TAF-ChIP: An ultra-low input approach for genome wide chromatin immunoprecipitation assay

Junaid Akhtar*[1], Piyush More[2], Apurva Kulkarni[1,4], Federico Marini[3], Waldemar Kaiser[1], and Christian Berger*[1]

[1] Institute for Developmental Biology and Neurobiology, University of Mainz, Johannes-Joachim-Becherweg 32, 55099 Mainz, Germany

[2] Department of Pharmacology, University Medical Center, University of Mainz, 55131 Mainz, Germany

[3] Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), 55101 Mainz, Germany

[4] present address: Indian Institute of Science Education and Research (IISER) Pune, Maharashtra 411008, India

*Corresponding authors: akhtar@uni-mainz.de and bergerc@uni-mainz.de

Abstract

Chromatin immunoprecipitation (ChIP) followed by next generation sequencing is an invaluable and powerful technique to understand transcriptional regulation. However, ChIP is currently limited by the requirement of large amount of starting material. This renders studying rare cell populations very challenging, or even impossible. Here, we present a tagmentation-assisted fragmentation ChIP (TAF-ChIP) and sequencing method to generate high-quality datasets from low cell numbers. The method relies on Tn5 transposon activity to fragment the chromatin that is immunoprecipitated, thus circumventing the need for sonication or MNAse digestion to fragment. Furthermore, Tn5 adds the sequencing adaptors while fragmenting the chromatin resulting in one-step generation of PCR ready library, which makes it an extremely simple approach with minimum hands-on time. The method can be directly implemented on sorted cells, and does not require de-multiplexing strategies to generate the profile from limited cell numbers. This can be very useful when the access to the starting material is restricted. Using our approach we generated the H3K4Me3 and H3K9Me3 profiles from 100 K562 cells and 1000 sorted neural stem cells (NSC) from *Drosophila.* We benchmarked our TAF-ChIP datasets from K562 cells against the ENCODE datasets. For validating the TAF-ChIP datasets obtained from *Drosophila* NSCs we performed conventional ChIP-Seq experiments in identical cell types. The epigenetic profiles obtained from both conventional ChIP-Seq and TAF-ChIP showed high degree of agreement, thereby underlining the utility of this approach for generating ChIP-Seq profiles from very low cell numbers.

Introduction

Chromatin immunoprecipitation coupled with next generation sequencing is a powerful and unbiased approach to study genome-wide DNA-protein interactions and epigenetic modifications [1]. However, the prerequisite of huge starting material (millions of cells) limits its utility in studying rare cell types [2]. There are few major limitations in implementation of this method to rare cell types; first, sonication that is by far the most popular method for fragmentation in ChIP-Seq experiments can destroy the epitope used for immunoprecipitation especially when the material is limited [3]. The alternative approach of micrococcal nuclease based digestion

(MNase) is harder to control in its efficacy and saturation, and shows some degree of sequence dependent biases [4, 5, 6]. Second, the addition of sequencing adaptors and the generation of final libraries involves steps where the limitation of ligation and loss of material during purification steps can result in a low-complexity libraries. In spite of this, there have been several attempts to scale-down ChIP-Seq protocols to address these limitations and to implement it for rare cell types [7 , 8]. One such method is FARP-ChIP [9]. In this method the cells are co-sonicated with bacterial cells, used for protection during sonication. To prevent the loss of DNA during library preparation, a biotinylated synthetic DNA (biotin-DNA) is used as a carrier DNA. This approach was successfully implemented for 500 mESC cells [9], however, a reasonable coverage requires deep sequencing runs (approx. 100 million reads) and the number of reads mapping to the DNA of the target cell type was very low (~16%), which makes this method less feasible for many applications and cost intensive. Another recently applied method is Mint-ChIP that uses prior ligation of barcoded adaptors to the chromatin digested by MNAse, and then using *in silico* demutliplexing strategy to obtain the profiles from low cell numbers [10]. The barcoding strategy does dramatically reduce the number of cells required for each profile however the method still initially requires 10000-100000 cells as starting material. Another method called ChIPmentation uses Tn5 transposon mediated tagmentation for preparation of libraries as an alternative to the ligation based library preparation methods [8]. However, the approach uses sonication for fragmenting the chromatin prior to immunoprecipitation. Moreover, this method still employs a larger batch preparation of chromatin, and uses subsequent splitting of the sample to generate the profile from 10000 cells. Here we describe an alternative approach for ChIP with tagmentation-assisted fragmentation of chromatin (TAF-ChIP) with hyperactive Tn5 transposase from Illumina, completely circumventing the need of sonication for fragmentation. We have used this approach to generate high quality datasets from as few as 100 human and 1000 *Drosophila* cells. This approach does not involve labor-intensive library preparation workflow and is readily applicable to any type of cells. Furthermore, so far there have been no direct comparisons between low amount ChIP-Seq methods and conventional ChIP-Seq approach. We have used ENCODE datasets and conventional ChIP-Seq performed in identical cell types to benchmark our TAF-ChIP datasets, and to demonstrate the utility of this approach. We expect our

approach to be very useful in conditions where the amount of sample is frequently the limiting factor, such as material isolated from animals or clinical samples.

Results

Method Overview

There are two critical steps in generating high quality ChIP-Seq datasets from very low number of cells. First, the chromatin should be fragmented without compromising the integrity of the associated proteins. This becomes far more challenging when the starting point of the protocol is low cell number samples. Second, the purified DNA needs to undergo multiple manipulation steps, namely end-repair, ligation of the sequencing adaptors, and PCR amplification to generate Illumina compatible sequencing libraries. In these steps, there is also requirement of beads-based purification of non-amplified DNA. Any potential loss at this step, especially when the amount of DNA is in picogram (pg) range, can severely compromise the completion of successful libraries. To overcome these limitations, we employed tagmentation as a tool to fragment the DNA. Although, tagmentation has been previously used for preparation of ChIP libraries and genomic DNA libraries. However, it has been used so far at the very last step for addition of sequencing adaptors. Here, we have used Tn5 activity to fragment the intact chromatin post immunoprecipitation (Figure 1). There is two major advantage of this approach. First, there is no need of sonication to fragment the chromatin. Therefore this strategy prevents potential loss of DNA-protein interactions during sonication, especially when starting with low cell numbers. Furthermore, sonication is extremely variable between different machines of same specifications and depending on different cycle of their usage. Second, the tagmentation reactions employ the hyperactive Tn5 transposomes that are preloaded with sequencing adaptors [11]. This results in one-step PCR compatible DNA library generation with overcoming the limitation of ligation efficiency [12] and avoiding intermediate bead purification steps, thus preventing potential loss of material. After PCR amplification the amplified libraries were beads-purified to generate the final libraries (Supplementary Figure 1A). The tagmentation reaction is also easier to control compared to MNAse, and resulted in highly reproducible libraries when compared between replicates (Supplementary

Figure 1B, 1C, 1D, and 1E). The method was also robust enough across different amount of starting materials.

For TAF-ChIP samples, the cells were directly sorted into RIPA buffer owing to the low FACS sheath fluid volume, and directly proceeded to nuclear lysis with low energy sonication. Prior to immunoprecipitation the antibody and the magnetic beads were first coupled in the presence of blocking buffer containing BSA, glycogen and tRNA. This step is critical to prevent non-specific binding of DNA to the antibody-coupled beads. The unfragmented genome was then immunoprecipitated with the addition of blocked antibody coupled beads overnight at 4°C. After initial wash with home made tagmentation buffer, the immunoprecipitated chromatin was then tagmented at 37°C to fragment concomitant with addition of sequencing adaptors. Tagmentation enzymes as well as background regions were washed away with subsequent high stringency washes. DNA was purified and PCR amplified to generate Illumina compatible DNA libraries (see methods section for further details) (Figure 1).

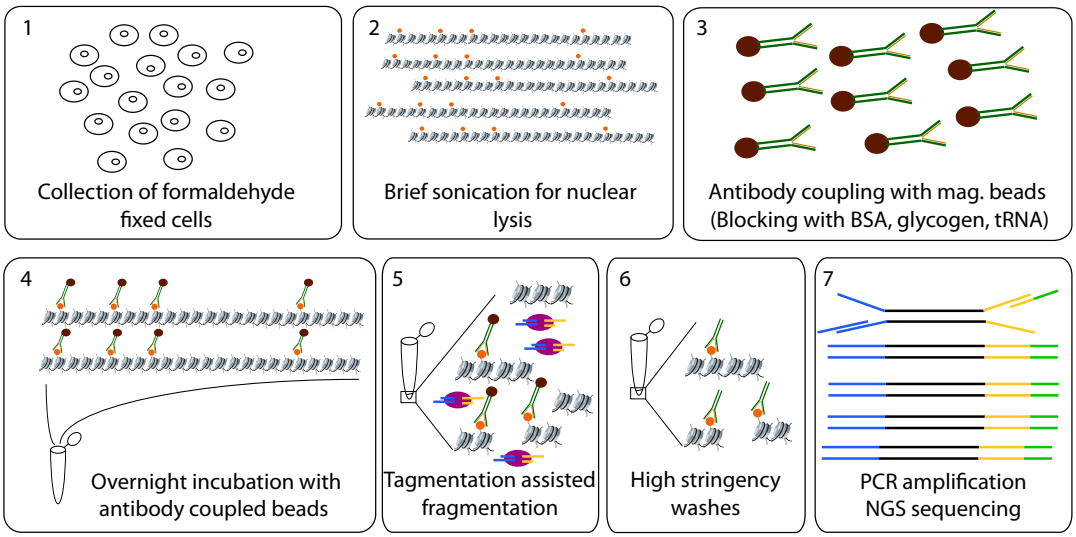

Figure 1. Schematic overview of TAF-ChIP (1) Formaldehyde fixed cells were directly sorted into RIPA buffer (see methods section for details). **(2)** Cells were briefly sonicated at low intensity to break open the nuclei. **(3)** Antibodies were coupled to magnetic beads in presence of blocking reagents. **(4)** Cleared cell lysate were added to the antibody coupled beads and incubated overnight at 4°C. **(5)** Tagmentation reaction was performed after initial washes with home made tagmentation buffer. **(6)** Tagmentation reaction, and the background regions (not anchored by antibody interaction) was washed away with subsequent high stringency washes. **(7)** The DNA was isolated after reverse cross-linking and PCR amplified to generate Illumina compatible libraries.

The conventional ChIP-Seq samples from *Drosophila* were sorted, pelleted and resuspended in RIPA buffer, as described earlier [13]. Upon immunoprecipitation with specific antibodies the DNA was extracted and converted into DNA libraries.

For the purpose of this study, we used two different type of starting material. The type II neural stem cells (NSCs) from *Drosophila* larval brain and human K562 cells, a human immortalised myelogenous leukemia line. We used formaldehyde to fix freshly dissected *Drosophila* larval brains or harvested K562 cells. The dissected *Drosophila* larval brains expressed a GFP that was used to sort NSCs from a primary cell suspension as described earlier [14]. Since *Drosophila* brains consist of approximately 400 NSCs only, FACS-sorting of wild type NSCs is not applicable to obtain 1 million cells necessary to generate a conventional ChIP-Seq dataset. Thus, in order to compare the TAF-ChIP dataset with the conventional ChIP-Seq dataset, we took advantage of a UAS line expressing a constitutively active Notch protein ($Notch^{intra}$) under the control of a type II NSCs specific binary expression system (UAS/GAL4 system; *wor*-Gal4; *ase*-Gal80 fly line), also expressing UAS-CD8-GFP [15]. The expression of constitutively active $Notch^{intra}$ protein in type II NSCs results in a massive over proliferation and thousands of supernumerous type II NSCs, making it amenable to sort cells for conventional ChIP-Seq [16, 17]. We sorted type II NSCs from this line with identical settings, for TAF-ChIP (1000 cells) as well as conventional ChIP-Seq (1.2 million cells). For obtaining 100 K562 cells, we stained the cells with Hoechst dye and used FACS for collecting samples with the precise number of cells. To benchmark our TAF-ChIP data sets from K562 cells we used publicly available dataset from ENCODE project [18].

TAF-ChIP performed with *Drosophila* NSCs shows high degree of agreement with conventional ChIP-Seq

For validating TAF-ChIP results we also generated conventional ChIP-Seq datasets for H3K4Me3 as well as H3K9Me3, as described above, from identical cell types. All of the TAF-ChIP generated datasets showed high degree of agreement with corresponding conventional ChIP-Seq datasets, as visualized through genome browser tracks (Figure 2 A and 2B). The TAF-ChIP data also showed high degree of mappability with low level of sequence duplication, and good concordance between the replicates (Supplementary Figure 1B, 1C and 1F). Furthermore, the MACS2 analysis for saturation of peak recovery showed higher recovery of peaks for

H3K4Me3 at shallow sequencing depth where as for K9Me3 the number of peaks continued to increase with increasing sequencing depths (Supplementary Figure 1G). This was consistent with the observed tendencies for point-source histone modifications (such as H3K4Me3) and histone modifications with broad domains of enrichments (such as H3K9Me3) [19]. We next calculated the binding matrix with scores based on read counts in every sample, using the DiffBind software package [20]. The correlation heatmap generated with this method showed that all the

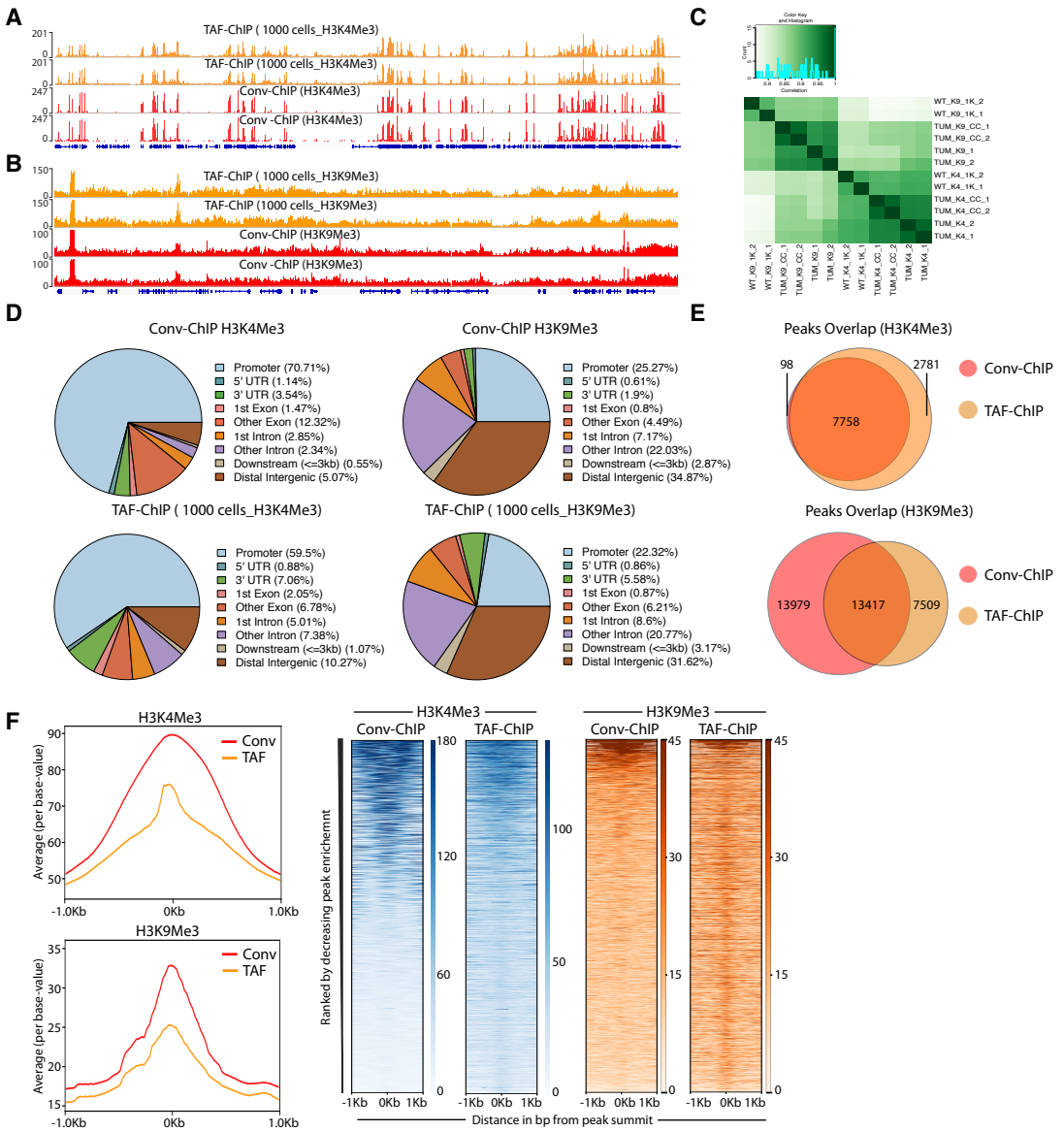

Figure 2. TAF-ChIP results from low number of NSCc are comparable to conventional ChIP-Seq **(A)** Genome browser track example of H3K4Me3 ChIP performed in FACS sorted NSCs with conventional ChIP-Seq (1.2 million cells) and TAF-ChIP (1000 cells), as indicated in the labels. The label below the tracks shows the gene model, and the y-axis represents normalized read density in reads per million (rpm). **(B)** Genome browser track example of H3K9Me3 ChIP performed in FACS

sorted NSCs with conventional ChIP-Seq (1.2 million cells) and TAF-ChIP (1000 cells), as indicated in the labels. The label below the tracks shows the gene model, and the y-axis represents normalized read density in reads per million (rpm). **(C)** The correlation heatmap showing binding matrix based on read counts. The replicates derived from TAF-ChIP cluster together, and also cluster with the corresponding conventional ChIP-Seq dataset. **(D)** Pie chart showing the distribution of annotated peaks obtained from conventional ChIP-Seq approach and TAF-ChIP approach, for indicated histone marks. Note the majority of H3K4Me3 peaks are at the promoters and H3K9Me3 at the intergenic regions, consistent with the expectation. **(E)** The overlap between the peaks identified from conventional ChIP-Seq and TAF-ChIP datasets, for the indicated histone modifications. MACS2 with identical parameters was used to identify the peaks and those present in both the replicates were further considered for the overlap. **(F)** Average profile of TAF-ChIP and conventional ChIP-Seq centered at the peaks for the indicated histone modifications. The Y-axis depicts average per base-value into the peaks while X-axis depicts genomic coordinates centered at the peaks. **(G)** Heatmaps of indicated histone modifications from conventional ChIP-Seq and TAF-ChIP method, centered at the peaks (-1kb to +1kb). Rows indicate all the peaks and are sorted by decreasing affinities in the conventional ChIP-Seq data sets. The color labels to the right indicate the level of enrichment.

replicates belonging to the same condition cluster together, furthermore TAF-ChIP datasets also clustered with conventional ChIP-Seq datasets (Figure 2C). The annotation of peaks obtained from TAF-ChIP for H3K4Me3 and H3K9Me3 marks also showed that their overlapping genomic features were very similar to the ones obtained from conventional ChIP-Seq (Figure 2D). Furthermore, consistent with the expectation the large fraction of H3K4Me3 peaks was at the promoters where as the majority of H3K9Me3 peaks were at the distal intergenic regions. The peaks called for conventional ChIP-Seq and TAF-ChIP datasets, using the MACS2 program [21], revealed a high degree of overlap between them, both for H3K4Me3 peaks and H3K9Me3 peaks (Figure 2E). The read distribution in the peaks was comparable between TAF-ChIP and conventional ChIP-Seq datasets (Figure 2F). The heatmaps generated for all the peaks identified in conventional ChIP-Seq datasets and sorted according to the intensity in the conventional ChIP-Seq, resulted in similar and comparable profile for TAF-ChIP datasets (Figure 2G). Finally, the distribution of the distances of the peaks relative to the TSS (transcriptional starting site) was as previously reported, both for H3K4Me3 and H3K9Me3 modifications (Supplementary Figure 2A and 2B) [18].

TAF-ChIP is robust across different numbers of cells used as starting material

After establishing TAF-ChIP on low number of cells and its subsequent benchmarking against conventional ChIP-Seq performed in identical cells, we next assayed whether TAF-ChIP can give comparable results with similar resolution, when variable numbers of cells are used as starting material. Towards this goal we sorted to wild type NSCs, also to ensure that the method itself is applicable in normal physiological context. We sorted two samples containing 1000 and 5000 NSCs from dissected *Drosophila* larval brains expressing a GFP-tagged Deadpan (Dpn) protein under the control of its endogenous enhancer, which is a transcription factor only present in neural stem cells in the brain [22]. The TAF-ChIP generated datasets from 1000 and 5000 NSCs resulted in identical profiles, as visualized through genome browser tracks (Figure 3A). The heatmaps generated from 1000 NSCs and 5000 NSCs were also very similar (Figure 3B). The read distribution in the peaks for samples with 1000 NSCs and 5000 NSCs were also comparable (Figure 3C). Altogether these results highlight the robustness of the TAF-ChIP method when applied in the context of variable number of cells as starting material.

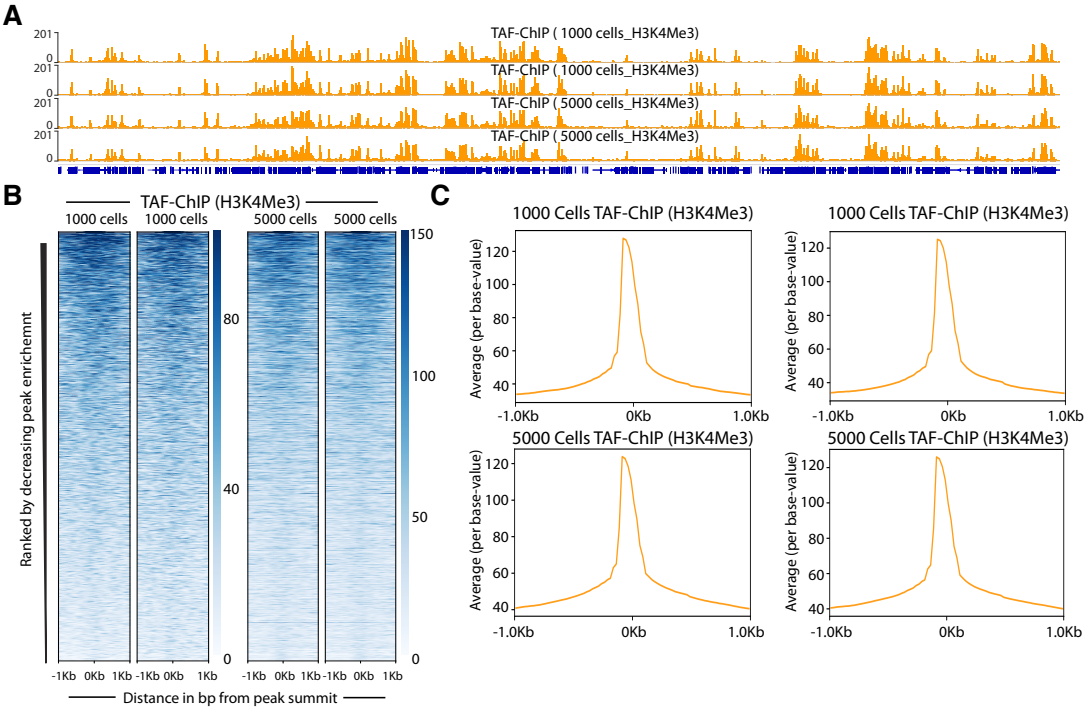

Figure 3. TAF-ChIP results are robust across different amount of starting material (number of cells) (A) Genome browser track example of H3K4Me3 ChIP performed in 1000 and 5000 FACS sorted NSCs cells with TAF-ChIP, as indicated in the labels. The label below the tracks shows the gene model, and the y-axis represents normalized read density in reads per million (rpm). **(B)** Heatmaps of

TAF-ChIP data from indicated number of NSCs, centered at the peaks (-1kb to +1kb). Rows indicate all the peaks and are sorted by decreasing affinities in the TAF-ChIP data from 5000 NSCs. The color labels to the right indicate the level of enrichment. **(C)** Average profile of TAF-ChIP data from 1000 and 5000 NSCs, centered at the peaks for the H3K4Me3 modification. The Y-axis depicts average per base-value into the peaks while X-axis depicts genomic coordinates centered at the peaks.

TAF-ChIP is widely applicable and can generate high quality dataset from different cell types

Encouraged by our results of investigating the epigenetic profile in *Drosophila*, we next wanted to apply our method in a different cell type to test its universal applicability. Towards this end, we used human K562 cells since conventional ChIP-Seq datasets from the ENCODE project can be used for benchmarking. With approximately 3234 million bases the human genome is roughly 15 times larger than the *Drosophila* genome. Reasoning this we decided to perform our TAF-ChIP approach for H3K4Me3 and H3K9Me3 modifications from 100 sorted human K562 cells. The H3K4Me3 TAF-ChIP and H3K9Me3 TAF-ChIP from 100 cells showed similar profiles when compared with the ENCODE datasets, as visualized through genome browser tracks (Figure 4A and 4B), and also showed good agreement between the replicates (Supplementary Figure 1D and 1E). The heatmap showing binding matrix with scores based on read counts in every sample also resulted in all the replicates belonging to same TAF-ChIP experiment clustering together. Furthermore, TAF-ChIP datasets also clustered with corresponding ENCODE ChIP-Seq datasets with starting material larger by several orders (Figure 4C). The annotation of peaks identified in TAF-ChIP dataset and the corresponding one from ENCODE showed similarity in distribution of overlapping genomic features, both for H3K4Me3 and H3K9Me3 datasets (Figure 4D). The peaks called for ENCODE ChIP-Seq datasets and TAF-ChIP datasets, using the MACS2 program with identical parameters, also have reasonable degree of overlap between them (Figure 4E). This highlights the utility of our method in identifying the bound genomic regions even when the starting material is significantly limited. The fraction of reads in peaks called with the TAF-ChIP dataset have also similar distribution profile when compared to the ENCODE ChIP-Seq datasets, however the level of this distribution was smaller in case of TAF-ChIP (Figure 4F). Nonetheless, the heatmaps generated for all the peaks identified in ENCODE ChIP-Seq datasets and sorted according to the

intensity in the ENCODE ChIP-Seq, showed similar and comparable profile for TAF-ChIP datasets from 100 K562 cells (Figure 4G). The peaks obtained from TAF-ChIP for H3K4Me3 and H3K9Me3 marks also showed that their distribution with respect to the distance from the TSS was as expected, with H3K4Me3 predominantly present closer and H3K9Me3 mostly distant from the TSS. This observation was also consistent with the one obtained from ENCODE ChIP-Seq datasets (Supplementary Figure 2C and 2D).

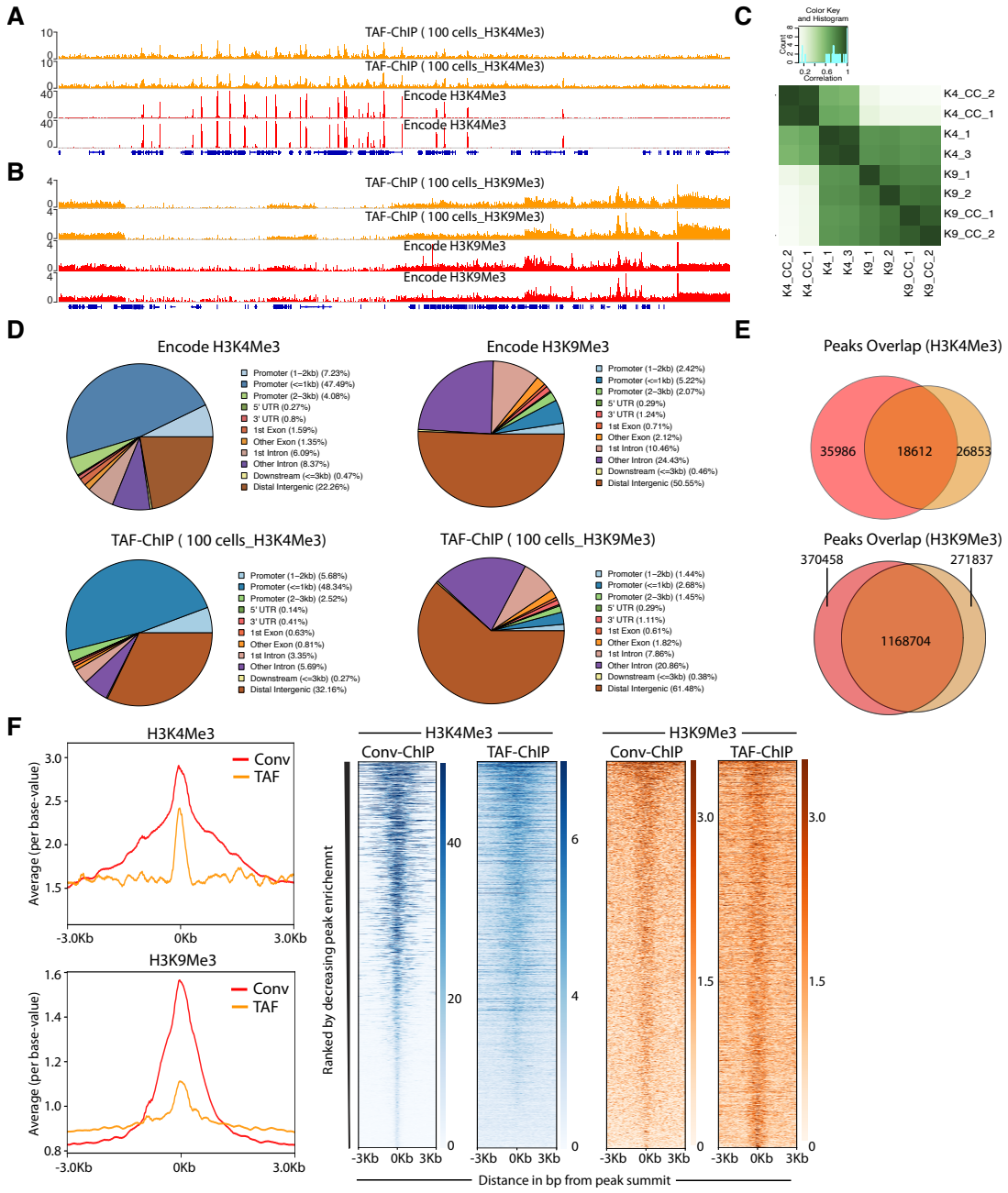

Figure 4. TAF-ChIP results from 100 K562 cells are comparable to ENCODE ChIP-Seq datasets **(A)** Genome browser track examples of H3K4Me3 ChIP performed in 100 FACS sorted K562 cells

with TAF-ChIP approach and corresponding K562 conventional ChIP-Seq datasets from ENCODE project, as indicated in the labels. The label below the tracks shows the gene model, and the y-axis represents normalized read density in reads per million (rpm). **(B)** Genome browser track examples of H3K4Me3 ChIP performed in 100 FACS sorted K562 with TAF-ChIP approach and corresponding conventional ChIP-Seq datasets from ENCODE project, as indicated in the labels. The label below the tracks shows the gene model, and the y-axis represents normalized read density in reads per million (rpm). **(C)** The correlation heatmap showing binding matrix based on read counts. The replicates derived from TAF-ChIP (100 K562 cells) cluster together, and also cluster with the corresponding ENCODE ChIP-Seq datasets. **(D)** Pie chart showing the distribution of annotated peaks obtained from ENCODE datasets and TAF-ChIP approach (100 K562 cells), for indicated histone marks. Note the majority of H3K4Me3 peaks are at the promoters and H3K9Me3 at the intergenic regions, consistent with the expectation. **(E)** The overlap between the peaks identified from ENCODE datasets and TAF-ChIP datasets, for the indicated histone modifications. MACS2 with identical parameters was used to identify the peaks and those present in both the replicates were further considered for the overlap. **(F)** Average profile of TAF-ChIP and corresponding ENCODE ChIP-Seq centered at the peaks for the indicated histone modifications. The Y-axis depicts average per base-value into the peaks while X-axis depicts genomic coordinates centered at the peaks. **(G)** Heatmaps of indicated histone modifications from ENCODE ChIP-Seq and TAF-ChIP method, centered at the peaks (-3kb to +3kb). Rows indicate all the peaks and are sorted by decreasing affinities in the ENCODE ChIP-Seq data sets. The color labels to the right indicate the level of enrichment.

Discussion

Here, we have presented a user-friendly, tagmentation-assisted fragmentation ChIP (TAF-ChIP) and sequencing method to generate high-quality datasets from low cell numbers. Compared to other protocols TAF-ChIP truely starts from ultra-low primary material (100 human cells), right from the initial collection of the samples until the final steps. The workflow of TAF-ChIP is extremely simple with minimum hands-on-time during library preparation preventing loss of material and potential user introduced variability. It can be applied to both fixed and unfixed cells, and thus can be used in different experimental settings. Unlike MNase based approach TAF-ChIP is easier to implement and does not lead to over digestion of chromatin, and results in one step generation of Illumina compatible libraries. Also, the tagmentation does not show any sequence dependent biases unlike other restriction based protocols. Furthermore, the approach does not need any specialized equipment and thus can be implemented in a standard molecular biology lab. We have used here the Tn5 transposase from Nextera XT DNA library kit, however it could be easily implemented with Tn5 loaded with different unique molecular indices [11]. This

could result in massively parallel ChIP-Seq applications, which may even further decrease the required staring material. Moreover, as this approach is amenable for various cell types it could be also implemented with a non-target cell type used as "spike-in" and carrier DNA.

The analysis of the TAF-ChIP datasets showed robustness across replicates and identified identical binding regions when compared to conventional ChIP-Seq approaches with larger amounts of starting material by several orders of magnitude. The peaks identified by sharp peak calling algorithms, such as MACS2, identified largely overlapping peaks in TAF-ChIP datasets when compared to sonication based conventional ChIP-Seq approaches. Although, the genome browser profiles obtained from 100 K562 cells showed slightly inferior signal to noise ratio when compared to the conventional datasets from the ENCODE project yet the peaks identified were mostly overlapping. Inspite of this the TAF-ChIP datasets perform reasonably well across several parameters, when compared to conventional datasets generated from millions of cells. Furthermore, we suspect that the signal to noise ratio can be improved by pooling the samples tagmented with different indices, prior to washes and following the demultiplexing strategy to obtain the data. The only limitation of the TAF-ChIP we envision is availability of a good antibody and reasonable number of binding sites in the genome, in order to generate a ChIP-Seq profile from low amount of cells. Given that our TAF-ChIP approach directly starts from as low as 100 isolated cells without requiring prior isolation of nuclei and with an extremely easy and straightforward workflow, it will be very useful when access to higher number of cells is not possible.

Author Contributions

J.A and C.B designed the experiments. J.A performed the experiments and analyzed the data, P.M and F.M helped in analysis of the data. A.K and W.K provided the *Drosophila* NSC samples. J.A and C.B wrote the paper.

Acknowledgments

We thank the Bloomington *Drosophila* Stock Center for the fly lines used in this study. The EMBL Genomics Core Facility (Heidelberg, Germany), especially Vladimir Benes for all the sequencing runs. We thank Jean-Yves Roignant, Prof. Leszek Wojnowski, Guillaume Junion, Joachim Urban, and Prasad Chitke for critically reading the manuscript. Research in the laboratory of C.B is supported by the Deutsche Forschungsgemeinschaft (DFG) and the Univeristy of Mainz. P.M. and A.K are supported by the International PhD Programme (IPP), Institute of Molecular Biology, Mainz.

Accession number

All the ChIP-Seq data used in this study is submitted to the GEO database under the accession number GSE112633.

Disclosure Declaration

The authors declare no conflict of interests.

Methods

Fly lines: GAL4 line expressing GFP and Notch^{intra} in type II NSCs: *wor*-GAL4, *ase*-GAL80; UAS-CD-*gfp*, UAS-N^{intra}, DPN-GFP has been generated in the Berger lab by insertion of GFP into the DPN gene locus.

Antibodies: The following antibodies were used in this study. For H3K4Me3 ChIP, antibody from Abcam (Cat No-ab8580) was used. H3K9Me3 ChIP was performed using an antibody from Active Motif (Cat No- 39161).

Fixation and cell sorting from *Drosophila* larval brain: Briefly: required number of larval brains after 48h of larval hatching were dissected in PBS. After dissection, larval brains were fixed with 1% formaldehyde in PBS for 10 minutes at room temperature, followed by quenching of the fix with 125mM glycine. The larval brains were dissociated and resuspended according to the previously established method [14]. The cells were then sorted on BD FACSAria™ according to the strength of GFP and size of the NSCs, resulting in a pure population of neural stem cells.

Fixation and cell sorting of K562 cells: K562 cells cultured in RPMI medium (supplemented with 10% Fetal Bovine Serum), at 37°C and 5% CO_2 were fixed for 10 minutes at room temperature with 1% formaldehyde. The crosslink was quenched with 125 mM gylcine, and sorted on BD FACSAria™ cell sorter using Hoechst stain. 100 K562 cells were directly sorted in RIPA 140mM (10mM Tris-Cl pH 8.0, 140mM NaCl, 0.5mM EDTA pH 8.0, 1% Triton-X 100 and 0.1% SDS).

Conventional ChIP-Seq and library preparation: Fixed cells (1.2 million FACS sorted NSCs per replicate) were resuspended in 140mM RIPA (10mM Tris-Cl pH 8.0, 140mM NaCl, 0.1mM EDTA pH8.0, 1% Triton X-100, and 0.1% SDS) and subjected to 14 cycles of sonication on a bioruptor (Diagnode), with 30 Secs "ON"/ "OFF" at high settings. After sonication, samples were centrifuged at 14,000 g for 10 minutes at 4°C and supernatant was transferred to a fresh tube. The extracts were incubated overnight with 2 µg of specific antibody at 4°C with head-over tail rotations. After overnight incubations 20 µl of blocked Protein A and G Dynabeads were added to the tubes and further incubated for 3 hours to capture the antibodies. The beads were separated with a magnetic rack and were washed as following; once with 140mM

RIPA (10mM Tris-Cl pH 8.0, 140mM NaCl, 0.1mM EDTA pH8.0, 1% Triton X-100, and 0.1% SDS), four times with 250mM RIPA (10mM Tris-Cl pH 8.0, 250mM NaCl, 0.1mM EDTA pH8.0, 1% Triton X-100, and 0.1% SDS) and twice with TE buffer pH8.0 (10mM Tris-Cl pH 8.0 and 0.1mM EDTA pH8.0). After the immunoprecipitation samples were RNase-treated (NEB) and subjected to Proteinase-K treatment for reversal of cross-links, 12 hours at 37°C and at least 6 hours at 65°C. The samples after proteinase K treatment were subjected to phenol chloroform extraction. After precipitating and pelleting, DNA was dissolved in 30 µl of TE buffer pH 8.0. The recovered DNA was converted into libraries using NebNext Ultra II DNA library preparation kit, following manufacturer's protocol.

TAF-ChIP and library preparation: Fixed cells were directly sorted in 240 µl of 140mM RIPA (10mM Tris-Cl pH 8.0, 140mM NaCl, 0.1mM EDTA pH8.0, 1% Triton X-100, and 0.1% SDS), and sonicated with 3 cycles at low power settings for breaking the nuclei. 15 µl of Protein A and G Dynabeads were coupled to 1 µg of specific antibody in the blocking buffer (RIPA 140mM supplemented with 0.2 mg/ml BSA, 0.05 mg/ml of glycogen and 0.2 mg/ml of yeast t-RNA) for 3 hrs at 4°C. The unfragmented chromatin were centrifuged at 14000 rpm for 10 minutes at 4°C, and the supernatant was transferred to the tube with blocked and antibody coupled beads. The samples were incubated at 4°C overnight with head over tail rotations. The samples were then washed twice briefly with 300 µl of home made tagmentation buffer (20 mM Tris(hydroxymethyl)aminomethane pH 7.6; 10 mM MgCl2; 20% (vol/vol) dimethylformamide) using magnetic rack for beads separation. The washed beads were resupended in 20 µl of 1X tagmentation DNA buffer (Nextera XT Kit) containing 1 µl of Nextera DNA tagmentation enzyme and incubated at 37 °C for 40 minutes with constant shaking in a thermoblock at 500 rpm. Following the tagmentation, the beads were washed as following; once with 140mM RIPA (10mM Tris-Cl pH 8.0, 140mM NaCl, 0.1mM EDTA pH8.0, 1% Triton X-100, and 0.1% SDS), four times with 250mM RIPA (10mM Tris-Cl pH 8.0, 250mM NaCl, 0.1mM EDTA pH8.0, 1% Triton X-100, and 0.1% SDS) and twice with TE buffer pH8.0 (10mM Tris-Cl pH 8.0 and 0.1mM EDTA pH8.0). The samples after proteinase K treatment for reverse-crosslinking were subjected to phenol chloroform extraction. After precipitating and pelleting, DNA was dissolved in 30 µl of TE buffer pH 8.0. The DNA was amplified in a 100 µl reaction with 1X NEBNext High-Fidelity PCR

Mix with primers containing molecular indices (listed in Table 1) with following program; 72 °C for 3min, {98 °C for 10 seconds, 63 °C for 30 seconds, 72 °C for 30 seconds} for 12 cycles, 72 °C for 5 minutes, and hold at 4 °C. The PCR reaction was purified with beads based size selection to remove any fragment that might be of larger than 1000 bp size. Ampure Xp beads were added to the PCR reaction in a ratio of 0.2X ratio to bind larger fragments. The magnetic beads were separated with the help of magnetic rack and supernatant was transferred to a fresh tube. Ampure Xp beads were added to the PCR reaction in a ration of 0.8X to bind the target library. After PCR purification, libraries were analyzed on Agilent Bioanalyzer for size distribution and the concentration was measured using Qubit fluorometer. The finished libraries were pooled in equimolar amounts and sequenced on illumina NextSeq 500.

Demultiplexing, and Mapping: De-multiplexing and fastq file conversion was performed using blc2fastq (v.1.8.4)**.** Reads from ChIP-Seq libraries were mapped using bowtie2 (v. 2.2.8) [23], and filtered for uniquely mapped reads. The genome build and annotation used for all Drosophila samples was BDGP6 (ensemble release 84). The genome build and annotation used for the K562 samples was hg38 (ensemble release 84).

Normalization and Peak calling: The mapped bam files were normalized to RPKM using deepTools, and bigwig coverage file was generated. Peak calling was performed using macs (v 2.1.1-20160309) [21]. The resulting peaks were annotated with the ChIPseeker package on Bioconductor, using nearest gene to the peak summit as assignment criteria [24].

References

1 Orlando, V., Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* 25 (3), 99-104 (2000).

2 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 13 (18), 2847-2852 (2014).

3 Stathopulos, P.B. *et al.*, Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci* 13 (11), 3017-3027 (2004).

4 Mieczkowski, J. *et al.*, MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun* 7, 11485 (2016).

5 Gutierrez, G. *et al.*, Subtracting the sequence bias from partially digested MNase-seq data reveals a general contribution of TFIIS to nucleosome positioning. *Epigenetics Chromatin* 10 (1), 58 (2017).

6 Dingwall, C., Lomonossoff, G.P., & Laskey, R.A., High sequence specificity of micrococcal nuclease. *Nucleic Acids Res* 9 (12), 2659-2673 (1981).

7 Adli, M., Zhu, J., & Bernstein, B.E., Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat Methods* 7 (8), 615-618 (2010).

8 Schmidl, C., Rendeiro, A.F., Sheffield, N.C., & Bock, C., ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* 12 (10), 963-965 (2015).

9 Zheng, X. *et al.*, Low-Cell-Number Epigenome Profiling Aids the Study of Lens Aging and Hematopoiesis. *Cell Rep* 13 (7), 1505-1518 (2015).

10 van Galen, P. *et al.*, A Multiplexed System for Quantitative Comparisons of Chromatin Landscapes. *Mol Cell* 61 (1), 170-180 (2016).

11 Picelli, S. *et al.*, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24 (12), 2033-2040 (2014).

12 Seguin-Orlando, A. *et al.*, Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PLoS One* 8 (10), e78575 (2013).

13 Ghavi-Helm, Y., Zhao, B., & Furlong, E.E., Chromatin Immunoprecipitation for Analyzing Transcription Factor Binding and Histone Modifications in Drosophila. *Methods Mol Biol* 1478, 263-277 (2016).

14 Berger, C. *et al.*, FACS purification and transcriptome analysis of drosophila neural stem cells reveals a role for Klumpfuss in self-renewal. *Cell Rep* 2 (2), 407-418 (2012).

15 Neumuller, R.A. *et al.*, Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell Stem Cell* 8 (5), 580-593 (2011).

16 Bowman, S.K. *et al.*, The tumor suppressors Brat and Numb regulate transit-amplifying neuroblast lineages in Drosophila. *Dev Cell* 14 (4), 535-546 (2008).

17 Wang, H. *et al.*, Aurora-A acts as a tumor suppressor and regulates self-renewal of Drosophila neuroblasts. *Genes Dev* 20 (24), 3453-3463 (2006).

18 An integrated encyclopedia of DNA elements in the human genome. *Nature* 489 (7414), 57-74.

[19] Landt, S.G. *et al.*, ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* 22 (9), 1813-1831 (2012).

[20] Ross-Innes, C.S. *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481 (7381), 389-393 (2012).

[21] Zhang, Y. *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9 (9), R137 (2008).

[22] Bier, E., Vaessin, H., Younger-Shepherd, S., Jan, L.Y., & Jan, Y.N., deadpan, an essential pan-neural gene in Drosophila, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev* 6 (11), 2137-2151 (1992).

[23] Langmead, B., Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* Chapter 11, Unit 11 17 (2010).

[24] Yu, G., Wang, L.G., & He, Q.Y., ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31 (14), 2382-2383 (2015).

Supplementary Figure legends

Supplementary Figure 1. TAF-ChIP libaries have high mappability and low replicate variabilities (A) Bionalyzer profile of a representative TAF-ChIP library showing the size distribution of final library fragments. **(B)** Table showing the percentage of mapped reads from TAF-ChIP experiments. The read duplicates were qunatified using Picard Tools. **(C)** and **(D)** Pearson correlation between the indicated replicates of TAF-ChIP samples from Drosophila 1000 cells, across equal sized bins of 500 bp **(E)** and **(F)** Pearson correlation between the indicated replicates of TAF-ChIP samples from 100 K562 cells, across equal sized bins of 5000 bp. **(G)** The percentage of peaks recovered for H3K4Me3 and H3K9Me3 TAF-ChIP samples from *Drosophila*, at various sequencing depths. The trend line fitting the data points is shown in grey.

Supplementary Figure 2. Distribution of peaks identified in the conventional ChIP-Seq and TAF-ChIP experiments with respect to their distance from the TSS (A) H3K4Me3, TAF-ChIP (1000 NSCs) and conventional ChIP-Seq (1.2 million cells). **(B)** H3K9Me3, TAF-ChIP (1000 NSCs) and conventional ChIP-Seq (1.2 million cells). **(C)** H3K4Me3, TAF-ChIP (100 K562 cells) and dataset from ENCODE project. **(D)** H3K9Me3, TAF-ChIP (100 K562 cells) and dataset from ENCODE project.