nature portfolio

https://doi.org/10.1038/s41596-022-00692-9

Supplementary information

Chromatin accessibility profiling by ATAC-seq

In the format provided by the authors and unedited

Supplementary Information for:

Chromatin accessibility profiling by ATAC-seq

Fiorella C. Grandi[1,2,3], Hailey Modi[1,2,3], Lucas Kampman[1,2,3], M. Ryan Corces[1,2,3,*]

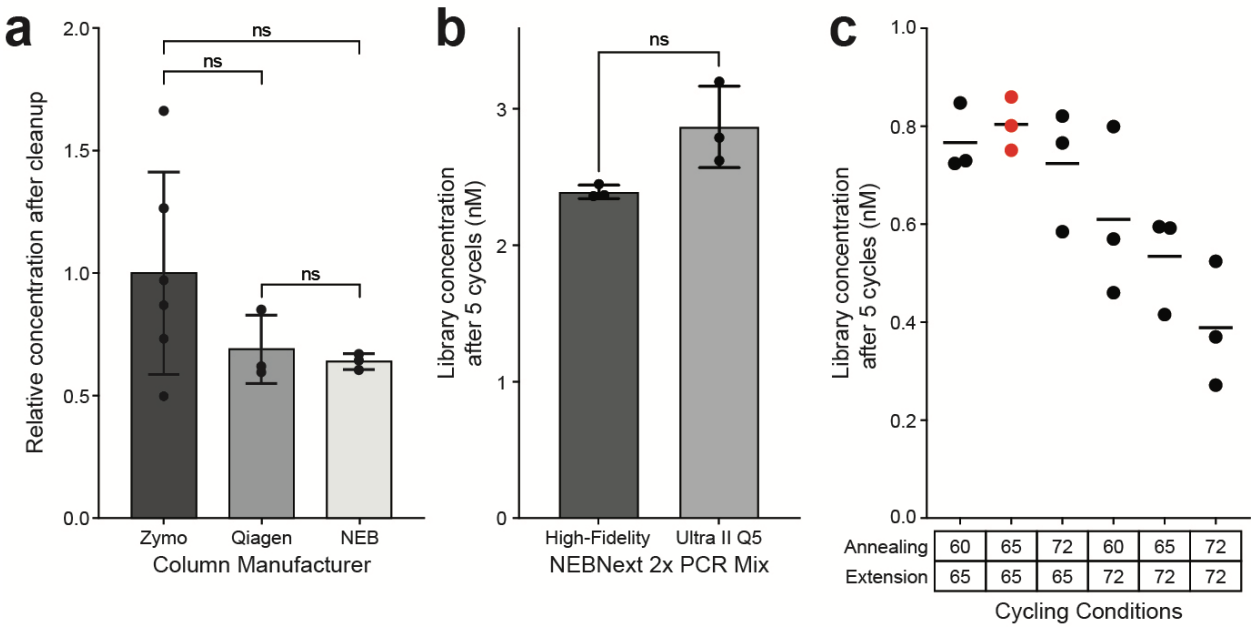
[1]Gladstone Institute of Neurological Disease, San Francisco, CA 94158
[2]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA 94158
[3]Department of Neurology, University of California San Francisco, San Francisco, CA 94158

*Correspondence should be addressed to M.R.C. (ryan.corces@gladstone.ucsf.edu)

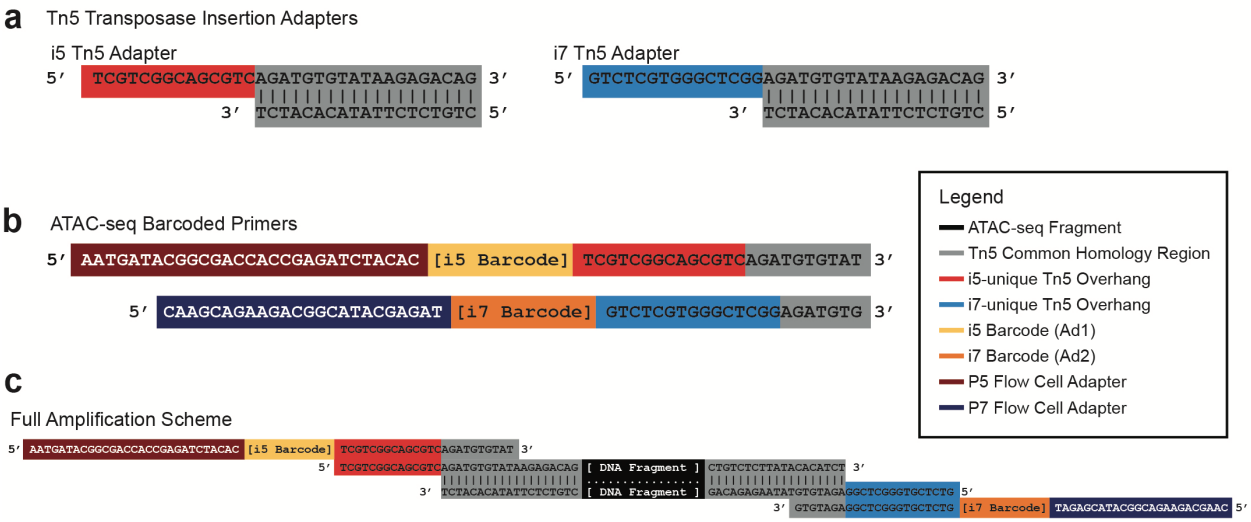
This document contains the following:

Alignment statistics for various ATAC-seq read lengths can be found in **Supplementary Table 1**, which is not included in this document.

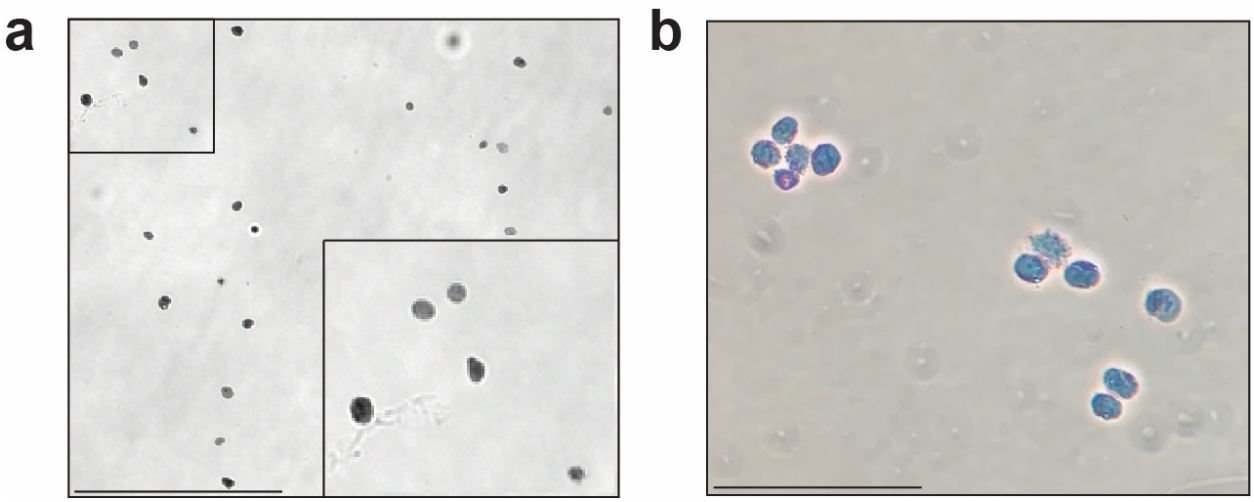
The adapter sequencers can be found in **Supplementary Table 2**, which is not included in this document.



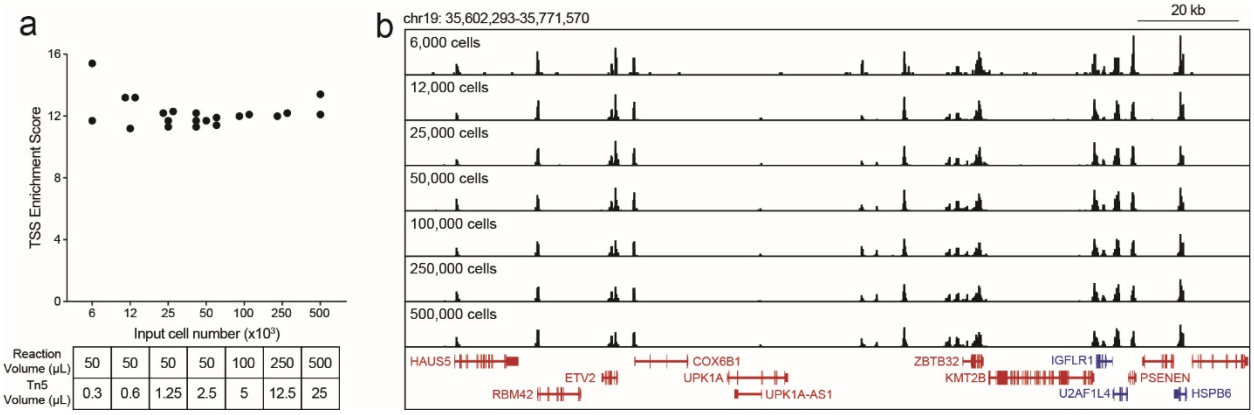
Supplementary Figure 1: Validation of alterations to Omni-ATAC protocol
a. Quantitative real-time PCR data from ATAC-seq libraries prepared using either the Zymo Research DNA Clean & Concentrator-5 kit (Zymo Research, cat. no. D4014), the Qiagen MinElute PCR Purification kit (Qiagen, cat. no. 28004), or the NEB Monarch PCR & DNA Cleanup kit (New England Biolabs, cat. no. T1030S). Concentration values for each sample were normalized to the average of the trials from the the Zymo DNA Clean & Concentration-5 kit (n=6, 3, and 3, respectively. n.s. not significant after one-way ANOVA; bars indicate standard deviation). **b.** ATAC-seq library concentration after pre-amplification for 5 cycles with either the NEBNext High-Fidelity 2x PCR Master Mix (New England Biolabs, cat. no. M0541L) or the NEBNext Ultra II Q5 2x Master Mix (New England Biolabs, cat. no. M0544L). Library concentration was calculated using the NEBNext Library Quant kit (n=3; n.s. not significant after a two-tailed student's t-test). **c.** Identification of optimal cycling conditions for the barcoding PCR using the NEBNext Ultra II Q5 2x Master Mix. Library concentration was measured after 5 cycles of amplification using the NEBNext Library Quant kit. The final selected cycling conditions used in this protocol (anneal at 65 °C and extend at 65 °C) are highlighted by red dots. Bars indicate the average of each trial.



Supplementary Figure 2: Sequences of ATAC-seq adapters and barcodes
a. Tn5 transposase insertion adapters which are complexed to the Tn5 enzyme and pasted into the native chromatin during the Tn5 transposition reaction in **Step 10 b** of the main protocol**.** ATAC-seq barcoded primers used in the barcoding PCR in **Step 25 c** of the main protocol**.** Full amplification scheme showing how the barcoded primers overlap with the Tn5 transposase insertion adapters.

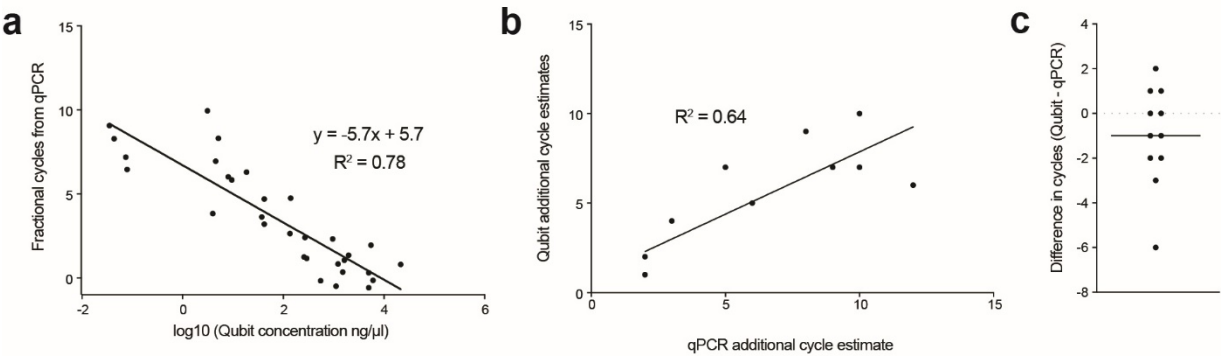


Supplementary Figure 3: Images of nuclei prior to ATAC-seq
a. Nuclei derived from human post-mortem brain samples after extraction. Note that most cell types should yield nuclei that appear round with little-to-no cytoplasm attached. Scale bar on the bottom left is 0.56 mm. Box in the bottom-right corner shows a zoom in of the box shown in the upper left. **b.** Image of GM12878 cells after 3 minutes of lysis and washing, immediately following **Step 7** in the main protocol. Cells/nuclei were stained with trypan blue. In contrast to live cells, the cell membrane is permeabilized, and we see trypan blue staining, rather than exclusion. Note that not all cells will appear lysed after the 3-minute incubation but this will likely not affect sample quality as lysis will continue during the transposition reaction. Scale bar on the bottom left is 0.56 mm.

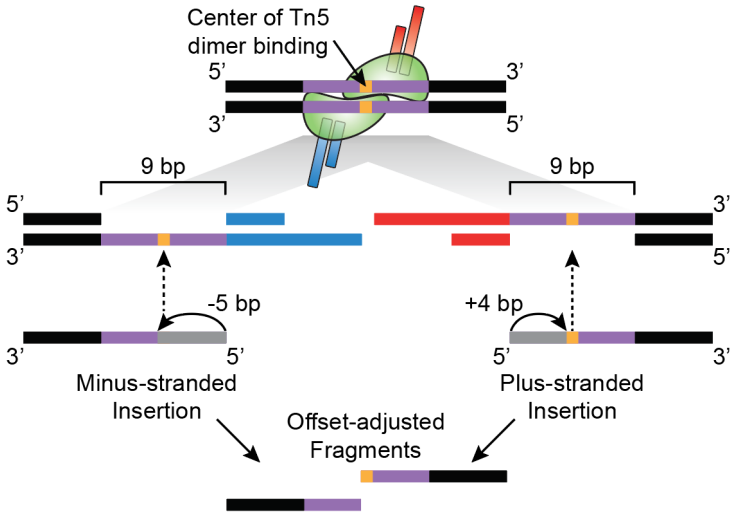


Supplementary Figure 4: Scaling of Tn5 concentration for different quantities of input cells

a. TSS Enrichment Scores for different quantities of input GM12878 cells. At least two independent trials were performed for each number of input cells. The reaction volume and Tn5 volume used is shown below the graph. **b.** Representative normalized ATAC-seq signal tracks for one of the replicates from each different number of cells shown in **(a**). Each track is scaled to the same height to allow for direct comparison across tracks. Genes shown in red are on the plus strand while genes shown in blue are on the minus strand.



Supplementary Figure 5: Derivation and validation of Qubit equation
a. Data used to derive the equation shown in **Box 3**, collected from ATAC-seq libraries (n=31) at a variety of different input cell numbers. Note that the final equation presented in the manuscript adds an additional cycle to ensure that all samples are adequately amplified. **b.** Correlation between cycles calculated by the qPCR method and the Qubit method for a different test set of ATAC-seq libraries from a variety of input cell numbers (n=11). **c.** Median difference in additional amplification cycles required between the Qubit-based and qPCR-based estimates for the same test data as in (**b)**. Each point represents the difference in cycles for a unique ATAC-seq library quantified by both methods (n=11). The median difference (-1) is noted by a solid line.



Supplementary Figure 6: Schematic of Tn5 offset adjustment
Adjusting ATAC-seq fragments for the Tn5 offset is an important component of early ATAC-seq data processing. When the Tn5 dimer binds, it inserts the adapter sequences, depicted in red and blue here, and duplicates the 9-bp DNA sequence at the center of the insertion site, depicted in purple and gold. In order to ensure that the start position of each ATAC-seq read represents the precise center of the region of chromatin accessibility being tagged by Tn5 (here shown in gold), each read must be offset to account for this duplicated 9-bp DNA region. For the minus-strand insertion, the fragment is trimmed by 5 bp, while on the plus-stranded insertion, the fragment is extended by 4 base-pairs. In this schematic, the Tn5 dimer is shown for illustrative purposes and is not drawn to scale.

Supplementary Notes

Supplementary Note 1: Transposition, chromatin accessibility, and TF dynamics

In the context of ATAC-seq, we use the term "accessibility" to describe chromatin that can be accessed by the Tn5 transposase. We refrain from using the commonly adopted "open" vs "closed' chromatin dichotomy. Often, "open chromatin" is interchangeably used with the terms "accessible chromatin" or euchromatin, and may be used to imply or depict DNA sequences with lower nucleosome density. Similarly, "closed chromatin" is often used interchangeably with the terms "inaccessible chromatin" or heterochromatin, to mean regions of epigenetic silenced DNA containing densely packed nucleosomes. When thinking about these terms in the context of ATAC-seq assays, we find these terms misleading. ATAC-seq, by definition, is an assay of sites *accessed* by Tn5 – regardless of the underlying biochemical reason, regions of chromatin with an enrichment of Tn5 insertions are by definition *accessible* to Tn5. It is worth noting that these regions of Tn5-accessible chromatin can occur in chromatin marked by repressive histone modifications such as H3K27me3 but more frequently occur in chromatin marked by active histone modifications such as H3K27ac or H3K4me1 (i.e. "closed" and "open" chromatin in colloquial terms)[1]. Because of this, it is important to not conflate Tn5-accessibility with the histone modification state of a genomic region.

When interpreting ATAC-seq peak data, it is important to remember two aspects of how the assay works. First, ATAC-seq peaks represent an enrichment of Tn5 transposition in a certain location over the background transposition that occurs throughout the genome. However, there is little to no chromatin that is truly "inaccessible" to Tn5 across all cells in a population; we are merely measuring regions where transposition is more likely to occur, due to the biochemical and biophysical properties of that chromatin. While it is well-appreciated that TF-bound regions of chromatin are enriched for transposition, the precise molecular mechanism and properties that govern the ability of Tn5 to access specific regions of chromatin remain an active area of investigation. One frequent explanation for the mechanism of Tn5 transposition is that TFs bind to chromatin and create adjacent nucleosome-free regions that are accessed by Tn5. While this provides an intuitive lens through which to think about ATAC-seq data, it may be an oversimplification. Along these lines, recent *in vitro* experiments have shown that Tn5 preferentially transposes nucleosomal linker sequences of a particular length as well as nucleosome entry/exit sites[2].

Second, even though ATAC-seq peak regions represent a clear enrichment in signal over the background, most peaks are defined by a few hundred transposition events. Thus, transposition at this location did not occur in the vast majority of the 50,000 cells in the assay. This likely relates to (i) the limiting amount of Tn5 transposase in the assay, and (ii) the dynamics of TF binding. In the ATAC-seq reaction, the Tn5 transposase is non-catalytic and each Tn5 transposase molecule can only perform a single transposition reaction. As such, the transposase is rate-limiting in the reaction. This favors transposition at sites where many proteins are bound together (for example TSSs) or at sites where the TF binds very strongly with longer residence. More specifically, TFs are not continuously bound to chromatin[3,4] but rather interact with the DNA through binding reactions that are governed by biophysical properties, including their dissociation constant (K_D), the local concentration of that TF and other chromatin modulating enzymes, the modifications on nearby histones, and other biophysical properties of long DNA polymers[5–9]. Current thinking suggests that TFs are constantly binding and releasing chromatin, replacing and being replaced by nucleosomes, a process additionally facilitated by ATP-dependent chromatin remodelers. Orthogonal assays to observe TF binding at single-molecule resolution using catalytic enzymes have shown that more than 50% of chromatin fibers overlapping a given ATAC-seq peak show evidence of TF binding[10]. Thus, it is important to understand that the static representation that we capture in bulk ATAC-seq is an ensemble average of the sensitivity of chromatin to Tn5 across all cells profiled.

Supplementary Note 2: Calculating Tn5-to-cell Ratio

The ratio of Tn5-to-cells should be kept constant across ATAC-seq libraries that will be compared to each other. When the input cell number is less than 50,000 cells/nuclei, we recommend maintaining the total reaction volume at a 50 µL while reducing the volume of Tn5 proportionally and replacing the difference in total reaction volume with water. When performing ATAC-seq on fewer than 5,000 cells, we recommend reducing the ATAC-seq reaction volume to 10 µL and reducing the volume of Tn5 proportionally to the cell number used. Because small volumes of Tn5 can be impossible to accurately pipette, a same-day working solution of 1:10 diluted Tn5 can be prepared immediately prior to use in 1x TD Buffer. When performing ATAC-seq on more than 50,000 cells, we recommend scaling the total volume to be 1/1000th of the cell quantity, and scaling the volume of Tn5 with the cell number used, as above. Example calculations are provided below. Expected TSS Enrichment Scores and example peak tracks are provided in **Supplementary Figure 4** for these scaled reactions.

Cell number	**1000**	**6,000**	**12,000**	**25,000**	**50,000**	**100,000**	**250,000**	**500,000**
Total volume of transposition reaction (µL)	10	50	50	50	50	100	250	500
TDE1 Tagment DNA Enzyme (Tn5 Transposase) (µL)	0.05	0.3	0.6	1.25	2.5	5	12.5	25

Supplementary Note 3: Tn5 Enzyme Production

We strongly recommend purchasing the TDE1 Tagment DNA Enzyme (Tn5 transposase) and TD Buffer from Illumina as these are core components of the ATAC-seq reaction. However, protocols for producing, purifying, and complexing the Tn5 transposase enzyme have been published[11] and we find these protocols to yield Tn5 transposase of equivalent quality to commercial products when performed by someone experienced in protein purification and biochemistry.

If Tn5 transposase will be home-made, then the 2x TD Buffer should also be made by hand. For 100 mL of 2x TD Buffer, combine 2 mL of 1M Tris-HCl pH 7.6 (Teknova, cat no. T1076), 1 mL of 1M MgCl2, and 70 mL of UltraPure Distilled Water. Adjust pH to 7.6 using 100% Glacial Acetic Acid (Millipore-Sigma, cat. no. A6283) and then add 20 mL of Dimethylformamide (Millipore-Sigma, cat. no. 227056). Lastly, bring the total volume to 100 mL with UltraPure Distilled Water. The final composition of 2x TD Buffer is 20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, and 20% vol/vol Dimethylformamide in water. Aliquot in 1-2 mL volumes and store at -20 °C for up to 2 years. Note that polystyrene is soluble in Dimethylformamide and the use of polystyrene serological pipettes to transfer Dimethylformamide will result in an unusable white cloudy solution.

Supplementary Note 4: Details about TF Footprinting Analysis

TF Footprinting analysis can be performed in a motif-centric method (either supervised or unsupervised) or in an agnostic *de novo* method. Both methods have strengths and drawbacks, and the specific choice may depend on the question. In general, supervised, motif-centric methods have the greatest sensitivity, but the tradeoff is that if they were not trained on a specific cell type and TFs of interest, the predictions may be less relevant. *De novo* and

unsupervised motif-centric methods provide this sample type flexibility, but must be implemented and corrected carefully.

Among the motif-centric methods, implementation can either be supervised or unsupervised. Supervised methods utilize previously collected ChIP-seq and DNase-seq (or ATAC-seq) datasets, for example those generated by ENCODE, to train their models and use this model to then detect footprints in new data. There are several tools which have been trained using DNase-seq data, and two, DeFCoM[12] and BaGFoot[13], which have been trained using ATAC-seq data. In particular, BaGFoot, or Bivariate Genomic Footprinting, is designed to measure two aspects of the footprint – the footprint depth and the flanking accessibility (200 base pairs on either side of the motif) between two conditions, and is more sensitive to subtle changes in the TF activity. This method was shown to be robust to the assay method (DNase-seq or ATAC-seq) and to the peak caller used. Although it does not explicitly correct for the Tn5 bias, BaGFoot was shown to be resistant to Tn5 bias, even in uncorrected data, because it looks for the difference between two conditions; ergo any bias should be equally found in both samples.

Unsupervised motif-centric methods do not rely on previous ChIP-seq data. Rather, they take a database of PWMs and scan the genome for all the TF binding sites (TFBSs). They will then classify the putative TFBSs as “bound” or “unbound” based on the features extracted from the genomic regions (distance to TSS, PWM match score, sequence conservation score) as well as from the ATAC-seq reads (read number, shape distribution around the putative TFBS). Two such tools, which have been applied to ATAC-seq data are CENTIPEDE[14] and PIQ[15], although neither natively corrects for Tn5 bias.

Finally, rather than relying on previously curated ChIP-seq or motif PWM databases, some programs will call footprinting motifs *de novo. De novo* motif calling is algorithmically and mathematically complicated[16]. For researchers interested in trying this analysis on their data, we recommend using either HINT-ATAC[17] or TOBIAS[18], both of which are *de novo* footprint callers that model the Tn5-specific bias and can perform differential footprinting between conditions. HINT-ATAC, which is derived from a sister program called HINT for DNase data, deals with the Tn5 bias by using sparse local inhomogeneous mixture (Slim) models[19] to predict the probable Tn5 bias at a particular site. These Slim models have been used to estimate additive dependencies on TF binding, and HINT-ATAC verifies their effectiveness for predicting cleavage biases on both ATAC-seq and DNase-seq data. HINT-ATAC also allows the user to test for differential footprints between two conditions and offers a motif matching analysis to determine which TF might be associated with the *de novo* motif predicted by HINT-ATAC. More recently TOBIAS was developed, which encompasses a suite of footprinting tools, including bias correction, differential footprinting between conditions, and visualization. TOBIAS shares many of the benefits of HINT-ATAC, with added support for performing comparisons between more than 2 conditions and can analyze footprints at both the meta and single-site levels. Each of the TOBIAS tools can be run separately, but are also integrated into a standardized Nextflow or Snakemake pipeline for greater usability, which will provide a full-featured output starting from BAM or FASTQ files. TOBIAS performs Tn5 bias correction by using a dinucleotide weight matrix to estimate the background bias of the Tn5 and also calculates the same footprint depth metric as BaGFoot. Their correction method was able to identify both cases where the Tn5 bias hid the footprint motif, as well as cases where the sequence is highly unfavorable to Tn5 integration, thereby creating a false-positive footprint in the uncorrected data.

When considering TF footprinting, it is important to note that footprinting at a single-site as opposed to across the genome, is very challenging[20]. Indeed, many footprinting methods aggregate reads from across the genome to make “meta” footprints. Second, there are several reported instances of TFs with verified binding that do not show strong footprints via these methods, likely due to technical limitations[13,21]. Finally, although supervised motif-centric methods generally outperform their unsupervised or *de novo* counterparts, their generalizability

is unclear. This is because the models are often trained on a different data type (DNase-seq versus ATAC-seq versus ChIP-seq) or using data from a different cell or tissue type. Finally, the sequencing depth guidelines for footprinting are still not formalized and generally saturation analysis is necessary to provide a suggestion for a particular dataset and TF. As a starting point, 200 million read-pairs per sample for ATAC-seq data is recommended.

Supplementary Protocol 1 – Isolation of nuclei from frozen tissues

Here we present a detailed protocol for isolation of nuclei from human or mouse tissues. We maintain a version of this protocol on Protocols.io where readers can post questions and read comments on protocol use.
https://dx.doi.org/10.17504/protocols.io.6t8herw

Input Material Preparation

This protocol has been extensively tested on flash frozen tissue. The health or quality of the tissue at the time of freezing will impact the quality of nuclei and the resultant ATAC-seq data. For example, high ischemic times, long postmortem intervals, or failure to maintain the tissue frozen will adversely affect data quality. We have found that this protocol also works well on thick (>= 30 micron) cryosections as long as the tissue is maintained frozen during cutting; specifically, the presence of some optimal cutting temperature (OCT) medium does not interfere with nuclei isolation. Lastly, we have found that slow-freezing tissue in BAM Banker cryopreservative instead of flash freezing can improve data quality and we therefore recommend this for new tissue collection where possible. To do this, place a 20 – 50 mg chunk of fresh tissue in a cryotube, add sufficient BAM Banker media to cover the tissue, and slow freeze in a designated freezing container. When thawing, thaw on ice until the tissue can be extracted and transfer the tissue to a Dounce (**Step 2** of this nuclei isolation protocol) using forceps. Tissue should be stored at or below -80 °C.

Necessary expertise

Basic molecular biology skills are necessary to perform this nuclei isolation.

Limitations

This protocol has been used to isolate nuclei from a wide variety of frozen human and mouse tissues. Certain tissue types, especially those with high extracellular matrix content, may be difficult to homogenize using a Dounce homogenizer and may benefit from the use of different nuclei isolation procedures. See the section titled *Alternative Nuclei Isolation Methods*, located in the main text, for additional information.

MATERIALS

Biological Materials

- Input material of interest.

Reagents and Equipment also Used in Main Protocol

Reagents

- 1M Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat no. 15-567-027)
- 5M NaCl (Corning, cat. no. 46-032-CV)
- 1M $MgCl_2$ (Invitrogen, cat. no. AM9530G)
- UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher Scientific, cat. no. 10977015)
- 10% Tween-20 wt/vol (Millipore-Sigma, cat. no. 11332465001)
  CRITICAL: 100% Tween-20 can be difficult to dilute accurately. For best results, purchase 10% Tween-20.
- 10% Nonidet P40 Substitute wt/vol (Millipore-Sigma, cat. no. 11332473001)
  CAUTION: NP40 can cause skin and eye irritation. Handle using appropriate protective gloves.

CRITICAL: 100% Nonidet P40 Substitute can be difficult to dilute accurately. For best results, purchase 10% NP40.
- BAM Banker Cryopreservative (Fisher Scientific, cat. no. NC9582225)
- ATAC-RSB (See main protocol for recipe)

Equipment
- 2 µL, 20 µL, 200 µL, and 1,000 µL single-channel Pipettes (Rainin, cat. nos. 17014393, 17014392, 17014391, and 17014382)
- 20 µL, 200 µL, and 1,000 µL filter tips (Rainin, cat. nos. 17014961, 17014963, and 17014967)
- 5 mL, 10 mL, 25 mL, and 50 mL Serological pipettes (Thermo Fisher Scientific, cat. nos. 07-200-573, 07-200-574, 07-200-575, and 07-200-576)
- Pipet-Aid XP (Drummond, cat. no. 4-000-101)
- 1.5 mL DNA LoBind tubes (Eppendorf, cat. no. 4043-1021)
  CRITICAL: LoBind tubes should be used to maximize sample recovery of nucleic acids as they reduce sample-to-surface binding.
- Refrigerated fixed-angle microcentrifuge (Eppendorf, cat. no. 5404000413)
- 15 mL and 50 mL conical tubes (Thermo Fisher Scientific, cat. nos. 05-538-53F and 05-538-55A)
- 1.5 mL Microcentrifuge tubes (Thermo Fisher Scientific, cat. no. 14-222-155)
- 0.22 µm 250 mL cellulose nitrate sterilizing filter (Corning, cat. no. 430756)

Reagents and Equipment not Used in Main Protocol

Reagents
- 60% Iodixanol / Optiprep (Millipore-Sigma, cat no. D1556-250ML)
- Sucrose (Millipore-Sigma, cat. no. S7903-250G)
- Tricine (Millipore-Sigma, cat. no. T0377-25G)
- Potassium Hydroxide (Millipore-Sigma, cat. no. P5958-250G)
- cOmplete Protease Inhibitors (Millipore-Sigma, cat. no. 11697498001)
- Spermidine Trihydrochloride (Millipore-Sigma, cat. no. S2501-1G)
- Spermine (Millipore-Sigma, cat. no. S3256-1G)
- 2 M KCl (Thermo Fisher Scientific, cat. no. AM9640G)
- Dithiothreitol / DTT (Thermo Fisher Scientific, cat. no. R0861)
- RNasin Plus RNase Inhibitor (Promega, cat no. N2611)
- Trypan Blue Solution 0.4% (Thermo Fisher Scientific, cat. no. 15250061)

Equipment
- 2 mL DNA LoBind tubes (Eppendorf, cat. no. 022431048)
- 1.8 mL Nunc Biobanking Cryogenic tubes (Thermo Fisher Scientific, cat. no. 375418)
- 70 µm Flowmi cell strainers (Fisher Scientific, cat. no. 03-421-228)
- Dounce tissue grinder set (Millipore-Sigma, cat. no. D8938-1SET)
- 0.2 µm pore size PTFE syringe filter (Millipore-Sigma, cat. no. SLLG025SS)
- 10 mL luer-lock syringes (Becton Dickinson, cat. no. 309604)
- Refrigerated swinging-bucket centrifuge (Eppendorf, cat. nos. 022620689 & 5427757007)
- 2 L polypropylene beaker or similar (Fisher Scientific, cat. no. 02-591-33)
- 70 µm cell strainer (Corning, cat. no. 352350)
- Kimwipe delicate task wipes (Kimberly-Clark, cat. no. 34120)

- Disposable hemocytometer slides, Neubauer Improved Grid (Fisher Scientific, cat. no. 22-600-100)
- CoolCell cell freezing containers (Corning, cat. no. 432003)

REAGENT SETUP

There are multiple stock buffers which should be prepared ahead of time. All stock buffers should be sterile filtered using a 0.22 micron filter and stored at the temperatures indicated below. An approximate shelf life and storage temperature is also given for each solution.

1.034x Homogenization Buffer	**For 200 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
1 M Sucrose	250 mM	4	50 mL
2 M KCl	25 mM	80	2.5 mL
1 M MgCl2	5 mM	200	1 mL
0.75 M Tricine-KOH pH 7.8	20 mM	37.5	5.33 mL
Water	-	-	141.17 mL
Store at 4 °C for 1 year		Total Volume	200 mL

Diluent Buffer	**For 100 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
2 M KCl	150 mM	13.33	7. 5 mL
1 M MgCl2	30 mM	33.33	3 mL
0.75 M Tricine-KOH, pH 7.8	120 mM	6.25	16 mL
Water	-	-	73.5 mL
Store at 4 °C for 1 year		Total Volume	100 mL

50% Iodixanol Solution	**For 50 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
Diluent Buffer	-	-	8.33 mL
60% Iodixanol	50%	1.20	41.67 mL
Store at 4 °C for 1 month		Total Volume	50 mL

1M Sucrose	**For 250 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
Sucrose (Powder)	1 M	-	85.58 g
H2O			~196.3 mL
Store at 4 °C for 1 year		Total Volume	250 mL

150 mM Spermine	**8.24 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
Spermine (Powder)	150 mM	-	1 g
H2O			To 8.24 mL
Aliquot and store at -20 °C for 1 year		Total Volume	8.24 mL

500 mM Spermidine	**1.96 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
Spermidine trihydrochloride (Powder)	500 mM	-	0.25 g
H2O			To 1.96 mL
Aliquot and store at -20 °C for 1 year		Total Volume	1.96 mL

750 mM Tricine-KOH pH 7.8	**186 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
Tricine (Powder)	750 mM	-	25 g
Potassium Hydroxide / KOH (Pellets)			As needed*
H2O			To 186 mL
Store at 4 °C for 1 year		Total Volume	186 mL

*Add KOH pellets as needed to obtain the desired pH of 7.8. Then bring the volume to 186 mL with water.

1 M DTT	**3.24 mL stock solution**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Total*
DTT (Powder)	1 M	-	0.5 g
H2O			~2.89 mL
Aliquot and store at -20 °C for 1 year		Total Volume	3.24 mL

PROCEDURE

This protocol and the provided time estimates assume the user is processing 12 samples. All steps should be performed on ice or at 4 °C. There are no pause points in this protocol but isolated nuclei can be cryopreserved and used for ATAC-seq at a later date.

Before starting the protocol:

1) Pre-chill a swinging bucket centrifuge and a fixed angle centrifuge to 4 °C.
2) Pre-chill all Dounces and pestles to 4 °C in a refrigerator.
3) Pre-chill and label all tubes. This should include the following per sample:
    a. One 2 mL round-bottom LoBind tube for gradient separation
    b. One 1.5 mL LoBind tube for RNA homogenate
    c. One 1.8 mL Nunc Cryotube for extra nuclei
    d. (optional) One 50 mL conical for filtration step
4) Fill a 2 L beaker with 500 mL sterile water which will be used to soak the used Dounces and pestles.
5) Prepare all same-day buffers as indicated below and chill to 4 °C.

Same-day Buffers:
The 1x homogenization buffer unstable solution, the 30% and 40% iodixanol solutions and the RSB-T buffer should be made fresh each time nuclei are isolated.
The numbers given in the below table are designed for 12 samples and can be scaled proportional to the number of samples being processed.

1x Homogenization Buffer Unstable Solution (HBUS)		**2083 µL per sample**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Per sample*	*Total*
1.034x Homogenization Buffer	1 M	1.03	2015 µL	24175 µL
1 M DTT	1 mM	1000	2.08 µL	25 µL
500 mM Spermidine	0.5 mM	1000	2.08 µL	25 µL
150 mM Spermine	0.15 mM	1000	2.08 µL	25 µL
10% NP40	0.3%	33.33	62.5 µL	750 µL
cOmplete Protease Inhibitor Tablets*	-	-	0.0417 tabs	0.50 tabs
		Total Volume	2083.33 µL	25000 µL

*Note that it is not feasible to accurately cut these protease inhibitor tablets into portions smaller than ¼. For this reason, we recommend making this buffer in increments of 6 samples, which requires ¼ tablet. For faster dissolution of the tablet, crush the tablet fragment to a powder prior to addition to the 1x Homogenization Buffer Unstable Solution. Ensure that all of the protease inhibitor has dissolved prior to using this 1x HBUS to make other same-day buffers.

30% Iodixanol Solution		**600 µL per sample**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Per sample*	*Total*
1x Homogenization Buffer	-	-	240 µL	3060 µL
50% Iodixanol Solution	30%	1.67	360 µL	4590 µL
		Total Volume	600 µL	7650 µL

40% Iodixanol Solution		**600 µL per sample**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Per sample*	*Total*
1x Homogenization Buffer	-	-	120 µL	1530 µL
50% Iodixanol Solution	40%	1.25	480 µL	6120 µL
		Total Volume	600 µL	7650 µL

RSB-Tween (RSB-T) Buffer		**1500 µL per sample**		
Reagent Name	*Final Conc.*	*Fold Dilution (x)*	*Per sample*	*Total*
ATAC-RSB	-	-	1485 µL	18.56 mL
10% Tween-20	0.1%	100.00	15 µL	187.5 µL
		Total Volume	3000 µL	37500 µL

Dounce Homogenization of Tissue (Timing: 1.5 h for 12 samples)

1) Remove samples from LN2 or -80 °C storage and place on dry ice until ready.
2) Transfer approximately 20 mg of frozen tissue into a pre-chilled 2 mL Dounce containing 1 mL of cold 1x HBUS. Ensure that the tissue is submerged in the 1x HBUS and fits inside of the Dounce.

CRITICAL STEP: For 10-20 mg tissue, use 1 mL of 1x HBUS. For 30-50 µm cryosections, use 0.5 mL of 1x HBUS. For 30-50 mg tissue, use 2 mL 1x HBUS. Use of more than 50 mg tissue will require scaling up the volumes used throughout this protocol, especially during the density gradient steps where large amounts of nuclei can overload the capacity of the gradient.

3) Allow tissue to thaw for 5 minutes in the 1x HBUS.
4) If planning to preform RNA-based assays on these nuclei, add 5 µL RNasin Plus RNase Inhibitor per mL of 1x HBUS and mix well. Note: for some sensitive applications of RNA, it may be necessary to add RNase inhibitor to all the buffers, at the same ratio (40 units/mL).
5) Dounce with the "A" loose pestle for 10 strokes until resistance goes away. If more strokes are necessary for the particular tissue type being used, perform up to 20 strokes. The number of strokes should be kept consistent across all samples.
   CRITICAL STEP: If residual un-homogenized tissue makes it difficult to Dounce, filter the homogenate through a 70 µm bucket filter into a pre-chilled 50 mL conical prior to proceeding.
6) Gently place the "A" pestle into the water beaker to soak for cleaning later.
7) Dounce with the "B" tight pestle for ~20 strokes.
   CRITICAL STEP: The exact number of strokes necessary is tissue-specific and should be optimized for each specific application, before performing it on all experimental conditions. The first time handling a new tissue, vary the number of strokes of pestle "B" between 10-25. The release of nuclei can be verified using a hemacytometer and Trypan Blue stain. Nuclei should appear as trypan positive spheres, amid the tissue debris.
   TROUBLESHOOTING
8) Gently place the "B" pestle into the water beaker to soak for cleaning later.
9) Load the homogenate into a 1 mL pipette tip, affix a 70 µm Flowmi strainer to the end of the tip, and filter the homogenate into a pre-chilled 2 mL LoBind tube. Repeat as necessary until all volume has been filtered and transferred.
10) Gently place the Dounce into the water beaker to soak for cleaning later.
11) Pellet the nuclei by spinning at 350 g for 5 min at 4 °C in a pre-chilled fixed-angle microcentrifuge.
12) Remove all but 50 µL of the supernatant, which contains cytoplasmic components. If planning to use this cytoplasmic fraction (for example, for bulk qPCR) transfer the supernatant to a pre-chilled 1.5 mL LoBind tube. This homogenate can be purified using the Zymo Research Direct-zol RNA Microprep kit (cat. no. R2061).
    CRITICAL STEP: If the pellet is not clearly visible or the remaining 50 µL are difficult to remove, more supernatant maybe be left in the tube. Up to 400 µL of the supernatant may be left behind, however, proportionally less of the 1x HBUS should be added in the next step.
13) Gently resuspend the nuclei (now in 50 µL) by adding 350 µL of 1x HBUS and pipetting up and down until the nuclei are fully resuspended without clumps. The final volume should be ~400 µL.

Density Gradient Centrifugation (Timing: 1.5 h for 12 samples)

14) Ensure that all Iodixanol solutions are well mixed by inverting the tubes 10 times.
15) Add 1 volume (400 µL) of 50% Iodixanol Solution to the resuspended nuclei from step 13 and mix well by pipetting. This will result in a ~25% iodixanol solution with nuclei.
16) Slowly layer 600 µL of 30% Iodixanol solution under the 25% mixture from step 15. To avoid mixing of layers, wipe the side of the pipette tip with a Kimwipe to remove residual Iodixanol solution from the external surfaces of the pipette tip.
    TROUBLESHOOTING

17) Slowly layer 600 µL of 40% Iodixanol solution under the 30% mixture. To avoid mixing of layers, wipe the side of the pipette tip with a Kimwipe to remove residual Iodixanol solution from the external surfaces of the pipette tip.
CRITICAL STEP: During this step, to gradually draw the pipette tip up as the 40% solution is pipette out to avoid overflowing the tube.
TROUBLESHOOTING
18) Centrifuge the tubes containing triple-layer gradients at 3,000 g for 20 min at 4 °C in a pre-chilled swinging bucket centrifuge with the centrifuge brake turned off. Handle tubes gently so as to not disturb the gradient.
CRITICAL STEP: It is critical to use a swinging-bucket centrifuge for this step and to have the brake turned off. Please see the **Appendix** of this protocol for suggestions on how to use common lab centrifuges for this step if a benchtop swinging-bucket centrifuge is not available.
19) Using a vacuum aspirator, remove liquid from the top of the gradient down to within 200-300 µL of the nuclei band which should be visible at the 30%-40% interface.
TROUBLESHOOTING
20) Using a 200 µL volume, collect the nuclei band and transfer it to a pre-chilled 1.5 mL LoBind tube. To do this, place the pipette tip just above the 30%-40% interface and carefully circle the tip back and forth while slowly releasing the plunger to “vacuum” up the nuclei band.
CRITICAL STEP: Do not collect more than 200 µL at this step. Inclusion of more homogenate will inevitably include more debris and unwanted materials. It is preferable to leave nuclei behind rather than collect additional volume.
21) Dilute nuclei solution by adding at least 200 µL of RSB-T Buffer and mix gently by pipetting. Depending on how many nuclei are recovered, add up to 1200 µL of RSB-T Buffer. The goal is to dilute the nuclei solution so that when they are counted on a hemocytometer, there are about 20-100 nuclei per large square.
22) Transfer 10 µL of the nuclei solution to a clean 1.5 mL tube and add 10 µL of Trypan Blue stain). Mix well by pipetting and load into a disposable hemocytometer slide.
CRITICAL STEP: We do not recommend use of automated cell counters to count nuclei as they are typically designed for counting of live cells.
23) Count the number of intact nuclei (examples shown in **Supplementary Figure 3**) in the 4 large corner squares. Take the total number in all 4 squares and multiply by 5000 to obtain the concentration of nuclei per mL.
TROUBLESHOOTING

These nuclei are now ready for ATAC-seq and can be used as input to **Step 8** of the main ATAC-seq protocol. These nuclei can also be used for single-nucleus ATAC-seq and other epigenomic and transcriptomic assays or cryopreserved for later use. We find minimal effect on ATAC-seq signal with freeze-thaw of nuclei when cryopreserved as described below.

Cryopreservation of Nuclei and Freezing Down Tubes (Timing: 1 h for 12 samples)

24) Pellet remaining nuclei by centrifugation at 500 g for 10 min at 4 °C in a fixed-angle centrifuge.
25) Aspirate all the supernatant using two pipetting steps. First, aspirate down to 100 µL with a p1000 pipette. Then, remove the final 100 µL with a p200 pipette.
CRITICAL STEP: Make sure to avoid the visible cell pellet when pipetting. Optimal removal of supernatant and minimal disruption of the cell pellet is attained when the removal of the final 100 µL is performed in a consistent and fluid motion without starting and stopping.

26) Gently resuspend the pellet, containing up to 1 million nuclei, in 100 µL of cold BAM Banker cryopreservation media and transfer to a pre-chilled 1.8 mL Nunc cryovial. If more nuclei are present in the pellet, increase the volume of BAM Banker cryopreservation media proportionally.
27) Transfer the cryotubes to a freezing container and place at -80 °C to allow the nuclei to slowly freeze.
28) Tubes can be removed from the freezing container and transferred to other -80 °C storage or liquid nitrogen storage after 2 hours. See the "Cryopreserved cells or nuclei" section of the main protocol for instructions on how to use these cryopreserved nuclei in downstream assays.
29) If homogenate was collected in **Step 12** of this nuclei isolation protocol, store this homogenate at -80 °C.
30) Clean all Dounces and pestles by thoroughly spraying them individually with water followed by 70% ethanol.

TROUBLESHOOTING

Step	Problem	Possible Reason	Solution
7	There is a lot of tissue debris after Douncing with pestle "A" and this makes it hard to use pestle "B".	The tissue has a lot of ECM or other hard to break up components.	Before moving onto the "B" pestle, filter the cell lysate through a 70 µm cell strainer.
7	Clear nuclei are not observed after 15 strokes of the B pestle.	The tissue has a lot of ECM or other hard to break up components.	We recommend calibrating the number of A and B strokes used for a particular cell type. Inspect each sample after using the B pestle and determine if released nuclei are observed. If not, add additional "B" strokes; titrate the number of strokes in increments of 5, checking the nuclei integrity on a pilot tissue sample. Once the optimal number of strokes has been observed, it should be kept consistent between all samples processed.
16-17	Iodixanol layers appear to "float" upwards as the layering is being performed.	Iodixanol solutions were not well mixed.	Mix all iodixanol solutions well before use, by inverting at least 10 times.

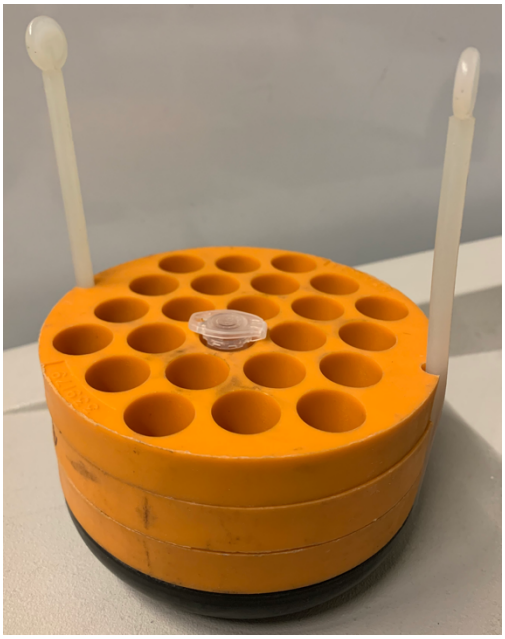
19	A clear band of nuclei is not observed at the 30-40% interface.	The input material was less than 10 mg or contained very few nuclei.	If input materials are very small, it may be difficult to observe nuclei in a band. The 30-40% interface can be marked by height on the tube on a practice tube where the layers are colored using a dye.
19	A clear band of nuclei is not observed at the 30-40% interface.	The iodixanol solution is old or not properly diluted.	Make fresh 50% solutions every month.
23	The nuclei have a large amount of tissue debris after isolation.	Too much tissue was loaded into the isolation and overwhelmed the gradient or recovery of the nuclei band inadvertently included material below or above the band.	Ensure that ~10-20 mg of tissue is loaded per 1 mL of homogenization buffer. For larger tissue pieces, increase the number of gradient spin tubes utilized per sample or increase the total volume of each gradient layer by using larger tubes.
24	The cell or nuclei pellet is hard to visualize after centrifugation.	Nuclei and certain cell types can be hard to properly pellet. For 50,000 cells, pellets should be easily observed, even for small cell types such as B cells.	Add sterile BSA to a final concentration of 0.5% wt/vol or Tween-20 to a final concentration of 0.1% wt/vol to help pellet the cells/nuclei.

TIMING (for 12 samples)
Steps 1-13, Dounce Homogenization of Tissue, 1.5 h
Steps 14-23, Density Gradient Centrifugation, 1.5 h
Steps 24-30, Cryopreservation of Nuclei and Freezing Down Tubes, 1h

ANTICIPATED RESULTS
The yield of nuclei will depend on the particular tissue sample and its cellular density as well as the amount of extra cellular matrix (ECM). However, in general, nuclei should appear round, stain positive for trypan blue, and be free of any cytoplasm or cytoplasmic membrane. Nuclei that appear “raindrop-shaped” (i.e. a droplet with a cytoplasmic tail) or as crescent shapes indicate a low-quality input sample, as does the abundance of small cellular debris. For examples of the expected nuclei shape, refer to **Supplementary Figure 3.**

APPENDIX

Most labs are not equipped with a benchtop swinging bucket rotor. However, standard bucket rotors can be used with disc adapters for 5 mL flow cytometry tubes which fit the 2 mL LoBind tubes quite well. If using this setup, make sure that some of the disc adapters are removed such that the bottom of the 2 mL LoBind tube is resting on the bottom of the adapter.



Supplementary Methods

DNA clean up column optimization

ATAC-seq libraries were prepared as per the protocol using GM12878 cells, using either the Zymo Research DNA Clean & Concentrator-5 kit (Zymo Research, cat. no. D4014), the QIAGEN MinElute PCR Purification kit (Qiagen, cat. no. 28004), or the NEB Monarch PCR & DNA Cleanup kit (New England Biolabs, cat. no. T1030S). The concentration of each library after amplification was determined using qPCR using the NEBNext Library Quant kit (New England Biolabs, cat. no. E7630). Concentration values for each sample were normalized to the average of the trials from the Zymo DNA Clean & Concentration-5 kit.

Optimization of PCR enzyme mix and cycling conditions

ATAC-seq library concentration was determined after pre-amplification for 5 cycles with either the NEBNext High-Fidelity 2x PCR Master Mix (New England Biolabs, cat. no. M0541L) or the NEBNext Ultra II Q5 2x Master Mix (New England Biolabs, cat. no. M0544L). Library concentration was calculated using the NEBNext Library Quant kit. After choosing the NEBNext Ultra II Q5 2x Master Mix, the cycling conditions were optimized by testing the annealing and extension temperatures listed in **Supplementary Figure 1c**. Library concentration was measured after 5 cycles of amplification using the NEBNext Library Quant kit.

Tn5 concentration scaling

GM12878 cells were cultured in RPMI + 10% FBS and 1% pen-strep and aliquoted into different input quantities (6,000, 12,000, 25,000, 50,000, 100,000, 250,000 and 500,000). Tn5 reactions were scaled according to **Supplementary Note 2** and prepared for sequencing. The resulting libraries were sequenced on a NextSeq500 sequencer using a high-output flow cell and 75-cycle kit according to the manufacturer's instructions. The following read lengths were used: 33 bp Read 1, 33 bp Read 2, 8 bp Index 1, and 8 bp Index 2. ATAC-seq data was processed using PEPATAC with Bulker (container version 1.0.8) and the hg38 reference genome. PEPATAC was used to calculate the TSS enrichment score. For plotting of ATAC-seq signal tracks, a custom script was used to generate the depth normalized bigWig files using a bin size of 100 bp.

Qubit estimation of cycle number

To derive the equation used to estimate the required number of additional cycles from Qubit concentration readings (see **Box 2** of the main text) ATAC-seq libraries were prepared from GM12878 cells as described, using a variety of starting cell numbers to give different library concentrations. Before qPCR library quantification (in **Steps 28-34**) library concentration was determined using the Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q33231). 1 µL of the library (undiluted) was used to determine the library concentration, as per the manufacturer's instructions. The samples were then quantified using the qPCR approach and the additional cycle number was determined using the equation in **Step 34.** To derive the Qubit concentration to cycle number equation, 75% of all collected data points (42 total trials, 31 used for training) were used to derive the equation. The Qubit concentration was log10 transformed and a linear regression was performed between log10(concentration) and cycle number, deriving the equation in **Supplementary Figure 5A**. The remaining 25% of the data (11 samples) were used to test the accuracy of this equation.

Peak Calling Comparison

Peaks used: Data from five distinct hematopoietic cell types from *Corces & Buenrostro et al. 2016* were used to illustrate various peak merging procedures. FASTQ files were obtained from GEO (GSE75384) and aligned to the hg38 reference genome using PEPATAC[22] with Bulker[23]

(container version 1.0.8). Peaks were called using MACS2 with parameters "--shift -75 --extsize 150 --nomodel --call-summits --nolamda --keep-dup all -p 0.01". *Raw overlap:* "bedtools merge" (v2.30.0) was used with default parameters to create the final merged peak set. *Clustered overlap:* "bedtools cluster" (v2.30.0) was used with default parameters. After clustering, peaks were sorted by the -log10(qValue) and the most significant peak per cluster was retained in the final merged peak set. *Iterative overlap:* The custom script provided with this protocol (see the **Code Availability** section in the main text) was used to perform iterative overlap peak merging as described previously[24]. This procedure involves both a filtering step based on "score-per-million" as described in the main text as well as the iterative overlap procedure itself. Data was visualized using bigWigs tracks, with a bin size of 100 bp.

ATAC-seq read length vs. mapping efficiency

Data from four sources was used to determine the effect of read length on the mapping efficiency of ATAC-seq data. Data for the GM12878 and K562 cell lines (this work) have been deposited on GEO. The lung adenocarcinoma data was derived from *Corces et al. 2018* [24] (available through the Genomic Data Commons Portal via data access request https://portal.gdc.cancer.gov/) and the postmortem human brain sample from GSE147672. Data was then processed with PEPATAC using the --trimmer pyadapt option for each read length.

Supplementary References

1. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
2. Sato, S. *et al.* Biochemical analysis of nucleosome targeting by Tn5 transposase. *Open Biol.* **9**, 190116 (2019).
3. Zeitlinger, J. Seven myths of how transcription factors read the cis-regulatory code. *Curr. Opin. Syst. Biol.* **23**, 22–31 (2020).
4. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* **217**, 1181–1191 (2018).
5. Fudenberg, G. & Mirny, L. A. Higher order chromatin structure: bridging physics and biology. *Curr. Opin. Genet. Dev.* **22**, 115–124 (2012).
6. Xin, B. & Rohs, R. Relationship between histone modifications and transcription factor binding is protein family specific. *Genome Res.* **28**, 321–333 (2018).
7. Guertin, M. J. & Lis, J. T. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Curr. Opin. Genet. Dev.* **23**, 116–123 (2013).
8. Huertas, J. & Cojocaru, V. Breaths, Twists, and Turns of Atomistic Nucleosomes. *J. Mol. Biol.* **433**, 166744 (2021).
9. Lobbia, V. R., Trueba Sanchez, M. C. & van Ingen, H. Beyond the Nucleosome: Nucleosome-Protein Interactions and Higher Order Chromatin Structure. *J. Mol. Biol.* **433**, 166827 (2021).
10. Abdulhay, N. J. *et al.* Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).
11. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
12. Quach, B. & Furey, T. S. DeFCoM: analysis and modeling of transcription factor binding sites using a motif-centric genomic footprinter. *Bioinformatics* **33**, 956–963 (2017).
13. Baek, S., Goldstein, I. & Hager, G. L. Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. *Cell Rep.* **19**, 1710–1722 (2017).
14. Raj, A., Shim, H., Gilad, Y., Pritchard, J. K. & Stephens, M. msCentipede: Modeling Heterogeneity across Genomic Sites and Replicates Improves Accuracy in the Inference of Transcription Factor Binding. *PLOS ONE* **10**, e0138030 (2015).
15. Sherwood, R. I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
16. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* **21**, 22 (2020).
17. Bajic, M., Maher, K. A. & Deal, R. B. Identification of Open Chromatin Regions in Plant Genomes Using ATAC-Seq. *Methods Mol. Biol. Clifton NJ* **1675**, 183–201 (2018).
18. Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
19. Keilwagen, J. & Grau, J. Varying levels of complexity in transcription factor binding motifs. *Nucleic Acids Res.* **43**, e119–e119 (2015).
20. He, H. H. *et al.* Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73–78 (2014).
21. Swinstead, E. E. *et al.* Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell* **165**, 593–605 (2016).
22. Smith, J. P. *et al.* PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genomics Bioinforma.* **3**, lqab101 (2021).

23. Sheffield, N. C. Bulker: a multi-container environment manager. (2019) doi:10.31219/osf.io/natsj.
24. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).