

# nMOWChIP-seq: low-input genome-wide mapping of non-histone targets

Zhengzhi Liu<sup>1</sup>, Lynette B. Naler<sup>2</sup>, Yan Zhu<sup>2</sup>, Chengyu Deng<sup>2</sup>, Qiang Zhang<sup>2</sup>, Bohan Zhu<sup>2</sup>, Zirui Zhou<sup>2</sup>, Mimosa Sarma<sup>2</sup>, Alexander Murray<sup>3</sup>, Hehuang Xie<sup>1,3</sup> and Chang Lu<sup>1,2,\*</sup>

<sup>1</sup>Department of Biomedical Engineering and Mechanics, Virginia Tech, Blacksburg, VA, USA, <sup>2</sup>Department of Chemical Engineering, Virginia Tech, Blacksburg, VA, USA and <sup>3</sup>Department of Biomedical Sciences & Pathobiology, Virginia Tech, Blacksburg, VA, USA

Received September 05, 2021; Revised March 18, 2022; Editorial Decision March 22, 2022; Accepted March 25, 2022

## ABSTRACT

**Genome-wide profiling of interactions between genome and various functional proteins is critical for understanding regulatory processes involved in development and diseases. Conventional assays require a large number of cells and high-quality data on tissue samples are scarce. Here we optimized a low-input chromatin immunoprecipitation followed by sequencing (ChIP-seq) technology for profiling RNA polymerase II (Pol II), transcription factor (TF), and enzyme binding at the genome scale. The new approach produces high-quality binding profiles using 1,000–50,000 cells. We used the approach to examine the binding of Pol II and two TFs (EGR1 and MEF2C) in cerebellum and prefrontal cortex of mouse brain and found that their binding profiles are highly reflective of the functional differences between the two brain regions. Our analysis reveals the potential for linking genome-wide TF or Pol II profiles with neuroanatomical origins of brain cells.**

## INTRODUCTION

Protein-DNA interactions are widely and critically involved in the regulations of gene transcription and expression (1). RNA polymerase II (Pol II) transcribes protein-coding genes into mRNA following several steps: formation of preinitiation complex, promoter-proximal pausing, elongation and termination. Pol II binding occurs throughout the genome, with enrichment at regions being either actively expressed or readied for imminent transcription upon environmental cues (2,3). This latter phenomenon of Pol II pausing is the rate limiting step on more than 70% of metazoan genes and has been shown to play key biological roles (3–6). Pol II is a key ChIP-seq target and the bindings of its subtypes reveal different states and stages of transcription in the genome (7,8). Similarly, transcription factors (TFs)

bind to DNA or cofactors and participate in gene regulations in a sequence-specific manner (1,9). TFs control key aspects of cellular biology including cell differentiation, development patterning, and immune response (8,10). They have long been profiled using ChIP-seq in order to identify their roles in gene activities (8). Finally, enzymes such as histone acetyl transferases (HATs) and histone deacetylases (HDACs) closely interact with the genome to effect epigenetic modifications (histone acetylations) that are critically involved in gene silencing and activation (11). They are also involved in processes beyond histone modification, such as memory formation and synaptic plasticity in brain (12,13). Targeting a single histone acetylation marker such as H3K27ac does not provide the full scope of information if the role of specific HAT/HDAC are being investigated (14). The binding profiles of these proteins are often indicative of genes that are activated or readied in a process. Difference in the binding profile among various tissue samples reveals underlying genome-wide molecular dynamics and variations involved in development or disease.

Chromatin immunoprecipitation coupled with sequencing (ChIP-seq) is a simple and direct approach to profile *in vivo* genome-wide binding of proteins. Although ChIP-seq has been generating reliable and high-quality data on histone modifications (i.e. the interaction between a modification histone and the genome) (15–17), ChIP-seq results on other types of protein-DNA interactions tend to be much more challenging. Compared to the robust histone-DNA interaction, the interaction between DNA and other protein molecules such as Pol II, TFs, and enzymes may be harder to preserve even after treatment such as formaldehyde crosslinking. Thus conventional ChIP-seq requires a large quantity of starting material (>10<sup>7</sup> cells per assay) for probing genomic binding of proteins that are not histones and the results tend to have lower reproducibility compared to histone ChIP-seq (18). In recent years, significant efforts have been made in developing low-input ChIP-seq methods (15–17,19–26). However, the vast majority of these methods were only demonstrated on examining histone modifi-

\*To whom correspondence should be addressed. Email: changlu@vt.edu

cations, with a few exceptions that also profiled transcription factors (27–29).

As an alternative to ChIP, IP-free technologies such as CUT&RUN (30,31) and ChIL-seq (32) were developed to profile factor binding to the genome. CUT&RUN maps Pol II and TF binding by cutting and releasing DNA fragments that interact with antibody-targeted Pol II and TFs into supernatant, requiring as few as 1,000 cells (30,31). The latest variation of CUT&RUN (CUT&Tag) (33) have been used to profile single cells. However, CUT&RUN data appear to exhibit lower correlation with the gold-standard ChIP-seq data (e.g. ENCODE data) compared to low-input ChIP-seq technologies (30,31,34). CUT&RUN datasets may also be contaminated by DNA sequences that are from unknown sources associated with the process (e.g. sequences containing (TA)<sub>n</sub>) (35).

Here we establish micrococcal nuclease (MNase)-digestion-based native MOWChIP-seq as a general tool for low-input profiling of protein binding to genome. MNase digestion was previously applied in native ChIP to probe mostly histone modifications (36) and also non-histone proteins including Pol II and TFs (26,37–41). However, these studies were all conducted using a large number of cells ( $10^5$ – $10^8$ ). Our approach, referred to as native MOWChIP-seq or nMOWChIP-seq, combines MNase digestion of native chromatin with a microfluidics-based low-input ChIP-seq technology (MOWChIP-seq (15,17)) that was previously applied to examine histone modifications only. We show that MNase digestion is effective for preserving the links between protein and genome, and essential for application of MOWChIP-seq to Pol II, TFs and enzymes (hence the name ‘nMOWChIP-seq’). We generated high-quality ChIP-seq data using as few as 1,000 cells for studying Pol II, 5,000 cells for TF EGR1, and 50,000 cells for HDAC2. We applied this method to study genome-wide binding of Pol II, EGR1 and MEF2C in two functional regions of the mouse brain: prefrontal cortex (PFC) and cerebellum. Extensive variations in RNA Pol II and TF binding were identified between these two brain regions and these profiles reveal the involvement of these regulatory molecules in the functional difference.

## MATERIALS AND METHODS

### Cell culture

GM12878 cells were obtained from Coriell Institute for Medical Research. Cells were cultured in RPMI-1640 medium (30–2001, ATCC) with 15% fetal bovine serum (16000–044, Gibco) and 1% pen-strep (Invitrogen) at 37°C, 5% CO<sub>2</sub>. Cells were subcultured every 3 d to maintain exponential growth.

### Mouse strain and brain dissection

C57BL/6J mice were purchased from Jackson Laboratory and maintained in the animal facility with 12-h light/12-h dark cycles and food and water ad libitum. 8-week old male mice were sacrificed by compressed CO<sub>2</sub> followed by cervical dislocation. Mouse brains were rapidly dissected, frozen on dry ice and stored at –80°C. This study was ap-

proved by the Institutional Animal Care and Use Committee (IACUC) at Virginia Tech.

### Nuclei isolation from brain tissues

A mouse brain was put on ice and PFC and cerebellum were dissected for nuclei isolation. The following steps were performed on ice and centrifugation performed at 4°C. Tissue was placed in 3 ml of ice-cold nuclei extraction buffer [0.32 M sucrose, 5 mM CaCl<sub>2</sub>, 3 mM Mg(Ac)<sub>2</sub>, 0.1 mM EDTA, 10 mM tris-HCl, and 0.1% Triton X-100, with 30 μl of PIC (P8340, Sigma-Aldrich), 3 μl of 100 mM PMSF, and 3 μl of 1 M dithiothreitol added before use]. Tissue was homogenized in the grinder set (D9063, Sigma-Aldrich) by slowly douncing 15 times with pestle A and 25 times with pestle B. Homogenate was filtered through a 40 μm cell strainer into a 15 ml tube and centrifuged at 1000g for 10 min. The supernatant was removed and the pellet was resuspended in 500 μl nuclei extraction buffer and transferred to a 1.5 ml tube. 750 μl of 50% iodixanol, 7.5 μl of PIC, 0.75 μl of 100 mM PMSF and 0.75 μl of 1M dithiothreitol were added and mixed by pipetting. The mixture was centrifuged at 10,000g for 20 min and the supernatant was removed. If mixed nuclei (without separation of neurons and glia) were used for nMOWChIP directly, the nuclei pellet was resuspended in 200 μl of Dulbecco’s phosphate-buffered saline (DPBS). If nuclei labeling and FACS sorting were conducted, 500 μl of 2% normal goat serum (50062Z, Life Technologies) in DPBS was added to the nuclei pellet and incubated for 10 min before resuspending. Anti-NeuN antibody conjugated with Alexa 488 (MAB377X, EMD Millipore) was diluted with DPBS to 2 ng/μl. 8 μl of anti-NeuN was added to each 500 μl of nuclei suspension and incubated at 4°C for 1 h on a rotator. The labeled nuclei were then sorted using FACS (BD FACSAria, BD Biosciences). 8 μl of non-labeled nuclei were saved as unstained control prior to addition of anti-NeuN antibody. The concentration of nuclei suspension after FACS was typically low ( $\sim 1.2 \times 10^5$ /ml). The nuclei were re-concentrated by adding 200 μl of 1.8 M sucrose, 5 μl of 1M CaCl<sub>2</sub> and 3 μl of 1M Mg(Ac)<sub>2</sub> to 1 ml of the nuclei suspension. The mixture was incubated on ice for 15 min and centrifuged at 1800g for 15 min. Supernatant was removed and the nuclei pellet was resuspended in DPBS to generate a suspension containing  $4 \times 10^6$  nuclei/ml.

### MNase digestion of chromatin

This protocol is scalable in volume and can digest cell/nuclei suspension with concentration up to  $4 \times 10^6$ /ml. Our experiments usually started with  $4 \times 10^5$  cells/nuclei suspended in 100 μl of DPBS. 1 μl of PIC, 1 μl of 100 mM PMSF and 100 μl lysis buffer [4% Triton X-100, 100 mM tris-HCl, 100 mM NaCl, and 30 mM MgCl<sub>2</sub>] were added, mixed by vortexing and incubated at room temperature for 10 min. 10 μl of 100 mM CaCl<sub>2</sub> and 2.5 μl 100U MNase (88216, Thermo Fisher Scientific) were added, mixed by vortexing and incubated at room temperature for 10 min. 22 μl of 0.5 M EDTA was then added, mixed by vortexing and incubated on ice for 10 min. The solution was centrifuged at 16,100g for 5 min at 4°C. Supernatant containing fragmented chromatin was collected into a new 1.5 ml tube

and placed on ice for use. Volumes equal to 100,000 and 50,000 cells/nuclei were portioned from the final chromatin (~220  $\mu$ l) for assays. For assays using 10,000 cells/nuclei or less, 100  $\mu$ l of suspension containing 40,000 cells/nuclei was used in the first step and the same procedure as above was followed.

### Preparation of immunoprecipitation (IP) beads

5  $\mu$ l of protein A Dynabeads (10001D, Invitrogen) were used in each MOWChIP assay. The beads were washed twice with IP buffer (20 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% (w/v) sodium deoxycholate, 0.1% SDS, 1% (v/v) Triton X-100) and resuspended in 150  $\mu$ l of IP buffer. 1  $\mu$ g of Pol II or TF antibody was added into the bead suspension for each assay using  $10^5$  cells/nuclei, while 0.5  $\mu$ g of antibody was added for each assay using  $5 \times 10^4$  or fewer cells/nuclei. We used the following antibodies in this work: anti-Pol II-total (ab817, lot GR3271868-2, Abcam), anti-Pol II-S5 (ab5131, lot GR3202335-5, Abcam), anti-Pol II-S5 4H8 (ab5408, lot GR3325973-4, Abcam), anti-EGR1 (sc-101033, lot A1618, Santa Cruz), anti-MEF2C (sc-365862, lot B0818, Santa Cruz), anti-HDAC2 (ab124974, lot GR97402-7, Abcam). The suspension was incubated on a rotator at 4°C for 2 h. The beads were then washed with IP buffer twice and resuspended in 5  $\mu$ l of IP buffer for loading into the device chamber.

### MOWChIP-seq

The microfluidics-based MOWChIP-seq process, including microfluidic device fabrication and operation, was conducted following our published protocol (17).

### Purification of ChIP DNA

After nMOWChIP, IP beads were rinsed once with IP buffer and resuspended in 200  $\mu$ l of DNA elution buffer (10 mM Tris-HCl, 50 mM NaCl, 10 mM EDTA, and 0.03% SDS). 2  $\mu$ l of 20 mg/ml proteinase K was added and the suspension was incubated at 65°C for 1 h. DNA was then extracted and purified by phenol-chloroform extraction and ethanol precipitation. DNA pellet was resuspended in 8  $\mu$ l of low EDTA TE buffer. Input DNA was purified with the same process, by digesting chromatin solution directly using proteinase K, followed by the same extraction and purification process.

### Library preparation, quantification and sequencing

Libraries were constructed using Accel-NGS 2S Plus DNA Library kit (Swift Biosciences) following the manufacturer's instructions.  $1 \times$  EvaGreen dye (Biotium) was added to the amplification mixture to monitor the PCR amplification. Library was eluted to 7  $\mu$ l of low EDTA TE buffer. Library fragment size was examined using TapeStation (Agilent) and the concentration was quantified with KAPA Library Quantification kit (Kapa Biosystems). We also examined the enrichment of the libraries using qPCR and primers listed in Supplementary Table S3. Libraries were pooled for sequencing by Illumina HiSeq 4000 SR50 mode.

### ChIP-seq data analysis

Raw sequencing FASTQ files were trimmed using Trim Galore! (0.4.1) with default settings. Reads that passed the quality check were mapped to reference genomes hg38 (human) or mm10 (mouse) with bowtie (1.1.2) (42). Uniquely mapped reads were filtered for known blacklisted genome regions using samtools (1.3.1) (43) and bedtools (2.29.2) (44) to remove ChIP-seq artifacts. The reference genome was then divided into 100-bp bins and ChIP-seq signal for each bin was counted for both ChIP and input samples. Normalized ChIP-seq signal for each bin was calculated using the following equation:

$$Signal_{100bp} = \left( \frac{ChIP\ signal}{No.\ of\ unique\ ChIP\ reads} - \frac{Input\ signal}{No.\ of\ unique\ input\ reads} \right) \times 1,000,000$$

Both ChIP and input reads were then extended by 100 bp on either ends to compute normalized signal for each bin in the same manner, and visualized as tracks in genome browser IGV (2.11.2) (45). Peak calling was conducted using MACS2 (2.1.1.20160309, default settings). Differential binding analysis was carried out using DiffBind (3.14) (46) with default settings. Selected genomic regions of interest were further analyzed for gene ontology terms using GREAT (47). Pearson's correlation between datasets was calculated via DiffBind, using *dba.count* command with default parameters to process MACS2 peaks and uniquely mapped reads and calculate consensus peaks and affinity score. Consensus peaks between technical replicates are identified by bedtools intersect with the additional setting of *-f 0.5*. Overlap between consensus peaks are plotted using Intervene (48).

ROC curves were created by calculating the true positive rate (TPR) and false positive rate (FPR) at 100 different peak-calling thresholds from 0.99 and  $10^{-15}$ . Thresholds were applied to both the EGR1 and Pol II sets, and AUC was calculated with the R package ROC. Subsampling of EGR1 bam files to between 25% to 95% of original reads was performed using samtools. Five separate, but consistent, seeds were used at each subsampling percentage. Matthew's correlation coefficient (MCC) was calculated in R (mltools) at each percentage and each seed using a MACS2 q-value cutoff of 0.05 for both EGR1 and Pol II peaks.

## RESULTS

### Profiling genome-wide binding of RNA Pol II, TFs, and enzymes

RNA Pol II, TF or enzyme binding has been conventionally studied after crosslinking using reagent such as formaldehyde that firmly immobilizes the protein to the genomic DNA (49,50). However, our initial results showed that our low-input MOWChIP-seq technology (15,17) yielded low-quality results when at least 100,000 to 1 million cells were crosslinked and sonicated to create the chromatin fragments before Pol II-S5 ChIP (Supplementary Figure S1). The data quality decreased when less cells were used (Supplementary Figure S1A). 53,817 and 17,561 peaks were yielded in the two technical replicates of the sonicated crosslinked samples using 100,000 cells, compared to 114,811 and 140,695

peaks generated by nMOWChIP-seq using 10,000 cells per assay. Data on crosslinked samples had a Jaccard index of 0.23 between replicates, lower than nMOWChIP-seq samples (0.48). We verified the size distribution of sonicated crosslinked chromatin and found no abnormality in terms of fragmentation efficiency (Supplementary Figure S1D). Because crosslinking and sonication have been successfully used in conventional ChIP-seq of Pol II using millions of cells, it is likely that the data quality issues here are specific to low-input assays (<100,000 cells per assay). In our Pol II-S5 in GM12878 profiling experiments, crosslinked MOWChIP with 50,000 cells yielded an average of 0.74 ng ChIP DNA, while nMOWChIP with 50,000 cells yielded an average of 0.51 ng ChIP DNA.

We then applied nMOWChIP-seq to profile protein binding to the genome. MNase digestion was previously applied in native ChIP to probe mostly histone modifications (36) and also non-histone proteins including Pol II and TFs (26,37–41). In our process (Figure 1A), tissues were first mechanically homogenized to extract nuclei and cultured cells were directly used. Nuclei or cells were lysed and digested with MNase to yield chromatin fragments with size range appropriate for ChIP (150–600 bp). That was followed by the MOWChIP process (15,17). Briefly, in a microfluidic chamber with a partially closed sieve valve, antibody-coated magnetic IP beads were packed into a dense bed. All chromatin fragments were forced through the bead bed, drastically increasing the adsorption efficiency of targeted fragments. Oscillatory washing was then applied to remove non-specific binding, and the beads with bound chromatins were transferred out of the chamber for DNA elution, library preparation and sequencing.

Using GM12878, a lymphoblastoid and ENCODE tier 1 cell line, we showed that our method could generate high quality ChIP-seq data with input as little as 1,000 cells per assay for Pol II-S5 (ab5131) (Figure 1B). Cells were first digested in tube by MNase and volumes equal to the desired cell numbers were portioned for nMOWChIP assays. Pearson correlations between technical replicates were 0.87, 0.84, 0.60, and 0.54 for 50,000-, 10,000-, 2,000-, and 1,000-cell samples on RNA Pol II-S5, respectively (Figure 1C). We benchmarked our data against ENCODE data obtained using conventional ChIP-seq method and 20 million cells per assay. We examined the number of peaks called and the fraction of reads in peaks (FRiP) for data quality (Supplementary Table S1). The nMOWChIP-seq Pol II data produced averagely 136,504, 127,753, 65,076 and 26,724 peaks with 50,000, 10,000, 2,000 and 1,000 cells per assay, respectively, compared to 111,350 peaks from the ENCODE data obtained using 20 million cells. We examined the overlap of peaks produced by these assays using decreasing number of cells (Supplementary Figure S2). For each dataset associated with a specific cell number, common peaks were identified by merging peaks that overlap for more than 50% of their peak region between technical replicates. 47% of the common peaks produced from the 1,000-cell samples overlapped with those from all the other 3 datasets (2,000, 10,000, and 50,000 cells per assay). Only 2% of the common peaks from 1,000-cell dataset do not overlap with any common peaks from other conditions. We observed that most common peaks

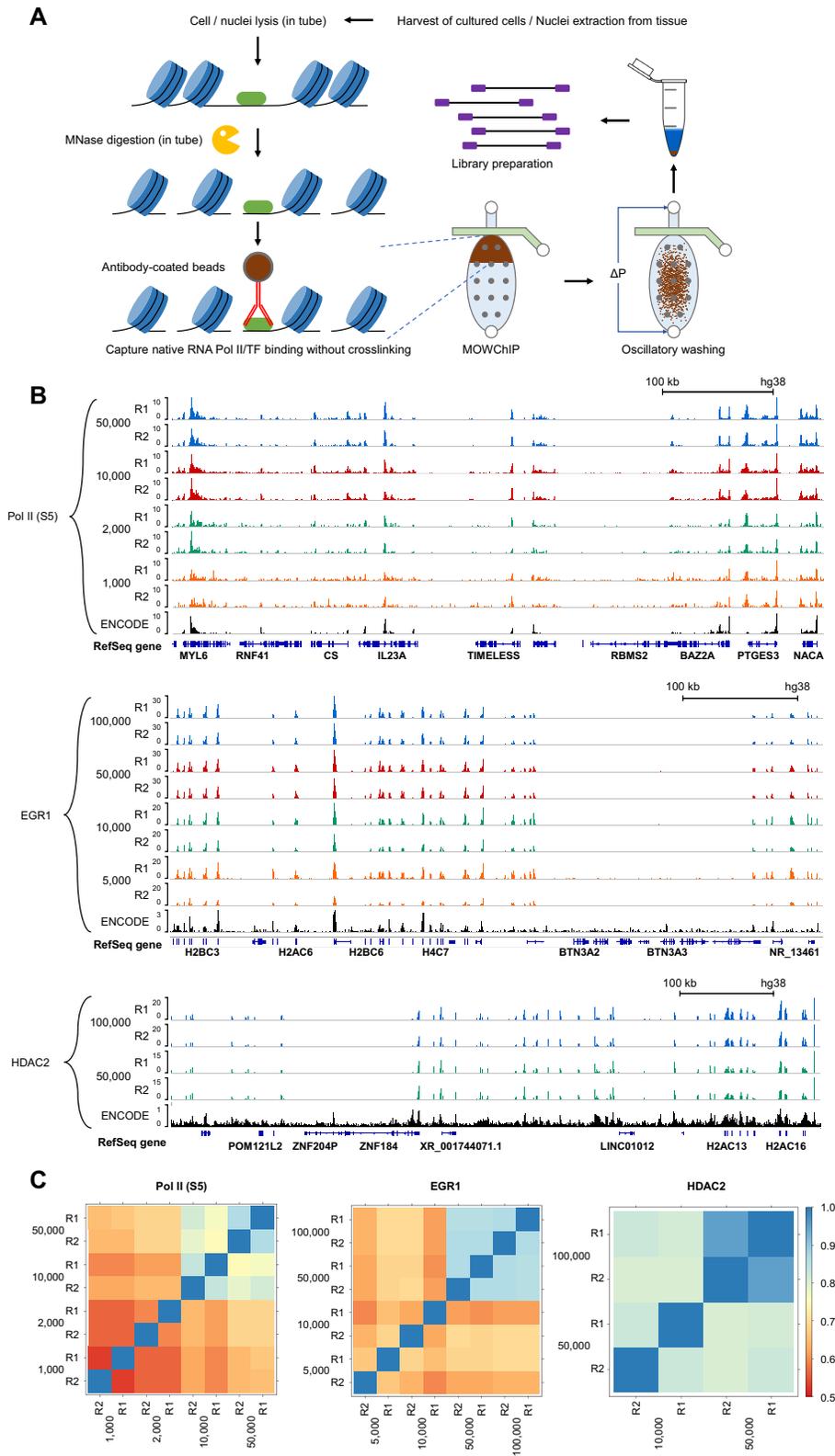
identified by the low-cell-number assays coincided with the strongest peaks in the large-cell-number datasets (Figure 1B). FRiP measures the amount of background in the data and nMOWChIP-seq Pol II data showed a decrease from 49.8% to 9.4% with cells per assays decreasing from 50,000 to 1,000, far exceeding the 1% threshold recommended by ENCODE (51).

We also determined that the binding of Pol II was stable enough to be preserved at  $-80^{\circ}\text{C}$ , a unique property not observed in histone modifications or TFs. From our testing, MNase-digested chromatin can be frozen at  $-80^{\circ}\text{C}$  or on dry ice for 2 d without substantial degradation in data quality for Pol II. We digested 40,000 cells and stored the chromatin on dry ice for 2 d. Volumes equal to 10,000 cells were then portioned for MOWChIP-seq and the data is shown as ‘10,000-st’ (Supplementary Figure S3).

It is worth noting that our nMOWChIP-seq Pol II-S5 profile shows higher signal than the corresponding ENCODE profile at regions that are close to the transcription ending site (TES) (Supplementary Figure S4). We conducted nMOWChIP-seq using two different Pol II-S5 antibodies (ab5131 and ab5408, with the latter used in the ENCODE data). For example, compared to the ENCODE data taken using conventional ChIP-seq technology, extra peaks near the TES regions can be observed in both nMOWChIP-seq datasets at genes *PTGES3* and *NACA* (Supplementary Figure S4A). Similar genome-wide trend can be seen in the average ChIP-seq signal per gene over promoter and gene body (Supplementary Figure S4B).

We profiled a transcription factor, early growth response protein 1 (*EGR1*), using as few as 5,000 GM12878 cells (Figure 1B and C, Supplementary Table S1). Pearson correlation coefficients between the technical replicates were 0.87, 0.87, 0.63, and 0.67 for 100,000-, 50,000-, 10,000-, and 5,000-cell samples, respectively (Figure 1C). *EGR1* data generally show lower peak numbers and FRiP than Pol II data, as shown by both nMOWChIP-seq and ENCODE data. FRiP ranged from 8.4% to 3.2% in our nMOWChIP-seq *EGR1* data. We also profiled another TF *MEF2C* (Supplementary Table S1). A large percentage of the *EGR1* (85%) and *MEF2C* (75%) peaks overlap with Pol II peaks (Supplementary Figure S5).

Finally, we applied nMOWChIP-seq to examine the binding of histone deacetylase *HDAC2* in GM12878 cells (Figure 1B and C, Supplementary Table S1). At least 50,000 cells were required to generate good quality data on *HDAC2* binding. An average of 1,394 (FRiP = 1.2%) and 8527 peaks (FRiP = 3.1%) were generated using 50,000 and 100,000 cells per assay, compared to 1820 peaks of ENCODE data obtained using 10 million cells (FRiP = 0.3%). The nMOWChIP-seq FRiP values still compare favorably to those of ENCODE when they are calculated from overlapping peaks between technical replicates (2.06% and 0.93% for 100,000 and 50,000 cells compared to 0.03% of ENCODE). We also examined *HDAC2* binding in mouse brain cells (Supplementary Figure S6). Both unsorted nuclei mixture and neuronal nuclei were used to verify that nMOWChIP can profile *HDAC2* in tissue samples and sorted nuclei. 100,000 mixed nuclei and 80,000 FACS-sorted NeuN + neuronal nuclei from PFC were used in each assay and an average of 3193 (FRiP ~13.1%) and 2112



**Figure 1.** Overview of the low-input nMOWChIP process and data on RNA polymerase II, EGR1 and HDAC2 binding in GM12878 cells. (A) Steps to generate nMOWChIP-seq libraries. Cells/nuclei are lysed and digested by MNase to release chromatin fragments in tube. ChIP on antibody-coated beads and washing are conducted in the MOWChIP device. DNA elution and library preparation are done off the device. (B) Normalized RNA Pol II-S5 ChIP signals generated using 1,000 to 50,000 cells, EGR1 signals generated using 5,000 to 100,000 cells, and HDAC2 signals generated using 50,000 and 100,000 cells. ENCODE data (Pol II: GSM803485; EGR1: GSM803434; HDAC2: GSE105521) are included for comparison. (C) Pearson's correlations of affinity score calculated from consensus MACS2 peaks and uniquely mapped reads using DiffBind, among samples of various cell numbers for RNA Pol II, EGR1 and HDAC2.

(FRiP  $\sim 10.4\%$ ) peaks were produced (Supplementary Table S1).

### Comparison of Pol II and Pol II-S5 binding profiles

We used two antibodies to differentiate the binding of Serine 5 phosphorylated subset of Pol II (Pol II-S5, ab5131) and all Pol II regardless of its phosphorylation status (Pol II-total, ab817, clone 8WG16). Our data showed the distinction (Figure 2). The two profiles on GM12878 cells were similar in a large fraction of the genome (Figure 2A). However, some genes (*ACTG1* and *MDM2* as examples) with Pol II binding over the entire gene body showed that Pol II-total profile captured the initial binding of unphosphorylated Pol II at TSS, which was absent in Pol II-S5 profile in comparison (Figure 2B). We examined the signal strength at promoter region and gene body, computed the ratio for all genes and compared between Pol II-total and Pol II-S5 profiles. 51% of mutual genes with both Pol II-total and Pol II-S5 bindings showed a higher ratio of promoter over gene body signal intensity (fold change  $> 2$ ) in Pol II-total data than in Pol II-S5 data. During the course of binding to a gene, Pol II is unphosphorylated during the pre-initiation stage, but undergoes phosphorylation once a short (20–60 bp) mRNA begins to transcribe (2). Being able to differentiate the distinct forms of Pol II is important when the exact Pol II binding status on specific genes is of interest (e.g. whether pre-initiation or activated Pol II dominates pausing at a TSS).

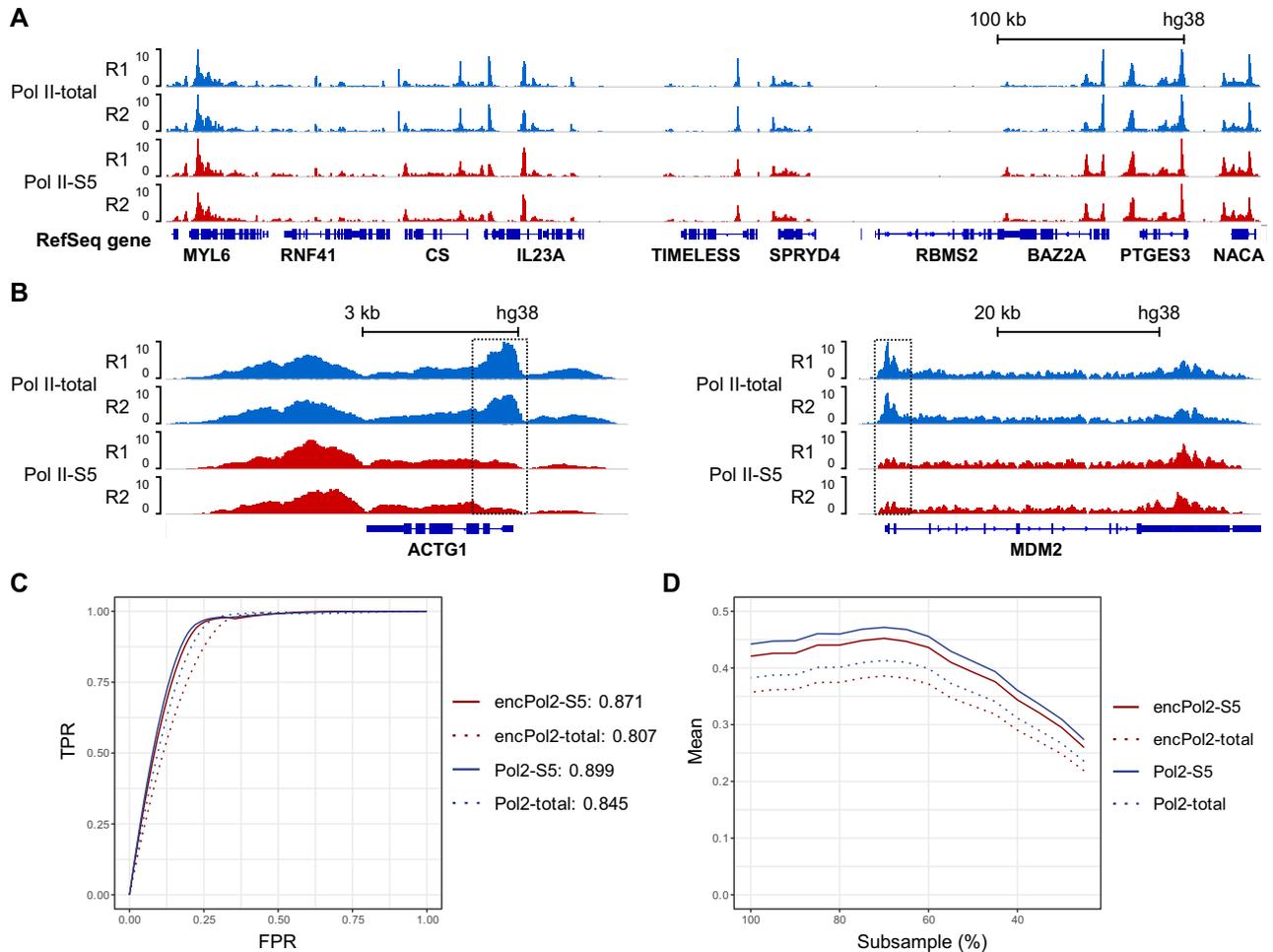
To determine whether Pol II-total or Pol II-S5 was more effective in predicting TF binding sites, we analyzed the overlap of either Pol II-total peaks or Pol II-S5 peaks with EGR1 peaks at varying peak-calling thresholds, shown as receiver operating characteristic (ROC) curves that display the true positive rate (TPR) versus the false positive rate (FPR) (Figure 2C). To quantify the predictive quality, we calculated the area under the curve (AUC) for each of the ROC curves using our EGR1 binding data as the gold standard. We found that Pol II-S5 profile had a higher predictive value than Pol II-total (0.899 vs 0.845 using nMOWChIP-seq data, and 0.871 vs 0.807 using ENCODE data). We also tested if Pol II-S5 was more robust than Pol II-total with samples of lower quality (Figure 2D). For this, we subsampled our EGR1 data from 25% to 95% of the original reads, with 5 replicates at each sampling. We then calculated the average Matthew's correlation coefficient (MCC), a robust single value quantifier of classifier quality, of the five replicates at each subsampling percentage. Pol II-S5 outperformed Pol II-total, both in our data and in ENCODE's, regardless of TF data quality.

### Differential RNA Pol II binding in prefrontal cortex and cerebellum of mouse brain

We applied nMOWChIP-seq to profile Pol II binding in mouse PFC and cerebellum. PFC have roles in cognitive functions, decision-making and short-term memory (52,53), while cerebellum controls motor functions and coordination (54). We reasoned that the functional difference between the two regions should reflect on the binding of Pol II and TFs that have recognized roles in the brain. There have not been published ChIP-seq data confirming this.

We mapped Pol II-total binding in B6 mouse brain using nuclei extracted from PFC and cerebellum with 50,000 nuclei used per assay (Figure 3). The genome-wide Pol II profiles were substantially different between cerebellum and PFC with Pearson's correlation based on DiffBind affinity score being 0.67, compared to an average of 0.96 between technical replicates (Figure 3A and B). DiffBind analysis identified 3021 peaks with higher levels of Pol II binding in PFC than in cerebellum, and 1197 peaks having higher binding in cerebellum (fold change  $> 2$ ,  $P < 10^{-5}$ ). Differentially bound peaks for Pol II-total are listed in Supplementary Dataset 1. Gene ontology (GO) analysis of these regions showed that the regions with high Pol II binding intensity in PFC were enriched in terms associated with memory, learning and anxiety-related response (Figure 3C). Synaptic plasticity, one of the fundamentals of learning and memory, was also enriched along with its key components: long term potentiation and depression (Figure 3C) (55). In contrast, the genomic regions with Pol II binding higher in cerebellum were enriched in terms that were specific to cerebellum (e.g. cerebellum morphology and development) (Figure 3C).

We also examined the Pol II pausing index for the Pol II-bound genes in PFC and cerebellum. Pol II pausing index (PI) refers to the ratio of Pol II read density between promoter proximal region (–30 to +300 bp of TSS) and gene body (+300 bp of TSS to TES), with higher PI indicating more Pol II binding near TSS (2,56). Pol II-bound genes can be divided into three categories based on the PI value: non-paused and expressed ( $PI < 2$ ), paused and expressed ( $2 < PI < 20$ ), and paused and unexpressed ( $PI > 20$ ). (2) For example, *Plk2*, which is a gene known to participate in rodent brain development and cell proliferation (57), and *Npas4*, which is involved in regulating reward-related learning and memory (58), both showed dramatically higher PI values in cerebellum than in PFC (Figure 3D). The data revealed that they were actively expressed in PFC but paused without expression in cerebellum. In contrast, *Slc46a1* and *Rsph9* showed higher PI values in PFC than cerebellum. While both were being expressed in the two tissues, there was significant pausing of Pol II at TSS in PFC. This suggests that PFC had more potential in transcribing these genes at a short notice, such as *Slc46a1*, which encodes a facilitative carrier for folate (59). Furthermore, we analyzed the distribution of the PI in PFC and cerebellum (Figure 3E). 36% of genes were being actively expressed in PFC ( $PI < 20$ ), compared to only 25% in cerebellum. Mouse brain displayed more pausing and less active transcription compared to GM12878 (52% actively expressed genes, Supplementary Figure S7A). This was also within our expectation because GM12878 cell line was maintained in log phase and actively dividing, unlike brain cells in adult mice. By examining the relationship between the PI and the gene expression level (derived from RNA-seq data), we show that the PI is highly suggestive of gene expression level (Supplementary Figure S7B). We identified Pol II-bound genes that displayed significantly different pausing indexes between PFC and cerebellum (fold change  $> 3$ , minimal read density  $> 0.02$  read/bp) (Supplementary Table S2). Among these, *Igfbp6* is involved in myelin formation during central nervous system (CNS) development (60), and *Slc1a2* en-



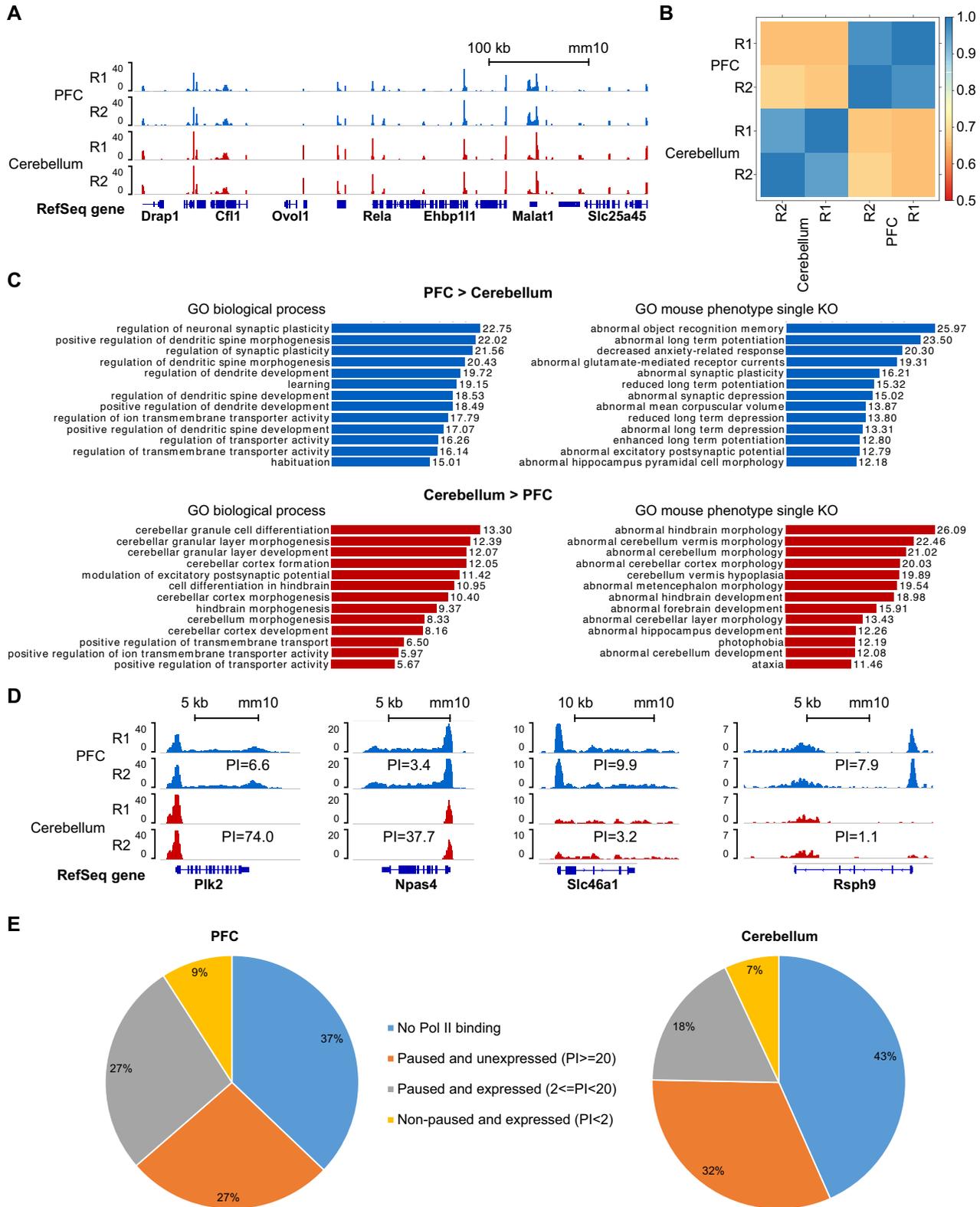
**Figure 2.** Comparison of Pol II-total and Pol II-S5 bindings in GM12878 cells. (A) Normalized Pol II-S5 and Pol II-total signal in GM12878 cells (50,000 cells per assay). (B) Normalized Pol II signal over genes *ACTG1* and *MDM2*, showing the difference in Pol II-total and Pol II-S5 binding. Pol II-total includes Pol II that is not phosphorylated. These non-active Pol II molecules tend to pause at promoter regions, while phosphorylated and active Pol II (S5 being the majority) binds to the entire gene body. (C) Receiver operating characteristic (ROC) curves using ENCODE (SRX100400 and SRX100530) and nMOWChIP-seq Pol II-total and Pol II S5 datasets predicting EGR1 binding peaks in GM12878 cells. (D) Matthew's correlation coefficient (MCC) plots of ENCODE and nMOWChIP-seq datasets being subsampled at different levels to predict EGR1 binding peaks in GM12878 cells.

codes excitatory amino acid transporter 2, responsible for reuptake of 90% glutamate in CNS (61). *Shank3* belongs to the *Shank* gene family that plays a role in synapse formation (62), and *Flrt2* encodes a member of the FLRT protein family that is shown to regulate signaling during mouse development (63).

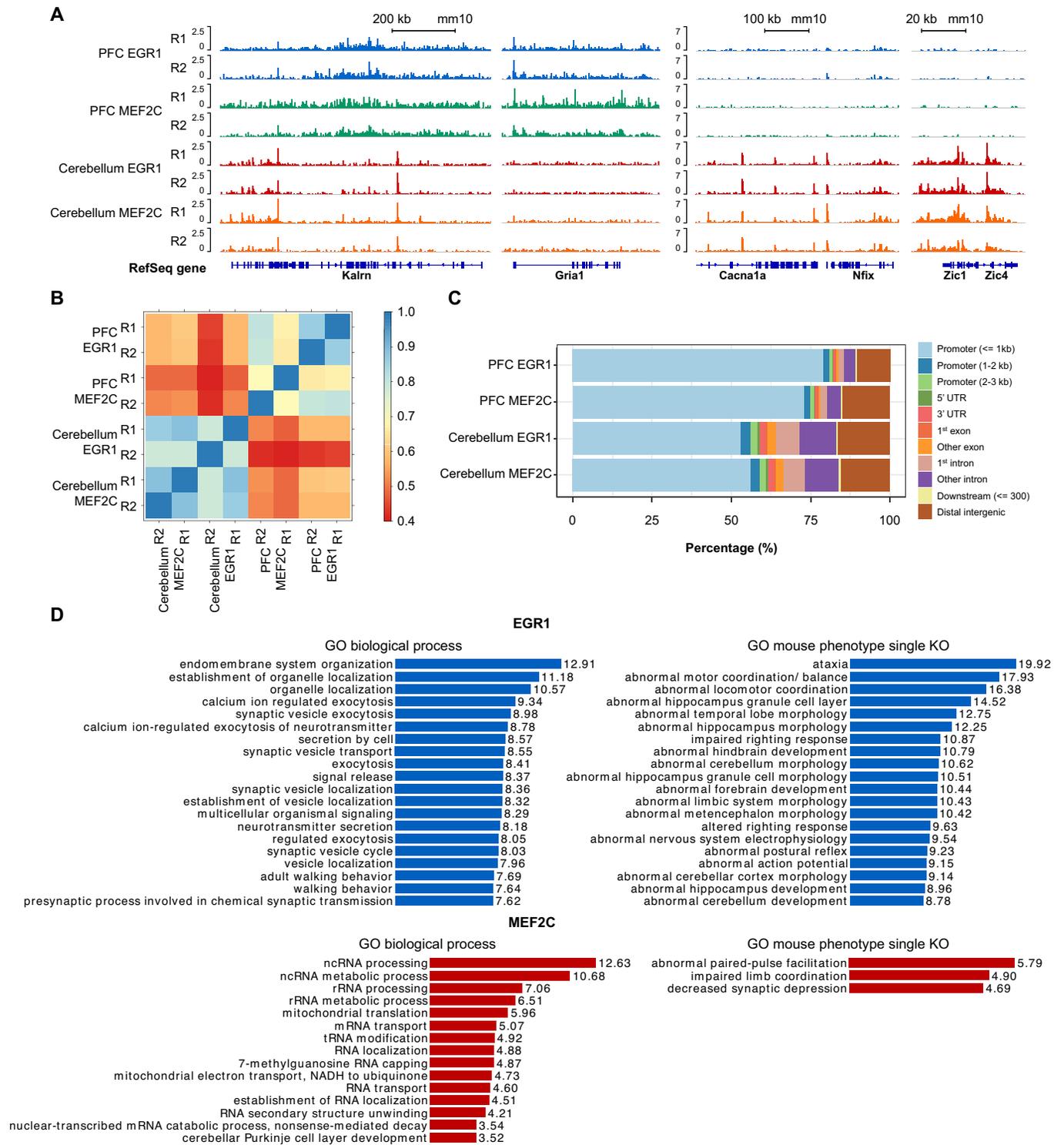
### Differential EGR1 and MEF2c binding in prefrontal cortex and cerebellum of mouse brain

We also examined EGR1 and MEF2C (myocyte enhancer factor-2 C) binding using nuclei extracted from mouse PFC and cerebellum. Chromatin from 100,000 nuclei were used in each assay, yielding high quality data that revealed differential binding between the two regions of brain (Figure 4A). We picked several genes as examples (Figure 4A). *Kalrn* (*Kalirin*) plays important roles in nerve growth (64). Higher EGR1/MEF2C activity in PFC was observed on *Gria1* (Glutamate receptor 1) which is involved in synaptic transmission (65). *Cacna1a* is involved in movement disorder

(66) and expression of *Nfix* can influence neural stem cell differentiation (67). *Zic1* and *Zic4* belong to the family of Zinc finger of the cerebellum (ZIC) protein family (68), whose loss of function can lead to Dandy-Walker malformation and incomplete cerebellar vermis (69). Higher EGR1 binding on these four genes were seen in cerebellum than in PFC. A large fraction of EGR1 and MEF2C peaks (68–89%) appeared to overlap with Pol II peaks, in both cerebellum and PFC (Supplementary Figure S8). We observed correlated EGR1 and MEF2C profiles, with an average of Pearson's correlation based on affinity score of 0.74 between the two TFs in PFC and 0.84 in cerebellum (Figure 4B). Such correlations were much higher than the one observed in GM12878 cells ( $r \sim 0.52$ ). On the other hand, EGR1 and MEF2C presented very different profiles between PFC and cerebellum, with the average correlation of 0.51 between the two brain regions for both TFs. We further examined the binding difference between PFC and cerebellum for these two TFs. Their binding sites were much more heavily situated at promoters in PFC (79%) than in cerebellum



**Figure 3.** Differential RNA polymerase II binding in prefrontal cortex and cerebellum of mouse brain. (A) Normalized Pol II-total signals generated using nuclei isolated from mouse prefrontal cortex (PFC) and cerebellum (50,000 nuclei per assay). (B) Pearson's correlations of affinity score calculated from consensus MACS2 peaks and uniquely mapped reads using DiffBind, among Pol II data on PFC and cerebellum. (C) GO biological processes and mouse phenotypes (single knockout) associated with regions having higher Pol II binding levels in either PFC or cerebellum ( $-\log_{10}$  binomial p value, fold change  $> 2$ ,  $P < 10^{-3}$ ). (D) Normalized Pol II-total signals at genes that have significantly different pausing indexes between PFC and cerebellum. (E) Distribution of Pol II pausing index in PFC and cerebellum.



**Figure 4.** Differential transcription factor binding in prefrontal cortex and cerebellum of mouse brain. (A) Normalized EGR1 and MEF2C signals at genes identified by DiffBind to have significantly different binding between PFC and cerebellum (fold change > 2,  $P < 10^{-5}$ ). (B) Pearson's correlations of affinity score calculated from consensus MACS2 peaks and uniquely mapped reads using DiffBind, among TF data on PFC and cerebellum. (C) Distribution of TF binding peaks ( $P < 10^{-5}$ ) in various genomic regions. Each peak set is the overlapping peaks from 2 replicates of the same sample. (D) GO biological process and mouse phenotype (single knockout) terms associated with regions having higher EGR1 and MEF2C binding levels in cerebellum than PFC (-log10 binomial p value, fold change > 2,  $P < 10^{-5}$ ).

lum (60%) (Figure 4C). We further analyzed the data using DiffBind to identify regions with significantly different level of EGR1/MEF2C binding between PFC and cerebellum (46). In total, DiffBind identified 1026 peaks with higher EGR1 binding in cerebellum than in PFC, and 1563 peaks with higher MEF2C binding in cerebellum than in PFC (fold change > 2,  $P < 10^{-5}$ ). These regions were further analyzed for GO term enrichment analysis using GREAT and linked to walking behavior, motor coordination, and cerebellum morphology and development in the case of EGR1, cerebellar development and limb coordination in the case of MEF2C (Figure 4D). In contrast, very few differential peaks (12 for EGR1 and 1 for MEF2C) were found to have higher binding signal in PFC than in cerebellum and no GO terms were found. Differentially bound peaks for TFs are listed in Supplementary Dataset 2.

## DISCUSSION

ChIP-seq profiling of proteins bound to the genome is generally much more challenging than that of modified histones. Histone ChIP-seq can be conducted with crosslinking and sonication or under native ChIP condition. Due to the robust interaction between histones and genome, native ChIP-seq for histone modifications can have very high efficiency. In contrast, previous ChIP-seq of protein bindings often involves crosslinking and sonication that immobilizes the protein to the interacting DNA sequence before breaking chromatin into fragments. Crosslinking is often considered necessary when TF, Pol II and enzyme interaction with the genome is studied, due to perceived needs and benefits for preservation of such interactions by crosslinking. However, crosslinking and sonication potentially cause epitope masking (70) and damage, respectively, and both affect antibody-antigen interaction critically involved in ChIP assays. Furthermore, crosslinking may also create artifact peaks at highly transcribed regions due to protein-protein crosslinking when carried out for long durations (e.g. 1 h) (71). In comparison, there have also been reports of possible bias for AT-rich regions and open chromatin when MNase is used to fragmentize chromatin (72) although the effect is possibly limited (73).

In this work, by conducting nMOWChIP-seq without crosslinking and sonication, we demonstrate a low-input technology that works with as few as 1,000–50,000 cells per assay for profiling a wide range of protein-genome interactions. We show that nMOWChIP-seq method effectively preserves the interaction between Pol II/TFs/enzyme and the genome. In our approach, the required number of cells depends on the robustness of the protein binding to the genome under native ChIP conditions, the number of binding sites, and the quality of the antibody. Compared to the state-of-the-art ChIP-seq data taken using millions of cells per assay, our datasets generally show very high signal-to-noise ratio and low background, which are characterized by high FRiP values. Although side-by-side comparison is difficult due to the lack of matching cell type, input quantity and antibody, our method appears to be at least comparable to CUT&RUN (31,35) in terms of FRiP and peak numbers. For example, our data yielded 27,000–137,000 peaks with a FRiP of 9–50% using 1,000–50,000 GM12878 cells

on Pol II-S5, and 2,700–16,000 peaks with a FRiP of 3–8% using 5,000–100,000 GM12878 cells on EGR1 (Supplementary Table S1). In comparison, CUT&RUN yielded 23,000 peaks with a FRiP of 20% on Pol II-S5 of K562 cells (35), 16,000–35,000 peaks with a FRiP of 12–27% using 1,000–100,000 K562 cells on CTCF (31), and 6,000–11,000 peaks with a FRiP of 3% using 10,000 melanocytes on SOX10 (GEO dataset GSE172066). The workflow has been published as a detailed protocol with step-by-step guide on setting up and running the assays (17). While it requires certain specialized parts for fluid control, it is possible to set up with help of a technician with some engineering background. nMOWChIP-seq shows some biased signal towards TES when RNA Pol II is profiled. This was not observed with other technologies including conventional crosslinking-based ChIP-seq and CUT&Tag (33). This effect is possibly due to the absence of crosslinking for immobilization with nMOWChIP-seq, although similar observation was not made with CUT&Tag (33).

Compared to histone modification data, ChIP-seq data on Pol II, TFs, enzyme in tissues are very scarce. We applied nMOWChIP-seq to profile Pol II, and key TFs EGR1 and MEF2C in a brain-region-specific manner in cerebellum and prefrontal cortex of mouse brain. We found that Pol II and TF profiles are highly characteristic of the brain regions. The Pol II binding profiles had 4,218 differential peaks between cerebellum and PFC, while EGR1/MEF2C profiles had > 1,000 differential peaks between the two brain regions. The peaks with their intensity high in PFC and low in cerebellum were highly enriched in functions including cognition, learning or memory, while the peaks that were high in cerebellum and low in PFC were enriched in GO terms including walking behavior, motor coordination and cerebellar cortex formation. The fact that the binding profiles of these functional molecules are highly characteristic of the functions of various brain regions indicate that Pol II and the key TFs are critically involved in the molecular dynamics associated with the spatial configuration of brain functions. Our results suggest the possibility of deciphering genome-wide non-histone molecular binding profiles to establish the connections between cells and their functions. In the case of brain cells, the approach potentially allows us to establish the neuroanatomical origin of a brain tissue.

## DATA AVAILABILITY

The ChIP-seq data sets are deposited in the Gene Expression Omnibus (GEO) repository with the following accession number GSE172224.

## SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

## ACKNOWLEDGEMENTS

*Author Contribution:* C.L. designed and supervised the study. Z.L. conducted all nMOWChIP-seq experiments and analyzed the data. L.B.N., Y.Z., C.D., Q. Z., B.Z., Z.Z., M.S. helped with data analysis and experiments. A.M. and H.X. helped with experiments on the TFs. Z.L., L.B.N.

and C.L. wrote the manuscript. All authors proofread the manuscript and provided feedback.

## FUNDING

US National Institutes of Health (NIH) grants R33 CA214176 (C.L.), R01EB017235 (C.L.), R01 CA243249 (C.L.), P30 CA012197 (C.L.), R01NS094574 (H.X.), R21MH120498 (H.X.), and a seed grant from Virginia Tech Institute for Critical Technology and Applied Science (C.L.).

*Competing interests.* C.L. holds a US patent on MOWChIP-seq. The other authors declare no competing interests.

## REFERENCES

- Stormo, G.D. and Zhao, Y. (2010) Determining the specificity of protein-DNA interactions. *Nat. Rev. Genet.*, **11**, 751–760.
- Adelman, K. and Lis, J.T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.*, **13**, 720–731.
- Liu, X., Kraus, W.L. and Bai, X. (2015) Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci.*, **40**, 516–525.
- Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J. *et al.* (2013) Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell*, **153**, 976–987.
- Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G. and Adelman, K. (2015) Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell*, **58**, 311–322.
- Shao, W. and Zeitlinger, J. (2017) Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.*, **49**, 1045–1051.
- Phatnani, H.P. and Greenleaf, A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.*, **20**, 2922–2936.
- Farnham, P.J. (2009) Insights from genomic profiling of transcription factors. *Nat. Rev. Genet.*, **10**, 605–616.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. and Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
- Lambert, S.A., Jolma, A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. and Weirauch, M.T. (2018) The human transcription factors. *Cell*, **172**, 650–665.
- Ropero, S. and Esteller, M. (2007) The role of histone deacetylases (HDACs) in human cancer. *Mol. Oncol.*, **1**, 19–25.
- Guan, J.S., Haggarty, S.J., Giacometti, E., Dannenberg, J.H., Joseph, N., Gao, J., Nieland, T.J., Zhou, Y., Wang, X., Mazitschek, R. *et al.* (2009) HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature*, **459**, 55–60.
- Korzus, E., Rosenfeld, M.G. and Mayford, M. (2004) CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron*, **42**, 961–972.
- Tanaka, S., Ise, W., Inoue, T., Ito, A., Ono, C., Shima, Y., Sakakibara, S., Nakayama, M., Fujii, K., Miura, I. *et al.* (2020) Tet2 and Tet3 in B cells are required to repress CD86 and prevent autoimmunity. *Nat. Immunol.*, **21**, 950–961.
- Cao, Z.N., Chen, C.Y., He, B., Tan, K. and Lu, C. (2015) A microfluidic device for epigenomic profiling using 100 cells. *Nat. Methods*, **12**, 959–962.
- Ma, S., Hsieh, Y., Ma, J. and Lu, C. (2018) Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. *Sci. Adv.*, **4**, eaar8187.
- Zhu, B.H., Hsieh, Y.P., Murphy, T.W., Zhang, Q., Naler, L.B. and Lu, C. (2019) MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nat. Protoc.*, **14**, 3366–3394.
- Furey, T.S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.*, **13**, 840–852.
- Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S. *et al.* (2014) Chromatin state dynamics during blood formation. *Science*, **345**, 943–949.
- Cox, M.C., Deng, C.Y., Naler, L.B., Lu, C. and Verbridge, S.S. (2019) Effects of culture condition on epigenomic profiles of brain tumor cells. *ACS Biomater. Sci. Eng.*, **5**, 1544–1552.
- Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G.Q., Kuan, S., Li, B., Lee, A.Y. *et al.* (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*, **537**, 548–552.
- Zhang, B.J., Zheng, H., Huang, B., Li, W.Z., Xiang, Y.L., Peng, X., Ming, J., Wu, X.T., Zhang, Y., Xu, Q.H. *et al.* (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*, **537**, 553–557.
- Zheng, H., Huang, B., Zhang, B.J., Xiang, Y.L., Du, Z.H., Xu, Q.H., Li, Y.Y., Wang, Q.J., Ma, J., Peng, X. *et al.* (2016) Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Mol. Cell*, **63**, 1066–1079.
- Murphy, T.W., Hsieh, Y.P., Ma, S., Zhu, Y. and Lu, C. (2018) Microfluidic low-input fluidized-bed enabled chip-seq device for automated and parallel analysis of histone modifications. *Anal. Chem.*, **90**, 7666–7674.
- Deng, C., Murphy, T.W., Zhang, Q., Naler, L.B., Xu, A. and Lu, C. (2020) Multiplexed and ultralow-input chip-seq enabled by tagmentation-based indexing and facile microfluidics. *Anal. Chem.*, **92**, 13661–13666.
- Gutin, J., Sadeh, R., Bodenheimer, N., Joseph-Strauss, D., Klein-Brill, A., Alajem, A., Ram, O. and Friedman, N. (2018) Fine-Resolution mapping of TF binding and chromatin interactions. *Cell Rep.*, **22**, 2797–2807.
- Jakobsen, J.S., Bagger, F.O., Hasemann, M.S., Schuster, M.B., Frank, A.-K., Waage, J., Vitting-Seerup, K. and Porse, B.T. (2015) Amplification of pico-scale DNA mediated by bacterial carrier DNA for small-cell-number transcription factor ChIP-seq. *BMC Genomics*, **16**, 46.
- Gutin, J., Sadeh, R., Bodenheimer, N., Joseph-Strauss, D., Klein-Brill, A., Alajem, A., Ram, O. and Friedman, N. (2018) Fine-Resolution mapping of TF binding and chromatin interactions. *Cell Rep.*, **22**, 2797–2807.
- Shankaranarayanan, P., Mendoza-Parra, M.A., Walia, M., Wang, L., Li, N., Trindade, L.M. and Gronemeyer, H. (2011) Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat. Methods*, **8**, 565–567.
- Skene, P.J. and Henikoff, S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, **6**, e21856.
- Skene, P.J., Henikoff, J.G. and Henikoff, S. (2018) Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.*, **13**, 1006–1019.
- Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H. and Ohkawa, Y. (2019) A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.*, **21**, 287–296.
- Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K. and Henikoff, S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.*, **10**, 1930.
- Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K. and Henikoff, S. (2020) Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.*, **15**, 3264–3283.
- Meers, M.P., Bryson, T.D., Henikoff, J.G. and Henikoff, S. (2019) Improved CUT&RUN chromatin profiling tools. *Elife*, **8**, e46314.
- O'Neill, L. (2003) Immunoprecipitation of native chromatin: NChIP. *Methods*, **31**, 76–82.
- Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M. and Henikoff, S. (2011) Epigenome characterization at single base-pair resolution. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 18318–18323.
- Zentner, G.E., Tsukiyama, T. and Henikoff, S. (2013) ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet.*, **9**, e1003317.
- Roca, H. and Franceschi, R.T. (2008) Analysis of transcription factor interactions in osteoblasts using competitive chromatin immunoprecipitation. *Nucleic Acids Res.*, **36**, 1723–1730.

40. Teves, S.S. and Henikoff, S. (2011) Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Gene Dev*, **25**, 2387–2397.
41. Kasinathan, S., Orsi, G.A., Zentner, G.E., Ahmad, K. and Henikoff, S. (2014) High-resolution mapping of transcription factor binding sites on native chromatin. *Nat. Methods*, **11**, 203–209.
42. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
43. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and GenomeProjectData, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
44. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
45. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, **14**, 178–192.
46. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R. *et al.* (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389–393.
47. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
48. Khan, A. and Mathelier, A. (2017) Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinf.*, **18**, 287.
49. Sun, H., Wu, J., Wickramasinghe, P., Pal, S., Gupta, R., Bhattacharyya, A., Agosto-Perez, F.J., Showe, L.C., Huang, T.H. and Davuluri, R.V. (2011) Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids. Res.*, **39**, 190–201.
50. Sun, Z., Xu, X., He, J., Murray, A., Sun, M.A., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X. *et al.* (2019) EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat. Commun.*, **10**, 3892.
51. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P. *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.*, **22**, 1813–1831.
52. Gabrieli, J.D.E., Poldrack, R.A. and Desmond, J.E. (1998) The role of left prefrontal cortex in language and memory. *Proc. Natl. Acad. Sci.*, **95**, 906–913.
53. Miller, E.K., Freedman, D.J. and Wallis, J.D. (2002) The prefrontal cortex: categories, concepts and cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **357**, 1123–1136.
54. Fine, E.J., Ionita, C.C. and Lohr, L. (2002) The history of the development of the cerebellar examination. *Semin. Neurol.*, **22**, 375–384.
55. Caporale, N. and Dan, Y. (2008) Spike timing-dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.*, **31**, 25–46.
56. Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A. and Young, R.A. (2010) c-Myc regulates transcriptional pause release. *Cell*, **141**, 432–445.
57. Ma, S., Charron, J. and Erikson, R.L. (2003) Role of Plk2 (Snk) in mouse development and cell proliferation. *Mol. Cell. Biol.*, **23**, 6936–6943.
58. Funahashi, Y., Ariza, A., Emi, R., Xu, Y., Shan, W., Suzuki, K., Kozawa, S., Ahammad, R.U., Wu, M., Takano, T. *et al.* (2019) Phosphorylation of Npas4 by MAPK regulates reward-related gene expression and behaviors. *Cell Rep.*, **29**, 3235–3252.
59. Zhao, R. and Goldman, I.D. (2013) Folate and thiamine transporters mediated by facilitative carriers (SLC19A1-3 and SLC46A1) and folate receptors. *Mol. Aspects Med.*, **34**, 373–385.
60. Larsen, P.H., DaSilva, A.G., Conant, K. and Yong, V.W. (2006) Myelin formation during development of the CNS is delayed in matrix metalloproteinase-9 and -12 null mice. *J. Neurosci.*, **26**, 2207–2214.
61. Rao, P., Yallapu, M.M., Sari, Y., Fisher, P.B. and Kumar, S. (2015) Designing novel nanoformulations targeting glutamate transporter excitatory amino acid transporter 2: implications in treating drug addiction. *J Pers Nanomed*, **1**, 3–9.
62. Boeckers, T.M., Bockmann, J., Kreutz, M.R. and Gundelfinger, E.D. (2002) ProSAP/Shank proteins – a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *J. Neurochem.*, **81**, 903–910.
63. Haines, B.P., Wheldon, L.M., Summerbell, D., Heath, J.K. and Rigby, P.W.J. (2006) Regulated expression of FLRT genes implies a functional role in the regulation of FGF signalling during mouse development. *Dev. Biol.*, **297**, 14–25.
64. Chakrabarti, K., Lin, R., Schiller, N.I., Wang, Y., Koubi, D., Fan, Y.X., Rudkin, B.B., Johnson, G.R. and Schiller, M.R. (2005) Critical role for Kalirin in nerve growth factor signaling through TrkA. *Mol. Cell. Biol.*, **25**, 5106–5118.
65. Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J. and Yankner, B.A. (2004) Gene regulation and DNA damage in the ageing human brain. *Nature*, **429**, 883–891.
66. Papantrou, A., Danti, F.R., Spaull, R., Leuzzi, V., McTague, A. and Kurian, M.A. (2020) The expanding spectrum of movement disorders in genetic epilepsies. *Dev. Med. Child Neurol.*, **62**, 178–191.
67. Harris, L., Zalucki, O., Clément, O., Fraser, J., Matuzelski, E., Oishi, S., Harvey, T.J., Burne, T.H.J., Heng, J.I.-T., Gronostajski, R.M. *et al.* (2018) Neurogenic differentiation of hippocampal neural stem and progenitor cells is biased by NFIX expression. *Development*, **145**, dev155689.
68. Ali, R.G., Bellchambers, H.M. and Arkell, R.M. (2012) Zinc fingers of the cerebellum (Zic): transcription factors and co-factors. *Int. J. Biochem. Cell Biol.*, **44**, 2065–2068.
69. Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W.B. and Millen, K.J. (2004) Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat. Genet.*, **36**, 1053–1055.
70. Kidder, B.L., Hu, G. and Zhao, K. (2011) ChIP-Seq: technical considerations for obtaining high-quality data. *Nat. Immunol.*, **12**, 918–922.
71. Teytelman, L., Thurtle, D.M., Rine, J. and van Oudenaarden, A. (2013) Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 18602–18607.
72. Chereji, R.V., Bryson, T.D. and Henikoff, S. (2019) Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.*, **20**, 198.
73. Allan, J., Fraser, R.M., Owen-Hughes, T. and Keszenman-Pereyra, D. (2012) Micrococcal nuclease does not substantially bias nucleosome mapping. *J. Mol. Biol.*, **417**, 152–164.