Efficient transcription-coupled chromatin accessibility mapping *in situ*

Steven Henikoff[1,2]*, Jorja G. Henikoff[1], Hatice S. Kaya-Okur[1,3] and Kami Ahmad[1]

[1]Basic Sciences Division

Fred Hutchinson Cancer Research Center,

1100 Fairview Ave N, Seattle, Washington, USA 98109

[2]Howard Hughes Medical Institute

[3]Present address:

Altius Institute for Biomedical Sciences

2211 Elliott Ave #4, Seattle, WA 98121

Corresponding author: Steven Henikoff

Email: steveh@fhcrc.org

https://orcid.org/0000-0002-7621-8685

Abstract

Chromatin accessibility mapping is a powerful approach to identify potential regulatory elements. In the popular ATAC-seq method, Tn5 transposase inserts sequencing adapters into accessible DNA ('tagmentation'). CUT&Tag is a tagmentation-based epigenomic profiling method in which antibody tethering of Tn5 to a chromatin epitope of interest profiles specific chromatin features in small samples and single cells. Here we show that by simply modifying the tagmentation conditions for histone H3K4me2/3 CUT&Tag, antibody-tethered tagmentation of accessible DNA sites is redirected to produce accessible DNA maps that are indistinguishable from the best ATAC-seq maps. Thus, DNA accessibility maps can be produced in parallel with CUT&Tag maps of other epitopes with all steps from nuclei to amplified sequencing-ready libraries performed in single PCR tubes in the laboratory or on a home workbench. As H3K4 methylation is produced by transcription at promoters and enhancers, our method identifies transcription-coupled accessible regulatory sites.

Keywords

epigenomic profiling; chromatin accessibility; histone modifications; CUT&Tag

Introduction

Identification of DNA accessibility in the chromatin landscape has been used to infer active transcription ever since the seminal description of DNaseI hypersensitivity by Weintraub and Groudine more than 40 years ago (1). Because nucleosomes occupy most of the eukaryotic chromatin landscape and regulatory elements are mostly free of nucleosomes when they are active, DNA accessibility mapping can potentially identify active regulatory elements genome-wide. Several additional strategies have been introduced to identify regulatory elements by DNA accessibility mapping, including digestion with Micrococcal Nuclease (MNase) (2) or restriction enzymes (3), DNA methylation (4), physical fragmentation (5) and transposon insertion (6). With the advent of genome-scale mapping platforms, beginning with microarrays and later short-read DNA sequencing, mapping regulatory elements based on DNaseI hypersensitivity became routine (7, 8). Later innovations included FAIRE (9) and Sono-Seq (10), based on physical fragmentation and differential recovery of cross-linked chromatin, and ATAC-seq (11), based on preferential insertion of the Tn5 transposase. The speed and simplicity of ATAC-seq, in which the cut-and-paste transposition reaction inserts sequencing adapters in the most accessible genomic regions (tagmentation), has led to its widespread adoption in many laboratories for mapping presumed regulatory elements.

For all of these DNA accessibility mapping strategies, it is generally unknown what process is responsible for creating any particular accessible sites within the chromatin landscape. Furthermore accessibility is not all-or-none, with the median difference between an accessible and a non-accessible site in DNA estimated to be only ~20%, with no sites completely accessible or inaccessible in a population of cells (12, 13). Despite these uncertainties, DNA accessibility mapping has successfully predicted the locations of active gene enhancers and promoters genome-wide, with excellent correspondence between methods based on very different strategies (14). This is likely because DNA accessibility mapping strategies rely on the fact that nucleosomes have evolved to repress transcription by blocking sites of pre-initiation complex formation and transcription factor binding (15), and so creating and maintaining a nucleosome-depleted region (NDR) is a pre-requisite for promoter and enhancer function.

A popular alternative to DNA accessibility mapping for regulatory element identification is to map nucleosomes that border NDRs, typically by histone marks, including "active" histone modifications, such as H3K4 methylation and H3K27 acetylation, or histone variants incorporated during transcription, such as H2A.Z and H3.3. The rationale for this mapping strategy is that the enzymes that modify histone tails and the chaperones that deposit nucleosome subunits are most active close to the sites of initiation of transcription, which typically occurs bidirectionally at both gene promoters and enhancers to produce stable mRNAs and unstable enhancer RNAs. Although the marks left behind by active transcriptional initiation "point back" to the NDR, this cause-effect connection between the NDR and the histone marks is only by inference (16), and direct evidence is lacking that a histone mark is associated with an NDR.

Here we show that a simple modification of our Cleavage Under Targets & Tagmentation (CUT&Tag) method for antibody-tethered *in situ* tagmentation can identify NDRs genome-wide at regulatory elements adjacent to transcription-associated histone marks in human cells. We provide evidence that reducing the ionic concentration during tagmentation preferentially attracts Tn5 tethered to the H3K4me2 histone modification via a Protein A/G fusion to the nearby NDR, shifting the site of tagmentation from nucleosomes bordering the NDR to the NDR itself. Practically all transcription-coupled accessible sites correspond to ATAC-seq sites and vice-versa, and lie upstream of paused RNA Polymerase II (RNAPII). "CUTAC" (Cleavage Under Targeted Accessible Chromatin) is conveniently performed in parallel with ordinary CUT&Tag, producing accessible site maps from low cell numbers with signal-to-noise as good as or better than the best ATAC-seq datasets.

Results

Streamlined CUT&Tag produces high-quality datasets with low cell numbers. We previously introduced CUT&RUN, a modification of Laemmli's Chromatin Immunocleavage (ChIC) method (17), in which a fusion protein between Micrococcal Nuclease (MNase) and Protein A (pA-MNase) binds sites of antibody binding in nuclei or permeabilized cells bound to magnetic beads. Activation of MNase with Ca^{++} results in targeted cleavage releasing the antibody-bound fragment into the supernatant for paired-end DNA sequencing. More recently, we substituted the Tn5 transposase for MNase in a modified CUT&RUN protocol, such that addition of Mg^{++} results in a cut-and-paste "tagmentation" reaction, in which sequencing adapters are integrated around sites of antibody binding (18). In CUT&Tag, DNA purification is followed by PCR amplification, eliminating the end-polishing and ligation steps required for sequencing library preparation in CUT&RUN. Like CUT&RUN, CUT&Tag requires relatively little input material, and the low backgrounds permit low sequencing depths to sensitively map chromatin features.

We have developed a streamlined version of CUT&Tag that eliminates tube transfers, so that all steps can be efficiently performed in a single PCR tube (19). However, we had not determined the suitability of the single-tube protocol for profiling low cell number samples. During the COVID-19 pandemic, we adapted this CUT&Tag-direct protocol for implementation with minimal equipment and space requirements that uses no toxic reagents, so that it can be performed conveniently and safely on a home workbench (**Fig. S1**). To ascertain the ability of our CUT&Tag-direct protocol to produce DNA sequencing libraries at home with data quality comparable to those produced in the laboratory, we used frozen aliquots of native human K562 cell nuclei prepared in the laboratory and profiled there using the streamlined single-tube protocol. Aliquots of nuclei were thawed and serially diluted in Wash buffer from ~60,000 down to ~60 starting cells, where the average yield of nuclei was ~50%. We used antibodies to H3K4me3, which preferentially marks nucleosomes immediately downstream of active promoters, and H3K27me3, which marks nucleosomes within broad domains of

Polycomb-dependent silencing. Aliquots of nuclei were taken home and stored in a kitchen freezer, then thawed and diluted at home and profiled for H3K4me3 and H3K27me3. In both the laboratory and at home we performed all steps in groups of 16 or 32 samples over the course of a single day through the post-PCR clean-up step, treating all samples the same regardless of cell numbers. Whether produced at home or in the lab, all final barcoded sample libraries underwent the same quality control, equimolar pooling, and final SPRI bead clean-up steps in the laboratory prior to DNA sequencing.

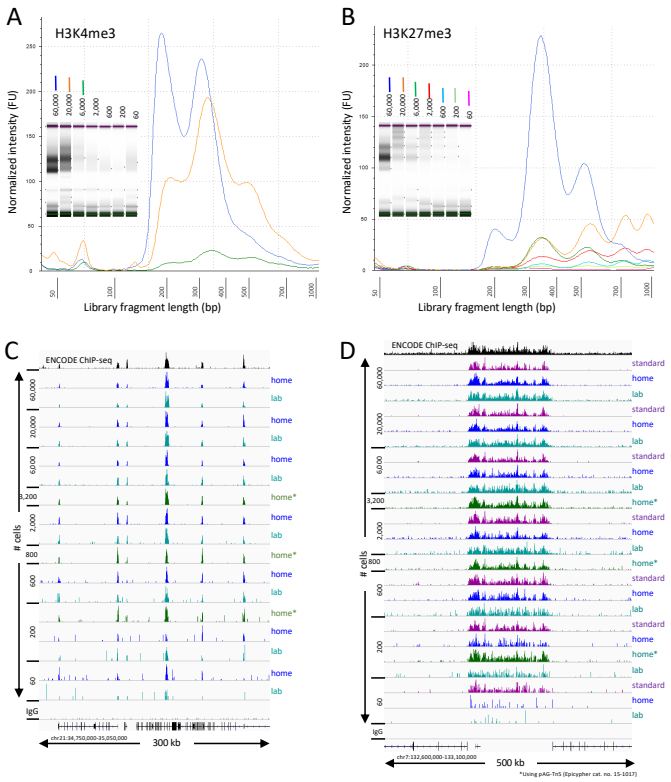

Figure 1. CUT&Tag-direct produces high-quality datasets on the benchtop and at home. Starting with a frozen human K562 cell aliquot, CUT&Tag-direct with amplification for 12 cycles yields detectable nucleosomal ladders for intermediate and low numbers of cells for both (**A**) H3K4me3 and (**B**) H3K27me3. **C**) Comparison of H3K4me3 CUT&Tag-direct results produced in the laboratory to those produced at home and to an ENCODE dataset (GSM733680). **D**) Same as (C) for H3K27me3 comparing CUT&Tag-direct results to CUT&Tag datasets using the standard protocol (18), and to an ENCODE dataset (GSM788088). pA-Tn5 was used except as indicated by asterisks for datasets produced at home using commercial pAG-Tn5 (Epicypher cat. no. 15-1017).

Tapestation profiles of libraries produced at home detected nucleosomal ladders down to 200 cells for H3K27me3 and nucleosomal and subnucleosomal fragments down to 2000 cells for H3K4me3 (**Fig. 1A-B**). Sequenced fragments were aligned to the human genome using Bowtie2 and tracks were displayed using IGV. Similar results were obtained for both at-home and in-lab profiles for both histone modifications (**Fig. 1C-D**) using pA-Tn5 produced in the laboratory, and results using commercial Protein A/Protein G-Tn5 (pAG-Tn5) were at least as good. All subsequent experiments reported here were performed at home using commercial pAG-Tn5, which provided results similar to those obtained using batches of homemade pA-Tn5 run in parallel.

NDRs attract Tn5 tethered to nearby nucleosomes during low-salt tagmentation. Because the Tn5 domain of pA-Tn5 binds avidly to DNA, it is necessary to use elevated salt conditions to avoid tagmenting accessible DNA during CUT&Tag. High-salt buffers included 300 mM NaCl for pA-Tn5 binding, washing to remove excess protein, and tagmentation at 37ºC. We have found that other protocols based on the same principle but that do not include a high-salt wash step result in chromatin profiles that are dominated by accessible site tagmentation (19).

To better understand the mechanistic basis for the salt-suppression effect, we bound pAG-Tn5 under normal high-salt CUT&Tag incubation conditions, then tagmented in low salt. We used either rapid 20-fold dilution with a prewarmed solution of 2 mM or 5 mM $MgCl_2$ or removal of the pAG-Tn5 incubation solution and addition of 50 µL 10 mM TAPS pH8.5, 5 mM $MgCl_2$. All other steps in the protocol followed our CUT&Tag-direct protocol (19) (**Fig. 2**). Tapestation capillary gel electrophoresis of the final libraries revealed that after a 20 minute incubation the effect of low-salt tagmentation on H3K4me2 CUT&Tag samples was a marked reduction in the oligo-nucleosome ladder with an increase in faster migrating fragments (**Fig. 3A**). CUT&Tag profiles using antibodies to most chromatin epitopes in the dilution protocol showed either little change or elevated levels of non-specific background tagmentation that obscured the targeted signal (**Fig. S2**), as expected considering that we had omitted the high-salt wash step needed to remove unbound pAG-Tn5. Strikingly, under low-salt conditions, high-resolution profiles of H3K4me3 and H3K4me2 showed that the broad nucleosomal distribution of CUT&Tag around promoters for these two modifications was mostly replaced by single narrow peaks (**Figs. 3B, S3-S4A-B**).

To evaluate the generality of peak shifts we used MACS2 to call peaks, and plotted the occupancy over aligned peak summits. For all three H3K4 methylation marks using normal CUT&Tag high-salt tagmentation conditions we observed a bulge around the summit representing the contribution from adjacent nucleosomes on one side or the other of the peak summit (**Fig. 3C**). In contrast, tagmentation under low-salt conditions revealed much narrower profiles for H3K4me3 and H3K4me2 (~40% peak width at half-height), less so for H3K4me1 (~60%), which suggests that the shift is from H3K4me-marked nucleosomes to an adjacent NDR.

To determine whether free pAG-Tn5 present during tagmentation contributes, we removed the pAG-Tn5 then added 5 mM $MgCl_2$ to tagment, and again observed narrowing of the H3K4me2 peak (**Fig. 3D "Removal"**). We also observed a narrowing if we included a stringent 300 mM washing step before low-salt tagmentation (**Fig. 3D, "Post-wash"**), which indicates that peak narrowing does not require free pAG-Tn5. However, we noticed that the peak narrowed further if following the stringent wash low-salt tagmentation we included a small amount of pAG-Tn5 and incubation was extended from 20 min to 1 hr (**Fig. 3D "Add-back"**). Because Tn5 is inactive once it integrates its payload of adapters, and each fragment is generated by tagmentation at both

ends, it is likely that a small amount of free pA(G)-Tn5 is sufficient to generate the additional small fragments where tethered pA(G)-Tn5 is limiting.

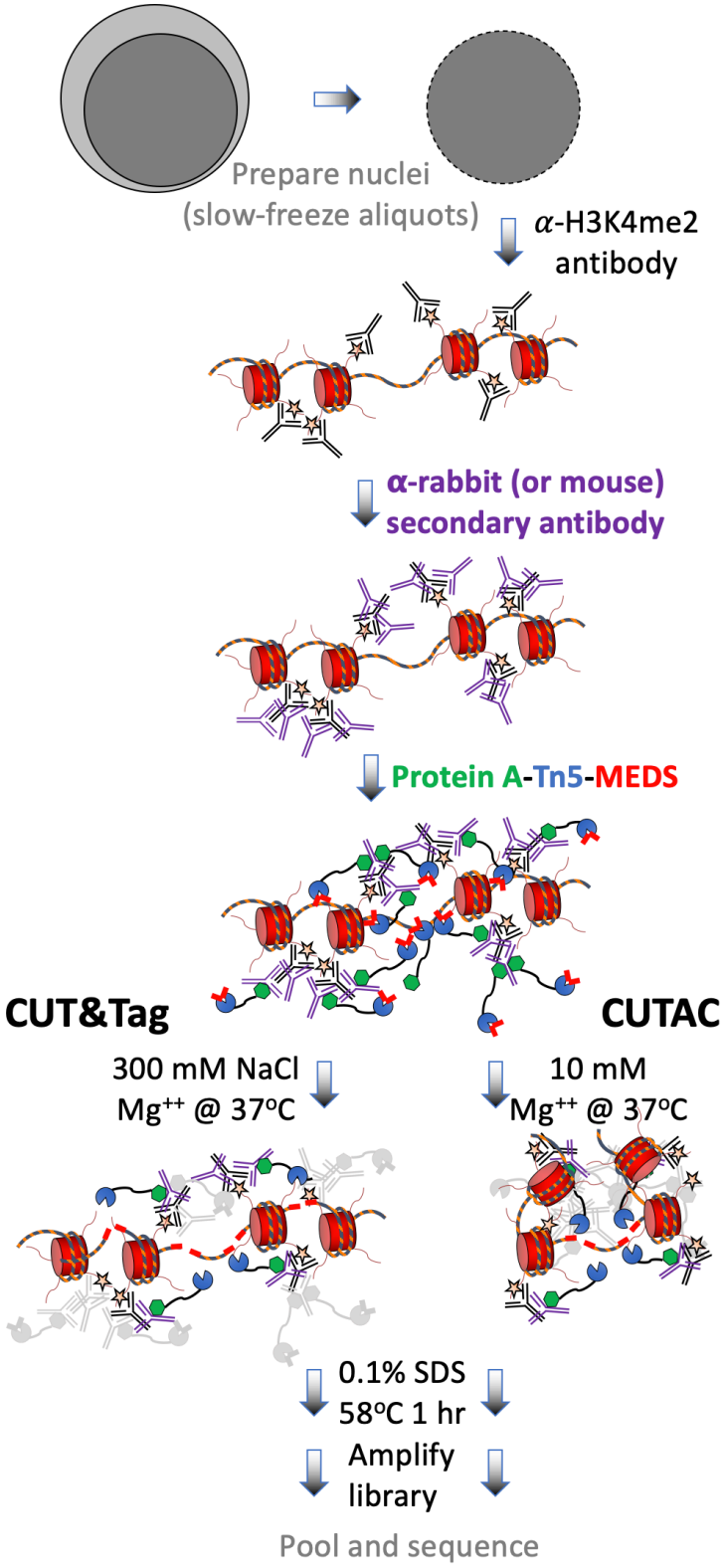

Figure 2. CUT&Tag with low-salt tagmentation (CUTAC). Steps in grey are lab-based and other steps were performed at home. Tagmentation can be performed by dilution, removal or post-wash. MEDS (Mosaic End Double-Stranded annealed oligonucleotides).

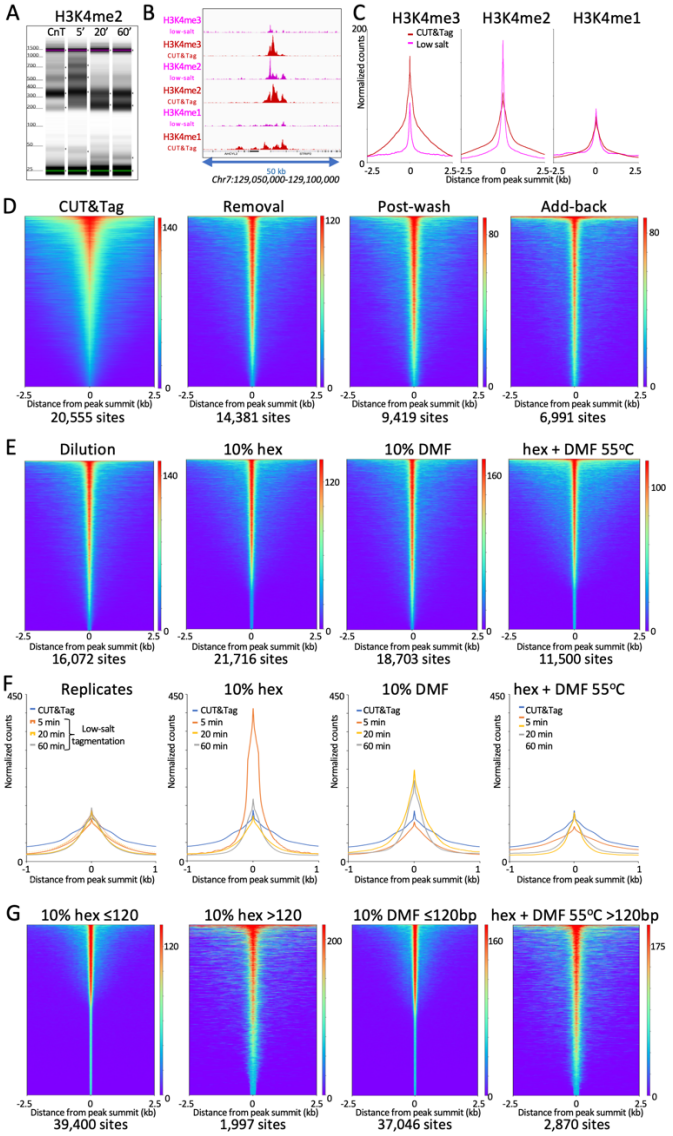

Figure 3. Low-salt tagmentation of H3Kme2/3 CUT&Tag samples sharpen peaks. A) Tapestation gel image showing the change in size distribution from standard CUT&Tag (CnT), tagmented in the presence of 300 mM NaCl with low-salt tagmentation using the dilution protocol. **B**) Representative tracks showing the shift observed with low-salt dilution tagmentation. **C**) Average plots showing the narrowing of peak distributions upon low-salt tagmentation using the dilution protocol. **D**) Heatmaps showing narrowing of H3K4me2 peaks after removing pAG-Tn5 (Removal), after a stringent wash (Post-wash), and after a stringent wash with low-salt tagmentation including a 1% pAG-Tn5 spike-in (Add-back). MAC2 was used to call peaks and heatmaps were ordered by density over the peak midpoints (sites). **E**) Heatmaps showing dilution tagmentation and further narrowing of H3K4me2 peak distributions upon low-salt tagmentation (after removal) for 20 minutes at 37°C in the presence of 10% 1,6-hexanediol (hex) and 10% dimethylformamide (DMF) or both for 1 hr at 55°C. **F**) Average plots showing effects of tagmentation with hex and/or DMF over time of low-salt tagmentation (after removal). **G**) Smaller fragments (≤120 bp) dominate NDRs. Comparisons of small (≤120 bp) and large (>120) fragments from CUTAC hex and DMF datasets show narrowing for small fragments around their summits. For each dataset a 3.2 million fragment random sample was split into small and large fragment groups, Removal of large fragments increases number of peaks called (sites).

Salt ions compete with protein-DNA binding and so we suppose that tagmentation in low salt resulted in increased binding of epitope-tethered Tn5 to a nearby NDR and then tagmentation. As H3K4 methylation is deposited in a gradient of tri- to di- to mono-methylation downstream of the +1 nucleosome from the transcriptional start site (TSS) (20), we reasoned that the closer proximity of di- and tri-methylated nucleosomes to the NDR than mono-methylated nucleosomes resulted in preferential proximity-dependent "capture" of Tn5. Consistent with this interpretation, we observed that the shift from broad to more peaky NDR profiles and heatmaps by H3K4me2 low-salt tagmentation was enhanced by addition of 1,6-hexanediol, a strongly polar aliphatic alcohol, and by 10% dimethylformamide, a strongly polar amide, both of which enhance chromatin accessibility (**Figs. 3E-F, S4C-D**). NDR-focused tagmentation persisted even in the presence of both strongly polar compounds at 55ºC. Enhanced localization by chromatin-disrupting conditions suggests improved access of H3K4me2-tethered Tn5 to nearby holes in the chromatin landscape during low-salt tagmentation. Localization to NDRs is more precise for small (≤120 bp) than large (>120) tagmented fragments, and by resolving more closely spaced peaks increases the number of peaks called (**Figs. 3D, S5-S6**).

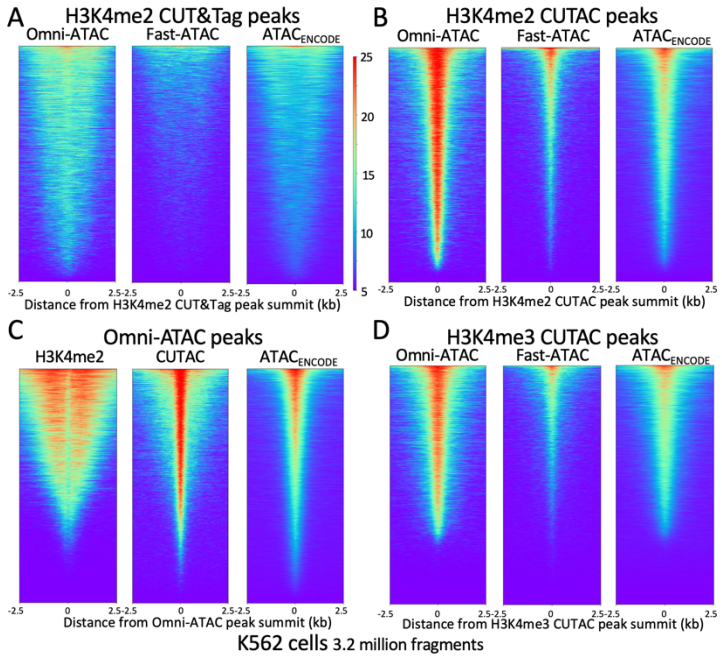

Figure 4. H3K4me2 CUTAC sites coincide with ATAC-seq sites. Heatmaps showing the correspondence between CUTAC and ATAC-seq sites. Headings over each heatmap denote the source of mapped fragments mapping to the indicated set of MACS2 peak summits, ordered by occupancy over the 5 kb interval centered over each site. CUT&Tag and CUTAC sites are from samples processed in parallel, where CUTAC tagmentation was performed by 20-fold dilution and 20 minute 37ºC incubation following pAG-Tn5 binding.

CUT&Tag low-salt tagmentation fragments coincide with ATAC-seq sites. Using CUT&Tag, we previously showed that most ATAC-seq sites are flanked by H3K4me2-marked nucleosomes in K562 cells (18). However, lining up ATAC-seq datasets over peaks called using H3K4me2 CUT&Tag data showed smeary heatmaps, reflecting the broad distribution of peak calls over nucleosome positions flanking NDRs (**Fig. 4A**). In contrast, alignment of ATAC-seq datasets over peaks called using low-salt tagmented CUT&Tag data produced narrow heatmap patterns for the vast majority of peaks (**Fig. 4B**). We confirmed these observations by the reciprocal comparisons, aligning H3K4me2 CUT&Tag and CUTAC datasets over peaks called from Omni-ATAC data (**Fig. 4C**). To reflect the close similarities between fragments released by H3K4me2-tethered low-salt tagmentation as by ATAC-seq using untethered Tn5, we will refer to low-salt H3K4me2/3 CUT&Tag tagmentation as Cleavage Under Targeted Accessible Chromatin (CUTAC).

To further evaluate the degree of similarity between CUTAC and ATAC-seq, we aligned the ENCODE ATAC-seq dataset over peaks called using Omni-ATAC and CUTAC, where all datasets were sampled down to 3.2 million mapped fragments with mitochondrial fragments removed. Remarkably, heatmaps produced using either Omni-ATAC or CUTAC peak calls for the same ENCODE ATAC-seq data showed occupancy of ~95% for both sets of peaks (compare right panels of **Fig 4B and 4C**). Using a window of 250 bp around the peak summit based on average peak width at half-height, we found ~50% overlap between ENCODE ATAC-seq peaks and peaks called from either Omni-ATAC (50.0%) or CUTAC (51.3%) data. This equivalence between H3K4me2 CUTAC and Omni-ATAC when compared to ENCODE ATAC-seq implies that CUTAC and Omni-ATAC are indistinguishable in detecting the same chromatin features. This conclusion does not hold for H3K4me3 CUTAC, because similar alignment of ENCODE ATAC-seq data resulted in only ~75% peak occupancy (**Fig. 4D**), which we attribute to the greater enrichment of H3K4me3 around promoters than enhancers relative to H3K4me2.

To sensitively evaluate signal-to-noise genome-wide, we called peaks using MACS2 and calculated the Fraction of Reads in Peaks (FRiP), a data quality metric introduced by the ENCODE project (24). For both ENCODE ChIP-seq and our published CUT&RUN data we measured FRiP ~ 0.2 for 3.2 million fragments, whereas for CUT&Tag, FRiP ~ 0.4, reflecting improved signal-to-noise relative to previous chromatin profiling methods (18). Using CUT&Tag-direct, H3K4me2 CUT&Tag FRiP = 0.41 for 3.2 million fragments and ~16,000 peaks (n=4), whereas tagmentation by dilution in 2 mM $MgCl_2$ resulted in FRiP = 0.18 for 3.2 million fragments and ~15,000 peaks (n=4) with similar values for tagmentation by removal [FRiP = 0.21, ~15,000 peaks (n=4)]. In add-back experiments, we measured lower FRiP values after stringent washing conditions whether or not some pAG-Tn5 was added back.

We also compared the number of peaks and FRiP values for CUTAC to those for ATAC-seq for K562 cells and observed that CUTAC data quality was similar to that for the Omni-ATAC method (22), better than ENCODE ATAC-seq (25), and much better than Fast-ATAC (21), a previous improvement over Standard ATAC-seq (11) (**Fig. 5A**). CUTAC is relatively insensitive to tagmentation times, with similar numbers of peaks

and similar FRiP values for samples tagmented for 5, 20 and 60 minutes (**Fig. 5A**). We attribute the robustness of CUT&Tag and CUTAC to the tethering of Tn5 to specific chromatin epitopes, so that when tagmentation goes to completion there is little untethered Tn5 that would increase background levels. When we measured peak numbers and FRiP values for ATAC-seq for K562 data deposited in the Gene Expression Omnibus (GEO) from multiple laboratories, we observed a wide range of data quality (**Fig. 5B**, even from very recent submissions (**Fig. S5**). We attribute this variability to the difficulty of avoiding background tagmention by excess free Tn5 in ATAC-seq protocols and subsequent release of non-specific nucleosomal fragments (26).

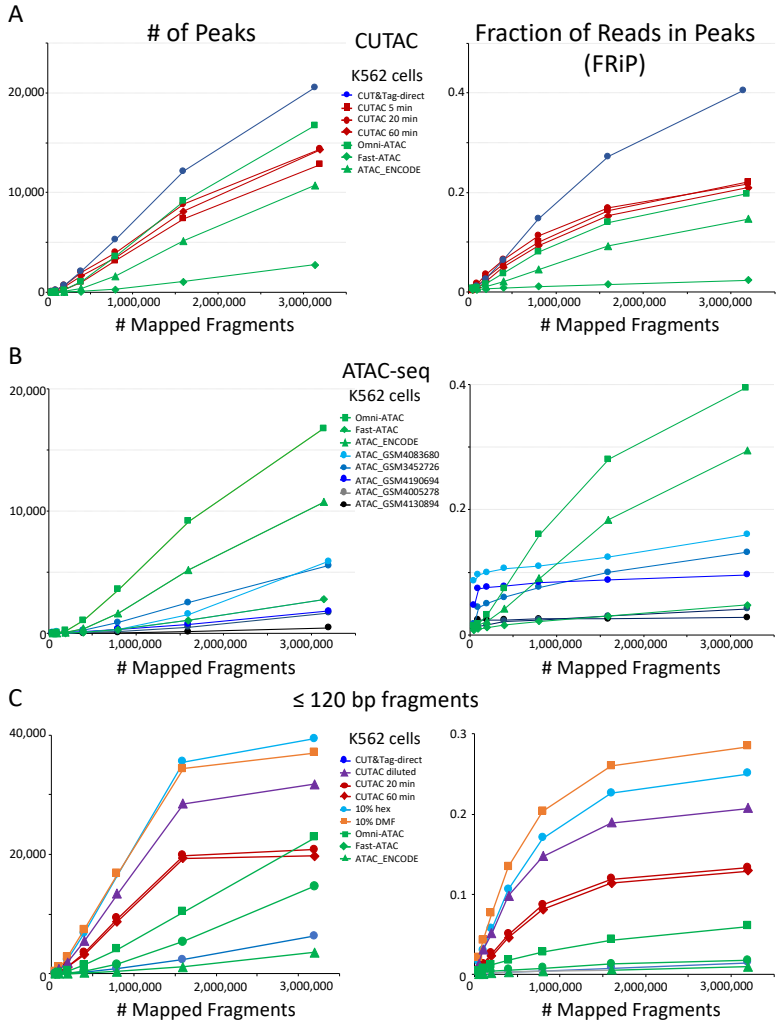

Figure 5. CUTAC data quality is similar to the best available ATAC-seq K562 cell data. Mapped fragments from the indicated datasets were sampled and mapped to hg19 using Bowtie2, and peaks were called using MACS2. **A**) Number of peaks (left) and fraction of reads in peaks for CUT&Tag (blue), CUTAC (red) and ATAC-seq (green). Fast-ATAC is an improved version of ATAC-seq that reduces mitochondrial reads (21), and Omni-ATAC is an improved version that additionally improved signal-to-noise (22). ATAC-seq_ENCODE is the current ENCODE standard (23). **B**) Five other K562 ATAC-seq datasets from different laboratories were identified in GEO and mapped to hg19. MACS2 was used to call peaks and peak numbers and FRiP values indicate a wide range of data quality found in recent ATAC-seq datasets. **C**) Small CUTAC fragments improved peak-calling.

If low-salt tagmentation sharpens peaks of DNA accessibility because tethering to neighboring nucleosomes increases the probability of tagmentation in small holes in the chromatin landscape, then we would expect smaller fragments to dominate CUTAC peaks. Indeed this is exactly what we observe for heatmaps (**Fig. S6**), tracks (**Fig. S7**), peak calls and FRiP values (**Fig. 5C**). Excluding larger fragments results in better resolution yielding more peaks and higher FRIP values, both of which approach a maximum with fewer fragments. Moreover, the addition of strongly polar compounds during tagmentation provides a substantial improvement in peak calling and FRiPs (**Fig. 5C**, turquoise and orange curves). Excluding large fragments did not improve ATAC-seq peak calls and FRiP values, which indicates that tethering to H3K4me2 is critical for maximum sensitivity and resolution of DNA accessibility maps.

CUTAC maps transcription-coupled regulatory elements. H3K4me2/3 methylation marks active transcription at promoters (27), which raises the question as to whether sites identified by CUTAC are also sites of RNAPII enrichment genome-wide. To test this possibility, we first aligned CUTAC data at annotated promoters displayed as heatmaps or average plots and found that CUTAC sites are located in the NDR between flanking H3K4me2-marked nucleosomes (**Fig. 6A**). CUTAC sites at promoter NDRs corresponded closely to promoter ATAC-seq sites, consistent with expectation for promoter NDRs.

To determine whether CUTAC sites are also sites of transcription initiation in general, we aligned CUT&Tag RNA Polymerase II (RNAPII) Serine-5 phosphate (RNAPIIS5P) CUT&Tag data over H3K4me2 CUT&Tag and CUTAC and Omni-ATAC peaks ordered by RNAIIS5P peak intensity. When displayed as heatmaps or average plots, CUTAC datasets display a conspicuous shift into the NDR from flanking nucleosomes (**Fig. 6B**).

Mammalian transcription also initiates at many enhancers, as shown by transcriptional run-on sequencing, which identifies sites of RNAPII pausing whether or not a stable RNA product is normally produced (28). Accordingly, we aligned RNAPII-profiling PRO-seq data for K562 cells over H3K4me2 CUT&Tag and CUTAC and Omni-ATAC sites, displayed as heatmaps and ordered by PRO-Seq signal intensity. The CUT&Tag sites showed broad enrichment centered ~1 kb around PRO-seq signals, whereas PRO-seq signals were tightly centered around CUTAC sites, with similar results for Omni-ATAC sites (**Fig. 6B**). Interestingly, alignment around TSSs, RNAPolIIS5P or PRO-seq data resolves immediately flanking H3K4me2-marked nucleosomes in CUT&Tag data, which is not seen for the same data aligned on signal midpoints (**Figs. 3, 5**). Such alignment of +1 and -1 nucleosomes next to fixed NDR boundaries is consistent with nucleosome positioning based on steric exclusion (29). Furthermore, the split in PRO-seq occupancies around NDRs defined by CUTAC and Omni-ATAC implies that the steady-state location of most engaged RNAPII is immediately downstream of the NDR from which it initiated. About 80% of the CUTAC sites showed enrichment of PRO-Seq signal downstream, confirming that the large majority of CUTAC sites correspond to NDRs representing transcription-coupled regulatory elements.

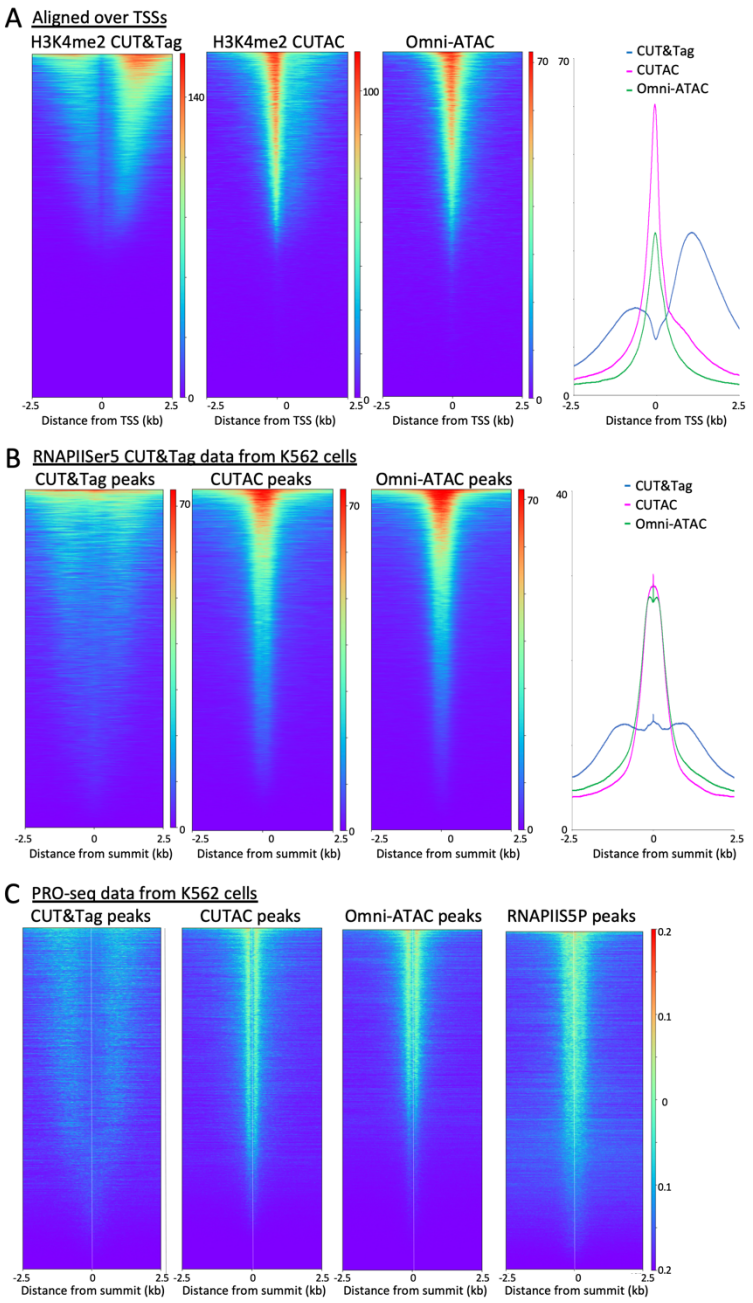

Figure 6. H3K4me2 CUTAC sites are coupled to transcription. **A**) H3K4me2 fragments shift from flanking nucleosomes to the NDR upon low-salt tagmentation, corresponding closely to ATAC-seq sites. **B**) The Serine-5 phosphate-marked initiation form of RNAPII is highly abundant over most H3K4me2 CUT&Tag, CUTAC and ATAC-seq peaks. **C**) Run-on transcription initiates from most sites corresponding to CUTAC and ATAC-seq peaks. Both plus and minus strand PRO-seq datasets downloaded from GEO (GSM3452725) were pooled and aligned over peaks called using 3.2 million fragments sampled from H3K4me2 CUT&Tag, CUTAC and Omni-ATAC datasets, and also from pooled CUT&Tag replicate datasets for K562 RNA Polymerase II Serine-5 phosphate.

Discussion

The correlation between sites of high DNA accessibility and transcriptional regulatory elements, including enhancers and promoters, has driven the development of several distinct methods for genome-wide mapping of DNA accessibility for nearly two decades (30). However, the processes that are responsible for creating gaps in the nucleosome landscape are not completely understood. In part this uncertainty is attributable to variations in nucleosome positioning within a population of mammalian cells such that there is only a ~20% median difference in absolute DNA accessibility between DNaseI hypersensitive sites and non-hypersensitive sites genome-wide (12). This suggests that DNA accessibility is not the primary determinant of gene regulation, and contradicts the popular characterization of accessible DNA sites as "open" and the lack of accessibility as "closed". Moreover, there are multiple dynamic processes that can result in nucleosome depletion, including transcription, nucleosome remodeling, transcription factor binding, and replication, so that the identification of a presumed regulatory element by chromatin accessibility mapping leaves open the question as to how accessibility was established and maintained. Our CUTAC mapping method now provides a physical link between a transcription-coupled process and DNA hyperaccessibility by showing that anchoring of Tn5 to a nucleosome mark laid down by transcriptional events immediately downstream identifies sites that are indistinguishable from the best ATAC-seq sites. The equivalence of CUTAC and ATAC at both enhancers and promoters provides support for the hypothesis that these regulatory elements are characterized by the same regulatory architecture (31, 32).

The mechanistic basis for asserting that H3K4 methylation is a transcription-coupled event is well-established (20). In all eukaryotes, H3K4 methylation is catalyzed by COMPASS/SET1 and related enzyme complexes, which associate with the C-terminal domain (CTD) of the large subunit of RNAPII when Serine-5 of the tandemly repetitive heptad repeat of the CTD is phosphorylated following transcription initiation. The enrichment of dimethylated and trimethylated forms of H3K4 is presumably the result of exposure of the H3 tail to SET1/MLL during RNAPII stalling just downstream of the TSS, so that these modifications are coupled to the onset of transcription. Therefore, our demonstration that Tn5 tethered to H3K4me2/3 histone tail residues efficiently tagments accessible sites, implies that accessibility at regulatory elements is created by events immediately following transcription initiation. This mechanistic interpretation is supported by the mapping of CUTAC sites just upstream of RNAPII, and is consistent with the recent demonstration that PRO-seq data can be used to accurately impute "active" histone modifications (16). Thus CUTAC identifies active promoters and enhancers that produce enhancer RNAs, which might help explain why ~95% of ATAC-seq peaks are detected by CUTAC and vice-versa (**Fig. 5B-C**).

CUTAC also provides practical advantages over other chromatin accessibility mapping methods. Like CUT&Tag-direct, all steps from frozen nuclei to purified sequencing-ready libraries for the data presented here were performed in a day in single PCR tubes on a home workbench. As it requires only a simple modification of one step in the CUT&Tag protocol, CUTAC can be performed in parallel with an H3K4me2 CUT&Tag positive control and other antibodies using multiple aliquots from each population of cells to

be profiled. We have shown that three distinct protocol modifications, dilution, removal and post-wash tagmentation yield high-quality results, providing flexibility that might be important for adapting CUTAC to nuclei from diverse cell types and tissues.

Although a CUT&Tag-direct experiment requires a day to perform, and ATAC-seq can be performed in a few hours, this disadvantage of CUTAC is offset by the better control of data quality with CUTAC as is evident from the large variation in ATAC-seq data quality between laboratories. In contrast, CUT&Tag is highly reproducible using native or lightly cross-linked cells or nuclei (19), and as shown here H3K4me2 CUTAC maps regulatory elements with sensitivity and signal-to-noise comparable to the best ATAC-seq datasets, even better when larger fragments are computationally excluded. Although H3K4me2 CUT&Tag datasets have lower background than CUTAC datasets run in parallel, the combination of the two provides both highest data quality (CUT&Tag) and precise mapping (CUTAC) using the same H3K4me2 antibody. Therefore, we anticipate that current CUT&Tag users and others will find the CUTAC option to be an attractive alternative to other DNA accessibility mapping methods for identifying transcription-coupled regulatory elements.

Materials and Methods

Biological materials. Human K562 cells were purchased from ATCC (CCL-243) and cultured following the supplier's protocol. H1 ES cells were obtained from WiCell (WA01-lot#WB35186) and cultured following NIH 4D Nucleome guidelines (https://data.4dnucleome.org/protocols/50f8300d-400f-4ce1-8163-42f417cbbada/). We used the following antibodies: Guinea Pig anti-Rabbit IgG (Heavy & Light Chain) antibody (Antibodies-Online ABIN101961 or Novus NBP1-72763), Rabbit anti-mouse (Abcam ab46540), H3K4me1 (Epicypher 13-0026, lot 28344001), H3K4me2 (Epicypher 13-0027 and Millipore 07-030, lot 3229364), H3K4me3 (Active Motif, 39159), H3K9me3 (Abcam ab8898, lot GR3302452-1), H3K27me3 (Cell Signaling Technology, 9733, Lot 14), H3K27ac (Millipore, MABE647), H3K36me3 (Epicypher #13-0031, lot 18344001) and NPAT (Thermo Fisher Scientific, PA5-66839). The pAG-Tn5 fusion protein used in many of these experiments was a gift from Epicypher, Inc. (#15-1117 lot #20142001-C1).

CUT&Tag-direct and CUTAC. Log-phase human K562 or H1 embryonic stem cells were harvested and prepared for nuclei in a hypotonic buffer with 0.1% Triton-X100 essentially as described (33). A detailed, step-by-step nuclei preparation protocol can be found at https://www.protocols.io/view/bench-top-cut-amp-tag-bcuhiwt6.

CUT&Tag-direct was performed as described (19), and a detailed step-by-step protocol including the three CUTAC options used in this study can be found at https://www.protocols.io/view/cut-amp-tag-direct-with-cutac-bmbfk2jn. Except as noted, all experiments were performed on a workbench in a home laundry room (**Fig. S1**) Briefly, nuclei were thawed, mixed with activated Concanavalin A beads and magnetized to remove the liquid with a pipettor and resuspended in Wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine and Roche EDTA-free protease inhibitor). After successive incubations with primary antibody (1-2 hr) and secondary antibody (0.5-1 hr) in Wash buffer, the beads were washed and resuspended in pA(G)-Tn5 at 12.5 nM in 300-Wash buffer (Wash buffer containing 300 mM NaCl) for 1 hr. Incubations were performed at room temperature either in bulk or in volumes of 25-50 µL in low-retention PCR tubes. For CUT&Tag, tagmentation was performed for 1 hr in 300-Wash buffer supplemented with 10 mM $MgCl_2$ in a 50 µL volume. For CUTAC, tagmentation was performed in low-salt buffer with varying components, volumes and temperatures as described for each experiment in the figure legends. In "dilution" tagmentation, tubes containing 25 µL of pA(G)-Tn5 incubation solution and 2 mM or 5 mM $MgCl_2$ solutions were preheated to 37°C. Tagmentation solution (475 µL) was rapidly added to the tubes and incubated for times and temperatures as indicated. In "removal" tagmentation, tubes were magnetized, liquid was removed, and 50 µL of ice-cold 10 mM TAPS, 5 mM $MgCl_2$ was added, followed by incubation for times and temperatures as indicated. In "add-back" tagmentation, beads were washed in 500 µL 300-wash buffer as in CUT&Tag, and then 50 µL of ice-cold 10 mM TAPS, 5 mM $MgCl_2$ was added, supplemented with pA(G)-Tn5 and incubated at 37°C for times as indicated.

Following tagmentation, CUT&Tag and CUTAC samples were chilled and magnetized, liquid was removed, and beads were washed in 50 µL 10 mM TAPS pH8.5, 0.2 mM EDTA then resuspended in 5 µL 0.1% SDS, 10 µL TAPS pH8.5. Following incubation at 58°C, SDS was neutralized with 15 µL of 0.67% Triton-X100, and 2 µL of 10 mM indexed P5 and P7 primer solutions were added. Tubes were chilled and 25 µL of NEBNext 2x Master mix was added with mixing. Gap-filling and 12 cycles of PCR were performed using an MJ PTC-200 Thermocycler. Clean-up was performed by addition of 65 µL SPRI bead slurry following manufacturer's instructions, eluted with 20 µL 1 mM Tris-HCl pH 8, 0.1 mM EDTA and 2 µL was used for Agilent 4200 Tapestation analysis. The barcoded libraries were mixed to achieve equimolar representation as desired aiming for a final concentration as recommended by the manufacturer for sequencing on an Illumina HiSeq 2500 2-lane Turbo flow cell.

Data processing and analysis. Paired-end reads were aligned to hg19 using Bowtie2 version 2.3.4.3 with options: -- end-to-end --very-sensitive --no-unal --no-mixed --no-discordant --phred33 -I 10 - X 700. Tracks were made as bedgraph files of normalized counts, which are the fraction of total counts at each basepair scaled by the size of the hg19 genome. Peaks were called using MACS2 version 2.2.6 callpeak -f BEDPE -g hs -p le-5 –keep-dup all –SPMR. Heatmaps were produced using deepTools 3.3.1. A detailed step-by-step Data Processing and Analysis Tutorial can be found at https://www.protocols.io/view/cut-amp-tag-data-processing-and-analysis-tutorial-bjk2kkye.

Author Contributions

Conceptualization S.H.; Methodology S.H., H.S.K., K.A.; Software J.G.H.; Validation S.H., K.A.; Formal analysis S.H., J.G.H.; Investigation S.H., H.S.K., K.A.; Resources S.H.; Data curation J.G.H.; Writing – original draft preparation S.H.; Writing – review & editing J.G.H., H.S.K., K.A.; Funding acquisition S.H., K.A.

Acknowledgments

We thank Terri Bryson, Christine Codomo for sample processing, the Fred Hutch Genomics Shared Resource for DNA sequencing, members of our laboratory for helpful discussions and Paul Talbert for critically reading the manuscript. S. H. is an Investigator of the Howard Hughes Medical Institute. This work was supported by the Howard Hughes Medical Institute (S.H.), grants R01 HG010492 (S.H.) and R01 GM108699 (K.A.) from the National Institutes of Health, and an HCA Seed Network grant from the Chan-Zuckerberg Initiative (S.H.).

References

1. Weintraub H & Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. *Science* 193:848–856.

2. Reeves R (1978) Nucleosome structure of Xenopus oocyte amplified ribosomal genes. *Biochemistry (Mosc)* 17:4908-4916.

3. Jack RS & Eggert H (1990) Restriction enzymes have limited access to DNA sequences in Drosophila chromosomes. *EMBO J* 9:2603-2609.

4. Gottschling DE (1992) Telomere-proximal DNA in Saccharomyces cerevisiae is refractory to methyltransferase activity in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 89:4062–4065.

5. Schwartz YB, Kahn TG, & Pirrotta V (2005) Characteristic low density and shear sensitivity of cross-linked chromatin containing polycomb complexes. *Mol Cell Biol* 25:432-439.

6. Bownes M (1990) Preferential insertion of P elements into genes expressed in the germ-line of Drosophila melanogaster. *Mol Gen Genet* 222:457-460.

7. Crawford GE*, et al.* (2004) Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci U S A* 101:992-997.

8. Dorschner MO*, et al.* (2004) High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods* 1:219-225.

9. Giresi PG, Kim J, McDaniell RM, Iyer VR, & Lieb JD (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* 17:877-885.

10. Auerbach RK*, et al.* (2009) Mapping accessible chromatin regions using Sono-Seq. *Proc Natl Acad Sci U S A* 106:14926-14931.

11. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, & Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10:1213-1218.

12. Chereji RV, Eriksson PR, Ocampo J, Prajapati HK, & Clark DJ (2019) Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Res* 29:1985-1995.

13. Oberbeckmann E*, et al.* (2019) Absolute nucleosome occupancy map for the Saccharomyces cerevisiae genome. *Genome Res* 29:1996-2009.

14. Karabacak Calviello A, Hirsekorn A, Wurmus R, Yusuf D, & Ohler U (2019) Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol* 20:42.

15. Kornberg RD & Lorch Y (2020) Primary Role of the Nucleosome. *Mol Cell* 79:371-375.

16. Wang Z*, et al.* (2020) Accurate imputation of histone modifications using transcription. *biorxiv* https://doi.org/10.1101/2020.04.08.032730.

17. Schmid M, Durussel T, & Laemmli UK (2004) ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* 16:147-157.

18. Kaya-Okur HS*, et al.* (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10:1930.

19. Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, & Henikoff S (2020) Efficient low-cost chromatin profiling with CUT&Tag. *Nature Protocols* http://dx.doi.org/10.1038/s41596-020-0373-x.

20. Henikoff S & Shilatifard A (2011) Histone modification: cause or cog? *Trends Genet* 27:389-396.

21. Corces MR*, et al.* (2016) Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 48:1193-1203.

22. Corces MR*, et al.* (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 14:959-962.

23. Moore JE*, et al.* (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583:699-710.

24. Landt SG*, et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* 22:1813-1831.

25. Zhang J*, et al.* (2020) An integrative ENCODE resource for cancer genomics. *Nat Commun* 11:3696.

26. Swanson E*, et al.* (2020) Integrated single cell analysis of chromatin accessibility and cell surface markers. *biorxiv* https://doi.org/10.1101/2020.09.04.283887.

27. Gilchrist DA*, et al.* (2012) Regulating the regulators: the pervasive effects of Pol II pausing on stimulus-responsive gene networks. *Genes Dev* 26:933-944.

28. Kaikkonen MU*, et al.* (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell* 51:310-325.

29. Chereji RV, Ramachandran S, Bryson TD, & Henikoff S (2018) Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol* 19:19.

30. Klein DC & Hainer SJ (2020) Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Res* 28:69-85.

31. Andersson R, Sandelin A, & Danko CG (2015) A unified architecture of transcriptional regulatory elements. *Trends Genet* 31:426-433.

32. Arnold PR, Wells AD, & Li XC (2019) Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers in cell and developmental biology* 7:377.

33. Skene PJ & Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 6:e21856.

Supplementary Figures

Figure S1. Equipment, supplies, reagents and solutions for CUT&Tag on a home workbench. All experiments starting with nuclei frozen in 10% DMSO in Mr. Frosty containers down to –80°C and held at –20°C were performed on a counter in a home laundry/utility room using stock solutions previously prepared in the lab. Following SPRI bead clean-up and liquid removal from the starting PCR tubes to fresh tubes, samples were brought into the lab for TapeStation analysis and equimolar mixing of barcoded samples followed by an SPRI clean-up of the pool and dilution for submission to the Fred Hutch Genomics Shared Resource for Illumina PE25x25 DNA sequencing. There are no hazardous materials or dangerous equipment used in the at-home protocol, however appropriate lab safety training is recommended.

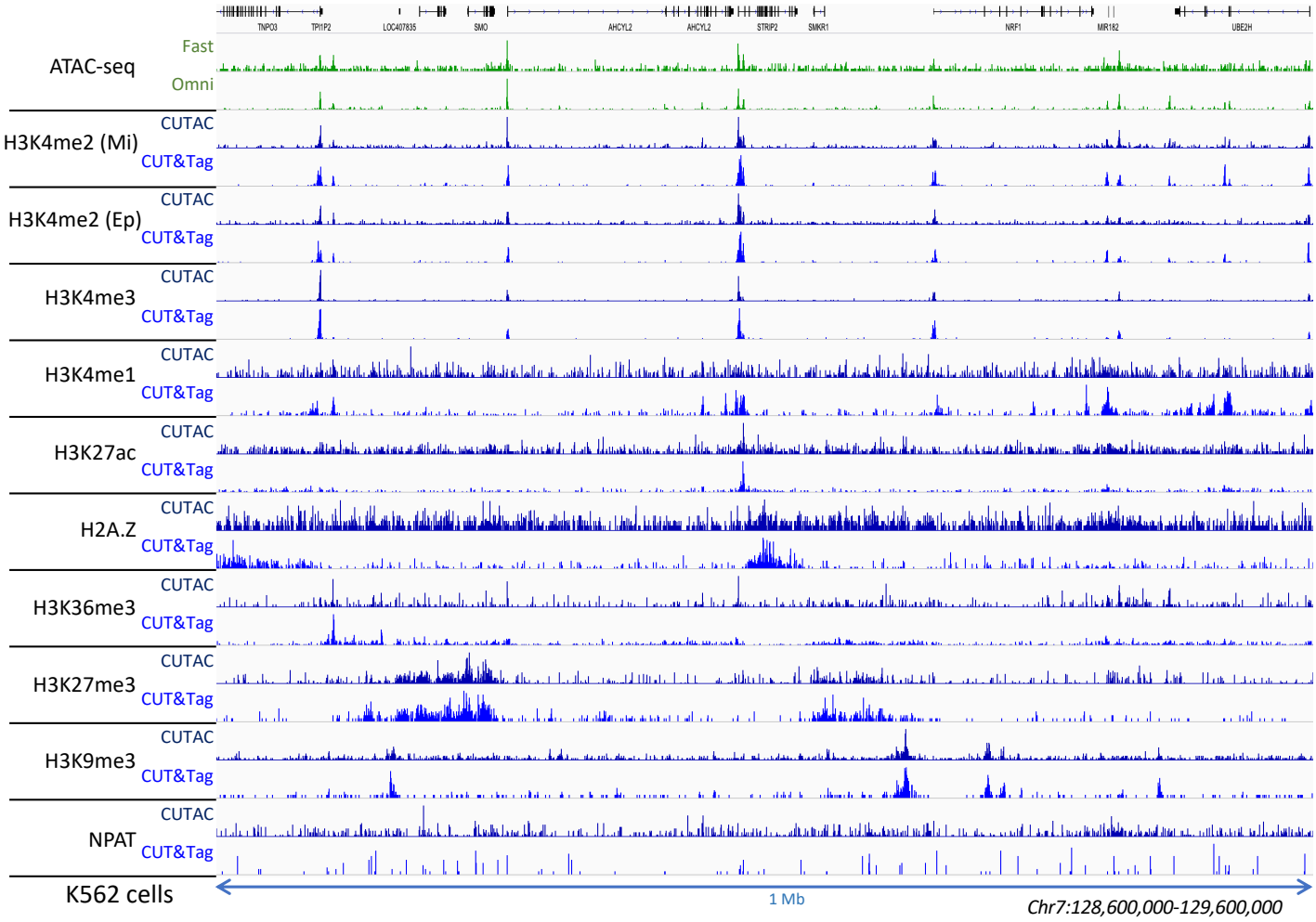

Figure S2: Low-salt tagmentation using various antibodies. Two H3K4me2 antibodies were used: Millipore 07-030 lot 3229364 (Mi) and Epicypher 13-0027 (Ep) and provided similar results. CUTAC was done using the Removal protocol and incubated 10 min 37ºC.

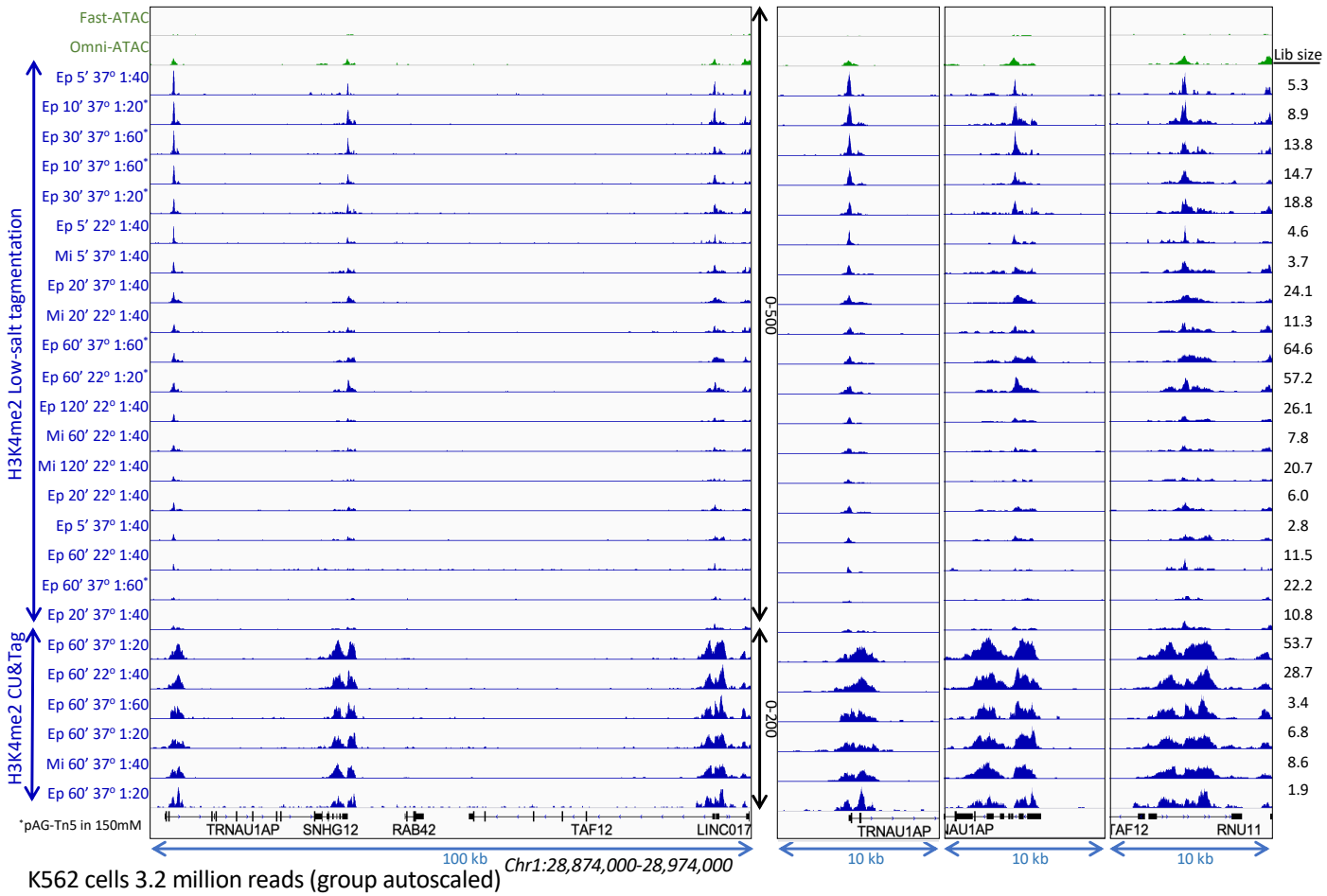

Figure S3: Optimization of low-salt tagmentation conditions: H3K4me2 CUT&Tag and low-salt tagmentation were performed using either a rabbit polyclonal [Millipore 07-030 lot 3229364 (Mi)] or rabbit monoclonal [Epicypher 13-0027 (Ep)] antibody with pAG-Tn5 (Epicypher 15-1117 lot #20142001-C1) at the indicated dilutions. Dilution tagmentation in 2 mM $MgCl_2$ was used at either 22°C or 37°C. Raw paired-end reads were sampled down to 3.2 million and mapped to hg19. A representative 100 kb region is shown (left) and expanded (right) around active promoters and group-autoscaled separately for low-salt tagmentation and standard CUT&Tag using IGV. Estimated library size (Lib size) was calculated by the Mark Duplicates program in Picard tools.

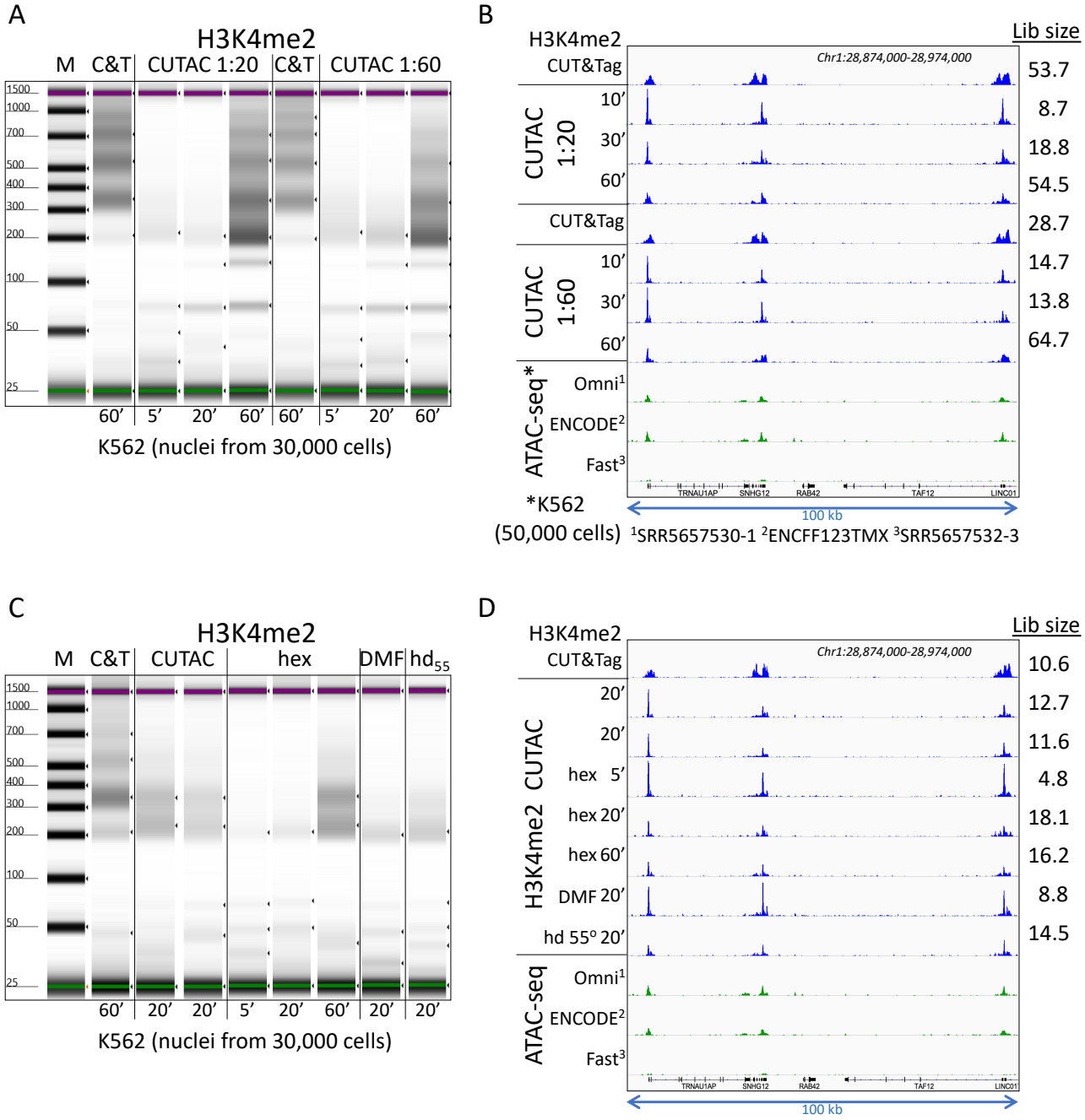

Figure S4: Low-salt tagmentation reduces fragment size, increasing peak resolution: A-B) H3K4me2 CUT&Tag (C&T) and low-salt tagmentations using the Dilution protocol were performed at 37°C using Epicypher 13-0027 antibody and Epicypher 15-1117 pAG-Tn5 for the times indicated using pA/G-Tn5 at either 1:20 (Manufacturer's recommendation) or 1:60 showing a bigger effect of tagmentation time than amount of pA/G-Tn5, and improvement in yield but reduction in signal-to-noise with longer tagmentations. A) Tapestation gel images showing time of tagmentation and yields based on loading 2 µL of each 20 µL sample and integrating over the 175-1000 bp range. M = markers, C&T = CUT&Tag; B) Group-autoscaled tracks showing fragment normalized count densities for sequenced libraries resolved in (A) and for published ATAC-seq data. **C-D**) Same as A-B using the Removal protocol with no additive (CUTAC), 10% 1,6-hexanediol (hex), 10% N,N-dimethylformamide (DMF) or 10% of both at 55°C (hd_{55}) for the times indicated. All datasets were sampled down to 3.2 million and mapped to hg19. A representative 100-kb region was group-autoscaled using IGV. Estimated library size (Lib size, millions of fragments) was calculated by the Mark Duplicates program in Picard tools.

A

Sample	Source	Year	Read_type	Raw_reads	hg19-mapped	% Chr M	# Peaks	FRiP %
CUT&Tag-direct K562	This study	2020	PE25	3,252,490	3,144,253	2	20,555	41
CUT&Tag-direct H1	This study	2020	PE25	2,612,112	2,598,018	0.5	20,308	36
CUTAC K562	This study	2020	PE25	5,973,063	4,785,931	3	14,381	22
Omni-ATAC K562 SRR5657531-2	Stanford	2017	PE75	4,407,706	3,181,110	13	16,737	20
CUTAC H1	This study	2020	PE25	2,770,901	2,734,092	1	14,369	18
ATAC H1 GSM3677783	Fred Hutch	2019	PE25	4,504,812	4,157,800	13	19,517	16
ATAC K562 ENCFF123TMX	Stanford (ENCODE)	2020	PE100	43,473,266	23,942,024	9	14,369	11
ATAC K562 GSM4083680	U. Texas-Southwestern	2019	SE74	29,193,873	17,612,609	43	5,894	8
ATAC K562 GSM3452726	Cornell U.	2018	PE36	86,907,625	83,038,866	24	1,555	6.2
ATAC K562 GSM4190694	Keio U.	2020	PE60	15,363,855	14,067,803	79	1,837	4.8
ATAC H1 GSM4130883	Stanford	2020	PE100	43,784,188	19,562,219	60	4,289	3.2
Fast-ATAC K562 SRR5657533-4	Stanford	2017	PE75	6,702,558	4,677,843	8	2,780	2.4
ATAC K562 GSM4005278	Penn State Hershey	2020	PE100	12,772,997	8,541,005	25	1691	2.1
ATAC K562 GSM4130894	Stanford	2020	PE100	45,122,834	19,021,462	86	449	1.4

B

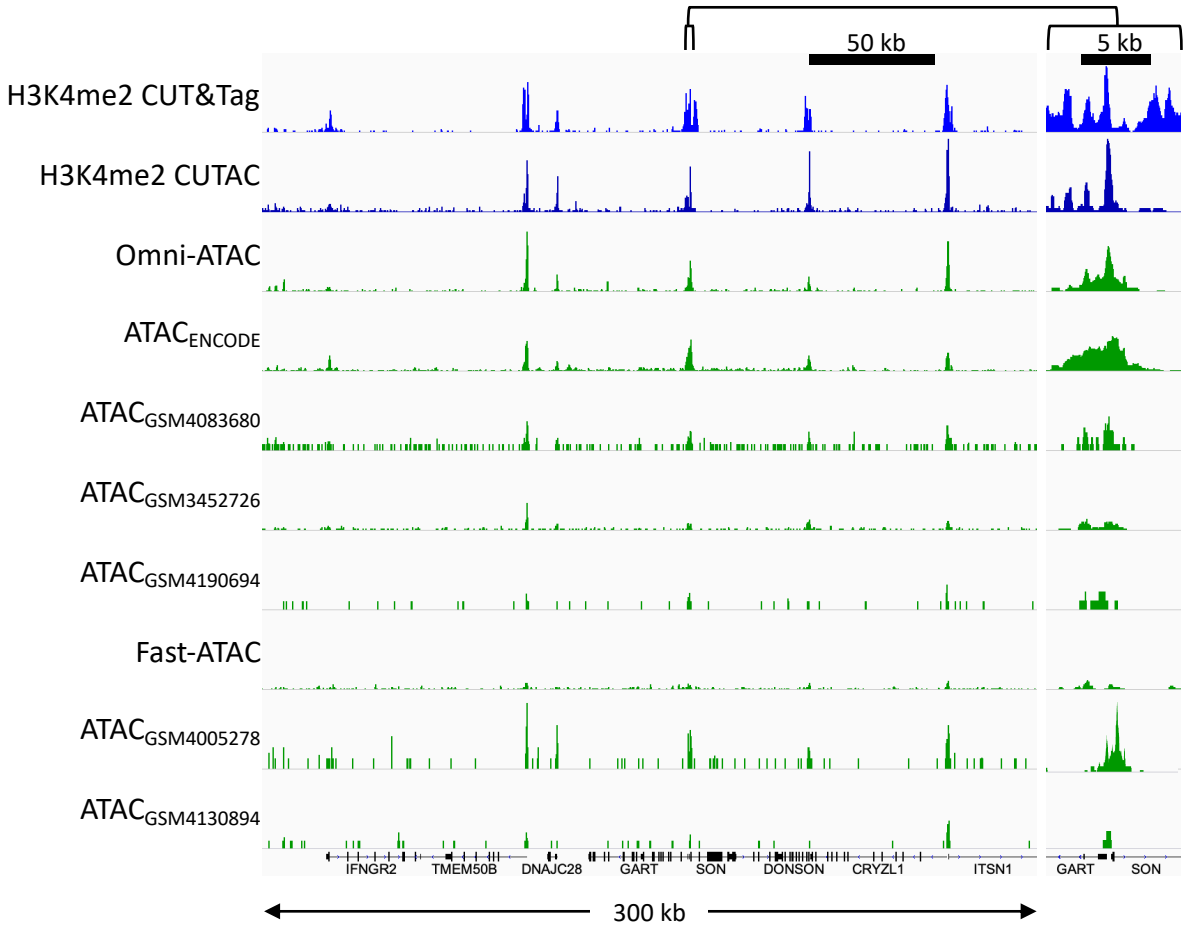

Figure S5: CUTAC data quality is similar to that of the best ATAC-seq datasets. Human K562 and H1 ES cell ATAC-seq datasets were downloaded from GEO, and Bowtie2 was used to map fragments to hg19. A sample of 3.2 million mapped fragments without Chr M was used for peak-calling by MACS2 to calculate FRiP values. Year of submission to GEO or SRA databanks is shown. % Chr M is percent of fragments mapped to Chr M (mitochondrial DNA). (**B**) Tracks over a representative region for K562 datasets listed in (A). Samples are ordered by decreasing FRiP.

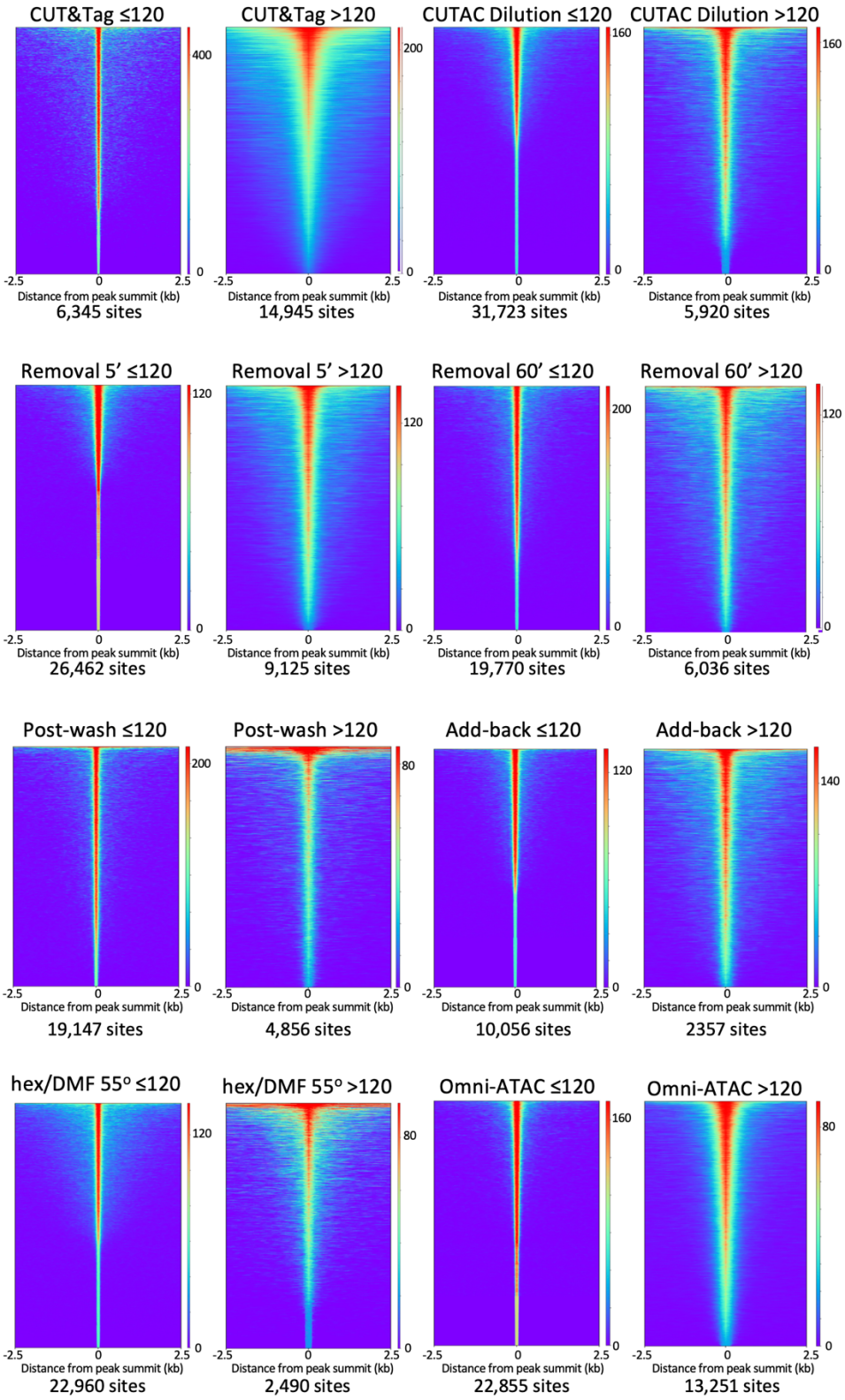

Figure S6: Smaller fragments (≤120 bp) dominate NDRs. See Fig. 2E. Additional comparisons of small (≤120 bp) and large (>120) fragments from diverse CUTAC datasets used in this study show consistent narrowing for small fragments around their summits. For each dataset a 3.2 million fragment random sample was split into small and large fragment groups, MAC2 was used to call peaks and heatmaps were ordered by density over the peak midpoints (sites).

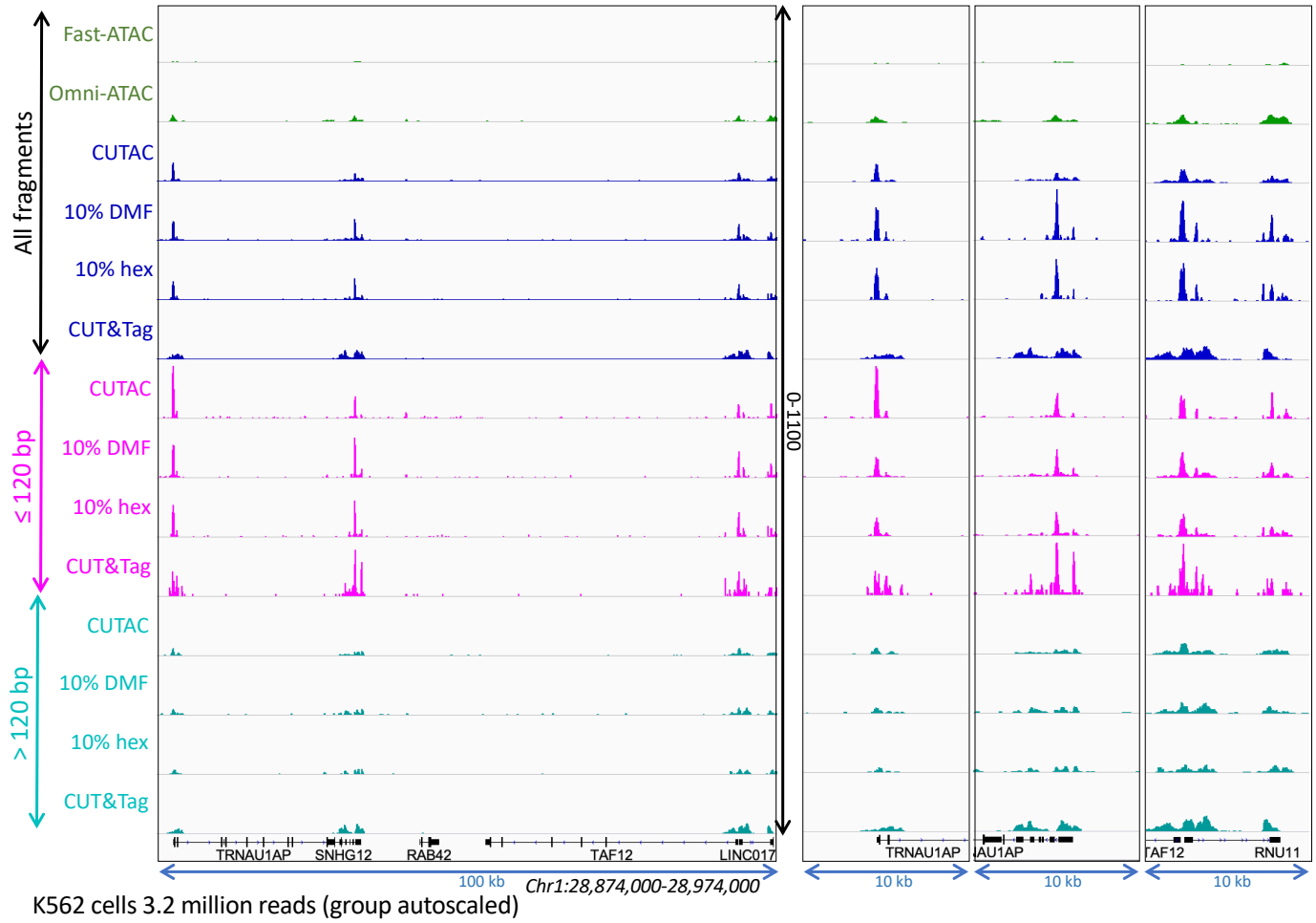

Figure S7: Small CUTAC fragments improve peak resolution. The representative region shown in Figs. S3-S4 is shown for all 3.2 million fragments and for ≤120 bp and >120 bp groups.