Automated chromatin profiling with spa-ChIP seq uncovers the impacts of condition variations

Yuwei Cao[1,2], Lauren Patel[1,3], Lauren Alcoser[4], Eric Mendenhall[5], Christopher Benner[6], Sven Heinz[6], Alon Goren[1,#]

[1]Department of Medicine, Division of Genomics & Precision Medicine, University of California San Diego, La Jolla, CA

[2]Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA

[3]Department of Bioengineering, Jacobs School of Engineering, University of California San Diego, La Jolla, CA

[4]Agilent Technologies, 5301 Stevens Creek Blvd, Santa Clara, CA 95051

[5]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA

[6]Department of Medicine, Division of Endocrinology & Metabolism, University of California San Diego, La Jolla, CA

[#]Correspondence should be addressed to agoren@ucsd.edu

ABSTRACT

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a well-established method for studying the genomic localization of DNA-associated proteins. Yet, while useful, most ChIP-seq protocols include multiple manual steps that can introduce inconsistency and make it burdensome to analyze large sample sets, limiting the inclusion of appropriate replicates and other controls. Although some of these challenges were addressed by incorporation of liquid handling platforms, most of those previous efforts have automated only a subset of the ChIP-seq steps. Further, using automation for efficiently mapping non-histone proteins, such as chromatin regulators, has been challenging. We recently developed a single-pot ChIP-seq protocol[1]. Here we established a liquid handler operation for this protocol and created an end-to-end fully automated version that is scalable from 8 to 96 ChIP-seq reactions. Our single-pot automated (spa-ChIP-seq) protocol requires about 3 days from cross-linked cells to library, and the costs to produce indexed libraries is approximately $70 per sample. We first benchmarked spa-ChIP-seq against manual ChIP-seq that was performed in parallel, showing a nearly indistinguishable signal-to-noise ratio between the two workflows. Next, we used spa-ChIP-seq to systematically evaluate multiple parameters including shearing and crosslinking conditions, buffer compositions, and antibody to cell-number ratios. Our method allowed us to identify optimal conditions for double crosslinked chromatin as well as easily and robustly conduct titrations and screening of different antibodies and reagents, eliminating many laborious and costly steps. We show, for the first time to our knowledge, that the effect of the ratio of antibody to cell-number is most pronounced in detecting weak genomic localization signals. In particular, while strong signal seems to be unimpacted by low antibody to cell-number ratio, weaker signals are sensitive to this ratio. We note the importance of maintaining a consistent antibody to cell-number ratio, especially when conducting comparative studies, e.g., between treatments such as small molecules, or individuals such as for chromatin-QTL mapping. Our automated ChIP-seq protocol is publicly available, including specific deck setups, software files and parameters. Lastly, we envision that our robust, cost-efficient protocol can advance research via multiple fronts, e.g., by (i) allowing the scaling up the number of replicates and conditions tested, (ii) improving quantification precision when using spike-in normalization in ChIP-seq experiments[2], (iii) enabling core facilities to provide high-throughput ChIP-seq as a service and (iv) being incorporated into antibody evaluation procedures, compound screening, population genomics and diagnostic frameworks.

INTRODUCTION

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a well-established method for studying the genomic localization of DNA-associated proteins. ChIP-seq is widely used to study histone modifications, chromatin regulators (CRs), transcription factors (TFs), and other DNA-associated proteins[3–6]. The commonly used ChIP-seq method has led to significant insights about principles of gene regulation and function of the epigenome in both normal and disease conditions[7,8]. While ChIP-seq is useful, the overall experimental workflow involves multiple steps that increase the risk of introducing inconsistency within experiments and between groups, making analysis across ChIP-seq datasets challenging. These challenges were partially addressed by the incorporation of automated liquid handling platforms to improve the robustness of the process[9–13]. Yet, most of the previous efforts[9,10,12] have automated only a subset of the ChIP-seq steps (**Table 1**). Additionally, the majority of current automated ChIP-seq protocols are limited in their ability to efficiently map non-histone proteins, such as chromatin regulators that may interact only indirectly with DNA (e.g., by being associated with modified histones or histone variants).

Commonly, ChIP-seq crosslinking involves a single step of formaldehyde (FA) to create covalent bonds between proteins and DNA. However, FA is a small molecule and only captures interactions in the range of 2.3–2.7 Å, which limits the distance of protein interactions captured[14]. Thus, to enable detection of protein interactions occurring between molecules that are further apart, larger FA-compatible crosslinkers, such as disuccinimidyl glutarate (DSG) that can form crosslinks of approximately 7.7 Å, can be combined with FA. Double crosslinking was shown to be a highly effective and reproducible method for capturing the binding of protein complexes or transcription factors (e.g.,NF-κB and STAT3) and identifying genomic targets for protein-DNA binding with high signal-to-noise ratios[15,16].

Moreover, effective chromatin solubilization is also critical in the ChIP-seq workflow as it directly influences the immunoprecipitation performance, sonication efficiency, and experiment reproducibility. A wide range of lysis buffers using different detergents have been developed to facilitate nuclear disruption and the release of crosslinked chromatin. These lysis buffers help to generate soluble chromatin preparation, balancing effective lysis and nuclear disruption with preservation of chromatin-protein interactions.

Another important aspect of ChIP-seq workflow is the ratio of antibody to cell-number, as this ratio can influence enrichment efficiency during immunoprecipitation. Usually, the vendor experimentally validates and titrates the antibody to yield an optimal titer for each antibody. Yet, different vendors may have different validation methods, and the available recommendations may not be suitable for all applications[17]. Using fixed amounts of antibody when cell numbers differ across samples or conditions can lead to suboptimal enrichment, increased background, and variability in ChIP-seq quality[18].

We recently developed and published a detailed single-pot ChIP-seq protocol that incorporates an on-bead library preparation step[1]. Here we adapted this single-pot ChIP-seq protocol to a liquid handler operation and created an end-to-end fully automated version that uses a 96-well sonicator. We benchmarked the automated

protocol against manual ChIP-seq to ensure high quality data is produced. This automated ChIP-seq protocol allows us to test dozens of samples in a time and cost-effective manner that would not be feasible with manual ChIP-seq. This allowed us to evaluate multiple parameters, including double crosslinking, shearing and crosslinking conditions, lysis buffer compositions, and antibody to cell-number ratios. We demonstrate the utility of our automated method by identifying optimal ChIP-seq conditions for both histone and non-histone proteins. We have streamlined the automation process and made the protocols generalizable, tailored the deck settings and packaged the individual protocols into a user-friendly interface. Our automated ChIP-seq protocol is publicly available, including the specific deck setups, software files and parameters (**Supplemental Files**). We envision that our robust, cost-efficient protocol can advance research efforts via multiple fronts. These include: (***i***) allowing the scaling up the number of replicates and conditions tested in each experiment, (***ii***) improving quantification precision when using spike-in normalization in ChIP-seq experiments[2], (***iii***) enabling core facilities to provide high-throughput ChIP-seq as a service, and (***iv***) being incorporated into antibody evaluation procedures[19–21], compound screening[22], population genomics[23] and diagnostic frameworks[24].

Method	Automated Steps	Sonication	Crosslinking	Library Prep	Protocol Availability
HT-ChIP (Garber *et al.*[9], Blecher-Gonen *et al.*[10])	Uses Agilent Bravo for library construction and DNA cleanup steps; the rest of the protocol is not automated	Branson Sonifier (probe sonicator)	Single crosslinking with 1% FA	Requires purification beforehand	Automated protocols are available as a written text in the Supplemental Experimental Procedures.
AHT-ChIP-seq (Aldridge *et al.*[11])	Uses Agilent Bravo NGS for ChIP process, Beckman Fx^p dual arm instrument for pre-PCR library construction, and Caliper Zephyr liquid handler for PCR cleanup	Misonix Sonicator 3000 (probe sonicator)	Single crosslinking with 1% FA	Requires purification beforehand	Agilent Vworks automated ChIP protocol files and Beckman library prep are available in Additional files.
R-ChIP (Gasper *et al.*[12])	Uses Tecan Freedom EVO 200 for ChIP process; library construction and DNA cleanup are not automated	Probe sonicator	Single crosslinking with 1% FA	Requires purification beforehand	Automated protocols are available as text in Methods and Supplementary Information
FloChIP (Dainese *et al.*[25])	Uses an “in house” custom built microfluidic device (non-commercial) for ChIP process; library construction and DNA cleanup are not automated	Covaris E220	Single crosslinking with 1% PFA	Requires purification beforehand	Automated protocols are available as text in Supplementary Information
spa-ChIP-seq (this study)	Uses Agilent Bravo NGS for entire ChIP-seq process including sequencing-ready library construction	Active Motif PIXUL	Double crosslinking with 2mM DSG and 1% FA	On-bead library preparation	User-friendly spa-ChIP-seq Agilent Vworks form and protocol files are available in Supplemental files.

Table 1. Comparison of spa-ChIP-seq protocol to other automated ChIP-seq approaches. For each liquid handler based ChIP-seq method, the table shows the automated steps, methods for sonication, crosslinking, and library prep, as well as protocol availability, noting the potential shortcomings that limits the high-throughput ability of these approaches: (1) Most of the methods shown only automated a subset of the entire ChIP-seq process, but spa-ChIP-seq is fully automated (from cross-linked cells to sequencing ready libraries). (2) For sonication, a probe sonicator shears one tube at a time and thus limits throughput. The Covaris E220 machine is costly and time consuming. Our method uses the PIXUL device[26], which is both low cost ($19/plate vs $640/plate for Covaris E220) and has a short turnaround (~60-108 min /96 samples vs ~9.5-10 hours (~600min) for Covaris E220). (3) The methods shown in the table are optimized for single crosslinking and does not include an option for double crosslinking. (4) All of these methods do not include on-bead library preparation and require reverse-crosslinking and DNA purification after immunoprecipitation prior to sequencing library construction. This leads to an increase in process time and cell loss.

Note, we do not discuss here there several variations of ChIP-seq (e.g., AutoRELACS) or other approaches for mapping DNA-associated proteins (e.g., AutoCUT&RUN, AutoCUT&Tag) due to key the differences between the experimental protocols.

RESULTS

Establishment and evaluation of a single-pot automated (spa)-ChIP-seq protocol

We built on our previous automated ChIP-seq setup with a liquid handling robot[13] and modified the parameters to accommodate our recent single-pot ChIP-seq protocol[1]. This protocol takes roughly 3 days (starting from fixed cells) to produce Illumina ready indexed libraries. Our automated process is scalable to between 8 to 96 ChIP-seq reactions (ideally in increments of 8) and at a cost of approximately $70 per sample (**Supplemental Table 1**). We first identified multiple experimental variables that could be optimized and tested (**Fig. 1**). We then assessed the performance of our automated ChIP-seq protocol by benchmarking it against the manual ChIP-seq using the exact same input material. To conduct this initial benchmarking, we focused on both H3K27ac, a histone modification that is associated with open chromatin, and H3K27me3, a histone mark that is linked to repressive heterochromatin. The ChIP-seq signal-to-noise ratio for H3K27ac and H3K27me3 were highly concordant between the manual and automated ChIP-seq workflows, validating the liquid handler settings we optimized and the reproducibility of our automated protocol (**Fig. 2a**).

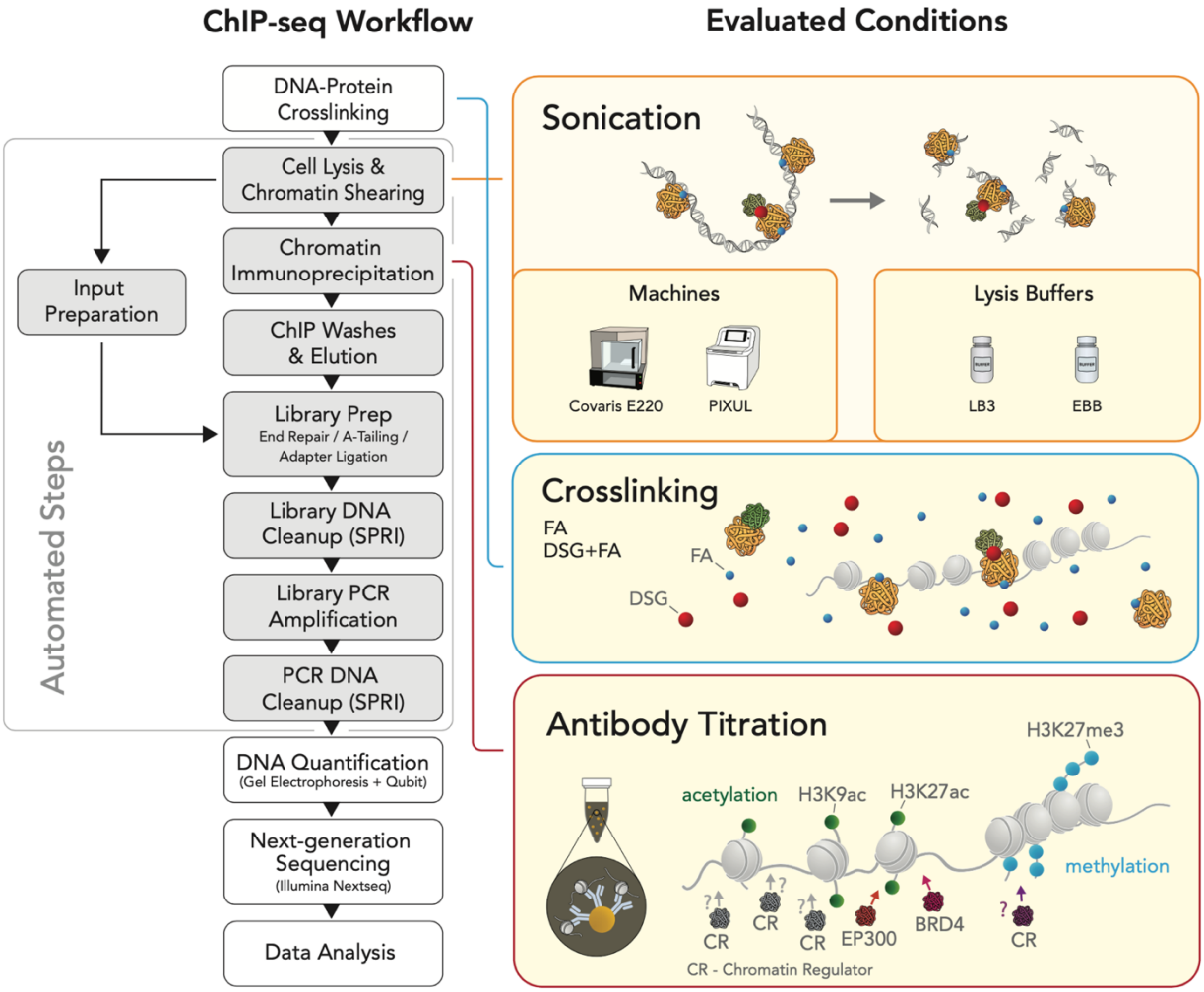


Fig. 1: Overview of the spa-ChIP-seq workflow and evaluated ChIP-seq experimental conditions. The automated steps of ChIP-seq workflow are highlighted in the gray boxes of the flowchart on the left and the evaluated experimental conditions in the yellow boxes. The crosslinking reagent formaldehyde (FA) is represented as a small blue dot, and disuccinimidyl glutarate (DSG) is represented as a large red dot.

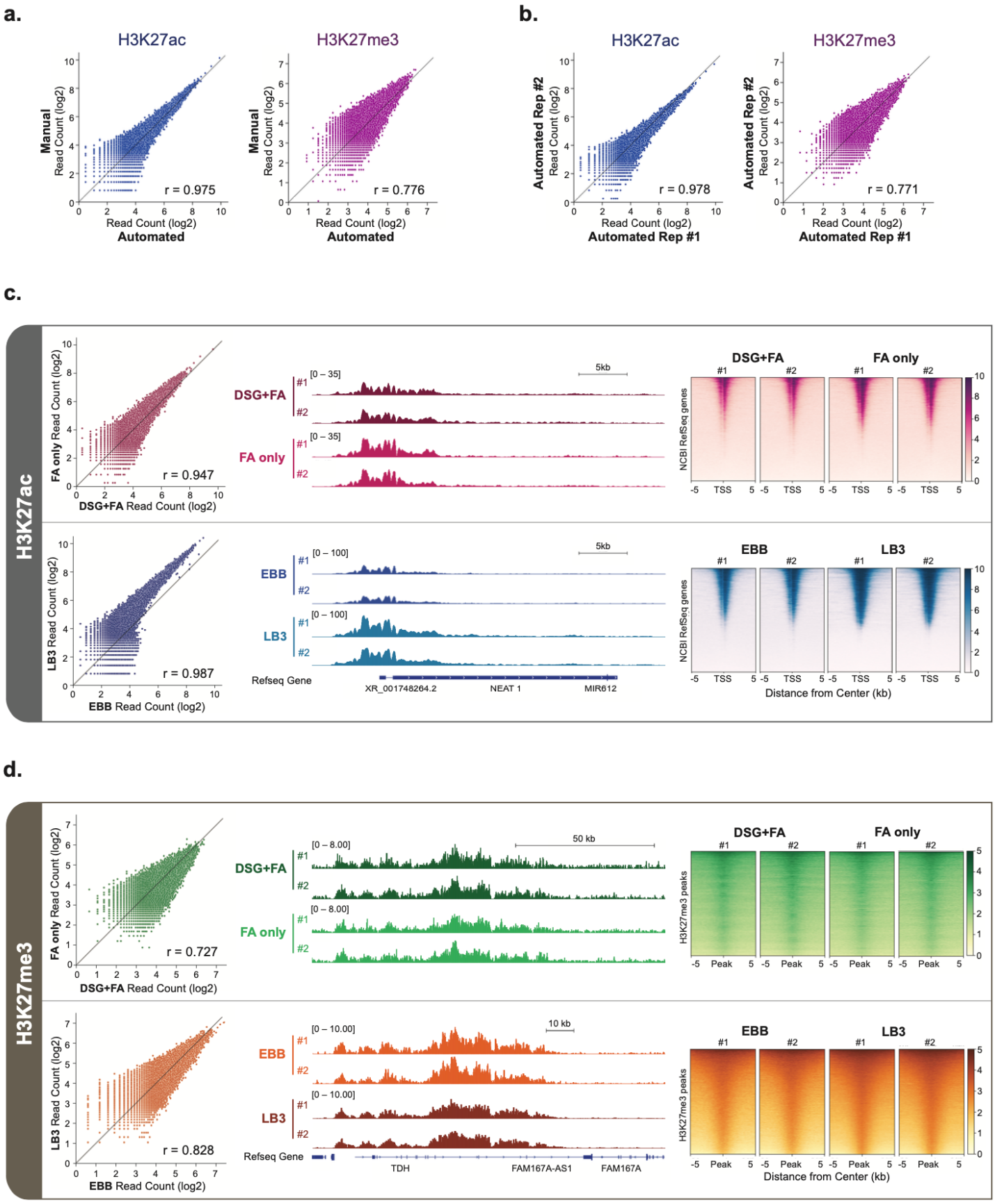


Fig. 2: Systematic evaluation of spa-ChIP-seq with comparison to manual workflow, cross- linking conditions and buffer compositions. a-b. Scatterplots comparing normalized read counts (log2) between (a) automated and manual ChIP-seq samples and (b) automated replicates at H3K27ac or H3K27me3 shared peaks. **c-d.** Scatterplots of read counts (log2) on shared peaks, IGV[27] genome

browser tracks and heatmaps of ±5 kb from the center of transcription start sites (TSSs) or peaks showing the automated ChIP-seq results of (c) H3K27ac and (d) H3K27me3 for evaluating crosslinking conditions, DSG+FA vs FA only (top), and buffer compositions, EBB vs LB3 (bottom). The Pearson r correlation coefficients are reported on the lower right corner of the scatterplots.

Identification of optimal shearing conditions, buffer composition and cross-linking using spa-ChIP-seq

Most of the initial automation efforts[9–11] has focused on using a probe sonicator such as Branson, which is limited to shearing one sample at a time. Here, we aimed to use a 96-well format sonicator. In previous work[13], we used the Covaris E220 device that allows shearing of up to 96 samples. However, as the Covaris E220 sonicates only one sample at a time, shearing of 96 samples requires approximately 10 hours. This limitation both leads to the samples being kept for a long time within the Covaris E220 before the addition of antibodies, as well as extends the time it takes to perform the protocol. A recently introduced sonication device, PIXUL[26] enables the shearing of 96 samples in only 1-2 hours, and also reduces the costs of the plate used for sonication by 30-fold (**Supplemental Table 1**). We decided to evaluate the ability of PIXUL to reproducibly shear samples in a manner that is at least comparable to the Covaris E220. For both devices we additionally tested the impact of placing samples at different locations within the 96-well plate (**Supplemental Fig. 1a**). While the samples sheared by the different devices had distinct size distributions (~250bp for Covaris E220 and ~350bp for PIXUL), both were within the range of 200-600 base pairs that is normally used for ChIP-seq (**Supplemental Fig. 1a**). As the PIXUL machine was significantly faster and about 30-fold lower in costs per plate, we proceeded to use this device for shearing in our tests of subsequent variables using our spa-ChIP-seq.

Next, we compared the ChIP-seq results of the histone modifications H3K27ac and H3K27me3 under single and double crosslinking conditions as well as two different lysis buffer compositions, EBB and LB3 (**Methods**). We used PIXUL to shear approximately 1 million crosslinked HeLa-S3 cells per well; to reduce variation, following the shearing step we merged the sheared chromatin from all wells. Next, we split the pool of sheared chromatin between the two conditions (manual or automated ChIP-seq). In total, for this initial evaluation, we performed 32 ChIP-seq experiments, comparing 16 different conditions, each in duplicates.

We observed that the samples treated with the additional crosslinker required longer sonication time in LB3, the lysis buffer we routinely use (**Supplemental Fig. 1b**). We evaluated the ability of EBB buffer to assist the shearing process. Sonication of the double crosslinked samples in EBB for 60 minutes provided optimal fragmentation patterns. This fits with prior work[28]. We overcame the inefficient fragmentation of double crosslinked samples in LB3 buffer (**Supplemental Fig. 1b**) by increasing the sonication time to 90-120 minutes. The longer shearing time provided optimal shearing patterns. Thus, both buffers were able to generate fragment sizes of 200-600bp and were subsequently used to determine their performance in ChIP-seq.

Next, we used the double crosslinked samples and performed ChIP-seq of H3K27ac and H3K27me3, either lysed with EBB or LB3. Overall, we observed that the samples sheared in EBB had lower signal-to-noise ratio (SNR) compared to the ones sonicated in LB3, potentially due to the high SDS levels in the EBB buffer (1% SDS; **Methods**; **Fig. 2b-c**). Given the lower SNR we obtained when using EBB, we conducted the rest of the experiments only with LB3 buffer and sonicated the double crosslinked samples for 108-120 minutes. Using

these conditions, the H3K27me3 and H3K27ac ChIP-seq results from the double crosslinked samples were mostly indistinguishable from the single crosslinked ones, providing evidence that the additional crosslinker or longer shearing time did not introduce bias in either active or repressive chromatin environments (**Fig. 2b-c**). Since double crosslinking with DSG and FA extends the range of captured interactions in ChIP-seq of DNA-associated proteins and does not negatively impact histone modification ChIP-seq results, using double-crosslinked cells for both target types offer a feasible and consistent protocol for experiments involving chromatin regulators, transcription factors and histone marks.

Identification of optimal conditions for genomic mapping of non-histone DNA-associated proteins

Next, we used the ChIP-seq conditions identified above to map non-histone proteins in double-crosslinked (DSG+FA) HeLa-S3 cells. As a proof of concept, we focused on 5 DNA-associated proteins (CTCF, HAT1, HDAC3, RPB1 (POLR2A), and YAP1) to perform ChIP-seq, each in two technical replicates (**Fig. 3**). Following our previous efforts to increase ChIP-seq standardization[13], we aimed to mainly focus on monoclonal antibodies and used our automated approach to evaluate between 2-4 different antibodies for each target protein. Technical replicates were highly reproducible SNR across six of the seven samples and the variability observed in the remaining sample was within the expected range of experimental variability (**Fig. 3**). Together, we were able to map all of the DNA-associated proteins we tested, although for a subset of the antibodies we obtained only a weak signal under the set of conditions we evaluated (**Supplemental Table 2**).

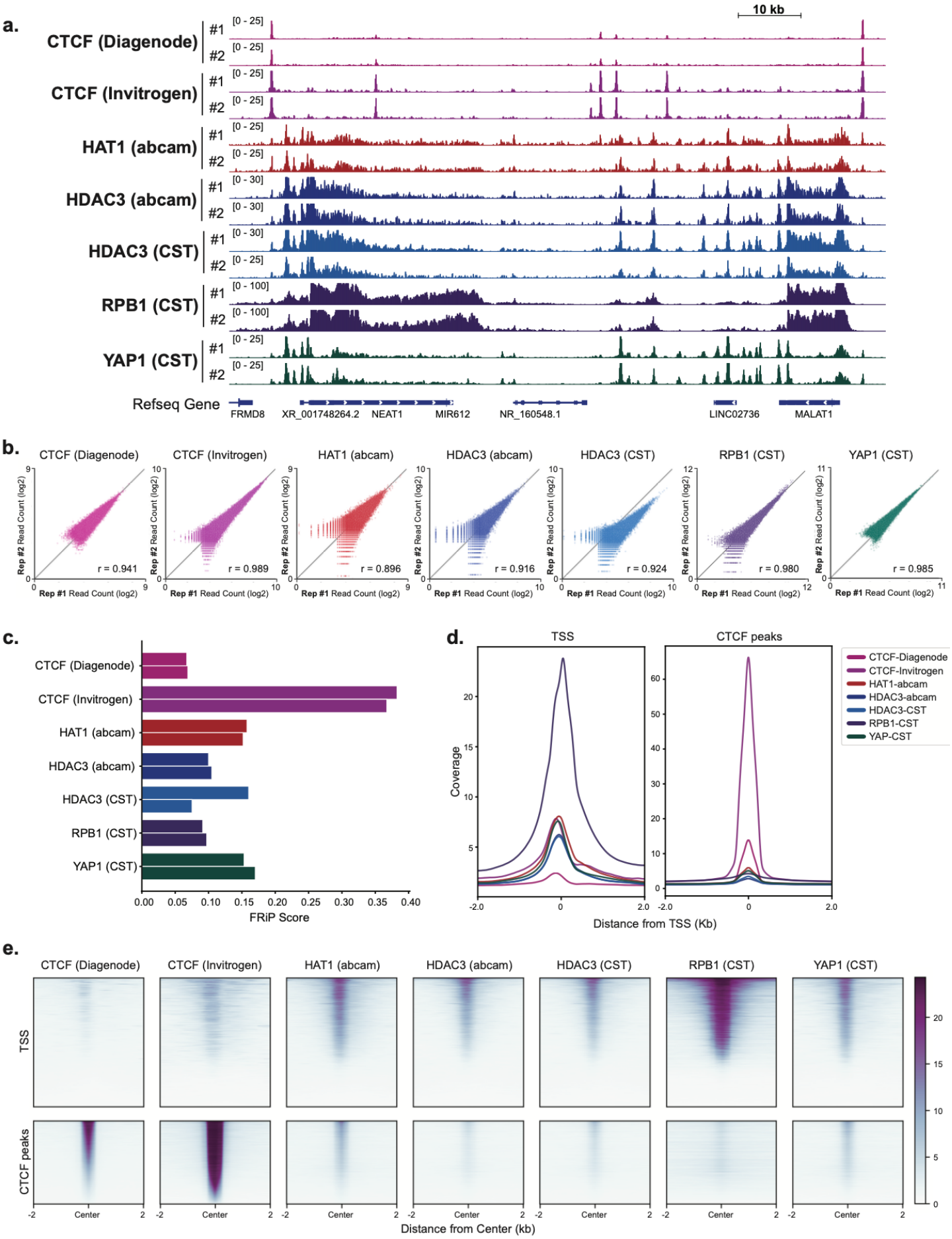

a.
10 kb
CTCF (Diagenode)
CTCF (Invitrogen)
HAT1 (abcam)
HDAC3 (abcam)
HDAC3 (CST)
RPB1 (CST)
YAP1 (CST)
#1
#2
Refseq Gene
FRMD8
XR_001748264.2
NEAT1
MIR612
NR_160548.1
LINC02736
MALAT1
b.
r = 0.941
r = 0.989
r = 0.896
r = 0.916
r = 0.924
r = 0.980
r = 0.985
Rep #1 Read Count (log2)
Rep #2 Read Count (log2)
c.
FRiP Score
d.
TSS
CTCF peaks
Coverage
Distance from TSS (Kb)
CTCF-Diagenode
CTCF-Invitrogen
HAT1-abcam
HDAC3-abcam
HDAC3-CST
RPB1-CST
YAP-CST
e.
Center
Distance from Center (kb)

Fig. 3: Mapping of DNA-associated proteins using spa-ChIP-seq. a. IGV genome browser tracks showing the ChIP-seq of DNA-associated proteins: CTCF, HAT1, HDAC3, Rpb1, and YAP1. CTCF and HDAC3 ChIP-seq experiments were performed with two antibodies from two different vendors. Each antibody condition was done in replicates. **b.** Scatterplots showing the comparison of the 2 technical replicates at shared peaks for each antibody condition. **c.** Bar plot showing Fraction of Reads in Peaks (FRiP) for the ChIP-seq of DNA-associated proteins. For each antibody, both replicates are plotted. **d.** Histograms showing the average normalized read coverage of the replicates at ± 2 kb from the center of TSSs. **e.** Heatmaps showing the average normalized read density of the replicates at ±2 kb from the center of TSSs (top row) and CTCF peaks (bottom row).

The antibody to cell-number ratio may impact ChIP-seq results

Another important, yet often overlooked, aspect of ChIP-seq is the ratio between the amount of antibody used for immunoprecipitation and the number of cells (or amount of chromatin) in the reaction. Our automated ChIP-seq protocol significantly simplifies the task of performing multiple antibody titrations which could be tedious otherwise. We first titrated H3K9ac antibody in HeLa-S3 cells. Initial evaluation of the genome browser tracks on IGV and FRiP scores show that different amounts of antibodies had minimal impact on the global ChIP-seq signal, and the overall trend is that higher amounts of antibodies may lead to only slightly higher ChIP enrichment (**Supplemental Fig. 4**). However, when we equally divided the ChIP-seq peaks into 4 quartiles based on the normalized read counts at peaks, we found that the amount of antibody had a greater impact on ChIP-seq peaks with weaker enrichment, whereas the enrichment signal is constant for the strongest peaks in the experiment (**Fig. 3a**).

We hypothesized that the weaker signals are more sensitive to the amount of antibody, and to investigate this further, we incorporated a cell line with a knock-in that allows an inducible degradation of the BRD4 protein (HEK293T-BRD4-FKBP12^{F36V}-2xHA[29]). Upon treatment of dTAG-13, the amount of BRD4 epitopes is greatly reduced[29]. We performed titration experiments for antibodies targeting HA-tag and BRD4 in both dTAG-13 and DMSO treated cells. Regardless of the amounts of antibody used, the ChIP-seq signal is reduced in dTAG-13 treated cells for both the HA-tag and BRD4 antibodies (**Fig. 3b**). When the DMSO peaks for each antibody were separated by quartiles based by their DMSO enrichment, the dTAG-13 treated samples showed an increasing higher signal as the ratio of antibody to cell-number increased in all 4 quartiles regardless of the trend in the DMSO samples (**Fig. 3c**).

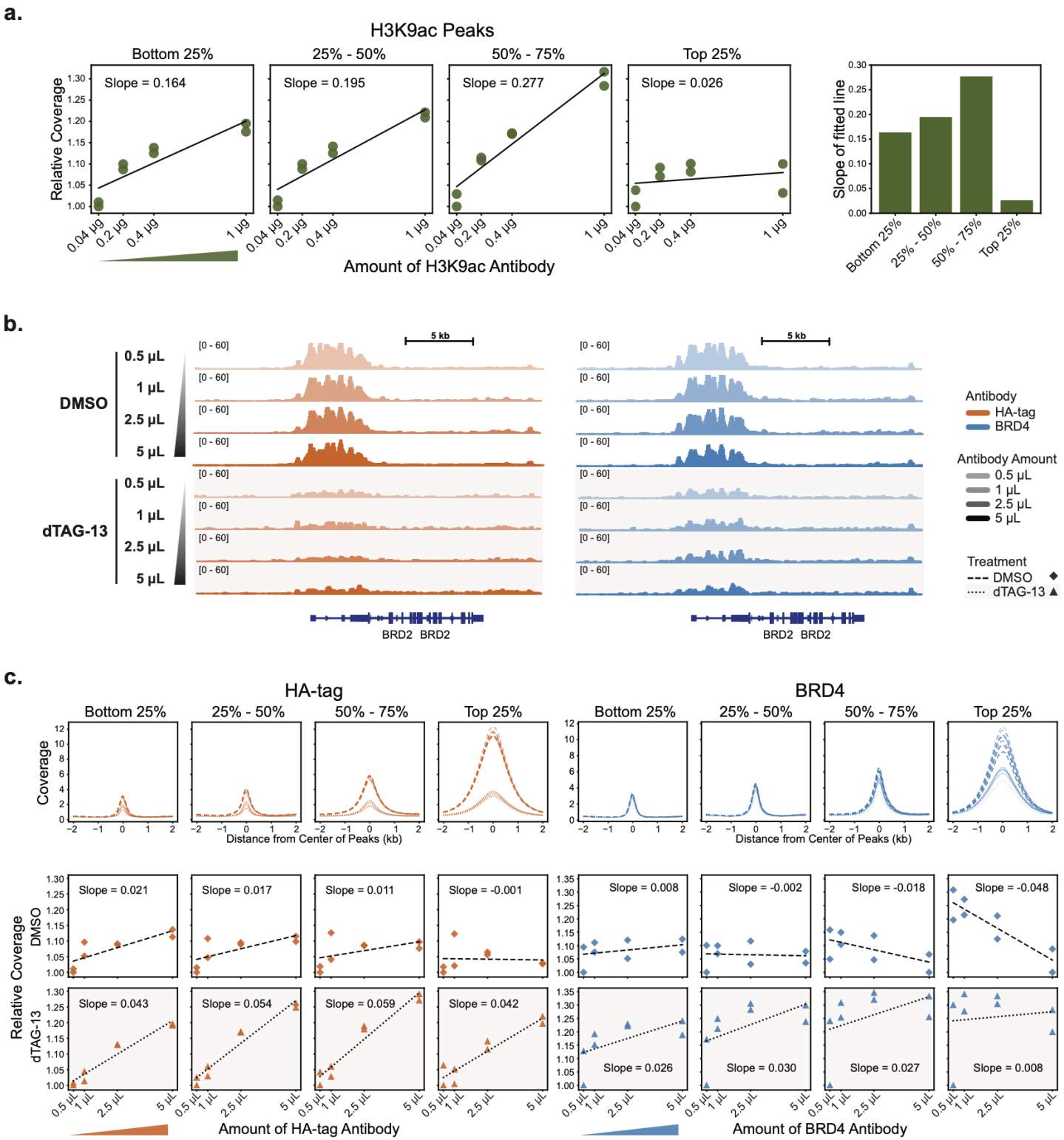


Fig. 4: Antibody titration with spa-ChIP-seq reveals the weaker signals are more sensitive to the amount of antibody. a. Relative total coverage of H3K9ac signal with 0.04 µg, 0.2 µg, 0.4 µg, and 1 µg of antibody for H3K9ac peaks separated into quartiles based on normalized read counts at peaks (the values are relative to the minimum, which is set to 1). Each condition has 2 technical replicates. Each plot is fitted with a linear regression line, and the corresponding slope of fitted line is shown in the bar plot on the left. **b.** IGV genome browser tracks for HA-tag and BRD4 ChIP-seq in dTAG-13 and DMSO treated HEK293T-BRD4-FKBP12^{F36V}-2xHA cells with 0.5 µL, 1 µL, 2.5 µL, and 5 µL of HA-tag and BRD4 antibodies. The tracks of 2 replicates for each condition are overlaid on one another. **c.** Histograms of normalized read coverage for HA-tag and BRD4 antibody titration ChIP-seq signals centered on peaks obtained by using the DMSO samples. The peaks are separated into quartiles according to normalized read counts at peaks. Relative total coverage of HA-

tag and BRD4 signals with 0.5 µL, 1 µL, 2.5 µL, and 5 µL of antibody for each quartile of DMSO peaks shown in the histogram above. The orange and blue triangles represent the increase in the ratio of antibody to cell-number in HA-tag and BRD4, respectively. Each condition has 2 technical replicates.

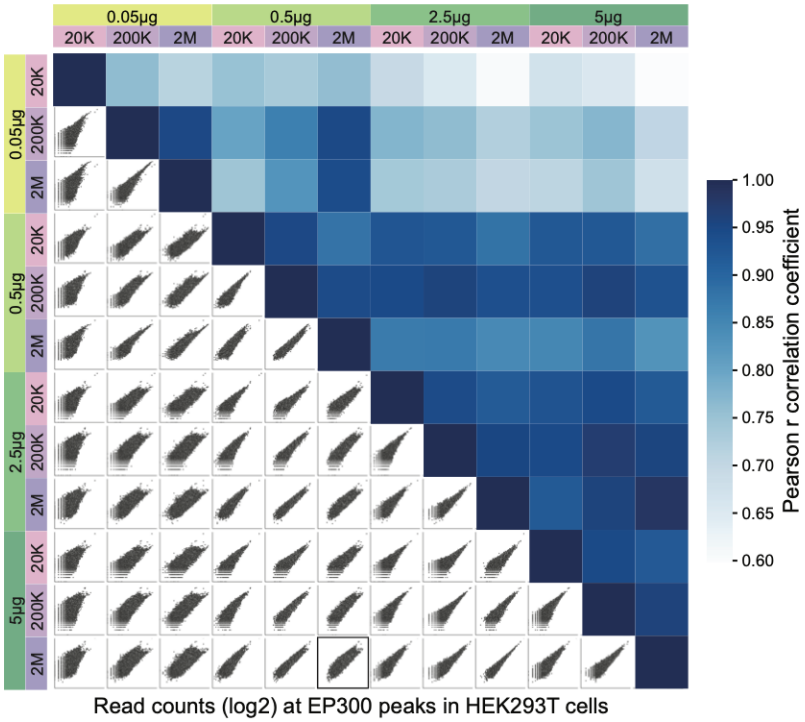
Next, to further study our hypothesis regarding the weaker signals being more sensitive to the antibody to cell-number ratio, we focused on the EP300, that binds strongly to non-promoter cis-regulatory elements and is weakly associated with promoter regions[30,31]. Using spa-ChIP-seq we conducted a comprehensive antibody titration experiment, where we varied both the number of cells and the amounts of EP300 antibody (**Fig. 4a**). In particular, we titrated 0.05 µg, 0.5 µg, 2.5 µg, and 5 µg of EP300 antibody with 20K, 200K, and 2M of HEK293T cells. To minimize variability, we used the same chromatin preparation that was split between the different conditions. We observed that, in general, the higher amount of antibody and higher number of cells tend to generate better ChIP-seq signal to noise ratios. Yet, when comparing the normalized read counts on shared EP300 peaks for the different combinations of antibody to cell-number ratios with each other, we noticed that samples with distinct ratios did not share similar set of peaks, and the Pearson r correlation coefficients varied across the combinations (**Fig. 4a**). Thus, given that the same original chromatin extract was used for all conditions, the ratio of EP300 antibody to cell-number appears to cause a discrepancy in the ChIP-seq results.

We conducted a similar antibody titration experiment using a different cell line, HeLa-S3. Here, we only varied the amount of antibody (ranging between 0.5 µg, 1 µg, 2.5 µg and 5 µg of EP300 antibody, **Supplemental Fig. 6**) with 1M HeLa-S3 cells, splitting the exact same chromatin extract between the different antibody amounts to minimize variability, as we did above. In a similar manner to our observations in HEK293T cells, we noticed a discrepancy between the genomic patterns between the different ratios of antibody to cell-number (**Supplemental Fig. 6b**).

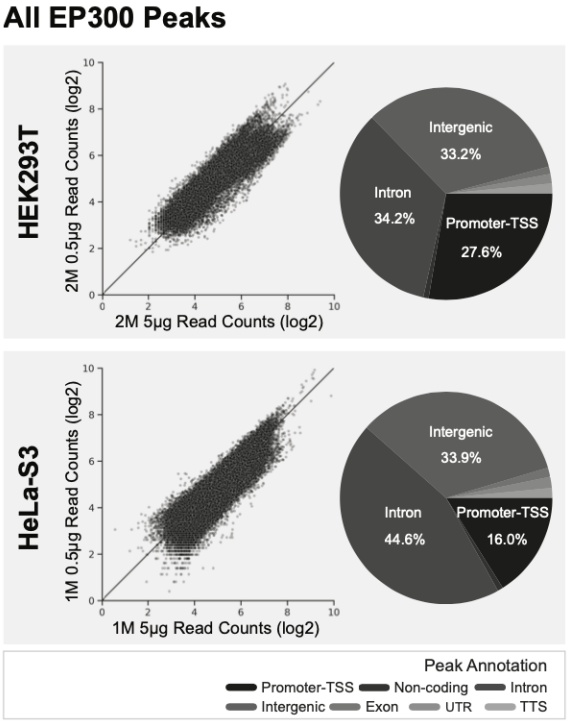
In both experiments – either when we used HEK293T or HeLa-S3 – we noticed that the major discrepancy between antibody to cell-number ratios was observed the regions with the higher EP300 read counts, in particular when comparing the lowest vs. the highest ratios (0.5 µg vs 5 µg EP300 antibody per 1M HeLa-S3 cells, respectively, **Fig. 4b**). To further study this discrepancy, we focused on the top 10th quantile and compared the normalized read counts between highest and lowest antibody to cell-number ratios. This analysis revealed a subset of peaks with stronger enrichment in the higher antibody to cell-number ratio (**Fig. 4c**). Next, we used PCA (principal component analysis) to separate out the subset of peaks that shoed an increased signal in samples with higher ratio of antibody to cell-number in both cell lines (**Supplemental Fig. 6d-e**). Annotation of the loci associated with the subset of peaks impacted by the antibody to cell-number ratio showed that these regions are predominantly present at promoter-TSS regions with 90% in HEK293T cells and 85% in HeLa-S3 cells, whereas the annotation of the entire set of EP300 peaks showed high localization at non-promoter intergenic or intron regions, and only 27% (HEK293T) and 16% (HeLa-S3) in promoter regions (**Fig. 4b-c**). Next, to better define the subset of peaks most impacted by the antibody to cell-number ratio, we conducted motif analysis, which identified a GFY motif commonly associated with some active promoter regions. Moreover, in

both cell lines, this subset of EP300 peaks was enriched for the same top known motifs (**Fig. 4c**) further supporting the notion that the weaker or less frequent localization of EP300 at promoter regions leads to the higher sensitivity of these peaks to the ratio of antibody to cell-number.

a.



b.



c.

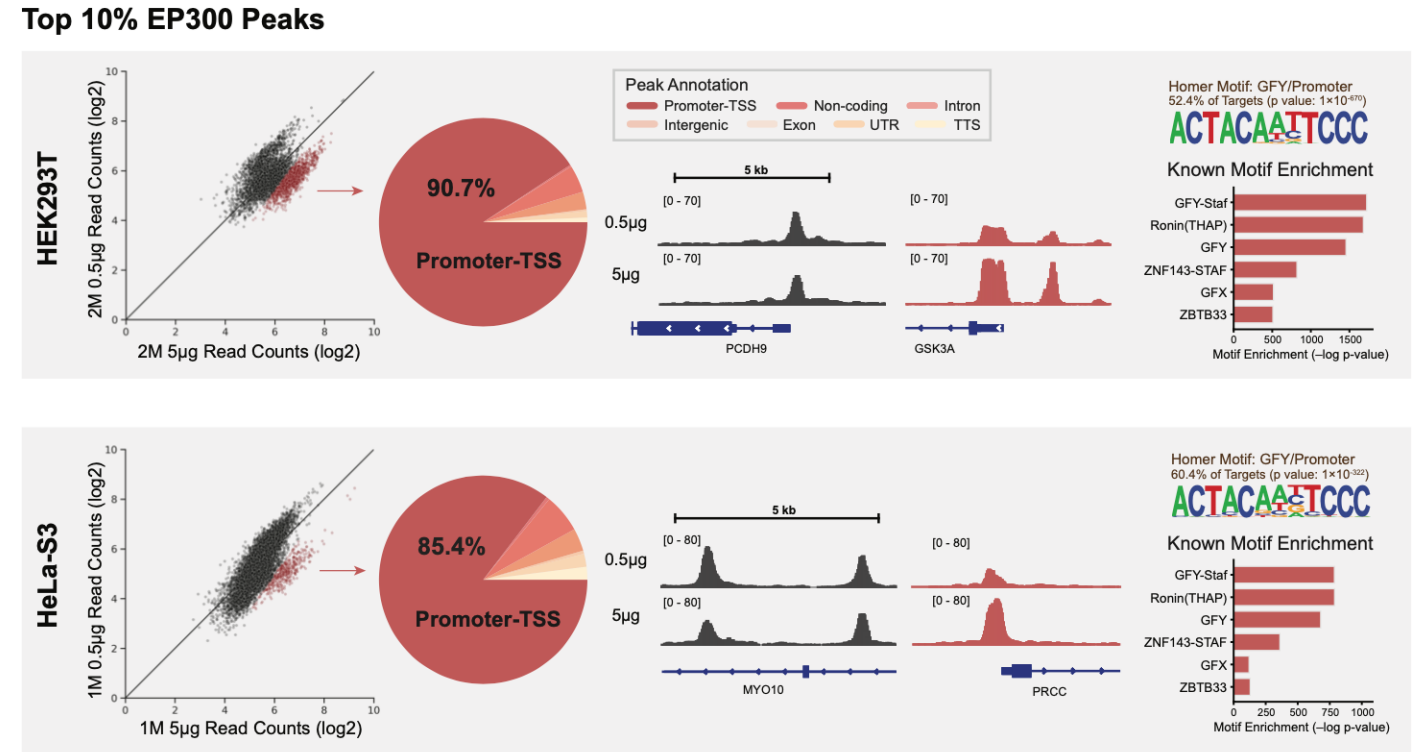


Fig. 5: EP300 antibody titration shows that higher antibody to cell-number ratio better captures weaker promoter signals. a. A matrix of scatterplots showing normalized read counts (log2) at EP300 peaks and heatmaps showing Pearson r correlation coefficients for EP300 antibody titrations in HEK293T cells with various antibody to cell-number ratios. We conducted a comprehensive comparison of conditions, with either 0.05 µg, 0.5 µg, 2.5 µg, and 5 µg of EP300 antibody (green shades) and 20K, 200K, and 2M of HEK293T cells (pink/purple shades). **b.** Scatterplot of normalized read counts (log2) at all EP300 peaks comparing 0.5 µg and 5 µg of antibody in 2M HEK293T cells (top) and 1M HeLa-S3 cells (bottom). The pie chart on the right shows the genomic annotations for all EP300 peaks. The peaks are annotated with HOMER and the annotations include promoter-TSS, non-coding, intron, intergenic, exon, UTR (5' UTR and 3' UTR), and TTS (see legend below). **c.** Scatterplot of normalized read counts (log2) at the top 10% of EP300 ChIP-seq peaks comparing 0.5 µg and 5 µg of antibody in 2M HEK293T cells (top) and 1M HeLa-S3 cells (bottom). The subset of peaks that has higher signal in 5 µg antibody than 0.5 µg antibody is highlighted in red. The pie chart shows the distribution of peak annotations in the red subset highlighted in the scatterplot to the left. IGV genome browser tracks show 2 peak loci in the red subset. HOMER motif analysis identified same top enriched known motifs in the red subset of peaks for both HeLa-S3 and HEK293T cells. The motif enrichments are shown in the bar plot as –log p-values.

DISCUSSION

We presented here spa-ChIP-seq, a single-pot automated protocol for chromatin profiling. We benchmarked spa-ChIP-seq by performing parallel manual ChIP-seq reactions and demonstrated a high reproducibility and nearly indistinguishable signal-to-noise ratio between manual and automated workflows. We then used spa-ChIP-seq we systematically evaluate several aspects of the ChIP-seq process, including shearing conditions, crosslinking methods, and buffer compositions.

In our hands, the optimal conditions included (***i***) shearing with the Active Motif PIXUL device[26], that provides a short incubation time and low cost per sample, (***ii***) lysing the cells with LB3. Although this buffer requires longer sonication time compared to EBB (potentially due to the weaker reagent), the ChIP-seq results had higher signal-to-noise ratio than EBB, (***iii***) using DSG/FA double crosslinking for all ChIP-seq profiling to allow a consistent the experimental procedure regardless of the target epitope. We demonstrated that for histone modifications the signal-to-noise ratio was indistinguishable between single (FA) and double (DSG/FA) crosslinking. As non-histone DNA-associated proteins such as chromatin regulators me be less amenable to get crosslinked to the DNA, double crosslinking by DSG/FA can yield ChIP-seq results with robust signal-to-noise ratios. While FA only forms crosslinks of 2.3–2.7 Å, DSG can form crosslinks of 7.7 Å thus potentially capturing longer interactions[14]. The combination of DSG and FA, DSG first stabilizes the longer protein-protein interactions and then FA crosslinks shorter DNA-protein interactions[15,16].

Using spa-ChIP-seq and our optimal set of conditions, we identified a set of mostly monoclonal antibodies for an array of non-histone DNA associated proteins, including CTCF, BRD4, HA-tag, HDAC3, HAT1 and YAP1. As we discussion in our previous work[13], the use of monoclonal antibodies is critical to allow reproducibility of studies e.g., between groups or over time, as polyclonal reagents are non-renewable and each batch may differ from the other. We thus envision that the validation of this set of antibodies and conditions to allow reproducible studies of these targets.

A key observation of our study relates to the impact of antibody to cell-number ratio in chromatin profiling. Using spa-ChIP-seq we conducted a set of comprehensive experiments to systematically interrogate the effect of this ratio. For each experiment, we used the exact same chromatin prep, thus only varying the ratios of the antibodies or the number of cells. First, we demonstrated that only the weaker quartiles of H3K9ac peaks are impacted by the antibody to cell-number ratio, while the upper quartile is not. We then used a BRD4 HA-tagged perturbation system to reduce the amount of epitope by degrading HA-tagged BRD4 and show that when we titrate either the HA-tag or BRD4 antibodies following depletion of the target, the normalized ChIP enrichment signal is impacted by the antibody to cell-number ratio much more then prior to depletion. Lastly, we focused on EP300 that has both strong (enhancer/cis regulatory elements) and weak (promoter) binding sites[30,31] and systematically titrated both the number of cells and the amount of antibody in two cell lines. This allowed us to identify a subset of the peaks that show enhanced sensitivity to the antibody to cell-number ratio and show that they are highly enriched for promoter signatures. Thus, our results highlight the importance of maintaining the ratio of antibody to cell-

number as constant as possible, in particular in chromatin profiling experiments aimed at studying the impact of perturbations (e.g., by small molecules or CRISPRi) to allow proper comparisons.

Altogether our study highlights the usefulness of spa-ChIP-seq in conducting comprehensive, systematic studies. The method reduces hands-on time and cost per sample, and the increased throughput (up to 96 samples) were key components in our ability to conduct our systematic studies. We anticipate that spa-ChIP-seq will allow the scaling up of ChIP-seq studies to improve statistical power and facilitate novel experimental work including assaying entire population cohorts and large compound libraries as well as potentially advance diagnostic approaches.

METHODS

Cell culture preparation

HeLa-S3 (ATCC CCL-2.2™), HEK293T (CRL-3216™), and HEK293T-BRD4-FKBP12^{F36V}-2xHA adherent cells were cultured in Dulbecco's Modified Eagle's Medium (Gibco™ DMEM, high glucose, GlutaMAX™ Supplement, with 4.5 g/L D-Glucose), supplemented with 10% fetal bovine serum (FBS), 110mg/L sodium pyruvate, 2 mM L-glutamine, and 1% Antibiotic-Antimycotic. Cells were incubated at 37°C with 5% CO_2 and 95% Air. HEK293T-BRD4-FKBP12^{F36V}-2xHA cells were treated with 500 nM of dTAG-13 (Tocris, cat #6605) or DMSO for 24 hours. When the cells were grown to confluency, they were harvested by adding 0.25% trypsin.

Cell fixation (crosslinking)

Double crosslinking: Per 1 million cells, cells were incubated first with at least 1 mL of 2mM DSG solution at room temperature for 30 minutes and then with 1% formaldehyde solution at room temperature for 10 minutes. The crosslinking reaction was quenched by adding 1/20th volume of 2.625 M glycine and 1/20th volume 10% BSA. The cells were washed 3 times with 0.5% BSA/PBS and centrifuged at 1000g for 5 minutes at 4°C. The washed cell pellets were stored in -80°C.

Single crosslinking: Per 1 million cells, cells were incubated with 1% formaldehyde solution in PBS at room temperature for 10 minutes. The crosslinking reaction was quenched by adding 1/20th volume of 2.625 M glycine and 1/20th volume 10% BSA. The cells were washed 3 times with 0.5% BSA/PBS and centrifuged at 1000g for 5 minutes at 4°C. The washed cell pellets were stored in -80°C.

Chromatin shearing

Cells were lysed with one of the following lysis buffers: SDS lysis buffer [1% SDS, 50 mM Tris-HCl (pH 8.1), 10 mM EDTA, 1× protease inhibitor cocktail], LB3 [10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.1% Deoxycholate, 0.5% N-Lauroylsarcosine, 1× protease inhibitor cocktail], or EBB lysis buffer [0.5% Empigen BB (Sigma-Aldrich, #30326), 1% SDS, 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1× protease inhibitor cocktail][28]. Cells were sheared either on the PIXUL sonicator for 60-120 minutes per plate, or on the Covaris E220 sonicator for 6 minutes per well. After the chromatin was sonicated, lysates sheared with SDS lysis buffer was diluted 10-fold with CDB (ChIP dilution buffer) [0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.mM Tris-HCl (pH 8.1), 167mM NaCl, 1 × protease inhibitor cocktail] and lysates sheared with EBB lysis buffer were diluted 2.5-fold in EBB dilution buffer [20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.5% Triton X-100, 2 mM EDTA, 1× protease inhibitor cocktail).

Input (WCE) preparation

Manual workflow: 1% of each sonicated lysate was saved as input sample. Per input sample, a reaction mix (0.5% SDS, 18.75mM EDTA, 280mM NaCl, 250 µg/mL Proteinase K, 125 µg/mL RNase A) was added and incubated in a thermocycler at 55°C for 1 hour and then 65°C for at least 2 hours. To clean up the input sample,

Speedbead mastermix [2µL SpeedBeads (Cytiva # 65152105050250), 12% PEG, 1M NaCl] was added to each input and incubated at room temperature for 10 minutes. Then, the input sample was washed twice with 80% EtOH and the Speedbeads pellet was air-dried. Lastly, the input sample was eluted with TT buffer [0.05% Tween 20, 10mM Tris pH 8.0).

Bravo workflow: Input plate was prepared with 1% of each lysate sample. Reaction mix plate was prepared by equally distributing the total required amount of reaction mix to each of the well in the first column and placed on Bravo deck 6. After running Bravo protocol *1a. Input Preparation*, the input plate was incubated in a thermocycler at 55°C for 1 hour and then 65°C for at least 2 hours. Speedbeads mastermix plate was prepared by equally distributing the total required amount of Speedbeads mastermix to each of the well in the first column and placed on Bravo deck 5. After incubation, the input plate was placed back on Bravo deck 4, and Bravo protocol *1b. Input PEG* was run. Ethanol reservoir was placed on deck 9, and TT buffer reservoir was placed on deck 3. Immediately after incubation with Speedbeads mastermix, Bravo protocol *1c. Input Speedbeads cleanup* was run to clean up the input samples. At the end of the protocols, the input plate should contain input samples eluted in TT buffer.

Chromatin immunoprecipitation

See **Supplemental Table 2** for all the antibodies used in the immunoprecipitation step.

Manual workflow: Dynabeads™ Protein A or G were resuspended either in LB3 and 1 % Triton X-100 for LB3 ChIP samples or in EBB dilution buffer for EBB ChIP samples. Resuspended beads were combined with each antibody to make antibody/beads mix. Per each ChIP sample, sheared chromatin (~500k cells) was combined with the antibody/bead mix.

Bravo workflow: The antibody/beads mix was made the same way as the manual workflow. The mix was transferred to tube strips, which were then placed on deck 4. PIXUL lysate plate was placed on deck 5. Bravo protocol *2. Chromatin Beads Antibody on Mag* was run to combine the chromatin lysate with the antibody/beads mix.

The ChIP samples were incubated overnight, rotating overhead on HulaMixer™ Sample Mixer at 4°C in the cold room.

ChIP Washes

Manual workflow: After overnight incubation, the sample was washed three times with WBI [10mM Tris pH7.5, 0.7% DOC, 1mM EDTA, 250mM LiCl, 1× protease inhibitor cocktail], three times with WBIII [20mM Tris pH7.5, 1% Triton X-100, 2mM EDTA, 150mM NaCl, 0.1% SDS, 1× protease inhibitor cocktail], and twice with TET [10mM Tris pH7.5, 0.2% Tween-20, 1mM EDTA, 1× protease inhibitor cocktail]. Each ChIP sample was eluted in 25µL of TT [10 mM Tris-HCl pH 8.0, 0.05% Tween 20].

Bravo workflow: Excessive amounts of wash buffers (WBI, WBII, and TET) and TT buffer were prepared. Waste reservoir was placed on deck 1, elution TT buffer reservoir was placed on deck 3, empty ChIP plate was placed

on deck 4, IP tube strips were placed on deck 6, and wash buffer reservoir was placed on deck 9. Bravo protocol *3. ChIP Wash Elution off Mag* was run. At the end of the protocol, each well of the ChIP plate should contain 25µL of ChIP sample eluted in TT buffer.

Library construction, reverse crosslinking, and library cleanup

For the IP and input samples, libraries were constructed with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (Catalog #E7645L). Indexing adapters were from NEXTflex DNA Barcodes: 48 (NOVA-514104) and NEXTflex Unique Dual Index Barcodes (8NT index, 97-192) (NOVA-514151).

Manual workflow: **(1) End prep:** End prep mastermix (NEBNext Ultra II End Prep Enzyme Mix and NEBNext Ultra II End Prep Reaction Buffer) was added to the ChIP and input samples and incubated for 30 minutes at 20°C, and then 30minutes at 65°C in a thermal cycler. **(2) Adapter ligation:** 0.625 µM of indexing adapters and ligation mastermix (NEBNext Ultra II Ligation Master Mix and NEBNext Ligation Enhancer) were added to the ChIP and input samples, which were then incubated for 15 minutes at 20°C. **(3) Reverse crosslinking:** Per sample, ligation reaction was stopped by adding 27µL of ligation stop solution mastermix [0.5% SDS, 18.75 mM EDTA, 250 µg/mL Proteinase K, 125 µg/mL RNase A] and 285mM NaCl. ChIP samples were incubated for 1 hour at 55°C and then at least 2 hours at 65°C. **(4) Library cleanup:** The supernatant of each sample was separated from Dynabeads Protein A/G and incubated with Speedbeads mastermix [2µL SpeedBeads, 8.6% PEG, and 0.8M NaCl] for 10 minutes at room temperature. Then, the input sample was washed twice with 80% EtOH and the Speedbeads pellet was air-dried. Lastly, the input sample was eluted with TT buffer.

Bravo workflow: **(1) End prep:** End prep mastermix plate was prepared by equally distributing the total required amount of the mastermix to each of the well in the first column and placed on Bravo deck 6. The ChIP sample plate was placed on deck 4. Bravo protocol *4. End Repair MM Dispense* was run and the ChIP plate was incubated for 30 minutes at 20°C, and then 30 minutes at 65°C in a thermal cycler. **(2) Adapter ligation:** Ligation mastermix plate was prepared by equally distributing the total required amount of the mastermix to each of the well in the first column and placed on Bravo deck 6. The ChIP sample plate was placed on deck 4, and the index plate was placed on deck 5. Bravo protocol *5. Adapter Ligation MM Dispense* was run and the ChIP plate was incubated for 15 minutes at 20°C. **(3) Reverse crosslink:** Ligation stop solution mastermix plate was prepared by equally distributing the total required amount of the mastermix to each of the well in the first column and placed on Bravo deck 6. The ChIP sample plate was placed on deck 4, and the 5M NaCl reservoir was placed on deck 9. was prepared as a Buffer Plate and the plate was placed on deck 5. Bravo protocol *6. Ligation Stop Solution Dispense* was run and the sample was incubated for 1 hour at 55°C and then at least 2 hours at 65°C. **(4) Library cleanup:** Speedbeads mastermix plate was prepared by equally distributing the total required amount of Speedbeads mastermix to each of the well in the first column and placed on Bravo deck 5. After incubation, the ChIP plate was placed back on Bravo deck 4, and Bravo protocol *7a. Adapter Ligation Cleanup* was run. Ethanol reservoir was placed on deck 9, and TT buffer reservoir was placed on deck 3. Immediately after

incubation with Speedbeads mastermix, Bravo protocol *7b. Adapter Ligation Cleanup* was run to clean up the input samples. At the end of the protocols, the ChIP plate should contain ChIP samples eluted in TT buffer.

PCR Amplification and final cleanup of ChIP and Input libraries

Manual workflow: **(1) PCR amplification:** The PCR mastermix [Ultra II Q5 2x Mastermix, and 25 μM sequencing platform-compatible forward and reverse primers (Solexa 1GA and Solexa 1GB)] was added to each library sample. A PCR program (Initial denaturation at 98°C for 30 seconds; 12 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 15 seconds, extension at 72°C for 30 seconds; and final extension at 72°C for 2 minutes) was run to amplify the libraries using a thermal cycler. **(2) PCR cleanup:** The PCR products were incubated with Speedbeads mastermix [2μL SpeedBeads, 8.5% PEG, and 1M NaCl] for 10 minutes at room temperature. Then, the input sample was washed twice with 80% EtOH and the Speedbeads pellet was air-dried. Lastly, the input sample was eluted with TT buffer.

Bravo workflow: **(1) PCR amplification:** PCR mastermix plate was prepared by equally distributing the total required amount of the mastermix to each of the well in the first column and placed on Bravo deck 6. The ChIP library plate was placed on deck 4. Bravo protocol *8. PCR Enrichment MM Dispense* was run. Then, a PCR program (Initial denaturation at 98°C for 30 seconds; 12 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 15 seconds, extension at 72°C for 30 seconds; and final extension at 72°C for 2 minutes) was run to amplify the libraries using a thermal cycler. **(2) Library cleanup:** Speedbeads mastermix plate was prepared by equally distributing the total required amount of Speedbeads mastermix to each of the well in the first column and placed on Bravo deck 5. After incubation, the ChIP library plate was placed back on Bravo deck 4, and Bravo protocol *9a. PCR Enrichment Cleanup* was run. Ethanol reservoir was placed on deck 9, and TT buffer reservoir was placed on deck 3. Immediately after incubation with Speedbeads mastermix, Bravo protocol *9b. PCR Enrichment Cleanup* was run to clean up the input samples. At the end of the protocols, the ChIP library plate should contain ChIP library samples eluted in TT buffer.

DNA Quantification and sequencing

The size of the DNA libraries was quantified on a 2% agarose gel pre-stained with GelGreen. The concentration was estimated with the Qubit HS DNA assay kit (Invitrogen). The libraries were pooled according to the concentration calculations. The pooled libraries were prepared and sequenced on Illumina NextSeq 500 or NextSeq 2000.

Sequencing data analysis

Samples were demultiplexed with Illumina's bcl2fastq program, adapters were removed with Trimmomatic[32], and FastQC was run for sequencing quality control. Reads were mapped to human reference genome hg38 using Bowtie2[33] with default parameters and duplicated reads were removed with SAMtools[34] (`samtools markdup`). HOMER[35] was used to make tag directories, which contain filtered read counts for primary alignments with

MAPQ > 10. HOMER makeBigWig was used to create bigwig files for visualization on IGV genome browser[27]. Peaks were called with HOMER findPeaks program with parameters "-style histone", "-style histone -size 1000 -minDist 2500" for broad peaks, and "-style factor" for transcription factors, using an input control as background. Histograms centered on Transcription Start Site (TSS) or peaks were generated with HOMER annotatePeaks.pl program set in "TSS" mode or with peak files (`tss or peaks, hg38, -size 4000, -hist 10`). The HOMER annotatePeaks.pl program was used to quantify the read-depth normalized read counts at the merged peaks from all samples for each antibody (made with mergePeaks program). Correlation scatterplots were created with python seaborn. Both HOMER programs automatically normalize each tag directory by the total number of mapped reads. Heatmaps of H3K27ac and H3K27me3 were generated using deepTools[36] computeMatrix and plotHeatmap with either the TSS or shared peaks. Other plots were made with customized python scripts.

DATA AVAILABILITY

Raw data are available on SRA with GEO Accession Number GSE304259.

ACKNOWLEDGEMENTS

Research reported in this publication was supported in part by NIH/NIMH grants R01MH127077 (A.G. and C.B.), R35GM149520 (C.B), R01GM129523 (S.H.), NSF grant 2003358 (A.G.). We thank T. Xu and T. Dishon for their input, and R. Wachs for her help with the illustrations. We thank the Stem Cell Genomics Core at the Sanford Stem Cell Institute and the Epigenomics Core at UC San Diego for providing sequencing services. We also thank N. Gray and B. Nabet (Dana-Farber Cancer Institute) for the HEK293T-BRD4-FKBP12^{F36V}-2xHA cell line and for the insightful input.

CONFLICT OF INTEREST

L.P., Y.C., C.B., S.H. and A.G. are inventors on related patent applications. L.A. is an employee of Agilent Technologies Inc.

REFERENCES

(1) Texari, L.; Spann, N. J.; Troutman, T. D.; Sakai, M.; Seidman, J. S.; Heinz, S. An Optimized Protocol for Rapid, Sensitive and Robust on-Bead ChIP-Seq from Primary Cells. *STAR Protocols* **2021**, *2* (1), 100358. https://doi.org/10.1016/j.xpro.2021.100358.

(2) Patel, L. A.; Cao, Y.; Mendenhall, E. M.; Benner, C.; Goren, A. The Wild West of Spike-in Normalization. *Nat Biotechnol* **2024**, *42* (9), 1343–1349. https://doi.org/10.1038/s41587-024-02377-y.

(3) Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **2007**, *129* (4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009.

(4) Johnson, D. S.; Mortazavi, A.; Myers, R. M.; Wold, B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* **2007**, *316* (5830), 1497–1502. https://doi.org/10.1126/science.1141319.

(5) Mikkelsen, T. S.; Ku, M.; Jaffe, D. B.; Issac, B.; Lieberman, E.; Giannoukos, G.; Alvarez, P.; Brockman, W.; Kim, T.-K.; Koche, R. P.; Lee, W.; Mendenhall, E.; O'Donovan, A.; Presser, A.; Russ, C.; Xie, X.; Meissner, A.; Wernig, M.; Jaenisch, R.; Nusbaum, C.; Lander, E. S.; Bernstein, B. E. Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells. *Nature* **2007**, *448* (7153), 553–560. https://doi.org/10.1038/nature06008.

(6) Furey, T. S. ChIP–Seq and beyond: New and Improved Methodologies to Detect and Characterize Protein–DNA Interactions. *Nat Rev Genet* **2012**, *13* (12), 840–852. https://doi.org/10.1038/nrg3306.

(7) Park, P. J. ChIP–Seq: Advantages and Challenges of a Maturing Technology. *Nat Rev Genet* **2009**, *10* (10), 669–680. https://doi.org/10.1038/nrg2641.

(8) Mundade, R.; Ozer, H. G.; Wei, H.; Prabhu, L.; Lu, T. Role of ChIP-Seq in the Discovery of Transcription Factor Binding Sites, Differential Gene Regulation Mechanism, Epigenetic Marks and Beyond. *Cell Cycle* **2014**, *13* (18), 2847–2852. https://doi.org/10.4161/15384101.2014.949201.

(9) Garber, M.; Yosef, N.; Goren, A.; Raychowdhury, R.; Thielke, A.; Guttman, M.; Robinson, J.; Minie, B.; Chevrier, N.; Itzhaki, Z.; Blecher-Gonen, R.; Bornstein, C.; Amann-Zalcenstein, D.; Weiner, A.; Friedrich, D.; Meldrim, J.; Ram, O.; Cheng, C.; Gnirke, A.; Fisher, S.; Friedman, N.; Wong, B.; Bernstein, B. E.; Nusbaum, C.; Hacohen, N.; Regev, A.; Amit, I. A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. *Molecular Cell* **2012**, *47* (5), 810–822. https://doi.org/10.1016/j.molcel.2012.07.030.

(10) Blecher-Gonen, R.; Barnett-Itzhaki, Z.; Jaitin, D.; Amann-Zalcenstein, D.; Lara-Astiaso, D.; Amit, I. High-Throughput Chromatin Immunoprecipitation for Genome-Wide Mapping of in Vivo Protein-DNA Interactions and Epigenomic States. *Nat Protoc* **2013**, *8* (3), 539–554. https://doi.org/10.1038/nprot.2013.023.

(11) Aldridge, S.; Watt, S.; Quail, M. A.; Rayner, T.; Lukk, M.; Bimson, M. F.; Gaffney, D.; Odom, D. T. AHT-ChIP-Seq: A Completely Automated Robotic Protocol for High-Throughput Chromatin Immunoprecipitation. *Genome Biology* **2013**, *14* (11), R124. https://doi.org/10.1186/gb-2013-14-11-r124.

(12) Gasper, W. C.; Marinov, G. K.; Pauli-Behn, F.; Scott, M. T.; Newberry, K.; DeSalvo, G.; Ou, S.; Myers, R. M.; Vielmetter, J.; Wold, B. J. Fully Automated High-Throughput Chromatin Immunoprecipitation for ChIP-Seq: Identifying ChIP-Quality P300 Monoclonal Antibodies. *Sci Rep* **2014**, *4* (1), 5152. https://doi.org/10.1038/srep05152.

(13) Busby, M.; Xue, C.; Li, C.; Farjoun, Y.; Gienger, E.; Yofe, I.; Gladden, A.; Epstein, C. B.; Cornett, E. M.; Rothbart, S. B.; Nusbaum, C.; Goren, A. Systematic Comparison of Monoclonal versus Polyclonal Antibodies for Mapping Histone Modifications by ChIP-Seq. *Epigenetics & Chromatin* **2016**, *9* (1), 49. https://doi.org/10.1186/s13072-016-0100-6.

(14) Sutherland, B. W.; Toews, J.; Kast, J. Utility of Formaldehyde Cross-Linking and Mass Spectrometry in the Study of Protein–Protein Interactions. *Journal of Mass Spectrometry* **2008**, *43* (6), 699–715. https://doi.org/10.1002/jms.1415.

(15) Nowak, D. E.; Tian, B.; Brasier, A. R. Two-Step Cross-Linking Method for Identification of NF-κB Gene Network by Chromatin Immunoprecipitation. *BioTechniques* **2005**, *39* (5), 715–725. https://doi.org/10.2144/000112014.

(16) Tian, B.; Yang, J.; Brasier, A. R. Two-Step Cross-Linking for Analysis of Protein–Chromatin Interactions. In *Transcriptional Regulation: Methods and Protocols*; Vancura, A., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2012; pp 105–120. https://doi.org/10.1007/978-1-61779-376-9_7.

(17) Marx, V. What to Do about Those Immunoprecipitation Blues. *Nat Methods* **2019**, *16* (4), 289–292. https://doi.org/10.1038/s41592-019-0365-3.

(18) Caride, A.; Jang, J. S.; Shi, G.-X.; Lenz, S.; Zhong, J.; Kim, K. H.; Allen, M.; Robertson, K. D.; Farrugia, G.; Ordog, T.; Ertekin-Taner, N.; Lee, J.-H. Titration-Based Normalization of Antibody Amount Improves Consistency of ChIP-Seq Experiments. *BMC Genomics* **2023**, *24* (1), 171. https://doi.org/10.1186/s12864-023-09253-0.

(19) Landt, S. G.; Marinov, G. K.; Kundaje, A.; Kheradpour, P.; Pauli, F.; Batzoglou, S.; Bernstein, B. E.; Bickel, P.; Brown, J. B.; Cayting, P.; Chen, Y.; DeSalvo, G.; Epstein, C.; Fisher-Aylor, K. I.; Euskirchen, G.; Gerstein, M.; Gertz, J.; Hartemink, A. J.; Hoffman, M. M.; Iyer, V. R.; Jung, Y. L.; Karmakar, S.; Kellis, M.; Kharchenko, P. V.; Li, Q.; Liu, T.; Liu, X. S.; Ma, L.; Milosavljevic, A.; Myers, R. M.; Park, P. J.; Pazin, M. J.; Perry, M. D.; Raha, D.; Reddy, T. E.; Rozowsky, J.; Shoresh, N.; Sidow, A.; Slattery, M.; Stamatoyannopoulos, J. A.; Tolstorukov, M. Y.; White, K. P.; Xi, S.; Farnham, P. J.; Lieb, J. D.; Wold, B. J.; Snyder, M. ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia. *Genome Res.* **2012**, *22* (9), 1813–1831. https://doi.org/10.1101/gr.136184.111.

(20) Wardle, F. C.; Tan, H. A ChIP on the Shoulder? Chromatin Immunoprecipitation and Validation Strategies for ChIP Antibodies. F1000Research July 13, 2015. https://doi.org/10.12688/f1000research.6719.1.

(21) Mendoza-Parra, M.-A.; Saravaki, V.; Cholley, P.-E.; Blum, M.; Billoré, B.; Gronemeyer, H. Antibody Performance in ChIP-Sequencing Assays: From Quality Scores of Public Data Sets to Quantitative Certification. F1000Research March 11, 2016. https://doi.org/10.12688/f1000research.7637.2.

(22) Zou, Z.; Iwata, M.; Yamanishi, Y.; Oki, S. Epigenetic Landscape of Drug Responses Revealed through Large-Scale ChIP-Seq Data Analyses. *BMC Bioinformatics* **2022**, *23* (1), 51. https://doi.org/10.1186/s12859-022-04571-8.

(23) Waszak, S. M.; Delaneau, O.; Gschwind, A. R.; Kilpinen, H.; Raghav, S. K.; Witwicki, R. M.; Orioli, A.; Wiederkehr, M.; Panousis, N. I.; Yurovsky, A.; Romano-Palumbo, L.; Planchon, A.; Bielser, D.; Padioleau, I.; Udin, G.; Thurnheer, S.; Hacker, D.; Hernandez, N.; Reymond, A.; Deplancke, B.; Dermitzakis, E. T. Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell* **2015**, *162* (5), 1039–1050. https://doi.org/10.1016/j.cell.2015.08.001.

(24) Yan, H.; Tian, S.; Slager, S. L.; Sun, Z. ChIP-Seq in Studying Epigenetic Mechanisms of Disease and Promoting Precision Medicine: Progresses and Future Directions. *Epigenomics* **2016**, *8* (9), 1239–1258. https://doi.org/10.2217/epi-2016-0053.

(25) Dainese, R.; Gardeux, V.; Llimos, G.; Alpern, D.; Jiang, J. Y.; Meireles-Filho, A. C. A.; Deplancke, B. A Parallelized, Automated Platform Enabling Individual or Sequential ChIP of Histone Marks and Transcription Factors. *Proceedings of the National Academy of Sciences* **2020**, *117* (24), 13828–13838. https://doi.org/10.1073/pnas.1913261117.

(26) Bomsztyk, K.; Mar, D.; Wang, Y.; Denisenko, O.; Ware, C.; Frazar, C. D.; Blattler, A.; Maxwell, A. D.; MacConaghy, B. E.; Matula, T. J. PIXUL-ChIP: Integrated High-Throughput Sample Preparation and Analytical Platform for Epigenetic Studies. *Nucleic Acids Research* **2019**, *47* (12), e69. https://doi.org/10.1093/nar/gkz222.

(27) Robinson, J. T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E. S.; Getz, G.; Mesirov, J. P. Integrative Genomics Viewer. *Nat Biotechnol* **2011**, *29* (1), 24–26. https://doi.org/10.1038/nbt.1754.

(28) Métivier, R.; Penot, G.; Hübner, M. R.; Reid, G.; Brand, H.; Koš, M.; Gannon, F. Estrogen Receptor-α Directs Ordered, Cyclical, and Combinatorial Recruitment of Cofactors on a Natural Target Promoter. *Cell* **2003**, *115* (6), 751–763. https://doi.org/10.1016/S0092-8674(03)00934-6.

(29) Nabet, B.; Roberts, J. M.; Buckley, D. L.; Paulk, J.; Dastjerdi, S.; Yang, A.; Leggett, A. L.; Erb, M. A.; Lawlor, M. A.; Souza, A.; Scott, T. G.; Vittori, S.; Perry, J. A.; Qi, J.; Winter, G. E.; Wong, K.-K.; Gray, N. S.; Bradner, J. E. The dTAG System for Immediate and Target-Specific Protein Degradation. *Nat Chem Biol* **2018**, *14* (5), 431–441. https://doi.org/10.1038/s41589-018-0021-8.

(30) Heintzman, N. D.; Stuart, R. K.; Hon, G.; Fu, Y.; Ching, C. W.; Hawkins, R. D.; Barrera, L. O.; Van Calcar, S.; Qu, C.; Ching, K. A.; Wang, W.; Weng, Z.; Green, R. D.; Crawford, G. E.; Ren, B. Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome. *Nat Genet* **2007**, *39* (3), 311–318. https://doi.org/10.1038/ng1966.

(31) Mendenhall, E. M.; Bernstein, B. E. Chromatin State Maps: New Technologies, New Insights. *Current Opinion in Genetics & Development* **2008**, *18* (2), 109–115. https://doi.org/10.1016/j.gde.2008.01.010.

(32) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

(33) Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat Methods* **2012**, *9* (4), 357–359. https://doi.org/10.1038/nmeth.1923.

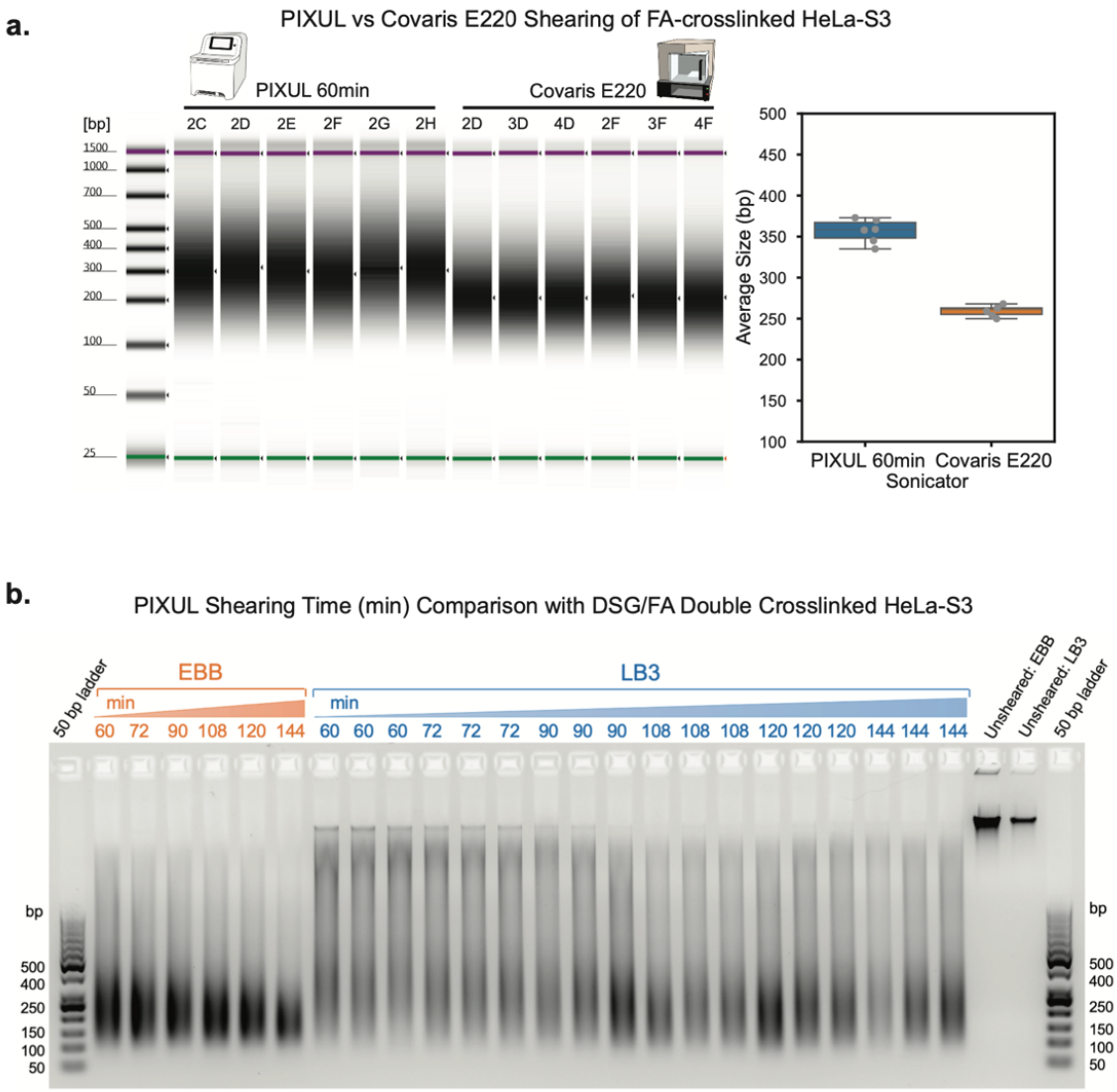
(34) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

(35) Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y. C.; Laslo, P.; Cheng, J. X.; Murre, C.; Singh, H.; Glass, C. K. Simple Combinations of Lineage-Determining Transcription Factors Prime *Cis*-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **2010**, *38* (4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

(36) Ramírez, F.; Ryan, D. P.; Grüning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A. S.; Heyne, S.; Dündar, F.; Manke, T. deepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis. *Nucleic Acids Research* **2016**, *44* (W1), W160–W165. https://doi.org/10.1093/nar/gkw257.

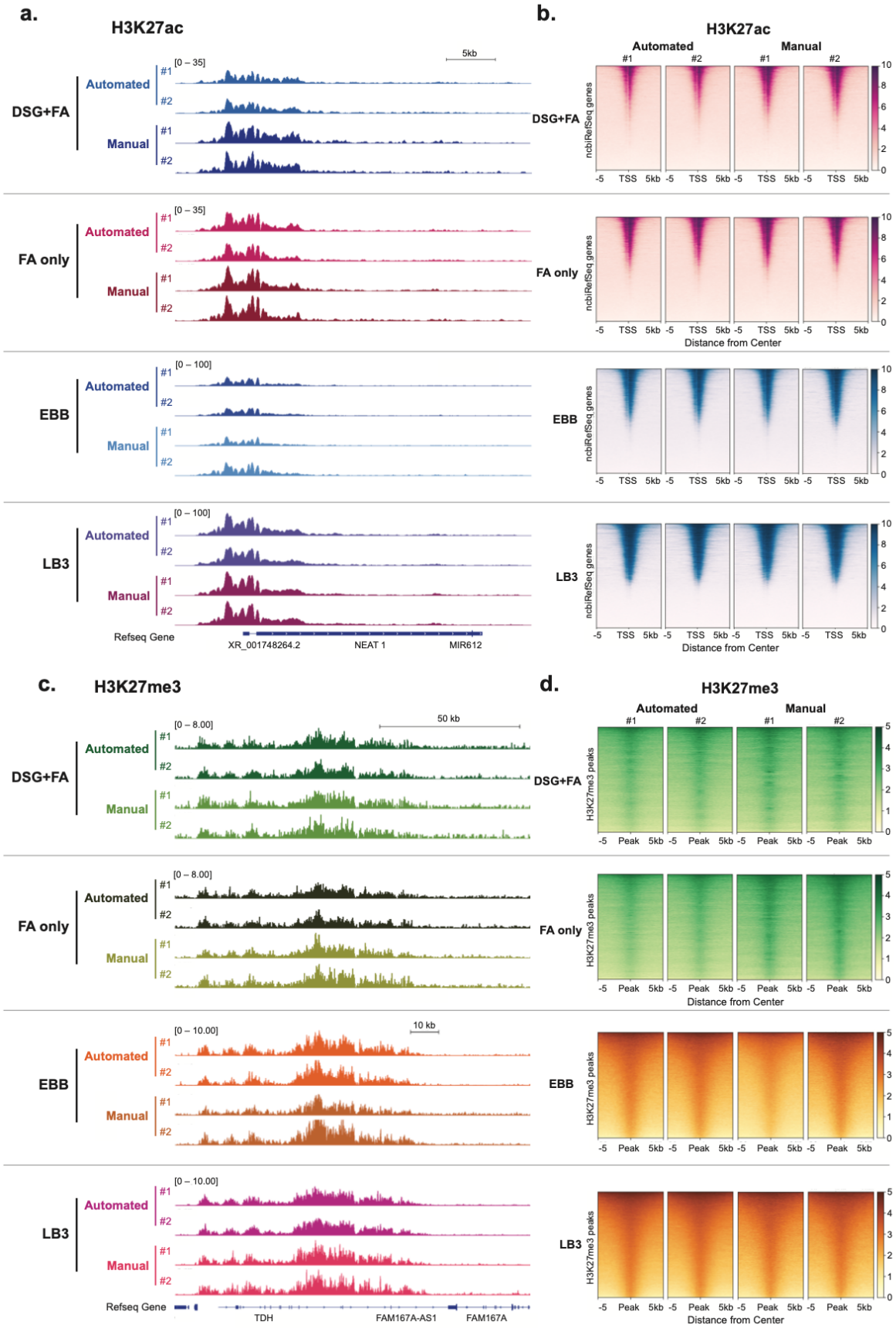
SUPPLEMENTAL FIGURES

Supplemental Fig. 1

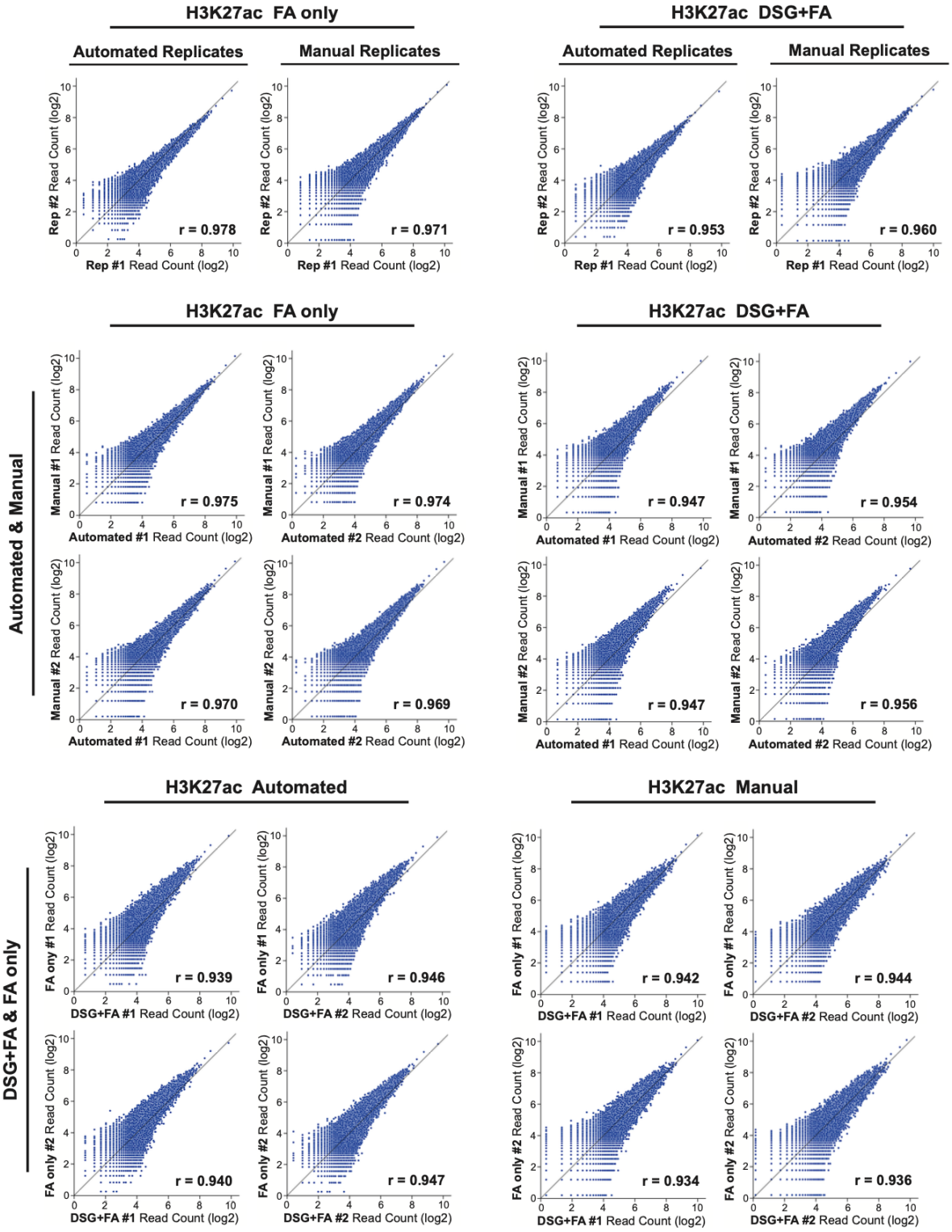


Supplemental Fig. 1: Comparison of 96-well sonicators and lysis buffers for shearing of crosslinked cells. a. TapeStation analysis comparing the two 96-well sonicators: PIXUL and Covaris E220 for shearing of FA-crosslinked HeLa-S3. Boxplot on the right shows the average size (bp) of sheared chromatin on each device based on TapeStation analysis. Error bars represent standard deviations. **b.** 2% agarose gel showing the comparison of sonication time on the PIXUL for lysis buffers EBB and LB3 with DSG/FA double crosslinked HeLa-S3 cells.

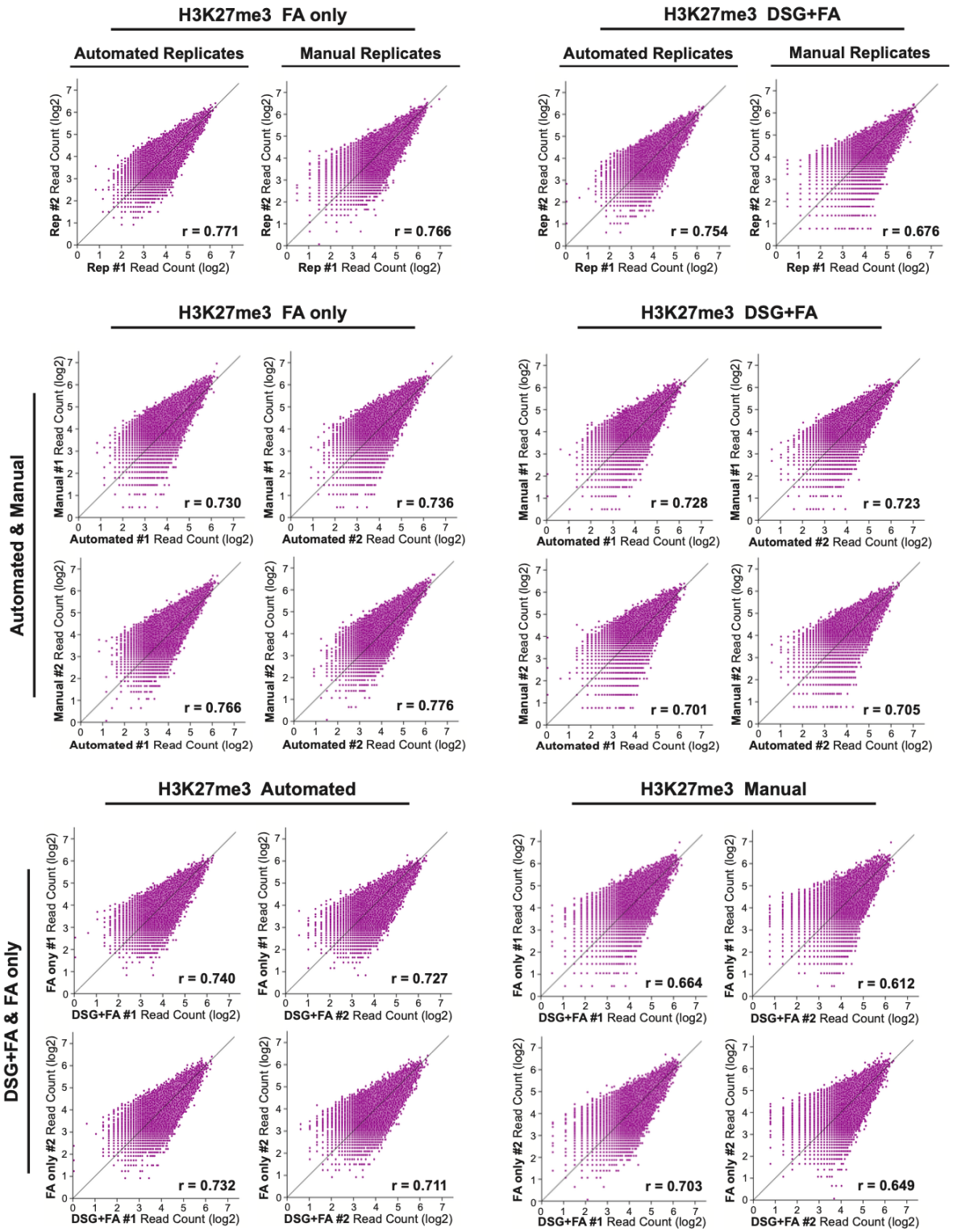
Supplemental Fig. 2



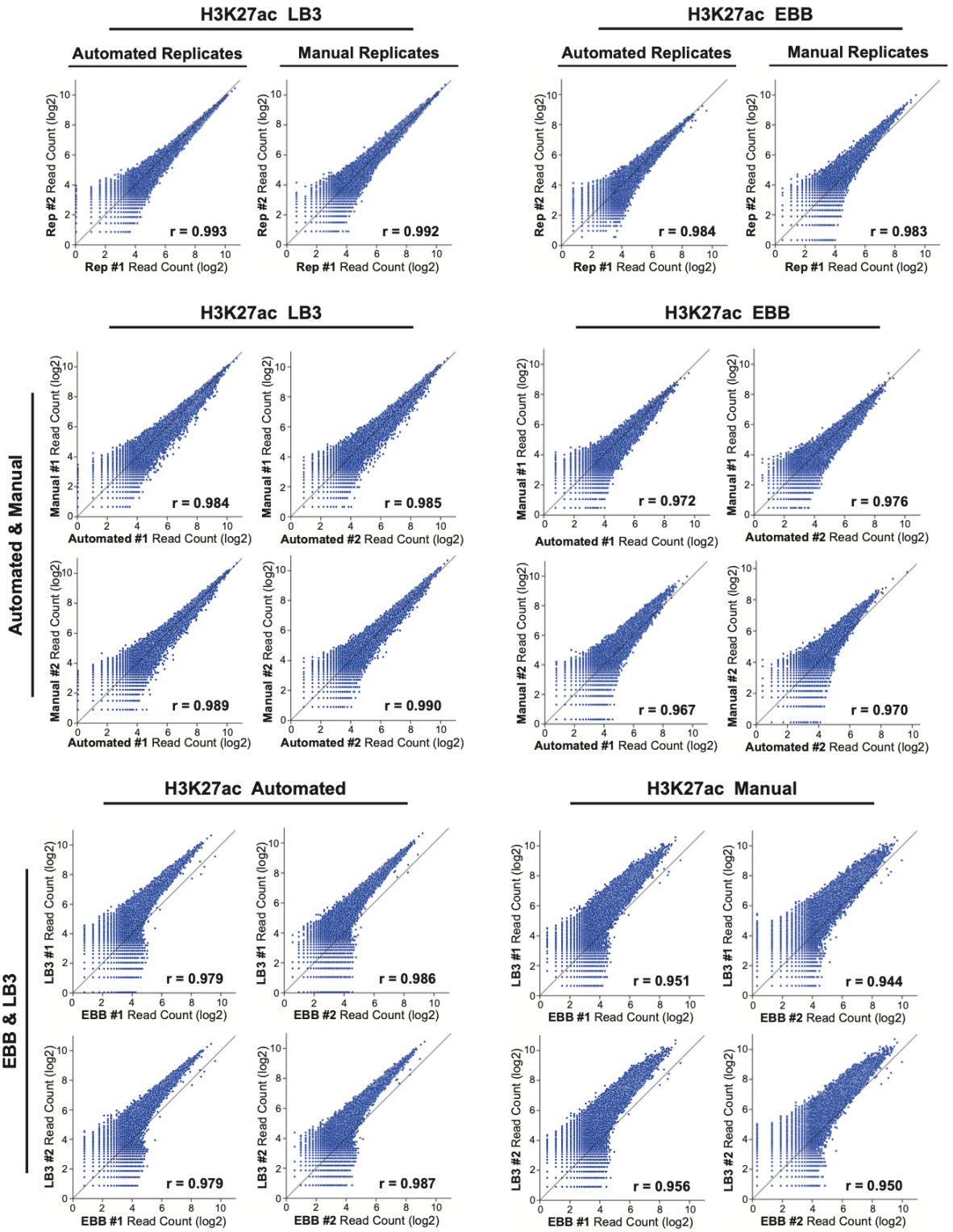
e.


H3K27ac FA only
Automated Replicates
Manual Replicates
H3K27ac DSG+FA
Automated Replicates
Manual Replicates
Rep #2 Read Count (log2)
Rep #1 Read Count (log2)
r = 0.978
r = 0.971
r = 0.953
r = 0.960
Automated & Manual
H3K27ac FA only
H3K27ac DSG+FA
Manual #1 Read Count (log2)
Manual #2 Read Count (log2)
Automated #1 Read Count (log2)
Automated #2 Read Count (log2)
r = 0.975
r = 0.974
r = 0.947
r = 0.954
r = 0.970
r = 0.969
r = 0.947
r = 0.956
DSG+FA & FA only
H3K27ac Automated
H3K27ac Manual
FA only #1 Read Count (log2)
FA only #2 Read Count (log2)
DSG+FA #1 Read Count (log2)
DSG+FA #2 Read Count (log2)
r = 0.939
r = 0.946
r = 0.942
r = 0.944
r = 0.940
r = 0.947
r = 0.934
r = 0.936

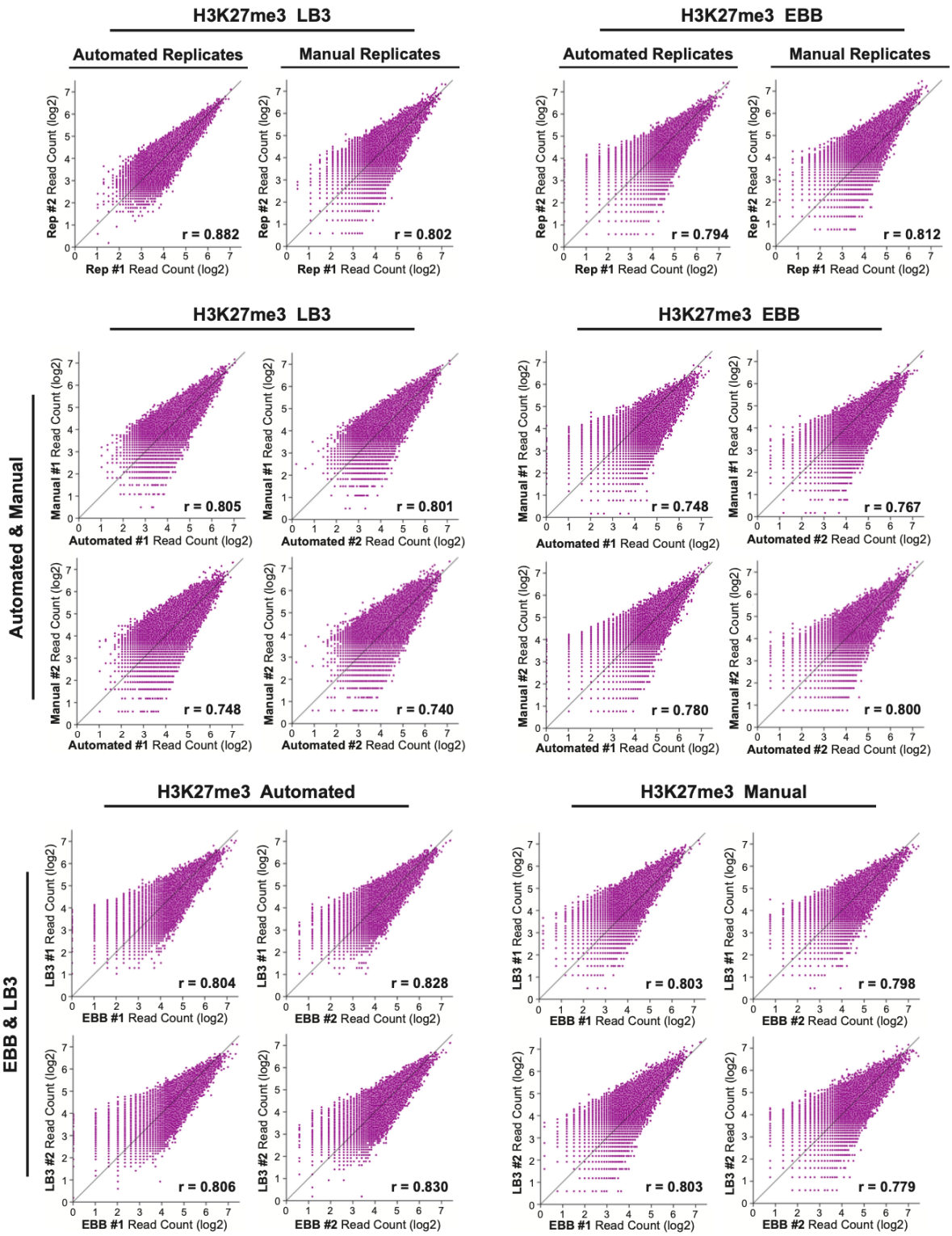
f.


H3K27me3 FA only
Automated Replicates
Manual Replicates
H3K27me3 DSG+FA
Automated Replicates
Manual Replicates
Rep #2 Read Count (log2)
Rep #1 Read Count (log2)
r = 0.771
r = 0.766
r = 0.754
r = 0.676
Automated & Manual
H3K27me3 FA only
H3K27me3 DSG+FA
Manual #1 Read Count (log2)
Manual #2 Read Count (log2)
Automated #1 Read Count (log2)
Automated #2 Read Count (log2)
r = 0.730
r = 0.736
r = 0.728
r = 0.723
r = 0.766
r = 0.776
r = 0.701
r = 0.705
DSG+FA & FA only
H3K27me3 Automated
H3K27me3 Manual
FA ony #1 Read Count (log2)
FA only #2 Read Count (log2)
DSG+FA #1 Read Count (log2)
DSG+FA #2 Read Count (log2)
r = 0.740
r = 0.727
r = 0.664
r = 0.612
r = 0.732
r = 0.711
r = 0.703
r = 0.649

g.


H3K27ac LB3
Automated Replicates
Manual Replicates
H3K27ac EBB
Automated Replicates
Manual Replicates
Rep #2 Read Count (log2)
Rep #1 Read Count (log2)
r = 0.993
r = 0.992
r = 0.984
r = 0.983
Automated & Manual
H3K27ac LB3
H3K27ac EBB
Manual #1 Read Count (log2)
Manual #2 Read Count (log2)
Automated #1 Read Count (log2)
Automated #2 Read Count (log2)
r = 0.984
r = 0.985
r = 0.989
r = 0.990
r = 0.972
r = 0.976
r = 0.967
r = 0.970
EBB & LB3
H3K27ac Automated
H3K27ac Manual
LB3 #1 Read Count (log2)
LB3 #2 Read Count (log2)
EBB #1 Read Count (log2)
EBB #2 Read Count (log2)
r = 0.979
r = 0.986
r = 0.979
r = 0.987
r = 0.951
r = 0.944
r = 0.956
r = 0.950

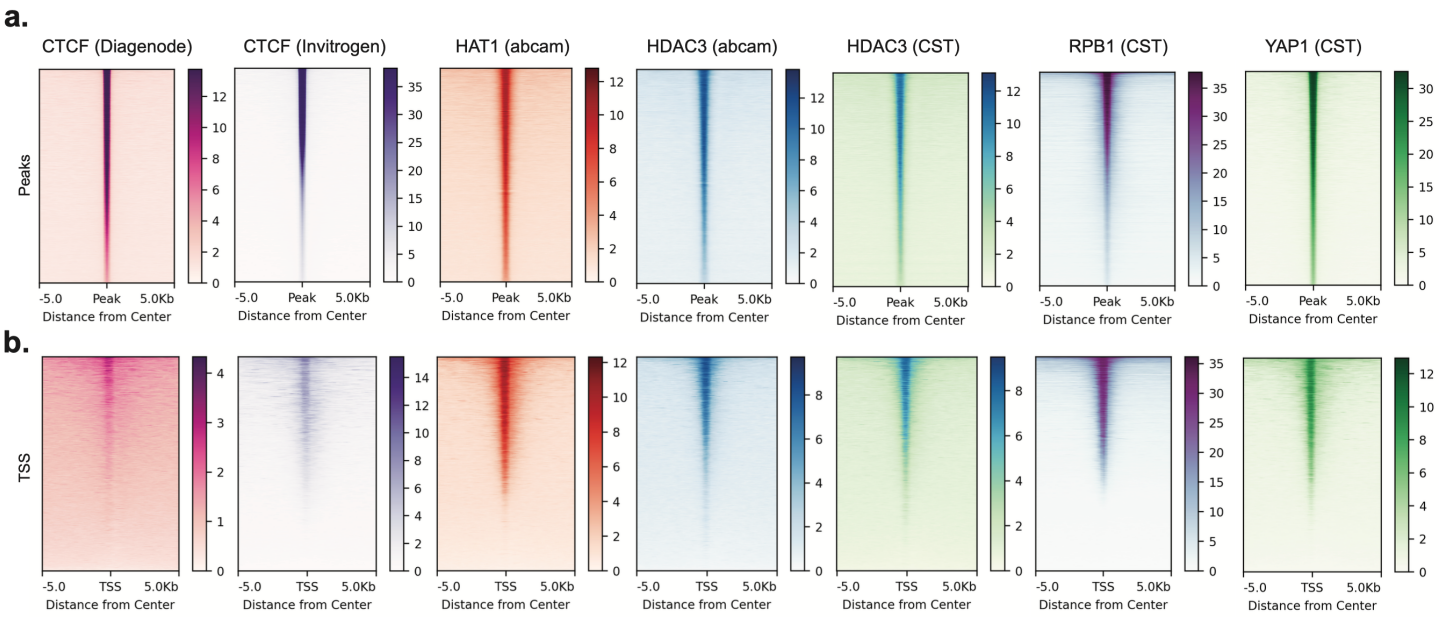
h.



Supplemental Fig. 2: Comprehensive data of our systematic evaluations of spa-ChIP-seq by comparing automated vs. manual workflows, crosslinking conditions and buffer compositions. a-b. IGV genome browser tracks (a) and heatmaps of ±5 kb from the center of TSSs (b) showing the automated and manual H3K27ac ChIP-seq results for comparing crosslinking conditions, DSG+FA and FA only, and buffer compositions, EBB and LB3. **c-d.** IGV genome browser tracks (c) and heatmaps of ±5 kb from the center of peaks (d) showing the automated and manual H3K27me3 ChIP-seq results for comparing crosslinking conditions, DSG+FA and FA only, and buffer compositions, EBB and LB3. **e.** Scatterplots of normalized read counts (log2) at shared H3K27ac peaks comparing automated replicates, manual replicates, automated and manual workflows for FA only and DSG+FA ChIP-seq samples, as well as comparing FA only and

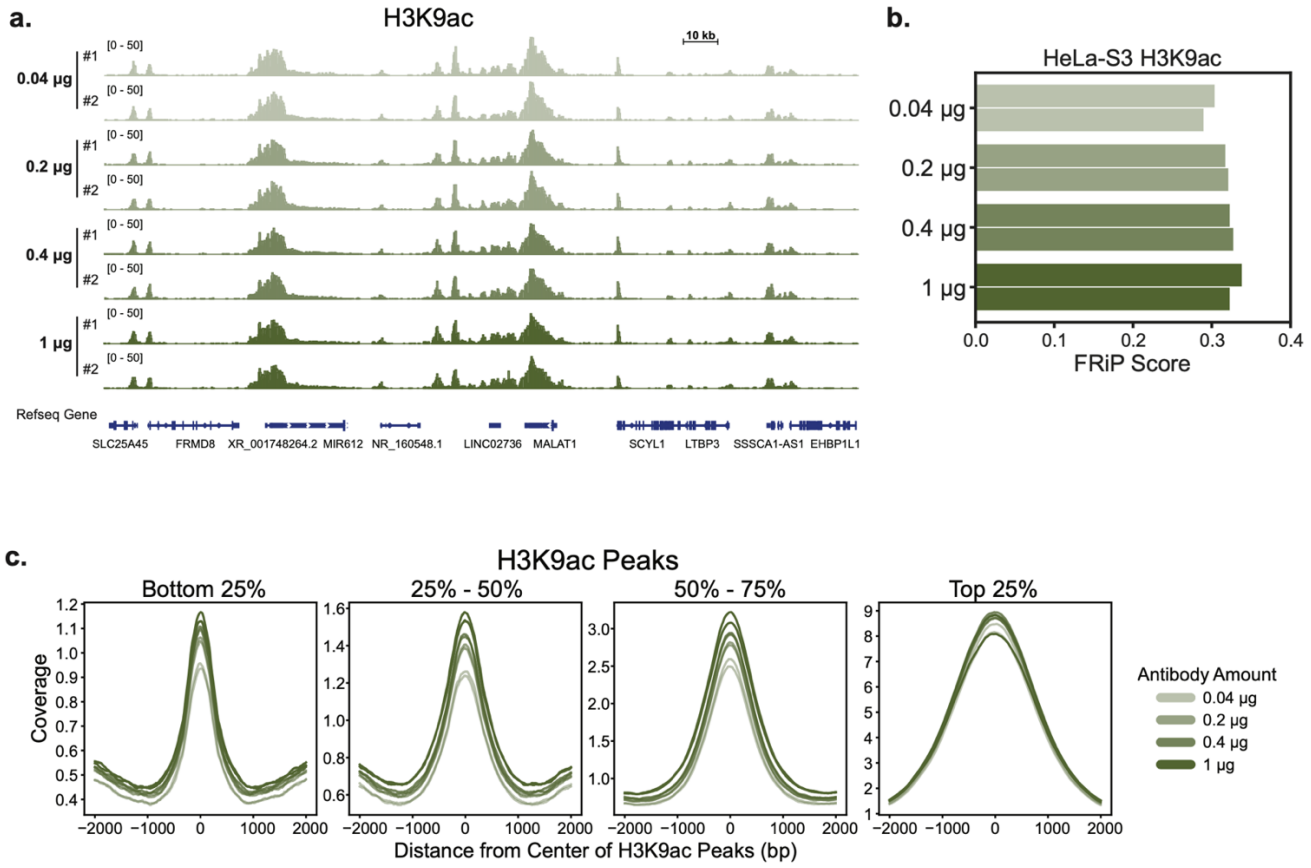
DSG+FA samples under automated and manual workflows. **f.** Scatterplots of normalized read counts (log2) at shared H3K27me3 peaks comparing automated replicates, manual replicates or automated vs. manual workflows for FA only and DSG+FA ChIP-seq samples, as well as comparing FA only and DSG+FA samples under automated vs. manual workflows. **g.** Scatterplots of normalized read counts (log2) at shared H3K27ac peaks comparing automated replicates, manual replicates, automated and manual workflows for LB3 and EBB ChIP-seq samples, as well as comparing LB3 and EBB samples under automated and manual workflows. **h.** Scatterplots of normalized read counts (log2) at shared H3K27me3 peaks comparing automated replicates, manual replicates, automated and manual workflows for LB3 and EBB ChIP-seq samples, as well as comparing LB3 and EBB samples under automated vs. manual workflows. The Pearson r correlation coefficients are reported on the lower right corner of the scatterplot.

Supplemental Fig. 3



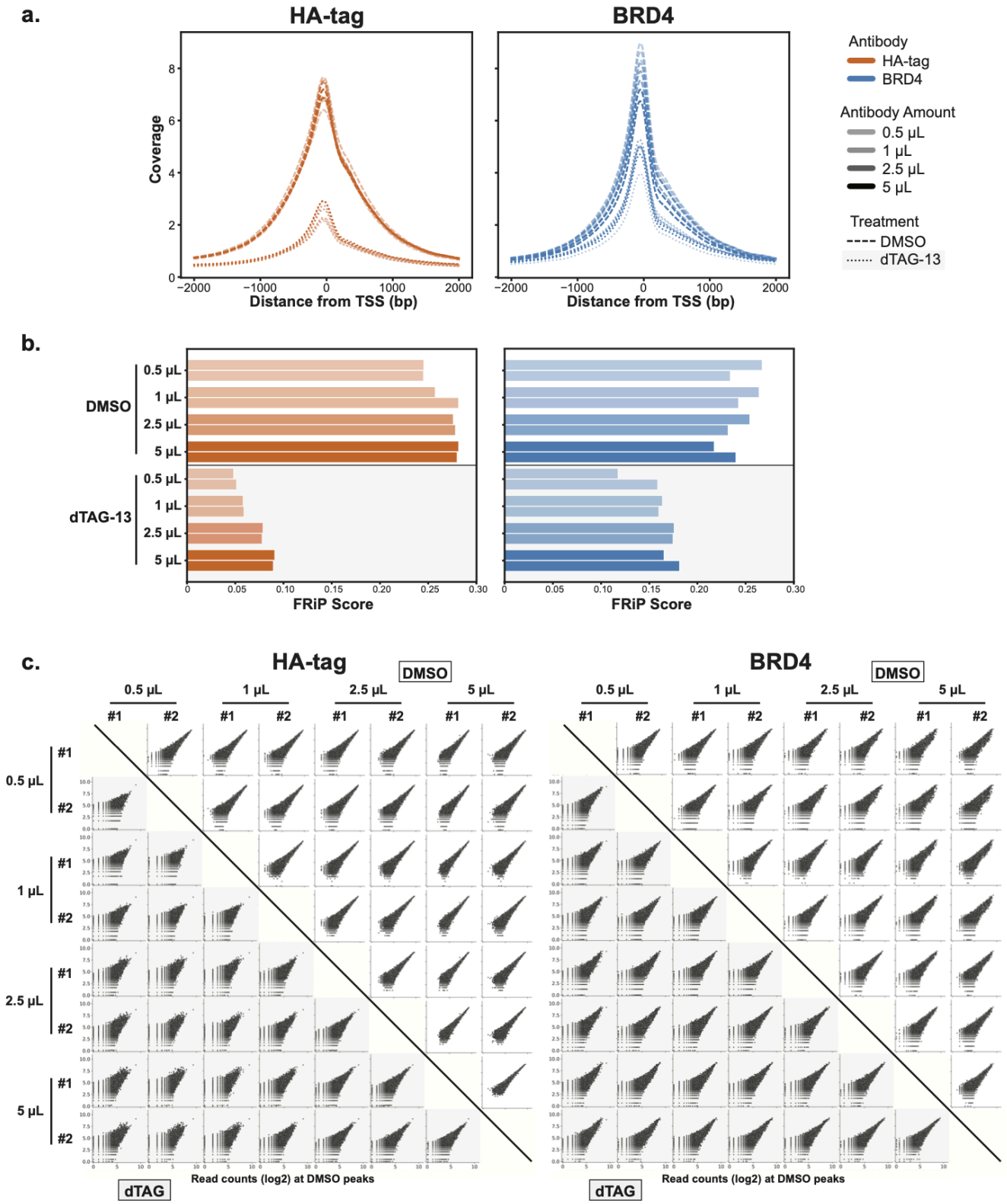
Supplemental Fig. 3: Extended data for mapping of DNA-associated proteins using spa-ChIP-seq. a-b. Heatmaps showing average normalized read density of the replicates at ±5 kb from the center of (a) peaks and (b) TSSs for ChIP-seq of DNA-associated proteins: CTCF, HAT1, HDAC3, Rpb1, and YAP1.

Supplemental Fig. 4



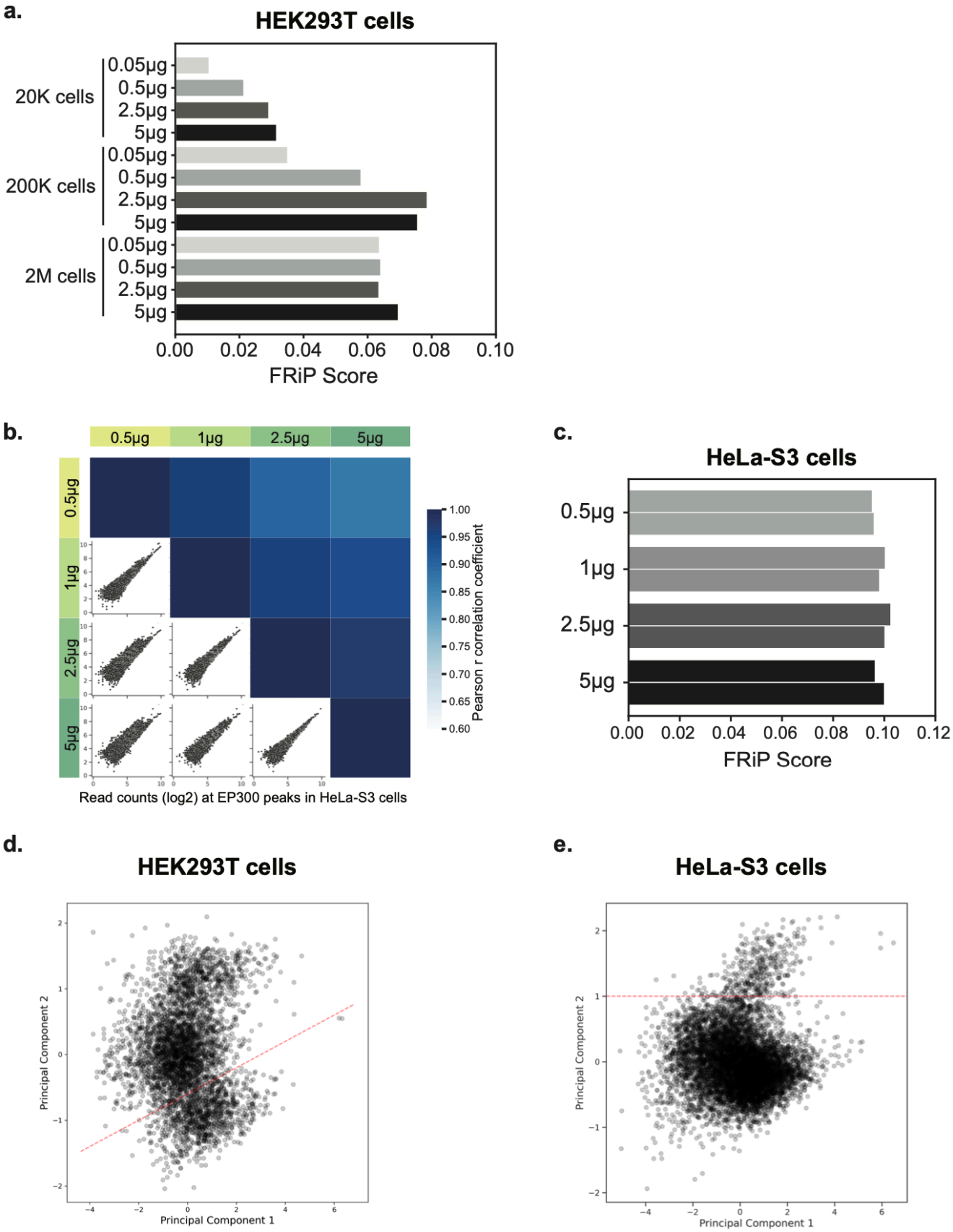
Supplemental Fig. 4: Extended data for H3K9ac antibody titration ChIP-seq experiments. a. IGV genome browser tracks for H3K9ac ChIP-seq done with various amounts of antibody (0.04 µg, 0.2 µg, 0.4 µg, 1 µg), each in replicates. **b.** Bar plot of FRiP scores for ChIP-seq of H3K9ac antibody titrations shown in (a). **c.** Read coverage histograms for H3K9ac antibody titration ChIP-seq signal centered on H3K9ac peaks, separated into quartiles based on the normalized read counts at the peaks.

Supplemental Fig. 5



Supplemental Fig. 5: Extended data for HA-tag and BRD4 antibody titration ChIP-seq experiments. a. Metagene plot of ±2000 bp from the center of TSS for BRD4 and HA-tag ChIP-seq in dTAG-13 and DMSO treated HEK293T-BRD4-FKBP12^{F36V}-2xHA cells with 0.5 µL, 1 µL, 2.5 µL, and 5 µL of HA-tag and BRD4 antibodies. **b.** Bar plot of FRiP scores for ChIP-seq of HA-tag and BRD4 antibody titrations shown in (a). **c.** Scatterplots showing normalized read counts (log2) at DMSO peaks for HA-tag and BRD4 antibody titration. The DMSO samples are in the upper right corner and dTAG samples are in the lower left corner shaded with a light gray background.

Supplemental Fig. 6



Supplemental Fig. 6: Extended data for EP300 antibody titration ChIP-seq experiments in HeLa-S3 and HEK293T cells. a. Bar plot of FRiP scores using the peaks called in each sample for ChIP-seq of EP300 antibody titrations with 0.05 µg, 0.5 ug, 2.5 ug, and 5

ug of antibody and 20K, 200K, and 2M of HEK293T cells. **b.** A matrix of scatterplots showing normalized read counts (log2) at EP300 peaks and heatmaps showing Pearson r correlation coefficients for EP300 antibody titrations in 1M HeLa-S3 cells with various antibody amounts 0.05 µg, 0.5 µg, 2.5 µg, and 5 µg of EP300 antibody (green shades). **c.** Bar plot of FRiP scores using the merged peaks of the replicates for each condition for ChIP-seq of EP300 antibody titrations with 0.05 µg, 0.5 µg, 2.5 µg, and 5 µg of EP300 and 1M of HeLa-S3 cells. Each condition is done in replicates represented as two bars with the same shading. **c-d.** Principal component analysis (PCA) on the normalized read counts (log2) at the top 10% of EP300 ChIP-seq peaks for 0.5 µg and 5 µg of antibody in 2M HEK293T cells (c) and 1M HeLa-S3 cells (d). The red dashed line shows how the red subset of peaks in **Fig. 4c** is isolated for each of the samples (the region under the line in d and the region above the line in e).