Native MOWChIP-seq: Genome-wide profiles of key protein bindings reveal functional differences among various brain regions

Zhengzhi Liu¹, Lynette B. Naler², Yan Zhu², Chengyu Deng², Qiang Zhang², Bohan Zhu², Zirui Zhou², Mimosa Sarma², Alexander Murray³, Hehuang Xie³, Chang Lu^{2,*}

¹ Department of Biomedical Engineering and Mechanics, Virginia Tech, Blacksburg, VA

² Department of Chemical Engineering, Virginia Tech, Blacksburg, VA

³ Department of Biomedical Sciences & Pathobiology, Virginia Tech, Blacksburg, VA

Corresponding author: changlu@vt.edu

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, termed native MOWChIP-seq, produces high-quality binding profiles using 1000-50,000 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 expression¹. 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 Similarly, transcription factors (TFs) bind to DNA or cofactors and participate in gene regulations in a sequence-specific manner.^{1,7} TFs control key aspects of cellular biology including cell differentiation, development patterning, and immune response.^{8,9} 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 activation¹⁰. For example, HDAC2 binding and activity are involved in memory formation and synaptic plasticity in brain¹¹. Genome-wide binding profiles of these proteins are highly indicative of genes being 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)¹²⁻¹⁴, 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 ($>10^7$ 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.¹⁵ These limitations makes it impractical to examine animal and human tissue samples available only in small quantities. In recent years, significant efforts have been made in developing low-input ChIP-seq methods^{12-14,16-22}. However, the vast majority of these methods were only been demonstrated on examining histone modifications.

As an alternative to ChIP, IP-free technologies such as CUT&RUN^{23,24} and ChIL-seq²⁵ 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^{23,24}. The latest variation of CUT&RUN (CUT&Tag)²⁶ have been used to profile single cells. However, CUT&RUN requires immobilization of cells to bead surface via concanavalin A and glycolipid interaction and 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^{23,24}. CUT&RUN datasets may also be contaminated by DNA sequences that are from unknown sources associated with the process (e.g. sequences containing (TA)_n)²⁷.

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 modifications²⁸ and also non-histone proteins including pol II and TFs²⁹⁻³³. However, these studies were all conducted using a large number of cells (10^7 - 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^{12,14}) 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 facilitating low-input profiling. 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.

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^{34,35}. However, our initial results showed that even using our low-input MOWChIP-seq technology which has proven high efficiency for collecting DNA-protein complexes^{12,14}, low-quality results were yielded when 100,000 cells were crosslinked and sonicated to create the chromatin fragments before Pol II ChIP (Supplementary Fig. S1). 54,000 and 17,000 peaks were yielded in the two technical replicates, compared to 115,000 and 141,000 peaks generated by nMOWChIP-seq using 10,000 cells per assay. Crosslinking and subsequent sonication may damage RNA Pol II's ability to form antigen-antibody complex and seriously limits the ChIP's efficiency for collecting targeted fragments.

We then applied nMOWChIP-seq to profile protein binding to the genome. MNase digestion (or native ChIP) had primarily been applied to studies associated with histone modifications where the strong interaction between genomic DNA and histone guarantees the pulldown of DNA without crosslinking^{13,23,36}. In our process (Fig. 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^{12,14}. 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 cell line GM12878, we show that our method can generate high quality ChIP-seq data with input as little as 1,000 cells per assay for Pol II (phospho S5) (Fig. 1B). 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, 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 137,000, 128,000, 65,000 and 27,000 peaks with 50,000, 10,000, 2,000 and 1,000 cells per assay, respectively, compared to 110,000 peaks from the ENCODE data obtained using 20 million cells. 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 ENCODE³⁷.

We also determined that the binding of Pol II was stable enough to be preserved at -80 °C, a unique property not observed in histone modifications or TFs (Supplementary Figure S2). From our testing, MNase-digested chromatin can be frozen at -80 °C or on dry ice for 2 d without substantial degradation in data quality for Pol II.

We profiled a transcription factor, early growth response protein 1 (EGR1), using as few as 5,000 GM12878 cells (Fig. 1B and 1C, 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 (Fig. 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 S3).

Finally, we applied nMOWChIP-seq to examine the binding of histone deacetylase HDAC2 in GM12878 cells (Fig. 1B and 1C, 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%). We also examined HDAC2 binding in mouse brain cells (Supplementary Fig. S4). 100,000 mixed nuclei and 800,000 FACS-sorted NeuN+ neuronal nuclei were used in each assay and an average of 3193 (FRiP ~13.1%) and 2112 (FRiP ~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 phosphor-5 activated 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 show the distinction (Fig. 2). The two profiles on GM12878 cells are similar in a large fraction of the genome (Fig. 2A). However, some genes (ACTG1 and MDM2 as examples) with Pol II binding over the entire gene body show that Pol II-total profile captures the initial binding of unphosphorylated Pol II at TSS, which is absent in Pol II-S5 profile in comparison (Fig. 2B). 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) (Fig. 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 (Fig. 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,^{38,39} while cerebellum controls motor functions and coordination.⁴⁰ 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 (Fig. 3). The genome-wide Pol II profiles were substantially different between cerebellum and PFC with Pearson's correlation being 0.67, compared to an average of 0.96 between technical replicates (Fig. 3A and 3B). 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, $p < 10^{-5}$). 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 (Fig. 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).⁴¹ 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) (Fig. 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^{2,42}. 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,⁴³ and *Npas4*, which is involved in regulating reward-related learning and memory,⁴⁴ both showed dramatically higher PI values in cerebellum than in PFC (Fig. 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.⁴⁵ Furthermore, we analyzed the distribution of the PI in PFC and cerebellum (Fig. 3E). 36% of genes were being actively expressed in PFC ($PI < 20$), compared to only 25% in cerebellum. Mouse brain displays more pausing and less active transcription compared to GM12878 (52% actively expressed genes, Supplementary Fig. S5). This was also within our expectation because GM12878 cell line was maintained in log phase and actively dividing, unlike brain cells in adult mice. 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,⁴⁶ and *Slc1a2* encodes excitatory amino acid transporter 2, responsible for reuptake of 90% glutamate in CNS.⁴⁷ *Shank3* belongs to the Shank

gene family that plays a role in synapse formation,⁴⁸ and Flrt2 encodes a member of the FLRT protein family that is shown to regulate signaling during mouse development.⁴⁹

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 (Fig. 4A). Kalrn (Kalirin) plays important roles in nerve growth⁵⁰. Higher EGR1/MEF2C activity in PFC was observed on Gria1 (Glutamate receptor 1) which is involved in synaptic transmission.⁵¹ Cacna1a is involved in movement disorder⁵² and expression of Nfix can influence neural stem cell differentiation.⁵³ Zic1 and Zic4 belong to the family of Zinc finger of the cerebellum (ZIC) protein family,⁵⁴ whose loss of function can lead to Dandy-Walker malformation and incomplete cerebellar vermis⁵⁵. Higher EGR1 binding on these four genes were seen in cerebellum than in PFC. A large fraction of EGR1 and MEF2C peaks (68-89%) appear to overlap with Pol II peaks, in both cerebellum and PFC (Supplementary Fig. S6). We observed correlated EGR1 and MEF2C profiles, with an average of Pearson's correlation of 0.74 between the two TFs in PFC and 0.84 in cerebellum (Fig. 4B). Such correlations were much higher than the one observed in GM12878 cells ($r \sim 0.52$). On the other hand, EGR1 and MEF2C present 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 (60%) (Fig. 4C). We further analyzed the data using DiffBind to identify regions with

significantly different level of EGR1/MEF2C binding between PFC and cerebellum.⁵⁶ 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.

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 masking⁵⁷ 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 crosslinking⁵⁸.

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. Our method is also comparable to CUT&RUN^{24,27} in terms of data quality.

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 molecular binding profiles to establish the neuroanatomical origin of brain tissues for the first time.

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% CO₂. 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 CO₂ followed by cervical dislocation. Mouse brains were rapidly dissected, frozen on dry ice and stored at -80°C. This study was approved 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 are 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 CaCl₂, 3 mM Mg(Ac)₂, 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 MOWChIP 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 (~1.2×10⁵/ml). The nuclei were re-concentrated by adding 200 µl of 1.8 M sucrose, 5 µl of 1M CaCl₂ and 3 µl of 1M Mg(Ac)₂ 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.

MNase digestion of chromatin

This protocol is scalable in volume and can digest cell/nuclei suspension with concentration up to 4×10^6 /ml. Our experiments usually started with 4×10^5 cells/nuclei suspended in 100 μ l of DPBS. 1 μ l of PIC, 1 μ l of 100 mM PMSF and 100 μ l lysis buffer [4% Triton X-100, 100 mM Tris-HCl, 100 mM NaCl, and 30 mM MgCl₂] were added, mixed by vortexing and incubated at room temperature for 10 min. 10 μ l of 100 mM CaCl₂ and 2.5 μ l 100U MNase (88216, Thermo Fisher Scientific) were added, mixed by vortexing and incubated at room temperature for 10 min. 22 μ l 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.

Preparation of immunoprecipitation (IP) beads

5 μ 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 μ l of IP buffer. 1 μ g of Pol II or TF antibody was added into the bead suspension for each assay using 10^5 cells/nuclei, while 0.5 μ g of antibody was added for each assay using 5×10^4 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-EGR1 (sc-101033, lot A1618, Santa Cruz), anti-MEF2C (sc-365862, lot B0818, Santa Cruz), anti-HDAC2 (ab124974, lot GR97402-7, Santa Cruz). 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 μ 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 protocol¹⁴.

Purification of ChIP DNA

After MOWChIP, IP beads were rinsed once with IP buffer and resuspended in 200 μ l of DNA elution buffer (10 mM Tris-HCl, 50 mM NaCl, 10 mM EDTA, and 0.03% SDS). 2 μ 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 μ 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 μ 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! with default settings. Reads that passed the quality check were mapped to reference genomes hg38 (human) or mm10 (mouse) with Bowtie.⁵⁹ Uniquely mapped reads were filtered for known blacklisted genome regions using Samtools⁶⁰ and bedtools⁶¹ 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.⁶² Peak calling was conducted using MACS2 (default settings). Differential binding analysis was carried out using DiffBind⁵⁶ with default settings. Selected genomic regions of interest were further analyzed for gene ontology terms using GREAT.⁶³ 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.

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.

Data Availability

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

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE172224>

Acknowledgements

This work was supported by 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.).

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.

Competing interests

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

References

- 1 Stormo, G. D. & Zhao, Y. Determining the specificity of protein-DNA interactions. *Nat Rev Genet* **11**, 751-760, (2010).
- 2 Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-731, (2012).
- 3 Liu, X., Kraus, W. L. & Bai, X. Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem Sci* **40**, 516-525, (2015).
- 4 Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D. S., Zeitlinger, J. & Levine, M. S. Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* **153**, 976-987, (2013).
- 5 Williams, L. H., Fromm, G., Gokey, N. G., Henriques, T., Muse, G. W., Burkholder, A., Fargo, D. C., Hu, G. & Adelman, K. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol Cell* **58**, 311-322, (2015).
- 6 Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* **49**, 1045-1051, (2017).
- 7 Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-263, (2009).
- 8 Farnham, P. J. Insights from genomic profiling of transcription factors. *Nat Rev Genet* **10**, 605-616, (2009).
- 9 Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R. & Weirauch, M. T. The Human Transcription Factors. *Cell* **172**, 650-665, (2018).
- 10 Roper, S. & Esteller, M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol* **1**, 19-25, (2007).
- 11 Guan, J. S., Haggarty, S. J., Giacometti, E., Dannenberg, J. H., Joseph, N., Gao, J., Nieland, T. J., Zhou, Y., Wang, X., Mazitschek, R., Bradner, J. E., DePinho, R. A., Jaenisch, R. & Tsai, L. H. HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature* **459**, 55-60, (2009).
- 12 Cao, Z. N., Chen, C. Y., He, B., Tan, K. & Lu, C. A microfluidic device for epigenomic profiling using 100 cells. *Nature Methods* **12**, 959-962, (2015).
- 13 Ma, S., Hsieh, Y., Ma, J. & Lu, C. Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. *Sci Adv* **4**, eaar8187, (2018).
- 14 Zhu, B. H., Hsieh, Y. P., Murphy, T. W., Zhang, Q., Naler, L. B. & Lu, C. MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nature Protocols* **14**, 3366-3394, (2019).
- 15 Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* **13**, 840-852, (2012).
- 16 Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D. A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., Friedman, N. & Amit, I. Chromatin state dynamics during blood formation. *Science* **345**, 943-949, (2014).
- 17 Cox, M. C., Deng, C. Y., Naler, L. B., Lu, C. & Verbridge, S. S. Effects of Culture Condition on Epigenomic Profiles of Brain Tumor Cells. *Acs Biomater Sci Eng* **5**, 1544-1552, (2019).
- 18 Dahl, J. A., Jung, I., Aanes, H., Greggains, G. D., Manaf, A., Lerdrup, M., Li, G. Q., Kuan, S., Li, B., Lee, A. Y., Preissl, S., Jermstad, I., Haugen, M. H., Suganthan, R., Bjoras, M., Hansen, K., Dalen, K. T., Fedorcsak, P., Ren, B. & Klungland, A. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552, (2016).
- 19 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., Liu, W. Q., Kou, X. C., Zhao, Y. H., He, W. T., Li, C., Chen, B., Li, Y. Y., Wang, Q. J., Ma, J., Yin, Q. Z., Kee, K., Meng, A. M., Gao, S. R., Xu, F., Na, J. & Xie, W. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557, (2016).

- 20 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., Xu, F. & Xie, W. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell* **63**, 1066-1079, (2016).
- 21 Murphy, T. W., Hsieh, Y. P., Ma, S., Zhu, Y. & Lu, C. Microfluidic Low-Input Fluidized-Bed Enabled ChIP-seq Device for Automated and Parallel Analysis of Histone Modifications. *Anal Chem* **90**, 7666-7674, (2018).
- 22 Deng, C., Murphy, T. W., Zhang, Q., Naler, L. B., Xu, A. & Lu, C. Multiplexed and Ultralow-Input ChIP-seq Enabled by Tagmentation-Based Indexing and Facile Microfluidics. *Anal Chem* **92**, 13661-13666, (2020).
- 23 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).
- 24 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019, (2018).
- 25 Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H. & Ohkawa, Y. A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat Cell Biol* **21**, 287-296, (2019).
- 26 Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K. & Henikoff, S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930, (2019).
- 27 Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**, (2019).
- 28 O'Neill, L. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76-82, (2003).
- 29 Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M. & Henikoff, S. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A* **108**, 18318-18323, (2011).
- 30 Zentner, G. E., Tsukiyama, T. & Henikoff, S. ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet* **9**, e1003317, (2013).
- 31 Roca, H. & Franceschi, R. T. Analysis of transcription factor interactions in osteoblasts using competitive chromatin immunoprecipitation. *Nucleic Acids Research* **36**, 1723-1730, (2008).
- 32 Teves, S. S. & Henikoff, S. Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Gene Dev* **25**, 2387-2397, (2011).
- 33 Kasinathan, S., Orsi, G. A., Zentner, G. E., Ahmad, K. & Henikoff, S. High-resolution mapping of transcription factor binding sites on native chromatin. *Nat Methods* **11**, 203-209, (2014).
- 34 Sun, H., Wu, J., Wickramasinghe, P., Pal, S., Gupta, R., Bhattacharyya, A., Agosto-Perez, F. J., Showe, L. C., Huang, T. H. & Davuluri, R. V. Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids Res* **39**, 190-201, (2011).
- 35 Sun, Z., Xu, X., He, J., Murray, A., Sun, M. A., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X., Li, L., Zhu, J., Chen, J., Morozov, A., Pickrell, A. M., Theus, M. H. & Xie, H. EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat Commun* **10**, 3892, (2019).
- 36 Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M. M. & Lorincz, M. C. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun* **6**, 6033, (2015).
- 37 Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglu, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., Iyer, V. R., Jung, Y. L., Karmakar, S., Kellis, M., Kharchenko, P. V., Li, Q., Liu, T., Liu, X. S., Ma, L., Milosavljevic, A., Myers, R. M., Park, P. J., Pazin, M. J., Perry, M. D., Raha, D., Reddy, T. E., Rozowsky, J., Shores,

- N., Sidow, A., Slaterry, M., Stamatoyannopoulos, J. A., Tolstorukov, M. Y., White, K. P., Xi, S., Farnham, P. J., Lieb, J. D., Wold, B. J. & Snyder, M. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831, (2012).
- 38 Gabrieli, J. D. E., Poldrack, R. A. & Desmond, J. E. The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences* **95**, 906, (1998).
- 39 Miller, E. K., Freedman, D. J. & Wallis, J. D. The prefrontal cortex: categories, concepts and cognition. *Philos Trans R Soc Lond B Biol Sci* **357**, 1123-1136, (2002).
- 40 Fine, E. J., Ionita, C. C. & Lohr, L. The History of the Development of the Cerebellar Examination. *Semin Neurol* **22**, 375-384, (2002).
- 41 Caporale, N. & Dan, Y. Spike timing-dependent plasticity: a Hebbian learning rule. *Annu Rev Neurosci* **31**, 25-46, (2008).
- 42 Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A. & Young, R. A. c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445, (2010).
- 43 Ma, S., Charron, J. & Erikson, R. L. Role of Plk2 (Snk) in mouse development and cell proliferation. *Mol Cell Biol* **23**, 6936-6943, (2003).
- 44 Funahashi, Y., Ariza, A., Emi, R., Xu, Y., Shan, W., Suzuki, K., Kozawa, S., Ahammad, R. U., Wu, M., Takano, T., Yura, Y., Kuroda, K., Nagai, T., Amano, M., Yamada, K. & Kaibuchi, K. Phosphorylation of Npas4 by MAPK Regulates Reward-Related Gene Expression and Behaviors. *Cell Rep* **29**, 3235-3252 e3239, (2019).
- 45 Zhao, R. & Goldman, I. D. Folate and thiamine transporters mediated by facilitative carriers (SLC19A1-3 and SLC46A1) and folate receptors. *Mol Aspects Med* **34**, 373-385, (2013).
- 46 Larsen, P. H., DaSilva, A. G., Conant, K. & Yong, V. W. Myelin formation during development of the CNS is delayed in matrix metalloproteinase-9 and -12 null mice. *J Neurosci* **26**, 2207-2214, (2006).
- 47 Rao, P., Yallapu, M. M., Sari, Y., Fisher, P. B. & Kumar, S. Designing Novel Nanoformulations Targeting Glutamate Transporter Excitatory Amino Acid Transporter 2: Implications in Treating Drug Addiction. *J Pers Nanomed* **1**, 3-9, (2015).
- 48 Boeckers, T. M., Bockmann, J., Kreutz, M. R. & Gundelfinger, E. D. ProSAP/Shank proteins – a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *Journal of Neurochemistry* **81**, 903-910, (2002).
- 49 Haines, B. P., Wheldon, L. M., Summerbell, D., Heath, J. K. & Rigby, P. W. J. Regulated expression of FLRT genes implies a functional role in the regulation of FGF signalling during mouse development. *Developmental Biology* **297**, 14-25, (2006).
- 50 Chakrabarti, K., Lin, R., Schiller, N. I., Wang, Y., Koubi, D., Fan, Y. X., Rudkin, B. B., Johnson, G. R. & Schiller, M. R. Critical role for Kalirin in nerve growth factor signaling through TrkA. *Mol Cell Biol* **25**, 5106-5118, (2005).
- 51 Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J. & Yankner, B. A. Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883-891, (2004).
- 52 Papandreou, A., Danti, F. R., Spaul, R., Leuzzi, V., McTague, A. & Kurian, M. A. The expanding spectrum of movement disorders in genetic epilepsies. *Dev Med Child Neurol* **62**, 178-191, (2020).
- 53 Harris, L., Zalucki, O., Clement, O., Fraser, J., Matuzelski, E., Oishi, S., Harvey, T. J., Burne, T. H. J., Heng, J. I., Gronostajski, R. M. & Piper, M. Neurogenic differentiation by hippocampal neural stem and progenitor cells is biased by NFIX expression. *Development* **145**, (2018).
- 54 Ali, R. G., Bellchambers, H. M. & Arkell, R. M. Zinc fingers of the cerebellum (Zic): transcription factors and co-factors. *Int J Biochem Cell Biol* **44**, 2065-2068, (2012).

- 55 Grinberg, I., Northrup, H., Ardinger, H., Prasad, C., Dobyns, W. B. & Millen, K. J. Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat Genet* **36**, 1053-1055, (2004).
- 56 Stark, R. & Brown, G. DiffBind: differential binding analysis of ChIP-seq peak data. (2019).
- 57 Kidder, B. L., Hu, G. & Zhao, K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol* **12**, 918-922, (2011).
- 58 Teytelman, L., Thurtle, D. M., Rine, J. & van Oudenaarden, A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A* **110**, 18602-18607, (2013).
- 59 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, (2009).
- 60 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, (2009).
- 61 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, (2010).
- 62 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, (2013).
- 63 McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. & Bejerano, G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, (2010).

Figures and captions

Fig 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. (B) Normalized RNA Pol II (S5) 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.

Fig 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.

Fig 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, $p < 10^{-5}$). (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.

Fig 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 ($-\log_{10}$ binomial p value, fold change > 2, $p < 10^{-5}$).