplexed and non-multiplexed experiments (Fig. 1D, E), and showed mutual exclusivity (Fig. 1F, G; Extended Data Fig. 1C). This suggests that multiplexed NTT-seq results in highly accurate localization of chromatin marks genome-wide. Then, we tested our isotype-specific nb-Tn5 profiling of three primary antibodies in a single experiment, repeating similar experiments using K562 cells staining with mouse IgG1 antibody against H3K27me3, mouse IgG2a antibody against H3K27ac, and including an additional rabbit IgG antibody for RNA Polymerase II (RNAPII) with phosphorylated Serine 2 and Serine 5 (elongating RNAPII, enriched on actively transcribed genes) (13). In comparison with a control experiment in which each of the three targets was profiled individually, multiplexed NTT-seq again produced comparable target enrichment specificity in peaks (Fig. 1H, I, J; Extended Data Fig. 1D), demonstrating the ability to profile three targets simultaneously, as well as the ability to profile non-histone proteins.

Encouraged by the results obtained in bulk cells, we next applied NTT-seq to characterize multimodal chromatin states at single-cell resolution using the 10x Genomics scATAC-seq kit (Fig. 2A). We profiled H3K27me3, H3K27ac and elongating RNAPII in a mixture of 8,617 K562 and HEK293 cells. We obtained on average 743 (s.d. 699) fragments for H3K27me3, 382 (s.d. 282) fragments for H3K27ac and 542 (s.d. 350) fragments for RNAPII per cell, outperforming the recently-developed multiCUT&Tag method (6) in terms of sensitivity and specificity (Extended Data Fig. 2A, B, C; Extended Data Table 1). We projected cells into a low-dimensional space using latent semantic indexing (LSI) and UMAP (14,15), and clustered cells using a weighted combination of all three data modalities (16) (Fig. 2B). We identified two groups of cells corresponding to K562 and HEK293 cells. The genomic distribution of reads for each mark obtained in the multiplexed single-cell experiment was highly similar to data from the same cell lines where each feature was profiled individually in bulk (Fig. 2C, Extended Data Fig. 2B). Examining the distribution of fragments at ATAC (17), H3K27me3, H3K27ac, and RNAPII peaks further showed the co-occupancy of RNAPII and H3K27ac in open chromatin regions, while the signal for H3K27me3 was mutually exclusive with the other profiled marks (Fig. 2D, E). Furthermore, multiplexed single-cell-derived signals were highly correlated with bulk-cell signal for each assay profiled individually (Fig. 2D). Using a combination of cellular modalities provided the strongest separation of the two cell types in low-dimension space. When constructing a neighbor graph, we observed a higher fraction of a cell's neighbors belonging to the same cell type as that cell when using multiple modalities (Fig. 2F). This highlights the value

of multimodal chromatin data in measuring cellular states, and together these results show that NTT-seq is an effective method for profiling multiple chromatin modalities at single-cell resolution.

We next sought to extend the NTT-seq method to enable simultaneous measurement of cell surface protein expression alongside multimodal chromatin states at single-cell resolution. Building on the recently-developed CUT&Tag-pro method (18), we stained a population of mobilized PBMCs with an oligonucleotide-conjugated panel of 173 antibodies targeting immune-relevant cell surface proteins. Cells were then crosslinked, permeabilized, and incubated with antibodies against H3K27me3 and H3K27ac, and our standard NTT-seq protocol followed to generate single-cell libraries. This resulted in a dataset of 4,684 cells with a mean of 2,854 H3K27me3 and 412 H3K27ac fragments per cell (s.d. 2,953, 356 respectively), with similar sensitivity and specificity to PBMC scCUT&Tag (19) (Extended Data Fig. 3A). We further quantified 690 antibody-derived tag (ADT) counts per cell (s.d. 613), achieving a sensitivity similar to the recently demonstrated scCUT&Tag-pro method (Extended Data Fig. 3B) (18). We clustered cells using a weighted combination of each modality (16) and annotated cell clusters based on their patterns of protein expression (Fig. 3A). Protein expression patterns were concordant with cell clusters determined from a chromatin-based clustering, and we observed uniform expression of CD3 in T cells, mutually exclusive expression of CD4 and CD8, expression of CD14 in monocytes, CD19 in B cells, and IL2RB in NK cells (Fig. 3B). Pseudobulk H3K27me3 and H3K27ac NTT-seq profiles were highly correlated with individual single-cell CUT&Tag-pro (18) profiles for human PBMCs for the same histone marks (Fig. 3C). Consistent with our previous results, we also observed an extremely low coefficient of determination (R^2=0.00028) between H3K27me3 and H3K27ac levels within peaks (Fig. 3D), further supporting the accuracy of multiplexed NTT-seq single-cell profiles when applied to complex tissues. We observed consistency between chromatin states and protein expression patterns for each cell type, supporting accurate cell-surface protein quantification. For example, the *PAX5* locus was repressed in non-B cells with low CD19 protein expression, and active in B cells with high CD19 expression (Fig. 3E). Similarly, the *CD33* locus was active in monocytes with high CD33 protein expression and repressed in B cells with low CD33 expression. To evaluate the accuracy of our cell type classifications and multimodal chromatin landscapes measured by NTT-seq, we compared the results of our single-cell NTT-seq experiment with FACS-sorted ChIP-seq profiles for CD14 monocytes, CD34+ CMPs, and B cells previously published by the ENCODE consortium (17). Pseudobulk profiles generated from our NTT-seq cell types recapitulated the expected cell-type-specific ENCODE ChIP-seq profiles (Extended Data Fig. 3C). To evaluate the reproducibility of single-cell chromatin profiles measured by scNTT-seq, we generated a second scNTT-seq dataset measuring H3K27me3 and H3K27ac in human PBMCs (Extended Data Fig. 3D). This dataset achieved a similar level of sensitivity and specificity (Extended Data Fig. 3E,F, Extended Data Table 1), and was highly correlated with the genome-wide chromatin profiles obtained in our first PBMC dataset (Extended Data Fig. 3G), supporting the reproducibility of the assay.

While cell-surface protein expression information provides a powerful method of studying immune cells, these methods are of limited value outside of the immunology field. To test whether a low-dimensional structure similar to that obtained using protein expression could

be learned using the chromatin data alone, we compared the neighbor graphs obtained using protein expression data to that obtained using individual or combined chromatin modalities. While individual chromatin marks were unable to faithfully recapitulate the low-dimensional structure observed when including protein expression data, the combination of H3K27me3 and H3K27ac modalities provided a similar low-dimensional neighbor structure (Fig. 3F). This again highlights the unique power of multimodal chromatin data in resolving cellular states, and indicates that multiplexed NTT-seq may be a powerful method capable of characterizing heterogeneous tissues without the need for cell surface protein measurements.

We next sought to apply NTT-seq in a complex tissue that contains differentiating cells to capture chromatin remodeling dynamics that shape cellular identity. We profiled H3K27me3 and H3K27ac in human bone marrow mononuclear cells (BMMCs) (Fig. 3G). This resulted in 5,236 cells with a mean of 1,217 and 326 fragments per cell for H3K27me3 and H3K27ac respectively (Fig. 3H, Extended Data Table 1). We annotated cell clusters using a combination of label transfer using an annotated BMMC scATAC-seq dataset (20,21) using the H3K27ac assay, and manual annotation inspecting the presence of active and repressive histone marks at key marker genes for each cell type. We identified the expected cell types present in the immune system, including hematopoietic stem and progenitor cells (HSPCs) (Fig. 3G). Consistent with results obtained using cells in culture and PBMCs, we observed mutual exclusivity between H3K27ac and H3K27me3 across regions of the genome for BMMCs, and a mean fraction of fragments in ENCODE peaks of 0.18 and 0.26 for H3K27me3 and H3K27ac, respectively (Extended Data Fig. 4A, B). To study how multimodal chromatin states may change during cell development, we ordered cells belonging to the B cell lineage, including HSPCs, common lymphoid progenitors (CLPs), pre-B, B, and plasma cells along a developmental pseudotime trajectory using Monocle 3 (22) (Fig. 3I).

While the H3K27ac data was more sparse than the H3K27me3 data, combining data from both modalities enabled a trajectory to be identified that revealed the expected ordering of cells in a trajectory leading from HSPCs through CLP, pre-B, B, and plasma cells. To identify regions of the genome that changed their H3K27me3 and H3K27ac state across this trajectory, we quantified fragment counts for each cell in 10 kb bins spanning the entire genome for each chromatin modality. We identified genome bins with signal correlated with pseudotime (Pearson correlation >0.2, Bonferroni-corrected p-value < 1e-08), and identified a set of 514 regions with opposing relationships between H3K27me3 and H3K27ac signal (>0.5 difference in Pearson correlation between the marks). Sorting these regions by the point at which they reached maximal H3K27me3 signal revealed an ordered sequence of sites that became repressed or activated during B cell development (Fig. 3J). The genome bin with the strongest gain in H3K27ac and loss of H3K27me3 signals across pseudotime was located at the *PAX5* promoter (H3K27me3 $r = -0.70$, H3K27ac $r = 0.53$), a B-cell-specific transcription factor. Of the 514 dynamic sites, we further identified 87 of these sites that displayed dynamic H3K27me3 and H3K27ac states across the B cell trajectory, but were static in their DNA accessibility profile ($|r| < 0.05$, Bonferroni-corrected $p > 0.01$), as quantified in an existing BMMC scATAC-seq dataset (20). This suggests that additional chromatin state dynamics can be identified using multimodal epigenomic data generated by scNTT-seq. Further experimental analysis will be required to fully characterize

the function of these chromatin-dynamic sites in B cell development. To systematically assess the cell-type-specific expression pattern of genes located near genomic bins that were repressed or activated along the B cell pseudotime trajectory, we examined a published scRNA-seq dataset for healthy human BMMCs. We identified the closest gene to each pseudotime-correlated genome bin, and classified these as activated (positive correlation between H3K27ac and pseudotime) or repressed (positive correlation between H3K27me3 and pseudotime). Examining the expression of repressed and activated genes in the scRNA-seq dataset revealed concordant patterns of gene expression, with chromatin-activated genes becoming expressed later in B cell development, and repressed genes being expressed in HSPCs but turned off later in B cell development ($p < 2.2e-16$, t-test; Fig. 3K).

Discussion

Together these analyses demonstrate that NTT-seq datasets provide accurate multimodal chromatin landscapes at single-cell resolution, contain sufficient information to identify major cell types and states in primary human tissues, provide profiles that reflect high-quality bulk ChIP-seq data (17), and can be generated in conjunction with accurate cell-surface protein expression measurements. Existing multimodal chromatin technologies require complex experimental workflows and have not been demonstrated to work with complex tissue samples (6,7), or are strictly limited in the chromatin states that they can measure (23). NTT-seq overcomes both of these limitations, providing a streamlined experimental workflow applicable to complex tissues.

Current limitations of this method, as well as other tagmentation-based chromatin profiling methods, include the need to perform tagmentation in high salt conditions to avoid open chromatin bias (1). This may preclude the measurement of some DNA binding proteins, including some transcription factors. Furthermore, the small number of currently available secondary nanobodies limits the number of different marks that can be profiled simultaneously.

We anticipate future reagent development and protocol improvements will enable us to scale NTT-seq to profiling of more than three marks simultaneously, and are actively working on the generation of additional nb-Tn5s targeting antibodies raised in different species such as goat, rat, sheep, guinea pig and multiple IgG isotypes within the same species. This will expand the portfolio of reagents for multimodal chromatin profiling. The application of computational integration methods (18,21) may also enable composite profiles for many aspects of chromatin states to be generated *in silico*, beyond what is feasible to measure in a single experiment. Moreover, we anticipate that the use of dual-barcoded nb-Tn5 can be implemented in our protocol to investigate intra-locus interactions between different chromatin features, such as bivalent promoters or enhancers. We believe that the simplicity with which NTT-seq achieves simultaneous profiling of chromatin features makes this approach particularly appealing, and could represent the standard for multifactorial chromatin mapping in the future.

Conclusions

In this study we have developed a novel multifactorial chromatin-profiling method, NTT-seq, capable of measuring the genome-wide distribution of up to three different chromatin marks in bulk- and single-cell samples. NTT-seq uses a set of engineered nanobody-Tn5 fusion proteins to guide Tn5 transposition to specific sites in the genome, where sequence-barcoded DNA sequencing adaptors are inserted by Tn5. Our results demonstrate the high accuracy of multiplexed chromatin profiles obtained by NTT-seq in comparison to non-multiplexed CUT&Tag or ChIP-seq experiments, compatibility with simultaneous cell-surface protein expression measurement, and the application of NTT-seq to human tissues.

Methods

Cell culture

K562 cells were acquired from ATCC (nos. CCL-243). HEK293FT cells were acquired from Thermo Fisher (no. R70007). HEK293FT cells were maintained at 37°C and 5% CO_2 in D10 medium (DMEM with high glucose and stabilized L-glutamine (Caisson, no. DML23) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher, no. 16000044)). K562 cells were maintained at 37°C and 5% CO_2 in R10 medium (RPMI with stabilized L-glutamine (Thermo Fisher, no. 11875119) supplemented with 10% FBS).

Primary cells acquisition and processing

Fresh mobilized peripheral blood mononuclear cells (PBMCs) used for scNTT-seq with cell surface protein measurement were isolated within 48 hours of blood collection utilizing a Ficoll (Thermo Fisher Scientific, #45-001-750) gradient according to manufacturer's recommendations and cryopreserved. Isolated mononuclear cells were thawed and stained according to standard procedures, beginning with resuspension in staining buffer (Biolegend, #420201) and incubation with Human TruStain FxC (10 minutes at 4°C; Biolegend, #422302) to block Fc receptor-mediated binding. Cells were then stained with a CD34-PE-Vio770 antibody (20 minutes at 4°C; Miltenyi Biotec, clone AC136, #130-113-180) and DAPI (Invitrogen, #D1306). The samples were then sorted for DAPI-negative, CD34-positive cells using a BD Influx cell sorter. Live CD34-positive and CD34-negative were mixed 1:10 and processed with NTT-seq. BMMCs and PBMCs profiled by scNTT-seq without cell surface protein measurement were purchased from AllCells. After thawing into DMEM with 10% FBS, the cells were spun down at 4°C for 5 min at 400 g and washed twice with PBS with 2% BSA. After centrifugation, the cell pellet was resuspended in staining buffer (2% BSA and 0.01% Tween in PBS).

Cloning of nb-Tn5 plasmid constructs

Previously published sequences coding for secondary nanobodies (8) were synthesized as a gene fragment (IDT) flanked by restriction enzyme sites NcoI and EcoRI. To replace protein-A with a nanobody, 3XFlag-pA-Tn5-Fl (addgene #124601) and gene fragments were digested with NcoI and EcoRI 1h at 37°C, ligated overnight at 16°C and subsequently transformed into competent cells (NEB C2992H).

Nanobody-Tn5 transposase production

The pTXB1-nbTn5 vector was transformed into BL21(DE3)-competent Escherichia coli cells (NEB, no. C2527), and nb-Tn5 was produced via intein purification with an affinity chitin-binding tag (24). 400 mL of Luria broth (LB) culture was grown at 37°C to optical density (OD600) = 0.6. nb-Tn5 expression was then induced with isopropyl-ß-d-thiogalactopyranoside (IPTG) 0.25 mM at 22°C 6 hours. After induction, cells were pelleted and then frozen at −80°C overnight. Cells were then lysed by sonication in 100 mL pf HEGX (20 mM HEPES-KOH pH 7.5, 0.8 M NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton X-100) with a protease inhibitor cocktail (Roche, no. 04693132001). The lysate was pelleted at 30,000g for 20 min at 4°C. The supernatant was transferred to a new tube, and 3 µL of neutralized 8.5% polyethylenimine (Sigma-Aldrich, P3143) was added dropwise to each 100 µL of bacterial extract, gently mixed and centrifuged at 30,000g for 30 min at 4°C to precipitate DNA. The supernatant was loaded on four 2 mL chitin columns (NEB, no. S6651S). Columns were washed with 10 mL of HEGX, then 1.5 mL of HEGX containing 100 mM DTT was added to the column with incubation for 48 h at 4°C to allow cleavage of nb-Tn5 from the intein tag. nb-Tn5 was eluted directly into two 30 kDa molecular-weight cutoff (MWCO) spin columns (Millipore, no. UFC903008) by the addition of 2 mL of HEGX. Protein was dialyzed in five dialysis steps using 15 mL of 2x dialysis buffer (100 HEPES-KOH pH 7.2, 0.2 M NaCl, 0.2 mM EDTA, 2 mM DTT, 20% glycerol) and concentrated to 1 mL by centrifugation at 5,000g. The protein concentrate was transferred to a new tube and mixed with an equal volume of 100% glycerol. nb-Tn5 aliquots were stored at −80°C.

Transposome assembly

We obtained barcoded Tn5 adaptors from IDT, as described by Amini et al. (11) with 8 bp barcode sequences designed using FreeBarcodes (25). To produce mosaic-end, double-stranded (MEDS) oligos, we annealed each barcoded T5 tagmentation oligo with the pMENT common oligo (100 µM each) as follows, in TE buffer: 95°C for 5 min then cooling at 0.2°C per second to 4°C (bcMEDS-A). The same process was used to anneal a single T7 tagment oligo with the pMENT common oligo (MEDS-B; Extended Data Table 2). bcMEDS-A and MEDS-B were mixed 1:1 and 6 µL was transferred to a new tube and mixed with 10 µL of nb-Tn5 enzyme. After 1 hour at room temperature to allow for transposome assembly. Adapter sequences are shown in Extended Data Table 2.

Antibodies

Antibodies used were H3K27ac (1:50, Active Motif, 39133), H3K27ac (1:50, Active Motif, 91193), H3K27ac (1:50, AbCam, ab4729), H3K27me3 (1:50, Active Motif, 61017), Phospho-Rpb1 CTD (Ser2/Ser5) (1:50, Cell Signaling, 13546). For NTT-seq with surface markers readout on primary cells, the TotalSeq-A conjugated Human Universal Cocktail v1.0 panel was obtained from BioLegend (399907).

NTT-seq

We performed NTT-seq using similar methods to those described previously by Kaya-Okur et al. (1) (dx.doi.org/10.17504/protocols.io.bcuhiwt6), described in detail below.

Antibody staining

For NTT-seq with surface markers readout on primary cells, 1 million thawed PBMCs were resuspended in 200 µL staining buffer (2% BSA and 0.01% Tween in PBS) and incubated for 15 min with 20 µL Fc receptor block (TruStain FcX, BioLegend) on ice. Cells were then washed three times with 1 mL staining buffer and pooled together. The panel of oligo-conjguated antibodies was added to the cells to incubate for 30 min on ice. After staining, cells were washed three times with 1 mL staining buffer and resuspended in 100 µL staining buffer. After the final wash, cells were resuspended 200 µL PBS ready for fixation.

Fixation and permeabilization

For human cell lines, nuclei were extracted as previously described (26) and resuspended in 150 µL of PBS. Then, 16% methanol-free formaldehyde (Thermo Fisher Scientific, PI28906) was added for fixation (final concentration: 0.1%) at room temperature for 3 min. The cross-linking reaction was stopped by addition of 12 µL 1.25 M glycine solution. Subsequently, nuclei were washed once with 150 µL antibody buffer (20 mM HEPES pH 7.6, 150 mM NaCl, 2 mM EDTA, 0.5 mM spermidine, 1% BSA, 1× protease inhibitors).

For NTT-seq on PBMCs and BMMCs, 16% methanol-free formaldehyde (Thermo Fisher Scientific, PI28906) was added for fixation (final concentration: 0.1%) at room temperature for 5 min. The cross-linking reaction was stopped by addition of 12 µL 1.25 M glycine solution. Subsequently, cells were washed twice with PBS. The permeabilization was performed by adding isotonic lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM $MgCl_2$, 0.1% NP40, 0.1% Tween-20, 1% BSA, 1× protease inhibitors) on ice for 7 min. Subsequently, 1 mL of cold wash buffer (20 mM HEPES pH 7.6, 150 mM NaCl, 0.5 mM spermidine, 1× protease inhibitors) was added, and cells were centrifuged at 800g for 5 min at 4°C.

Tagmentation

Nuclei or permeabilized cells were directly suspended with 150 µL antibody buffer (20 mM HEPES pH 7.6, 150 mM NaCl, 2 mM EDTA, 0.5 mM spermidine, 1% BSA, 1× protease inhibitors) with a cocktail of primary antibodies and incubated overnight on a rotator at 4°C. The next day cells were washed twice with 150 µL wash buffer to remove the remaining antibodies. The cells were then resuspended in 150 µL high salt wash buffer (20 mM HEPES pH 7.6, 300 mM NaCl, 0.5 mM spermidine, 1× protease inhibitors) with 2.5 µL nb-Tn5 for each target of interest and incubated for 1 h on a rotator at room temperature. The cells were then washed twice with high salt wash buffer and resuspended in 50 µL tagmentation buffer (20 mM HEPES pH 7.6, 300 mM NaCl, 0.5 mM spermidine, 10 mM $MgCl_2$, 1× protease inhibitors). The samples were incubated for 1 h at 37°C. Tagmentation steps were performed in 0.2 mL tubes to minimize cell loss.

NTT-seq Bulk

To stop tagmentation, 1 µL of 0.5 M EDTA, 1 µL of 10% SDS and 0.25µL of 20 mg/mL Proteinase K was added to the sample, incubated at 55°C for 1 hour. DNA was extracted with Chip DNA clean & Concentrator kit (Zymo Research, D5201) following manufacturer instructions. To amplify libraries, 21 µL DNA was mixed with 2 µL of a universal i5 and

a uniquely barcoded i7 primer, using a different barcode for each sample. A volume of 25 μL NEBNext HiFi 2× PCR Master mix was added and mixed. The sample was placed in a Thermocycler with a heated lid using the following cycling conditions: 72°C for 5 min (gap filling); 98°C for 30 s; 14 cycles of 98°C for 10 s and 63°C for 30 s; final extension at 72°C for 1 min and hold at 8°C. Post-PCR clean-up was performed by adding 1.1× volume of Ampure XP beads (Beckman Coulter), and libraries were incubated with beads for 15 min at RT, washed twice gently in 80% ethanol, and eluted in 30 μL 10 mM Tris pH 8.0.

NTT-seq single cell encapsulation, PCR, and library construction

After tagmentation, cells were centrifuged for 5 min at 1,000g and the supernatant was discarded. Cells were resuspended with 30 μL 1× Diluted Nuclei Buffer (10x Genomics, #2000207), counted, and diluted to a concentration based on the targeted cell number. The transposed cell mix was prepared as following: 7 μL of ATAC buffer and 8 μL cells in 1× Diluted Nuclei Buffer. All remaining steps were performed according to the 10x Chromium Single Cell ATAC protocol. For NTT-seq with surface markers readout on primary cells, the library construction method was adapted from ASAP-seq (27). Briefly, 0.5 μL of 1 μM bridge oligo A (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNVTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT/3InvdT/) was added to the barcoding mix. Linear amplification was performing using the following PCR program: (40°C for 5 min, 72°C for 5 min, 98°C for 30 s; 12 cycles of 98°C for 10 s, 59°C for 30 s and 72°C for 1 min; ending with hold at 15°C). The remaining steps were performed according to the 10x Genomics scATAC-seq protocol (v1.1), with the following additional modifications:

Antibody-derived tags: during silane bead elution (Step 3.1s), beads were eluted in 43.5 μL of elution solution I. The extra 3 μL was used for the surface protein tags library. During SPRI cleanup (Step 3.2d), the supernatant was saved and the short DNA derived from antibody oligos was purified with 2x SPRI beads. The eluted DNA was combined with the 3 μL left aside after the silane purification to be used as input for protein tag amplification. PCR was set up to generate the protein tag library with Kapa Hifi Master Mix (P5 and RPI-x primers): 95°C for 3 min; 14–16 cycles of 95°C for 20 s, 60°C for 30 s and 72°C for 20 s; followed by 72°C for 5 min and ending with hold at 4°C.

RPI-x primer:

CAAGCAGAAGACGGCATACGAGATxxxxxxxxGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA.

P5 Primer:

AATGATACGGCGACCACCGAGATCTACAC

Sequencing

The final libraries were sequenced on NextSeq 550 by using custom primers (Extended Data Table 2) with the following strategy: i5: 38bp, i7: 8bp, read1: 60bp, read2: 60bp (for PBMC single-cell NTT-seq without cell surface proteins, read1: 50bp, read2: 50bp).

Bulk-cell data analysis

Bulk-cell data for the cell culture and PBMC datasets were mapped to the hg38 analysis set using bwa-mem2 with default parameters (28). Output BAM files were sorted and indexed using samtools (29), and bigwig files created using the deeptools bamCoverage function with the –normalizeUsing BPM option set. Fragment files were created using the Sinto (https://github.com/timoast/sinto), which uses the Pysam and htslib packages (29). Multi-NTT-seq heatmaps were generated in DeepTools (30). ChIP-seq peak coordinates for H3K27me3 and H3K27ac for bulk PBMCs, and for H3K27me3, H3K27ac, and RNAPII serine-2 and serine-5 phosphate for K562 cells were downloaded from ENCODE (17). We counted sequenced DNA fragments falling within each peak region for each bulk-cell PBMC or K562-cell NTT-seq dataset using custom R code and the scanTabix function in Rsamtools, and normalized counts according to the total number of mapped reads for each dataset (counts per million mapped reads normalization). The coefficient of determination (R^2) between peak counts across pairs of experiments was computed using the lm function in R.

Single-cell data analysis

Cell culture dataset

Read mapping: Reads were mapped to the hg38 analysis set using bwa-mem2 (28) with default parameters, the output sorted and indexed using samtools (29), and the resulting BAM file used to create a fragment file using the Sinto package (https://github.com/timoast/sinto). We ran the sinto fragments command with the --barcode_regex "[^:]*" parameter set to extract cell barcodes from the read name. Output files were coordinate-sorted, bgzip-compressed and indexed using tabix (31), and the resulting fragment files used as input to downstream analyses.

Quantification, quality control, and dimension reduction: Genomic regions were quantified using the AggregateTiles function in Signac (14) with binsize=10000 and min_counts=1, using the hg38 genome. Cells with <10,000 total counts, >75 H3K27ac counts, >150 H3K27me3 counts, and >100 RNAPII counts were retained for further analysis. Each assay was processed by performing TF-IDF normalization on the count matrix for the assay, followed by latent semantic indexing (LSI) using the RunTFIDF and RunSVD functions in Signac with default parameters. Two-dimensional visualizations were created for each assay using UMAP, using LSI dimensions 2 to 10 for each assay. Weighted nearest neighbor (WNN) analysis was performed using the FindMultiModalNeighbors function in Seurat, with reduction.list = list("lsi.k27ac", "lsi.k27me", "lsi.pol2") and dims = list(2:10, 2:10, 2:10) to use LSI dimensions 2 to 10 for each assay. Cell clustering was performed using the resulting WNN graph using the Smart Local Moving community detection algorithm (32) by running the FindClusters function in Seurat, with algorithm=3, graph.name="wsnn", and resolution=0.05. This resulted in two cell clusters, which were assigned as HEK or K562 based on their correlation with bulk-cell chromatin data for HEK and K562 cells.

Specificity analysis: K562-cell bulk ChIP-seq peaks for H3K27ac, H3K27me3, and RNA Pol2 Ser-2 and Ser-5 phosphate were downloaded from ENCODE (17). Since the fraction of reads in peaks metric can be sensitive to the peak set used, we opted to use previously-reported ENCODE peaks throughout our analysis as much as possible. Ser-2 and Ser-5 phosphate peaks were combined using the reduce function from the GenomicRanges R package. Fragment counts for K562 cells in the bulk and single-cell dataset were quantified for each peak using the scanTabix function in the Rsamtools R package, with counts normalized according to the total sequencing depth for each dataset. To assess the targeting specificity in single-cell NTT-seq, we computed the coefficient of determination (R^2) between peak counts for each pair of assays, and between bulk and single-cell data for the same assay. We visualized relative peak counts for each assay for each peak by creating a ternary plot using the ggtern R package (33). To assess the low-dimensional neighbor structure obtained using each assay or combinations of assays, we computed the fraction of *k*-nearest neighbors for each cell *i* that belonged to the same cell type classification as cell *i* (*k*=50 for single-modality neighborhoods, variable *k* per-cell for multimodal neighbor graph due to the weighted nearest neighbor method).

multi-CUT&Tag comparison: To create a fragment file for the published multi-CUT&Tag dataset, raw sequencing data from Gopalan et al. (6) were downloaded from NCBI SRA and split into separate FASTQ files according to their Tn5 barcode using a custom Python script. Reads were mapped to the hg38 genome using bwa-mem2 and fragment files created as described above for the NTT-seq datasets. Code to reproduce this analysis is available on GitHub: https://github.com/timoast/multi-ct. We ran the CountFragments function in Signac to count the total number of fragments per cell for each multi-CUT&Tag assay, and retained cells with >200 total counts for further analysis, as described in the original publication (6). For mixed-barcode fragments we counted ½ count to the total of each assay matching the pair of Tn5 barcodes. To compute the targeting specificity, we downloaded published ENCODE ChIP-seq peaks for H3K27me3 and H3K27ac for mESCs (ENCFF008XKX and ENCFF360VIS), and computed the fraction of fragments in peak regions using the scanTabix function in the Rsamtools R package, normalizing counts according to the total sequencing depth for the dataset. We also computed the R^2 between H3K27me3 and H3K27ac as described above, using the ENCODE peak regions.

PBMC datasets

Read mapping: Genomic reads were mapped and processed as described above for the cell culture single-cell dataset. Antibody-derived tag (ADT) reads were processed using Alevin (34). We first created a salmon index (35) for the BioLegend TotalSeq-A antibody panel, with the --features -k7 parameters. We quantified counts for each ADT barcode using the salmon alevin command with the following parameters: --naiveEqclass, --keepCBFraction 0.8, --bc-geometry 1[1-16], --umi-geometry 2[1-10], --read-geometry 2[71-85].

Quantification, quality control, and dimension reduction: Genomic bins were quantified using the AggregateTiles function in Signac, with binsize=5000 and min_counts=1 to quantify 5 kb bins genome-wide, retaining bins with at least one count. We retained cells with <40,000 and >300 H3K27me3 counts, <10,000 and >100 H3K27ac counts,

and <10,000 and >100 antibody-derived tag (ADT) counts. We normalized the ADT data using a centered log ratio transformation using the NormalizeData function in Seurat, with normalization.method="CLR" and margin=2. We reduced the dimensionality of the ADT assay by first scaling and centering the protein expression values, and running PCA (ScaleData and RunPCA functions in Seurat). We computed a 2-dimensional UMAP visualization using the first 40 principal components (PCs), and clustered cells using the Louvain community detection algorithm. We identified and removed two low-quality clusters containing higher overall ADT counts, as well as higher counts for naive IgG antibodies included in the staining panel. After removing low-quality ADT clusters, we reduced the dimensionality of the H3K27me3 and H3K27ac assays using LSI (FindTopFeatures, RunTFIDF, RunSVD functions in Signac) and created 2-dimensional UMAPs using LSI dimensions 2 to 30 for each chromatin assay. To construct a low-dimensional representation using all three data modalities, we ran the weighted nearest neighbors (WNN) algorithm, using the first 40 ADT PCs, and LSI dimensions 2 to 30 for H3K27me3 and H3K27ac (FindMultiModalNeighbors function in Seurat). We clustered cells using the WNN neighbor graph using the Smart Local Moving algorithm(32) (FindClusters function in Seurat with algorithm=3 and resolution=1). Cell clusters were manually annotated as cell types using the protein expression information. To compare the low-dimensional structure obtained using individual chromatin modalities or combinations of modalities, we computed for each cell *i* the fraction of neighboring cells annotated as the same cell type as cell *i*. We repeated this computation using neighbor graphs computed using single data modalities, or weighted combinations of modalities computed using the WNN method.

ENCODE data comparison: Peaks and genomic coverage bigWig files for H3K27me3 and H3K27ac ChIP-seq published by the ENCODE consortium (17) for B cells, CD34+ CMPs, and CD14+ monocytes were downloaded from the ENCODE website (https://www.encodeproject.org/). We created bigWig files for each corresponding cell type identified in the single-cell multiplexed NTT-seq PBMC dataset by writing sequenced fragments for those cells to a separate BED file, creating a bedGraph file using the bedtools genomecov command(36), and creating a bigWig file using the UCSC bedGraphToBigWig tool. We computed the genomic coverage for NTT-seq datasets and ChIP-seq datasets within H3K27me3 and H3K27ac regions using the deeptools multiBigwigSummary function (30) with the –outRawCounts option set to output the raw correlation matrix as a text file. We computed the correlation between peak region coverage in NTT-seq and ENCODE ChIP-seq datasets using the cor function in R with method="spearman". We computed the fraction of fragments per cell falling in ENCODE H3K27me3 and H3K27ac ChIP-seq peak regions for PBMCs for each assay as described above.

CUT&Tag-pro data comparison: Processed CUT&Tag-pro H3K27me3 and H3K27ac datasets for human PBMCs were downloaded from Zenodo: https://zenodo.org/record/5504061. We compared the number of antibody-derived tag (ADT) counts in NTT-seq and scCUT&Tag-pro datasets by extracting the total number of ADT counts per cell from the scCUT&Tag-pro and NTT-seq Seurat objects and plotting the distribution of total ADT counts per cell for each dataset. We created bigWig files for each scCUT&Tag-pro dataset

by first creating a bedGraph file using the bedtools genomecov function, and then creating a bigWig file using the UCSC bedGraphToBigWig function. We computed the coverage for scCUT&Tag-pro datasets within H3K27me3 and H3K27ac PBMC ENCODE peaks using the multiBigwigSummary function in deeptools as described above for the ENCODE data comparison.

BMMC dataset

Read mapping: Raw genomic reads were mapped and processed as described above for the cell culture single-cell dataset.

Quantification, quality control, and dimension reduction: Genomic bins were quantified using the AggregateTiles function in Signac, with binsize=5000 and min_counts=1 to quantify 5 kb bins genome-wide, retaining bins with at least one count. We retained cells with <10,000 and >100 H3K27me3 counts, and <10,000 and >75 H3K27ac counts for further analysis. We normalized the counts and reduced dimensionality for each assay by running the RunTFIDF, RunSVD, and RunUMAP functions in Signac and Seurat for each assay. We computed a WNN graph for H3K27me3 and H3K27ac using the FindMultiModalNeighbors function in Seurat, with reduction=list("lsi.me3", "lsi.ac") and dims.list=list(2:50, 2:80) to use LSI dimensions 2 to 50 and 2 to 80 for H3K27me3 and H3K27ac, respectively. A 2-dimensional UMAP was created using the WNN graph by running the RunUMAP function in Seurat with nn.name="weighted.nn" to use the pre-computed neighbor graph. We clustered cells using the WNN graph using the Smart Local Moving community detection algorithm (32) (FindClusters function in Seurat with algorithm=3, resolution=3, graph.name="wsnn"). We computed the fraction of fragments per cell falling in ENCODE PBMC H3K27me3 and H3K27ac ChIP-seq peak regions for each assay as described above.

Cell annotation: To annotate cell types we performed label transfer (21) using the H3K27ac assay and a previously published scATAC-seq dataset containing healthy human bone marrow cells (20). As the original publication mapped reads to the hg19 genome, we re-processed the original reads using the 10x Genomics cellranger-atac v2 software with default parameters, aligning to the hg38 genome. Code to reproduce this analysis is available on GitHub: https://github.com/timoast/MPAL-hg38. To transfer cell type labels from the scATAC-seq dataset to our multimodal NTT-seq dataset, we quantified scATAC-seq peaks using the H3K27ac assay, then performed TF-IDF normalization on the resulting count matrix using the IDF value from the scATAC-seq dataset. We performed LSI on the scATAC-seq BMMC dataset using the RunTFIDF and RunSVD functions in Signac with default parameters. We next ran the FindTransferAnchors function in Seurat, with reduction="lsiproject", dims=2:30, and reference.reduction="lsi" to project the query data onto the reference scATAC-seq LSI using dimensions 2 to 30, and find anchors between the reference and query dataset. We ran TransferData with weight.reduction=bmmc_ntt[["lsi.me3"]] dims=2:50 to weight anchors using LSI dimensions 2 to 50 from the H3K27me3 assay. We used these unsupervised cell type predictions as a guide when assigning cell clusters to cell types.

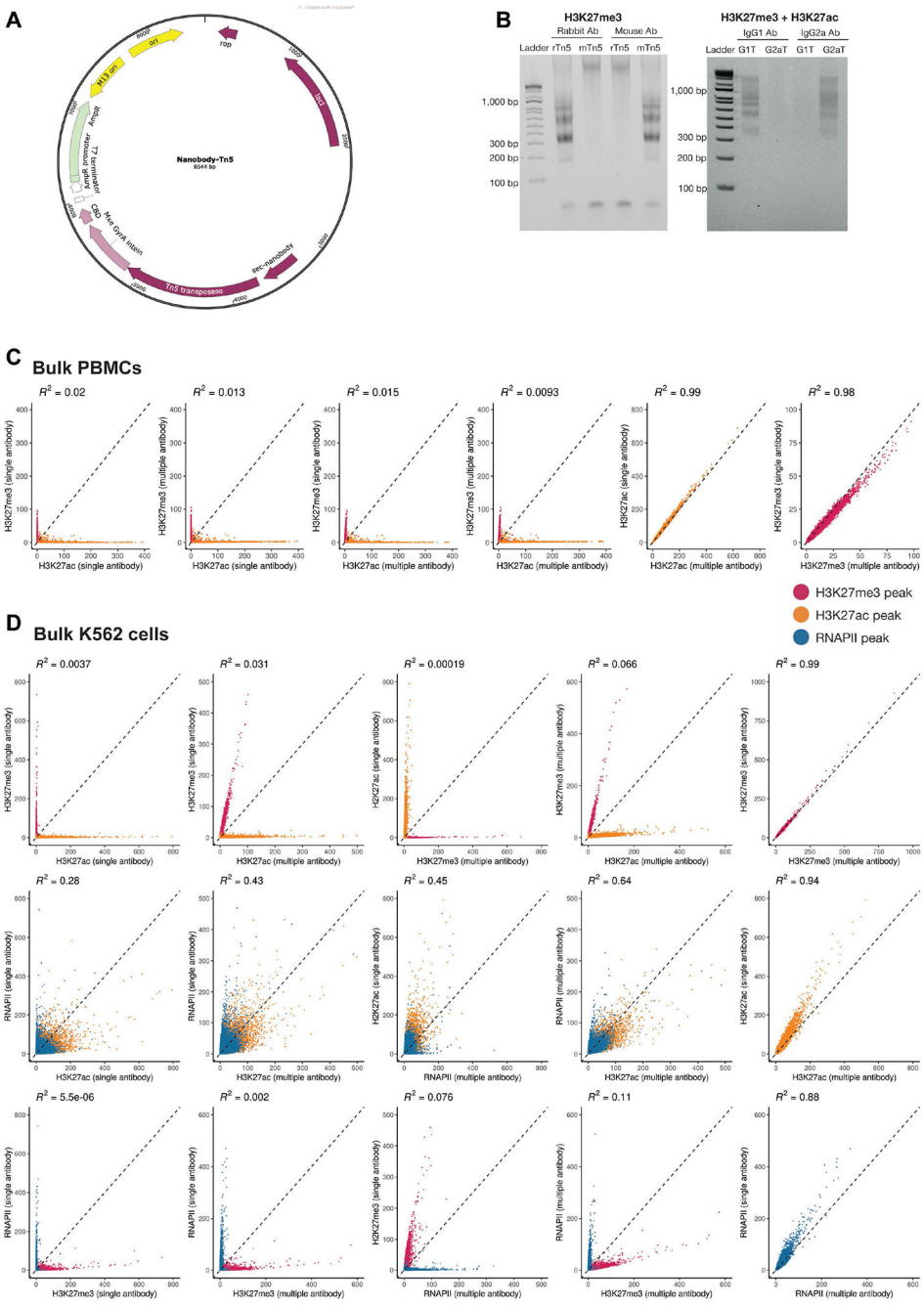
Trajectory analysis: We subsetted the BMMC dataset to contain cells annotated as HSPC, GMP/CMP, Pre-B, B, or Plasma cells. Using the subset object, we constructed a new UMAP dimension reduction by running FindTopFeatures, RunTFIDF, and RunSVD in Signac, followed by RunUMAP in Seurat with reduction="lsi", for each assay. We then constructed a joint low-dimensional space using the WNN method by running the FindMultiModalNeighbors function in Seurat. We converted the Seurat object containing these cells to a SingleCellExperiment object using the as.cell_data_set function in the SeuratWrappers package (https://github.com/satijalab/seurat-wrappers). We next ran Monocle 3 (22) using the pre-computed UMAP dimension reduction constructed using both chromatin modalities by running the cluster_cells, learn_graph, and order_cells functions, setting the HSPC cells as the root of the trajectory. To find genomic features in each assay whose signal depended on pseudotime state, we quantified fragment counts for each cell in each 10 kb genome bin for the H3K27me3 and H3K27ac assays. To reduce the sparsity of the measured signal, we averaged counts for each genomic region across the cell's 50 nearest neighbors, defined using the H3K27me3 neighbor graph with LSI dimensions 2 to 20, and normalized the fragment counts by the total neighbor-averaged counts per cell. For each genomic region we computed the Pearson correlation between the signal in the genomic region and the cell's position in pseudotime. To find regions that underwent coordinated activation or repression we selected regions with a Pearson correlation >0.2 or <−0.2 and a difference in Pearson correlation between the H3K27me3 and H3K27ac assays greater than 0.5 (e.g., −0.25 correlation for H3K27me3 and +0.25 for H3K27ac). To display genomic regions in a heatmap representation we ordered cells based on their pseudotime rank and ordered genomic regions based on the position in pseudotime showing maximal H3K27me3 signal. For the purpose of visualization, we smoothed the signal for each genomic region by applying a rolling sum function with cells ordered based on pseudotime, summing the signal over 100-cell windows. This was performed using the roll_sum function in the RcppRoll R package (version 0.3.0).

We used the ClosestFeature function in Signac to identify the closest gene to each genomic region correlated with pseudotime. Genomic regions where the closest gene was >50,000 bp away were removed (21 genes for H3K27me3 and 7 genes for H3K27ac). To examine the gene expression patterns of these genes, we downloaded a previously integrated and annotated scRNA-seq dataset for the human bone marrow, produced as part of the HuBMAP consortium (https://zenodo.org/record/5521512) (20,37,38). We subset the scRNA-seq object to contain the same cell states that we examined in the NTT-seq data (HSC, LMPP, CLP, pro-B, pre-B, transitional B, naive B, mature B, plasma) and computed a gene module score for the active and repressed genes using the AddModuleScore function in Seurat.

To compare changes in scATAC-seq signal across the B cell developmental trajectory, we also downloaded a previously published BMMC scATAC-seq dataset (20), and subset the cells belonging to the B cell trajectory using the published cell type annotations provided by the original authors. We quantified the same set of genomic regions used in the scNTT-seq BMMC analysis, and created a similar B cell developmental trajectory by assigning a numeric value to each B cell type according to its relative position along the known

developmental trajectory (1 = HSC, 2 = CMP/LMPP, 3 = CLP, 4 = B, 5 = Plasma), and computed the Pearson correlation between each genomic region and the B cell trajectory.

Extended Data



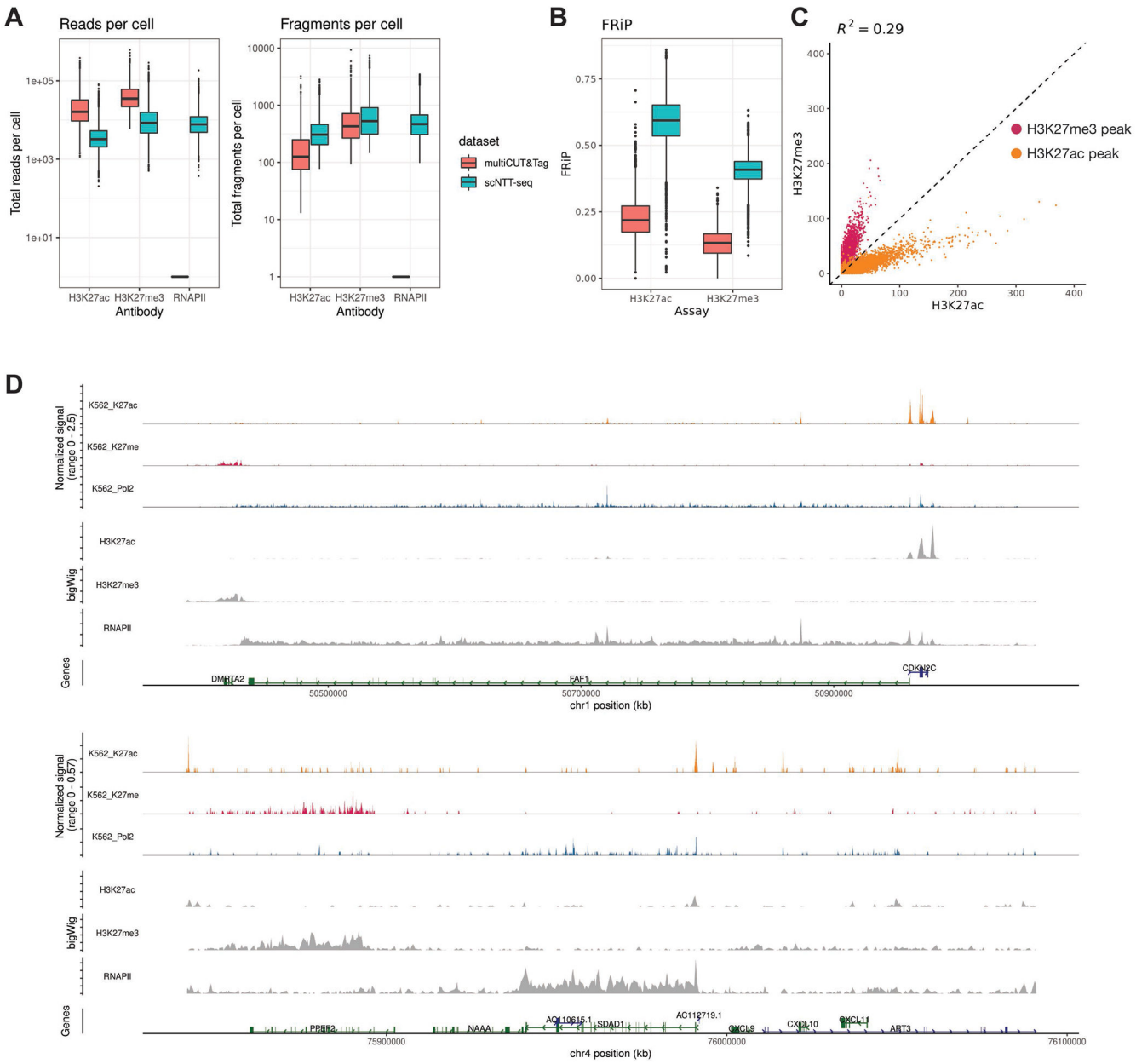
Extended Data Figure 1. Design and evaluation of nb-Tn5

A) Nanobody-Tn5 fusion protein plasmid map schematic showing position of Tn5 and secondary nanobody sequences.

B) Agarose DNA gel showing size-separation of PCR-amplified DNA sequencing library products for different combinations of nb-Tn5 and primary IgG antibody. Rabbit Ab: rabbit primary IgG antibody; Mouse Ab: mouse primary IgG antibody; IgG1 Ab: mouse IgG subtype 1 primary antibody; IgG2a Ab: mouse IgG subtype 2a primary antibody; rTn5: anti-rabbit IgG secondary nanobody-Tn5 fusion; mTn5: anti-mouse IgG secondary nanobody-Tn5 fusion; G1T: anti-mouse IgG1 secondary nanobody-Tn5 fusion; G2aT: anti-mouse IgG2a secondary nanobody-Tn5 fusion. Gels shows expected library amplification product (bands between 200 and 1,000 bp) in lanes where the nb-Tn5 fusion matches the primary IgG antibody (rabbit Ab + rTn5; mouse Ab + mTn5; IgG1 Ab + G1T; IgG2a Ab + G2aT). Replicates were not performed.

C) Scatterplots showing normalized fragment counts for H3K27me3 and H3K27ac peaks defined by ENCODE (17) for bulk multiplexed and non-multiplexed NTT-seq experiments in human PBMCs. Peaks are colored according to their chromatin modality (red: H3K27me3 peak, yellow: H3K27ac peak). Coefficient of determination (R^2) between experiments are shown above each scatterplot.

D) Scatterplots showing normalized fragment counts for H3K27me3, H3K27ac, and RNAPII peaks defined by ENCODE (17) for bulk multiplexed and non-multiplexed NTT-seq experiments in K562 cells. Peaks are colored according to their chromatin modality (red: H3K27me3; yellow: H3K27ac; blue: RNAPII).

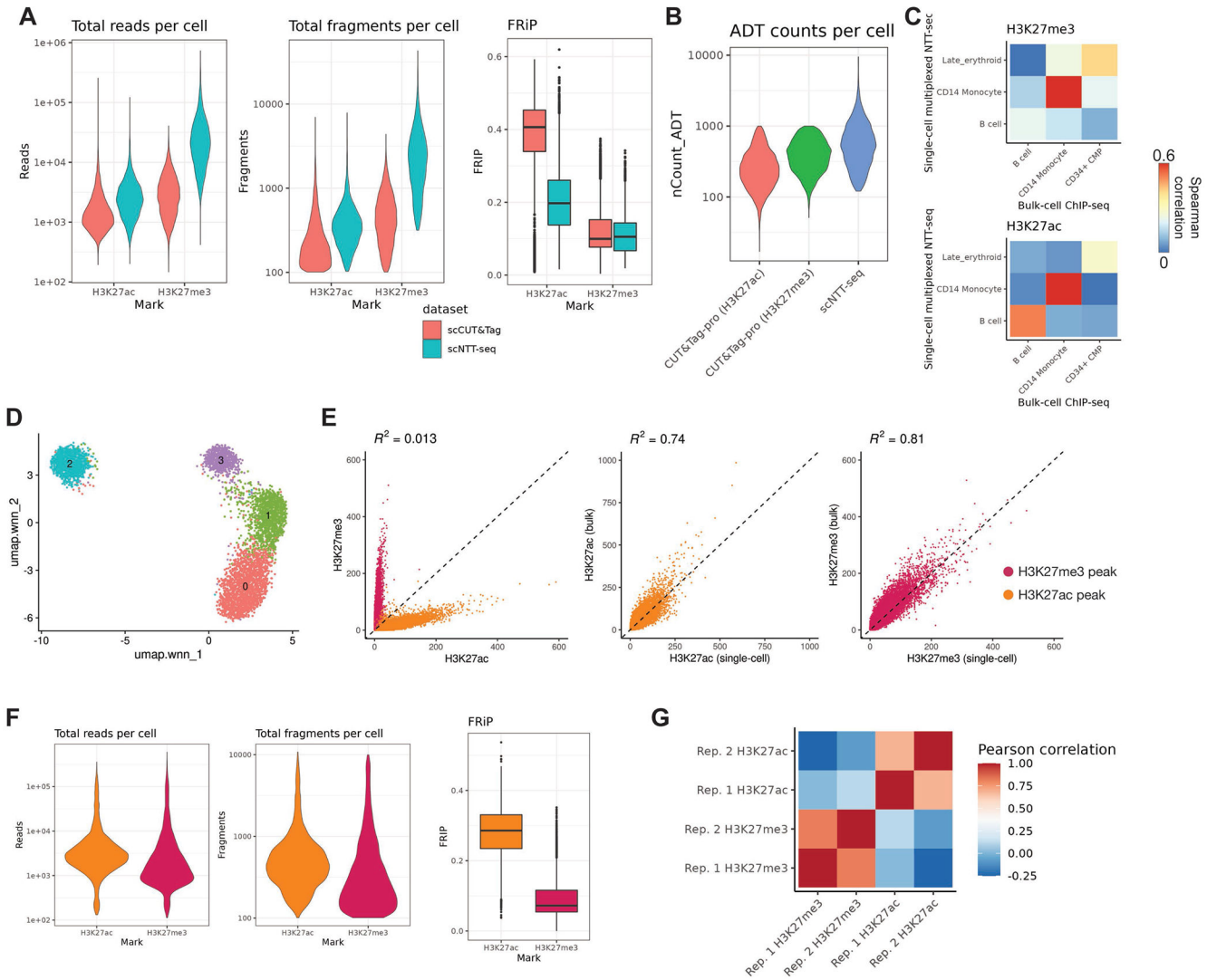


Extended Data Figure 2. Data sensitivity comparison across multimodal chromatin profiling methods

A) Total reads and fragment counts per cell for multiCUT&Tag (6) and scNTT-seq. Read and fragment counts on y-axis are on a log10 scale. multiCUT&Tag profiled only two marks, H3K27ac and H3K27me3, and so do not have RNAPII counts. Box-plot lower and upper hinges represent first and third quartiles. Upper/lower whiskers extend to the largest/smallest value no further than 1.5x the interquartile range. Data beyond the whiskers are plotted as single points.

B) Fraction of fragments falling in ENCODE peak regions for H3K27me3 and H3K27ac marks, for multiCUT&Tag (red) and scNTT-seq (blue). Box-plots constructed as for panel A.

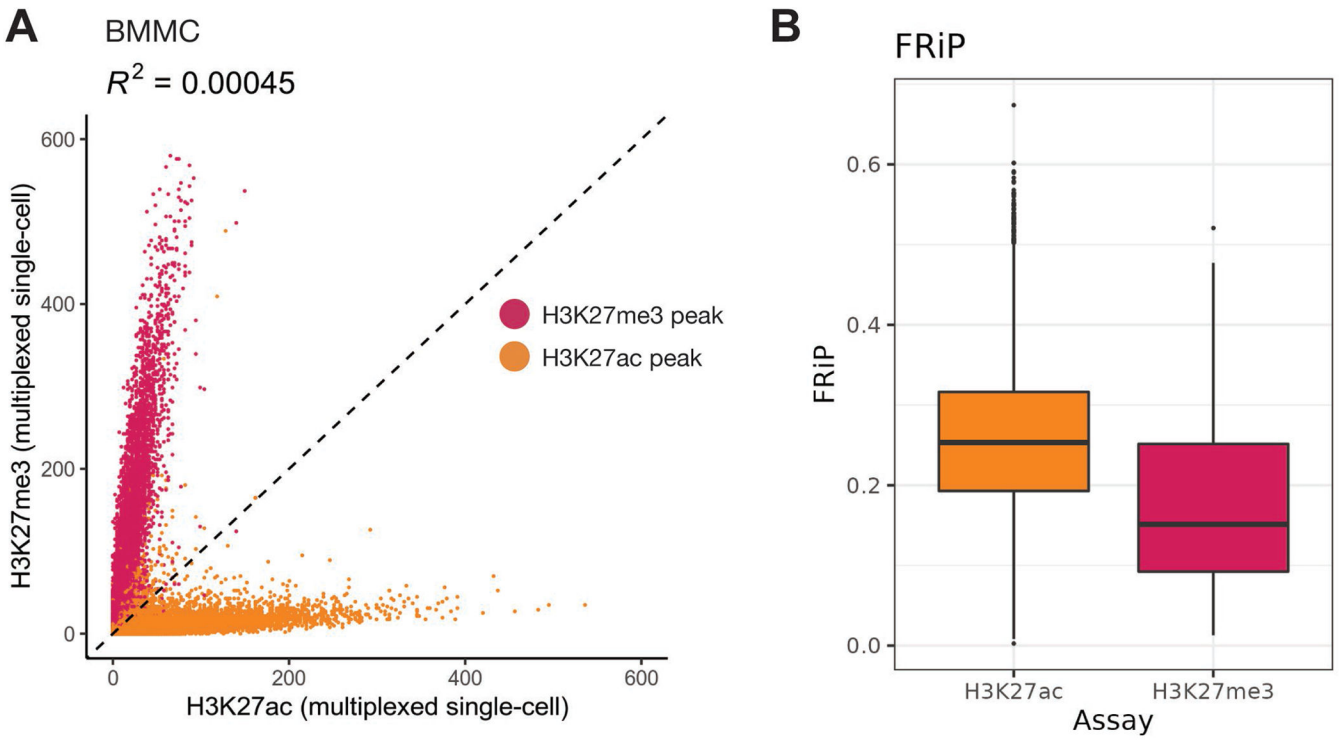
C) Scatterplot showing the normalized insertion counts in H3K27me3 and H3K27ac ENCODE peak regions for the multiCUT&Tag mESC single-cell dataset.
D) Multimodal genome browser view of a representative genomic locus, for K562 cells. Top three tracks show H3K27ac, H3K27me3, and RNAPII profiled simultaneously in a single-cell experiment. Lower three tracks show H3K27ac, H3K27me3, and RNAPII profiled individually in bulk-cell NTT-seq experiments using K562 cells.



Extended Data Figure 3. Sensitivity and reproducibility of scNTT-seq

A) Total read and fragment counts per cell and fraction of fragments in peaks (FRiP) for scCUT&Tag and scNTT-seq PBMC datasets. Box-plot lower and upper hinges represent first and third quartiles. Upper/lower whiskers extend to the largest/smallest value no further than 1.5x the interquartile range. Data beyond the whiskers are plotted as single points.
B) Comparison of total unique antibody-derived tag (ADT) counts sequenced per cell for CUT&Tag-pro (18) and scNTT-seq.

C) Spearman correlation between H3K27me3 counts (top) or H3K27ac counts (bottom) for cells profiled using multiplexed single-cell NTT-seq, or FACS-sorted bulk ChIP-seq profiled by ENCODE (17).
D) Two-dimensional UMAP projection and clustering for a second PBMC scNTT-seq replicate profiling H3K27me3 and H3K27ac. UMAP representation was constructed using both modalities, using the weighted nearest neighbors (WNN) method.
E) Scatterplots showing the number of fragment counts per H3K27me3 and H3K27ac ENCODE peak region for each assay profiled in the second PBMC scNTT-seq replicate dataset.
F) Total read and fragment count and FRiP distributions for H3K27me3 and H3K27ac assays profiled in the second PBMC scNTT-seq replicate dataset.
G) Pearson correlation between H3K27me3 and H3K27ac marks across PBMC scNTT-seq replicate datasets.



Extended Data Figure 4. Accuracy of scNTT-seq applied to human BMMCs.
A) Scatterplot showing the number of counts per H3K27me3 and H3K27ac peak for each assay, for BMMC cells profiled using single-cell multiplexed NTT-seq. Peaks are colored according to their assay (red: H3K27me3 peaks; yellow: H3K27ac peaks).
B) Fraction of fragments in ENCODE peaks per cell, for H3K27ac and HK27me3 marks. Box-plot lower and upper hinges represent first and third quartiles. Upper/lower whiskers extend to the largest/smallest value no further than 1.5x the interquartile range. Data beyond the whiskers are plotted as single points.

Extended Data Table 1 |

Quality metrics of datasets generated in this work

		Mean fragments per cell			Total fragments standard deviation			Mean fraction of fragments in ENCODE peaks		
Dataset	**Total cells**	**H3K27me3**	**H3K27ac**	**RNAPII**	**H3K27me3**	**H3K27ac**	**RNAPII**	**H3K27me3**	**H3K27ac**	**RNAPII**
K562	8617	743	382	542	699	282	350	0.4	0.59	0.2
PBMC + protein	4684	2854	412	-	2953	356	-	0.11	0.21	-
PBMC	4770	670	731	-	1243	1035	-	0.1	0.28	-
BMMC	5236	1217	326	-	1274	334	-	0.18	0.26	-

Extended Data Table 2 |

nb-Tn5 adapter and custom oligo sequences

Oligo name	Oligo sequence	Barcode seque
MEDSA_1	TCGTCGGCAGCGTCGGATTGCTGCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GGATTGCT
MEDSA_2	TCGTCGGCAGCGTCGTAATGCAGCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GTAATGCA
MEDSA_3	TCGTCGGCAGCGTCGTCAAGGAGCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GTCAAGGA
MEDSA_4	TCGTCGGCAGCGTCGTGAGCGTGCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GTGAGCGT
MEDSA_5	TCGTCGGCAGCGTCGTGTGACCGCGATCGAGGACGGCAGATGTGTATAAGAGACAG	GTGTGACC
MEDSA_6	TCGTCGGCAGCGTCTAAGGTGGGCGATCGAGGACGGCAGATGTGTATAAGAGACAG	TAAGGTGG
MEDSB	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	
Custom R1	GCGATCGAGGACGGCAGATGTGTATAAGAGACAG	
Custom i5	CTGTCTCTTATACACATCTGCCGTCCTCGATCGC	

Acknowledgements

This work was supported by the National Institutes of Health (K99HG011489 to T.S.; RM1HG011014-01 to R.S.,D.L.). B.Z. is a postdoctoral fellow of the Jane Coffin Childs Memorial Fund for Medical Research. This investigation has been aided by a grant from the Jane Coffin Childs Memorial Fund for Medical Research. We thank members of the Satija and NYGC Technology Innovation laboratories for feedback on the manuscript.

Data availability

The datasets generated in this study are available from NCBI GEO (GSE212588) and SRA (SRP395379). Processed single-cell R objects are available from Zenodo (https://zenodo.org/record/7102159). Data collected from PBMCs with cell-surface protein expression are available from dbGaP (phs003068.v1.p1). Nanobody-Tn5 fusion plasmids developed in this study are available from AddGene (184285, 184286, 184287, 184288). The following publicly available datasets were used in this study: GSE195725, GSE157910, GSE139369, GSM5227096.

References

1. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun [Internet]. 2019 Apr 29;10(1):1930. Available from: 10.1038/s41467-019-09982-5 [PubMed: 31036827]
2. Carter B, Ku WL, Kang JY, Hu G, Perrie J, Tang Q, et al. Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). Nat Commun [Internet]. 2019 Aug 20;10(1):3747. Available from: 10.1038/s41467-019-11559-1 [PubMed: 31431618]
3. Wang Q, Xiong H, Ai S, Yu X, Liu Y, Zhang J, et al. CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. Mol Cell [Internet]. 2019 Oct 3;76(1):206–16.e7. Available from: 10.1016/j.molcel.2019.07.015 [PubMed: 31471188]
4. Janssen SM, Lorincz MC. Interplay between chromatin marks in development and disease. Nat Rev Genet [Internet]. 2022 Mar [cited 2021 Oct 4];23(3):137–53. Available from: 10.1038/s41576-021-00416-x [PubMed: 34608297]
5. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A [Internet]. 2010 Dec 14;107(50):21931–6. Available from: 10.1073/pnas.1016071107 [PubMed: 21106759]
6. Gopalan S, Wang Y, Harper NW, Garber M, Fazzio TG. Simultaneous profiling of multiple chromatin proteins in the same cells. Mol Cell [Internet]. 2021 Nov 18;81(22):4736–46.e5. Available from: 10.1016/j.molcel.2021.09.019 [PubMed: 34637755]
7. Meers MP, Janssens DH, Henikoff S. Multifactorial chromatin regulatory landscapes at single cell resolution [Internet]. bioRxiv. 2021 [cited 2022 Mar 3]. p. 2021.07.08.451691. Available from: 10.1101/2021.07.08.451691v1
8. Pleiner T, Bates M, Görlich D. A toolbox of anti-mouse and anti-rabbit IgG secondary nanobodies. J Cell Biol [Internet]. 2018 Mar 5;217(3):1143–54. Available from: 10.1083/jcb.201709115 [PubMed: 29263082]
9. Saha K, Bender F, Gizeli E. Comparative study of IgG binding to proteins G and A: nonequilibrium kinetic and binding constant determination with the acoustic waveguide device. Anal Chem [Internet]. 2003 Feb 15;75(4):835–42. Available from: 10.1021/ac0204911 [PubMed: 12622374]
10. Hassanzadeh-Ghassabeh G, Devoogdt N, De Pauw P, Vincke C, Muyldermans S. Nanobodies and their potential applications. Nanomedicine [Internet]. 2013 Jun;8(6):1013–26. Available from: 10.2217/nnm.13.86 [PubMed: 23730699]
11. Amini S, Pushkarev D, Christiansen L, Kostem E, Royce T, Turk C, et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet [Internet]. 2014 Dec;46(12):1343–9. Available from: 10.1038/ng.3119 [PubMed: 25326703]
12. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, et al. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development [Internet]. 2009 Sep [cited 2022 Feb 3];136(18):3131–41. Available from: 10.1242/dev.037127 [PubMed: 19700617]
13. Zaborowska J, Egloff S, Murphy S. The pol II CTD: new twists in the tail. Nat Struct Mol Biol [Internet]. 2016 Sep 6;23(9):771–7. Available from: 10.1038/nsmb.3285 [PubMed: 27605205]
14. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nat Methods [Internet]. 2021 Nov [cited 2021 Nov 1];18(11):1333–41. Available from: 10.1038/s41592-021-01282-5 [PubMed: 34725479]
15. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol [Internet]. 2018 Dec 3; Available from: 10.1038/nbt.4314
16. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell [Internet]. 2021 Jun 24;184(13):3573–87.e29. Available from: 10.1016/j.cell.2021.04.048 [PubMed: 34062119]
17. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature [Internet]. 2012 Sep 6;489(7414):57–74. Available from: 10.1038/nature11247

18. Zhang B, Srivastava A, Mimitou E, Stuart T, Raimondi I, Hao Y, et al. Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. Nat Biotechnol [Internet]. 2022 Mar 24; Available from: 10.1038/s41587-022-01250-0
19. Wu SJ, Furlan SN, Mihalas AB, Kaya-Okur HS, Feroze AH, Emerson SN, et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. Nat Biotechnol [Internet]. 2021 Jul;39(7):819–24. Available from: 10.1038/s41587-021-00865-z [PubMed: 33846646]
20. Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat Biotechnol [Internet]. 2019 Dec;37(12):1458–65. Available from: 10.1038/s41587-019-0332-7 [PubMed: 31792411]
21. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell [Internet]. 2019 Jun 13 [cited 2019 Jun 6];177(7):1888–902.e21. Available from: 10.1016/j.cell.2019.05.031 [PubMed: 31178118]
22. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods [Internet]. 2017 Oct;14(10):979–82. Available from: 10.1038/nmeth.4402 [PubMed: 28825705]
23. Tedesco M, Giannese F, Lazarević D, Giansanti V, Rosano D, Monzani S, et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. Nat Biotechnol [Internet]. 2022 Feb [cited 2021 Oct 11];40(2):235–44. Available from: 10.1038/s41587-021-01031-1 [PubMed: 34635836]
24. Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res [Internet]. 2014 Dec;24(12):2033–40. Available from: 10.1101/gr.177881.114 [PubMed: 25079858]
25. Hawkins JA, Jones SK Jr, Finkelstein IJ, Press WH. Indel-correcting DNA barcodes for high-throughput sequencing. Proc Natl Acad Sci U S A [Internet]. 2018 Jul 3;115(27):E6217–26. Available from: 10.1073/pnas.1802640115 [PubMed: 29925596]
26. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost chromatin profiling with CUT&Tag. Nat Protoc [Internet]. 2020 Oct;15(10):3264–83. Available from: 10.1038/s41596-020-0373-x [PubMed: 32913232]
27. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. Nat Biotechnol [Internet]. 2021 Oct [cited 2021 Jun 3];39(10):1246–58. Available from: 10.1038/s41587-021-00927-2 [PubMed: 34083792]
28. Vasimuddin M, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) [Internet]. ieeexplore.ieee.org; 2019. p. 314–24. Available from: 10.1109/IPDPS.2019.00041
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 Aug 15;25(16):2078–9. Available from: 10.1093/bioinformatics/btp352 [PubMed: 19505943]
30. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res [Internet]. 2014 Jul;42(Web Server issue):W187–91. Available from: 10.1093/nar/gku365 [PubMed: 24799436]
31. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics [Internet]. 2011 Mar 1;27(5):718–9. Available from: 10.1093/bioinformatics/btq671 [PubMed: 21208982]
32. Waltman L, van Eck NJ. A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys J B [Internet]. 2013 Nov 13;86(11):471. Available from: 10.1140/epjb/e2013-40829-0
33. Hamilton NE, Ferry M. ggtern: Ternary Diagrams Using ggplot2. J Stat Softw [Internet]. 2018 Dec 20 [cited 2022 Mar 3];87:1–17. Available from: http://statistik-jstat.uibk.ac.at/index.php/jss/article/view/v087c03

34. Srivastava A, Malik L, Smith T, Sudbery I, Patro R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol [Internet]. 2019 Mar 27;20(1):65. Available from: 10.1186/s13059-019-1670-y [PubMed: 30917859]

35. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods [Internet]. 2017 Apr;14(4):417–9. Available from: 10.1038/nmeth.4197 [PubMed: 28263959]

36. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics [Internet]. 2010 Mar 1;26(6):841–2. Available from: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20110278&retmode=ref&cmd=prlinks [PubMed: 20110278]

37. Oetjen KA, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, et al. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. JCI Insight [Internet]. 2018 Dec 6;3(23). Available from: 10.1172/jci.insight.124928

38. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature [Internet]. 2019 Oct [cited 2019 Oct 9];574(7777):187–92. Available from: 10.1038/s41586-019-1629-x

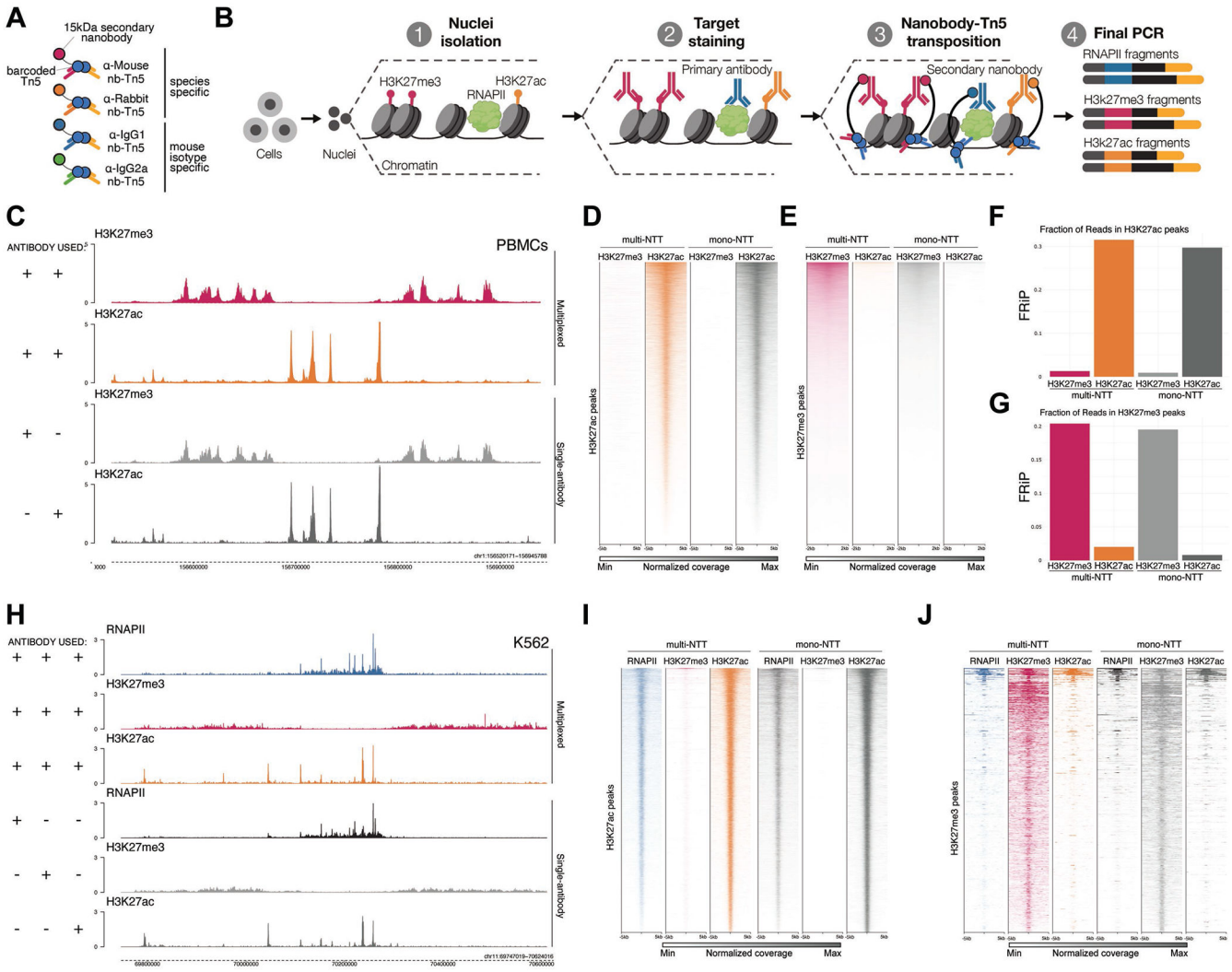


Figure 1. Bulk-cell NTT-seq enables simultaneous profiling of multiple chromatin marks.

A) Schematic representation of nanobody-Tn5 fusion proteins loaded with barcoded DNA adaptors.

B) Overview of the NTT-seq protocol. Nuclei are extracted from cells and stained with a mixture of IgG primary antibodies for targets of interest. Nanobody-Tn5 fusion proteins are then added and tagment the genomic DNA surrounding primary antibody binding sites. Released DNA fragments are amplified by PCR to obtain a sequencing library harboring barcode sequences specific for each nb-Tn5 protein used.

C) Genome browser tracks for a representative region of the human genome. NTT-seq was performed on PBMCs for H3K27me3 alone (light grey), H3K27ac alone (dark grey) or for both together in a multiplexed experiment (red/yellow). Sequencing data were normalized as bins per million mapped reads (BPM)

D) Heatmap displaying coverage within 33,205 H3K27ac peaks identified using MACS2, for multiplexed (multi) and non-multiplexed (mono) NTT-seq PBMC experiments.

E) As for D, for 67,459 H3K27me3 peaks.

F) Fraction of reads in H3K27ac peaks for multiplexed and non-multiplexed NTT-seq PBMC datasets.

G) As for F, for H3K27me3 peaks.

H) Genome browser tracks for a representative region of the human genome for multiplexed and non-multiplexed NTT-seq K562 cell datasets. Sequencing data were normalized as bins per million mapped reads (BPM), as for the PBMC datasets.

I) Heatmap displaying coverage centered on H3K27ac peaks for multiplexed and non-multiplexed NTT-seq experiments using K562 cells, for RNAPII, H3K27ac, and H3K27me3 modalities.

J) As for I, for H3K27me3 peaks.

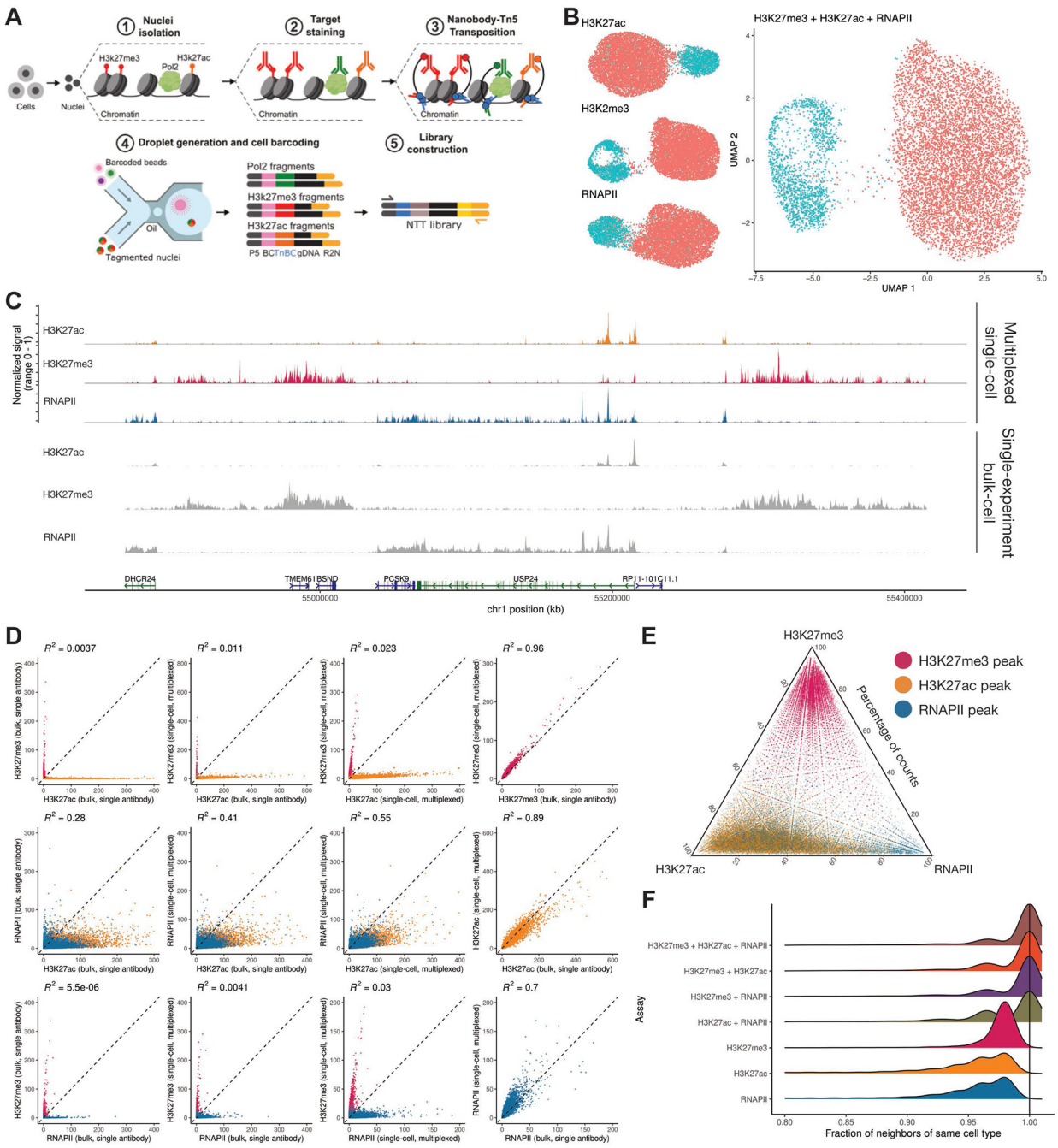


Figure 2. NTT-seq provides accurate single-cell multimodal chromatin profiles

A) Schematic overview of the single-cell NTT-seq protocol. Cells are tagmented and processed in bulk (steps 1-3), and are encapsulated in droplets to attach cell-specific barcode sequenced to transposed DNA fragments (steps 4-5).

B) UMAP representations of cells profiled using multiplexed single-cell NTT-seq. Individual UMAP representations built using each assay are shown (left side), along with a visualization constructed incorporating information from all three chromatin modalities (WNN UMAP, right side). Cells are colored by their predicted cell type.

C) Multimodal genome browser view of a representative genomic locus, for K562 cells. Fragment counts for each assay are shown, scaled to the maximal value for each assay within the locus. Top three tracks show H3K27ac, H3K27me3, and RNAPII profiled simultaneously in a single-cell experiment. Lower three tracks show H3K27ac, H3K27me3, and RNAPII profiled individually in bulk-cell NTT-seq experiments using K562 cells.
D) Scatterplots showing normalized fragment counts for H3K27me3, H3K27ac, and RNAPII peaks defined by ENCODE (17), for bulk and single-cell multiplexed NTT-seq experiments, for K562 cells. Peaks are colored according to their chromatin modality (red: H3K27me3 peak, yellow: H3K27ac peak, blue: RNAPII peak). Coefficient of determination (R^2) between experiments are shown above each scatterplot.
E) Ternary plot showing the relative frequency of H3K27me3, H3K27ac, and RNAPII fragment counts within H3K27me3, H3K27ac, and RNAPII peak regions defined by ENCODE ChIP-seq datasets.
F) Fraction a cell's nearest-neighbors belonging to the same predicted cell type, for neighbor graphs defined using a single chromatin modality or a weighted combination of modalities.

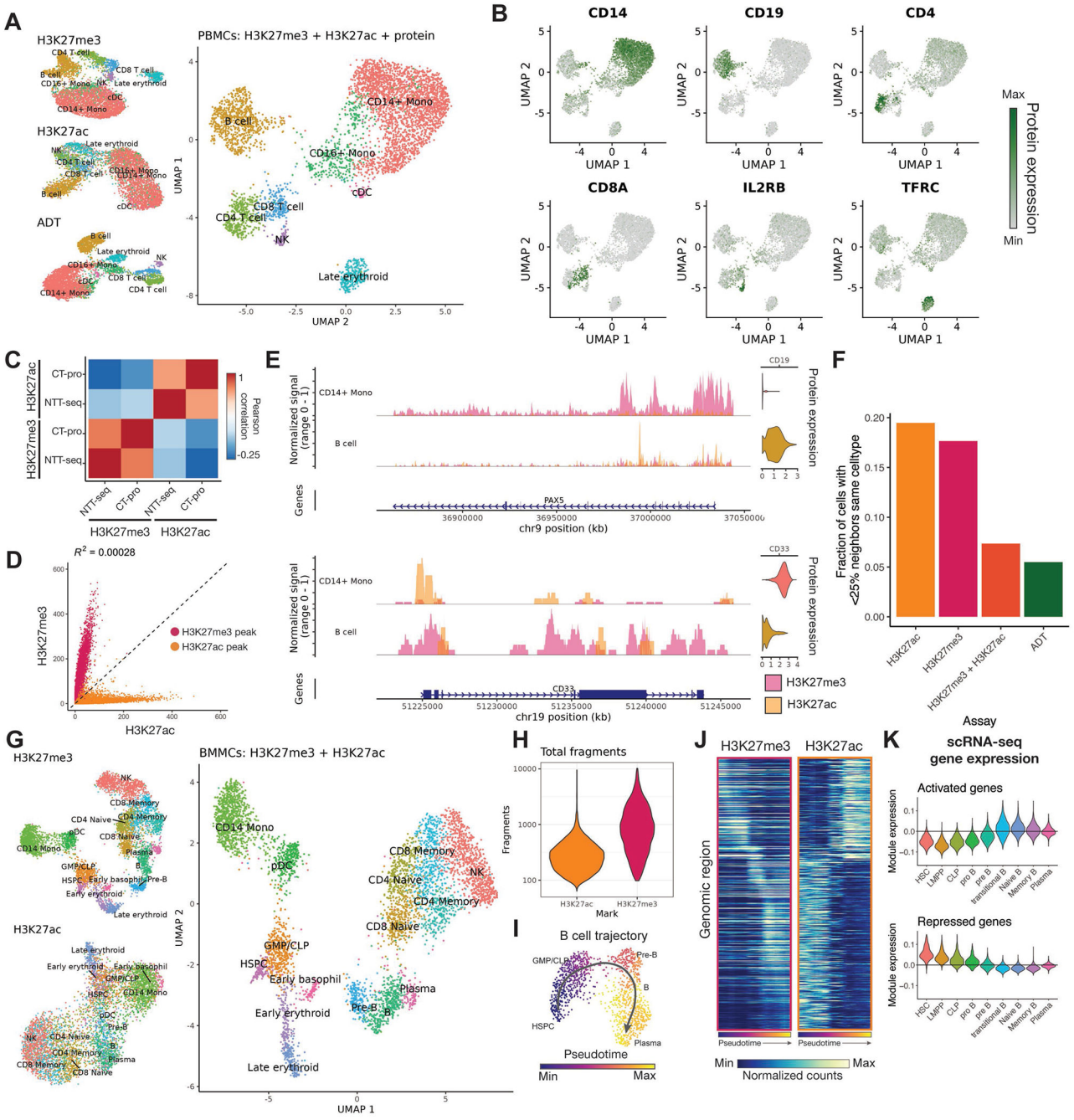


Figure 3. Application of multiplexed single-cell NTT-seq to human tissues.

A) UMAP representation of PBMCs profiled using NTT-seq with protein expression. UMAPs for each assay are shown (left side), along with a multimodal UMAP constructed using all modalities (right side). Cells are colored and labeled by cell types.

B) Patterns of cell-surface-protein expression in PBMCs profiled using NTT-seq.

C) Pearson correlation between NTT-seq and scCUT&Tag-pro (CT-pro) signal in PBMCs within H3K27me3 and H3K27ac peaks.

D) Scatterplot showing the number of counts per H3K27me3 and H3K27ac peak for each assay, for PBMCs profiled by NTT-seq. Peaks are colored according to their assay (red: H3K27me3; yellow: H3K27ac). Coefficient of determination (R^2) is shown above. Axes: total fragment counts per million.

E) Genome browser view of the *PAX5* and *CD33* loci for B cells and CD14+ monocytes. Normalized protein expression values are shown alongside coverage tracks for each cell type for CD19 and CD33 protein. H3K27me3 and H3K27ac histone modification profiles are overlaid, with the signal for each scaled to the maximal signal within the genomic region shown.

F) Fraction of cells with <25% of neighbors belonging to the same cell type, for neighbor graphs defined using individual chromatin modalities, cell-surface protein expression, or a combination of chromatin modalities.

G) UMAP of BMMCs profiled using NTT-seq. Separate UMAPs for H3K27me3 and H3K27ac are shown (left side), and a UMAP using both H3K27me3 and H3K27ac is shown (right). Cells colored and labeled by their cell type. HSPC: hematopoietic stem and progenitor cells; GMP/CLP: granulocyte monocyte progenitor / common lymphoid progenitor; CD14 Mono: CD14+ monocyte; pDC: plasmacytoid dendritic cell; NK: natural killer cell.

H) Distribution of total fragment counts per cell for H3K27ac and H3K27me3.

I) Pseudotime trajectory for B cell development. Cells are colored by their pseudotime value and labeled by their annotated cell type.

J) Heatmap showing H3K27me3 and H3K27ac signal for 10 kb genomic bins correlated with B cell pseudotime progression. Heatmaps show the same genomic regions for both assays, with identical ordering of genomic regions.

K) Expression of genes close to activated (gain H3K27ac, upper plot) or repressed (gain H3K27me3, lower plot) genomic regions in a separate scRNA-seq BMMC dataset, for cells in the B cell developmental trajectory.