

Microfluidics for low input epigenomic analysis and application to oncology and brain neuroscience

Zhengzhi Liu

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State

University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Biomedical Engineering

Chang Lu, Chair

Rafael Davalos

Liwu Li

Zhen Bouman Chen

Scott Verbridge

May 22, 2023

Blacksburg, VA

Keywords: microfluidics, epigenome, chromatin immunoprecipitation, next generation sequencing, histone modification, RNA polymerase II, transcription factor, chromatin isolation by RNA purification, epigenomic tomography, non-small cell lung cancer

Microfluidics for low input epigenomic analysis and application to oncology and brain neuroscience

Zhengzhi Liu

Abstract

Microfluidics is a versatile tool with many applications in biology. Its ability to manipulate small volumes of liquid precisely has led to the development of many microfluidic assay platforms. They could handle small amounts of samples and carry out analysis with high sensitivity and throughput. Microfluidic assays have provided new insights into scarce biological samples at higher resolution. In this thesis, we developed microfluidic tools to conduct low input ChIP-seq and ChIRP-seq. We applied them to a variety of samples profiling different targets. The native MOWChIP-seq platform was developed to map RNA polymerase II, transcription factors and histone deacetylase binding in 1,000-50,000 cells. We examined mouse prefrontal cortex and cerebellum using this technology. We found extensive differences that correlated with distinct neurological functions of the brain regions. The same platform and workflow were used to profile five key histone modifications in human lung tumor and normal tissue samples. Integrative analysis with gene expression data revealed extensive chromatin remodeling in lung tumor. Spatial histone modification mapping was conducted in mouse neocortex in a similar fashion. We generated an epigenomic tomography that demonstrated the molecular state of the brain in 3D. Lastly, we developed a microfluidic version of the ChIRP-seq process which successfully conducted the assay using only 500K cells. This improvement makes ChIRP-seq in tissue samples feasible.

Microfluidics for low input epigenomic analysis and application to oncology and brain neuroscience

Zhengzhi Liu

General audience abstract

Microfluidics is a type of technology that can control small volumes of liquid in a miniature system. It can carry out reactions on very small scales with higher precision and sensitivity than conventional methods. Microfluidics has found many uses in the field of biology, especially dealing with samples available in limited quantities. These low input microfluidic platforms have helped researchers gain new knowledge on many complex questions. In this thesis, we developed microfluidic tools to carry out low input ChIP-seq and ChIRP-seq. These are two established techniques used to map where certain targets are located on the genome of an organism. These targets include specific chemical modifications to the wrapper protein of DNA (histone modification), proteins that take part in transcription and expression of genes (RNA polymerase II, transcription factors) and other molecules. Our nMOWChIP-seq system removed the need for fixation by chemicals. It was able to examine RNA polymerase II, transcription factors and other enzymes using 1,000-50,000 cells. Traditional ChIP-seq requires more than 10 million cells and time-consuming chemical treatment steps. Our technology greatly improved sensitivity and ease of use. We also used this platform to test five important histone modifications in human lung tumors and healthy tissues. We constructed a spatial map of histone modification in mouse brain by analyzing slices of the cortex. Finally, we developed a microfluidic version of ChIRP-seq process to map locations of long non-coding

RNAs in cultured human cells. The cells needed for a successful test were reduced to 500K from 20 million of the original workflow.

Acknowledgements

Graduate school has been an exciting and rewarding experience for me. This journey was made possible by the help and guidance I received along the way. I am grateful for the many people who played different roles in the past six years. I could not have arrived at this destination without them.

First, I must thank my advisor, Dr. Chang Lu, for the great work he has done. He accepted me into the group directly after a phone interview across the ocean and has been fully supportive since. He is a keen researcher who always has a clear picture of the work in the field. He is a kind mentor who has taught me important lessons not only about research, but also about being a better person. I appreciated his wisdom, humor, and straightforward way in our many discussions over the years on topics of work and life. Above all, he instilled the idea to never discount a new idea at face value, to always make a trial worthwhile, and to extract maximum information from the result no matter if it's a success or failure. A failure is never worthless if one can learn from it and use it to move one step closer to success. These are both guidelines for research and life mottos to live by.

I want to thank my committee members, Dr. Rafael Davalos, Dr. Liwu Li, Dr. Zhen Bouman Chen, and Dr. Scott Verbridge. They offered valuable suggestions for my work and helped me see its strengths and shortcomings in fresh perspectives. Their support during my preliminary exam and final defense made the processes much more enjoyable. Additionally, Dr. Chen has been a long-term collaborator on our ChIRP-seq work. I also worked with Dr. Hehuang Xie, Dr. Wei Wang, Dr. Lina Zheng, Dr. Xiaoting Jia, and Dr. Shan Jiang closely on my other projects, whose expertise in their fields made my research work possible.

I am grateful to have the support of two Virginia Tech departments: the Department of Chemical Engineering where I worked, and the School of Biomedical Engineering and Sciences, where I was a student of. Staff members on both sides have helped me deal with the issues that came up along the way. My fellow grad students were a great source of joy and encouragement, especially in my first two years here.

I would like to thank my lab mates over the years: Sai Ma, Yan Zhu, Travis Murphy, Mimosa Sarma, Chengyu Deng, Yuan-Pang Hsieh, Qiang Zhang, Bohan Zhu, Lynette Naler, Zirui Zhou, Tom Hadlock, Gaoshan Li, Jenna Catalano, Jacob Neice, and Xing Zhang. They made up the collaborative and happy lab I worked in for the past six years. In particular, I wish to thank Yan for showing me how to work with mice, Travis for training me to be the lab manager, Mimosa and Lynette for the guidance and fun conversations, Yuan-Pang, Qiang and Bohan for training me on the key techniques in the lab. Our group was an effective team to work in and a fun family to hang out with.

Finally, I must thank my family for their continued support. My parents, Caijin Liu and Fengya Hu, always have my back and help me in every way they can. Their understanding and support never changed over the six years even when I was unable to visit home once, first due to work, then international tensions and finally the pandemic. My wife, Dr. Suzanne Laliberte, gave me her unconditional love and care, without which I would not have been able to achieve my goals. She made me a member of the new family here in the and I will be thankful forever. Lastly, I would like to thank my cat, Chips (full name MEOW CHIP), for showing me that there are few problems in life which a good nap and a full belly will not fix.

Table of contents

Abstract	ii
General audience abstract.....	iii
Acknowledgements.....	v
Table of contents	vii
List of abbreviations.....	ix
Chapter 1: Overview	1
Chapter 2: Native MOWChIP-seq for profiling RNA polymerase II, transcription factors and histone deacetylases.....	7
Introduction	7
Results and discussion.....	12
Materials and methods	26
Chapter 3: Comprehensive histone modification and gene expression profiling in human lung tumor samples	34
Introduction	34
Results and discussion.....	38
Materials and methods	51

Chapter 4: Spatially resolved epigenetic mapping of mouse neocortex.....	58
Introduction	58
Results and discussion.....	61
Materials and methods	80
Chapter 5: ChIRP-seq profiling of lncRNA binding in cell lines.....	86
Introduction	86
Results and discussion.....	88
Materials and methods	92
Chapter 6: Summary and outlook	98
References	101
Publications	125
Appendix.....	127
Primers used in this thesis.....	127
Custom scripts used in this thesis	130

List of abbreviations

ncRNA	Non-coding RNA
Pol II	RNA polymerase II
TF	Transcription factor
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
lncRNA	Long non-coding RNA
Ser5P	Serine 5-phosphorylation
ChIP-seq	Chromatin immunoprecipitation coupled with sequencing
MNase	Micrococcal nuclease
MOWChIP	Microfluidic oscillatory washing–based chromatin immunoprecipitation
ATAC-seq	Assay for transposase-accessible chromatin coupled with sequencing
PFC	Prefrontal cortex
ChIRP-seq	Chromatin isolation by RNA purification coupled with sequencing
ENCODE	Encyclopedia of DNA Elements
HDAC2	Histone deacetylase 2
CUT&RUN	Cleavage under targets and release using nuclease
EGR1	Early growth response factor 1
MEF2C	Myocyte-specific enhancer factor 2C
FRiP	Fraction of reads in peaks
TSS	Transcription start site
ROC	Receiver operating characteristic

AUC	Area under curve
MCC	Matthew's correlation coefficient
GO	Gene ontology
TES	Transcription end site
PI	Pausing index
CNS	Central nervous system
PDMS	Polydimethylsiloxane
DMER	Differentially modified epigenomic region
DEG	Differentially modified gene
RAM	Regulation associated module
qPCR	Quantitative polymerase chain reaction
HOTAIR	HOX transcript antisense RNA
GFP	Green fluorescent protein
TERC	Telomerase RNA component
FACS	Fluorescence-activated cell sorting
PIC	Protein inhibitor cocktail
PMSF	Phenylmethylsulfonyl fluoride
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(β -aminoethyl ether)-N,N,N',N'-tetraacetic acid
RFU	Relative fluorescence unit
SDS	Sodium dodecyl sulfate

Tris	Tris(hydroxymethyl)aminomethane
TE	Tris-EDTA

Chapter 1: Overview

Gene expression can change without the DNA sequence being altered, through the process of epigenetics. Epigenetic changes occur naturally in organisms throughout their life cycle. They control key biological processes such as cell differentiation, development, and diseases.

Various environmental factors affect epigenetic regulation. There are three main categories of epigenetic changes: histone modification, DNA methylation and non-coding RNA (ncRNA).

They exert effects through various mechanisms that control the transcription level of protein-coding genes. RNA polymerase II (Pol II) transcribes genes into messenger RNA, which are then further coded into functional products following the central dogma of molecular biology.

Various proteins such as transcription factors (TFs) take part in this process and further regulate the transcription process. Enzymes including histone acetyl transferases (HATs) and histone deacetylases (HDACs) regulate histone modification levels which in turn affect the transcription of protein-coding genes. Some of the non-coding regions of the genome also transcribe into ncRNA. They do not have protein products yet serve varied biological functions.

At the start of this project, our lab had developed the first generation of microfluidic oscillatory washing-based chromatin immunoprecipitation (MOWChIP-seq). The traditional chromatin immunoprecipitation (ChIP) method relies on free binding between antibody and chromatin fragments in a set volume of suspension inside a tube. Free binding limits the efficiency of target chromatin capture. Due to this, traditional ChIP requires a very large amount of input material to generate enough ChIP DNA. MOWChIP addresses this shortcoming by employing a microfluidic chamber where magnetic beads carrying antibody are packed into a dense bed at the exit. Chromatin sample is forced to go through the packed bed of beads in its entirety at a slow flow rate. This shortens diffusion distance and significantly increases the efficiency of

target capture. We demonstrated capture of histone modification approaching the theoretical limit, and successful ChIP-seq assays from 100 cells.¹

While histone modifications are the common focus for ChIP-seq assays, other non-histone targets including Pol II, TFs and enzymes are also of interest. They have a lower occurrence on the genome compared to histone modification and are more difficult to profile. After the success with histone modification using MOWChIP, we turned our attention to low input profiling of Pol II and TFs. However, initial trials using the same process of chemical crosslinking and sonication to prepare chromatin proved inadequate. MOWChIP maintained its high capture efficiency, but the datasets had low reproducibility and high noise levels. At this time, we started employing MNase to digest chromatin in other ongoing projects to very consistent results. We believed crosslinking and sonication was the issue for MOWChIP dealing with non-histone targets. Thus, we tested MNase digestion of native chromatin followed by MOWChIP. This resulted in much higher data quality comparable to published standards. The results reinforced our hypothesis that chemical fixation was not necessary for ChIP-seq assays when a gentle fragmentation method like MNase digestion was used. We tested with cell line samples and found the practical limit to be 1,000 cells for Pol II MOWChIP-seq. Our collaborators from Dr. Rong Li's group at GWU followed our progress on Pol II closely and sent us both treated human cell line and mouse blood cell samples. We profiled Pol II binding in these samples using the native MOWChIP-seq workflow. We generated good quality data that captured the varying states of Pol II binding in different conditions. We further tested mouse brain PFC and cerebellum and achieved good results. The difference in Pol II binding was clearly shown in the datasets and reflected the distinct functions of the brain regions. We next moved onto testing EGR1 and MEF2C, two TFs that participate in brain development, at

the suggestion of Dr. Hehuang Xie. Cell line trials revealed the practical limit at 5,000 cells per assay. We found that compared to Pol II, TF binding was more transient and introduced more variance to data quality. Native MOWChIP-seq was able to distinguish between mouse PFC and cerebellum and capture their functional differences. We found EGR1 and MEF2C to be closely associated with each other and Pol II in the samples tested. Finally, we tested the histone modifying enzyme HDAC2, which proved to be the most difficult. The reproducible limit was found to be 50,000 cells.

After establishing the MOWChIP-seq technology and later optimizing it for multiple parallel runs,² we sought to apply it to tissue samples derived from patients and lab animals. Our collaboration project with Dr. Wei Wang's group at UCSD was aimed precisely at this goal: make use of MOWChIP-seq's capability in low input assays to examine multiple histone modifications in clinical tumor samples. Frozen lung tissue samples, both tumor and non-neoplastic, were acquired from Moores Cancer Center at UCSD. We set out to examine five key histone modifications in parallel using MOWChIP-seq, and gene expression via RNA-seq. We adapted the nuclei extraction protocol previously used for brain samples to the lung samples. Less than 15 mg of tissue per sample was needed to conduct all MOWChIP-seq assays for each tissue. In addition, we found out that lung tissue is more delicate in the context of RNA-seq and came up with a suitable process to preserve its mRNA using RNAlater-ICE, followed by direct RNA extraction from tissue slices. After initial optimization, we were able to process samples at a steady rate of one tissue per day, completing nuclei extraction, digestion and MOWChIP of five histone modifications (ten assays). Data analysis showed extensive changes in histone modification and gene expression levels between normal and tumor tissues. Integrated analysis of the six different types of data revealed systemic chromatin

reorganization brought about by tumor and distinct clusters of genomic regions marked by their specific co-modification patterns. The application of the tested and matured MOWChIP-seq technology enabled us to profile five major histone modifications simultaneously. Two of the five marks were repressive ones that provide information outside the typical open chromatin regions mapped by similar studies. The high-quality dataset would be a valuable resource for epigenetic studies into lung cancer and proved the power of MOWChIP-seq in dealing with scarce patient derived tissue samples difficult for conventional methods.

Aside from cancer, brain and neuroscience have been another major focus of our lab. We have applied low input ChIP-seq and reduced representation bisulfite sequencing to investigate different brain regions.^{3,4} We have also examined the effect of psychedelic exposure on the brain using MOWChIP-seq.⁵ Mouse brain sampling, nuclei extraction, FACS sorting and MOWChIP-seq had become a combined routine pipeline in our lab. This enabled us to carry out epigenetic profiling of selected cell/nuclei population from very small amounts of sample. We were inspired by a report of spatial transcriptomic profiling in zebrafish embryo, termed “RNA tomography”,⁶ to construct an epigenomic tomography in mouse brain with spatial resolution. Considering the tools available, we decided on processing mouse brain into 0.5 mm-thick coronal slices using a slicing matrix. We conducted complete profiling of H3K27me3 of neurons and glia for both left and right hemispheres of neocortex in two mice. Results in these biological replicates showed similar patterns in the two hemispheres of neocortex, and that glia remained largely uniform in the spatial context. Knowing this, we further profiled one mouse for gene expression using RNA-seq, concentrating on the right hemisphere and neurons. By comparing the transcriptomic data with the repressive H3K27me3 data, we found a distinct subgroup of genes which promoter H3K27me3 signal and

gene expression signal followed inverse spatial trends. Finally, we examined both H3K27ac and H3K27me3 in a pair of injected mice, one with kainic acid to induce seizure. We found that the epigenetic effect of induced seizure in neocortex to be concentrated in the anterior and posterior parts. The most disrupted genomic regions were related to key functions in neurotransmission, learning and memory. Apart from these biological insights into the fine details of neocortex, we also optimized a clustering pipeline for analyzing such spatial epigenomic datasets and visualizing them in an easy-to-understand way.

Finally, we continued our effort in expanding the selection of epigenomic targets for our low input technology. Long non-coding RNA (lncRNA) is the last main category and research interest in it has been rising in recent years. We collaborated with Dr. Zhen Chen's group, who had been studying lncRNA extensively. We worked from the basis of the chromatin isolation by RNA purification (ChIRP-seq) protocol developed by Howard Chang's group. We decided to adapt it for our MOWChIP device, utilizing the strength of microfluidics to reduce the amount of sample needed. We kept the basic steps of cell fixation and sonication from the original method, and switched the hybridization steps around so the biotinylated probes were adsorbed onto magnetic beads first. The coated beads were then incubated with chromatin sample on the device and hybridized via gentle oscillating pressure pulse to achieve capture, followed by washing and chromatin isolation. We were able to produce datasets with consistent quality and matched the published results using 500K cells per assay. This was a significant improvement over the original method which recommended 20 million cells as input material. Through our trials we also found out that the design of the probe affected the resulting data of the same target lncRNA molecule when all other conditions were the same. This project is ongoing, and we have laid the foundation for a microfluidic, low input profiling technology capable of

examining lncRNA binding in animal and human tissue samples. This will expand the scope of ChIRP-seq to include samples with more biological relevance. Low input ChIRP-seq would help researchers reach beyond cultured cells in their continued effort to understand lncRNA's function. It also completes our lab's arsenal of microfluidic platforms for low input epigenomic mapping, covering all major aspects including DNA methylation, histone modifications, non-histone protein binding and lncRNA.

Chapter 2: Native MOWChIP-seq for profiling RNA polymerase II, transcription factors and histone deacetylases

Introduction

The genetic information of eukaryotes are packed into chromatin, a complex of DNA wrapped around octamers of histone proteins.⁷ Chemical modifications to either DNA or histone proteins affects chromatin structure, regulating gene expression in the process.⁸ As a result, histone modification and DNA methylation have received much research interest as the two main categories of epigenetics, with lncRNA's role only emerging recently. However, it is important to note the roles of non-histone protein targets that also bind to DNA and are heavily involved in regulating transcription and expression of genes.⁹ These proteins include RNA polymerases, transcription factors and histone modifying enzymes.

RNA polymerase II (Pol II) is one of the three RNA polymerases found in eukaryotic nucleus.¹⁰ It transcribes mRNA from protein-coding genes with the help of many co-factors through a number of steps: preinitiation complex formation, promoter-proximal pausing, productive elongation and termination.¹¹ Once bound to DNA, Pol II acquires serine 5-phosphorylation (Ser5P) and enters a paused state, being held in place by two pausing factors. A short (< 60 nt) mRNA has been produced at this time. The paused Pol II can either be released without generating full length mRNA or wait for a cue to complete transcription. This pausing is critical as it controls the rate of expression for 70% of metazoan genes.¹² Pol II pausing also serves key biological functions, such as coordinating tissue morphogenesis, regulating developmental potential and inhibition of new transcription.¹³⁻¹⁵

Transcription factors (TFs) are proteins that control the rate of transcription of genes. They bind to the genome in a sequencing-specific manner and initiate the process of Pol II binding and mRNA production.^{16,17} Through regulating the expression of genes, TFs control important biological processes in cells including differentiation, development patterning and immune response.^{16,18} They are a large group with complex functionalities, with early estimations of about 3,000 DNA-binding TFs and 1,391 manually curated for detailed analysis by Vaquerizas *et al.* in 2009.¹⁹ Interest in TFs had been growing since, along with efforts to establish resources that examine and compile their detailed information.²⁰ Because of their central role in key cellular functions, TF activity are often dysregulated in diseases, and researchers have been investigating their potential as therapeutic targets.^{21,22} Another interesting property of a TF is its ability to regulate the expression of both itself and other TFs. Because of this, the concept of TF network has been used to describe the circuitry of complex TF interactions. These networks function to maintain normal cellular functions and are often disrupted in diseases.²³

Histone modifying enzymes are enzymes that participate in conferring various types of post-translational modifications to histone substrates.^{24,25} Histone modification is one major component of the eukaryotic epigenome, regulating gene expression and key cellular functions. As such, histone modifying enzymes serve as the upstream source of histone modification and play important roles in the overall epigenetic regulation of cell biology. Major types of histone modifying enzymes include histone acetyltransferases (HATs), histone deacetylases (HDACs), histone lysine methyltransferases (HKMTs), histone arginine methyltransferases (HRMTs) and histone demethylases (HDMases).²⁶ They possess modulating functions in development, cell fate decisions and diseases including cancer.²⁷⁻³⁰

Because of their involvement in epigenetic regulation and cellular functions, genome-wide profiling of the above non-histone protein targets is of great value. Like histone modification, non-histone targets are also examined primarily with ChIP-seq methods. While the standard practice of ChIP-seq applies to both histone modification and non-histone targets, some differences exist. The linkage between non-histone proteins and DNA are usually weaker and more transient than that between histone proteins and DNA. This leads to more technical difficulties, more input materials per assay and lower overall data quality.³¹

Chromatin immunoprecipitation coupled with sequencing (ChIP-seq) is the gold standard method to profiling genome-wide state of histone modification and binding of non-histone proteins.³¹ Fixation of protein-DNA binding with chemicals such as formaldehyde had been a standard practice in conventional ChIP-seq. However, the process was cumbersome, involving multiple steps of treatment, incubation and washing.³² Large amounts of starting material was usually required ($\sim 10^7$ cells per assay), making it impractical for use on human and small animal sample which are limited in quantity.³¹ In recent years, significant efforts have been made in developing low-input ChIP-seq methods.^{3,33-41} However, most of these methods have only been demonstrated on examining histone modifications.

The necessity of fixation was in part due to the harsh conditions of DNA fragmentation by sonication, which can easily damage the transient binding between proteins and DNA in their native state and render ChIP-seq impossible. Sonication also required extensive optimization as difference in cell lines, tissues, fixation processes, model of sonicators and their programs can all change the results significantly.⁴² Our lab had years of experience on fixation and sonication. The optimization process was long and costly for each project as circumstances varied, and the results were rarely consistent. Because of these shortcomings, an alternative

DNA fragmentation strategy with gentle conditions became the next step towards better ChIP-seq methods.

Micrococcal nuclease (MNase) is isolated from *Staphylococcus aureus* and able to digest DNA. Under the correct conditions, MNase will cleave chromatin into fragments at the spaces between nucleosomes, leaving DNA fragments sized between 180bp and 550bp (1 to 3 nucleosomes). Histone proteins are left intact and bound to DNA. In ChIP-seq, MNase has been used as the replacement of sonication to fragment DNA. Since it is a much gentler treatment process compared to sonication, chromatin does not need chemical fixation.^{43,44} Native ChIP-seq has become the preferred method for profiling histone modification due to its higher resolution and lack of artifacts.⁴⁵ In terms of practicality, MNase digestion shortens the process by up to several working days as the complex and time-consuming steps of chemical crosslinking, sonication and reverse-crosslinking are replaced with a 30-minute process. The need for a specialized and expensive sonicator is also eliminated.

MNase digestion has also seen increasing use in the development of low-input ChIP-seq assays. While early attempt such as nano-ChIP-seq⁴⁶ and LinDA-ChIP-seq⁴⁷ still used chemical fixation, recent developments in the field have moved on towards native ChIP-seq and MNase digestion.⁴⁸ In particular, CUT&RUN, a low-input ChIP-seq method now in early commercialization, is based solely on MNase. Their work went back to using MNase as simply a fragmentation tool on fixed chromatin to improve ChIP-seq resolution.⁴⁹ Later, they combined the fragmentation of MNase with the specificity of immunoprecipitation by developing custom MNase with protein A anchor point for antibody binding.⁵⁰ With the ChIP antibody in tow, the MNase selectively cleave only the target sections of the genome and release them from nuclei for recovery. This greatly increased the efficiency of ChIP-seq and enabled CUT&RUN to

profile 100 cells for H3K27me3 and 1,000 cells for CTCF.^{51,52} The newer CUT&Tag method further incorporates the function of tagmentation by Tn5 to improve sensitivity and reduce sequencing cost.⁵³ The latest version of CUT&Tag⁵⁴ 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^{55,56}. 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).⁵⁷

Aside from the advantages MNase offers in the pure molecular biology sense, its small footprint, lower cost and ease of use contributed significantly to this popularity. From our lab's experience, MNase digestion produces high quality results consistently and we have found little to no need for optimization when switching to a different cell line or tissue sample. This in turn saves valuable time and cost and is one of reasons we were able to conduct several large-scale projects involving hundreds of ChIP-seq assays each. While our low-input MOWChIP-seq was first developed with fixation and sonication steps,¹ later developments have made use of MNase exclusively, including Surface-ChIP-seq³ and parallel native MOWChIP-seq on both histone modification and non-histone targets.^{2,58}

In this work, we demonstrated the use of native MOWChIP-seq based on MNase digestion for low-input profiling of Pol II, TFs and HDAC2 binding in various samples. It was believed that these proteins needed to be chemically fixed to DNA prior to ChIP assays. However, our trials revealed that crosslinking drastically reduced the performance of MOWChIP-seq on these targets. On the other hand, MNase digestion was shown to preserve the binding between proteins and DNA effectively, enabling low-input profiling. We tested the system with GM12878

cells and generated high quality ChIP-seq data using as few as 1,00 cells for Pol II, 5,000 cells for TF EGR1, and 50,000 cells for HDAC2. The method was then applied to map genome-wide binding of Pol II, EGR1 and MEF2C in mouse prefrontal cortex (PFC) and cerebellum.

Significant variations in Pol II and TF binding were identified between these two functional regions of the mouse brain. Further analysis of the differential regions and their associated genes yielded biological functions and pathways that agree with the distinct functions of PFC and cerebellum.

Author contribution

Chang Lu designed and supervised the study. Zhengzhi Liu conducted native MOWChIP-seq experiments and data analysis. Lynette Naler, Yan Zhu, Chengyu Deng, Qiang Zhang, Bohan Zhu, Zirui Zhou and Mimosa Sarma helped with experiments and data analysis. Alexander Murray and Hehuang Xie helped with experiments on TFs. Zhengzhi Liu, Lynette Naler and Chang Lu wrote the manuscript.

Results and discussion

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

RNA Pol II, TF and enzyme binding has been conventionally studied after crosslinking using reagent such as formaldehyde that firmly immobilizes the protein to the genomic DNA.^{59,60} Our low-input MOWChIP-seq technology which has proven high efficiency for collecting DNA-protein complexes.^{39,40} However, we generated low-quality results when 100,000 crosslinked cells were used for MOWChIP-seq. 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 limit ChIP efficiency.

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.^{3,55,61} 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.^{39,40} 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 chromatin were removed from the chamber for DNA elution, library preparation and sequencing.

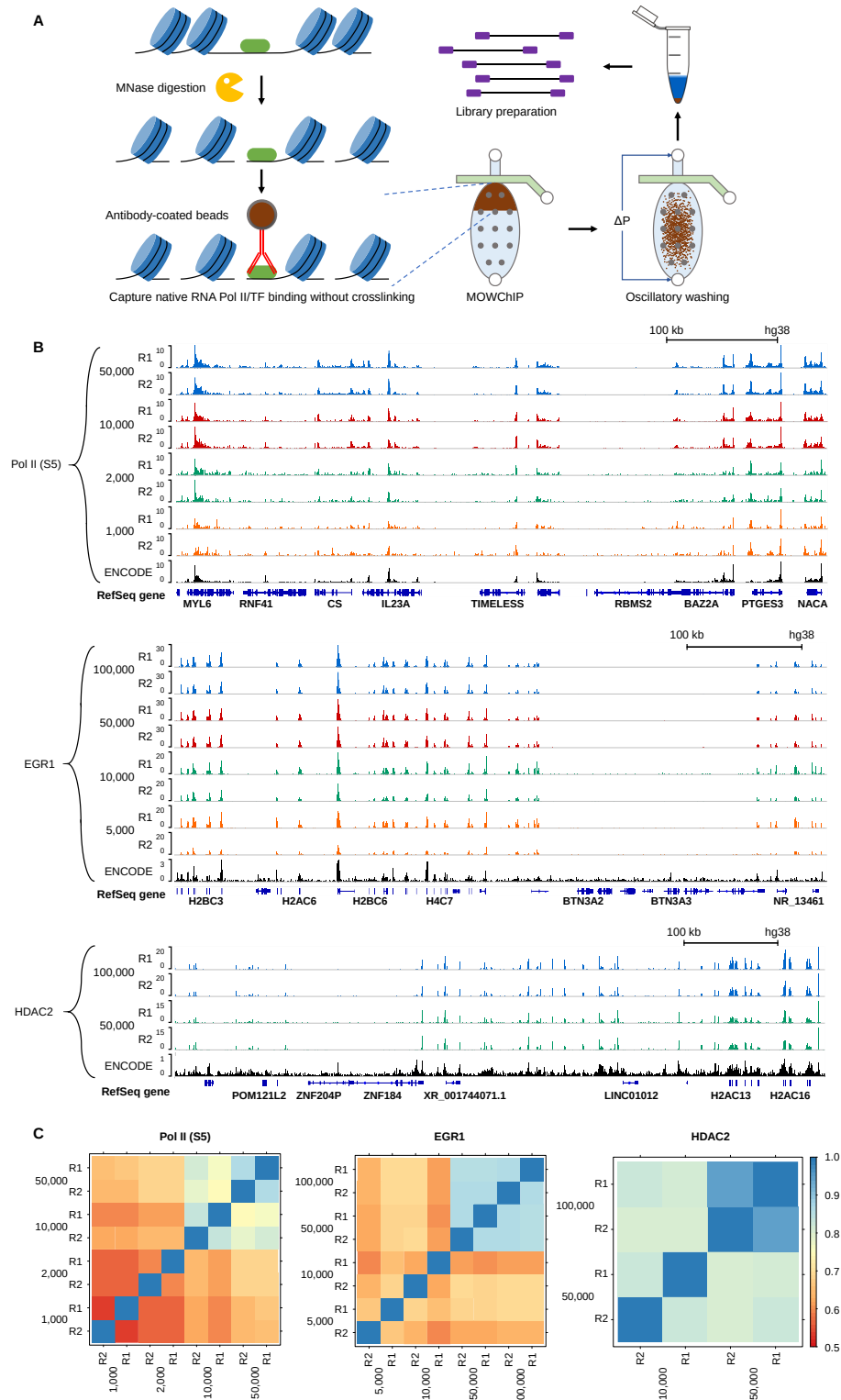


Figure 1: Overview of the low input nMOWChIP process and data on RNA polymerase II, EGR1 and HDAC2 binding in GM12878 cells. (A) Steps to generate nMOWChIP-seq libraries. (B) Normalized RNA Pol II (S5) signals generated using 1,000 to 50,000 cells, EGR1 signals generated using 5,000 to 100,000 cells, 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.

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 (phosphor S5) (Figure 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 and 20 million cells per assay. We examined the number of peaks called and the fraction of reads in peaks (FRiP) for data quality (Table 1). The nMOWChIP-seq Pol II data produced an average of 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 ENCODE.⁶²

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 (Figure 2). 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 (Figure 1). 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 (Table

1). A large percentage of the EGR1 (85%) and MEF2C (75%) peaks overlap with Pol II peaks.

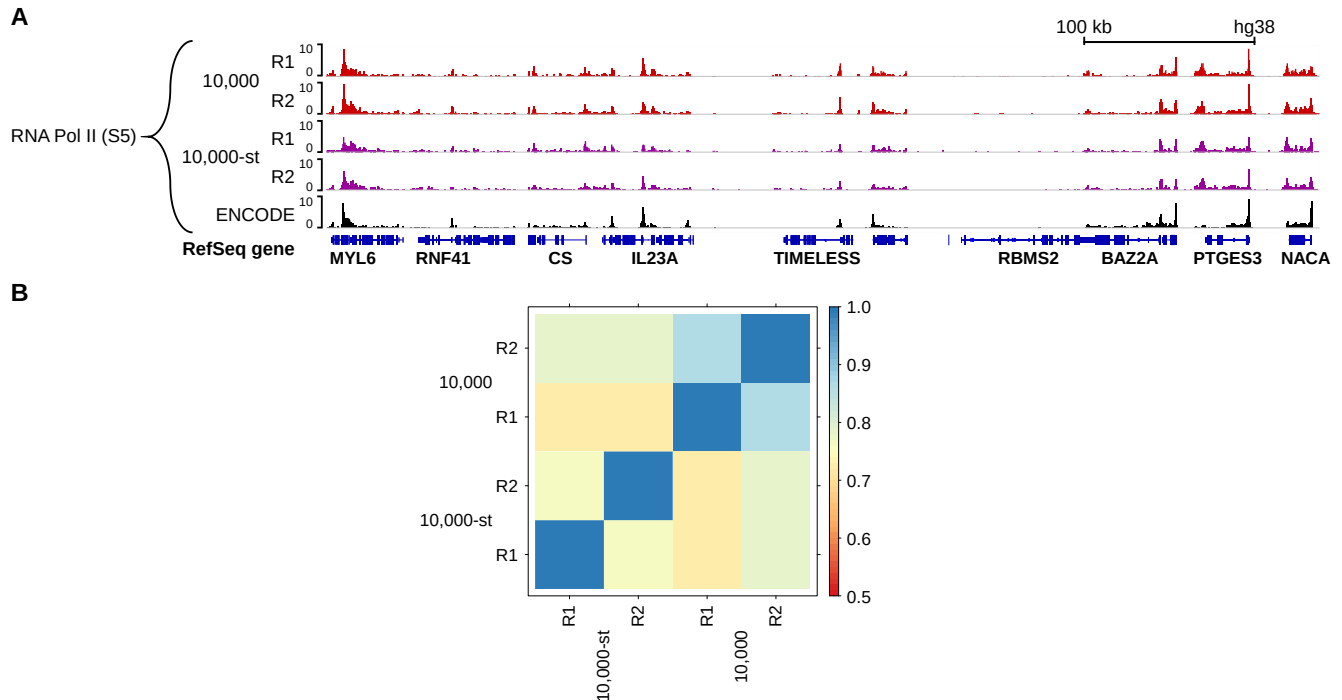


Figure 2: Storage at -80 °C does not change RNA pol II binding profile. (A) Normalized Pol II-S5 signal of GM12878 cells. “10,000-st” denotes chromatin samples (each containing chromatin fragments from 10,000 cells) that were stored at -80 °C for 2 d and then used for nMOWChIP-seq. “10,000” refers to samples processed without such storage. (B) Pearson’s correlation matrix of Pol II-S5 nMOWChIP-seq data of fresh and -80 °C stored GM12878 chromatin, computed using DiffBind affinity score method.

Finally, we applied nMOWChIP-seq to examine the binding of histone deacetylase HDAC2 in GM12878 cells (Figure 1B and 1C, Table 1). 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. 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 (Table 1).

Table 1: Summary of nMOWChIP-seq data

File	Mapped reads	Mapping Rate	Unique reads	Unique rate	MACS2	FRIP
GM12878_Pol2-total_50K_1	16432424	98%	16204517	99%	133015	57.1%
GM12878_Pol2-total_50K_2	20219467	98%	19820603	98%	129219	55.5%
GM12878_Pol2-S5_50K_1	16006273	98%	15570715	97%	142339	52.4%
GM12878_Pol2-S5_50K_2	15203135	98%	14377380	95%	130669	47.3%
GM12878_Pol2-S5_10K_1	12271356	96%	12061753	98%	114811	41.3%
GM12878_Pol2-S5_10K_2	21780356	96%	21550914	99%	140695	48.0%
GM12878_Pol2-S5_10K_st_1	9646207	95%	9354458	97%	82320	27.8%
GM12878_Pol2-S5_10K_st_2	13775137	96%	13338787	97%	79800	24.9%
GM12878_Pol2-S5_2K_1	14243331	90%	13432201	94%	59713	17.7%
GM12878_Pol2-S5_2K_2	15242929	87%	14400694	94%	70439	22.0%
GM12878_Pol2-S5_1K_1	15572762	87%	14923018	96%	33627	10.7%
GM12878_Pol2-S5_1K_2	9194036	86%	8777057	95%	19820	8.1%
GM12878_Pol2-S5_100K_cl_1	31745569	98%	31406005	99%	53817	8.8%
GM12878_Pol2-S5_100K_cl_2	31111394	98%	30545134	98%	17561	2.7%
GM12878_EGR1_100K_1	14323523	96%	12199195	85%	16614	8.3%
GM12878_EGR1_100K_2	17458334	96%	14961426	86%	15584	7.9%
GM12878_EGR1_50K_1	14910367	95%	12800282	86%	16465	8.6%
GM12878_EGR1_50K_2	16669698	95%	14480504	87%	15159	8.2%
GM12878_EGR1_10K_1	9072849	84%	7725182	85%	2626	3.8%
GM12878_EGR1_10K_2	12461029	94%	10785293	87%	2781	2.6%
GM12878_EGR1_5K_1	10330104	92%	8383719	81%	8245	5.0%
GM12878_EGR1_5K_2	12660454	93%	11175451	88%	1423	1.9%
GM12878_MEF2C_100K_1	24369737	76%	21133894	87%	14716	4.5%
GM12878_MEF2C_100K_2	24642860	85%	20705206	84%	19884	6.3%
GM12878_HDAC2_100K_1	21068312	84%	18801502	89%	11221	3.8%
GM12878_HDAC2_100K_2	19862691	85%	18127467	91%	5832	2.3%
GM12878_HDAC2_50K_1	8767036	93%	8093468	92%	1204	1.3%
GM12878_HDAC2_50K_2	12215492	93%	11358639	93%	1584	1.1%
PFC_Pol2-total_50K_1	13064208	97%	12803650	98%	66123	48.6%
PFC_Pol2-total_50K_2	16395371	99%	16251965	99%	80138	55.4%
PFC_Pol2-total_10K_1	8574476	95%	7748982	90%	24423	26.5%
PFC_Pol2-total_10K_2	13007858	93%	12338679	95%	53291	46.4%
Cerebellum_Pol2-total_50K_1	10617272	99%	10444570	98%	43387	52.0%
Cerebellum_Pol2-total_50K_2	9805738	98%	9553592	97%	36762	48.5%
Cerebellum_Pol2-total_10K_1	15483210	97%	15200342	98%	54534	67.1%
Cerebellum_Pol2-total_10K_2	14308273	98%	14139656	99%	54278	69.5%
PFC_EGR1_100K_1	9285737	91%	8104052	87%	6212	10.7%
PFC_EGR1_100K_2	8557087	51%	7510446	88%	6921	11.0%
Cerebellum_EGR1_100K_1	9882524	95%	8739064	88%	17795	16.2%

Cerebellum_EGR1_100K_2	13141514	95%	11799729	90%	22866	15.4%
PFC_MEF2C_100K_1	8478934	82%	7497951	88%	1354	7.5%
PFC_MEF2C_100K_2	11407867	91%	10012928	88%	5362	9.1%
Cerebellum_MEF2C_100K_1	11268711	95%	10016208	89%	24063	19.0%
Cerebellum_MEF2C_100K_2	11481247	95%	10025093	87%	17393	15.6%
PFC_HDAC2_100K_1	14011639	85%	11164030	80%	2929	13.6%
PFC_HDAC2_100K_2	15354891	85%	12429080	81%	3465	12.6%
PFC_NeuN+_HDAC2_80K_1	12258720	80%	10188120	83%	1382	11.2%
PFC_NeuN+_HDAC2_80K_2	12448796	80%	10639071	85%	2842	9.6%

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 showed the distinction. The two profiles on GM12878 cells were similar in a large fraction of the genome (Figure 3A). 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 3B). 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.¹¹ 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 3C).

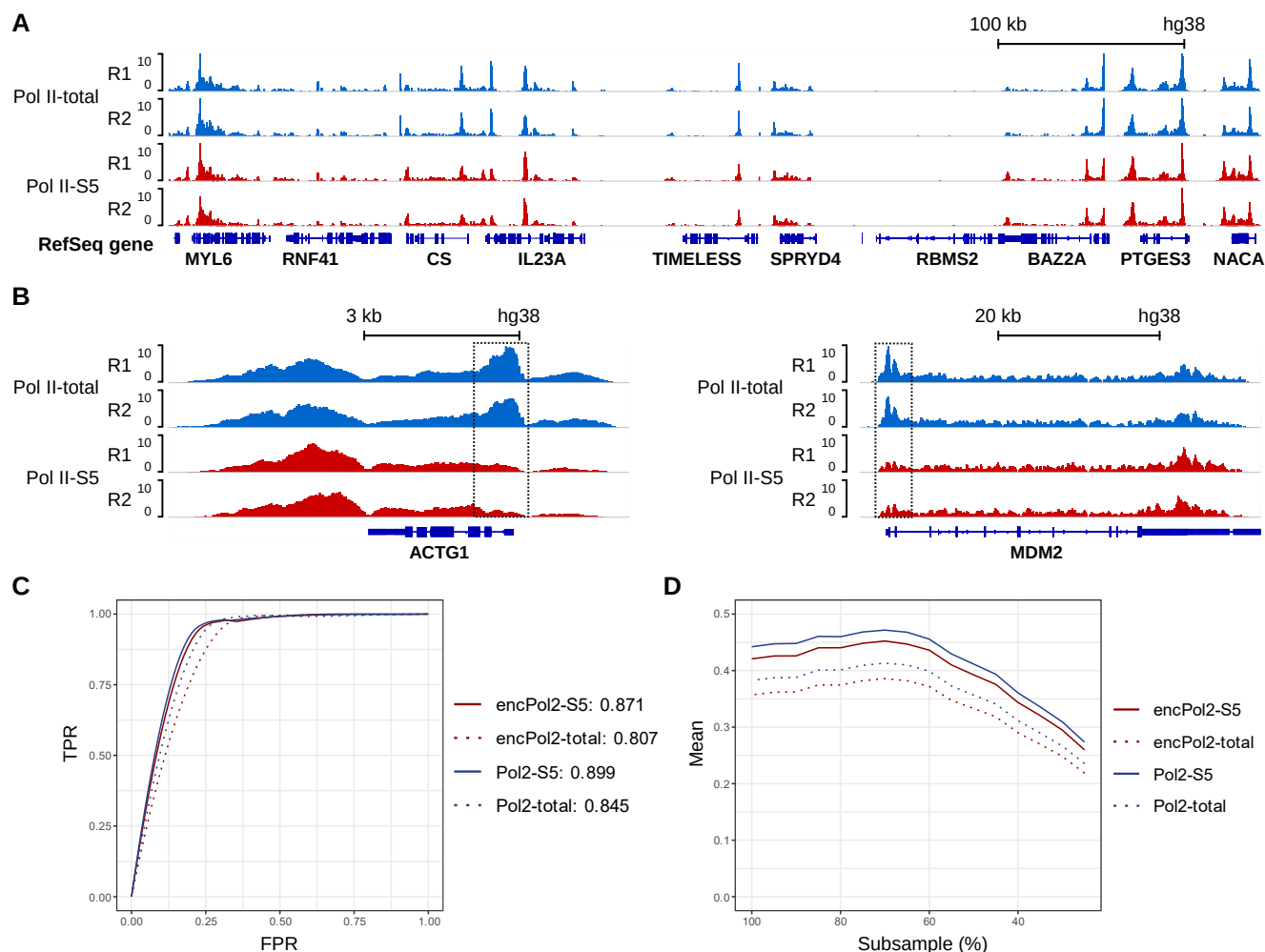


Figure 3: Comparison of Pol II-total and Pol II-S5 bindings in GM12878 cells. (A) Normalized Pol II-S5 and Pol II-total signal in GM12878 cells (50,000 cells per assay). (B) Normalized Pol II signal over genes ACTG1 and MDM2, showing the difference in Pol II-total and Pol II-S5 binding. Pol II-total 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.

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 3D). 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,^{63,64} while cerebellum controls motor functions and coordination.⁶⁵ 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 has 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 (Figure 4). The genome-wide Pol II profiles were substantially different between cerebellum and PFC with Pearson's correlation at 0.67, compared to an average of 0.96 between technical replicates (Figure 4A-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, $p < 10^{-5}$). Gene ontology (GO) analysis showed regions with high Pol II binding intensity in PFC were enriched in memory, learning and anxiety-related response. Synaptic plasticity, one of the fundamentals of learning and memory, was also enriched along with its key components: long term potentiation and depression (Figure 4C).⁶⁶ In contrast, the genomic regions with higher Pol II binding in cerebellum were enriched in cerebellum specific processes (e.g. cerebellum morphology and development) (Figure 4C).

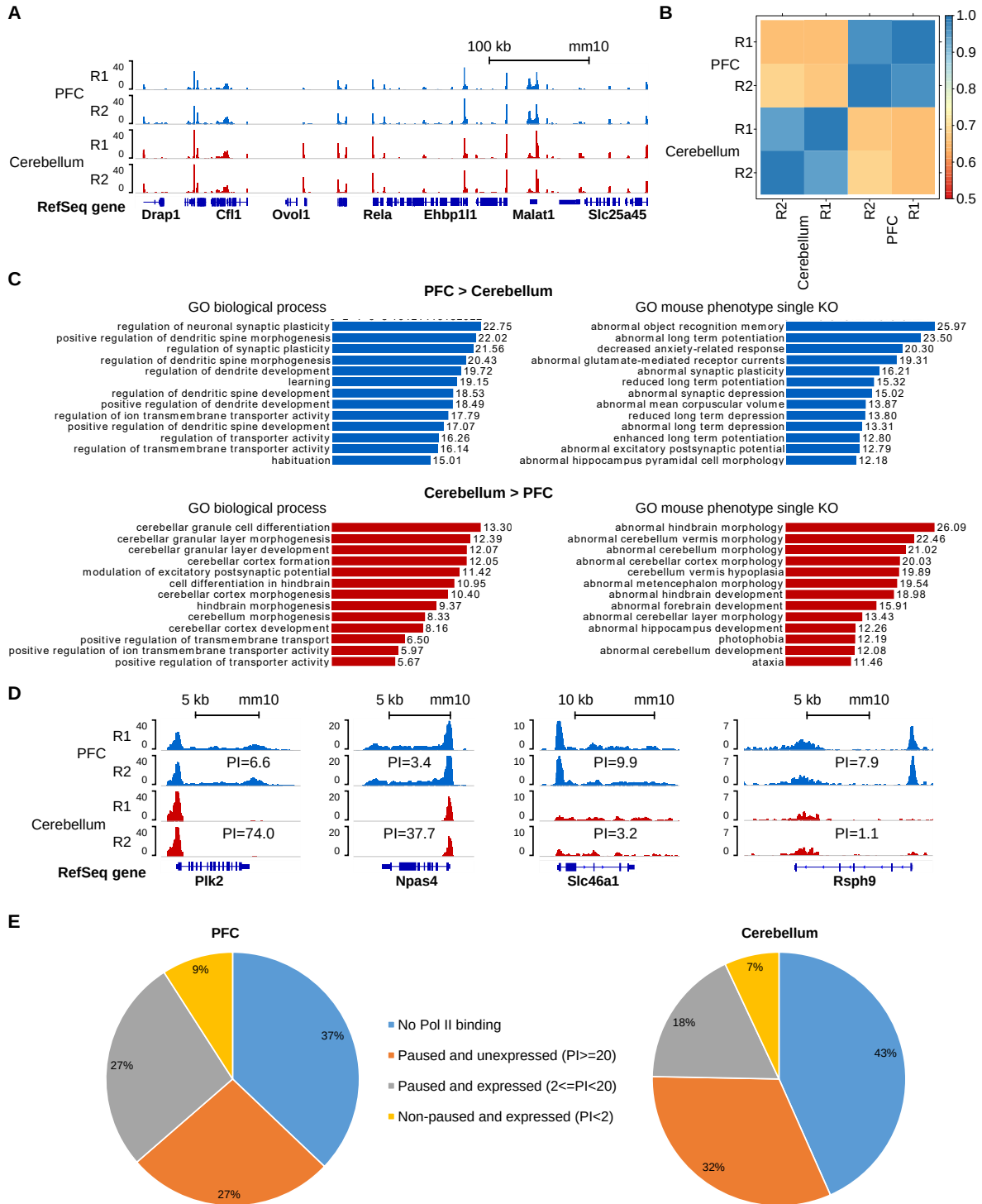


Figure 4: Differential RNA polymerase II binding in prefrontal cortex and cerebellum of mouse brain. (A) Normalized Pol II-total signals generated using nuclei isolated from mouse prefrontal cortex (PFC) and cerebellum (50,000 nuclei per assay). (B) Pearson's correlations of affinity score calculated 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.

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.^{11,67} 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 ($PI > 20$).¹¹ For example, *Plk2*, which is a gene known to participate in rodent brain development and cell proliferation,⁶⁸ and *Npas4*, which is involved in regulating reward-related learning and memory,⁶⁹ both showed dramatically higher PI values in cerebellum than in PFC (Figure 4D). 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.⁷⁰ Furthermore, we analyzed the distribution of the PI in PFC and cerebellum (Figure 4E). 36% of genes were being actively expressed in PFC ($PI < 20$), compared to only 25% in cerebellum. Mouse brain displays more pause and less active transcription compared to GM12878 (52% actively expressed genes). This was also within our expectation because GM12878 cells were 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). Among these, *Igfbp6* is involved in myelin formation during central nervous system (CNS) development,⁷¹ and *Slc1a2* encodes excitatory amino acid transporter 2, responsible for reuptake of 90% glutamate in CNS.⁷² *Shank3* belongs to the Shank gene family that plays a

role in synapse formation,⁷³ and *Flrt2* encodes a member of the FLRT protein family that is shown to regulate signaling during mouse development.⁷⁴

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 5A). We picked several genes as examples. *Kalrn* (Kalirin) plays important roles in nerve growth⁷⁵. Higher EGR1/MEF2C activity in PFC was observed on *Gria1* (Glutamate receptor 1) which is involved in synaptic transmission.⁷⁶ *Cacna1a* is involved in movement disorder⁷⁷ and expression of *Nfix* can influence neural stem cell differentiation.⁷⁸ *Zic1* and *Zic4* belong to the family of Zinc finger of the cerebellum (ZIC) protein family,⁷⁹ whose loss of function can lead to Dandy-Walker malformation and incomplete cerebellar vermis.⁸⁰ 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. We observed correlated EGR1 and MEF2C profiles, with an average of Pearson's correlation of 0.74 between the two TFs in PFC and 0.835 in cerebellum (Figure 5B). 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 0.513 and 0.508 between the two brain regions, respectively. 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%) (Figure 5C). We further analyzed the data using DiffBind to identify regions with significantly different level of EGR1/MEF2C binding between PFC and cerebellum.⁸¹

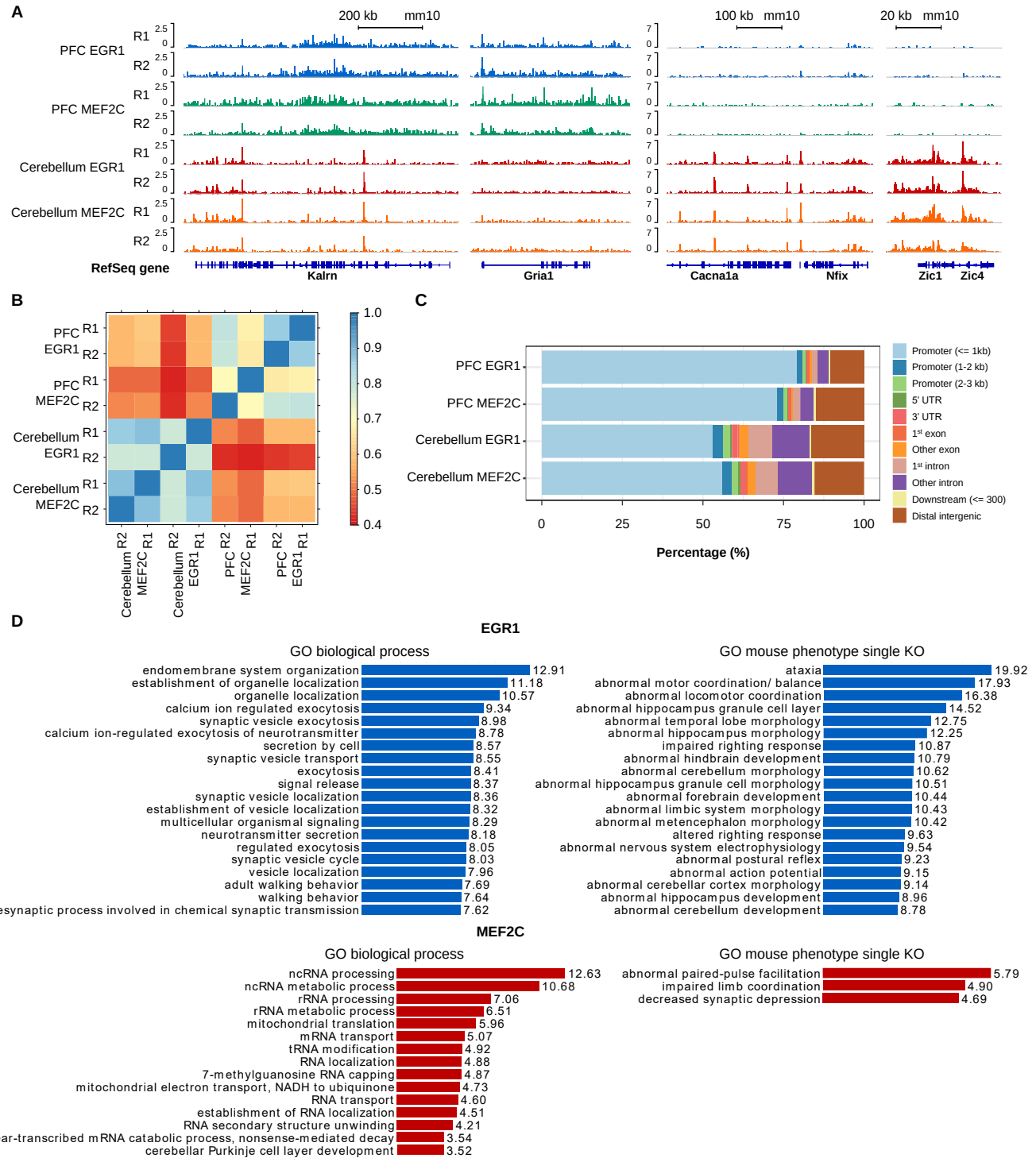


Figure 5: Differential transcription factor binding in prefrontal cortex and cerebellum of mouse brain. (A) Normalized EGR1 and MEF2C signals at genes identified by DiffBind to have significantly different binding between PFC and cerebellum (fold change > 2, $p < 10^{-5}$). (B) Pearson's correlations 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, $p < 10^{-5}$).

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 5D). 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 Pol II, TF 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 masking⁸² 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.⁸³

In this work, we conducted nMOWChIP-seq without crosslinking and sonication on non-histone targets. We demonstrated that our low-input technology could work with as few as 1000-

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 a very high signal-to-noise ratio and low background, characterized by high FRiP values.

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 were highly characteristic of the brain regions. The Pol II binding profiles had 4,200 differential peaks between cerebellum and PFC, while EGR1/MEF2C profiles had 2,100 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.

Materials and methods

Fabrication of microfluidic device

Microfluidic ChIP device was fabricated using polydimethylsiloxane (PDMS) via soft lithography. Photomasks of fluidic layer and control layer with microscale patterns were designed with FreeHand MX (Macromedia) and printed on high-resolution (5,080 dpi) transparencies. For fluidic layer master, photoresist (SU-8 2025, Microchem) was spun on a 3-inch silicon wafer (978, University Wafer) at 500 rpm for 10s and 2500 rpm for 30s and baked at 95°C for 8 min. For control layer master, SU-8 was spun at 500 rpm for 10s and 1500 rpm for 30s and baked at 95°C for 8 min. Each master was covered with its photomask and exposed to 580mW UV for 17s, and followed by baking at 95°C for 8 min. They were then developed in SU-8 developer for 2 min, rinsed with IPA and D.I. water and air blown to dry. The masters were then baked at 140°C to strengthen the bond and prevent features from detaching. To make fluidic layer, PDMS (General Electric silicone RTV 615, MG chemicals) with a mass ratio of A:B = 5:1 was thoroughly mixed and vacuumed for 1 hr. It was then poured onto the fluidic layer master in a Petri dish to a thickness of ~5 mm. To make control layer stamp, PDMS with a mass ratio of A:B = 20:1 was mixed, vacuumed for 1 hr, and spun onto the control layer master at 1100 rpm for 35s. Both stamps were partially cured at 80 °C for 13 min. The fluidic layer was then peeled off from the fluidic layer master and aligned to the control layer. Two-layer PDMS were thermally bonded at 80 °C for 90 min and peeled off from control layer master. The inlets and outlets of the device were punched by a 2 mm hole puncher. Glass slides washed thoroughly, and air blown to dry. The two-layer PDMS and a pre-cleaned glass slide were treated with oxygen plasma cleaner (PDC-32G, Harrick Plasma) and brought together to form closed channels and chamber. The device was then baked at 80 °C for 4 hours to strengthen the bonding between PDMS and glass.

Cell culture

GM12878 cells were obtained from Coriell Institute for Medical Research. Cells were cultured in RPMI-1640 medium (30-2001, ATCC) supplemented with 15% fetal bovine serum (1600-044, Gibco) and 1% pen-strep (Invitrogen) at 37 °C, 5% CO₂. Cells were subcultured every 3 days to maintain exponential growth.

Mouse strain and brain dissection

Male C57BL/6J mice were obtained from Jackson Laboratory. They were kept in the animal facility with 12-h light/12-h dark cycles and food and water ad libitum. 8-week-old male mice were euthanized with compressed CO₂ followed by cervical dislocation. Brains were rapidly dissected, flash frozen on dry ice and stored at -80 °C. The study was approved by the Institutional Animal Care and Use Committee (IACUC) at Virginia Tech.

Nuclei isolation from mouse brain tissue

A mouse brain was put on dry 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, 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 μ l of Dulbecco's phosphate-buffered saline (DPBS). If nuclei labeling and FACS sorting were conducted, 500 μ l 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/ μ l. 8 μ l of anti-NeuN was added to each 500 μ l of nuclei suspension and incubated at 4°C for 1 h on a rotator. The labeled nuclei were then sorted using FACS (BD FACS Aria, BD Biosciences). 8 μ l 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 μ l of 1.8 M sucrose, 5 μ l of 1M CaCl_2 and 3 μ l of 1M $\text{Mg}(\text{Ac})_2$ 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.

Chromatin preparation by MNase digestion

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 MgCl_2] were added, mixed by vortexing and incubated at room temperature for 10 min. 10 μ l of 100 mM CaCl_2 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.

Immunoprecipitation beads preparation

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, Abcam). The suspension was incubated on a rotator at 4°C overnight. The beads were then washed with IP buffer twice and resuspended in 5 μ l of IP buffer for loading into the device chamber.

Operation of MOWChIP device

The device was controlled by a customized LabVIEW program (National Instruments) managing solenoid valves (18801003-12V, ASCO Scientific) via a data acquisition card (NI SCB-68, National Instruments). The valves are connected to compressed air for closing the sieve valve and conducting oscillatory washing. The control channel was filled with water first and connected to a valve with 30 psi of pressure. When the valve was opened, the pressure deformed the control layer membrane and partially closed the sieve valve in the device. This prevented the beads from leaking but allowed the solution to pass. The device was then rinsed with IP buffer at 50 μ l/min using a syringe pump (Fusion 400, Chemyx) to remove air from the chamber. IP beads were loaded into the inlet of the chamber guided by magnet. 60 μ l of chromatin solution from the appropriate number of cells (diluted from MNase-digested

chromatin samples with IP buffer where necessary) was loaded into the chamber at 1.5 μ l/min, and beads were packed against the partially closed sieve valve after 2 min of sample loading. Sample loading was completed after 60 min. 2 washing tubes filled with low salt buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% (v/v) Triton X-100) were then connected to both inlet and outlet of the device, and oscillatory washing was conducted for 5 min at ~1 psi with alternating pulses of 0.5 s with the sieving valve open. The same washing was repeated with high salt buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% (v/v) Triton X-100). Beads were then flushed out of the chamber using IP buffer at 200 μ l/min and collected in a 1.5 ml tube.

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 2S Plus DNA Library kit (IDT) 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 Appendix. 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.⁸⁴ Uniquely mapped reads were filtered for known blacklisted genome regions using Samtools⁸⁵ and bedtools⁸⁶ 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.⁸⁷ Peak calling was conducted using MACS2 (default settings). Differential binding analysis was carried out using DiffBind⁸¹ with default settings. Selected genomic regions of interest were further analyzed for gene ontology terms using GREAT.⁸⁸ 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 9.9×10^{-1} 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.

Chapter 3: Comprehensive histone modification and gene expression profiling in human lung tumor samples

Introduction

The importance of epigenetic changes in cancer biology has been well-established. Significant efforts have been made investigating the cancer epigenome and decoding it for insight into possible intervention. Abnormalities and misregulation of epigenetic pathways are common in cancer. A large body of evidence suggest these changes play key roles in oncogenesis and progression and hold potential for drug development.⁸⁹⁻⁹⁴ Studies of primary tumor samples have revealed valuable information on risk loci and regulatory element reprogramming in cancer. The Cancer Genome Atlas has conducted several large scale profiling of primary cancer samples, including whole-genome bisulfite sequencing of 39 tumors which identified previously unknown hypomethylation in partially methylated domains.⁹⁵ Their extensive ATAC-seq profiling of 410 tumor samples across 23 cancer types discovered numerous risk loci as DNA regulatory elements in cancer.⁹⁶ Other works on epigenetic profiling of tumors have revealed important aspects of the disease, such as the reactivation of epigenomic programs for tissue development during prostate cancer metastasis,⁹⁷ and abnormal activity in and regulators of colorectal cancer gained enhancers.⁹⁸⁻¹⁰⁰

Profiling epigenetic landscape, often combined with transcriptomic assays in primary, patient-derived samples have proven valuable. However, the scope of such studies is often limited by the scarcity of primary tissue from patients, available from a handful of tissue biobanks typically in the milligram range. This reality contrasts sharply with the starting material requirement for epigenetic and transcriptomic assays, especially when the desire of examining

multiple markers and the need of replicates. Furthermore, due to the complex and highly variable nature of patient samples, a relatively large number of samples is often necessary to draw reliable conclusions. This compounds the discrepancy between sample availability and assay requirement. Usually, no more than three distinct marks are profiled directly in patient samples.¹⁰⁰⁻¹⁰² Xenograft on lab animals and *in vitro* culture are commonly used to increase sample amount,^{97-99,103,104} but these procedures have more stringent requirements for samples (e.g. being freshly extracted and not frozen) and will introduce non-native interferences.

Lung cancer is the most common cause among cancer death as of 2020. Both cases and deaths are on the rise, with 2.09 million new cases and 1.76 million deaths estimated for 2018 alone.¹⁰⁵ Its high mortality rate is attributed to patients often being diagnosed at advanced stages and poor five-year survival rate of only 15%.¹⁰⁶ Epigenetic changes such as DNA methylation, histone modification and non-coding RNA expression have all been shown to play important roles in the initiation and progression of lung cancer, and understanding the underlying mechanisms will contribute to better early detection, diagnosis and therapeutics.¹⁰⁶⁻¹⁰⁹ For instance, it was demonstrated that histone modification levels can predict survival probability in lung cancer patients.¹¹⁰

The maturation of next-generation sequencing technology, in particular its accessibility and reduced cost, has renewed interest in large scale epigenetic profiling of patient samples in cancer. The Cancer Genome Atlas (TCGA) led such an effort in 2018 in which they conducted ATAC-seq on 410 patient tumor samples from 23 types of cancer to map their chromatin accessibility, a key marker of active DNA regulatory elements.⁹⁶ Such large scale profiling of statistically significant number of samples revealed abundant information useful for various

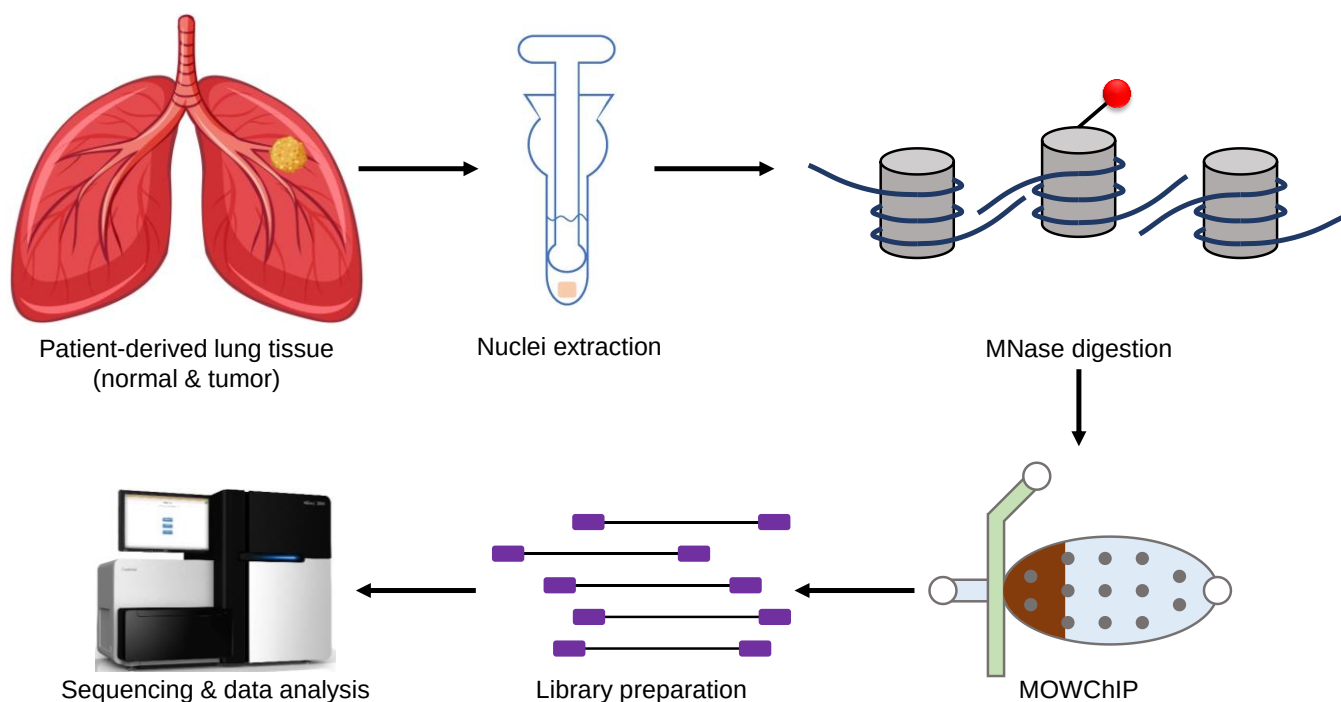


Figure 6: Process of profiling multiple histone modifications in lung tissue samples.

research into cancer. Profiling multiple histone modifications across large number of tumor samples would provide even more new knowledge since different histone modifications have their unique effect on transcription, while ATAC-seq only map the openness of chromatin structure. However, tumor samples from patients are usually available in no more than 50 mg to individual research labs. Work by TCGA was done on 20 mg per sample, ideally yielding more than 100,000 nuclei. This was enough for ATAC-seq assays, but would be far from the required number for typical ChIP-seq assay once multiple histone marks and technical replicates are factored in. Our MOWChIP-seq method is capable of consistently generating high quality ChIP-seq data using 10,000 or less nuclei per assay input material and is suitable for conducting a large-scale profiling of multiple histone modifications in patient samples limited in available weight.

In this work, we used native MOWChIP-seq system to map five key histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K27me3 and H3K9me3) in patient derived lung tissues. A total of 38 lung tissue samples were obtained by Dr. Wei Wang at UCSD from tissue bank. 18 were tumor samples and 20 were non-neoplastic control samples. We generated high quality ChIP-seq data of the five histone marks for 38 tissues, using about 20 mg of sample on average. 8 pairs of samples were 'matched', meaning that the tumor and control tissue come from the same patient. RNA-seq was conducted by extracting total RNA from tissue slices incubated with RNAlater-ICE reagent, followed by a modified Smart-seq2 workflow. Data analysis showed that the libraries produced were of high quality and reproducibility. Numerous differentially modified epigenomic regions (DMER) and differentially expressed genes (DEG) were identified between tumor and non-neoplastic samples. Downstream analyses revealed differential TFs and consensus regulation associated modules (cRAM) specific to tissue conditions. Genome-wide clustering found 6 distinct sections with specific co-modification patterns. Integrated analysis showed tumor and non-neoplastic specific cRAMs to be key regions related to chromatin reorganization with functional implication.

Author contribution

This work was a collaboration with Wei Wang's group at University of California San Diego. Dr. Wang's group acquired the samples, supplied major reagents and funded all sequencing. Lina Zheng coordinated the acquisition and shipping of samples and conducted all bioinformatic analyses of the sequencing data. Zhengzhi Liu carried out histone modification MOWChIP-seq and library preparation. Zirui Zhou conducted RNA extraction and library preparation. All

authors wrote the manuscript.

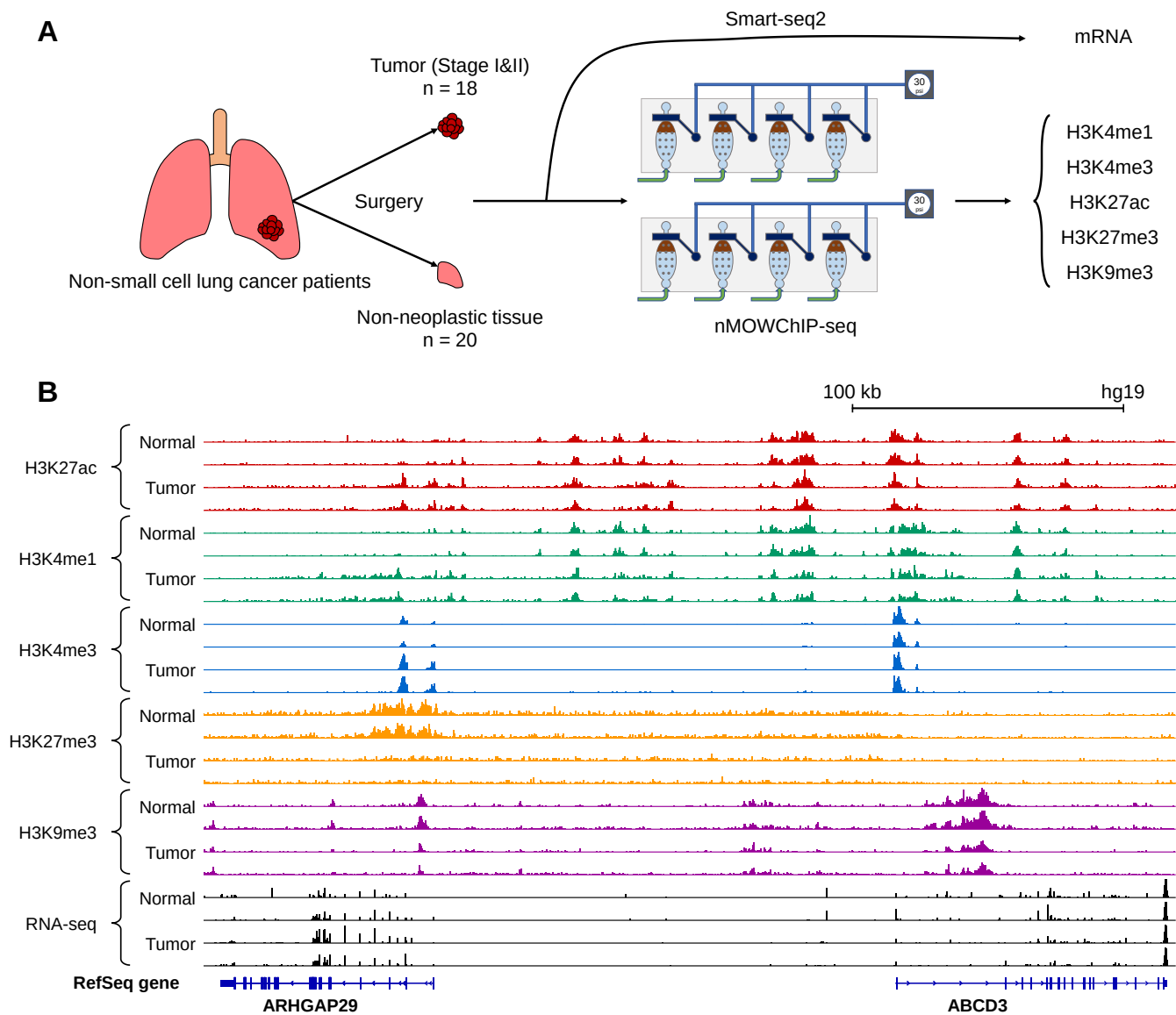


Figure 7: (A) Illustration of the lung ChIP-seq and RNA-seq workflow. (B) Normalized ChIP-seq and RNA-seq signal tracks from normal and tumor samples.

Results and discussion

Tissue processing optimization

We adapted our standard nuclei extraction protocol developed for brain tissues to extract nuclei from lung tissue samples from patients. Tissues were stored in liquid nitrogen vapor

phase. All transportation and handling were done on dry ice to prevent possible sample degradation. Due to the lung tissues being denser and containing more connective tissue compared to brain, we cut thin slices off the tissue piece and diced them into smaller pieces before loading into the tissue grinder. Extra douncing was done to the lung tissue pieces to ensure complete release of nuclei into the extraction solution (20 times with pestle A and 30 times with pestle B, compared to 15 times with pestle A and 25 times with pestle B for brain samples). Nuclei yield per unit of weight also varied across a wide range though we kept the amount of sample used in each nuclei extraction consistent.

Parallel MOWChIP assays and chromatin storage

We processed one sample per day and ran all 10 MOWChIP-seq runs in parallel (5 histone modifications, each having 2 technical replicates). Typically, chromatin from 30,000 nuclei were used for each MOWChIP assay. Extra chromatin was stored at 4 °C as backup in case any library constructed from the ChIP DNA didn't pass qPCR quality check. During the project, we used MNase-digested chromatin stored at 4 °C for 1 month for re-runs and produced high quality data.

RNA extraction in lung tissue sample

Our lab has previously developed an RNA extraction workflow that ran in parallel with the nuclei extraction process from brain tissue. A fraction of the nuclei extracted and purified from sample was set aside and used for RNA extraction by RNeasy Mini kit. The same protocol was initially employed for the lung tissue samples. However, the cDNA we produced was too small for RNA-seq library preparation. The small size indicated severe RNA degradation during the extraction process. After side-by-side comparison with brain tissue samples, we concluded

ChIP-seq

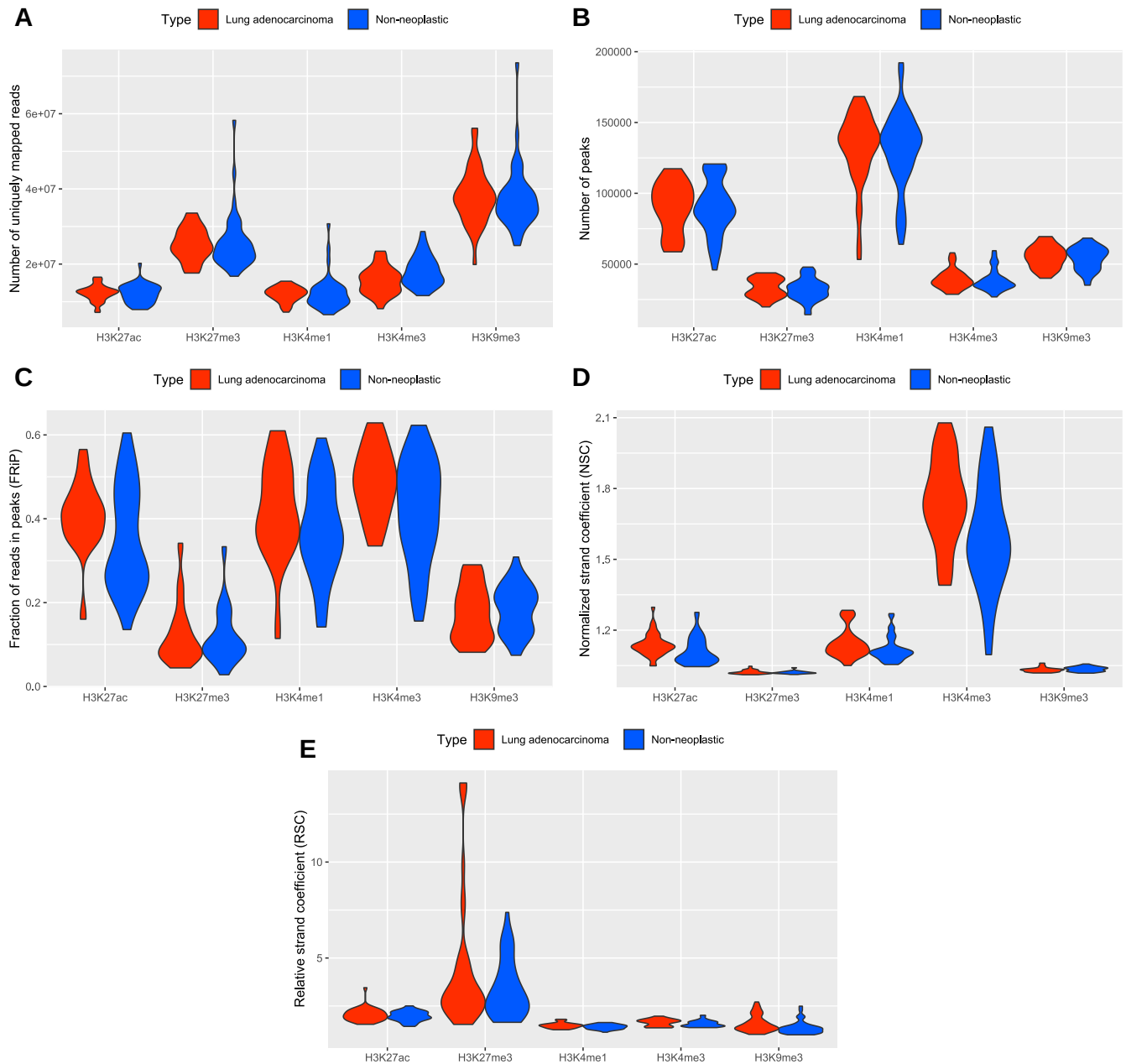


Figure 8: Distribution of five key quality control matrices for ChIP-seq data. (A) Number of uniquely mapped reads. (B) Number of called peaks. (C) Fraction of reads in peaks (FRiP). (D) Normalized strand cross-correlation coefficient (NSC). (E) Relative strand cross-correlation coefficient (RSC).

that the lung cells contained more molecules that could degrade RNA than brain cells. These molecules were released into the suspension during nuclei extraction and degraded the mRNA in nuclei before RNA extraction. Therefore, we used the RNAlater-ICE reagent to protect RNA

in the lung tissue. The reagent was prechilled to -80 °C. Thin tissue slices cut from the sample were transferred into RNAlater-ICE and allowed to incubate at -20 °C overnight. The tissue slices permeated with RNAlater-ICE were then homogenized and lysed for RNA extraction. This way we were able to produce cDNA of normal size distribution.

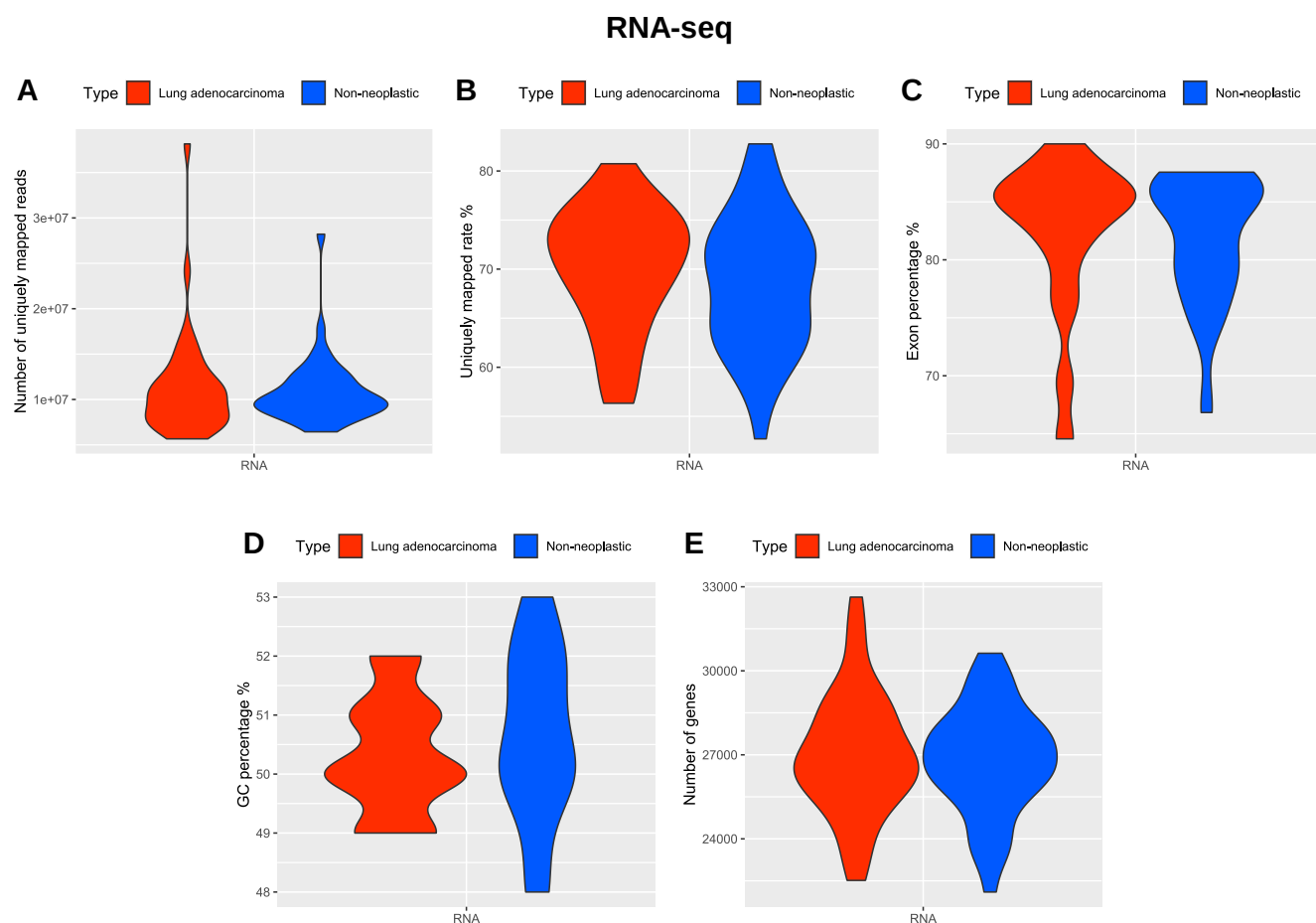


Figure 9: Distribution of five key quality control matrices for RNA-seq data. (A) Number of uniquely mapped reads. (B) Uniquely mapped rate. (C) Exon percentage. (D) GC percentage. (E) Number of genes covered.

Sequencing data processing and quality

We generated an average of 12.2, 17.0, 12.2, 25.5, 37.8 and 11.1 million uniquely mapped reads per library for H3K27ac, H3K4me3, H3K4me1, H3K27me3, H3K9me3 and RNA-seq. The MOWChIP-seq process gave a high signal-to-noise ratio, with the average FRiP for the five histone modifications at 0.377, 0.450, 0.396, 0.122 and 0.181 (Figure 8). The ChIP-seq

libraries also showed high reproducibility, with the average Pearson's correlation between technical replicates at 0.93, 0.95, 0.93, 0.92 and 0.95. RNA-seq libraries had an average mapping rate of 69.3% and exon rate of 82.2% (Figure 9).

Differential binding and TF regulation analysis

ChIP-seq data was analyzed using DiffBind⁸¹ to reveal DMERs in each of the five histone modifications. 27,233 H3K27ac (13,415 gain, 13818 loss in tumor compared to non-neoplastic), 44,677 H3K4me1 (21,132 gain, 23,545 loss), 13,005 H3K4me3 (8,725 gain, 4,280 loss), 5,091 H3K27me3 (2,877 gain, 2,214 loss), 18,589 H3K9me3 (11,211 gain, 7,378 loss) DMERs were identified. RNA-seq data was analyzed using DESeq2.¹¹¹ 1,053 DEGs were identified. The top 96 DEGs showed a clear separation in signal intensity pattern between tumor and non-neoplastic samples (Figure 10A). We used g:Profiler¹¹² to examine their related biological function and found the top DEGs to be highly enriched in cancer and cell proliferation/apoptosis pathways (Figure 10B). DMERs from the five types of histone modifications were also examined the same way. In particular, “pathways in cancer” was identified in four out of the five types. Other cancer and cell proliferation related pathways were also found here, including “Ras signaling pathway”, PI3K-Akt signaling pathway”, “MAPK signaling pathway” and “cAMP signaling pathway” (Figure 10C).

We used *Taiji* and *EpiTensor* pipeline to investigate the underlying transcription regulation and look for change between tumor and non-neoplastic samples.^{113,114} H3K27ac and RNA-seq data from each sample were integrated to assemble a network of TF-gene interaction. *Taiji* used the H3K27ac data to predict TF binding sites with known motifs in active promoter and enhancer regions. Enhancers were then linked to their interacting promoters predicted by

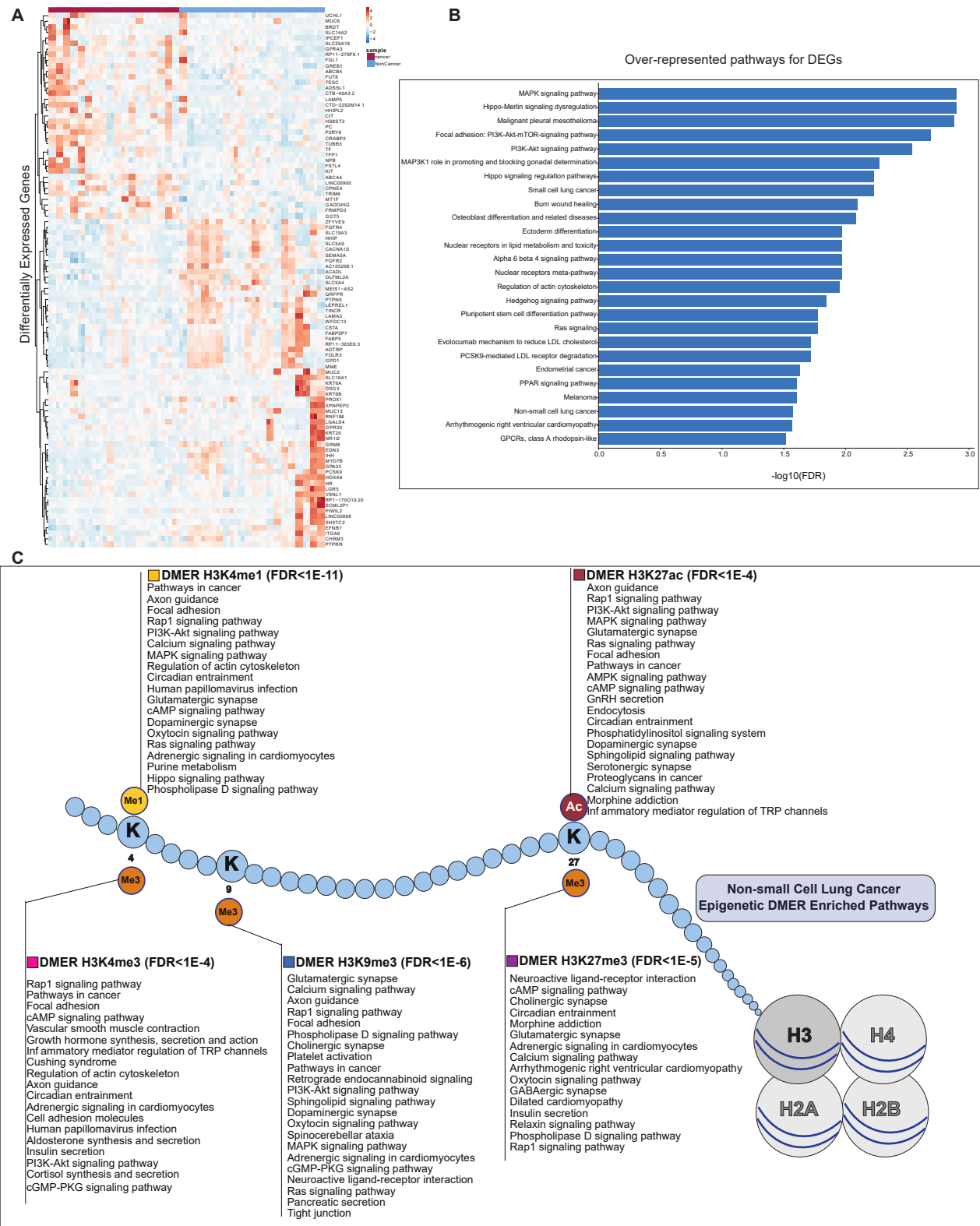


Figure 10: Differential binding analysis of tumor and non-neoplastic lung tissues. (A) Heatmap of top ranked DEGs. (B) Over-represented pathways of top ranked DEGs. (C) Enriched pathways for DMERs in each histone modification.

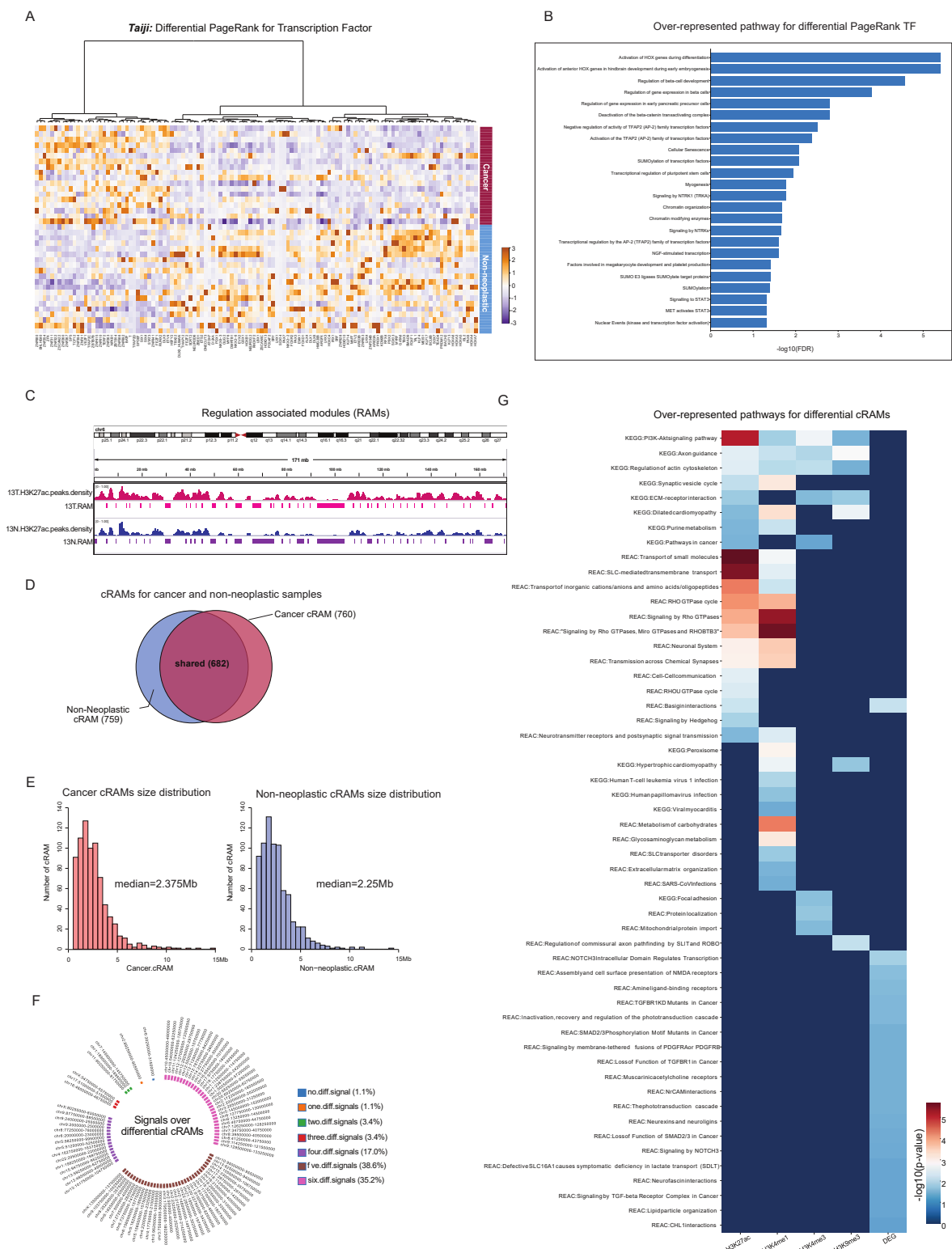


Figure 11: Differential regulatory analysis of tumor and non-neoplastic lung tissue. (A) Differential page rank TF heatmap. (B) Over-represented pathways of differential TFs. (C) Example track showing individual RAMs (D) Venn diagram of cRAM from tumor and non-neoplastic samples. (E) cRAM size distribution. (F) DMER and DEG distribution in viable cRAMs. (G) Over-represented pathway in differential cRAMs with DMER and DEG signal.

EpiTensor. The TF-gene pairs were combined into a network for each sample and PageRank score for each node was computed to reflect its importance. We compared the PageRank score for TFs in tumor and non-neoplastic samples and identified 118 differential TFs (Figure 11 A). The over-represented pathways for the differential TFs were also identified, with chromatin organization being of particular interest (Figure 11B).

Regulation associated module analysis

Strong association between changes in chromatin organization and cancer progression have been previously identified.¹¹⁵⁻¹¹⁷ In order to further examine the state of chromatin organization in our data, we applied the software suite *findRAM* to locate regulation associated modules (RAM). RAMs are genome wide modular patterns that reflect diverse chromatin activities at a better scale. Deletion of RAM boundaries would severely affect chromatin organization and cell fitness.¹¹⁸ Peaks from H3K27ac data were used to identify RAMs in individual samples, and consensus RAMs (cRAM) were then assembled from either tumor or non-neoplastic samples. 760 cRAMs were identified in tumor samples and 759 cRAMs were identified in non-neoplastic samples. Among these, 682 cRAMs were shared between the two groups, with more than 50% mutual overlap (Figure 11D). 78 cRAMs were specific to tumor samples and 77 cRAMs were specific to non-neoplastic samples. Overlap between these two groups of cRAMs were merged and 88 cRAMs remained, referred to as “viable cRAMs”. cRAMs specific to one of the two conditions indicated the changes in chromatin organization related to the pathological state. By mapping DMERs and DEGs identified from differential binding analysis onto viable cRAMs, we found that 91% of the viable cRAMs overlap with four or more types of DMERs and DEGs (Figure 11F). g:Profiler revealed that DMERs and DEGs that overlap with viable cRAMs were enriched in a variety of pathways (Figure 11G). New terms aside from the common ones

related to cancer emerged. Several terms were of particular interest, including “ECM-receptor interaction”, “loss of function of TGFBR1 in cancer”, “signaling by TGF-beta receptor complex in cancer”, and “loss of function of SMAD2/3 in cancer”. Remodeling of ECM pathways has been associated with cancer cell survival and metastasis.¹¹⁹⁻¹²² TGF-beta and SMAD2/3 both play key roles in non-small cell lung cancer.^{123,124}

Global co-modification patterns clustering

Aside from the differential analysis between tumor and non-neoplastic samples, we also sought to examine the global histone modification pattern in all samples, particularly the interaction between the 6 markers profiled. An unbiased, unsupervised integrative analysis was done using *EpiSig*.^{125,126} *EpiSig* scanned the entire human genome in 5kb windows and retained windows with significant signal for any one of the five histone modifications or gene expression across all samples. Windows with signal were then clustered based on similarity of co-modification patterns, resulting in 429 clusters in total (average 66 windows per cluster). k-means clustering was then applied to combine these clusters into 6 large sections, each with an overall co-modification pattern of epigenetic and transcriptomic signals (Figure 12A). Section I had strong H3K27ac and H3K4me1 signal, with weaker H3K4me3 signal and gene expression, and were in intron and intergenic regions, indicating active enhancers. Section II had a strong H3K4me3 signal with noticeable H3K27ac, H3K4me1 and gene expression signal, indicating active promoters. This section distributed itself across major types of genomic annotations except 3' UTR. Section III had noticeable gene expression and was concentrated in intron and coding exon regions. Section VI was similar to section III except for higher gene expression levels. Both were likely transcribed regions of the genebody.

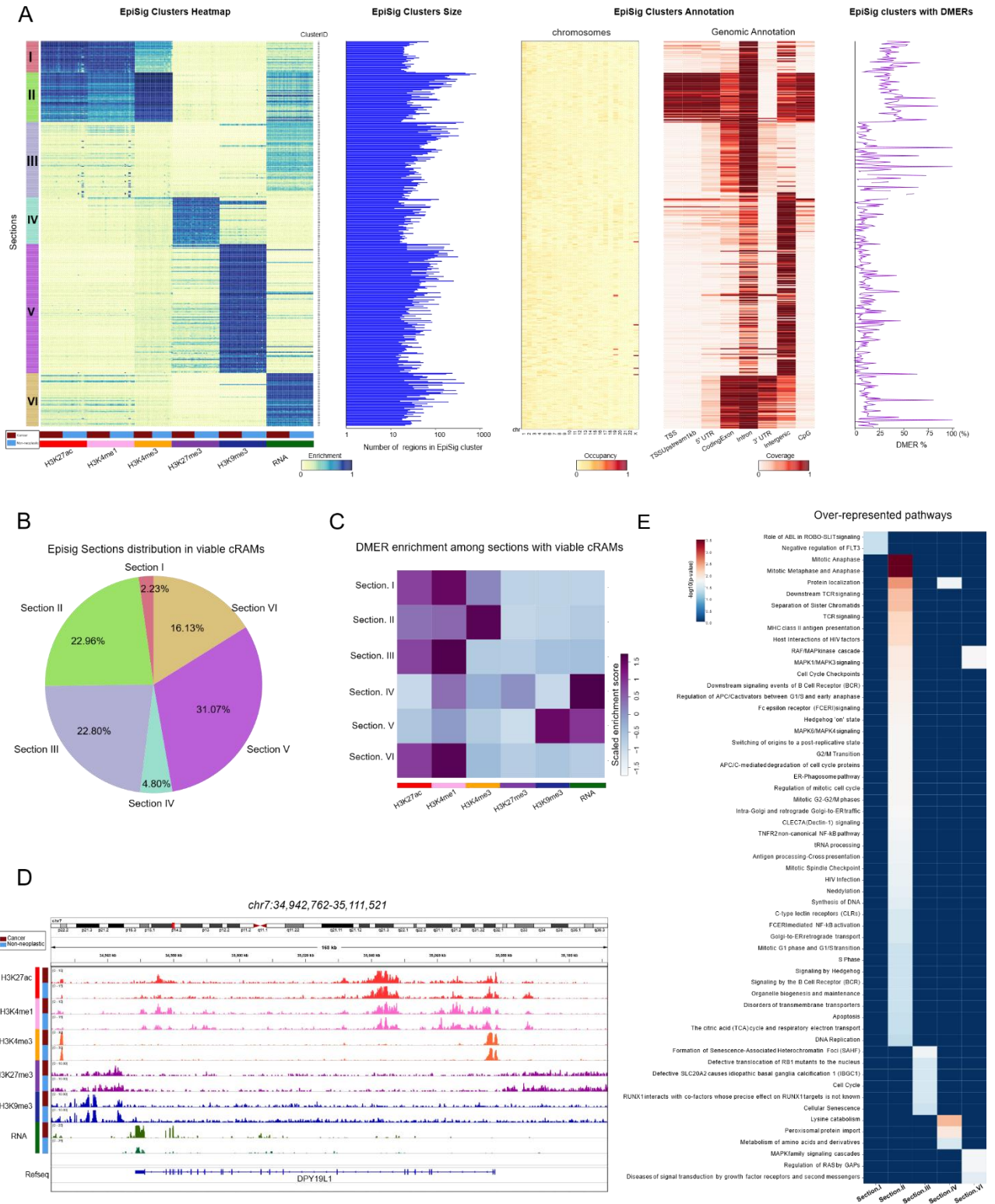


Figure 12: Integrated analysis of co-modification pattern and viable cRAM. (A) EpiSig co-modification clusters and annotation. (B) Distribution of co-modification sections in viable cRAMs. (C) DMER and DEG enrichment in sections with viable cRAM overlap. (D) An example showing DMER and DEG within a viable cRAM region. (E) Over-represented pathways in sections with viable cRAMs.

Section IV and V were both repressed regions enriched in H3K27me3 and H3K9me3, respectively, and no significant signal from active markers or gene expression. They were distributed mostly in intergenic regions and showed bias towards chromosome 19 and X. These two sections also showed a lower percentage of DMER and DEG coverage compared to the other four sections.

We linked global co-modification pattern back to viable cRAMs by mapping the 6 *EpiSig* sections over the viable cRAM regions. The distribution of sections within viable cRAMs were not to the ratio of their relative size (Figure 12B). DMER and DEG distribution within these overlapping regions also showed a distinct pattern (Figure 12C): section I, III and VI were enriched with H3K27ac and H3K4me1 DMERs, section II enriched with all three active histone mark DMERs, section IV enriched with RNA DEGs, and section V enriched with H3K9me3 DMERs and RNA DEGs. Pathway analysis of each section's overlap with viable cRAMs revealed distinct functionalities except section V (Figure 12E). Section I showed enrichment "Negative regulation of FLT3" in tumor samples. This is interesting because FLT3 is a biomarker in hematopoiesis study and important for the normal development of immune system and stem cells.¹²⁷ Section II was enriched with critical cancer related pathways including "TCR signaling", "MAPK1/MAPK3 signaling", "MAPK6/MAPK4 signaling", "signaling by the B cell receptor", "TNFR2 non-canonical NF-kB pathway", "apoptosis", "DNA replication" and "cell cycle checkpoints", highlighting the importance of these loci in non-small cell lung cancer.

Discussion

Here we present comprehensive multi-omics profiling of human lung tumor and non-neoplastic tissues. Our low-input MOWChIP-seq technology allowed a large number (5) of histone marks

and gene expression to be profiled simultaneously, while using very small quantities of tissue sample (~20 mg) and achieving high data quality. The bench top workflow and subsequent bioinformatics pipeline could serve as a template for future large-scale studies of scarce tissue samples where multiple histone modifications and other targets are to be mapped.

Furthermore, the data we generated will be a valuable resource to the field of lung cancer research because of the detailed epigenomic and transcriptomic characterization of both normal lung tissue and non-small cell lung tumor, similar to the examination of prostate cancer by Li *et al.*¹⁰² The data generated in this study represents the first sizable effort in comprehensive epigenomic and transcriptomic mapping of both open and closed chromatin in lung cancer, conducted using a proven low input ChIP-seq technology.

Previous epigenomic studies of patient derived tumor samples focused on the open chromatin state⁹⁶ and active histone modifications.^{97,100} Repressive histone modifications associated with downregulation of gene expression and heterochromatin formation were usually left out. The reason could be a combination of technical difficulty in generating high quality ChIP-seq data on repressive marks and the presumption that they yield less information compared to active marks. We included two repressive marks (H3K27me3, H3K9me3) in our study apart from the three active marks. This allowed us to define chromatin state in various genomic regions with higher precision for both open and closed chromatin regions, and to integrate six different data types together and separate epigenomic regions into distinct groups based on their co-modification pattern. We identified 6 main sections each with a distinct pattern of histone modifications and gene expression, with 4 sections highly enriched in DMERs and DEGs. We found H3K27me3 and H3K9me3 to occupy two clusters of regions with a small amount of overlap, outside of the genome marked by active modifications and gene expression. While

both are repressive marks in nature, H3K9me3 is seen as a permanent mark of repression while H3K27me3 is more temporary.¹²⁸ Co-modification pattern analysis also showed that H3K9me3 covered a much larger portion of the genome compared to H3K27me3. We also found 18,589 differentially modified regions in H3K9me3 compared to 5,091 in H3K27me3, suggesting that epigenomic change between normal tissue and tumor can be reliably captured by both temporary and permanent repressive marks. Overall, our results demonstrate that not only is high quality mapping of repressive marks feasible with suitable tools, but also necessary to gain a full picture of epigenomic landscape.

Solid tumors are complex assemblies of cancerous cells and a variety of noncancerous cells that either promote or inhibit tumor development.¹²⁹ Our experiment did not separate the cells into various subpopulations. It is likely that the difference we observed between normal and tumor samples represented the combination of tumorigenesis and the promoting/inhibiting effect from other cell types. This “bulk” approach captures the full range of cell types found in tissue samples. However, the contribution of different cell populations cannot be easily separated. Corces *et al.* theorized that it may be possible to dissect the complex data from tumors to reveal the contribution of different cell types.⁹⁶ However, methods such as cellular deconvolution have so far only been developed for transcriptomic and methylome mapping data.^{130,131} A similar workflow optimized for multiple histone modifications will assist in understanding the roles of different cell types in tumor. It is useful to capture the big picture in an early-stage study and provide the full information as a resource. This will guide potentially more refined studies where the scope is narrowed down. It will be worthwhile to investigate specific cell types, isolated via techniques such as FACS, or move onto single cell methods that could distinguish subpopulations of cells within the bulk tissues.^{132,133}

We previously identified a modular pattern in histone modifications, referred to as regulation associated modules (RAM), that reflected spatial chromatin organization. Somatic mutations were consistently enriched in RAMs in cancer patients. This suggests a possible link between tumorigenesis and changes in RAM, which have major impact on chromatin organization (23). Here, we identified 88 viable cRAMs composed of 78 tumor specific and 77 non-neoplastic specific cRAMs, apart from 682 cRAMs shared by both conditions. The viable cRAMs should effectively sum up the change in RAM boundaries from tumorigenesis. Indeed, we found that 91% of the viable cRAMs overlap with four or more types of DMERs and DEGs out of the six types examined, indicating high correlation between differential epigenomic and transcriptomic signal and tumor/normal-specific cRAMs. If only the viable cRAMs were considered, the DMER and DEG enrichment pattern of the six sections would change significantly. Further pathway analysis of sections within viable cRAMs presented unique functions, including many cancer related pathways, for five of the sections. This demonstrates the ability of viable cRAMs to extract information from the genome-wide trend due to their high relevance to the difference between tumor and non-neoplastic samples. The above analyses correlated well with each other and further supported the robustness of RAMs as modules of chromatin organization. The common patterns in chromatin organization resided in cRAMs shared by tumor and non-neoplastic samples, while pathologically important changes concentrated in viable cRAMs specific to either sample type. Our results are evidence that RAMs could provide a useful guide in analyzing large epigenomic and transcriptomic datasets. It can condense numerous DMERs and DEGs scattered as small windows into a much smaller number of functionally important regions.

Materials and methods

Human lung tissue selection

De-identified and archived human non-small cell lung cancer (N = 18) and non-neoplastic lung tissue (N = 20) were obtained from Biorepository and Tissue Technology Shared Resource at Moores Cancer Center, UCSD. Both UCSD and Virginia Tech IRB office reviewed the study protocol and deemed the project as “Not Human Subject Research”. All tumor samples were of pathological stage I and II. Frozen samples were sealed in 1.8 ml cryovials and stored in liquid nitrogen vapor phase before use.

Fabrication of microfluidic device

Same as in Methods of Chapter 2.

Nuclei isolation from human lung tissue

Slices of tissues were dissected off the large piece on dry ice using a sharp scalpel and moved into a tissue grinder set (D9063, Sigma-Aldrich) pre-chilled on ice. The following steps are performed on ice and centrifugation performed at 4°C. 3 ml of ice-cold nuclei extraction buffer [0.32 M sucrose, 5 mM CaCl_2 , 3 mM $\text{Mg}(\text{Ac})_2$, 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] was added into the tissue grinder. Tissue slices were homogenized by slowly douncing 20 times with pestle A and 30 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. Purified nuclei pellet was resuspended in 100 µl of Dulbecco's phosphate-buffered saline (DPBS).

MNase digestion of chromatin

Resuspended nuclei were counted to determine concentration and diluted to 4×10^6 /ml. Our experiments usually started with 4×10^5 cells/nuclei suspended in 100 µl of DPBS (protocol is scalable in volume). 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_2] were added, mixed by vortexing and incubated at room temperature for 10 min. 10 µl of 100 mM CaCl_2 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. 30 µl of chromatin solution (~50K nuclei) was used for each MOWChIP assay.

Immunoprecipitation beads preparation

5 µl 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 µl of IP buffer. We used the following antibodies in this work: anti-H3K27ac (ab4729, Abcam), 0.125 µg/assay; anti-H3K4me1 (39297, Active Motif), 0.5 µg/assay; anti-H3K4me3 (ab8580, Abcam), 0.125 µg/assay; anti-H3K27me3 (39155, Active Motif), 0.5 µg/assay; anti-H3K9me3 (ab8898, Abcam), 0.125 µg/assay. The suspension was

incubated on a rotator at 4°C overnight. The beads were then washed with IP buffer twice and resuspended in 5 µl of IP buffer for loading into the device chamber.

Operation of MOWChIP device

Same as in Methods of Chapter 2.

Purification of ChIP DNA

Same as in Methods of Chapter 2.

Library preparation, quantification and sequencing

Same as in Methods of Chapter 2.

RNA-seq

Tissue slices were placed in 200 µl of RNAlater-ICE solution (AM7030, Thermo Fisher Scientific) pre-chilled to -80°C, and then incubated overnight at -20°C. Each library was prepared from 10 ng RNA extract. Two technical replicates were prepared per sample. RNA was extracted from the incubated tissue slices using RNeasy Mini Kit (74104, Qiagen) and RNase-free DNase Set (79254, Qiagen). Extracted RNA was resuspended in 4.5 µl of RNase-free water with 5% RNase inhibitor added. 2 µl of oligo-dT primer (10 µM), 2 µl of dNTP mix (10 mM) and 4.3 µl of total RNA solution were mixed, incubated at 72°C for 3 minutes and immediately put on ice. Then 1 µl of SuperScript II reverse transcriptase (200 U/µl), 0.5 µl of RNase inhibitor (40 U/µl), 4 µl of Superscript II first-strand buffer, 1 µl of DTT (100 mM), 4 µl of 5 M Betaine, 0.12 µl of 1 M MgCl₂, 0.2 µl of TSO (100 µM), 0.88 µl of nuclease-free water were added to the 8.3 µl mixture mentioned above. The reverse transcription (RT) reaction mix was then incubated at 42°C for 90 minutes, followed by 10 cycles of 50°C for 2 minutes and 42°C

for 2 minutes, and lastly 70°C for 15 minutes. The RT reaction mix was mixed with 25 µl KAPA HiFi HotStart ReadyMix, 0.5 µl IS PCR primers (10 µM), 2.5 µl Evagreen dye, and 2 µl nuclease-free water, and amplified with: 98°C for 1 minute, 9–11 cycles of 98°C 15 seconds, 67°C 30 seconds, 72°C 6 minutes. The cDNA was purified using 30 µl of SPRIselect beads and eluted in 5 µl of low EDTA TE buffer.

The RNA-seq library was made using tagmentation. 3 µl of home-made Tn5 was mixed with 3 µl of 50 µM pre-annealed P5/P7 transposon and incubated at 37°C for 1 hour. Assembled P5 and P7 transposons were then mixed well to form Tn5 mix. 1 µl cDNA (600 pg) and 1 µl of tagmentation buffer (TNP92110, Lucigen) were added to 8 µl Tn5 mix, and the mixture was incubated at 37°C for 1 hour. 1 µl of 10X stop buffer (TNP92110, Lucigen) was then added to the mixture to quench tagmentation. 11 µl SPRIselect beads were used to purify tagmented cDNA. The cDNA was then eluted in 9.5 µl of low EDTA TE buffer. 25 µl of KAPA HiFi HotStart ReadyMix was incubated at 98°C for 30 seconds, mixed with 9.5 µl tagmented cDNA and incubated at 72°C for 5 minutes. 1.5 µl of 100 µM P5 primer, 1.5 µl of 100 µM P7 primer, 10 µl nuclease-free water and 2.5 µl Evagreen were then added and amplified with: 98°C for 30 seconds; 10-12 cycles of 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 30 seconds. Amplified cDNA libraries were purified with 50 µl SPRIselect beads and eluted in 8 µl of low EDTA TE buffer. RNA-seq libraries were sequenced with Illumina HiSeq 4000 platform.

ChIP-seq data analysis

Raw sequencing FASTQ files were trimmed using Trim Galore! with default settings. Trimmed reads were then aligned to hg19 genome using bwa.¹³⁴ After deduplication, uniquely mapped reads with MAPQ ≥ 10 were retained for processing. Narrow peaks (H3K27ac, H3K4me1, H3K4me3) were called using MACS2 (2.2.7.1) with q-value ≤ 0.05 . Broad peaks (H3K27me3,

H3K9me3) were called using Homer *findPeaks* function in histone mode with -p-value ≤ 0.05 . Differential binding peaks for histone modifications were identified using DiffBind with default setting.⁸¹

RNA-seq data analysis

Raw sequencing FASTQ files were trimmed using Trim Galore! with default settings. Trimmed reads were aligned to the hg19 genome using STAR.¹³⁵ Reads with Q30 were kept for further processing. *featureCounts* was used to quantify gene expression in each sample. DEGs were identified using DESeq2.

Transcription factor analysis using *Taiji*

Processed H3K27ac peaks and RNA-seq data were loaded into *Taiji* pipeline with default parameters. 3D contacts for the regulatory elements were identified using unsupervised learning algorithm *EpiTensor*. Putative TF binding motifs were curated from the CIS-BP database.¹³⁶

Regulatory associated modules (RAM) analysis

H3K27ac peak density was calculated with a sliding window of step size = 250 kb, flanking size = 500 kb for each window in all samples. RAM boundaries and peaks were detected using *findRAM*. RAMs from the 18 tumor samples and 20 non-neoplastic samples were combined into their respective groups. Regions with more than 25% presenting frequency in each group were identified as consensus RAM (cRAM) for the group. cRAMs were filtered to keep those larger or equal to 500 kb in size. cRAM boundaries located within 250 kb were merged.

Genome-wide co-modification pattern analysis

The genome-wide co-modification of epigenetic and gene expression signal was clustered using the unsupervised learning method EpiSig. Sequencing data were normalized, adjusting for total sequencing depth. Regions within the ENCODE blacklist were removed. ChIP-seq signals were subtracted for input noise. ChIP-seq and RNA-seq data were then loaded into EpiSig pipeline. EpiSig divided the genome into 5kb bins and detected signal enrichment of histone modifications and gene expression in all bins among all samples. Bins with similar signal patterns were clustered into one EpiSig cluster. For better interpretation of the data, individual EpiSig clusters were further concentrated using k-means clustering. The optimal number of clusters was determined by Elbow method.

Chapter 4: Spatially resolved epigenetic mapping of mouse neocortex

Introduction

Gene expression is influenced by multiple levels of epigenetic and transcriptomic control. Epigenomic and transcriptomic profiling reveal rich information on the regulation of important biological processes involved in development and disease. While many epigenomic and transcriptomic studies were done using cell lines, tissue samples provide superior *in vivo* information and reflect conditions inside an organism which cell cultures cannot reproduce. One such aspect is the spatial complexity of tissue and organ, and the three-dimensional interactions between different cell populations and organization of tissue structure. Because of this, spatial resolution is an important merit to consider when conducting epigenetic and transcriptomic mapping of tissue samples with complex structures. Various transcriptomic methods have been developed with spatial resolution up to single cells, providing key insights into the cell organization, tissue microenvironment and cytoarchitecture.¹³⁷⁻¹⁴³ In comparison, few reports on a similar method exist for epigenetic mapping of tissue samples. Conventional assays require millions of cells as input material per run. As a result, they are typically optimized for cell lines which are uniform in nature. The complexity of real tissue samples, be it the spatial arrangement of different cell types, their interactions and other variability are unlikely to be replicated in cell culture. Tissue samples can be processed into small enough pieces with relative ease and become spatially resolved. However, the individual ‘pixels’ of the tissue usually cannot produce enough material at this stage for meaningful epigenetic examination. Recent works based on single-cell methyl-seq¹⁴⁴ and CUT&Tag¹⁴⁵ demonstrate

that certain spatial resolution can be achieved for epigenomic markers. However, the complexity and cost of single-cell assays make them impractical for examining a large volume of tissue.

This dilemma is reflected in that while there have been several reports on spatially resolved transcriptomic profiling of tissue sample with highly interesting results,^{6,146} no work has been reported on histone modification in the same manner. A recent study by Joseph Ecker's group demonstrated the intricacy of spatially resolved brain regions in terms of DNA methylation state using single-cell bisulfite sequencing.¹⁴⁴ Even with their mature single-cell pipeline, pooling of samples from 6 – 30 mice was necessary to generate the dataset.

The development of low-input, MNase-based ChIP-seq methods makes it practical to profile histone modification and other targets in spatially resolved tissue pieces. The yield from each piece would be enough to conduct ChIP-seq with replicates and possibly more than one marker. The use of MNase makes the process much shorter while minimizing the loss of material and variability in fixation/sonication. Our group have previously applied this approach to demonstrate the extensive differences between mouse PFC and cerebellum across multiple histone marks.³ ChIP-seq profiling of histone modifications of brain tissue in a spatially resolved manner is the next logical extension.

Epilepsy is a neurological condition characterized by recurrent seizures as a result of abnormal and excessive neuron firing.¹⁴⁷ Research into both animal models and human samples reveal that epilepsy causes significant changes to the epigenome of the brain, including histone modification. These changes then lead to regulation of gene expression and cascading downstream effects in the brain.¹⁴⁸⁻¹⁵⁰ Insights into the change in genome-wide histone

modification profile are important to further our understanding of the progression of epilepsy and potential treatments.

In this project, we conducted spatially resolved mapping of histone modifications and mRNA expression in mouse neocortex. The whole mouse brain was processed into 0.5 mm-thick slices in a brain matrix and neocortex portions were dissected and separated into left and right portions. Nuclei were extracted from the tissue samples and sorted into neuronal (NeuN+) and glial (NeuN-) portions. A total of 5 mice were used in this study. We examined polycomb repression mark H3K27me3 in neurons and glia from both left and right neocortices using MOWChIP-seq (M1-2). We conducted RNA-seq on the neurons from the right neocortex using Smart-seq2 (M3). We further examined the effect of epilepsy by inducing seizures via kainic acid injection (M4) and compared the H3K27ac and H3K27me3 profiles of neurons from right neocortex with a control mouse (M5).

Our analysis revealed distinct spatial patterns of histone modification in mouse neocortex and significant changes because of induced seizure. The differences were concentrated in specific regions of the neocortex. Differentially marked genomic regions associated strongly with brain functions and biological processes important in epilepsy. Using k-means clustering, we were able to sort the top features into individual functional groups based on their spatial distribution. GO analysis of the biological process terms associated with each group indicated their specific role in brain function.

Author contribution

Zhengzhi Liu conducted all mouse brain dissection and isolation of the neocortex slices.

Chengyu Deng carried out MOWChIP-seq and RNA-seq experiments of mouse M1-M3.

Zhengzhi Liu carried out MOWChIP-seq experiments of mouse M4-M5 and data analysis. Injections to produce mouse M4 and M5 were conducted by Jiang Shan with generous help from Xiaoting Jia's group at Virginia Tech. The project received helpful advice from Melissa Makris, who operated the FACS sorting service at Virginia Tech.

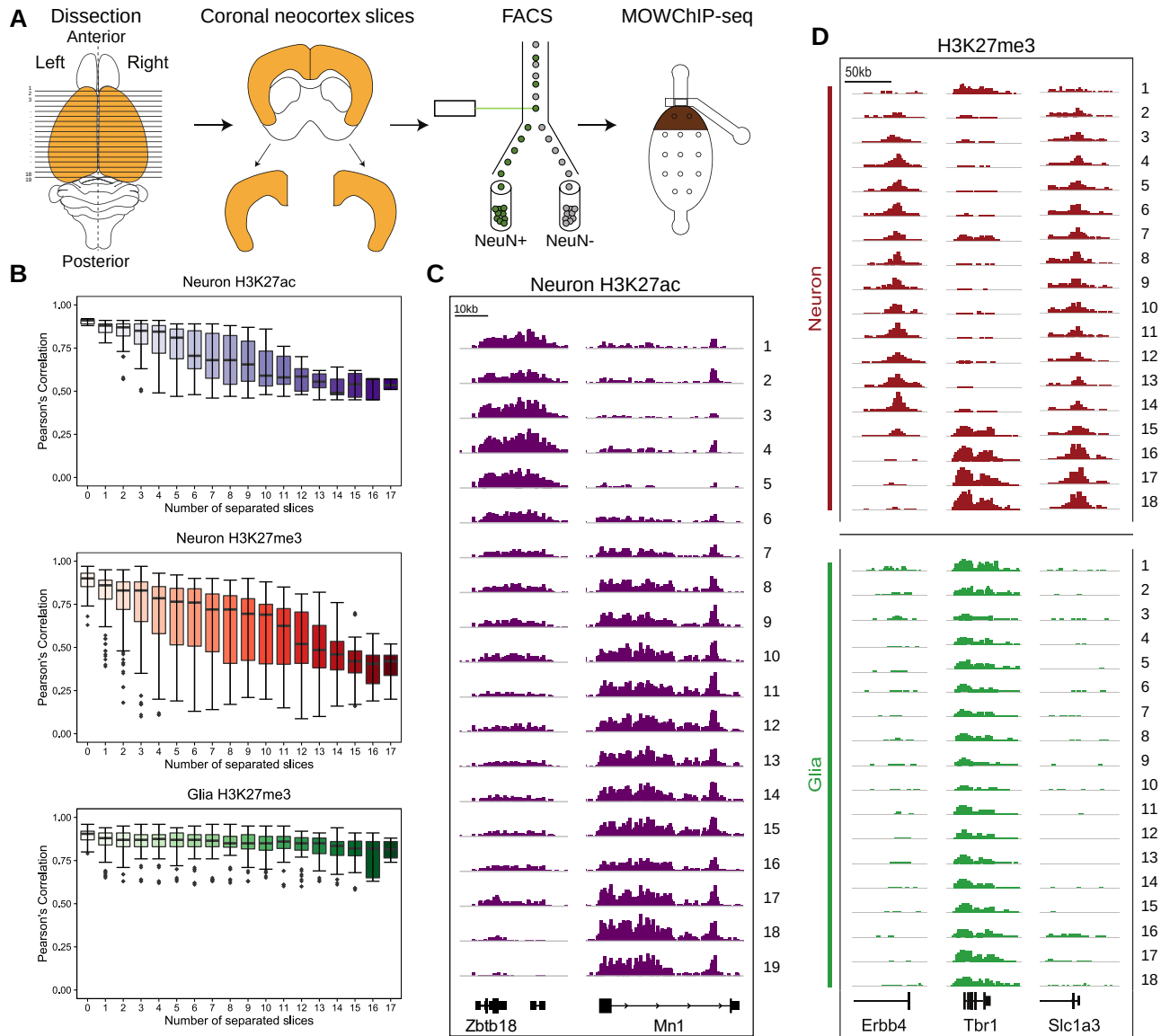


Figure 13: Overall process of spatial resolved profiling of mouse neocortex. (A) Illustration of workflow. (B) Pearson's correlation between samples separated by different number of slices. (C) Normalized H3K27ac ChIP-seq signal in mouse neurons (M5). (D) Normalized H3K27me3 ChIP-seq signal in neurons and glia in mouse neurons and glia (M2).

Results and discussion

Sample processing and MOWChIP-seq

Brain samples were extracted from 10-week male C57BL/6J mice. M1-3 were euthanized by compressed CO₂. M4 was injected with kainic acid to induce seizure. M5 was injected with saline as control. M4-5 were euthanized by direct cervical dislocation to avoid changes to epigenetic profiles induced by CO₂ suffocation. Whole brains were dissected and cut into coronal slices of 0.5 mm thickness in a mouse brain slicing matrix. Each slice was examined, and the neocortex portion was separated into left and right portions. Neocortex slices were then used for nuclei extraction, FACS sorting and MOWChIP-seq. Each assay was done using around 4,000 sorted nuclei. 2 technical replicates per sample were prepared (Figure 13A). MOWChIP-seq yielded consistently high-quality data (Table 2). For example, M4-5 averaged 11069 peaks and FRiP of 18.4% for H3K27me3 and 43951 peaks and 14.4% FRiP for H3K27ac.

Table 2: MOWChIP-seq data summary of M4 and M5.

Mouse	Slice	Mark	Rep	Mapped Reads	Mapping Rate	Unique Reads	Unique Rate	Peaks	FRiP
M4	1	H3K27ac	1	15129600	89.3%	14974041	99.0%	55721	29.3%
M4	1	H3K27ac	2	14851829	91.4%	14746704	99.3%	58955	34.4%
M4	2	H3K27ac	1	20787168	92.7%	20695123	99.6%	47508	17.5%
M4	2	H3K27ac	2	18215617	93.2%	18146039	99.6%	54318	19.9%
M4	3	H3K27ac	1	13834773	92.2%	13785668	99.6%	48784	17.1%
M4	3	H3K27ac	2	16739229	90.1%	16629112	99.3%	42173	14.5%
M4	4	H3K27ac	1	16373079	91.4%	16331849	99.7%	46710	14.8%
M4	4	H3K27ac	2	16439446	91.2%	16396689	99.7%	45496	14.3%
M4	5	H3K27ac	1	15419107	77.4%	15364342	99.6%	31150	7.3%
M4	5	H3K27ac	2	16331106	75.6%	16257442	99.5%	30305	7.2%
M4	6	H3K27ac	1	15541069	83.7%	15492350	99.7%	41314	11.0%
M4	6	H3K27ac	2	15441248	82.5%	15383458	99.6%	41078	11.0%
M4	7	H3K27ac	1	15143393	86.2%	15092721	99.7%	40304	10.8%
M4	7	H3K27ac	2	15149694	84.7%	15077798	99.5%	37434	10.4%
M4	8	H3K27ac	1	14487281	85.0%	14440781	99.7%	42127	10.6%
M4	8	H3K27ac	2	14344921	83.4%	14295697	99.7%	43645	11.2%
M4	9	H3K27ac	1	14720207	85.9%	14655758	99.6%	48058	13.9%
M4	9	H3K27ac	2	17282495	84.9%	17190306	99.5%	49403	14.4%
M4	10	H3K27ac	1	28911856	81.7%	28796178	99.6%	36941	9.8%
M4	10	H3K27ac	2	25165867	81.9%	25079172	99.7%	35370	9.7%

M4	11	H3K27ac	1	21527910	86.3%	21468975	99.7%	39536	10.7%
M4	11	H3K27ac	2	21360600	84.7%	21298679	99.7%	39506	10.6%
M4	12	H3K27ac	1	16965122	91.8%	16834267	99.2%	44382	12.2%
M4	12	H3K27ac	2	20985104	90.5%	20838728	99.3%	35347	10.0%
M4	13	H3K27ac	1	24589243	91.3%	24531131	99.8%	37382	10.7%
M4	13	H3K27ac	2	25250034	92.0%	25177828	99.7%	36034	10.1%
M4	14	H3K27ac	1	21884955	81.2%	21825688	99.7%	13495	3.4%
M4	14	H3K27ac	2	18715186	80.7%	18676086	99.8%	27399	6.5%
M4	15	H3K27ac	1	24485662	88.3%	24443394	99.8%	25257	6.3%
M4	15	H3K27ac	2	20311336	88.9%	20273892	99.8%	26673	6.6%
M4	16	H3K27ac	1	29376050	84.8%	29298148	99.7%	22869	5.9%
M4	16	H3K27ac	2	29570056	89.9%	29481959	99.7%	21741	5.6%
M4	17	H3K27ac	1	17587478	82.1%	17534456	99.7%	10775	2.7%
M4	17	H3K27ac	2	18199589	82.7%	18155806	99.8%	13949	3.5%
M4	18	H3K27ac	1	15132723	77.6%	15029502	99.3%	36209	8.5%
M4	18	H3K27ac	2	16052819	76.9%	15968296	99.5%	29500	7.2%
M4	19	H3K27ac	1	17712640	70.9%	17614735	99.4%	34225	8.9%
M4	19	H3K27ac	2	19386261	81.5%	19209984	99.1%	44958	14.8%
M4	20	H3K27ac	1	18062999	89.9%	17941073	99.3%	33942	10.0%
M4	20	H3K27ac	2	17576948	89.8%	17446994	99.3%	32561	9.5%
M4	21	H3K27ac	1	20257691	86.8%	20180299	99.6%	48125	25.9%
M4	21	H3K27ac	2	22301542	86.5%	22188933	99.5%	49753	25.9%
M4	1	H3K27me3	1	18709424	92.2%	18140925	97.0%	10174	25.0%
M4	1	H3K27me3	2	20036276	92.4%	19426692	97.0%	10257	26.5%
M4	2	H3K27me3	1	22187518	89.6%	21387422	96.4%	12361	21.9%
M4	2	H3K27me3	2	24082302	90.1%	23236961	96.5%	12314	23.1%
M4	3	H3K27me3	1	24944713	89.5%	24134923	96.8%	12752	23.1%
M4	3	H3K27me3	2	23317393	89.1%	22554507	96.7%	12501	22.1%
M4	4	H3K27me3	1	37549043	89.3%	36311276	96.7%	15099	23.4%
M4	4	H3K27me3	2	24414280	89.0%	23629545	96.8%	12068	19.7%
M4	5	H3K27me3	1	25943329	76.5%	25148226	96.9%	9788	16.5%
M4	5	H3K27me3	2	24990710	79.0%	24217735	96.9%	9863	16.7%
M4	6	H3K27me3	1	25615493	87.7%	24802326	96.8%	12098	16.4%
M4	6	H3K27me3	2	22545182	84.2%	21851125	96.9%	11536	14.5%
M4	7	H3K27me3	1	26872969	87.8%	26034588	96.9%	13678	19.3%
M4	7	H3K27me3	2	24461264	87.7%	23712080	96.9%	13282	17.8%
M4	8	H3K27me3	1	20043494	86.8%	19420243	96.9%	9237	16.0%
M4	8	H3K27me3	2	20838709	85.2%	20174155	96.8%	9412	16.8%
M4	9	H3K27me3	1	21904356	88.1%	21245137	97.0%	12033	23.4%
M4	9	H3K27me3	2	22902579	82.6%	22196320	96.9%	13797	17.5%
M4	10	H3K27me3	1	23059222	81.5%	22325299	96.8%	10626	19.7%
M4	10	H3K27me3	2	28491617	86.7%	27569683	96.8%	13184	17.6%
M4	11	H3K27me3	1	26802541	88.6%	25960905	96.9%	12539	16.0%
M4	11	H3K27me3	2	29312553	85.1%	28499696	97.2%	12926	16.3%
M4	12	H3K27me3	1	18974653	89.3%	18378962	96.9%	9109	15.2%
M4	12	H3K27me3	2	19837408	61.7%	19209285	96.8%	10046	17.5%
M4	13	H3K27me3	1	31862152	88.1%	30858384	96.8%	14983	23.8%
M4	13	H3K27me3	2	31472656	87.4%	30452602	96.8%	14546	22.7%
M4	14	H3K27me3	1	31676278	84.1%	30710402	97.0%	12042	16.4%
M4	14	H3K27me3	2	29280366	84.3%	28392688	97.0%	11611	15.4%
M4	15	H3K27me3	1	31929693	87.9%	31009420	97.1%	13712	21.6%
M4	15	H3K27me3	2	30710030	87.0%	29795170	97.0%	13534	20.2%
M4	16	H3K27me3	1	25063537	87.9%	24230550	96.7%	11887	16.4%

M4	16	H3K27me3	2	19124793	88.1%	18458499	96.5%	8489	15.1%
M4	17	H3K27me3	1	23147599	75.2%	22175749	95.8%	8352	16.0%
M4	17	H3K27me3	2	30484696	80.0%	29191691	95.8%	10846	14.8%
M4	18	H3K27me3	1	24685800	79.4%	23726492	96.1%	10727	15.0%
M4	18	H3K27me3	2	22413208	68.0%	21555570	96.2%	9092	18.6%
M4	19	H3K27me3	1	28919722	75.9%	27819481	96.2%	11449	16.6%
M4	19	H3K27me3	2	25549785	72.8%	24598605	96.3%	10587	14.4%
M4	20	H3K27me3	1	20215140	86.8%	19517710	96.6%	9927	19.4%
M4	20	H3K27me3	2	17130170	87.2%	16548195	96.6%	8722	15.5%
M4	21	H3K27me3	1	23439859	85.8%	22674860	96.7%	15990	27.7%
M4	21	H3K27me3	2	22669206	84.8%	21929556	96.7%	13399	31.7%
M5	1	H3K27ac	1	12824937	94.7%	12785362	99.7%	52923	19.4%
M5	1	H3K27ac	2	13156040	94.6%	13106409	99.6%	55988	21.3%
M5	2	H3K27ac	1	15430657	94.6%	15394083	99.8%	57332	18.0%
M5	2	H3K27ac	2	12701056	92.3%	12662989	99.7%	49718	15.9%
M5	3	H3K27ac	1	12111403	94.6%	12074775	99.7%	63617	26.4%
M5	3	H3K27ac	2	11710093	95.6%	11682491	99.8%	60761	25.2%
M5	4	H3K27ac	1	14512453	94.9%	14466073	99.7%	58122	24.9%
M5	4	H3K27ac	2	12767002	94.6%	12723822	99.7%	60417	26.5%
M5	5	H3K27ac	1	15771802	94.5%	15691987	99.5%	49106	21.5%
M5	5	H3K27ac	2	16505143	95.6%	16418826	99.5%	50654	21.1%
M5	6	H3K27ac	1	13330460	93.7%	13243061	99.3%	47997	13.8%
M5	6	H3K27ac	2	10842581	92.9%	10743297	99.1%	44546	13.0%
M5	7	H3K27ac	1	12787154	94.4%	12650006	98.9%	40038	10.7%
M5	7	H3K27ac	2	11822896	95.6%	11748402	99.4%	52506	14.2%
M5	8	H3K27ac	1	10392321	95.7%	10355446	99.6%	46629	13.7%
M5	8	H3K27ac	2	10635102	92.9%	10570423	99.4%	49359	15.0%
M5	9	H3K27ac	1	11396298	95.1%	11312315	99.3%	52855	16.0%
M5	9	H3K27ac	2	11072639	94.7%	10937845	98.8%	47663	14.2%
M5	10	H3K27ac	1	15526509	94.8%	15447067	99.5%	56503	17.6%
M5	10	H3K27ac	2	12291850	85.0%	12228070	99.5%	50461	16.4%
M5	11	H3K27ac	1	15928403	95.6%	15839544	99.4%	45097	13.9%
M5	11	H3K27ac	2	18556511	95.7%	18471813	99.5%	48005	14.7%
M5	12	H3K27ac	1	13759565	96.4%	13702962	99.6%	52612	17.1%
M5	12	H3K27ac	2	12156584	95.4%	12054021	99.2%	50874	16.0%
M5	13	H3K27ac	1	13894457	93.4%	13801115	99.3%	58068	17.1%
M5	13	H3K27ac	2	17542582	93.0%	17356054	98.9%	49880	14.6%
M5	14	H3K27ac	1	13334405	89.3%	13234246	99.2%	45187	13.4%
M5	14	H3K27ac	2	12237034	88.2%	12161300	99.4%	51916	15.3%
M5	15	H3K27ac	1	14367313	93.8%	14288414	99.5%	42942	12.6%
M5	15	H3K27ac	2	14149645	93.4%	14010206	99.0%	39451	11.5%
M5	16	H3K27ac	1	18055790	89.4%	17827002	98.7%	37194	10.5%
M5	16	H3K27ac	2	17515972	93.8%	17344945	99.0%	35202	9.7%
M5	17	H3K27ac	1	17112231	88.7%	16956272	99.1%	56548	20.6%
M5	17	H3K27ac	2	18744747	87.4%	18549890	99.0%	62238	22.4%
M5	18	H3K27ac	1	16296728	79.8%	16130842	99.0%	56551	19.4%
M5	18	H3K27ac	2	15972213	78.1%	15791020	98.9%	52705	18.1%
M5	19	H3K27ac	1	18418657	83.1%	18261896	99.1%	57745	21.2%
M5	19	H3K27ac	2	19612274	80.6%	19369323	98.8%	46263	17.1%
M5	1	H3K27me3	1	20309683	91.9%	19676379	96.9%	10517	23.0%
M5	1	H3K27me3	2	11749316	92.1%	11375996	96.8%	7451	12.8%
M5	2	H3K27me3	1	21461444	92.1%	20831862	97.1%	11484	16.7%
M5	2	H3K27me3	2	25270894	92.7%	24580806	97.3%	12735	20.7%

M5	3	H3K27me3	1	26353246	93.1%	25640953	97.3%	13810	24.8%
M5	3	H3K27me3	2	24544272	92.8%	23900082	97.4%	13822	24.1%
M5	4	H3K27me3	1	23021334	92.4%	22403597	97.3%	14160	23.7%
M5	4	H3K27me3	2	19037064	92.7%	18531468	97.3%	11619	25.9%
M5	5	H3K27me3	1	17867358	91.8%	17346566	97.1%	11096	24.0%
M5	5	H3K27me3	2	20774013	91.7%	20159082	97.0%	11035	25.6%
M5	6	H3K27me3	1	18257408	92.6%	17738085	97.2%	9583	19.3%
M5	6	H3K27me3	2	19338831	92.4%	18801333	97.2%	9842	20.5%
M5	7	H3K27me3	1	23486390	92.3%	22763174	96.9%	11348	17.3%
M5	7	H3K27me3	2	17118012	91.4%	16587002	96.9%	8122	15.8%
M5	8	H3K27me3	1	22689859	92.8%	22041060	97.1%	11862	17.6%
M5	8	H3K27me3	2	21844579	91.6%	21130571	96.7%	10348	14.6%
M5	9	H3K27me3	1	18429915	92.0%	17831713	96.8%	8632	16.3%
M5	9	H3K27me3	2	26553501	91.9%	25670272	96.7%	12302	18.8%
M5	10	H3K27me3	1	26148779	92.3%	25311518	96.8%	11787	17.5%
M5	10	H3K27me3	2	20255329	92.7%	19665926	97.1%	9398	18.8%
M5	11	H3K27me3	1	17336062	93.0%	16865008	97.3%	8155	15.0%
M5	11	H3K27me3	2	15041369	85.7%	14624427	97.2%	6949	11.8%
M5	12	H3K27me3	1	16611418	93.6%	16185429	97.4%	9553	17.3%
M5	12	H3K27me3	2	21326216	93.0%	20723441	97.2%	12039	16.4%
M5	13	H3K27me3	1	23248815	90.4%	22536427	96.9%	10714	15.2%
M5	13	H3K27me3	2	24938511	90.9%	24082228	96.6%	10442	15.2%
M5	14	H3K27me3	1	17345110	61.0%	16773835	96.7%	5768	10.2%
M5	14	H3K27me3	2	24899332	90.3%	24149452	97.0%	10022	14.2%
M5	15	H3K27me3	1	20599933	91.2%	19906107	96.6%	8236	16.5%
M5	15	H3K27me3	2	26246977	91.3%	25362618	96.6%	11143	16.7%
M5	16	H3K27me3	1	25669176	89.9%	24780120	96.5%	11020	16.7%
M5	16	H3K27me3	2	26933154	91.7%	26069016	96.8%	12031	19.2%
M5	17	H3K27me3	1	23678042	90.9%	22948012	96.9%	11856	17.4%
M5	17	H3K27me3	2	22979648	90.5%	22274263	96.9%	11433	16.4%
M5	18	H3K27me3	1	23431240	90.8%	22733916	97.0%	8995	12.9%
M5	18	H3K27me3	2	21416764	90.6%	20783198	97.0%	7854	16.7%
M5	19	H3K27me3	1	23128612	90.6%	22429890	97.0%	9178	13.6%
M5	19	H3K27me3	2	21753842	91.3%	21112805	97.1%	8615	12.6%

We computed Pearson's correlation between replicates of the same slice and slices spaced apart by different numbers of slices (Figure 13B). Overall, correlation coefficient values dropped as slices spaced further apart. Neurons also showed a higher level of variation compared to glia. In fact, M1-2 showed that glia remained largely uniform compared to neurons (Figure 14). It was also observed that left and right neocortices had little difference between them. Due to this, we focused our effort on neurons from one side of the neocortex in the subsequent experiments on M3-5. Spatial variation at genes related brain functions were

also observed by examining the ChIP-seq signal in genome browser (Figure 11C-D).

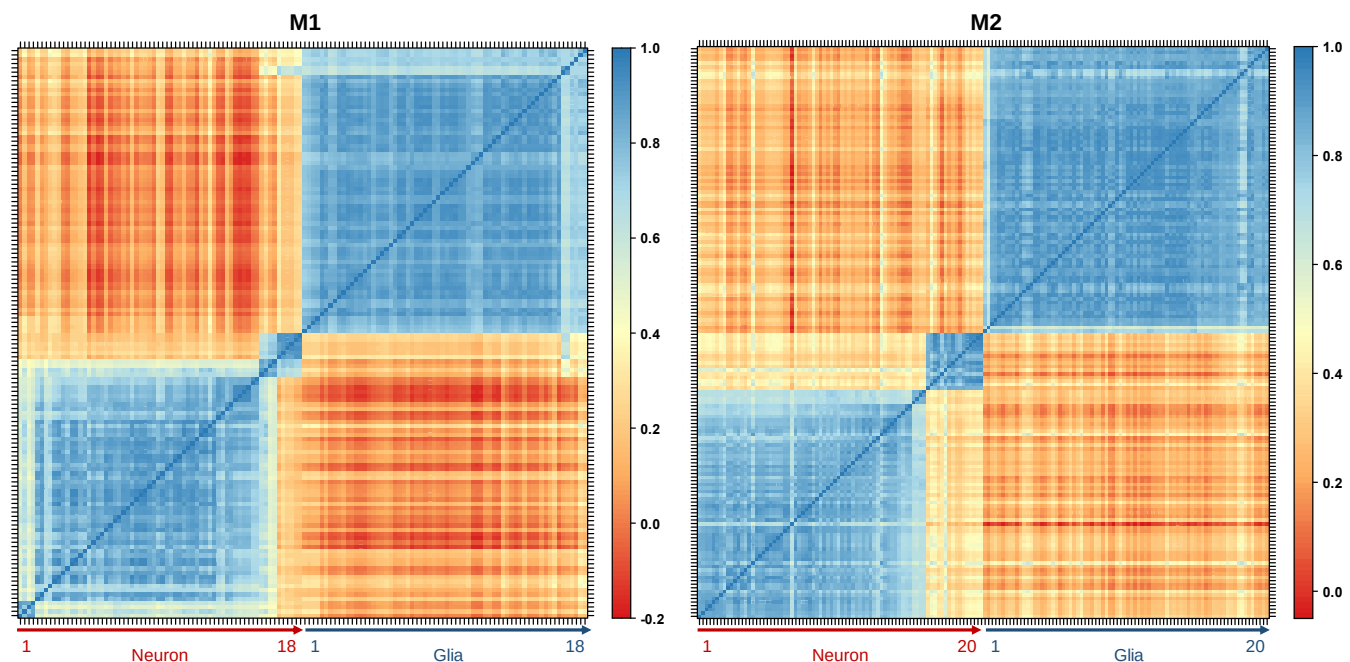


Figure 14: Pearson's correlation plots of neuron and glia samples in M1 and M2. Distinct portions can be observed in neuron samples.

Spatial profiles of histone modifications in mouse brain neurons

We plotted the overall correlation matrix of M4 and M5 for both histone marks using DiffBind (Figure 15). Similar spatial profiles of correlation were observed for both histone marks, and the same general trend was seen in both normal and seizure mice. Starting from slice 1 which corresponds to the front of the brain, three distinct portions with relatively high correlation can be observed (M4: 1-4, 5-20, 21; M5: 1-6, 7-17, 18-19). We divided the brain into the anterior, mid, and posterior portions based on this observation. We observed higher correlation between seizure and control mice in the mid portion, and lower correlation between the different conditions in the anterior and posterior portions of the brain. A similar trend in tissue morphology can be observed in the mouse brain neocortex according to Allen Brain Atlas.¹⁵¹ The dominant functional brain area changes accordingly, starting with primary and secondary

motor areas, and gradually transitions to somatosensory areas, auditory and visual areas.

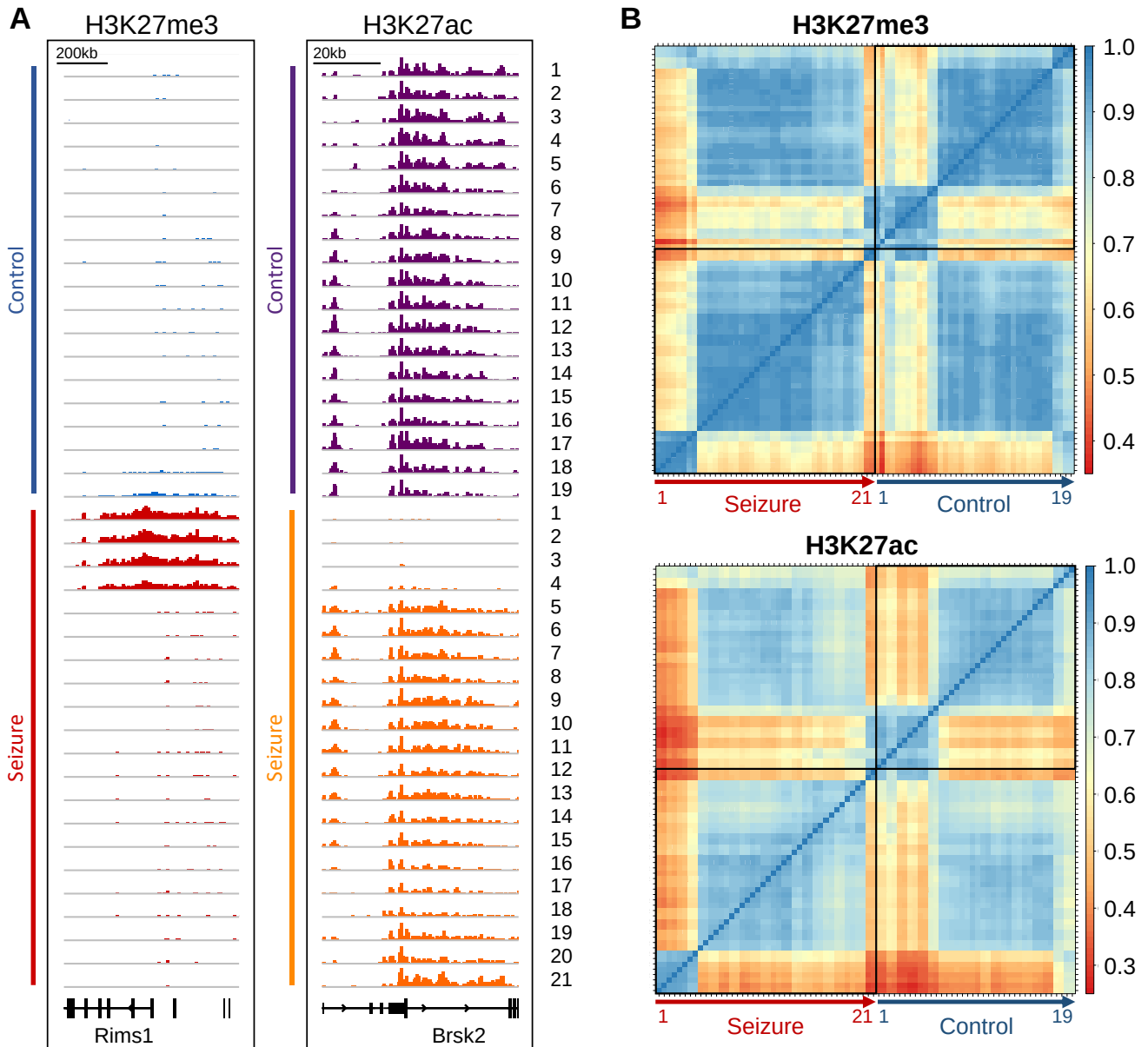


Figure 15: Spatial resolved profiling reveal changes induced by seizure in mouse neocortex. (A) Normalized ChIP-seq signal at selected genes showing significant difference between control and seizure mice. (B) Pearson's correlation of control and seizure mice for both H3K27ac and H3K27me3. 3 distinct spatial portions (frontal, medial, caudal) can be seen and correlation between control and seizure is low in frontal and caudal portions.

We used DiffBind to analyze the difference in histone modification levels. First, we treated the samples within the three portions of the brain as replicates of the same tissue. We then compared the groups between normal and seizure conditions. Differentially marked regions

marked by DiffBind were further filtered (fold > 2, $p < 1E-5$) and separated into two groups (seizure > normal, normal > seizure). GO analyses were conducted on these groups to determine their related biological processes and phenotypes.

Significant differences were identified between the anterior and posterior portions of the brain between the two conditions. For H3K27ac, DiffBind analysis identified 18061 peaks with higher binding in control and 15867 peaks with higher binding in seizure within the frontal section of the brain; 7911 peaks with higher binding in control and 11075 peaks with higher binding in seizure were identified in the back section. For H3K27me3, 1496 peaks with higher binding in control and 1428 peaks with higher binding in seizure were identified in the frontal section; 313 peaks with higher binding in control and 2364 peaks with higher binding in seizure were identified in the back section (fold change > 2, $P < 1E-5$). These peak sets were fed to GREAT gene ontology analysis to check for their related biological processes (Figure 16). Most terms identified were related to signal transmission and general brain functions, including exocytosis, neurotransmitter secretion, ion transporter activity, synaptic plasticity, neuron differentiation, cognition, learning, and memory.

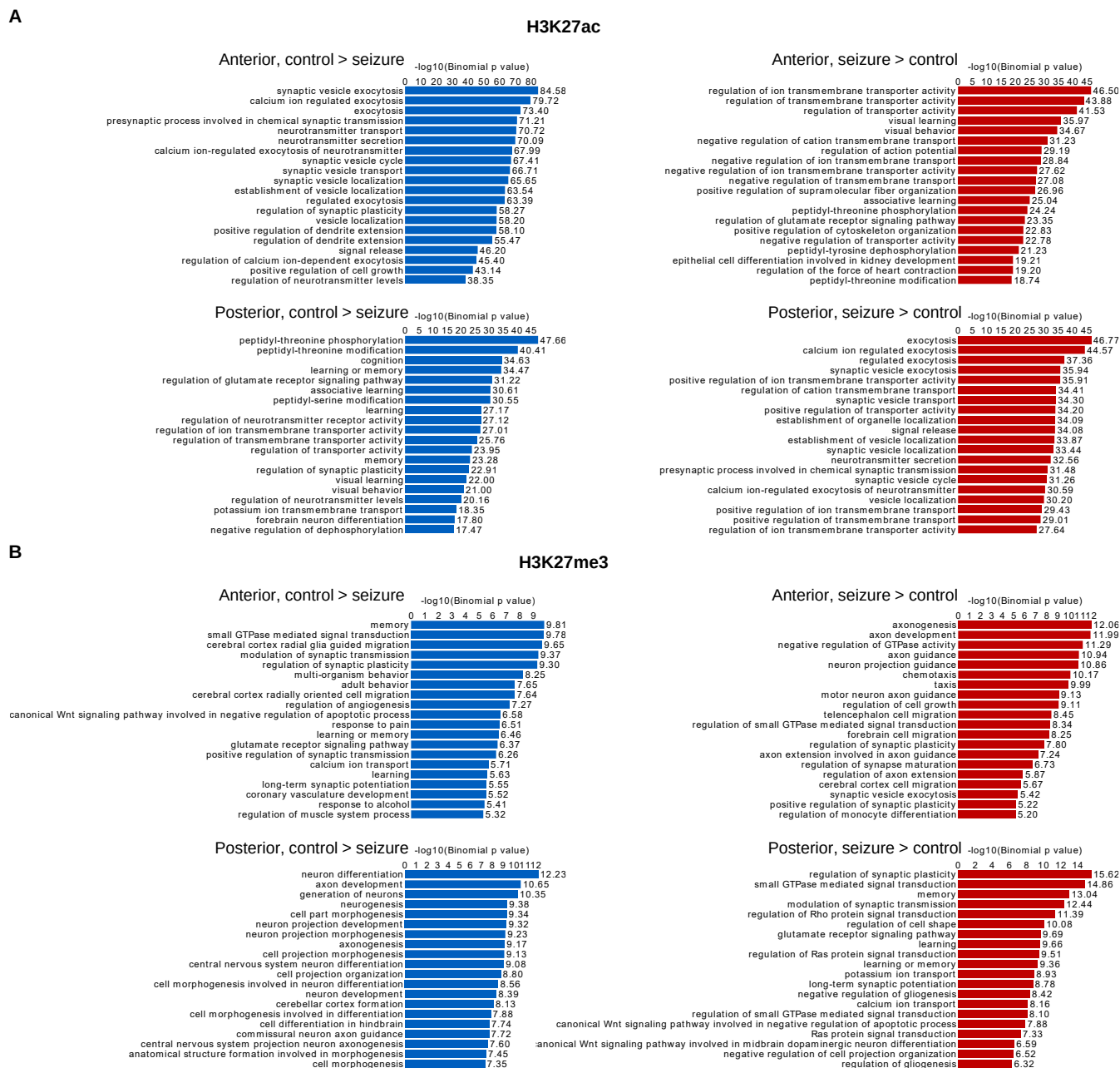


Figure 16: GO biological process terms from differentially modified peaks between control and seizure mice in anterior and posterior portions of the neocortex for (A) H3K27ac and (B) H3K27me3.

Consensus peak clustering in untreated mice

We selected the top 30% of peaks ranked by confidence score from each sample and picked from them a set of consensus peaks (peaks that occur in at least 2 samples). For H3K27me3, promoter peaks were retained. The average normalized read coverage for each peak was

computed using Deeptools (*multiBigWigSummary*). The resulting matrix in which the x-axis lists the samples and y-axis lists the consensus peaks was used for k-means clustering and visualized in heatmaps. The optimal number of clusters was determined by Elbow method.



Figure 17: Individual epigenomic tomography generated by k-means clustering of promoter peaks in neuron samples and their associated GO biological terms for each cluster. (A) M1 left; (B) M1 right; (C) M2 left; (D) M2 right. Top 10 GO terms are listed. (E) Overlap of peaks of the same cluster between different tomographies.

Peaks grouped into the same cluster were then used in GREAT gene ontology analysis to check for their enriched biological processes.

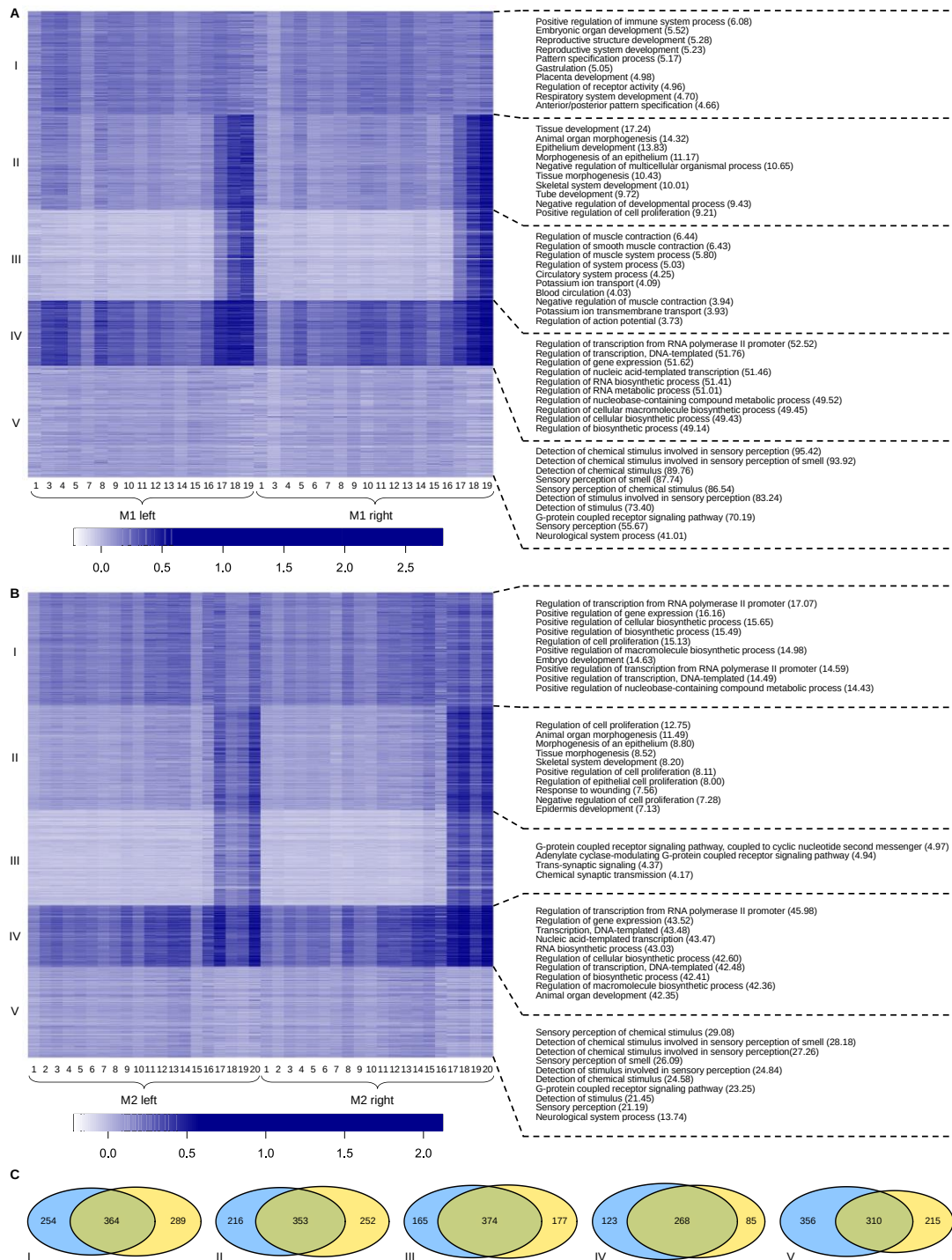


Figure 18: Double epigenomic tomography produced by k-means clustering of promoter H3K27me3 peaks from neurons of both left and right neocortices, and top GO biological process terms of each cluster. (A) M1 left and right. (B) M2 left and right. (C) Overlap between peaks of each cluster.

We produced 4 tomographies using the untreated mice neuron data (M1-2). While each varied in the number of peaks per cluster, the general patterns of spatial variation were highly similar, and the peaks shared much overlap (Figure 17). The clusters also shared many GO biological process terms among them. Cluster I and IV included gene expression regulation, development, and biosynthetic processes. Cluster II was linked to morphogenesis, development, and cell proliferation. Cluster V was associated with sensory perception and chemical stimulus.

We also produced combined neuronal tomographies by joining 2 sets of data together for the peak selection and clustering process. This way we could compare the biological replicates (M1 and M2), and the left and right portions of the neocortex. The results showed that left and right neocortices shared similar cluster patterns (Figure 18). 5 clusters with their distinct spatial distribution were identified in both cases. Cluster I and II were related to development and morphogenesis, cluster IV were related to transcription and RNA synthesis, and cluster V was related to sensory processes and chemical stimulus. The similarities between clustering results were seen again in a different mouse.

In comparison, glia data from untreated mice showed similar patterns in clustering results and much less spatial variation compared to neurons (Figure 19). Figures were clustered together mainly by uniform levels of signal intensity. Peak overlap between clusters from the different

sets of data and the sharing of GO terms remain the same (Figure 19).

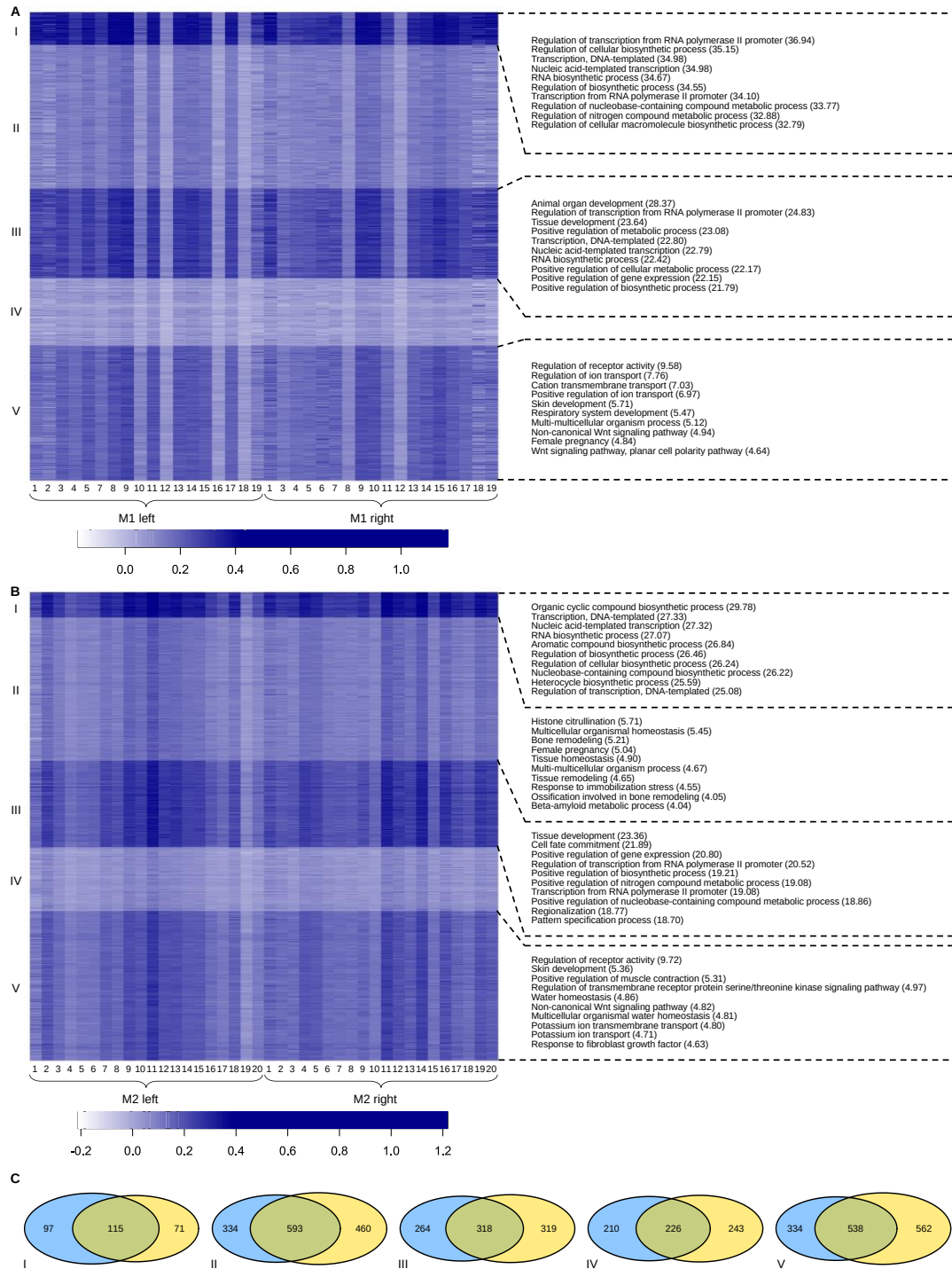


Figure 19: Double epigenomic tomography produced by k-means clustering of promoter H3K27me3 peaks from glia of both left and right neocortices, and top GO biological process terms of each cluster. (A) M1 left and right. (B) M2 left and right. (C) Overlap between peaks of each cluster.

Link between epigenomic tomography and RNA tomography

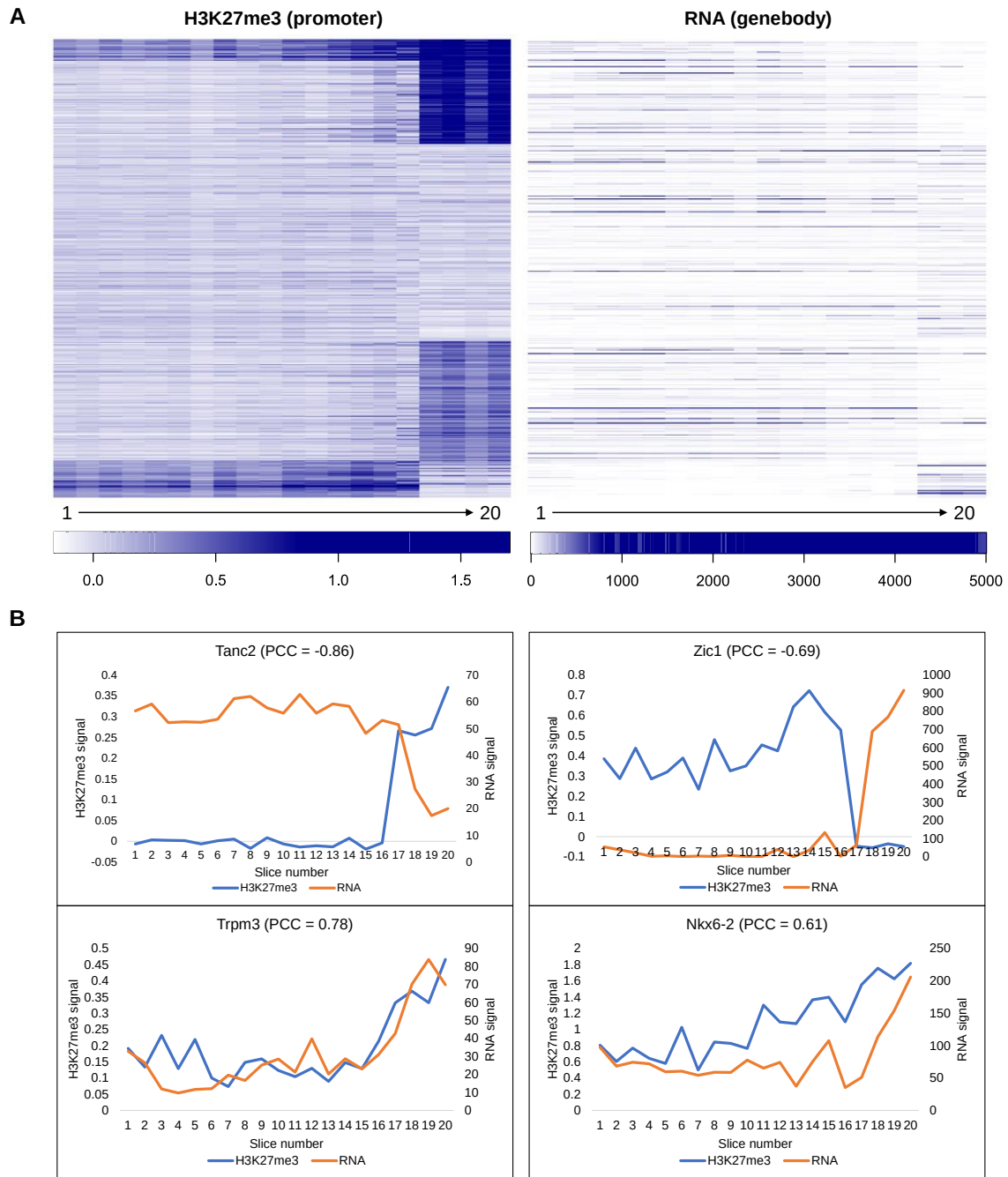


Figure 20: Correlation between promoter H3K27me3 (M2) and genebody RNA signal (M3). (A) *k*-means clustering of promoter H3K27me3 signal from 495 genes with inverse trend between H3K27me3 and RNA ($PCC < -0.5$). Genebody RNA signal for genes are ranked in the same order on the right. (B) Selected genes with strong inverse or positive correlation between promoter H3K27me3 and RNA signal.

We next examined the relationship between epigenomic tomography with gene expression.

Using Smart-seq2, we profiled the mRNA levels across the neurons of right neocortex slices of

mouse M3. The RNA-seq signal over genebody of all mouse RefSeq genes were computed. Pearson's correlations between genebody RNA-seq signal and promoter H3K27me3 signal across all slices were computed to test the correlation of two data types for the genes. A total of 4,443 genes passed the quality filter and 2,839 of those (64%) had negative correlation. 495 genes showed high negative correlation ($r < -0.5$). As H3K27me3 is repressive to gene expression, we expect genes involved in brain functions to have a negative correlation between H3K27me3 and RNA levels. For example, *Tanc2* and *Zic1* are associated with signaling during brain development (Figure 20).¹⁵²⁻¹⁵⁴

Consensus peak clustering in seizure and control mice

We processed the seizure (M4) and control (M5) mice data on H3K27ac and H3K27me3 in the same pipeline to cluster the consensus peaks. For H3K27ac, we further filtered the consensus peaks to retain those without overlap with promoters (active enhancers). We identified 4 clusters of enhancer peaks with distinct patterns that differ significantly between control and seizure mice (Figure 21). Regulation of transmembrane ion transporter activity, exocytosis and neurotransmitter secretion ranked top in the GO biological processes related to these peaks. Ion transportation across membrane and exocytosis form the basis of electrical and chemical signaling in nerve cells.¹⁵⁵ Seizure is an example of severely perturbed neural activity characterized by significant changes to extracellular and intracellular ion concentrations and increased neuronal firing.^{156,157} It is within our expectations that inducing seizure in mice caused histone modification changes to genomic regions related to these major biological processes. In cluster IV, locomotory behavior was identified as a top GO biological process.

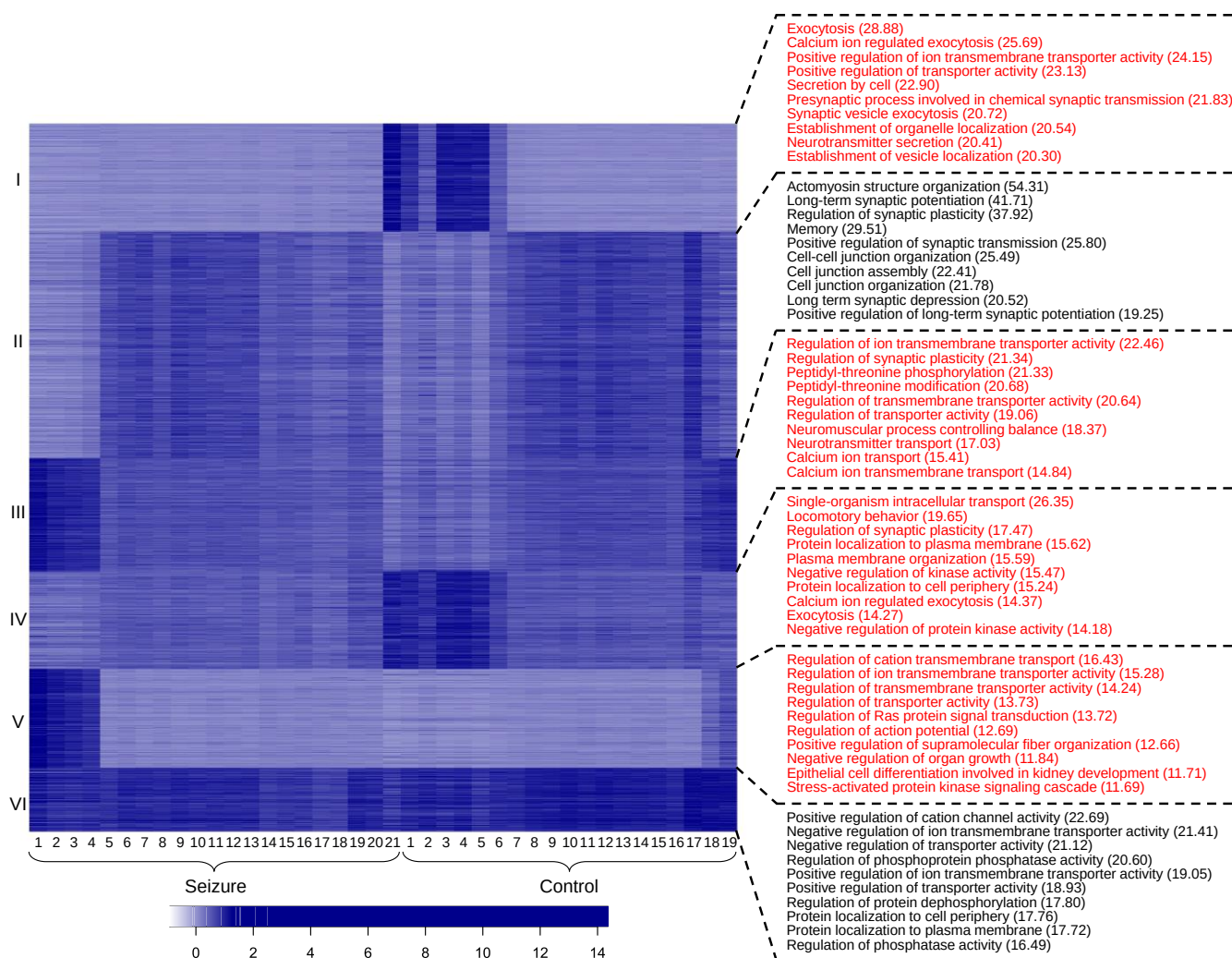


Figure 21: Double epigenomic tomography produced by k-means clustering of enhancer peaks from neuronal H3K27ac data on right neocortex of control and seizure mice and top GO biological process terms of each cluster.

The impact of seizures on rodent locomotory behavior through disrupting suprachiasmatic nucleus function has been well documented.¹⁵⁸⁻¹⁶⁰ We examined the peaks related to this GO term and found the three genes with the largest number of related peaks: *Grin2a*, *Gnao1* and *Negr1*. Mutation in GRIN2A causes epilepsy, language difficulties and delay of brain development in human.¹⁶¹⁻¹⁶³ Mutation in GNAO1 leads to epilepsy, movement disorder and development delay in human.^{164,165} Susceptibility to pentylenetetrazol-induced seizures was observed in *Negr1*-KO mice, together with learning deficits and impaired social behavior.¹⁶⁶

We examined the correlation between control H3K27ac ChIP-seq signal and RNA-seq signal from M3 for enhancer peaks to predict enhancer for these genes. We found that signal trend at predicted enhancers were altered significantly in seizure mouse, suggesting that our analysis captured the perturbed brain function and underlying biological process. Other genes in the top 20 by number of peaks related to locomotory behavior include *Apba2*¹⁶⁷, *Atp2b2*¹⁶⁸, *Lsamp*¹⁶⁹, *Mapt*^{170,171}, *Cacna1c*¹⁷², *Dab1*¹⁷³, *Fgf14*¹⁷⁴, *Cacnb4*¹⁷⁵, *Cntn2*¹⁷⁶, and *Lrrtm1*¹⁷⁷, all of which play important roles in brain functions and development. In comparison, the remaining clusters of peaks related to processes such as synaptic plasticity, RNA splicing and RNA processing.

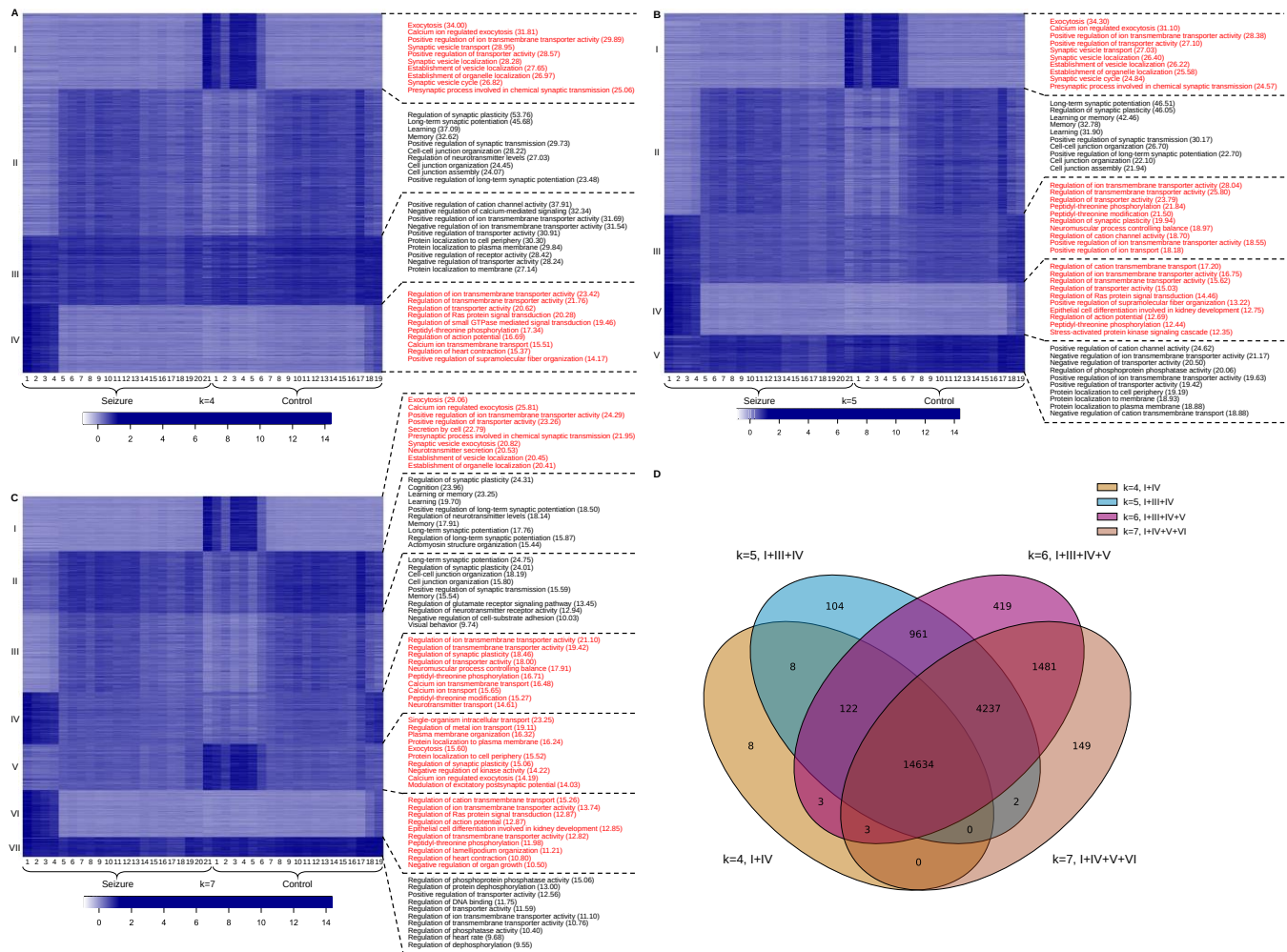


Figure 22: Effect of k value on double tomography of H3K27ac in seizure and control mice. (A) k=4; (B) k=5; (C) k=7. (D) Overlap of spatially differential peaks from clusters of varied k values.

We also tested the effect of cluster number (k) on the results of k-means clustering. Elbow method results in a range of optimal k values between 4 to 7. k-means clustering with cluster number set to these values yielded similar results in terms of distribution of clusters with spatial differences (Figure 22A-C). The total number of spatially differential peaks increased from 14,778 to 20,068 and 21,860 when k increased from 4 to 5 and 6. It then decreased to 20,506 when k was set to 7. Strong overlap was observed among all the differential peaks from different k values, with 14,634 peaks being common in all 4 scenarios (Figure 22D).

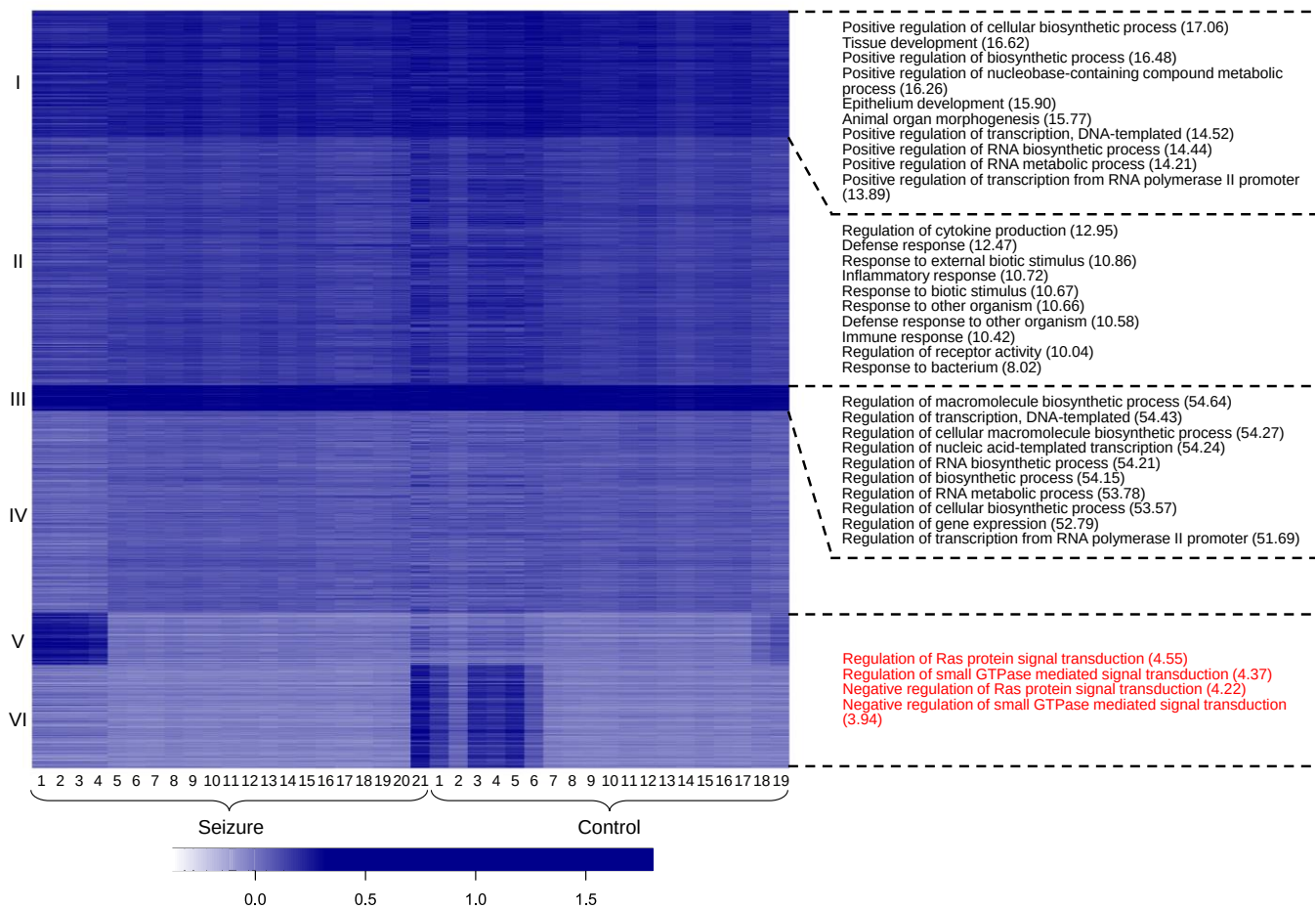


Figure 23: Double epigenomic tomography produced by k-means clustering of promoter H3K27me3 peaks from neurons in control and seizure mice, and their associated GO biological process terms.

In the case of H3K27me3 promoter peak clustering, we discovered 2 clusters with different spatial patterns between control and seizure mice (Figure 23). The difference was concentrated in the anterior part of the brain, similar to that in H3K27ac. GREAT analysis did not identify related biological processes from the clusters individually. When combined, the clusters yielded terms on GTPase mediated signal transduction and Ras protein signal transduction, agreeing with the terms found in H3K27ac clustering. Top terms from H2K27ac clustering such as transmembrane ion transport and exocytosis were not found in the H3K27me3 promoter peak clustering. Knowing the distinctly different mechanisms of the two histone modifications, it is reasonable to believe that they are linked to different biological processes.

Discussion

Our epigenomic tomographies showed that neurons possessed more spatial distribution differences than glia in mouse neocortex. The neocortex can be divided into three distinct portions based on histone modification status of the neurons. Spatial patterns within the blocks were mostly uniform and changed abruptly at the interfaces. Such division was not seen in glia. k-means clustering resulted in groups of peaks that shared spatial patterns and were enriched for similar GO biological process terms. This suggests that epigenomic regions that have similar functions are likely to distribute in the same spatial pattern throughout the neocortex.

Significant changes in the neocortex following kainic acid-induced seizure were observed. Both H3K27me3 and H3K27ac showed differentially modified regions in the anterior and posterior portions of the neocortex. These regions correspond to GO terms, many of which were associated with underlying processes involved in epilepsy. k-means clustering of promoter

H3K27me3 peaks and enhancer H3K27ac peaks yielded similar results, with the changes from seizure being concentrated in the anterior and posterior portions. These spatially different regions were also related to biological processes of seizure and epilepsy, and we discovered several genes with brain and epilepsy related functions.

We believe this method can be used as a general technique for spatial epigenomic analysis of tissue samples. It can be further improved by increasing spatial resolution since MOWChIP-seq has a practical lower limit between 100 – 1000 cells per assay. Thinner tissue slices can be produced via a cryostat and more cell types can be separated with more markers. Similar in concept to “RNA tomography”,⁶ our method also makes use of low-input technology to map spatial pattern in histone modification, revealing insights into the biological processes and their relationship with the spatial regions of the neocortex. Compared to spatial epigenomic profiling using single-cell pipelines, our work covered a larger area at a lower cost, although the spatial resolution was lower due to the limitation in tissue slicing. The limited resolution makes information on the microenvironment and cell-cell interaction unavailable. Still, the method is useful in filling the knowledge gap between bulk epigenomic assays and a potential organ atlas at single cell resolution, while remaining practical and cost-effective.

Materials and methods

Mouse strain, seizure induction and brain dissection

All procedures were conducted according to NIH guidelines and approved by Institutional Animal Care and Use Committee (IACUC) at Virginia Tech. 8 weeks old male C57BL/6J mice were obtained from Jackson Laboratory and kept in the animal facility with 12-h light/12-h dark cycles and food and water ad libitum. Mice were allowed 2 weeks to acclimate before

experiments. M1-M3 were euthanized using compressed CO₂ and death was confirmed by cervical dislocation. Brains were rapidly dissected and placed on ice. Mouse brain slicer matrix (BSMAS005-1, Zivic Instruments) was chilled on ice and the brain was placed inside. Double edge blades were inserted into the slots until the entire length of the brain was covered. Blades were pushed downward at the same time to complete the cutting process. Blades were then carefully extracted from the matrix. Individual slices (0.5 mm thickness) were visually examined and neocortex portions were separated into left and right hemisphere slices, transferred into 1.5 ml tubes and flash frozen on dry ice. Tissue slices were transported on dry ice and stored at -80 °C.

M4 and M5 were used for the study of seizure. Kainic acid (K0133, LKT Laboratories) was prepared fresh to 2 mg/ml in sterile saline. Mice were injected intraperitoneally with repeated 5 mg/kg doses at 30 min intervals until behavioral seizures were observed (2 doses used).¹⁷⁸ Control group were injected with same amounts of sterile saline at 30 min intervals. Mice were euthanized directly 3 d after injection by cervical dislocation and death was confirmed by bilateral thoracotomy. Brains were dissected and neocortex slices were isolated in the same manner as M1-M3.

Nuclei extraction from mouse brain slices

The following steps were performed on ice and centrifugation performed at 4 °C. A single brain slice was placed in a pre-chilled tissue grinder set (D9063, Sigma-Aldrich). 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] was added. Brain tissue was homogenized by slowly douncing 15 times with pestle A and 25 times with pestle B. Care was

taken to avoid generating foam. Homogenate was filtered through a 40 μ m strainer into a 15 ml tube and centrifuged at 1,000g for 10 min. Supernatant was removed and the crude nuclei pellet was resuspended in 500 μ l of nuclei extraction buffer and transferred into a 1.5 ml tube. 750 μ l of 50% iodixanol, 7.5 μ l of PIC, 0.75 μ l of 100 mM PMSF and 0.75 μ l of 1M dithiothreitol were added and mixed by gentle pipetting. The mixture was centrifuged at 10,000g for 20 min and supernatant was removed. 300 μ l of 2% normal goat serum (50062Z, Life Technologies) diluted using Dulbecco's phosphate-buffered saline (DPBS) was added and incubated for 10 min before resuspending. Anti-NeuN antibody conjugated with Alexa 488 (MAB377X, EMD Millipore) was diluted with DPBS to 2 ng/ μ l. 10 μ l of anti-NeuN antibody was added to the resuspended nuclei and incubated at 4 °C for 1 h on a rotator. The labeled nuclei were sorted using FACS (BD FACSAria, BD Biosciences) and NeuN+ neurons were collected into a 15 ml tube and placed on ice.

MNase digestion of chromatin

200 μ l of sorted neuron suspension was used to prepare chromatin (containing ~30,000 neurons). 2 μ l of PIC, 2 μ l of 100 mM PMSF and 200 μ l of 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. 4 μ l of 1 M CaCl₂ and 5 μ l of 100 U MNase (88216, Thermo Fisher Scientific) were added, mixed by vortexing and incubated at room temperature for 10 min. 44 μ l of 0.5 M EDTA was then added, mixed by vortexing and incubated on ice for 10 min. The mixture 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. Excess chromatin solution was stored at 4 °C. 70 μ l of chromatin was used for each MOWChIP-seq assay (equal to 4,600 neurons). In case of tissue slices generating less than 200 μ l of sorted neuron

suspension, the entire volume was used for digestion and volumes of reagents used were scaled down accordingly to keep concentrations the same.

Preparation of immunoprecipitation beads

5 µl of Dynabeads Protein A (10002D, Thermo Fisher Scientific) were used for each MOWChIP assay. 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 µl of IP buffer. The following antibodies were used in this work: anti-H3K27ac (ab4729, Abcam), 0.125 µg/assay; anti-H3K27me3 (39155, Active Motif), 0.5 µg/assay. The suspension was incubated on a rotator at 4 °C overnight. Beads were then washed with IP buffer twice and resuspended in 5 µl of IP buffer for loading into the MOWChIP device.

MOWChIP operation

Same as in Methods of Chapter 2.

Purification of ChIP DNA

Same as in Methods of Chapter 2.

Library preparation, quantification and sequencing

Libraries were constructed using 2S Plus DNA Library kit (10009877, IDT) and 2S CDI Adapter S1 indexing kit (10009895, IDT) following the manufacturer's instructions. 5% (v/v) EvaGreen (31000, Biotium) was added to the amplification mixture to monitor the progress of library amplification. Amplification was stopped once an increase of 2,000 RFU was reached. Library was eluted to 10 µl of low EDTA TE buffer. Library concentration was quantified with KAPA

Library Quantification Kit (07960140001, Roche). Libraries were pooled for sequencing by Illumina NovaSeq 6000 in S4 200 mode (pair-end, 100 bp).

ChIP-seq Data analysis

Raw FASTQ files were trimmed and QC filtered using Trim Galore! (0.4.1) with default settings. Filtered reads were mapped to reference genome mm10 with bowtie (1.1.2). Uniquely mapped reads were filtered for known blacklisted regions using samtools (1.3.1)⁸⁵ and bedtools (2.29.2)⁸⁶ to remove ChIP-seq artifacts. The reference genome was divided into 100-bp bins and signal for each bin was counted for ChIP and input samples. Normalized ChIP-seq signal per bin was calculated as follows:

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

Both ChIP and input reads were then extended by 100 bp on either ends to compute the normalized signal for each bin and visualized as genome browser tracks in IGV (2.11.2).

Peaks in H3K27ac samples were called by MACS2 (2.1.1.20160309) with default settings.

Peaks in H3K27me3 samples were called by epic2 (0.0.41) with the following settings: -fs 250 --gaps-allowed 4 -fdr 1e-20. Pearson's correlation between datasets was calculated using DiffBind (3.14)⁸¹ *dba.count* command to process peaksets and uniquely mapped reads.

Construction of epigenomic tomography

Peak files from all datasets of interest were sorted by confidence score generated by the peak caller. The top 30% peaks of each dataset were then extracted and loaded into DiffBind (3.14). Peaks that occurred in at least 2 datasets were combined into a set of consensus peaks of the group. Promoters were defined as ± 2 kb from transcription starting sites. Consensus peaks

were tested for their overlap with promoter regions using bedtools *intersect* function.

H3K27me3 peaks with overlap were extracted as promoter peaks for clustering. H3K27ac peaks without overlap were extracted as enhancer peaks for clustering.

Score-in-peak matrix for a group of samples were generated by computing the average normalized ChIP-seq signal in each peak region using deepTools (3.5.0) *multiBigwigSummary* function. The matrix was then clustered using *kmeans* function's default Hartigan method¹⁷⁹ in R with the following setting: *nstart* = 10, *iter.max* = 100. Elbow method was used to determine the optimal number of clusters with the *tot.withinss* function to compute the within-cluster-sum of squared errors for k values from 1 to 15. Individual clusters were analyzed using GREAT⁸⁸ with single nearest gene setting. Overlap of peaks and clusters was visualized using Intervene *venn* function.¹⁸⁰

Chapter 5: ChIRP-seq profiling of lncRNA binding in cell lines

Introduction

Less than 2% of the mammalian genome is protein-coding. The non-coding regions have important roles in cell biology, serving as elements of regulation including promoter, enhancer and silencer. Some non-coding regions also transcribe into non-coding RNA (ncRNA), mainly ribosomal RNA, transfer RNA, micro RNA and long non-coding RNA (lncRNA). Early works in the field have shown that ncRNA participate in epigenetic regulation, differentiation and development.¹⁸¹ As the research progress, the focus shifted onto lncRNA, which are more than 200 nt in length. lncRNA are a class of varied molecules that make up a distinct type of epigenetic regulation.¹⁸² Functionally, they are divided into guides, signals, decoys and scaffolds.¹⁸³ lncRNA can influence gene expression in a number of ways, resulting in either enhancement or inhibition.^{182,184} Production of lncRNA itself is also regulated by a variety of epigenetic components including transcription factors and chromatin state,¹⁸⁵ making lncRNA an important player in the overall epigenetic regulation network.

Hon *et al.* profiled 1,829 FANTOM5 cap analysis of gene expression datasets to generate a thorough atlas of human lncRNA genes, totaling 27,919 lncRNA genes.¹⁸⁶ This number exceeds the number of protein-coding genes in the human genome, reflecting the large number of distinct lncRNA molecules. lncRNA can affect both local gene expression and regulate distant genes through several mechanisms: sequence-specific activation, spliceosome, promoter and enhancer interaction, and regulation of chromatin state and Pol II activity.^{187,188} In terms of biological effect, lncRNA play roles in influencing telomere stability, transcription, X-chromosome inactivation, epigenetic modification and gene

expression.^{184,189,190} These functions lead to lncRNA being implicated in development and diseases,^{191,192} in particular cancer.¹⁹³⁻¹⁹⁵ Despite the considerable progress made, there is still much unknown in the mechanisms of lncRNA, in large part due to the limitations of sequencing-based profiling methods.

It is important to examine the interaction between lncRNA and the genome to understand their roles in cell regulation. Genome-wide mapping of specific lncRNA's binding is a key component of such investigation as it provides information to the effect of lncRNAs on specific genes and how they relate to other epigenetic processes like histone modification and DNA methylation. Chromatin isolation by RNA purification (ChIRP-seq) was developed to profile the binding of specific lncRNAs, which were first fixed to genomic DNA and then captured using sets of complementary oligonucleotide probes.¹⁹⁶ Since the first iteration of ChIRP-seq by Howard Chang's group in 2011, the method has largely remained the same. ChIRP has seen use in research of lncRNA in the years that followed.¹⁹⁷⁻¹⁹⁹ More recently, the method has been used regularly, sometimes in combination with other omics assays.²⁰⁰⁻²⁰³ While enabling researchers to directly map lncRNA binding over the target genome, ChIRP-seq has several major limitations. It requires a huge amount of input cells (~ 20 million cells per assay) to work due to the low abundance of lncRNAs bound to genome, and the low efficiency of lncRNA capture.¹⁹⁶⁻¹⁹⁸ This makes it impossible to study tissue from lab animals and human, and so far there has only been published ChIRP-seq datasets on cell lines. Many available datasets were also produced from cell lines transfected to overexpress the target lncRNA.^{196,198} While boosting lncRNA binding signals, this could also introduce artifacts and alter the native biological state of the cell. Mature low-input solutions (< 1 million cells per assay) have been available, some of which commercially, for other epigenetic and transcriptomic assays. By

comparison, ChIRP-seq is still stuck with the bulky and complex format with a massive input material barrier (> 10 million cells per assay) preventing its direct use on animal and patient samples. We have not seen progress reported on development of ChIRP-seq methods that lower this requirement. Seeing that there is plenty of room for exploration in the functions of lncRNA, especially in native tissues, the need for low-input ChIRP-seq technology is very present.

In this project, we demonstrated the feasibility of conducting ChIRP-seq in a microfluidic device. This increased the capture efficiency of ChIRP and reduced the amount of input material needed. We profiled lncRNA binding in cell lines without the use of lncRNA overexpression. Using two cell lines, MDA-MB-231 and HeLa, we compared our results to standard data published by the inventor of the ChIRP-seq method.

Author contribution

Initial work on this project was conducted by Yuan-Pang Hsieh in collaboration with Dr. Zhen Chen at City of Hope. Dr. Hsieh adapted the MOWChIP device for ChIRP-seq and fit the original workflow to the microfluidic system. Zhengzhi Liu took over the project in late 2019.

Results and discussion

ChIRP-seq of HOTAIR in MDA-MB-231 cells

We started mapping lncRNA HOTAIR binding in MDA-MB-231 cells with the aim of reproducing the data from Howard Chang's original protocol. Our lab has previously developed a microfluidic ChIRP-seq protocol using the device from MOWChIP-seq. ChIRP-seq began with harvesting cultured cells and fixing the binding between lncRNA and genomic DNA with glutaraldehyde. Cells were sonicated to release the chromatin and break it down to appropriate

size (100-500bp). For our microfluidic method we kept the exact same steps up until this point.

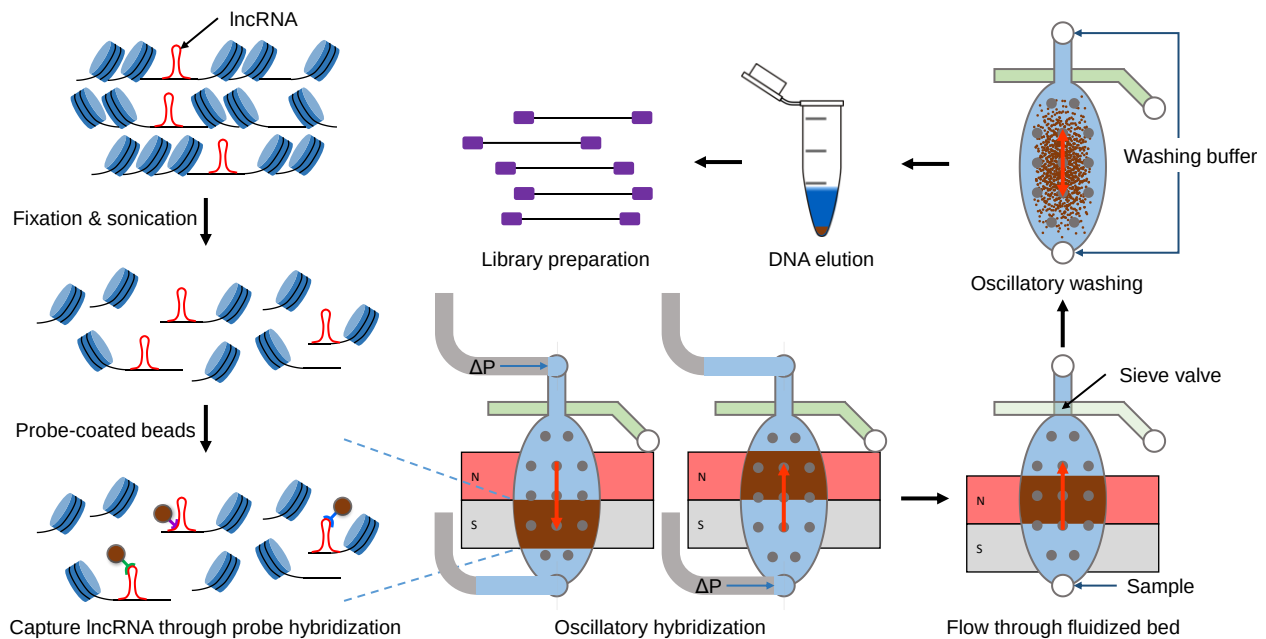


Figure 24: Process illustration of microfluidic ChIRP-seq.

The original ChIRP-seq method published by Howard Chang's group used short oligos designed to hybridize with target lncRNA sequence for specific binding of the target.¹⁹⁶ The oligos were biotinylated so that after being hybridized to the target lncRNA, they could be captured by streptavidin beads which interacts strongly with biotin. Repeated washing cycles were then conducted to remove non-specific binding. Proteins that linked lncRNA and genomic DNA together were digested with proteinase K and DNA fragments with lncRNA binding were eluted and purified for library preparation and sequencing. Multiple short sequence oligo probes were designed to hybridize with different parts of the lncRNA target. Probes were divided into odd and even groups and assays were conducted using each set. The agreeing peaks in a pair were designated 'true' peaks to reduce random noise. We adapted this process to work with our microfluidic device. First, we incubated the streptavidin beads with oligo probes and loaded the beads into the microfluidic device. Sonicated chromatin was then

loaded into the device for hybridization with the oligo probes attached to streptavidin beads. Two separate steps of hybridization were applied: first chromatin was cycled through the device with oscillating pressure pulses while the beads were held within the device with a magnet; then chromatin was passed through packed bed of beads held against the sieve valve. Oscillatory washing steps were used to remove nonspecific binding. DNA fragments were eluted in the same manner as the original in-tube method.

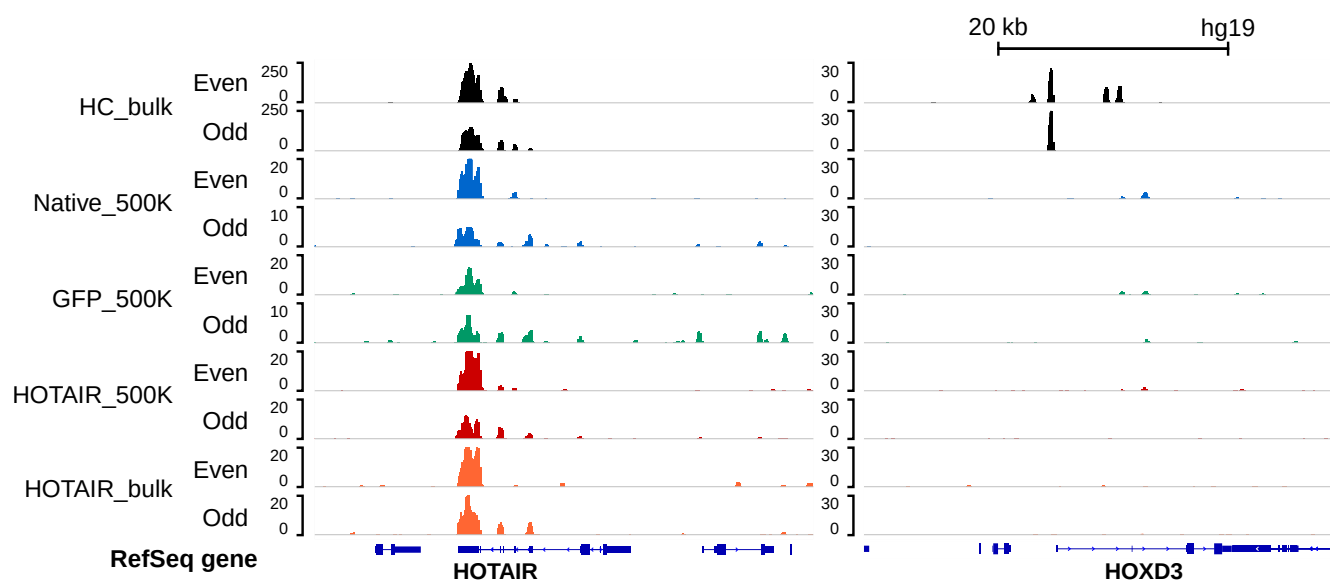


Figure 25: Normalized HOTAIR ChIRP-seq signal of MDA-MB-231 cells. While all samples of microfluidic ChIRP-seq (500K cells per assay) and our bulk in-tube trial using 8 million HOTAIR transfected cells (HOTAIR_bulk) show true peak at HOTAIR gene, our trials do not show true peak at locations where Howard Chang's standard data (HC_bulk) show significant true peak, such as at HOXD3 gene.

We did not copy the probe design of the original method since the software they used was for RNA-FISH probe design and defunct by this point. More mature tools designed for ChIRP-seq probe design were also available. We used the design tool from the probe manufacturer (Biosearch Technologies) to design our 17-probe set for HOTAIR. The original method used 48 probes in total. While the published data were generated using MDA-MB-231 cells that were transfected with lentivirus to stably overexpress HOTAIR, we started with native, non-overexpressed cells. We were able to generate consistent ChIRP-seq data on HOTAIR binding

in MDA-MB-231 cells using 500K cells as input material per assay. However, our data did not agree well with the published data, having few ‘true’ peaks that aligned with the standard outside the encoding region of HOTAIR, where the lncRNA is transcribed on the genome.

Suspecting that over-expression of HOTAIR may alter its genome-wide binding profile, we conducted our own lentivirus transfection to make a culture of MDA-MB-231 that overexpresses HOTAIR. We purchased two strains of replication-deficient lentivirus, one carrying both HOTAIR overexpression and GFP genes, and another that carrying only the GFP gene for control. Both transfections were successfully carried out, verified by fluorescence imaging. We conducted microfluidic ChIRP-seq on both cell strains. The results suggest that neither the transfection itself or HOTAIR overexpression changed HOTAIR binding profile, as both the GFP-only and the HOTAIR culture of MDA-MB-231 showed similar HOTAIR binding profile to native cells, and neither matched the published data well. Results from our in-tube bulk ChIRP-seq method also suggest that the microfluidic workflow was not missing binding peaks, since the binding profiles we generated using 8 million cells in tube were comparable to our microfluidic device data.

ChIRP-seq of TERC in HeLa cells

From the results of MDA-MB-231 cells, we suspected that probe design plays an understated role in the results of both bulk and microfluidic ChIRP-seq. In our bulk HOTAIR ChIRP-seq, the only difference from the original protocol is the probe design, yet our results differed significantly from the data published by Howard Chang’s group. Knowing this, we found another set of data they generated using human HeLa cells and copied their complete probe design for the lncRNA TERC. We first conducted bulk in-tube ChIRP on this target in HeLa

cells using 10 million and 2 million cells per assay, and successfully reproduced the published data with matching tracks and good correlation.

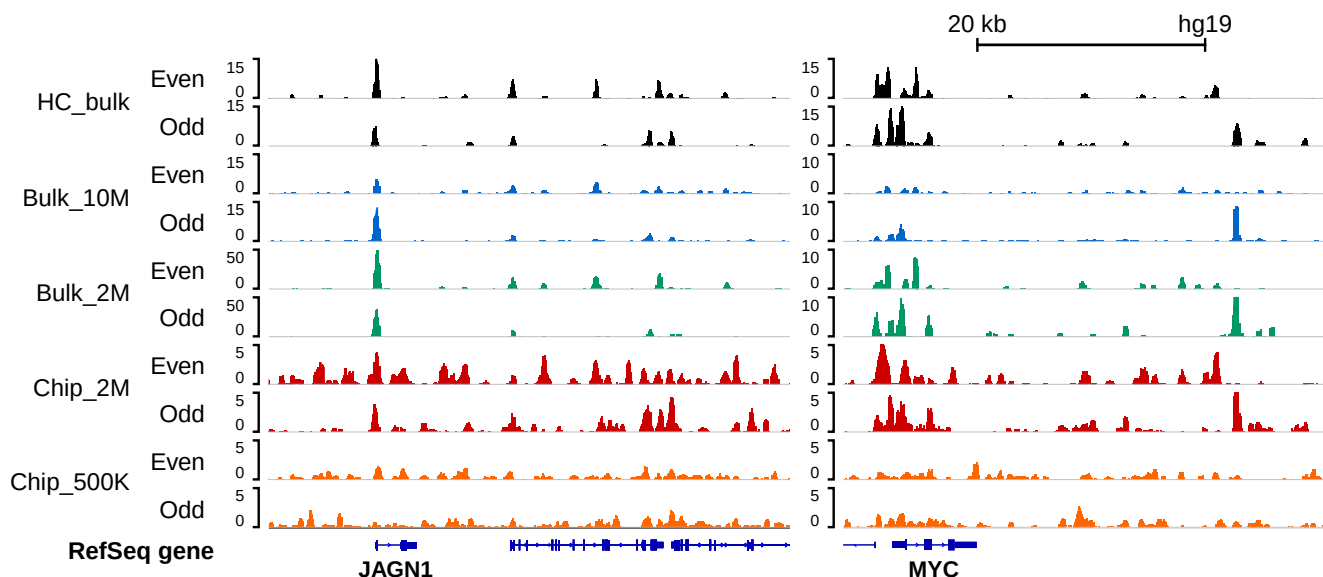


Figure 26: Normalized TERC ChIRP-seq signal of HeLa cells. Our bulk in-tube trials using 10 and 2 million cells (bulk_10M, bulk_2M) show similar true peaks as Howard Chang's standard HeLa cell data (HC_bulk) at genes such as JAGN1 and MYC. Microfluidic ChIRP-seq using 2 million and 500K cells (Chip_2M, Chip_500K) have true peaks at these locations, but the signal-to-noise ratio is lower.

Next, we applied the microfluidic ChIRP-seq process to HeLa cell chromatin. Trials using 2 million cells and 500 K cells were conducted. We observed binding profiles that generally matched the bulk ChIRP-seq data but noticed that the microfluidic datasets had much lower signal-to-noise ratio at 'true' peak locations. This suggests that our microfluidic workflow, while capable of capturing the same lncRNA binding as traditional ChIRP-seq with reduced input material, is not as specific in capturing the target. There is significant room to improve the specificity in lncRNA target capture.

Materials and methods

Cell culture

HeLa cell line and MDA-MB-231 cell line were obtained from American Type Culture Collection. Both cell lines were cultured in DMEM medium (11965084, Gibco) supplemented with 10% fetal bovine serum (1600-044, Gibco) and 1% pen-strep (Invitrogen) at 37 °C, 5% CO₂. Cells were monitored daily and subcultured at 80% confluency. Stocks were frozen in growth medium with 5% added DMSO in liquid nitrogen vapor phase.

Lentivirus transfection of MDA-MB-231

MDA-MB-231 cells were subcultured into 25 cm² flasks at 10⁴ cells/cm² and allowed to attach overnight. Medium was replaced with 5 ml of fresh DMEM medium with 8 µg/ml polybrene and 125 µl of lentivirus stock added. Medium was replaced with 5 ml of DMEM medium after 24 hr of incubation. GFP levels were monitored under a fluorescence microscope to confirm successful transfection. 5 µg/ml of puromycin was added to the medium for selecting cells with successful transfection. Medium with puromycin was removed after 24 h with dead cells that did not have puromycin resistance gene (transfected with GFP and HOTAIR overexpression genes). Transfected cells were subcultured for harvest and stocks were frozen in liquid nitrogen vapor phase storage.

Cell harvest and fixation

Confluent cells were recovered from the culture flask surface by treating with 0.25% trypsin-EDTA (25200-072, Gibco) for 5 min at room temperature. Trypsin was then neutralized by adding an equal volume of growth medium. Cells were flushed off the surface by pipetting and collected into a 50 ml tube. The cell suspension was centrifuged at 800g for 4 min. Medium was removed and cell pellet was resuspended in PBS (1 ml per million cells). The suspension was centrifuged again at 800g for 4 min. PBS was then removed.

Fresh 1% glutaraldehyde was diluted from 25% stock solution with room temperature PBS. 1 ml was used per 1 million cells. Cell pellet was resuspended in 1% glutaraldehyde, starting with a small volume to avoid clumps, and then filled to appropriate volume in a 50 ml tube. Cells were fixed for 10 min at room temperature on a rotator. Reaction was quenched with 1/10th volume of freshly prepared 1.25M glycine at room temperature for 5 min. The suspension was centrifuged at 2,000g for 5 min. Supernatant was removed and the pellet was washed with 20 ml of chilled PBS and centrifuged at 2,000g for 5 min. Supernatant was removed and the pellet was resuspended in 1 ml chilled PBS per 20 million cells. The suspension was transferred into 1.5 ml tubes and centrifuged at 2,000g for 3 min at 4 °C. Remaining PBS was removed and the pellet was frozen at -80 °C.

Chromatin sonication

Frozen pellet of fixed cells was thawed at room temperature and resuspended in 1 ml of lysis buffer [50 mM Tris-HCl, 10mM EDTA, 1% SDS, supplemented with 1% PIC, 1mM of PMSF and 0.5% SupersaseIN before use] per 30 million cells. Suspension was pipetted repeatedly and vortexed to break up cell clumps. Cell suspension was divided into 300 µl aliquots and transferred into 1.5 ml sonication tubes. Cells were sonicated using Bioruptor (Diagenode) with 4 °C water bath at HIGH setting (30 s ON, 45 s OFF pulse interval). Cell lysate was checked every 30 min and sonicated until it turned clear. Lysate was then centrifuged at 16,100g for 5 min at 4 °C. Supernatant containing sheared chromatin was transferred to a new 1.5 ml tube and stored at -80 °C.

Probe preparation

lncRNA target probes were purchased from Biosearch Technologies in plate format with 3' biotin modification and purified by cartridge. Upon receiving the plate of probes, 50 μ l of low EDTA TE buffer was added into each well containing the dried probe pellet. The plate was sealed and vortexed for 1 min to fully dissolve the pellet. The probe stocks were then transferred to individual 1.5 ml tubes and stored at -20 °C. 10 μ M working probe mix was made by combining probes from odd and even sets at equal volume and diluting the final mixture 10 times to 10 μ M with low EDTA TE buffer. The mixed working probe stock was stored at 4 °C and used within a month.

Bulk ChIRP-seq in tube

Frozen chromatin was thawed at room temperature. 320 μ l (10 million cells) of chromatin was used per assay. 640 μ l of hybridization buffer [750 mM NaCl, 1% SDS, 50 mM Tris-HCl, 1mM EDTA, 15% formamide, supplemented with 1% PIC, 1mM of PMSF and 0.5% SupraseIN before use] and 3.2 μ l of 10 μ M probe mixture (even or odd set) were added to the chromatin and mixed by vortexing. The tube was then placed on a rotator and incubated in the dark at 37 °C overnight. 32 μ l of streptavidin beads (washed with lysis buffer) was added to the tube the next day and incubated on rotator for 30 min at 37 °C. The tube was placed on magnetic rack and the supernatant was removed. Beads were then washed with 1 ml of washing buffer [2 $\times$ SCC buffer, 0.5% SDS, supplemented with 1 mM of PMSF before use] at 37 °C for 5 min with shaking. Washing was repeated for a total of 5 rounds.

Fabrication of microfluidic device

Same as in Methods in Chapter 2.

Microfluidic ChIRP-seq

The device was controlled by a customized LabVIEW program (National Instruments) managing solenoid valves (18801003-12V, ASCO Scientific) via a data acquisition card (NI SCB-68, National Instruments). The valves are connected to compressed air for closing the sieve valve and conducting oscillatory washing. The control channel was filled with water first and connected to a valve with 30 psi of pressure. When the valve was opened, the pressure deformed the control layer membrane and partially closed the sieve valve in the device. This prevented the beads from leaking but allowed the solution to pass. The device was then rinsed with lysis buffer at 50 μ l/min using a syringe pump (Fusion 400, Chemyx) to remove air from the chamber. Probe-coated streptavidin beads were loaded into the inlet of the chamber guided by magnet. Sample chromatin solution mixed with 2 times volume of hybridization buffer was then loaded into hybridization tubes and connected to both inlet and outlet of the device. Beads were held in place in the chamber of the device by magnet, and the device is heated at 37 °C. Oscillatory hybridization was conducted for 4 h in the dark at 0.5 psi with 20 s pulses. Sample was then passed through the beads at 2 μ l/min using a syringe pump. Beads were washed with washing buffer at 1 psi for 7.5 min with 0.3 s pulse for 2 rounds. Beads were then flushed out of the device with lysis buffer at 200 μ l/min and collected for DNA isolation.

ChIRP DNA isolation

Each sample of beads was resuspended in 100 μ l DNA elution buffer [50 mMNaHCO₃, 1% SDS, added 2 μ l of RNase A and 3.2 μ l RNase H per 1 ml of buffer before use] and incubated at 37 °C with shaking for 30 min. Supernatant was separated from beads on magnetic rack and transferred to a fresh 1.5 ml tube. Another 100 μ l of DNA elution was added to the beads and incubated at 37 °C for 30 min with shaking. Supernatant was separated from beads and combined with previous elution. 10 μ l of 20mg/ml proteinase K was added to the collected

DNA elution buffer and incubated at 50 °C with shaking. ChIRP DNA was then extracted and purified by phenol-chloroform extraction and ethanol precipitation. DNA pellet was resuspended in 8 µl of low EDTA TE buffer.

Library preparation, quantification and sequencing

Libraries were constructed using 2S Plus DNA Library kit (IDT) following the manufacturer's instructions. 1× EvaGreen dye (Biotium) was added to the amplification mixture to monitor the PCR amplification. Library was eluted to 7 µl 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). Libraries were pooled for sequencing by either Illumina HiSeq 4000 SR50 mode or NovaSeq 6000 PE100 mode.

ChIRP-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 hg19 with Bowtie.⁸⁴

Uniquely mapped reads were filtered for known blacklisted genome regions using samtools⁸⁵ and bedtools⁸⁶ 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 ChIRP-seq signal for each bin was calculated using the following equation:

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

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.⁸⁷ Peak calling was conducted using MACS2 (default settings).

Chapter 6: Summary and outlook

Advances in sequencing technology have reduced its unit cost and accessibility barrier for many new users. Assays like ChIP-seq and ChIRP-seq can now be conducted with relative ease. As the usage of these methods increased, the limitation of input materials became more apparent. In this thesis, we described the use of the microfluidic-based MOWChIP-seq platform to conduct low-input ChIP-seq and ChIRP-seq on a variety of samples and targets. We expanded the scope of MOWChIP-seq to include non-histone targets including Pol II, TFs and histone modification enzymes and showed that they could be profiled without chemical fixation. We also applied MOWChIP-seq workflow to conduct large scale histone modification profiling in patient-derived lung tumor samples and mouse neocortex. Through these projects MOWChIP-seq demonstrated excellent sensitivity and consistency for both histone and non-histone targets. The MOWChIP device itself and part of the operation scheme were used for conducting ChIRP-seq assays on cell line samples, successfully reducing the input material requirement to below 1 million cells.

The future direction of the abovementioned works will be determined by both the available technologies and needs of potential users. Sequencing technology is the major factor here because it dominates the cost and time considerations. Trend in the field of next-generation sequencing is larger throughput and lower cost per nucleotide. This is evident in the wide adoption of Illumina's NovaSeq 6000 by sequencing service providers in recent years. Their newest platform, NovaSeq X, continues this trend and delivers 2.5 times the reads per run compared to the already impressive NovaSeq 6000. After Illumina's decision to cease support for HiSeq platforms, all their current industry machines deliver pair-end 150 bp with more than 2 billion reads per run. However, due to the nature of ChIP-seq and ChIRP-seq, pair-end long

reads yield the same information as single-end short reads, i.e., the location of a particular DNA fragment. Our lab found that the most cost-effective sequencing format for ChIP-seq was HiSeq 4000 in single-end 50 bp mode, delivering the most reads per unit cost. NovaSeq 6000 in S2 pair-end 100 bp mode is the best currently available option, being close to the discontinued HiSeq 4000. Since NovaSeq 6000 delivers more than 2 billion reads per run in such configuration, large scale ChIP-seq and ChIRP-seq profiling involving hundreds of libraries are not only practical but also the rational way forward. The end users also need large scale profiling to reach reliable conclusions, as shown by our work in Chapter 3 and Chapter 4, as well as recent publications studying human samples. Patient samples are highly heterogeneous in nature due to the variability of individuals in both genomic and epigenomic makeup. Therefore, only a sufficiently large sample pool can gather enough data and gain insight beyond the random variance. In the case of animal study, while variance between individual samples may be reduced, the desire to increase resolution within a given organ or tissue would still expand the sample pool. The possibility of separating cell population into distinct fractions via FACS or similar techniques should also be considered. Additionally, the integrated analysis of multiple epigenetic markers has proven valuable. With the inclusion of multiple histone and non-histone marks, the scale of epigenetic mapping is also expanded. For future projects involving MOWChIP-seq, large scale profiling of multiple targets, tissue and cell types, with large biological sample pool will be the norm.

In the case of individual projects, there is also potential for further investigation. In Chapter 2, our work touched on the variation between subtypes of Pol II but did not go beyond the difference between total Pol II and Pol II-S5. We only mapped two TFs and one histone modifying enzyme, out of a large and diverse group of potential targets. It would be interesting

to see more data from animals or human samples on a larger group of non-histone markers. A list of potential targets will be guided by the latest results from the field. In Chapter 3, we profiled five key histone modifications in 38 lung cancer patient samples. A logical extension of this will be sampling a variety of different tumor samples, although our current workflow may need further optimization to complete such a project within a reasonable timeframe. Partial automation of the MOWChIP-seq process and downstream library preparation will greatly reduce the reliance on manual control and improve steady throughput, though this must not compromise the consistency of the workflow. In Chapter 4, we examined the possibility of further increasing the tissue sample resolution. Tissue samples could be sliced thinner with specialized equipment. However, the practical limit of MOWChIP-seq process may be reached before that, especially considering the need for cell/nuclei sorting.

Our lab is continuing the work on microfluidic ChIRP-seq method. We have updated the base design of MOWChIP device to better fit the needs of ChIRP-seq and experimented with other changes to the protocol including probe and beads concentrations, buffer proportions and hybridization conditions. The new phase of work has yielded noticeable improvements, and we plan to carry out trials on animal tissue samples. Based on data gathered so far, we are confident that microfluidic ChIRP-seq can reliably be conducted on 100K cells per assay. It may be worth investigating whether ChIRP process could be done without chemical fixation, similar to that of Pol II and TFs. While the transient binding nature presents a challenge, the potential benefit in simplifying the workflow, improving sensitivity, and reducing time cost is significant.

References

- 1 Cao, Z., Chen, C., He, B., Tan, K. & Lu, C. A microfluidic device for epigenomic profiling using 100 cells. *Nat Methods* **12**, 959-962, doi:10.1038/nmeth.3488 (2015).
- 2 Zhu, B. *et al.* MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nat Protoc* **14**, 3366-3394, doi:10.1038/s41596-019-0223-x (2019).
- 3 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).
- 4 Ma, S. *et al.* Cell-type-specific brain methylomes profiled via ultralow-input microfluidics. *Nature Biomedical Engineering* **2**, 183-194, doi:10.1038/s41551-018-0204-3 (2018).
- 5 de la Fuente Revenga, M. *et al.* Prolonged epigenomic and synaptic plasticity alterations following single exposure to a psychedelic in mice. *Cell Reports* **37**, doi:10.1016/j.celrep.2021.109836 (2021).
- 6 Junker, J. P. *et al.* Genome-wide RNA Tomography in the zebrafish embryo. *Cell* **159**, 662-675, doi:10.1016/j.cell.2014.09.038 (2014).
- 7 Rando, O. J. & Chang, H. Y. Genome-wide views of chromatin structure. *Annu Rev Biochem* **78**, 245-271, doi:10.1146/annurev.biochem.78.071107.134639 (2009).
- 8 Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* **10**, 295-304, doi:10.1038/nrg2540 (2009).
- 9 Stormo, G. D. & Zhao, Y. Determining the specificity of protein-DNA interactions. *Nat Rev Genet* **11**, 751-760, doi:10.1038/nrg2845 (2010).

- 10 Young, R. A. RNA POLYMERASE II. *Annual Review of Biochemistry* **60**, 689-715, doi:10.1146/annurev.bi.60.070191.003353 (1991).
- 11 Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-731, doi:10.1038/nrg3293 (2012).
- 12 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, doi:10.1016/j.tibs.2015.07.003 (2015).
- 13 Lagha, M. *et al.* Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* **153**, 976-987, doi:10.1016/j.cell.2013.04.045 (2013).
- 14 Williams, L. H. *et al.* Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol Cell* **58**, 311-322, doi:10.1016/j.molcel.2015.02.003 (2015).
- 15 Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* **49**, 1045-1051, doi:10.1038/ng.3867 (2017).
- 16 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665, doi:10.1016/j.cell.2018.01.029 (2018).
- 17 Pan, Y., Tsai, C.-J., Ma, B. & Nussinov, R. Mechanisms of transcription factor selectivity. *Trends in Genetics* **26**, 75-83, doi:<https://doi.org/10.1016/j.tig.2009.12.003> (2010).
- 18 Farnham, P. J. Insights from genomic profiling of transcription factors. *Nat Rev Genet* **10**, 605-616, doi:10.1038/nrg2636 (2009).

- 19 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, doi:10.1038/nrg2538 (2009).
- 20 Yusuf, D. *et al.* The Transcription Factor Encyclopedia. *Genome Biology* **13**, R24, doi:10.1186/gb-2012-13-3-r24 (2012).
- 21 Chen, A. & Koehler, A. N. Transcription Factor Inhibition: Lessons Learned and Emerging Targets. *Trends in Molecular Medicine* **26**, 508-518, doi:<https://doi.org/10.1016/j.molmed.2020.01.004> (2020).
- 22 Bushweller, J. H. Targeting transcription factors in cancer - from undruggable to reality. *Nat Rev Cancer* **19**, 611-624, doi:10.1038/s41568-019-0196-7 (2019).
- 23 Wilkinson, A. C., Nakauchi, H. & Göttgens, B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Systems* **5**, 319-331, doi:<https://doi.org/10.1016/j.cels.2017.07.004> (2017).
- 24 Morgan, M. A. J. & Shilatifard, A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nature Genetics* **52**, 1271-1281, doi:10.1038/s41588-020-00736-4 (2020).
- 25 Couture, J.-F. & Trievel, R. C. Histone-modifying enzymes: encrypting an enigmatic epigenetic code. *Current Opinion in Structural Biology* **16**, 753-760, doi:<https://doi.org/10.1016/j.sbi.2006.10.002> (2006).
- 26 Biel, M., Wascholowski, V. & Giannis, A. Epigenetics—An Epicenter of Gene Regulation: Histones and Histone-Modifying Enzymes. *Angewandte Chemie International Edition* **44**, 3186-3216, doi:<https://doi.org/10.1002/anie.200461346> (2005).

- 27 Lin, W. & Dent, S. Y. R. Functions of histone-modifying enzymes in development. *Current Opinion in Genetics & Development* **16**, 137-142, doi:<https://doi.org/10.1016/j.gde.2006.02.002> (2006).
- 28 Butler, J. S., Koutelou, E., Schibler, A. C. & Dent, S. Y. R. Histone-modifying enzymes: regulators of developmental decisions and drivers of human disease. *Epigenomics* **4**, 163-177, doi:10.2217/epi.12.3 (2012).
- 29 Zhang, K. & Dent, S. Y. R. Histone modifying enzymes and cancer: Going beyond histones. *Journal of Cellular Biochemistry* **96**, 1137-1148, doi:<https://doi.org/10.1002/jcb.20615> (2005).
- 30 Ropero, S. & Esteller, M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol* **1**, 19-25, doi:10.1016/j.molonc.2007.01.001 (2007).
- 31 Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* **13**, 840-852, doi:10.1038/nrg3306 (2012).
- 32 David, S. A. *et al.* An Assessment of Fixed and Native Chromatin Preparation Methods to Study Histone Post-Translational Modifications at a Whole Genome Scale in Skeletal Muscle Tissue. *Biol Proced Online* **19**, 10, doi:10.1186/s12575-017-0059-0 (2017).
- 33 Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943-949 (2014).
- 34 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).

- 35 Dahl, J. A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552 (2016).
- 36 Zhang, B. J. *et al.* Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557 (2016).
- 37 Zheng, H. *et al.* Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Molecular Cell* **63**, 1066-1079 (2016).
- 38 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, doi:10.1021/acs.analchem.8b01541 (2018).
- 39 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, doi:10.1038/Nmeth.3488 (2015).
- 40 Zhu, B. H. *et al.* MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nature Protocols* **14**, 3366-3394 (2019).
- 41 Deng, C. *et al.* Multiplexed and Ultralow-Input ChIP-seq Enabled by Tagmentation-Based Indexing and Facile Microfluidics. *Anal Chem* **92**, 13661-13666, doi:10.1021/acs.analchem.0c02550 (2020).
- 42 Raha, D., Hong, M. & Snyder, M. ChIP-Seq: A Method for Global Identification of Regulatory Elements in the Genome. *Current Protocols in Molecular Biology* **91**, 21.19.21-21.19.14, doi:<https://doi.org/10.1002/0471142727.mb2119s91> (2010).
- 43 Klein, D. C. & Hainer, S. J. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Research* **28**, 69-85, doi:10.1007/s10577-019-09619-9 (2020).

- 44 O'Neill, L. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76-82, doi:10.1016/s1046-2023(03)00090-2 (2003).
- 45 Kidder, B. L., Hu, G. & Zhao, K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nature Immunology* **12**, 918-922, doi:10.1038/ni.2117 (2011).
- 46 Adli, M. & Bernstein, B. E. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nature Protocols* **6**, 1656-1668, doi:10.1038/nprot.2011.402 (2011).
- 47 Shankaranarayanan, P. *et al.* Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat Methods* **8**, 565-567, doi:10.1038/nmeth.1626 (2011).
- 48 Fossli, M. *et al.* Going low to reach high: Small-scale ChIP-seq maps new terrain. *WIREs Systems Biology and Medicine* **12**, e1465, doi:<https://doi.org/10.1002/wsbm.1465> (2020).
- 49 Skene, P. J. & Henikoff, S. A simple method for generating high-resolution maps of genome-wide protein binding. *eLife* **4**, e09225, doi:10.7554/eLife.09225 (2015).
- 50 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856, doi:10.7554/eLife.21856 (2017).
- 51 Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314, doi:10.7554/eLife.46314 (2019).
- 52 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols* **13**, 1006-1019, doi:10.1038/nprot.2018.015 (2018).

- 53 Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nature Protocols* **15**, 3264-3283, doi:10.1038/s41596-020-0373-x (2020).
- 54 Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930, doi:10.1038/s41467-019-09982-5 (2019).
- 55 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, doi:10.7554/eLife.21856 (2017).
- 56 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, doi:10.1038/nprot.2018.015 (2018).
- 57 Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**, doi:10.7554/eLife.46314 (2019).
- 58 Liu, Z. *et al.* nMOWChIP-seq: low-input genome-wide mapping of non-histone targets. *NAR Genomics and Bioinformatics* **4**, lqac030, doi:10.1093/nargab/lqac030 (2022).
- 59 Sun, H. *et al.* Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids Res* **39**, 190-201, doi:10.1093/nar/gkq775 (2011).
- 60 Sun, Z. *et al.* EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat Commun* **10**, 3892, doi:10.1038/s41467-019-11905-3 (2019).
- 61 Brind'Amour, J. *et al.* An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun* **6**, 6033, doi:10.1038/ncomms7033 (2015).

- 62 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831, doi:10.1101/gr.136184.111 (2012).
- 63 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, doi:10.1073/pnas.95.3.906 (1998).
- 64 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, doi:10.1098/rstb.2002.1099 (2002).
- 65 Fine, E. J., Ionita, C. C. & Lohr, L. The History of the Development of the Cerebellar Examination. *Semin Neurol* **22**, 375-384 (2002).
- 66 Caporale, N. & Dan, Y. Spike timing-dependent plasticity: a Hebbian learning rule. *Annu Rev Neurosci* **31**, 25-46, doi:10.1146/annurev.neuro.31.060407.125639 (2008).
- 67 Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445, doi:10.1016/j.cell.2010.03.030 (2010).
- 68 Ma, S., Charron, J. & Erikson, R. L. Role of Plk2 (Snk) in mouse development and cell proliferation. *Mol Cell Biol* **23**, 6936-6943, doi:10.1128/mcb.23.19.6936-6943.2003 (2003).
- 69 Funahashi, Y. *et al.* Phosphorylation of Npas4 by MAPK Regulates Reward-Related Gene Expression and Behaviors. *Cell Rep* **29**, 3235-3252 e3239, doi:10.1016/j.celrep.2019.10.116 (2019).

- 70 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, doi:10.1016/j.mam.2012.07.006 (2013).
- 71 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, doi:10.1523/JNEUROSCI.1880-05.2006 (2006).
- 72 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).
- 73 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, doi:<https://doi.org/10.1046/j.1471-4159.2002.00931.x> (2002).
- 74 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, doi:<https://doi.org/10.1016/j.ydbio.2006.04.004> (2006).
- 75 Chakrabarti, K. *et al.* Critical role for Kalirin in nerve growth factor signaling through TrkA. *Mol Cell Biol* **25**, 5106-5118, doi:10.1128/MCB.25.12.5106-5118.2005 (2005).
- 76 Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883-891, doi:10.1038/nature02661 (2004).
- 77 Papandreou, A. *et al.* The expanding spectrum of movement disorders in genetic epilepsies. *Dev Med Child Neurol* **62**, 178-191, doi:10.1111/dmcn.14407 (2020).

- 78 Harris, L. *et al.* Neurogenic differentiation by hippocampal neural stem and progenitor cells is biased by NFIX expression. *Development* **145**, doi:10.1242/dev.155689 (2018).
- 79 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, doi:10.1016/j.biocel.2012.08.012 (2012).
- 80 Grinberg, I. *et al.* Heterozygous deletion of the linked genes ZIC1 and ZIC4 is involved in Dandy-Walker malformation. *Nat Genet* **36**, 1053-1055, doi:10.1038/ng1420 (2004).
- 81 Stark, R. & Brown, G. DiffBind: differential binding analysis of ChIP-seq peak data. (2019).
- 82 Kidder, B. L., Hu, G. & Zhao, K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol* **12**, 918-922, doi:10.1038/ni.2117 (2011).
- 83 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, doi:10.1073/pnas.1316064110 (2013).
- 84 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, doi:10.1186/gb-2009-10-3-r25 (2009).
- 85 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 86 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

- 87 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
- 88 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
- 89 Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27, doi:10.1016/j.cell.2012.06.013 (2012).
- 90 Dawson, M. A. The cancer epigenome: Concepts, challenges, and therapeutic opportunities. *Science* **355**, 1147-1152, doi:doi:10.1126/science.aam7304 (2017).
- 91 Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* **8**, 286-298, doi:10.1038/nrg2005 (2007).
- 92 Feinberg, A. P., Koldobskiy, M. A. & Gondor, A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet* **17**, 284-299, doi:10.1038/nrg.2016.13 (2016).
- 93 Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, doi:10.1126/science.aal2380 (2017).
- 94 Pfister, S. X. & Ashworth, A. Marked for death: targeting epigenetic changes in cancer. *Nat Rev Drug Discov* **16**, 241-263, doi:10.1038/nrd.2016.256 (2017).
- 95 Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nature Genetics* **50**, 591-602, doi:10.1038/s41588-018-0073-4 (2018).
- 96 Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, doi:10.1126/science.aav1898 (2018).

- 97 Pomerantz, M. M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat Genet* **52**, 790-799, doi:10.1038/s41588-020-0664-8 (2020).
- 98 Della Chiara, G. *et al.* Epigenomic landscape of human colorectal cancer unveils an aberrant core of pan-cancer enhancers orchestrated by YAP/TAZ. *Nat Commun* **12**, 2340, doi:10.1038/s41467-021-22544-y (2021).
- 99 Cohen, A. J. *et al.* Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat Commun* **8**, 14400, doi:10.1038/ncomms14400 (2017).
- 100 Li, Q. L. *et al.* Genome-wide profiling in colorectal cancer identifies PHF19 and TBC1D16 as oncogenic super enhancers. *Nat Commun* **12**, 6407, doi:10.1038/s41467-021-26600-5 (2021).
- 101 Geng, H. *et al.* Integrative epigenomic analysis identifies biomarkers and therapeutic targets in adult B-acute lymphoblastic leukemia. *Cancer Discov* **2**, 1004-1023, doi:10.1158/2159-8290.CD-12-0208 (2012).
- 102 Li, J. *et al.* A genomic and epigenomic atlas of prostate cancer in Asian populations. *Nature* **580**, 93-99, doi:10.1038/s41586-020-2135-x (2020).
- 103 Lomberk, G. *et al.* Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. *Nat Commun* **9**, 1978, doi:10.1038/s41467-018-04383-6 (2018).
- 104 Kelley, D. Z. *et al.* Integrated Analysis of Whole-Genome ChIP-Seq and RNA-Seq Data of Primary Head and Neck Tumor Samples Associates HPV Integration Sites with Open Chromatin Marks. *Cancer Res* **77**, 6538-6550, doi:10.1158/0008-5472.CAN-17-0833 (2017).

- 105 Bade, B. C. & Dela Cruz, C. S. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clin Chest Med* **41**, 1-24, doi:10.1016/j.ccm.2019.10.001 (2020).
- 106 Duruisseaux, M. & Esteller, M. Lung cancer epigenetics: From knowledge to applications. *Semin Cancer Biol* **51**, 116-128, doi:10.1016/j.semcancer.2017.09.005 (2018).
- 107 Rayleen V. BOWMAN, I. A. Y., Annalese B.T. SEMMLER, Kwun M. FONG. Epigenetics of lung cancer. *Respirology* **11**, 355-365, doi:10.1111/j.1400-1843.2006.00859.x (2006).
- 108 Langevin, S. M., Kratzke, R. A. & Kelsey, K. T. Epigenetics of lung cancer. *Transl Res* **165**, 74-90, doi:10.1016/j.trsl.2014.03.001 (2015).
- 109 Wen, J., Fu, J., Zhang, W. & Guo, M. Genetic and epigenetic changes in lung carcinoma and their clinical implications. *Mod Pathol* **24**, 932-943, doi:10.1038/modpathol.2011.46 (2011).
- 110 Seligson, D. B. *et al.* Global levels of histone modifications predict prognosis in different cancers. *Am J Pathol* **174**, 1619-1628, doi:10.2353/ajpath.2009.080874 (2009).
- 111 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 112 Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191-W198, doi:10.1093/nar/gkz369 (2019).
- 113 Zhang, K., Wang, M., Zhao, Y. & Wang, W. Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Science Advances* **5**, eaav3262, doi:10.1126/sciadv.aav3262 (2019).

- 114 Zhu, Y. *et al.* Constructing 3D interaction maps from 1D epigenomes. *Nature Communications* **7**, 10812, doi:10.1038/ncomms10812 (2016).
- 115 Kim, T. *et al.* Comparative characterization of 3D chromatin organization in triple-negative breast cancers. *Experimental & Molecular Medicine* **54**, 585-600, doi:10.1038/s12276-022-00768-2 (2022).
- 116 Vilarrasa-Blasi, R. *et al.* Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nature Communications* **12**, 651, doi:10.1038/s41467-020-20849-y (2021).
- 117 Brock, M. V., Herman, J. G. & Baylin, S. B. Cancer as a Manifestation of Aberrant Chromatin Structure. *The Cancer Journal* **13** (2007).
- 118 Zheng, L. & Wang, W. Regulation associated modules reflect 3D genome modularity associated with chromatin activity. *Nature Communications* **13**, 5281, doi:10.1038/s41467-022-32911-y (2022).
- 119 Parker, A. L. *et al.* Extracellular matrix profiles determine risk and prognosis of the squamous cell carcinoma subtype of non-small cell lung carcinoma. *Genome Medicine* **14**, 126, doi:10.1186/s13073-022-01127-6 (2022).
- 120 Naba, A. *et al.* Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. *Journal of Proteome Research* **16**, 3083-3091, doi:10.1021/acs.jproteome.7b00191 (2017).
- 121 Naba, A., Clauser, K. R., Lamar, J. M., Carr, S. A. & Hynes, R. O. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. *eLife* **3**, e01308, doi:10.7554/eLife.01308 (2014).

- 122 Hebert, J. D. *et al.* Proteomic Profiling of the ECM of Xenograft Breast Cancer Metastases in Different Organs Reveals Distinct Metastatic Niches. *Cancer Research* **80**, 1475-1485, doi:10.1158/0008-5472.CAN-19-2961 (2020).
- 123 Park, S. *et al.* Crizotinib attenuates cancer metastasis by inhibiting TGF β signaling in non-small cell lung cancer cells. *Experimental & Molecular Medicine* **54**, 1225-1235, doi:10.1038/s12276-022-00835-8 (2022).
- 124 Jeon, H.-S. & Jen, J. TGF- β Signaling and the Role of Inhibitory Smads in Non-small Cell Lung Cancer. *Journal of Thoracic Oncology* **5**, 417-419, doi:10.1097/JTO.0b013e3181ce3afd (2010).
- 125 Ai, R. *et al.* Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. *Nature Communications* **9**, 1921, doi:10.1038/s41467-018-04310-9 (2018).
- 126 Hon, G., Ren, B. & Wang, W. ChromaSig: A Probabilistic Approach to Finding Common Chromatin Signatures in the Human Genome. *PLOS Computational Biology* **4**, e1000201, doi:10.1371/journal.pcbi.1000201 (2008).
- 127 Gilliland, D. G. & Griffin, J. D. The roles of FLT3 in hematopoiesis and leukemia. *Blood* **100**, 1532-1542, doi:10.1182/blood-2002-02-0492 (2002).
- 128 Kim, J. & Kim, H. Recruitment and Biological Consequences of Histone Modification of H3K27me3 and H3K9me3. *ILAR Journal* **53**, 232-239, doi:10.1093/ilar.53.3-4.232 (2012).
- 129 Egeblad, M., Nakasone, E. S. & Werb, Z. Tumors as Organs: Complex Tissues that Interface with the Entire Organism. *Developmental Cell* **18**, 884-901, doi:10.1016/j.devcel.2010.05.012 (2010).

- 130 Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdag, P. & De Preter, K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nature Communications* **11**, 5650, doi:10.1038/s41467-020-19015-1 (2020).
- 131 Jeong, Y. *et al.* Systematic evaluation of cell-type deconvolution pipelines for sequencing-based bulk DNA methylomes. *Briefings in Bioinformatics* **23**, bbac248, doi:10.1093/bib/bbac248 (2022).
- 132 Wu, S. J. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature Biotechnology* **39**, 819-824, doi:10.1038/s41587-021-00865-z (2021).
- 133 Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature Biotechnology* **39**, 825-835, doi:10.1038/s41587-021-00869-9 (2021).
- 134 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 135 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2012).
- 136 Weirauch, Matthew T. *et al.* Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**, 1431-1443, doi:10.1016/j.cell.2014.08.009 (2014).
- 137 Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* **11**, 360-361, doi:10.1038/nmeth.2892 (2014).

- 138 Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090, doi:10.1126/science.aaa6090 (2015).
- 139 Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363, doi:10.1126/science.1250212 (2014).
- 140 Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, doi:10.1126/science.aat5691 (2018).
- 141 Eng, C. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239, doi:10.1038/s41586-019-1049-y (2019).
- 142 Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467, doi:10.1126/science.aaw1219 (2019).
- 143 Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681 e1618, doi:10.1016/j.cell.2020.10.026 (2020).
- 144 Liu, H. *et al.* DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598**, 120-128, doi:10.1038/s41586-020-03182-8 (2021).
- 145 Deng, Y. *et al.* Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681-686, doi:10.1126/science.abg7216 (2022).
- 146 Yvernogeau, L. *et al.* Multispecies RNA tomography reveals regulators of hematopoietic stem cell birth in the embryonic aorta. *Blood* **136**, 831-844, doi:10.1182/blood.2019004446 (2020).
- 147 Hauser, R. M., Henshall, D. C. & Lubin, F. D. The Epigenetics of Epilepsy and Its Progression. *Neuroscientist* **24**, 186-200, doi:10.1177/1073858417705840 (2018).

- 148 Conboy, K., Henshall, D. C. & Brennan, G. P. Epigenetic principles underlying epileptogenesis and epilepsy syndromes. *Neurobiol Dis* **148**, 105179, doi:10.1016/j.nbd.2020.105179 (2021).
- 149 Henshall, D. C. Epigenetic changes in status epilepticus. *Epilepsia* **59 Suppl 2**, 82-86, doi:10.1111/epi.14502 (2018).
- 150 Henshall, D. C. & Kobow, K. Epigenetics and Epilepsy. *Cold Spring Harb Perspect Med* **5**, doi:10.1101/cshperspect.a022731 (2015).
- 151 Allen Reference Atlas – Mouse Brain [brain atlas]. Available from atlas.brain-map.org.
- 152 Guo, H. *et al.* Disruptive mutations in TANC2 define a neurodevelopmental syndrome associated with psychiatric disorders. *Nature Communications* **10**, 4679, doi:10.1038/s41467-019-12435-8 (2019).
- 153 Inoue, T., Ota, M., Ogawa, M., Mikoshiba, K. & Aruga, J. Zic1 and Zic3 Regulate Medial Forebrain Development through Expansion of Neuronal Progenitors. *The Journal of Neuroscience* **27**, 5461, doi:10.1523/JNEUROSCI.4046-06.2007 (2007).
- 154 Kim, S.-G. *et al.* Tanc2-mediated mTOR inhibition balances mTORC1/2 signaling in the developing mouse brain and human neurons. *Nature Communications* **12**, 2695, doi:10.1038/s41467-021-22908-4 (2021).
- 155 The principles of nerve cell communication. (1997).
- 156 Raimondo, J. V., Burman, R. J., Katz, A. A. & Akerman, C. J. Ion dynamics during seizures. *Front Cell Neurosci* **9**, 419, doi:10.3389/fncel.2015.00419 (2015).
- 157 Gonzalez, O. C., Krishnan, G. P., Timofeev, I. & Bazhenov, M. Ionic and synaptic mechanisms of seizure generation and epileptogenesis. *Neurobiol Dis* **130**, 104485, doi:10.1016/j.nbd.2019.104485 (2019).

- 158 Murphy, P. & Burnham, W. M. The effect of kindled seizures on the locomotory behavior of long–evans rats. *Experimental Neurology* **180**, 88-92, doi:[https://doi.org/10.1016/S0014-4886\(02\)00056-0](https://doi.org/10.1016/S0014-4886(02)00056-0) (2003).
- 159 Möller, C. *et al.* Impact of repeated kindled seizures on heart rate rhythms, heart rate variability, and locomotor activity in rats. *Epilepsy & Behavior* **92**, 36-44, doi:<https://doi.org/10.1016/j.yebeh.2018.11.034> (2019).
- 160 Stewart, L. S. & Leung, L. S. Temporal lobe seizures alter the amplitude and timing of rat behavioral rhythms. *Epilepsy & Behavior* **4**, 153-160, doi:[https://doi.org/10.1016/S1525-5050\(03\)00006-4](https://doi.org/10.1016/S1525-5050(03)00006-4) (2003).
- 161 Lesca, G. *et al.* GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat Genet* **45**, 1061-1066, doi:10.1038/ng.2726 (2013).
- 162 Strehlow, V. *et al.* GRIN2A-related disorders: genotype and functional consequence predict phenotype. *Brain* **142**, 80-92, doi:10.1093/brain/awy304 (2019).
- 163 Venkateswaran, S. *et al.* Whole-exome sequencing in an individual with severe global developmental delay and intractable epilepsy identifies a novel, de novo GRIN2A mutation. *Epilepsia* **55**, e75-79, doi:10.1111/epi.12663 (2014).
- 164 Nakamura, K. *et al.* De Novo mutations in GNAO1, encoding a Gα subunit of heterotrimeric G proteins, cause epileptic encephalopathy. *Am J Hum Genet* **93**, 496-505, doi:10.1016/j.ajhg.2013.07.014 (2013).
- 165 Feng, H., Khalil, S., Neubig, R. R. & Sidiropoulos, C. A mechanistic review on GNAO1-associated movement disorder. *Neurobiol Dis* **116**, 131-141, doi:10.1016/j.nbd.2018.05.005 (2018).

- 166 Singh, K. *et al.* Neuronal Growth and Behavioral Alterations in Mice Deficient for the Psychiatric Disease-Associated Negr1 Gene. *Front Mol Neurosci* **11**, 30, doi:10.3389/fnmol.2018.00030 (2018).
- 167 Lin, A. Y. *et al.* A rare autism-associated MINT2/APBA2 mutation disrupts neurexin trafficking and synaptic function. *Sci Rep* **9**, 6024, doi:10.1038/s41598-019-42635-7 (2019).
- 168 Smits, J. J. *et al.* De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment. *Hum Genet* **138**, 61-72, doi:10.1007/s00439-018-1965-1 (2019).
- 169 Catania, E. H., Pimenta, A. & Levitt, P. Genetic deletion of Lsamp causes exaggerated behavioral activation in novel environments. *Behav Brain Res* **188**, 380-390, doi:10.1016/j.bbr.2007.11.022 (2008).
- 170 Strang, K. H., Golde, T. E. & Giasson, B. I. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. *Lab Invest* **99**, 912-928, doi:10.1038/s41374-019-0197-x (2019).
- 171 Shaw-Smith, C. *et al.* Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* **38**, 1032-1037, doi:10.1038/ng1858 (2006).
- 172 Nyegaard, M. *et al.* CACNA1C (rs1006737) is associated with schizophrenia. *Mol Psychiatry* **15**, 119-121, doi:10.1038/mp.2009.69 (2010).
- 173 Feng, L. & Cooper, J. A. Dual functions of Dab1 during brain development. *Mol Cell Biol* **29**, 324-332, doi:10.1128/MCB.00663-08 (2009).

- 174 Hoxha, E. *et al.* Emerging roles of Fgf14 in behavioral control. *Behav Brain Res* **356**, 257-265, doi:10.1016/j.bbr.2018.08.034 (2019).
- 175 Tadmouri, A. *et al.* Cacnb4 directly couples electrical activity to gene expression, a process defective in juvenile epilepsy. *EMBO J* **31**, 3730-3744, doi:10.1038/emboj.2012.226 (2012).
- 176 Stogmann, E. *et al.* Autosomal recessive cortical myoclonic tremor and epilepsy: association with a mutation in the potassium channel associated gene CNTN2. *Brain* **136**, 1155-1160, doi:10.1093/brain/awt068 (2013).
- 177 Francks, C. *et al.* LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Mol Psychiatry* **12**, 1129-1139, 1057, doi:10.1038/sj.mp.4002053 (2007).
- 178 Tse, K., Puttachary, S., Beamer, E., Sills, G. J. & Thippeswamy, T. Advantages of Repeated Low Dose against Single High Dose of Kainate in C57BL/6J Mouse Model of Status Epilepticus: Behavioral and Electroencephalographic Studies. *PLOS ONE* **9**, e96622, doi:10.1371/journal.pone.0096622 (2014).
- 179 Hartigan, J. A. & Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**, 100-108, doi:10.2307/2346830 (1979).
- 180 Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287, doi:10.1186/s12859-017-1708-7 (2017).
- 181 Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev* **19**, 1635-1655, doi:10.1101/gad.1324305 (2005).

- 182 Chen, Z., Li, S., Subramaniam, S., Shyy, J. Y. & Chien, S. Epigenetic Regulation: A New Frontier for Biomedical Engineers. *Annu Rev Biomed Eng* **19**, 195-219, doi:10.1146/annurev-bioeng-071516-044720 (2017).
- 183 Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43**, 904-914, doi:10.1016/j.molcel.2011.08.018 (2011).
- 184 Khorkova, O., Hsiao, J. & Wahlestedt, C. Basic biology and therapeutic implications of lncRNA. *Advanced Drug Delivery Reviews* **87**, 15-24, doi:<https://doi.org/10.1016/j.addr.2015.05.012> (2015).
- 185 Wu, Z. *et al.* Regulation of lncRNA expression. *Cellular & Molecular Biology Letters* **19**, 561-575, doi:10.2478/s11658-014-0212-6 (2014).
- 186 Hon, C. C. *et al.* An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**, 199-204, doi:10.1038/nature21374 (2017).
- 187 Kopp, F. & Mendell, J. T. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393-407, doi:10.1016/j.cell.2018.01.011 (2018).
- 188 Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**, 47-62, doi:10.1038/nrg.2015.10 (2016).
- 189 Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology* **22**, 96-118, doi:10.1038/s41580-020-00315-9 (2021).
- 190 Yao, R.-W., Wang, Y. & Chen, L.-L. Cellular functions of long noncoding RNAs. *Nature Cell Biology* **21**, 542-551, doi:10.1038/s41556-019-0311-8 (2019).
- 191 Kung, J. T., Colognori, D. & Lee, J. T. Long noncoding RNAs: past, present, and future. *Genetics* **193**, 651-669, doi:10.1534/genetics.112.146704 (2013).

- 192 Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749, doi:10.7554/eLife.01749 (2013).
- 193 Yang, G., Lu, X. & Yuan, L. LncRNA: A link between RNA and cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1839**, 1097-1109, doi:<https://doi.org/10.1016/j.bbagrm.2014.08.012> (2014).
- 194 Gutschner, T., Hammerle, M. & Diederichs, S. MALAT1- a paradigm for long noncoding RNA function in cancer. *J Mol Med* **91**, 791-801, doi:10.1007/s00109-013-1028-y (2013).
- 195 Liu, S. J., Dang, H. X., Lim, D. A., Feng, F. Y. & Maher, C. A. Long noncoding RNAs in cancer metastasis. *Nature Reviews Cancer* **21**, 446-460, doi:10.1038/s41568-021-00353-1 (2021).
- 196 Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**, 667-678, doi:10.1016/j.molcel.2011.08.027 (2011).
- 197 Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516-520, doi:10.1038/nature12210 (2013).
- 198 Long, J. *et al.* Long noncoding RNA Tug1 regulates mitochondrial bioenergetics in diabetic nephropathy. *J Clin Invest* **126**, 4205-4218, doi:10.1172/JCI87927 (2016).
- 199 Miao, Y. *et al.* Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat Commun* **9**, 292, doi:10.1038/s41467-017-02113-y (2018).

- 200 Zapparoli, E. *et al.* Comprehensive multi-omics analysis uncovers a group of TGF- β -regulated genes among lncRNA EPR direct transcriptional targets. *Nucleic Acids Research* **48**, 9053-9066, doi:10.1093/nar/gkaa628 (2020).
- 201 Vidovic, D. *et al.* ALDH1A3-regulated long non-coding RNA NRAD1 is a potential novel target for triple-negative breast tumors and cancer stem cells. *Cell Death & Differentiation* **27**, 363-378, doi:10.1038/s41418-019-0362-1 (2020).
- 202 van Solingen, C. *et al.* Long noncoding RNA CHROMR regulates antiviral immunity in humans. *Proceedings of the National Academy of Sciences* **119**, e2210321119, doi:10.1073/pnas.2210321119 (2022).
- 203 Lu, J. Y. *et al.* Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports* **30**, 3296-3311.e3295, doi:10.1016/j.celrep.2020.02.048 (2020).

Publications

1. Zhengzhi Liu, Chengyu Deng, Zirui Zhou, Ya Xiao, Shan Jiang, Bohan Zhu, Lynette B. Naler, Xiaoting Jia, Danfeng (Daphne) Yao, Chang Lu. “Epigenomic tomography for probing spatially-defined molecular state in the brain.” (2023); submitted
2. Zhengzhi Liu, Lina Zheng, Chang Lu, Wei Wang. “Comprehensive multi-mark and multi-omic profiling reveals epigenetic and transcriptional rewiring in lung tumors.” (2023); in preparation
3. Qiang Zhang, Sai Ma, Zhengzhi Liu, Bohan Zhu, Zirui Zhou, Gaoshan Li, J. Javier Meana, Javier González-Maeso, Chang Lu. “Droplet-based bisulfite sequencing for high-throughput profiling of single-cell DNA methylomes.” (2023); Under revision in Nature Communications
4. Zhengzhi Liu, Lynette B. Naler, Yan Zhu, Chengyu Deng, Qiang Zhang, Bohan Zhu, Zirui Zhou, Mimosa Sarma, Alexander Murray, Hehuang Xie and Chang Lu. “nMOWChIP-seq: low-input genome-wide mapping of non-histone targets.” *NAR Genomics and Bioinformatics* **4**, lqac030 (2022); doi: 10.1093/nargab/lqac030
5. Zhengzhi Liu, Niaz Banaei and Kangning Ren. “Microfluidics for Combating Antimicrobial Resistance.” *Trends in Biotechnology* **35**, 1129-1139 (2017); doi: 10.1016/j.tibtech.2017.07.008
6. Zhengzhi Liu, Han Sun and Kangning Ren. “Multiplexed, gradient-based, full-hydrogel microfluidic platform for rapid, high-throughput antimicrobial susceptibility testing.” *ChemPlusChem* **82**, 792-801 (2017); doi: 10.1002/cplu.201600654

7. Sun, Han, Zhengzhi Liu, Chong Hu and Kangning Ren. "Cell-on-hydrogel platform made of agar and alginate for rapid, low-cost, multidimensional test of antimicrobial susceptibility." *Lab on a Chip* **16**, 3130-3138 (2016); doi: 10.1039/C6LC00417B
8. Hu, Chong, Han Sun, Zhengzhi Liu, Yin Chen, Yangfan Chen, Hongkai Wu and Kangning Ren, "Freestanding 3-D microvascular networks made of alginate hydrogel as a universal tool to create microchannels inside hydrogels." *Biomicrofluidics* **10**, 044112 (2016); doi: 10.1063/1.4961969

Appendix

Primers used in this thesis

Sample type	Locus name	Primer sequence
GM12878 (Pol II, EGR1, MEF2C, HDAC2)	<i>ACTG1</i> (F)	CGG AAA GAT CGC CAT ATA TGG AC
	<i>ACTG1</i> (R)	ACC GGC AGA GAA ACG CGA
	<i>POLR2A</i> (F)	GAG AGA CAA ACT GCC GTA ACC
	<i>POLR2A</i> (R)	GGG AAA TAA GGA GCG AAA GGA G
	<i>AFM</i> (F)	GCA GAA CCT AGT TCC TCC TTC AAC
	<i>AFM</i> (R)	AGT CAT CCC TTC CTA CAG ACT GAG A
	<i>KCTD16</i> (F)	GCC ATA AAG AAA CAC ACA TGG AAA C
	<i>KCTD16</i> (R)	GCA AAG CAG GAA CCC TCA AA
Mouse PFC and cerebellum (Pol II, EGR1, MEF2C)	<i>Arc</i> (F)	CAG CAT AAA TAG CCG CTG GT
	<i>Arc</i> (R)	GTC GCC GCT GAA GCT AGA
	<i>Polr2a</i> (F)	AAA GAA GGG AGG AGA GGA GGA
	<i>Polr2a</i> (R)	GGG AGA GAC AAA CTG CCG TAA
	<i>Afm</i> (F)	CAT TTG ACC CAA ACT GCT AAG T
	<i>Afm</i> (R)	GTA TGT GAA GTG TCA GCA ATG G
	<i>Gcg</i> (F)	TCA ACC CAA AGT CCC TGA AG
	<i>Gcg</i> (R)	CCA AGA GGT TGC ATT GGA AG
Human lung tissue	<i>BAZ1A</i> (F)	TCT CAA CTC CGC TCC TCT CT
	<i>BAZ1A</i> (R)	TGG GCT GGG CTT CGT TT

(H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K9me3)	<i>ANXA5 (F)</i>	GCT ACA CTG TAC TCC TAA A
	<i>ANXA5 (R)</i>	TGA CTT CCT GTG GTA TTG
	<i>C9ORF3 (F)</i>	CCT CCT CAG TTC TCC CAG ACT
	<i>C9ORF3 (R)</i>	AGC TGA GGT GGT AAG ATG TGA C
	<i>HOTAIR (F)</i>	TGT AAC AGG CAG GTG GAT TC
	<i>HOTAIR (R)</i>	GTG TAG ACC CAG CCC AAT TTA
	<i>BRSK2 (F)</i>	GCT TCT GTT GGC AGT TTC TTC
	<i>BRSK2 (R)</i>	CAC AGG AGA ACC ACT CTC AAG
	<i>N1 (F)</i>	TCA TCT GCA AAT GGG GAC AA
	<i>N1 (R)</i>	AGG ACA CCC CCT CTC AAC AC
	<i>N-SLC (F)</i>	TTC CCA ACG TCA CAG AGT TAG
	<i>N-SLC (R)</i>	GAC AGT ACA GCA CAG AGG TTA G
Mouse Neocortex (H3K27ac, H3K27me3)	<i>Igsf8 (F)</i>	CCA TCT GGA GGC TGA ATG TT
	<i>Igsf8 (R)</i>	TAT GGA AAC TGG CGT TCT GG
	<i>Sim2 (F)</i>	GGG TGT ACA ATT GAA TGC TGT ATG
	<i>Sim2 (R)</i>	GCT GGC TAA CCT TCC AGA AAT A
	<i>Afm (F)</i>	CAT TTG ACC CAA ACT GCT AAG T
	<i>Afm (R)</i>	GTA TGT GAA GTG TCA GCA ATG G
MDA231 (HOTAIR)	<i>HOTAIR (F)</i>	TGT AAC AGG CAG GTG GAT TC
	<i>HOTAIR (R)</i>	GTG TAG ACC CAG CCC AAT TTA
	<i>PROKR2 (F)</i>	GTG GCA TAT TCT GGT CTC CTA C
	<i>PROKR2 (R)</i>	CAG ACA GGA AGA GTG GAA AGG

	<i>AFM (F)</i>	GCA GAA CCT AGT TCC TCC TTC AAC
	<i>AFM (R)</i>	AGT CAT CCC TTC CTA CAG ACT GAG A
HeLa (TERC)	<i>TERC (F)</i>	CTG ACA GAG CCC AAC TCT TC
	<i>TERC (R)</i>	GCC TTC CAC CGT TCA TTC TA
	<i>LCLAT1 (F)</i>	AGG CGT CAG GTG CTG AG
	<i>LCLAT1 (R)</i>	CTG AGG TCG CAG GAG GAA G
	<i>DUSP11 (F)</i>	CAG GAT AAG ACC CTA AAC AGG AAA
	<i>DUSP11 (R)</i>	TAC TGG CCT CTG GCT AAA GA
	<i>JAGN1 (F)</i>	GGC GCA AAT AGG GTC AGT
	<i>JAGN1 (R)</i>	CGC CTC ACT TCA TAC CTC ATC
	<i>AFM (F)</i>	GCA GAA CCT AGT TCC TCC TTC AAC
	<i>AFM (R)</i>	AGT CAT CCC TTC CTA CAG ACT GAG A

Custom scripts used in this thesis

Pausing index computation

```
for file in *.bam

do

    base="${file%.bam}"

    bedtools coverage -counts -b $PWD/$base".bam" -a $PWD/hg19_TSS.bed >
    $PWD/$base"_TSS.bed"

    bedtools coverage -counts -b $PWD/$base".bam" -a $PWD/hg19_GB.bed >
    $PWD/$base"_GB.bed"

    cut -d$'\t' -f1-3,5,7 $PWD/$base"_TSS.bed" > $PWD/$base"_TSS_extracted.bed"

    cut -d$'\t' -f1-3,5,7 $PWD/$base"_GB.bed" > $PWD/$base"_GB_extracted.bed"

    awk -F$'\t' 'BEGIN {OFS=FS}{$6=$5/($3-$2); print}' $PWD/$base"_TSS_extracted.bed" >
    $PWD/$base"_TSS_calculated.bed"

    awk -F$'\t' 'BEGIN {OFS=FS}{$6=$5/($3-$2); print}' $PWD/$base"_GB_extracted.bed" >
    $PWD/$base"_GB_calculated.bed"

    paste $PWD/$base"_TSS_calculated.bed" $PWD/$base"_GB_calculated.bed" >
    $PWD/$base"_combined.bed"

    cut -d$'\t' -f4,6,12 $PWD/$base"_combined.bed" > $PWD/$base"_TSS-GB.bed"

done
```

Single cell BS-seq data QC pipeline

```
for file in *_1.fastq

do

    base="${file%_1.fastq}"

    f2="_2.fastq"

# trimming in pairs (5min)

cutadapt -f fastq -q 20 -m 62 -a AGATCGGAAGAGCACACGTCTGAAC -A
AGATCGGAAGAGCGTCGTGTAGGGA -o $base"_1_trimmed.fastq" -p
$base"_2_trimmed.fastq" $file $base$f2

# trimming single reads (5min)

cutadapt -f fastq -u 16 -u -16 -m 30 -o $base"_1_Trimmed.fastq" $base"_1_trimmed.fastq"

cutadapt -f fastq -u 16 -u -16 -m 30 -o $base"_2_Trimmed.fastq" $base"_2_trimmed.fastq"

# bismark mapping of single reads (60min), this part requires large amounts of memory,
shortage of which causes processing being stuck

bismark --bowtie2 --pbat --genome /work/cascades/jerryliu/sc-Meth/hg19/
$base"_1_Trimmed.fastq" --multicore 16

bismark --bowtie2 --genome /work/cascades/jerryliu/sc-Meth/hg19/ $base"_2_Trimmed.fastq" -
-multicore 16

# combine mapped bam files from both reads (5-10min)
```

```

bamtools merge -in $base"_1_Trimmed_bismark_bt2.bam" -in
$base"_2_Trimmed_bismark_bt2.bam" -out $base"_aligned.bam"

# sort bam file (5min)

samtools sort -T $base"_temp.sorted" -o $base"_sorted.bam" $base"_aligned.bam"

# remove duplicate reads (3min)

java -jar /home/jerryliu/picard/picard.jar MarkDuplicates REMOVE_DUPLICATES=true
I=$base"_sorted.bam" O=$base"_marked.bam" M=$base"_dupmetrics.txt"

# filter mapped reads (5-10min)

methylypy bam-quality-filter --input-file $base"_marked.bam" --output-file $base"_filtered.bam" --
ref-fasta /work/cascades/jerryliu/sc-Meth/hg19/GRCh38.fa --min-mapq 30 --min-num-ch 3 --
max-mch-level 0.7 --buffer-line-number 100

# call methylated sites (15-30min)

methylypy call-methylation-state --input-file $base"_filtered.bam" --paired-end False --sample
$base --ref-fasta /work/cascades/jerryliu/sc-Meth/hg19/GRCh38.fa

done

```

Generating average bigwig profile from technical replicates

```

for file in *_1_Full_Normalized.bedGraph

do

base="${file%_1_Full_Normalized.bedGraph}"

```

```

cut -d$'\t' -f4 $PWD/$base"_1_Full_Normalized.bedGraph" > $PWD/$base"_1_Signal.bed"

cut -d$'\t' -f4 $PWD/$base"_2_Full_Normalized.bedGraph" > $PWD/$base"_2_Signal.bed"

paste $PWD/$base"_1_Signal.bed" $PWD/$base"_2_Signal.bed" >
$PWD/$base"_combined_Signal.bed"

awk -F$'\t' 'BEGIN {OFS=FS}{$3=($1+$2)/2; print}' $PWD/$base"_combined_Signal.bed" >
$PWD/$base"_average_Signal.bed"

cut -d$'\t' -f3 $PWD/$base"_average_Signal.bed" > $PWD/$base"_Average.bed"

paste ~/Data/Promotors/mm10/GenomeWindowed_sort.bed $PWD/$base"_Average.bed" >
$PWD/$base"_Average.bedGraph"

bedGraphToBigWig $PWD/$base"_Average.bedGraph" ~/Data/Promotors/mm10/Genome.bed
$PWD/$base"_Average.bw"

rm $PWD/$base"_1_Signal.bed" $PWD/$base"_2_Signal.bed"
$PWD/$base"_combined_Signal.bed" $PWD/$base"_average_Signal.bed"
$PWD/$base"_Average.bed" $PWD/$base"_Average.bedGraph"

done

```

Sorting peaks based on peak score

```

for f in *_peaks.narrowPeak

do

base=`basename "$f" _peaks.narrowPeak`

sort -r -g -k 5,5 < $base"_peaks.narrowPeak" > $base"_temp.out"

```

```

Incnt=$(wc -l < $base"_temp.out")

percent_linecount_infloat=$(echo "$Incnt*.3" | bc)

float2Int=$(printf %.0f "$percent_linecount_infloat")

head -"$float2Int" $base"_temp.out" > $base"_top03.narrowPeak"

rm $PWD/$base"_temp.out"

done

```

Generate consensus peakset in DiffBind

```

library(DiffBind)

K27ac_B<-dba(sampleSheet = "K27ac_B.csv",dir=system.file("extra", package="DiffBind"))

K27ac_B_consensus<-dba.peakset(K27ac_B,bRetrieve = TRUE)

write.csv(K27ac_B_consensus,file="C:/Users/Jerry Liu/Documents/K27ac_B_consensus.csv")

```

k-means clustering of consensus peak multibigwig summary matrix

```

K27ac_EB_enhancer<-read.csv("E:/R files/K27ac_EB_enhancer.csv",row.names = 1)

K27ac_EB_enhancer<-as.matrix(K27ac_EB_enhancer)

wss_K27ac_EB_enhancer<-sapply(1:15, function(k){kmeans(K27ac_EB_enhancer,k,nstart =
10,iter.max = 100)$tot.withinss})

plot(wss_K27ac_EB_enhancer,type="l")

K27ac_EB_enhancer_cluster_k5<-kmeans(K27ac_EB_enhancer,5,nstart = 10,iter.max = 100)

```

```

write.csv(K27ac_EB_enhancer_cluster_k5["cluster"],file="E:/R
files/K27ac_EB_enhancer_cluster_k5.csv")

# sort the original matrix based on the cluster number assignment from
"K27ac_EB_enhancer_cluster_k5.csv" and save the re-ordered matrix to
"K27ac_EB_enhancer_k5.csv"

K27ac_EB_enhancer_k5<-read.csv("E:/R files/K27ac_EB_enhancer_k5.csv",row.names = 1)

K27ac_EB_enhancer_k5<-as.matrix(K27ac_EB_enhancer_k5)

break_K27ac_EB_enhancer<-
seq(min(K27ac_EB_enhancer),max(K27ac_EB_enhancer),by=0.01)

palette_K27ac_EB_enhancer<- c(colorRampPalette(c("white",
"darkblue"))(220),colorRampPalette(c("darkblue",
"darkblue"))(length(break_K27ac_EB_enhancer)-220))

heatmap(K27ac_EB_enhancer_k5,Rowv = NA,Colv = NA,scale =
"none",col=palette_K27ac_EB_enhancer,break(break_K27ac_EB_enhancer),labRow =
NA,labCol = NA)

image(break_K27ac_EB_enhancer,1,as.matrix(1:length(break_K27ac_EB_enhancer)),col=palette_K27ac_EB_enhancer,ylab="",yaxt="n",xlab="")

```