Low-Input Multi-Omic Studies of Brain Neuroscience Involved in Mental Diseases

BOHAN ZHU

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
Chemical Engineering

Chang Lu, Chair
Richey M. Davis
Aaron Goldstein
Rong Tong

Aug 2, 2022

Blacksburg, VA

Keywords: Microfluidics, Epigenome, Transcriptome, Chromatin Immunoprecipitation, Next generation sequencing, RNA sequencing, Psychiatric disorders, Psychedelics, Schizophrenia, Antipsychotic treatment, Maternal Immune Activation

Low-Input Multi-Omic Studies of Brain Neuroscience Involved in Mental Diseases

Bohan Zhu

Abstract

Psychiatric disorders are believed to result from the combination of genetic predisposition and many environmental triggers. While the large number of disease-associated genetic variations have been recognized by previous genome-wide association studies (GWAS), the role of epigenetic mechanisms that mediate the effects of environmental factors on CNS gene activity in the etiology of most mental illnesses is still largely unclear. A growing body of evidence suggested that the abnormalities (changes in gene expression, formation of neural circuits, and behavior) involved in most psychiatric syndromes are preserved by epigenetic modifications identified in several specific brain regions. In this thesis, we developed the second generation of one of our microfluidic technologies (MOWChIP-seq) and used it to profile genome-wide histone modifications in three mental illness-related biological studies: the effect of psychedelics in mice, schizophrenia, and the effect of maternal immune activation in mice offspring. The second generation of MOWChIP-seq was designed to generate histone modification profiles from as few as 100 cells per assay with a throughput as high as eight assays in each run. Then, we applied the new MOWChIP-seq and SMART-seq2 to profile the histone modification H3K27ac and transcriptome, respectively, using NeuN+ neuronal nuclei from the mouse frontal cortex after a single

dose of psychedelic administration. The epigenomic and transcriptomic changes induced by 2,5-Dimethoxy-4-iodoamphetamine (DOI), a subtype of psychedelics, in mouse neuronal nuclei at various time points suggest that the long-lasting effects of the psychedelic are more closely related to epigenomic alterations than the changes in transcriptomic patterns. Next, we comprehensively characterized epigenomic and transcriptomic features from the frontal cortex of 29 individuals with schizophrenia and 29 individually matched controls (gender and age). We found that schizophrenia subjects exhibited thousands of neuronal vs. glial epigenetic differences at regions that included several susceptibility genetic loci, such as NRXN1, RGS4 and GRIN3A. Finally, we investigated the epigenetic and transcriptomic alterations induced by the maternal immune activation (MIA) in mice offspring's frontal cortex. Pregnant mice were injected with influenza virus at GD 9.5 and the frontal cortex from mice pups (10 weeks old) were examined later. The results offered us some insights into the contribution of MIA to the etiology of some mental disorders, like schizophrenia and autism.

Low-Input Multi-Omic Studies of Brain Neuroscience Involved in Mental Diseases

Bohan Zhu

General Audience Abstract

While this field is still in its early stage, the epigenetic studies of mental disorders present promise to expand our understanding about how environmental stimulates, interacting with genetic factors, contribute to the etiology of various psychiatric syndromes, like major depression and schizophrenia. Previous clinical trials suggested that psychedelics may represent a promising long-lasting treatment for patients with depression and other psychiatric conditions. These research presented the therapeutic potential of psychedelic compounds for treating major depression and demonstrated the capability of psychedelics in increasing dendritic density and stimulating synapse formation. However, the molecular mechanism mediating the clinical effectiveness of psychedelics remain largely unexplored. Our study revealed that epigenomic-driven changes in synaptic plasticity sustain psychedelics' long-lasting antidepressant action. Another serious mental illness is schizophrenia, which could affect how an individual feels, thinks, and behaves. Like most other mental disorders, schizophrenia results from a combination of genetic and environmental causes. Epigenetic marks allow a dynamic impact of environmental factors, including antipsychotic medications, on the access to

genes and regulatory elements. Despite this, no study so far has profiled cell-type-specific genome-wide histone modifications in postmortem brain samples from schizophrenia subjects or the effect of antipsychotic treatment on such epigenetic marks. Here we show the first comprehensive epigenomic characterization of the frontal cortex of 29 individuals with schizophrenia and 29 matched controls. The process of brain development is surprisingly sensitive to a lot of environmental insults. Epidemiological studies have recognized maternal immune activation as a risk factor that may change the normal developmental trajectory of the fetal brain and increase the odds of developing a range of psychiatric disorders, including schizophrenia and autism, in its lifetime. Given the prevalence of the coronavirus, uncovering the molecular mechanism underlie the phenotypic alterations has become more urgent than before, for both prevention and treatment.

Acknowledgments

I would like to thank to my advisors, family, and my friends. Without the help and support from them, I don't think I will be able to finish my research and meet the requirement for doctoral degree. They have essential roles in my growth and success during my time at Virginia Tech. I want to present my appreciation to all of them and wish to thank them for all the amazing things that they have done for me.

First, I must thank my advisor and the head of my committee, Dr. Chang Lu, for his enlightenment and support. He is not just a supportive advisor to me, but also a senior friend who is very kind and caring. He always speaks frankly with me and also with respect. Dr. Lu is the one who lead me into the beautiful and meaningful world of microfluidics and epigenomics. We have had tons of discussions over almost all the topics in my PhD journey. His mentorship, enthusiasm and critical thinking made me a more effective researcher and a mature person. His trust in me also let me build faith in myself. I am truly lucky to have had a reasonable, caring, and interesting mentor as I have had in Dr. Lu.

I would also like to thank my committee members, Dr. Aaron Goldstein, Dr. Richey Davis, and Dr. Rong Tong, for their practical and intuitive suggestions during my studies and preliminary presentation. Their support during my presentation in both seminar and symposium have been priceless. They have also been a source of enlightening and entertaining conversation.

I also want to thank my collaborators, Dr. Javier Gonzalez-Maeso and his postdoc Salvador Sierra, Dr. Wei Wang and his postdoc Richard I. Ainsworth and

Zengmiao Wang. Their knowledge in building mouse models and integrating multi-omic sequencing data helped me in my exploration in epigenomics and transcriptomics.

I am also indebted to the people in the Department of Chemical Engineering at Virginia Tech. The faculty members have provided lectures, advice, and support during my PhD life. The office staff, in particular Andrea Linkous and Diane Cannaday were a source of joy and positivity. We will definitely miss a lot of opportunities without the help from them.

I want to thank my lab mates who have supported me for many years: Dr. Sai Ma, Dr. Yan Zhu, Dr. Travis Murphy, Dr. Mimosa Sarma, Dr. Chengyu Deng, Dr. Yuan-Pang Hsieh, Dr. Lynette Naler, Qiang Zhang, Jerry (ZhengZhi) Liu, Zirui Zhou, Tom Hadlock, Gaoshan Li, Xin Zhang, Jenna Catalano, and Jacob Neice. They each were kindhearted, thoughtful, and amusing. This allows us to have a positive and enjoyable lab environment. They let me look forward to every day, even when I suffered from the failure of experiments.

In the end, I would like to thank my parents and friends for their support in allowing everything happened. My parents, Yiping Cai and Weimin Wang, advised and encouraged me over the years. They are the ones I always love and admire. My friends, who always support me to achieve my life goals and gave me positive and encouraging feedback to my graduate studies. Last but not least, my partner, Fei Wang, was a vital spring of strength in my life. She always encourages me to make decision based on my interest and support me with all the love.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BAM	Binary Alignment Map
BWA	Burrows-Wheeler Aligner
ChIP	Chromatin Immunoprecipitation
Ct	Cycle threshold
CTCF	CCCTC-Binding Factor
CUT&RUN	Cleavage Under Targets & Release Using Nuclease
DEG	Differentially Expressed Eenes
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
ERK	Extracellular signal-Regulated Kinase
FACS	Flourescence activated cell sorting
FBS	Fetal bovine serum
FC	Fold change
FDR	False Discovery Rate
FOXO1	Forkhead Box O1
GD	Gestation Day
GREAT	Genomic Regions Enrichment of Annotations Tool
GO	Gene ontology
GSEA	Gene set enrichment analysis
GWAS	Genome-wide association study
H3K36me3	Tri-methylation at the 36th lysine residue of the histone H3
H3K79me2	Bi-methylation at the 79th lysine residue of the histone H3
H3K9me3	Tri-methylation at the 9th lysine residue of the histone H3
HISAT2	Hierarchical Indexing for Spliced Alignment of Transcripts
IQGAP	IQ-motif containing Ras GTPase-Activating-like Protein
KEGG	Kyoto Encyclopedia of Genes and Genomes
LabVIEW	Laboratory Virtual Instrument Engineering Workbench
MACS2	Model-based Analysis of ChIP-Seq
MAPK	Mitogen-Activated Protein Kinases

MIA	Maternal immune activation
NSC	Normalized strand correlation
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PFA	Perfluoroalkoxy Alkanes
SCC	Spearman Correlation Coefficient
SDS	Sodium Dodecyl Sulfate
SNP	Single Nucleotide Polymorphism
SICER	Spatial Clustering for Identification of ChIP-Enriched Regions
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
TSS	Transcription start site
VEH	Vehicle

1 Overview

Recently, the coupling of epigenome and transcriptome has become a new direction that could help us better understand about the molecular mechanism that underlies the effect of environmental factors in the etiology of various psychiatric diseases[1-4]. The purpose of this research is to develop a new microfluidic protocol for profiling genome-wide histone modifications with low-input and high-throughput processing and use it to explore the role of epigenetic factors, which act as a bridge between various environmental factors and the phenotypic changes, in the development and treatment of various mental diseases.

It has been shown that epigenetic modifications, such as histone modifications and DNA methylation, play important roles in a variety of complex diseases, like cancer and mental illnesses[5,6]. The feature of genome-wide histone modifications can be profiled through chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq)[7,8]. In general, a ChIP-seq assay consists of several steps, including cell collection and lysis, chromatin fragmentation (e.g., sonication or enzyme digestion), immunoprecipitation, enriched DNA (ChIPed DNA) purification, sequencing library preparation, and next generation sequencing. There are largely three limitations of the currently available ChIP-seq protocols: **1)** the requirement of millions of cells to start from that excludes the application for the scarce amount of tissue (conventional ChIP-seq)[8], **2)** complex chemistry in indexing or surface immobilization that could be hard to be achieved (iChIP[9], ULi-ChIP[10] and surfaceChIP-seq[11]), **3)** low-throughput processing that makes the experiment time-consuming if it needs to deal with a large quantity of samples or contains multiple histone marks of interest (the original MOWChIP-seq[12]).

We demonstrated the second generation of the microfluidic oscillatory washing ChIP-seq (MOWChIP-seq) protocol for profiling histone modifications using as few as 100 cells per assay with a throughput as high as eight assays in one run[13] (Chapter 3).

Because the profile of a histone mark is highly specific to a cell type, cell isolation from tissues is often necessary to generate a homogeneous cell population, and such operations tend to yield a low number of cells[14]. The application of our new microfluidic technology enables us to generate genome-wide histone modification profiles with small number of cells isolated from primary tissues and further enhance our understanding of many complex multifactorial diseases, like mental disorders. We applied MOWChIP-seq and SMATR-seq2 to profile histone modification H3K27ac and transcriptome, respectively, using the small quantity of neuronal nuclei extracted from the mouse frontal cortex to examine whether a single dose of psychedelic leads to long-lasting epigenomic and transcriptomic changes in the mouse brain (Chapter 4). We also applied the combination of MOWChIP-seq and SMART-seq2 for low-input profiling of H3K27ac and H3K4me3 histone modifications and transcriptomes, respectively, to investigate the epigenetic and transcriptomic features of schizophrenia and alterations induced by repeated administration of antipsychotic medications (Chapter 5).

We also conducted the study of the epigenetic and transcriptomic alteration in mice offspring induced by maternal immune activation. We used the similar method above to profile two histone marks, H3K4me3 and H3K27ac and transcriptomic information using neuronal nuclei isolated from the brain tissue of seven experimental groups of mice. In this project, we not only examined the effect of maternal immune activation to the brain developmental process, but also tested the influence of MIA

affected siblings to the mice (Chapter 6). The focus of each chapter is summarized below:

CHAPTER 3 – MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications

Epigenetic modifications, like histone modifications, play critical roles in the controlling the structure of chromatin[14]. Profiling of various histone modifications at the genome scale using tissues from animal and human samples is an important step for functional studies of epigenomes and epigenomics-based precision medicine. While some advanced ChIP-seq protocols have been reported, this field is still suffered from the lack of technology that is low-input, multiplexed, and user-friendly. Here, we presented the second generation of MOWChIP-seq with high throughput (eight assays in parallel). Our process is also semi-automated to reduce labor and improve reproducibility. We used the protocol to study a number of histone modifications in various types of mouse and human tissues. The protocol can be conducted by a user who is familiar with molecular biology procedures and has basic engineering skills.

CHAPTER 4 – Investigation of prolonged epigenomic changes following single exposure to a psychedelic in mice brain

Clinical evidence suggests that rapid and long-term antidepressant action can be attained with a single exposure to psychedelics[15,16]. However, the biological substrates and key molecular mechanisms of psychedelics' enduring action remain unclear. Here, we show that a single does of DOI, a type of psychedelic, leads to changes in chromatin organization, particularly at enhancer regions of genes involved in synaptic assembly

that stretch for days after the psychedelic exposure. These DOI-induced alterations in the neuronal epigenome overlap with genetic loci associated with schizophrenia, depression, and attention deficit hyperactivity disorder. Together, these data support that epigenomic-driven changes in synaptic plasticity sustain psychedelics' long-lasting antidepressant action but also warn about potential substrate overlap with genetic risks for certain psychiatric conditions.

CHAPTER 5 – Examination of antipsychotic effect on cell-type-specific epigenomic landscapes in the frontal cortex of schizophrenia subjects

Schizophrenia results from a combination of genetic and environmental causes[17]. Epigenetic marks allow a dynamic impact of environmental factors, including antipsychotic medications, on access to genes and regulatory elements. However, no study so far has profiled cell-specific genome-wide histone modifications in postmortem brain samples from schizophrenia subjects or the effect of antipsychotic treatment on such epigenetic marks. Here we show the first comprehensive epigenomic characterization of the frontal cortex of 29 individuals with schizophrenia and 29 matched controls (42 males and 16 females, range from 18 to 74 years old). Comparing untreated and treated schizophrenia subjects with controls, our findings provide entirely new insights into differentially modified genes for unexpected pathways that are potential markers of antipsychotic treatment. Additionally, our results suggest that the effect of age on the epigenomic landscape is more pronounced in frontal cortex of antipsychotic-treated schizophrenics. This provides important evidence of epigenetic alterations in the frontal cortex of individuals with schizophrenia and remarks the impact of age and antipsychotic treatment on chromatin organization.

CHAPTER 6 – Deciphering epigenomic and transcriptomic changes in mice offspring brains induced by maternal immune activation

A large body of epidemiological research has suggested that a serious maternal infection during a specific period of pregnancy works as a critical environmental insult that could significantly increase the risk of serval neurodevelopmental and psychiatric disorders in offspring. However, our knowledge of the molecular mechanisms underlying the deleterious effects of this environmental exposures on the developmental trajectory of the fetal brain remains largely unclear. Given the growing impact of the coronavirus disease 2019 (COVID-19), it is vital for us to investigate all of the potential epigenetic or transcriptomic dysregulation that may be involved in the process. In this study, we used influenza virus or vehicle solution to stimulate the pregnant mice and activate their immune system. To examine the prenatal and postnatal effect of maternal immune activation, we swapped half of the infant mice right after their birth. In that case, each mother would raise both MIA affected baby and MIA unaffected baby, no matter if themselves were exposed to the virus or not. Overall, this project provides the first dataset with cell-type-specific epigenomic and transcriptomic features in mice frontal cortex after affected by maternal immune activation.

2 Introduction

2.1 Epigenomics

Epigenomics refers to the study of genome-wide epigenetic modifications which are biological mechanisms that regulate whether genes are switch on or off without any DNA sequence mutation[18]. Epigenetic modifications can influence the manufacture of proteins in each cell and play a substantial role in biological processes, like cell proliferation, differentiation, and death[19]. There are several common mechanisms, including DNA methylation, post-translational histone modification and non-coding RNA-associated gene silencing[18,20]. These modifications could induce alterations in chromatin structure and gene accessibility, and finally regulate the gene expression patterns. Different from the relatively stable genome, epigenetic marks such as histone modifications remain largely plastic during the development and adulthood period[21,22]. This allows a dynamic impact of environmental factors, including lifestyle, mental stress, organic pollutants, and drugs, on the access to genes and regulatory elements. In this dissertation, we will focus on histone modifications[23-26].

2.2 Chromatin structures

Within each eukaryotic cell, the length of DNA is about 2 meters[27]. In order to fit it into the nucleus of each cell, DNA has to be highly condensed and packed. Instead of being folded directly into the final chromatin structure, the packing of DNA consists of a series of hierarchies of organization. The first level of the organization is called nucleosome. Nucleosome is achieved by warping double strand DNA around an octamer of nuclear proteins (histone), forming a so-called ‘nucleosomal core particle’[28].

There are two DNA components in the particle, the first one is core DNA, the 146 bp DNA wrapped around a histone octamer, and the other one is called linker DNA, which is used to connect two nucleosomes together to form a 'beads on a string' fiber[29]. The histone octamer consists of two copies of four different core histones (H2A, H2B, H3, and H4), which have an affinity for DNA[29-31]. The next level of the organization is the 30 nm chromatin fiber. Besides those four histones mentioned above, the last member in the histone family is called H1, whose function is to bind to the linker DNA and sustain the formation of the 30 nm chromatin fiber[29,32]. The chromatin fiber is then further packed to form the final organization, a highly-ordered structure, around 700 nm metaphase chromosomes[33,34]. In general, chromatin is not a fixed structure, but a changeable DNA platform that can interact with environmental stimuli to regulate gene expression without any DNA mutation.

2.3 Histone modifications

The study of histone modifications roughly started in the early 1960s, after Vincent Allfrey's work about how histones are post-translationally modified[35]. Histone modifications play a role in distinct biological procedures, like transcriptional activation or repression, permissive euchromatin, DNA repair or meiosis[36-42]. To date, more than ten types of histone modification, including acetylation, methylation, phosphorylation, and ubiquitination[43-46]. There is an ever-growing number of histone modifications and they can be roughly divided into two groups, the first one is the group of histone modifications that has been found and studied so that their functional aspects are known; the other is the newly detected histone modifications with largely uncovered function in gene regulation[47]. So far, the best-understood histone modifications are

those on the histones N-terminal 'tail'. For example, the addition of trimethyl moieties to lysine 9 of histone H3 (H3K9me3) has been shown to increase the chromatin compaction, interfere with other histone modifications and finally suppress the expression of target gene. For acetylation, like acetylation of lysine 27 of histone 3 (H3K27ac), it is generally associated with open and available chromatin areas and viewed as an active enhancer mark[48].

2.4 ChIP-seq

Chromatin immunoprecipitation coupled with next generation sequencing (ChIP-seq) is a standard protocol to collect and examine the sequence of DNA that wrapped around a histone protein with the specific chemical modification or other protein-DNA binding cases. Using ChIP-seq, researchers can characterize the DNA sequences associated with various epigenetic markers from patient by comparing the sequencing data with reference genome database. In order to study the genome-wide occurrence of a specific histone modification, a particular antibody is used in the assay to capture the DNA that is wrapped around the histone protein with the target modification. ChIP-seq can be used to detect various types of epigenetic marks, including histone modification (mono-, di-, trimethylation, acetylation, and phosphorylation), histone variant (H3.3, H2A.Z, and H4.V)[49], and transcription factor (RNA polymerase II, CTCF, estrogen receptors alpha)[50-52].

2.4.1 Next generation sequencing (NGS)

Nucleic acid sequencing is a technique used to precisely identify the order of nucleotides in a certain length of DNA or RNA molecule. It has been almost two

decades since the discovery of the structure of DNA double helix. Sanger, Nicklen and Coulson first developed a dideoxy method and finished the first DNA sequence, which illustrated that sequencing information could provide deep view into the genome[53]. So, automated Sanger sequencing is considered as the first-generation sequencing[54]. With the help of Sanger sequencing, we finished the Human Genome Project (HGP), which started in 1990, took 13 years, spent over 3 billion US dollars. The completion of this work greatly inspired sequencing activities. However, there has been a fundamental alteration away from the operation of the first-generation sequencing in genome related research in recent years. Researchers were demanding a faster, cheaper, and more accurate sequencing technology.

Now, there are several commercially available next-generation sequencing (NGS) in the market, including 454 pyrosequencing from Roche Applied Science, solexa sequencing from Solexa, SOLiD chemistry from Applied Biosystems, and Hi-seq platform from Illumina. Among these technologies, the sequencing technology from Illumina is the most widely employed NGS technique throughout the world. The principle of illumina sequencing technology is the combination of cluster generation and sequencing by synthesis (SBS) technology, through which the technology can rapidly produce high throughput database[55]. The innovation and flexibility of this technology permits its application in the studies of genomics, transcriptomics, and epigenomics. So far, around 90% of the world's sequencing data is generated by illumina sequencing technology[56]. In order to conduct the sequence using this technology, the sample DNA needs to be end-repaired and ligated with specific adaptors on both end, 5' and 3' end. Then, through the denaturation of DNA by heating, single strand DNA with adaptor

could be immobilized and form a cluster on the surface of flow cell, which has been modified by adaptors with complementary DNA sequence. This ensures the nucleic acid chain would be present in a designed manner. Each immobilized nucleic acid chain could attach its free end (also ligated with an adaptor) with another cluster to form a bridge structure by hybridization. The immobilized nucleic acid chain would then be amplified through PCR reaction to generate abundant fluorescence signal later. After that, four fluorescentlylabeled nucleotides are introduced into the flow cell to sequence the tens of millions of clusters in parallel. In each round of sequencing, a single fluorescently labeled deoxynucleotide triphosphate (dNTP) would be added to the nucleic acid chain, followed by enzyme cleavage, and scanning the fluorescent dye signal to identify the type of the base.

Once the sequencing is completed, the sequencer could generate tens of millions of reads per sample. After the generation of sequencing data, these raw reads need to be firstly mapped to a known reference genome, e.g., hg19 or hg38 for human, mm9 or mm10 for mouse. To improve the mapping rate, raw data is usually trimmed with software to reduce bias. Whether the read can be uniquely aligned to the reference genome is the key factor in the alignment. After that, the DNA sequence of each read is replaced with the genomic coordinates and used in the downstream analysis.

2.4.2 Chromatin immunoprecipitation (ChIP)

In general, chromatin immunoprecipitation is a multi-step process, which includes cell/nuclei collection, immunoprecipitation, and DNA purification. The whole procedure of ChIP assay can be divided into six steps, harvest of cell/nuclei, cell/nuclei lysis, chromatin fragmentation, immunoprecipitation, protein digestion (using proteinase K),

and DNA purification. According to the method used to break chromatin, ChIP assay can be cataloged into two groups, cross-linked ChIP (XChIP)[57] and native ChIP (NChIP)[58]. For XChIP, the interactions between protein and DNA are firstly stabilized by introducing formaldehyde into cells. The addition of formaldehyde enables the formation of a covalent bond linking functional group between two different macromolecules, DNA and nearby protein[59]. Then, the stabilized chromatin can be sonicated into 200-700 bp chromatin fragments. For NChIP, the chromatin is directly digested by some special nucleases, micrococcal nuclease (MNase) or deoxyribonuclease I (DNase I), without any crosslinking. Through nuclease digestion, the chromatin can be fragmented to produce a ladder of DNA size, and the length of DNA is corresponding to the number of nucleosome cores (the length of DNA within one nucleosome core is around 170 bp). The target fragmented chromatin is then precipitated by the beads coated with a specific antibody because of the reaction between antigen and antibody. Enriched DNA is separated from protein by heating (reverse crosslinking) and proteinase K (digest attached protein), and then purified. The quality and quantity of purified target DNA can be characterized by polymerase chain reaction (PCR), real time PCR (qPCR), nanodrop spectrophotometer, and sequencing (ChIP-seq).

2.5 The state of the art of ChIP-seq

Traditionally, the protocol of ChIP-seq requires several millions of cells per assay, due to the low sensitiveness of the technology and the loss of material during the various steps, including cell/nuclei harvest and lysis, chromatin fragmentation, immunoprecipitation, DNA elution and purification, library preparation and sequencing. Recently, a number of advanced ChIP-seq techniques have been developed, like Nano-

ChIP (10,000 cells per assay)[60], LinDA (5,000 cells per assay)[61], iChIP (500–1,000 cells per assay)[9], MOWChIP-seq (100 cells per assay)[12,13], and surfaceChIP-seq (30–100 cells per assay)[11], that can be used to generate high-quality sequencing data from a small number of cell/nuclei. Drop-ChIP was reported that it enables the completion of the whole process of ChIP-seq at the single-cell level[62]. Furthermore, an alternative technology to ChIP, called CUT&RUN (requires 100 cells per assay), was demonstrated to map specific interaction between protein and DNA by digesting and releasing target protein-DNA complex into supernatant with the help of a special fusion nuclease[63-65].

Nano-ChIP, which was developed by Bradley E Bernstein's group, can be used for profiling histone modification features with 5,000 to 10,000 cells per assay. To minimize primer self-annealing, Nano-ChIP uses a primer with hairpin structure into the protocol. The hairpin structure here serves as a restriction site in the primers.

Single-tube linear DNA amplification (LinDA) was demonstrated to profile the histone modification (H3K4me3) with 10,000 cells, estrogen receptor-α (ERα) binding with 5000 cells and several transcription factors with 30 picogram of DNA because of its own high-efficiency amplification strategy. In LinDA, ChIPed DNA is first denatured to single strand oligonucleotide, and poly T (dT_{15}) end is ligated to the single nucleic acid chain and in vitro transcribed to RNA. Then, the new RNA strand is reverse transcribed by using RNA polymerase to its complementary DNA. The new produced DNA is amplified with the T7 promoter-BpmI-oligo(dA)15 primer. The additional steps for RNA transcription and reverse-transcription in the ChIP process enables LinDA to generate sequencing data with reasonably high quality from less cell.

Indexing-first ChIP (iChIP) was developed to ligate adaptor with various index to each sample. Here, the samples with different index are pooled together before ChIP process. The ChIPed DNA from indexed samples is used to generate sequencing library and then sequenced. Sequencing data is demultiplexed according to the ligated index sequence. Through this method, iChIP can yield a sensitivity of 500 cells per assay.

SurfaceChIP-seq, a microfluidic based ChIP process, relies on the antibody-coated channel surface for multiplexed and ultralow-input profiling of histone modifications. Because the specific histone-target antibody is immobilized on the surface, the manipulation of the immunomagnetic beads is not required in this technology. However, it requires users to have relatively high level of experimental skills, because of the chemical modification on the surface of glass slide, conjugation of single strand DNA to the antibody by covalent bond and immobilization of the antibody to the surface in a microfluidic chamber. Antibody is first conjugated with a certain length of single strand DNA. Then, the antibody-conjugated DNA is hybridized with a complimentary single nucleic acid chain and fixed on the surface of glass slide by the covalent bonding. The surface of microfluidic channel coated with high density of antibody enables this technology to be used to produce high quality of sequencing data from as few as 30 cells.

Drop-seq, which can be used to generated ChIP-seq data at single cell level, was published in 2015 on Nature Biotechnology by David A Weitz and Bradley E Bernstein's groups. The principle of this technology is to conduct the whole process of ChIP in droplets, starting from cell lysis and digestion. The enriched chromatins is attached with

unique DNA barcode in order to demultiplex the sequencing data later. Conducting ChIP assay in droplets enables the barcoding of each single cell, however, Drop-ChIP can only generate around 1,000 unique reads per cell, and the coverage of the genome is very sparse[62].

CUT&RUN was developed by Steven Henikoff's group. It is different from traditional ChIP-seq assay. This technique produces data with high signal-to-noise ratio and reasonable correlation between each replicates. However, compared with histone modification data generated by low-input ChIP-seq technologies, the data produced by CUT&RUN showed weak correlation with ChIP-seq data generated from conventional approaches (e.g., ENCODE data). The correlation also does not improve with increasing cell numbers (in the range of 100–6,000 cells per assay). Nevertheless, this IP-free method still provides powerful insights into epigenomic states for researchers.

MOWChIP-seq was developed in our lab and could be used to detect histone makers (H3K4me3, H3K27me3, H3K27ac, H3K9me3, etc.) from as little as 100 cells. This technology will be described in chapter 3.

2.6 Bioinformatics

After DNA sequencing of the library, tens of millions of certain length of short reads were generated by the sequencer, which is called raw data. Each read is the specific order of the four chemical blocks (A, T, C, G) that make up of parts of DNA strand. However, the short reads themselves cannot provide much information. Computational or statistical tools need to be applied to the sequenced results. With the help of these tools, we can obtain the large amount of information and increase our

understanding of various biological processes. This procedure of data analysis is called bioinformatics. In my thesis, the analysis can be further separated into two parts based on the type of sequencing data, one is ChIP-seq data analysis[66], and the other is RNA-seq data analysis[67].

2.6.1 ChIP-seq data analysis

The sequenced results (those short reads in certain length) are firstly mapped to a reference genome using software, like Bowtie2[68] or BWA[69]. The mapped reads were then stored in output files whose format could be BAM or SAM. Duplicated reads because of PCR bias during library preparation should be removed after reads alignment. Those reads that could be mapped to multiple positions should also be filtered. The next step is so-called peak calling step using tools such as MACS2[70]. Peaks are the genomic positions that are significantly enriched with mapped reads as compared to the sequenced data from background DNA. The positions of peaks represent the locations where a specific histone modification happened. By comparing the peak intensity between two conditions using differential analysis tools (DESeq2 or edgeR)[71,72], we could identify those differential peaks based on their modification levels, and further annotate the peaks to their controlled genes to investigate if there are any enriched meaningful biological pathways involved in.

Visualization of the ChIP-seq data is also an important step in the analysis. It could effectively help us to assess and analyze the sequenced data. Now, there are several interactive visualization software have been developed in the field, such as Integrated Genome Viewer (IGV)[73]. Overall, because of these tools, we were able to study the role of epigenome in many aspects, such as the development of mental disorders.

2.6.2 Transcriptomic data analysis

RNA-seq enables us to test and examine the transcriptome, which links the genetic information with its functional protein expression[74]. Different from ChIP-seq which is done by capturing the target wrapped DNA, RNA-seq is conducted by extracting messenger RNA (mRNA) and converting them into their complimentary DNA (cDNA). cDNA is then fragmented and ligated with adaptors. The ligated cDNA fragments were prepared as sequencing library after amplification, size selection and clean-up. Once the library is prepared, it is then analyzed by next generation sequencing. The pre-processing steps for sequenced results from RNA-seq is similar to the steps for ChIP-seq. However, the mapping tools used for RNA-seq sequenced reads are different. Instead of aligning short reads to genomic locations, tools for RNA-seq results, like STAR[75] or HISAT2[76], directly associate the reads with pre-indexed transcripts. The results from these methods could be accurate for the high-expressed transcripts, but the performance of these tools for low-expressed genes is worse than the accuracy of tools used for ChIP-seq analysis. Because of that, those transcripts with low read counts should be removed before differential analysis.

3 MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications

3.1 Project Summary

Epigenetic mechanisms such as histone modifications play critical roles in adaptive tuning of chromatin structures. Profiling of various histone modifications at the genome scale using tissues from animal and human samples is an important step for functional studies of epigenomes and epigenomics-based precision medicine. Because the profile of a histone mark is highly specific to a cell type, cell isolation from tissues is often necessary to generate a homogeneous cell population, and such operations tend to yield a low number of cells. In addition, high-throughput processing is often desirable because of the multiplexity of histone marks of interest and the large quantity of samples in a hospital setting. Here, we present the latest version of microfluidic oscillatory washing-based chromatin immunoprecipitation followed by sequencing (MOWChIP-seq) for profiling of histone modifications using as few as 100 cells per assay with a throughput as high as eight assays in one run.

3.2 Introduction

Besides genomic variations, epigenomic mechanisms also play an important role in regulation of gene activities and phenotypes[77,78]. Epigenomic mechanisms consist of several epigenetic modifications, like histone modifications, DNA methylation, and non-coding RNA regulation. It has been demonstrated that abnormal epigenetic modifications are involved in development of many human diseases, from

neurodevelopmental disorders to cancer[18]. Researchers could have a better understanding of epigenomic features of various diseases by comparing disease and reference epigenomes, and all these knowledge might potentially provide us some new therapeutic strategies. The US NIH Roadmap Epigenomic Mapping Consortium are engaging in building reference epigenomes for different types of tissue or cell under natural development and in disease condition[79]. More groups are currently conducting similar projects. Once the epigenomic features of references and various diseases are established, personalized treatment to each patient can really come true.

However, these efforts in uncovering epigenomic features of reference and diseases are often restricted by the small size of the available tissue samples. The limitation of sample size is always combined with the fact that many epigenetic modifications may need to be tested, especially for histone modifications. Among over 100 recognizable histone modifications, there are tens of histone modifications that are examined commonly[80]. Precise statement of chromatin conditions can only be predicted with the combination of advanced computational algorithms and exhaustive large-scale sequencing data sets on different histone modifications. Because of the limited quantity of tissue samples generally available for examination and lots of histone marks which are commonly interesting, it is demanding to have low-input and multiplexed technologies for profiling histone modifications.

The most common method for mapping genome-wide histone modification and protein binding is chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq)[8]. However, due to the poor efficiency of collection of target chromatin, millions of cells are required for each traditional ChIP-seq assay. To improve

the sensitivity of chromatin immunoprecipitation assay, our lab developed a low input microfluidic-based ChIP-seq protocol, microfluidic-oscillatory-washing-based ChIP-seq (MOWChIP-seq) in 2015[12]. For MOWChIP-seq, it increased the collection efficiency of specific DNA region by introducing a packed bed of antibody coated beads into a microfluidic chamber. Compared with other competing advanced ChIP-seq protocols, MOWChIP-seq provides several unique advantages. First of all, MOWChIP-seq can produce high quality ChIP-seq data, which rival those generated by conventional assay from tens of millions of cells per assay, from as few as 100 cells. For example, 10,000–30,000 peaks can be recognized from the profiling of H3K4me3 produced by MOWChIP-seq and 2–5 million unique reads can be generated when 100–1,000 GM12878 cells per assay are used. The sequencing results also show relatively high robustness of the technology among replicates, with Pearson's correlation coefficients larger than 0.96 for H3K4me3 and larger than 0.90 for H3K27ac from 1,000 cells per assay. Thus, MOWChIP-seq makes the generation of sequencing data using limited cell samples possible for use as reference epigenomes and patient profile. However, it is not multiplexed. The users can only conduct one ChIP assay each time with the current setup. This makes the experiment with many tissue samples very time-consuming. For example, an experienced user needs to spend over one and a half years to profile two histone modifications from 120 different samples with two replicates.

Here, we demonstrate an improved MOWChIP protocol, as shown in Fig. 3-1 for generating 8 single MOWChIP data sets simultaneously using as few as 1000 nuclei. In our protocol, we first designed and built a new pneumatic control system. This new system enables us to control each unit of MOWChIP with a separate pressure source,

both in valve control part and oscillatory washing part. Then, we switched the method used to cut chromatin from sonication to using specific nuclease, micrococcal nuclease (MNase). This adjustment enables us to apply MOWChIP to profile various genome-wide histone modifications, even some hard-to-profile histone marks, like H3K27me3. Lastly, we validated this modified MOWChIP protocol using 1000 nuclei isolated from mouse PFC for multiplexed profiling of different histone modifications. These ChIP-seq data produced 20K peaks and 13 million unique reads on various histone marks, like H3K27me3, H3K9me3, etc.

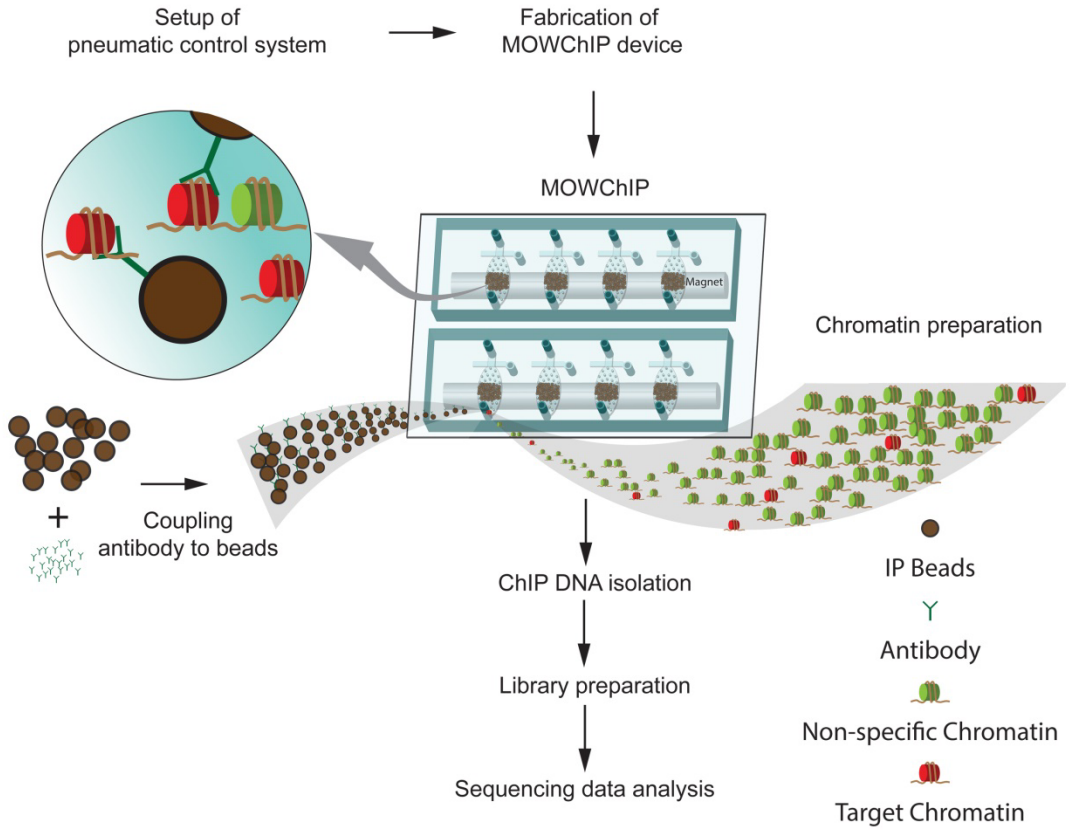

Figure 3-1 A schematic description of the MOWChIP-seq protocol. The whole process, starts from the setup of the pressure control system and ends at sequencing data

analysis, takes around 6 days, which excludes the processing time for receiving sequencing data back.

3.3 Materials and Methods

Design and setup of the pneumatic control system

A programmable pneumatic control system was created for operation of eight sieve valves (a monolithic pneumatic microvalve on the MOWChIP device) and the microfluidic oscillatory washing. The electrical and mechanical schematics shown in Fig. 3-2 were designed for pneumatic control system of the MOWChIP protocol. The design of the control system is similar to what is commonly built for the operation of pneumatic microvalves (also known as Quake-style valves). The electrical connection schematics (Fig. 3-2a) show the system consists of a computer, a data acquisition (DAQ) card (National Instruments, SCB-68), relays, solenoid valves, and a number of wires. A LabVIEW program was written in the computer and used to define the input and output signal of the DAQ card. The mechanical part of the system (Fig. 3-2b) demonstrates the connections among pressure gauges, pressure regulators, solenoid valves, and microfluidic devices. With the combination of 24 solenoid valves, 9 pressure gauges, 9 pressure regulators, and a LabVIEW program, we are able to run 8 MOWChIP experiments in parallel.

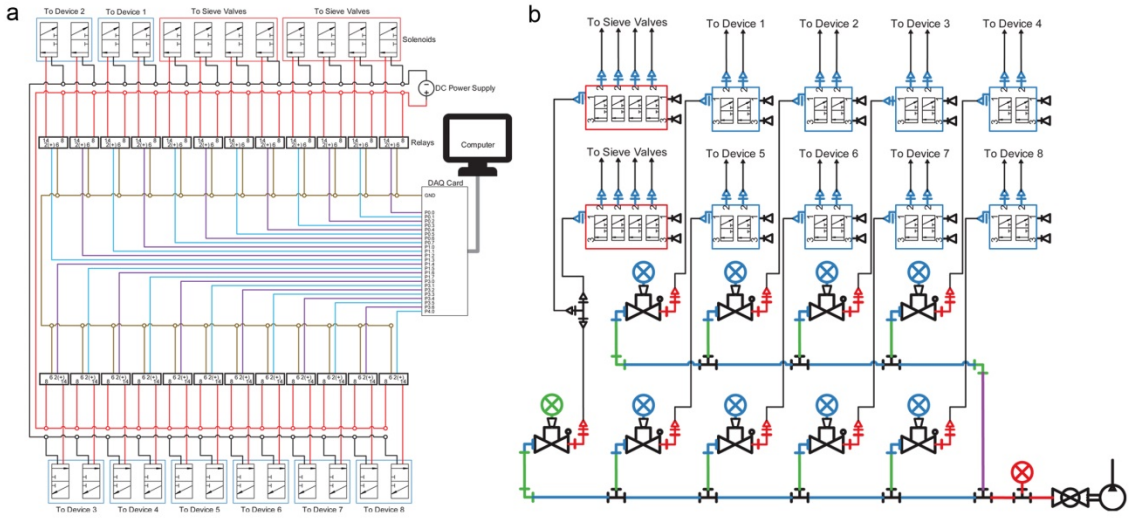

Figure 3-2 Schematics for electrical and pressure control systems associated with 8-unit MOWChIP devices. (a) Electrical connections for automating 24 solenoid valves associated with an 8-unit device. (b) Connection of various parts that constitute the pressure control system for an 8-unit device

Fabrication of the microfluidic devices

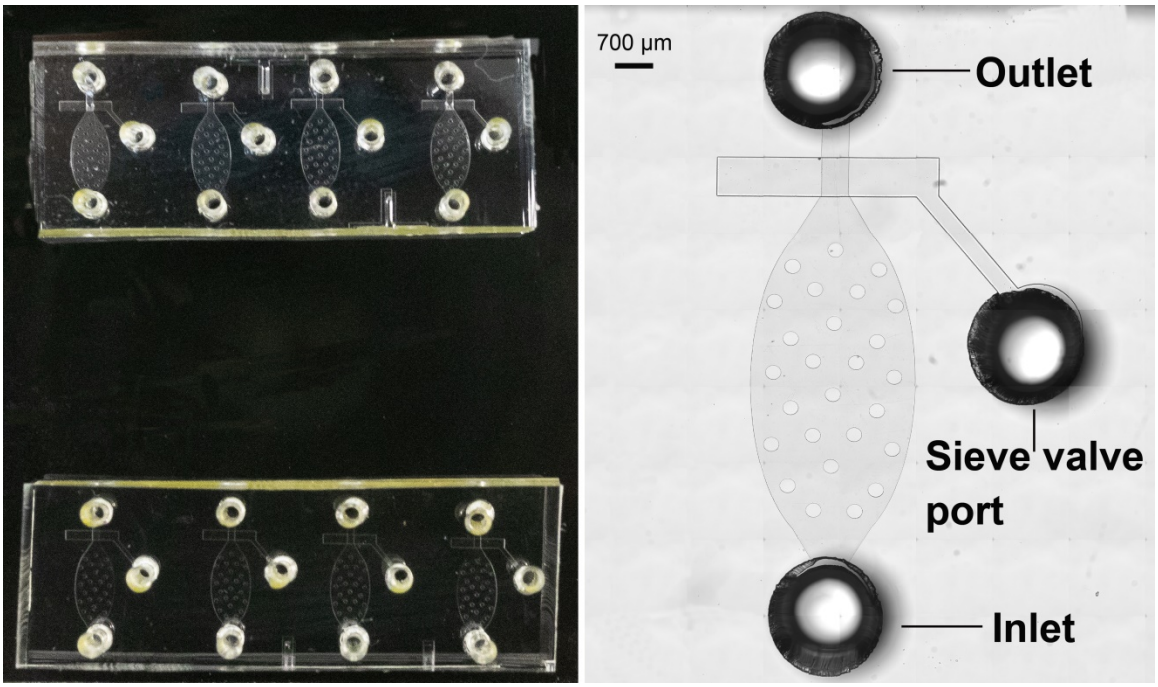

Figure 3-3 The 8 units device. Picture (left) and microscopic image (right) of an eight-unit MOWChIP device. In the microscopic image, the details of one unit are shown, including the inlet, outlet, and the sieve valve port of the device.

The two-layer microfluidic MOWChIP devices were made of polydimethylsiloxane (PDMS). The fabrication of the devices followed multilayer soft lithograph with minor optimization. First, two photomasks were prepared with the microscale patterns drawn by software Layout editor and printed on high-resolution transparencies (5080 dpi). Then the masters for fluidic layer (~40 μm) and for control layer (~60 μm) were made of photoresist SU-8 2025 on a 3-inch silicon wafer. After that, 36 g of PDMS mixture (Momentive, cat. no. RTV615) at ratio of 5:1 (reagent A: reagent B) was poured into a Petri dish that contains the fluidic layer master to form fluidic layer part of the devices (~6 mm). Similarly, 5 g of PDMS mixture at ratio of 20:1 was poured into another Petri dish contains the control layer master. Both of the Petri dishes with PDMS mixture were

cured at 80 °C for 12 min. Fluidic layer was peeled off and aligned to control layer according to the designed align-mark. The two-layer PDMS slab was baked at 80 °C for another 2 hours. The slab was then cut into several pieces, each piece contains 4 units. Three holes were punch for every unit, and the holes served as inlet, outlet, and sieve valve port later. Finally, all the PDMS pieces, in company with pre-cleaned glass slide, were treated by oxygen plasma (in a plasma cleaner, Harrick Plasma) and bound with treated glass slide to form MOWChIP chips (Fig. 3-3).

Functionalization of magnetic beads with an antibody

5 µl of Dynabeads Protein A were washed twice with ice-cold IP buffer (20 mM Tris-HCl (pH 8.0), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1%(w/v) sodium doxycholate, 0.1% SDS, and 1%(v/v) Triton-100X in Milli-Q water). Then, the beads were resuspended in 150 µl of IP buffer. Specific amount of antibody (from Active Motif) against a target histone modification (e.g., H3K4me3, H3K27ac, H3K9me3, etc.) was added into the bead suspension. The optimized amount of antibody against specific histone modification was determined by the ChIP-qPCR results. The mixture was incubated at 4 °C overnight on a rotator at 24 r.p.m. Beads coupled with antibody were washed three times with ice-cold IP buffer prior to be used in MOWChIP assays.

Chromatin fragmentation

We used nuclei isolated from frozen mouse brain tissue as example. The extraction of nuclei follows the protocol prepared by Dr. John M. Graham with minor modifications. Then 0.25 µl of protease inhibitor cocktail (PIC) and 0.25 µl of 100 mM Phenylmethanesulfonyl fluoride (PMSF) were freshly added into 25 µl of nuclei

suspension which contains 100-10,000 nuclei. The suspension was then mixed with 25 µl of 2x lysis buffer (100 mM tris (pH 8.0), 100 mM NaCl, 30 mM MgCl2, and 4% Triton X-100). The mixture was mixed by vortex, and then incubated at room temperature for 10 min. 2.5 µl of 100 mM $CaCl_2$ and 5.5 µl of 10 U/ µl MNase were added into the mixture, mixed well, and incubated at room temperature for 10 min. 5.5µl of 0.5 M EDTA was added into the mixture, mixed well, and incubated on ice for 10 min. The mixture was centrifuged at 16,100 r.c.f. for 5 min. The supernatant (~55 µl) was then transferred to a new 1.5 ml Eppendorf microcentrifuge tube and placed on ice. The supernatant was set aside on ice and used later.

MOWChIP

Our MOWChIP procedure consists of a number of steps (Fig. 3-4, four steps of MOWChIP). First, the microfluidic chambers of the device were filled up with IP buffer with the help of syringe pump at a flow rate of 200 µl/min. Then, the antibody coupled magnetic beads were flowed into the device via the combined force provided by pipette and a cylindrical permanent magnet. After plugging the long tubing assembly (10 cm C-Flex clear tubing connected with 3 cm PFA tubing) containing chromatin sample into the inlet of the device, the sieve valve was partially closed by turn on the pressure control system, which applied 30 psi on the valve. The magnetic beads were manipulated by the magnet and packed against the valve to form a packed bed while injecting the chromatin solution into the device at a flow rate of 1.5 µl/min. Under this flow rate, the immunoprecipitation (IP) part was completed in ~50 min. After that, the long tubing

assembly was disconnected from the device.

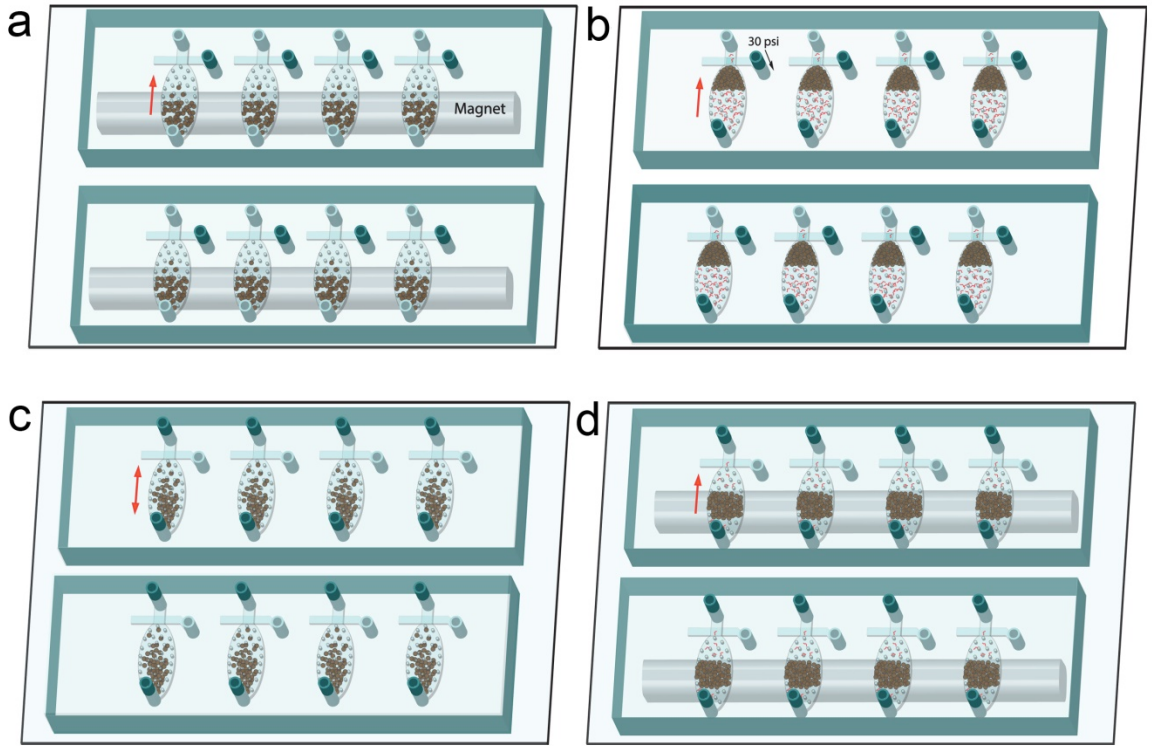

Figure 3-4 Schematic figures for the operation of the devices. (a) Loading of immunomagnetic beads (IP beads) into the chamber with the assistance of a magnet. The sieve valves are open during this step. (b) Flowing of chromatin fragments through the packed bed of IP beads. The sieve valves are closed under 30 p.s.i. pressure. (c) Oscillatory washing under alternating pressure pulses applied via the inlet and outlet. Low salt washing buffer and high salt washing buffer are used to remove non-specific bound chromatin. (d) Flushing out unbound chromatin and debris while retaining the beads with magnet. The single-direction red arrows indicate the liquid flow direction, and the bi-directional red arrow indicates that there is oscillatory movement of the liquid flow in the microchamber.

Oscillatory washing was conducted following the completion of IP. Low salt washing buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X)) was added into two short tubing assembly (6 cm C-Flex clear tubing connected with 3 cm PFA tubing), and then the tubing assemblies were plugged into the inlet and outlet of the microfluidic device unit. The PFA tubing of pressure control system was plugged into the other ends of the short tubing assemblies. Oscillatory washing parameters were set in a pre-written LabVIEW program. Both duration and interval were set as 0.5 s for the alternating pressure pulses operated by the two solenoid valves attached to the same device (In the LabVIEW program, enter 0.5 for both “inlet duration” and “outlet duration”). The number of oscillations was 300, which generated a total washing time of 5 min. The pressure was turned to ~0.65 psi by twisting the knob of pressure regulator. After finishing the washing with low salt washing buffer, low salt washing buffer was replaced by high salt washing buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X), and the process of washing was conducted again with the same washing parameter. The magnetic IP beads were retained in the microfluidic chamber during oscillatory movement by using permanent magnet. After the accomplishment of oscillatory washing (low salt and high salt washing buffer), non-specifically adsorbed chromatin fragments were removed from the IP beads surface. Magnetic IP beads were then eluted from the microfluidic chamber to a new 1.5 ml Eppendorf microcentrifuge tube with IP buffer at a flow rate of 200 µl/min. The microcentrifuge tube was then placed on a magnetic rack to collect beads. Supernatant was removed and discarded by pipette. 90 µl of elution buffer (10 mM tris-HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA, and 0.03% (wt/v) SDS)

and 4 µl of proteinase K (1 mg/ml) were added into the tube to resuspend the beads pellet. Fig 3-5 shows the whole experimental setup for 8-unit MOWChIP operation. The tube was incubated in a multi-therm heat/cool shaker incubator at 65 °C for 2 hours. Phenol chloroform cleanup and ethanol precipitation were performed after incubation to extract and purify ChIP DNA. Finally, ChIP was resuspended in 10 µl of low EDTA TE buffer.

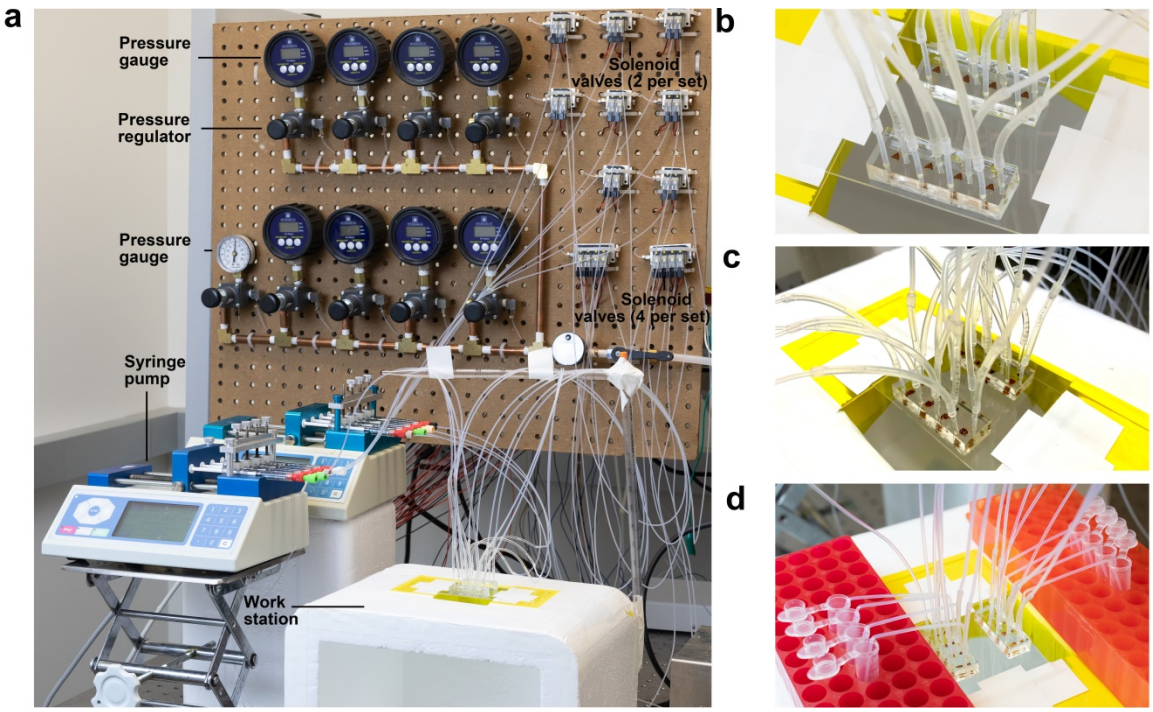

Figure 3-5 The experimental setup for MOWChIP operation. (a) An 8-unit MOWChIP device connected to its ancillary fluid and pressure control system. (b) Flow of chromatin samples through the packed IP bed. (c) Oscillatory washing of the beads. (d) Elution of the beads into tubes after oscillatory washing.

ChIP DNA quality check

The quality of ChIP DNA was firstly estimated by measuring the size of fragmentized chromatin with high sensitivity DNA analysis kit (Agilent Technologies, cat. no. 5067-5583) in a Tapestation (Agilent Technologies, model no. G2962/A). Then, a further examination for quality check was conducted by real-time PCR before sequencing. The reaction was conducted with iQ SYBRGreen Master Mix (Bio-rad, catlog no. 1708882) in wells of 96-well plate (Bio-Rad, cat. no. HSP9601) on a CFX Connect™ real-time PCR Detection System (Bio-Rad, cat. no. 1855200). The 20 µl reaction mixture is consists of 10 µl of SYBRGreen, 1.4 µl of primer pairs, and 8.6 µl of DNA template. All the sequencing information about primer pairs used for checking enrichment of ChIP DNA is provided in Table 3.1-1. All qPCR assays were performed using following thermal cycling profile: 95 °C for 2 min followed by 45 cycles of (95 °C for 15 s, 58°C for 20 s, and 72 °C for 20 s). The relative fold enrichment, which is the ratio of percent input between a positive locus and a negative locus, was calculated. The percent input for a specific locus was calculated using the following equation:

$$Percent\ input = 2\text{^}(Ct_{input} - \frac{\log\,(DF)}{log2} - Ct_{ChIP}) * 100\%$$

where Ct_{ChIP} and Ct_{input} are the Ct value of ChIP DNA and input, respectively. DF is the dilution factor, which is defined as (sample volume of input + sample volume of ChIP DNA) / sample volume of input. The relative fold enrichment was calculated using the following equation:

$$Enrichment = \frac{Percent\ input(positive\ locus)}{Percent\ input(negative\ locus)}$$

Species	Histone mark	Positive (P) or negative (N) locus	Locus name	Primer

Mouse	H3K27me3/H3K9me3	P	Hoxc10	Active Motif, Cat. No. 71019
		N	Neg Set 3	Active Motif, Cat. No. 71013
	H3K36me3/H3K79me2	P	Actb-1	Active Motif, Cat. No. 71015
		N	Neg Set 1	Active Motif, Cat. No. 71011

Table 3-1 Primers used in ChIP-qPCR for quality control

Library preparation and sequencing

All the ChIP-seq libraries were prepared using an Accel-NGS 2S Plus DNA Library kit (Swift Biosciences, cat. no. 21024). Library preparation started from 8 µl of ChIP DNA in low EDTA TE buffer. We followed the kit's manual with a little modification as described below. First, we used one-fifth of the volume of each reagent recommended in the instruction. And then, 2.5 µl of 20x EvaGreen dye was added into the amplification reagents MasterMix to monitor and quantify PCR amplification, the amplification reaction was stopped when the intensity of fluorescence increased ~3000 RFU. Finally, purified ChIP DNA library was resuspended in 7 µl low EDTA TE buffer.

The size of each ChIP DNA library was measured by using a High Sensitivity DNA Analysis kit (Agilent) in a TapeStation (Agilent) again (Fig. 3-9). The concentration of each library was quantified through a KAPA Library Quantification kit (Kapa Biosystems, cat. no. KK4824). All the ChIP DNA libraries were pooled at 10 nM for sequencing by Illumina HiSeq 4000 with single-end 50-nt read.

Sequencing data analysis

Some software for data analysis was installed at the beginning, like Trim Galore!, Bowtie2, fetchChromSizes, samtools, bedtools, MACS2 and SICER for trimming, mapping, and peak calling. All the sequencing reads were trimmed using Trim Galore! with default parameters. Then, the trimmed reads were mapped to a reference genome,

hg19 or mm9, by using Bowtie2. The aligned reads were processed for peak calling by using MACS2 (q-value < 0.05) for narrow peaks and using SICER with parameters (–windowSize 1000 –gaps-allowed 3) for broad peaks. Uniquely aligned reads (from both ChIP and input DNA datasets) were used to calculate a signal for each 100nt bin across the whole genome. Normalize the ChIP signal in each 100nt bin using the equation described below:

$$Signal_{100nt} = \left(\frac{ChIP\ Signal}{No.Unique\ ChIP\ Reads} - \frac{Input\ Signal}{No.Unique\ Input\ Reads}\right) * 1{,}000{,}000$$

Normalized signal in each bin across the whole genome can be used to calculate Pearson correlation coefficient later.

3.4 Results and Discussion

Among all the parameters in the protocol, we reasoned that the amount of antibody used to functionalize magnetic beads is the most crucial factor in the ChIP assay. Then, we optimized this parameter in MOWChIP protocol for various histone marks by criticizing the relative fold enrichment. The relative fold enrichment, which was calculated by normalizing the percent input at a specific locus against that at a chosen negative locus, tells the quality of ChIP DNA. As shown in Fig. 3-6, we introduced different amount of antibody into the magnetic beads suspension to vary the antibody coating condition. Using 2000 mouse prefrontal cortex nuclei, we demonstrated the performance of the 8-unit parallel MOWChIP device. We firstly used real-time PCR to quantify the percent input of ChIP DNA collected from nuclei at specific positive and negative loci for H3K27me3, H3K9me3, H3K36me3, and H3K79me2. ChIP assay started from MNase fragmented chromatin isolated from 2000 mouse prefrontal cortex

nuclei. The percent input was calculated based on the equation above. Error bars were generated from two or four ChIP technical replicates. The relative fold enrichment in Fig. 3-6 endorsed that the percent inputs were considerably higher on specific positive loci than specific negative loci. This shown that the ChIP DNA collected by our modified MOWChIP system faithfully reflects the histone modifications along various genomic loci. The qPCR results shown that the relative fold enrichment was the highest for H3K27me3 when the amount of antibody (anti-H3K27me3) was 0.4 μg; for H3K9me3 when the amount of antibody (anti-H3K9me3) was 0.6 μg, for H3K36me3 when the amount of antibody (anti-H3K36me3) was 0.25 μg, and for H3K79me2 when the amount of antibody (anti-H3K79me2) was 1 μg.

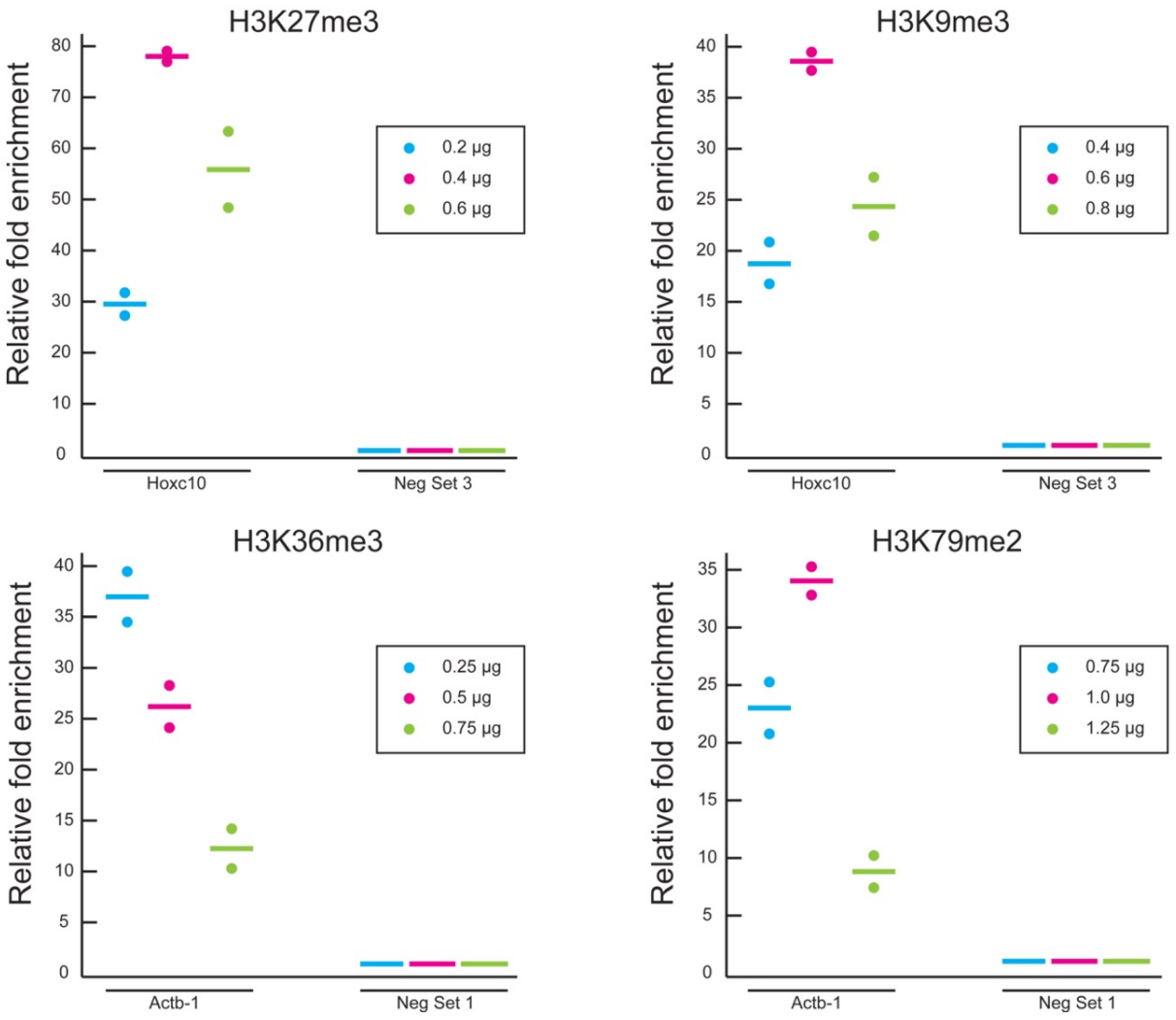

Figure 3-6 Optimization of antibody quantity using MOWChIP-qPCR for four histone marks. For each histone mark, 3 different quantities of antibody were used during IP bead coating and enrichment of ChIP DNA at positive (Hoxc10 for H3K27me3 and H3K9me3; Actb-1 for H3K36me3 and H3K79me2) and negative (Neg Set 3 for H3K27me3 and H3K9me3; Neg Set 1 for H3K36me3 and H3K79me2) loci. Enrichment was examined by qPCR under each condition. All experiments were conducted in duplicate and the horizontal lines represent the mean. The relative fold enrichment was normalized against that of Neg Set 3 or Neg Set 1.

We then successfully prepared sequencing libraries using Accel-NGS 2S Plus DNA Library kit. The library preparation started from ChIP DNA collected by our MOWChIP system. This ensured that ChIP DNA with good relative fold enrichment produced by our microfluidic device can be used for ChIP-seq project. The fragment size of ChIP DNA and ChIP DNA library were measured by using TapeStation and shown in Fig. 3-7. Two peaks named as ‘Lower’ and ‘Upper’ represent the location of 35 bp and 1000 bp respectively. The area of these two peaks also serves as a criterion for roughly quantifying the concentration of DNA at any range of size. The reason why DNA samples, both ChIP DNA and ChIP DNA library, have several peaks is because we used MNase to fragmentize chromatin. MNase will randomly cut chromosome into mononuclesome, di-nuclesomes, or tri-nuclesomes. The size of DNA in a

mononuclesome is around 170 bp. The average size of ChIP DNA library is around 370 bp at the range of 200 bp to 1000 bp.

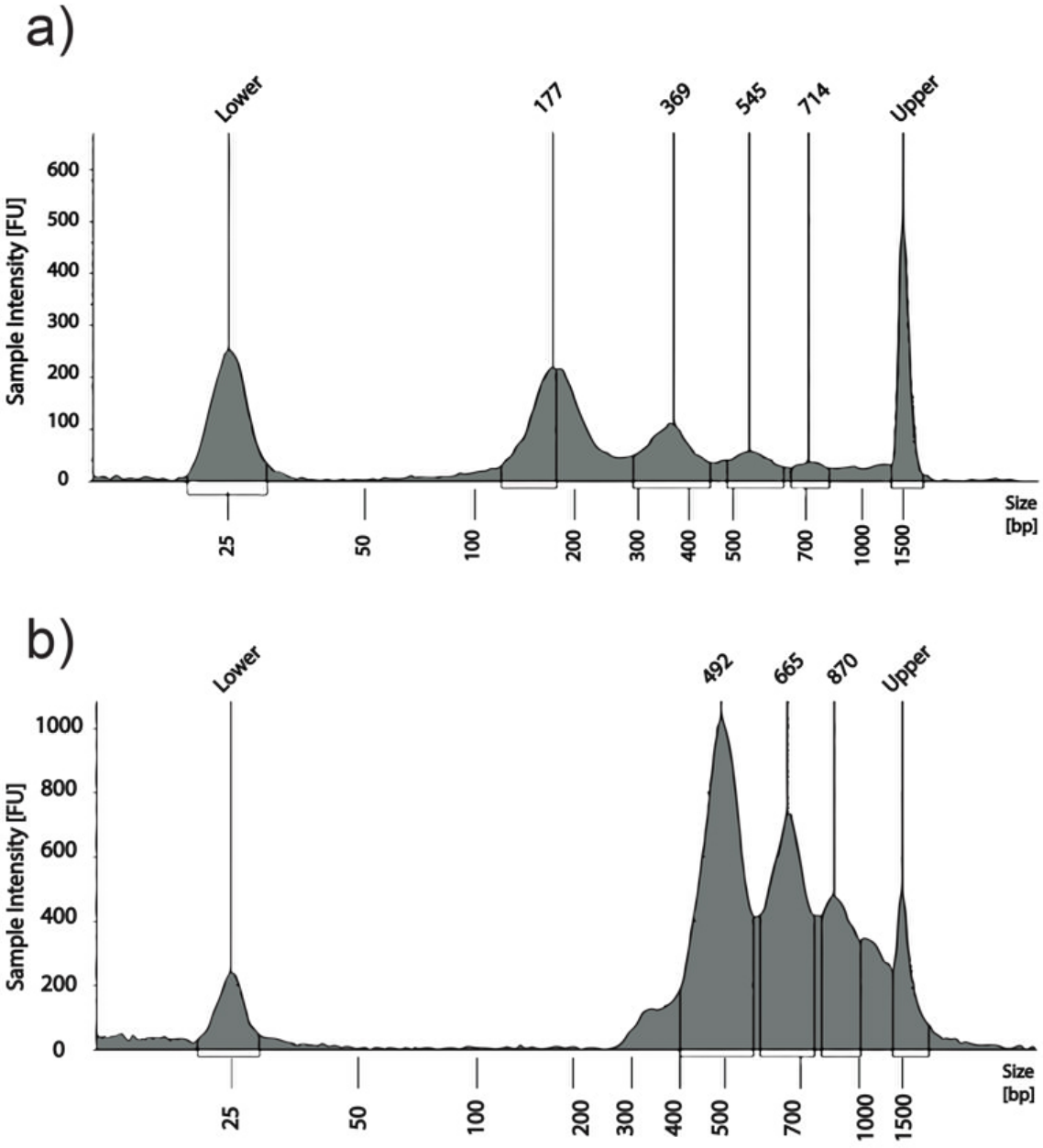

Figure 3-7 DNA fragment size distributions measured by a Tapestation. (a) DNA from 2000 nuclei isolated from mouse prefrontal cortex after MNase digestion. (b) ChIP DNA

sequencing library after H3K27me3 MOWChIP of 2000 mouse prefrontal cortex nuclei and library preparation.

After the completion of sequencing library preparation for each ChIP DNA sample, we quantified the concentration of DNA using KAPA library quantification kit. Following the instruction of the kit, 1 µl of DNA sample was diluted in three serial 1:100 dilutions and mixed with a master mix in a well of 96-well plate. The plate was then placed in a real-time PCR detection system. We performed qPCR using following thermal cycling profile: 95 °C for 5 min followed by 35 cycles of (95 °C for 30 s, and 60 °C for 45 s). The qPCR results, as shown in Fig. 3-8, were used to calculate the concentration of DNA. Finally, we pooled the DNA libraries together and sent them for sequencing.

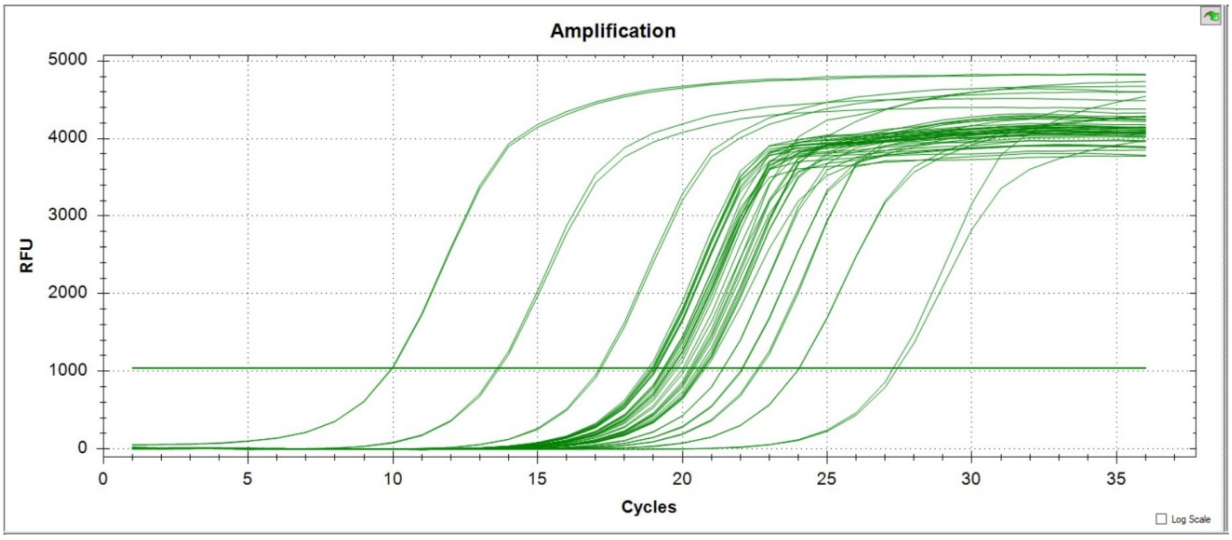

Figure 3-8 Real-time PCR results using KAPA library quantification kit.

ChIP DNA libraries were sequenced by Illumina HiSeq 4000 with single-end 50-bp read at University of Chicago. Once we got our sequencing data back, quality control (QC) analysis was conducted using software called Fastqc. Results were shown in Fig. 3-9.

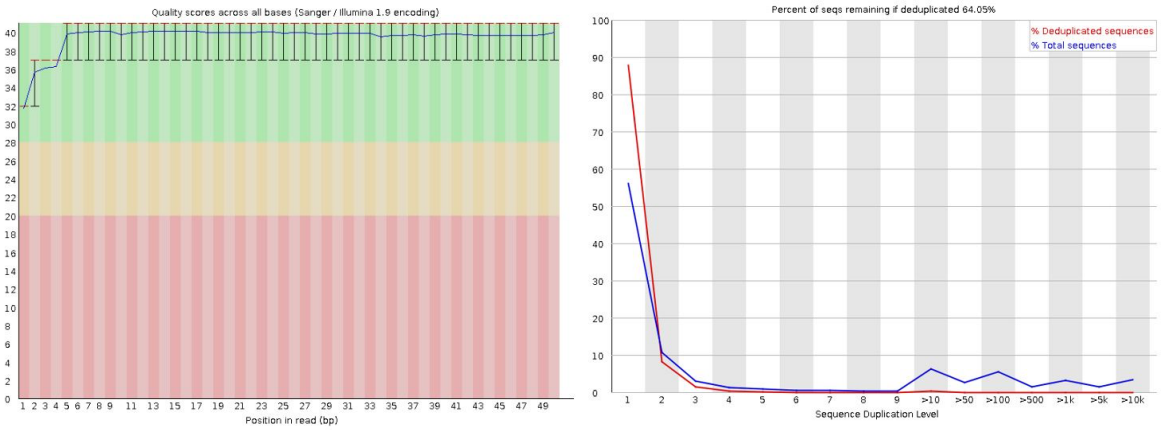

Figure 3-9 QC results from Fastqc. (left) Quality scores for all bases in reads. The blue line in the graph represents the average score for all reads in the dataset. Green region means all bases in it have good quality. (Right) Calculated deduplicated rate.

After QC, the datasets were processed with our analysis workflow. As shown in Table 3-2, we sequenced four technical replicates of ChIP DNA on H3K27me3, two replicates on H3K9me3, two replicates on H3K36me3, and two replicates on H3K79me2. Around 20 million reads were obtained for each sample. Sequencing data was firstly trimmed by trim galore! and aligned to mouse reference genome (mm9) using Bowtie2. Any read shorter than 20 bp will be removed from the dataset. Then, we used SICER to call peaks from mapped reads, and the average number of peaks for each dataset is 21K peaks for H3K27me3, 26K for H3K9me3, 22K for H3K36me3, and 22K for H3K79me2.

	Total reads	Trimmed reads	Aligned reads	Mapping rate	Unique reads	Peaks
H3K27me3_1	41013013	40930986	39748081	97.11%	22257112	20203
H3K27me3_2	25218638	25168200	24221876	96.24%	11470744	19232
H3K27me3_3	29207303	29148888	27921720	95.79%	11548715	18447
H3K27me3_4	27037435	26983360	24325499	90.15%	8999157	25226
H3K9me3_1	29399612	29340812	27853233	94.93%	15285575	24184
H3K9me3_2	17584707	17549537	16896694	96.28%	10662153	29231
H3K36me3_1	10839438	10817759	10485654	96.93%	8182766	22057
H3K36me3_2	26401662	26348858	25489885	96.74%	18963740	23774

H3K79me2_1	19466982	19428048	18378933	94.60%	8794087	23185
H3K79me2_2	27057229	27007719	26313621	97.43%	21291276	22541

Table 3-2 Summary of sequencing results of ChIP-seq assays with H3K27me3, H3K9me3, H3K36me3, and H3K79me2.

The mapped reads on each histone mark, H3K27me3, H3K9me3, H3K36me3, and H3K79me2, were visualized in Integrative Genomic Viewer (IGV) and shown in Fig. 3-10a. Reference genome at certain region was also shown in the graph. We also calculated the normalized sequencing data from both whole genome-wide and promoter regions, which is defined as +/- 2kb around transcription start sites (TSS). In Fig 3-10b, the ChIP DNA collected by using MOWChIP system was substantially correlated among their replicates, Pearson correlation coefficient > 0.87. Also, as shown in Fig 3-10b, the

cross correlation of MOWChIP-seq sample on different histone mark was relative low Pearson correlation, which is also as expected.

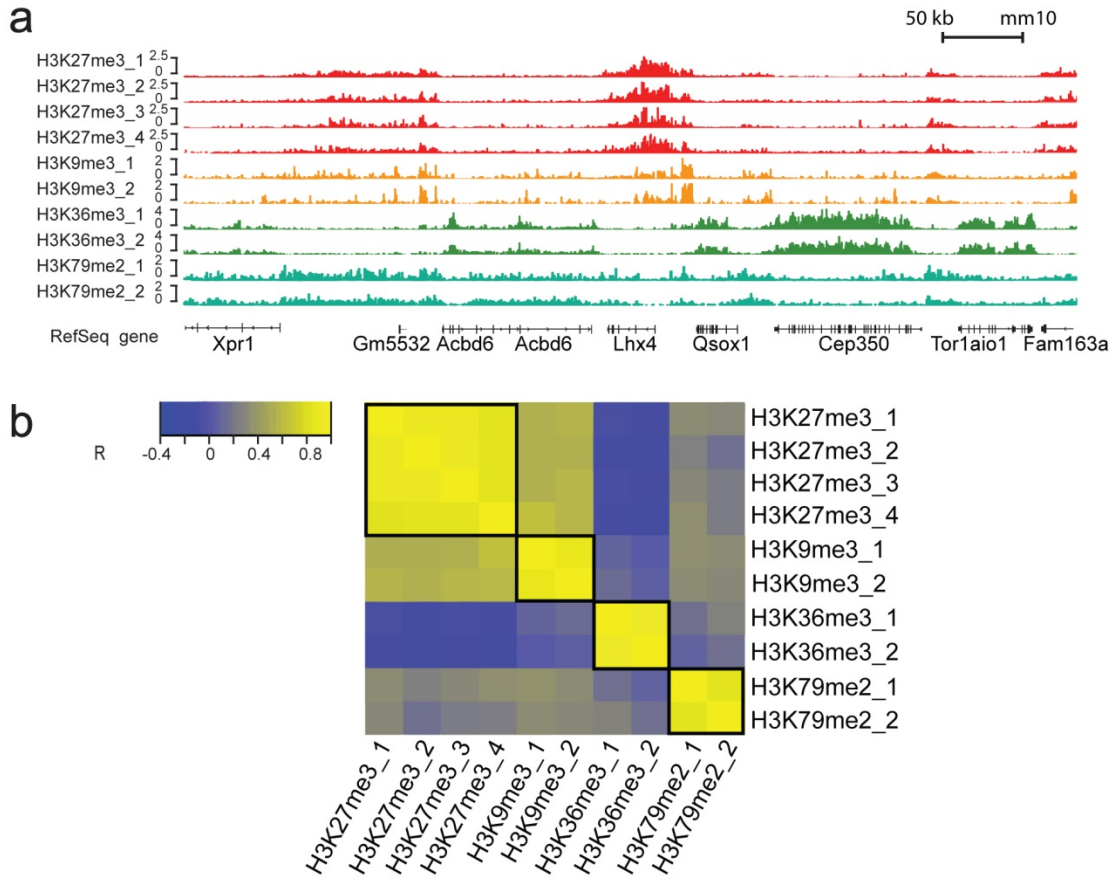

Figure 3-10 MOWChIP-seq data for H3K27me3, H3K9me3, H3K36me3 and H3K79me2 generated using 1,000 nuclei from mouse prefrontal cortex per assay. H3K27me3 data (four replicates) were produced by four units of an eight-unit MOWChIP device, whereas H3K9me3, H3K36me3 and H3K79me2 data (two replicates each) were produced by six units of another eight-unit MOWChIP device. 10-week-old CD-1 male mice were used in the study. The nuclei were not sorted and included neurons and glia. The animal protocol was approved by the Institutional Animal Care and Use Committee (IACUC) of Virginia Commonwealth University. (a) Normalized

H3K27me3, H3K9me3, H3K36me3 and H3K79me2 signals. (b) Cross-correlations of H3K27me3, H3K9me3, H3K36me3 and H3K79me2 ChIP-seq signals in the promoter regions. R represents Pearson correlation coefficient.

3.5 Conclusions

In conclusion, the modified MOWChIP protocol can be successfully applied for multiplexed profiling of genome-wide histone modifications, even on some hard-to-profile histone mark. The sequencing results show high cross correlation between their technical replicates MOWChIP-seq operation involves flowing of chromatin fragments through a packed bed of antibody-coated beads followed by a vigorous microfluidic oscillatory washing. The process is semi-automated for reduced labor and improved reproducibility. Using one 8-unit device, it takes 2 d to produce 8 sequencing libraries from chromatin samples. The technology is scalable. We used the protocol to study a number of histone modifications in various types of mouse and human tissues. The protocol can be conducted by a user who is familiar with molecular biology procedures and has basic engineering skills.

4 Investigation of prolonged epigenomic changes following single exposure to a psychedelic in mice brain

4.1 Project Summary

Clinical evidence suggests that rapid and sustained antidepressant action can be attained with a single exposure to psychedelics[81,82]. However, the biological substrates and key mediators of psychedelics' enduring action remain unknown. Here, we show that a single dose of DOI, a type of psychedelic, leads to changes in chromatin organization, particularly at enhancer regions of genes involved in synaptic assembly that stretch for days after the psychedelic exposure. These DOI-induced alterations in the neuronal epigenome overlap with genetic loci associated with schizophrenia, depression, and attention deficit hyperactivity disorder. Together, these data support that epigenomic-driven changes in synaptic plasticity sustain psychedelics' long-lasting antidepressant action but also warn about potential substrate overlap with genetic risks for certain psychiatric conditions.

4.2 Introduction

Psychiatric conditions including depression, anxiety, and stressor-related disorders affect the life of millions of individuals worldwide[83,84]. All currently available medications, including monoamine reuptake-based antidepressants such as fluoxetine, paroxetine, and citalopram require several weeks to months for the clinically relevant improvements to occur, and there is a high proportion of patients taking standard pharmacotherapies who remain treatment resistant[85]. In addition, these pharmacological interventions are

often accompanied by undesirable side effects[86]. There is therefore an urgent need for better antidepressant treatments, with a faster onset of action, which will also offer relief to patients who do not respond to classic antidepressants.

Psychedelics, which can be classified into two main groups: phenethylamines such as mescaline and the substituted amphetamine 1-(2,5-dimethoxy-4-iodophenyl)-2-aminopropane (DOI) and tryptamines such as psilocybin and lysergic acid diethylamide (LSD), are psychoactive compounds that profoundly affect various mental domains, particularly sensory perception and thought processes[15]. Most of the previous preclinical and human neuroimaging studies focused their efforts on the effects that occur within minutes to hours after psychedelic administration, with the aim of elucidating the molecular and neural circuit mechanisms responsible for their psychotic-like states[15,16]. However, recent pilot clinical trials suggest that psychedelics may represent a promising long-lasting treatment for patients with depression and other psychiatric conditions[87-90]. As an example, people with cancer demonstrated rapid increases in positive affect after psilocybin administration, with 60%–85% showing long-term improvements in anxiety and depression measures at a 6.5-month follow-up[89]. Despite these striking effects, their acute psychotic symptoms and drug abuse potential preclude the routine use of psilocybin and other psychedelics in daily clinical practice. Therefore, there is a clear need for basic and translational research focused on understanding the molecular mechanisms mediating the clinical effectiveness of psychedelics—with the ultimate goal of developing safer, more effective, and non-psychedelic treatment strategies.

Preclinical assays in rodent models have evaluated the effects of psychedelics as rapid-acting antidepressant medications. Notably, it has been suggested that a single

administration of psilocybin or LSD produces long-lasting (for 5 weeks post-administration) antidepressant-like effects in rats within models of behavioral despair or passivity such as the forced swim test. Additionally, it has been reported that a single dose of psychedelics such as 4-bromo-3,6-dimethoxybenzocyclobuten-1-yl)methylamine (TCB-2) and N,N-dimethyltryptamine (DMT) facilitates the extinction of cued fear memory in mice and rats, respectively[91,92].

Using classic gene-expression assays such as microarrays, previous studies demonstrated that a single administration of psychedelics including DOI and LSD alters the level of expression of several genes that showed maximal changes at ~60 min—returning to basal level at ~2 h after drug exposure[93,94]. Several of the genes showing transient induction upon acute psychedelic administration have previously been implicated in processes related to transcription and chromatin organization. However, these previous transcriptomic studies in bulk tissue samples suffered from lack of cell-type specificity in their profiling.

In this study, we tested whether a single dose of DOI leads to long-lasting epigenomic and transcriptomic alterations in the frontal cortex using high-resolution, cell-type-specific, and low-input ChIP-seq and RNA-seq measurements. We aimed to unveil the molecular mechanism responsible for the post-acute effects of the psychedelic phenethylamine DOI using mouse behavior models relevant to depression, anxiety, and stressor-related disorders, as well as its implication in epigenetic alterations (defined as covalent histone modifications on their amino tails that change chromatin structure and play an important role in regulating transcription), and associated gene expression.

4.3 Materials and Methods

Experimental model and subject details

Experiments were performed on adult (10 – 20 weeks old) male mice randomly allocated into the different pre-treatment groups. Animals were housed on a 12 h light/dark cycle at 23°C with food and water ad libitum, except during behavioral testing (always during the light cycle). All procedures were conducted in accordance with NIH guidelines and were approved by the Virginia Commonwealth University Animal Care and Use Committee. All efforts were made to minimize animal suffering and the number of animals used. Mice in experimental groups were injected with drugs or vehicle 4 times (control groups: 4 times of vehicle administration; 24h, 48h, and 7d groups: once of drug exposure and 3 times of vehicle administration). Drugs were administered (i.p.) diluted in saline (0.9%); vehicle refers to the saline solution alone. All the mice were sacrificed in the same day (Fig. 4-1). Experiments involving ChIP-seq and RNA-seq were conducted in 129S6/SvEv mice (Taconic).

Nuclei isolation and sorting via FACS

Nuclei isolation was conducted using a published. All steps were conducted on ice, and all centrifugation was conducted at 4°C. One piece of mouse frontal cortex tissue (6-10 mg) was placed in 3 mL of ice-cold nuclei extraction buffer (NEB) [0.32 M sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 0.1 mM EDTA, 10 mM tris-HCl, and 0.1%(v/v) Trition X-100] with 30 µL of freshly added protease inhibitor cocktail (PIC, Sigma-Aldrich), 3 µL of 100 mM phenylmethylsulfonyl fluoride (PMSF, Sigma-Aldrich) in Isopropyl alcohol, 3 µL of 1 M dithiothreitol (DTT, Sigma-Aldrich), and 4.5 µL of

ribonuclease (RNase) inhibitor (2313A,Takara Bio). The tissue was homogenized in a tissue grinder (D9063, Sigma-Aldrich). The homogenate was filtered with a 40 µm cell strainer (22-363-547, Thermo Fisher Scientific) and collected in a 15 mL centrifuge tube. The cell suspension was centrifuged at 1000 RCF for 10 min. The supernatant was discarded, and the pellet was resuspended in 0.5 mL of ice-cold NEB with 5 µL of freshly added PIC, 0.5 µL of PMSF, 0.5 µL of DTT, and 0.75 µL of RNase inhibitor. 500 µL of the sample was mixed with 750 µL of 50%(w/v) iodixanol (made by mixing 4 mL of OptiPrep™ gradient (Sigma-Aldrich) and 0.8 mL of diluent [150 mM KCl, 30 mM $MgCl_2$, and 120 mM tris-HCl]). The mixture was centrifuged at 10,000 RCF for 20 min. Then, the supernatant was removed and 300 µL of 2%(w/v) normal goat serum (50062Z, Life technologies) in Dulbecco's PBS (DPBS, Life technologies) was added to resuspend the nuclei pellet. To separate NeuN+ and NeuN- fractions, 6 µL of 2 ng/µl anti-NeuN antibody conjugated with Alexa 488 (MAB377X, Millipore) in DPBS was added into the nuclei suspension. The suspension was mixed well and incubated at 4°C for 1 h on a rotator mixer (Labnet). After incubation, the sample was sorted into NeuN+ and NeuN- fractions using a BD FACSAria™ cell sorter (BD Biosciences). The sorted NeuN+ nuclei were directly used for RNA-seq experiment. 200 µL of sorted NeuN+ nuclei suspension, containing ~26,000 nuclei, was added into 800 µL of ice-cold PBS for ChIP-seq experiment. 200 µL of 1.8 M sucrose solution, 10 µL of 1 M $CaCl_2$, and 3 µL of 1 M $Mg(Ac)_2$ were added into the mixture. The solution was mixed well and incubated on ice for 15 min. Then, the sample was centrifuged at 1800 RCF at 4°C for 15 min. The supernatant was discarded and the pellet was resuspended in 60 µL of PBS with 0.6 µL of freshly added PIC and 0.6 µL of PMSF and stored on ice until use for ChIP-seq.

Chromatin fragmentation

We used nuclei isolated from frozen mouse brain tissue as example. The extraction of nuclei follows the protocol prepared by Dr. John M. Graham with minor modifications. Then 0.25 µl of protease inhibitor cocktail (PIC) and 0.25 µl of 100 mM Phenylmethanesulfonyl fluoride (PMSF) were freshly added into 25 µl of nuclei suspension which contains 100-10,000 nuclei. The suspension was then mixed with 25 µl of 2x lysis buffer (100 mM tris (pH 8.0), 100 mM NaCl, 30 mM MgCl2, and 4% Triton X-100). The mixture was mixed by vortex, and then incubated at room temperature for 10 min. 2.5 µl of 100 mM CaCl2 and 5.5 µl of 10 U/ µl MNase were added into the mixture, mixed well, and incubated at room temperature for 10 min. 5.5µl of 0.5 M EDTA was added into the mixture, mixed well, and incubated on ice for 10 min. The mixture was centrifuged at 16,100 r.c.f. for 5 min. The supernatant (~55 µl) was then transferred to a new 1.5 ml Eppendorf microcentrifuge tube and placed on ice, for use later.

MOWChIP

Our MOWChIP procedure consists of a number of steps. First, the microfluidic chambers of the device were filled up with IP buffer with the help of syringe pump at a flow rate of 200 µl/min. Then, the antibody coupled magnetic beads were flowed into the device via the combined force provided by pipette and a cylindrical permanent magnet. After plugging the long tubing assembly (10 cm C-Flex clear tubing connected with 3 cm PFA tubing) containing chromatin sample into the inlet of the device, the sieve valve was partially closed by turn on the pressure control system, which applied 30 psi on the valve. The magnetic beads were manipulated by the magnet and packed against the valve to form a packed bed while injecting the chromatin solution into the device at a

flow rate of 1.5 µl/min. Under this flow rate, the immunoprecipitation (IP) part was completed in ~50 min. After that, the long tubing assembly was plugged out of the device.

Oscillatory washing was conducted following the completion of IP. Low salt washing buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X)) was added into two short tubing assembly (6 cm C-Flex clear tubing connected with 3 cm PFA tubing), and then the tubing assemblies were plugged into the inlet and outlet of the microfluidic device unit. The PFA tubing of pressure control system was plugged into the other ends of the short tubing assemblies. Oscillatory washing parameters were set in a pre-written LabVIEW program. Both duration and interval were set as 0.5 s for the alternating pressure pulses operated by the two solenoid valves attached to the same device (In the LabVIEW program, enter 0.5 for both “inlet duration” and “outlet duration”). The number of oscillations was 300, which generated a total washing time of 5 min. The pressure was turned to ~0.65 psi by twisting the knob of pressure regulator. After finishing the washing with low salt washing buffer, low salt washing buffer was replaced by high salt washing buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X), and the process of washing was conducted again with the same washing parameter. The magnetic IP beads were retained in the microfluidic chamber during oscillatory movement by using permanent magnet. After the accomplishment of oscillatory washing (low salt and high salt washing buffer), non-specifically adsorbed chromatin fragments were removed from the IP beads surface. Magnetic IP beads were then eluted from the microfluidic chamber to a new 1.5 ml Eppendorf microcentrifuge tube with IP buffer at a

flow rate of 200 µl/min. The microcentrifuge tube was then placed on a magnetic rack to collect beads. Supernatant was removed and discarded by pipette. 90 µl of elution buffer (10 mM tris-HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA, and 0.03% (wt/v) SDS) and 4 µl of proteinase K (1 mg/ml) were added into the tube to resuspend the beads pellet. The tube was incubated in a multi-therm heat/cool shaker incubator at 65 °C for 2 hours. Phenol chloroform cleanup and ethanol precipitation were performed after incubation to extract and purify ChIP DNA. Finally, ChIP was resuspended in 10 µl of low EDTA TE buffer.

Construction of ChIP-seq libraries

Approximately 10,000 NeuN+ nuclei were used to produce each ChIP-seq library and two technical replicates were generated for each brain sample. 6 brain samples were processed under each treatment condition. One input per sample was generated using ~4,000 nuclei. ChIP was carried out using multiplexed MOWChIP-seq with MNase digestion for chromatin fragmentation and anti-H3K27ac (39135, Active Motif) antibody, following a published protocol.

Construction of RNA-seq libraries

Approximately 5000 NeuN+ nuclei were used to produce each RNA-seq library and two technical replicates were generated for each brain sample. 6 brain samples were processed under each treatment condition. RNA extraction from 50 µL of nuclei suspension from FACS (containing 5000 sorted NeuN+ nuclei) was conducted using RNeasy Mini Kit (74104, QIAGEN) and RNase-Free DNase Set (79254, QIAGEN), following the manufacturer's instruction. Extracted mRNA in 30-µl volume was

concentrated by ethanol precipitation and resuspended in 4.6 µL of RNase-free water. cDNA was prepared using the SMART-seq2 protocol with minor modification. ~2 ng of mRNA in 4.6 µl of water was mixed with 2 µL of 100 µM oligo-dT primer and 2 µL of 10 mM dNTP mix. After being denatured at 72°C for 3 min, the mRNA solution was immediately placed on ice. 11.4 µL of reverse transcript mix [made from 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 (100mM), 4 µL of 5 M Betaine, 0.12 µL of 1 M $MgCl_2$, 0.2 µL of TSO (100 µM), 0.58 µL of nuclease-free water] was mixed with the mRNA solution and the mixture was incubated at 42°C for 90 min, followed by 10 cycles of (50°C for 2 min, 42°C for 2 min). The reaction was finally inactivated at 70°C for 15 min. 20 µL of first-strand mixture was then mixed with 25 µL of KAPA HiFi HotStart ReadyMix, 0.5 µL of (100 µM) IS PCR primers, 0.5 µL of Evagreen dye, and 4 µL of nuclease-free water. Generated cDNA was amplified by incubation at 98°C for 1 min, followed by 9-11 cycles of (98°C 15 s, 67°C 30 s, 72°C 6 min). After PCR amplification, 50 µL of PCR mix was purified by using 50 µL of SPRIselect beads. ~600 pg of purified cDNA was used to produce a RNA-seq library using Nextera XT DNA Library Preparation kit (FC-131-1024, Illumina), following the manufacturer instructions.

Sequencing

The fragment size of ChIP-seq and RNA-seq libraries was measured using high sensitivity DNA analysis kit (5067-4626, Agilent) on a TapeStation system (2200, Agilent). The concentration of each library was examined using a KAPA library quantification kit (KK4809, Kapa Biosystems), and then the quantified libraries were pooled at 10 nM. The libraries were sequenced by Illumina HiSeq 4000 with single-end

50-nt read. Around 15 million reads were generated for each ChIP-seq library, 10 million reads for each input library, and 11 million reads for each RNA-seq library.

ChIP-seq reads alignment and normalization

Once we got the sequencing data back, we conducted the pre-processing pipeline first. The procedure included read trimming (the sequencing reads were trimmed by Trim Galore! with default settings), filtering low-quality reads (SAMTools -view -bq 10), duplicated reads removal (SAMTools rmdup), and subtracting out the blacklisted regions defined by ENCODE (BEDTools subtract). After the pre-processing, uniquely mapped reads from ChIP and input samples were extended to 250 bp from their original length (50 bp) and stored in BED files. MACS2 was used to finish the peak calling step ($q< 0.05$). For visualization in IGV (Broad Institute), the number of reads was counted in 100bp windows over the whole genome and stored the output results as a bigWig file.

Differential enhancers

We first predicted enhancer regions (1000 bp in width with peak summit ± 500 bp) by identifying H3K27ac peaks that did not intersect with promoters (defined as TSS ± 2000 bp). Consensus enhancer sets were generated for each experimental group (VEH, 24 h, 48 h and 7d) using Diffbind and then combined to generate an overall consensus enhancer set. In the process, a majority rule was applied to identify the consensus enhancers when all technical/biological replicates were considered. For example, among all 12 replicates of H3K27ac datasets in one experimental group, a consensus enhancer must be present in at least 7. Differential enhancers between any two

experimental groups were identified from the overall consensus enhancer set using the bioconductor DESeq2 package (FDR < 0.05) with Benjamini-Hochberg method.

K-means clustering, GO term analysis, TF motif analysis

We also performed K-means clustering on differential enhancers and DEGs across experimental groups. In each row, the values (x) were normalized by dividing by $\sqrt{sum\ (x^2)}$. The optimal number of clusters was determined using Silhouette method. We created a list of the enhancer target genes by examining the correlation between gene expression (as reflected by RNA-seq data) and H3K27ac intensity at the candidate enhancers. Compared to assigning an enhancer to its nearest gene, this approach produces a much higher percentage of enhancer-gene pairs that match the results determined experimentally using techniques including Hi-C. All the enhancers that were not linked with their target genes after the step were then processed using ChIPSeeker with default condition and TxDb.Mmusculus.UCSC.mm10.knownGene to link to their nearest genes (as target genes). All the target genes identified, either by correlation between H3K27ac and RNA-seq or by linking with nearest genes, were combined for downstream analysis. GO biological processes enrichment test was performed using the bioconductor clusterProfiler package (with FDR adjusted p value < 0.01). Motif analysis was also performed to determine enriched transcription factor binding motifs among the enhancer regions using Homer (with options –size 200 –p 16, and q-value < 0.01).

GWAS loci association

Peak locations were converted from mm10 to hg19 through UCSC's liftOver tool. GWAS datasets were obtained from NHGRI-EBI's GWAS Catalog (https://www.ebi.ac.uk/gwas/home). GWAS sets were selected in three categories - psychiatric disorders (schizophrenia, depression, and ADHD), neurological disorders (Multiple Sclerosis, Alzheimer's), and other conditions that did not involve brain or neuron functions. Only sets that had at least 500 SNPs were selected in order to effectively perform the analysis. SNPs that did not have a reference SNP ID were removed. Only SNPs present in the 1000 Genomes Project Phase 3 European population were kept. For each GWAS set, if there were n SNPs in the GWAS set, then n SNPs would be randomly selected from all SNPs present in the 1000 Genomes European population set, retaining an identical distribution of SNPs across the chromosomes. The number of differential peaks that overlapped with this SNP test set was calculated and repeated for a total of 10,000 runs (Monte Carlo randomization). From this distribution, we used the R function pnorm, in conjunction with the actual overlap of the GWAS set with the differential peaks, to calculate a p value.

RNA-seq data analysis

Sequencing reads were trimmed under default settings using Trim Galore!. The trimmed reads were mapped to GRCm38 genome by hisat2. Mapped bam files were imported into SeqMonk v1.47.1 (Babraham Institute). Datasets with poor quality (percentage of reads aligned to exons < 40%) were discarded. FeatureCounts was used to count reads and DEGs were determined by pairwise comparison using DESeq2 (FDR < 0.05). The threshold of FDR < 0.05 was used to be consistent with that used in differential enhancers.

Weighted gene co-expression network analysis (WGCNA)

A signed co-expression network was built using the WGCNA package in R. We first normalized gene expression values using log2(FPKM) (Fragments Per Kilobase per Million mapped reads) and then genes were ranked based on their median absolute deviation (MAD). The genes with over 0.3 at MAD were selected as input for the network construction. We converted the normalized gene expression values into an adjacency matrix using the function $a_{ij} = \left|cor(x_i, x_j)\right|^{\beta}$, where x_i and x_j were the expression data of two genes. The value of β was determined to be 7 using the approximate scale-free topology. We further performed the topological overlap measure to transform the adjacency matrix into a topological overlap matrix using the blockwiseModules function in the package (with parameters corType = “bicor,” TOMType = “signed,” minModuleSize = 30, reassignThreshold = 0, mergeCutHeight = 0.15). Modules were then detected based on the topological overlap matrix and the gene expression level of each module was characterized by the module eigengene. The top 100 intramodular hub genes (the most highly connected genes within the module) were identified using the softConnedticity function in the package. The relationship between these hub genes were visualized using Cytoscape.

4.4 Results and Discussion

Frontal cortex is one of the brain regions that heavily involved in processed affected by psychedelics[95]. Because of that, we chose to examine the epigenetic alterations and transcriptomic dynamic of the frontal cortex from mice (n = 6 animals for each condition) that were injected (i.p.) with vehicle or DOI (Fig. 4-1). In the case of DOI

injection, frontal cortex samples were collected at 24h, 48h, or 7 days after DOI administration. Histone modification H3K27ac (acetylation of histone H3 at lysine 27) along with the transcriptome were profiled in NeuN positive ($NeuN^+$) neuronal nuclei isolated from the frontal cortex by fluorescence-activated cell sorting (FACS). We applied microfluidic oscillatory washing ChIP-seq[12,13] (MOWChIP-seq) and SMART-seq2[96,97] to profile H3K27ac and transcriptome, respectively, using the small quantity of neuronal nuclei (10,000 for ChIP-seq and 5,000 for RNA-seq) yielded from the mouse frontal cortex.

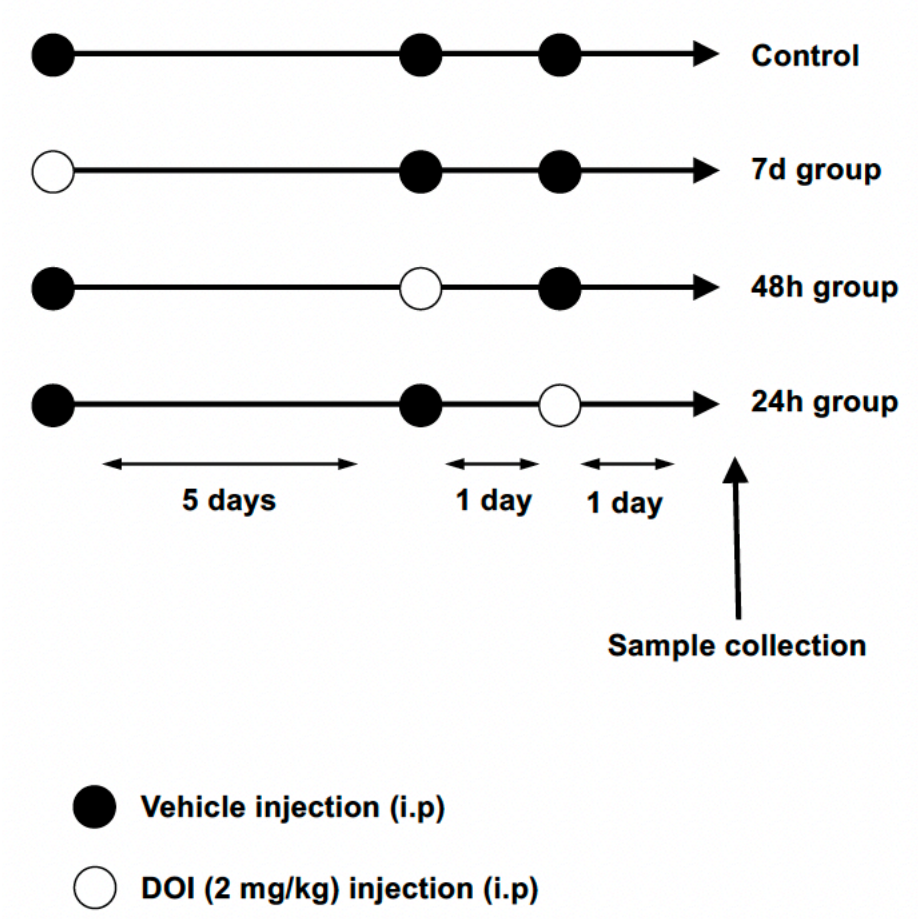

Figure 4-1 Experimental set up of treatment (i.p.) with DOI (2 mg/kg), or vehicle

We first calculated the pairwise Pearson's correlation coefficients across all the H3K27ac data sets, and the results are visualized in a heatmap (Fig. 4-2a). The technical and biological replicates are highly correlated within the same treatment group, with an average correlation coefficient of 0.939 for vehicle group, 0.917 for 24h group, 0.927 for 48h group, and 0.935 for 7d group. Among the treatment groups, 24h and 48h groups appear to be strongly correlated (r=0.913) while the vehicle group and the other three groups have lower correlations (0.854, 0.864, and 0.878 with 24h, 48h and 7d, respectively). The increasing trend in the correlation with the vehicle suggests that the treated mice generally recover over time to become more similar to the vehicle. The distribution of H3K27ac peaks in different genomic regions was relatively consistent across these groups. Less than 15% of peaks were associated with promoter regions (defined as 2 kb regions upstream and downstream of the transcription start site of RefSeq, hg19) (Fig. 4-2b). The majority of signals were located in either intergenic (around 30%) or intronic (around 50%) and approximately 5% were found in exons (Fig. 4-2b).

Using the vehicle as the reference, we identified 2925, 2087, and 2004 differential H3K27ac peaks in 24h, 48 h and 7d groups with false discovery rate (FDR) of less than 0.05, respectively (Fig. 4-2c-f). The vast majority (95.9%) of differential peaks in the 24h group had a decreased intensity compared to the vehicle group. However, the percentage of differential peaks with decreased intensity declined to 71.4% and 62.8%, for 48 h and 7d groups, respectively.

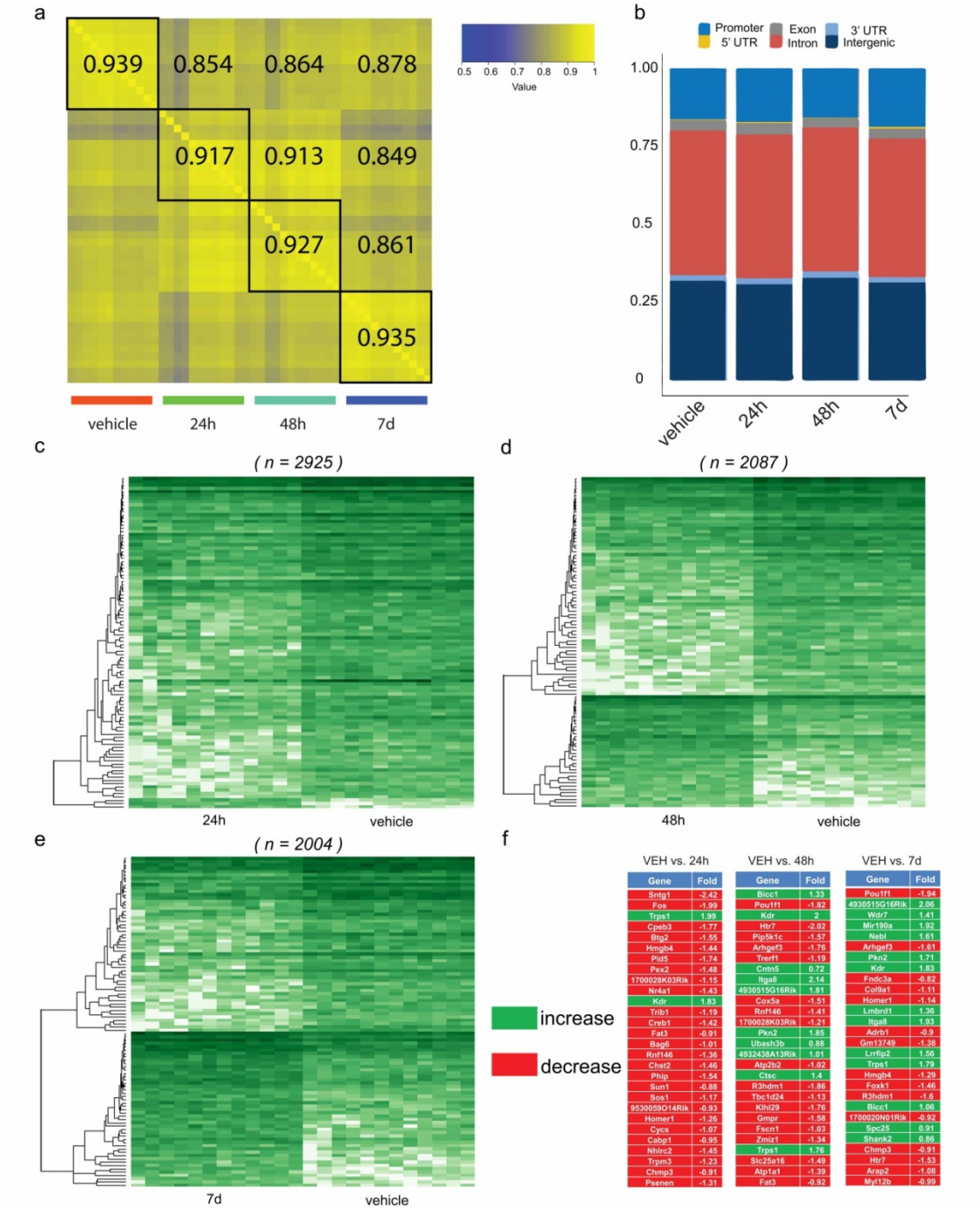

Figure 4-2 Summary of sequencing results. (a) Heatmap for Pearson's correlation coefficients comparing H3K27ac profiles for four mice groups. (b) Bar plot express proportions of H3K27ac peaks in relation to Refseq database, including Promoter,

Exon, Intron, 5' UTR, 3' UTR, and Intergenic. (c-e) Differential binding of H3K27ac in the comparison between vehicle and 24h, vehicle and 48h, and vehicle and 7d. (f) Top 28 differential annotated genes.

Enhancers are highly dynamic epigenomic regulatory elements with known involvement in plasticity and neurodevelopmental processes[98]. We predicted active enhancers by scanning the H3K27ac^{high} regions that did not intersect with promoters[12]. H3K27ac mark distinguishes active enhancers from poised and inactive enhancers[48]. H3K27ac and enhancers are also highly differentiating of various brain locations and functions[11]. We conducted K-means clustering of differential enhancers that showed change from the control (false discovery rate [FDR] <0.05) due to DOI exposure at one or more of the three time points, and examined gene ontology (GO) terms and enriched transcription factor (TF) binding motifs associated with the clusters (Fig. 4-3a). We divided all differential enhancers into 6 clusters with various patterns of variation after DOI injection (Fig. 4-3a). Clusters 1–3 represent enhancers exhibiting eventual increase after DOI administration, while clusters 4–6 show a decrease in DOI-injected mice as compared to vehicle-treated mice. Clusters 1 and 2 show an increase that peaks at 48 h before decreasing at 7 days, while cluster 3 experiences a much slower increase that shows the highest intensity at 7 days. Clusters 4 and 5 have the deepest decrease at 24 h, while enhancer intensity slowly recovers over 48 h and 7 days. Cluster 6 presents an initial decrease at 24 h followed by minor fluctuations at 48 h and 7 days. Interestingly, at least 32.7% of these differential enhancers (clusters 3 and 6) are still in their altered state at 7 days, suggesting long-lasting effects of DOI at the epigenomic level that outlast by several days the presence of the drug in native tissue.

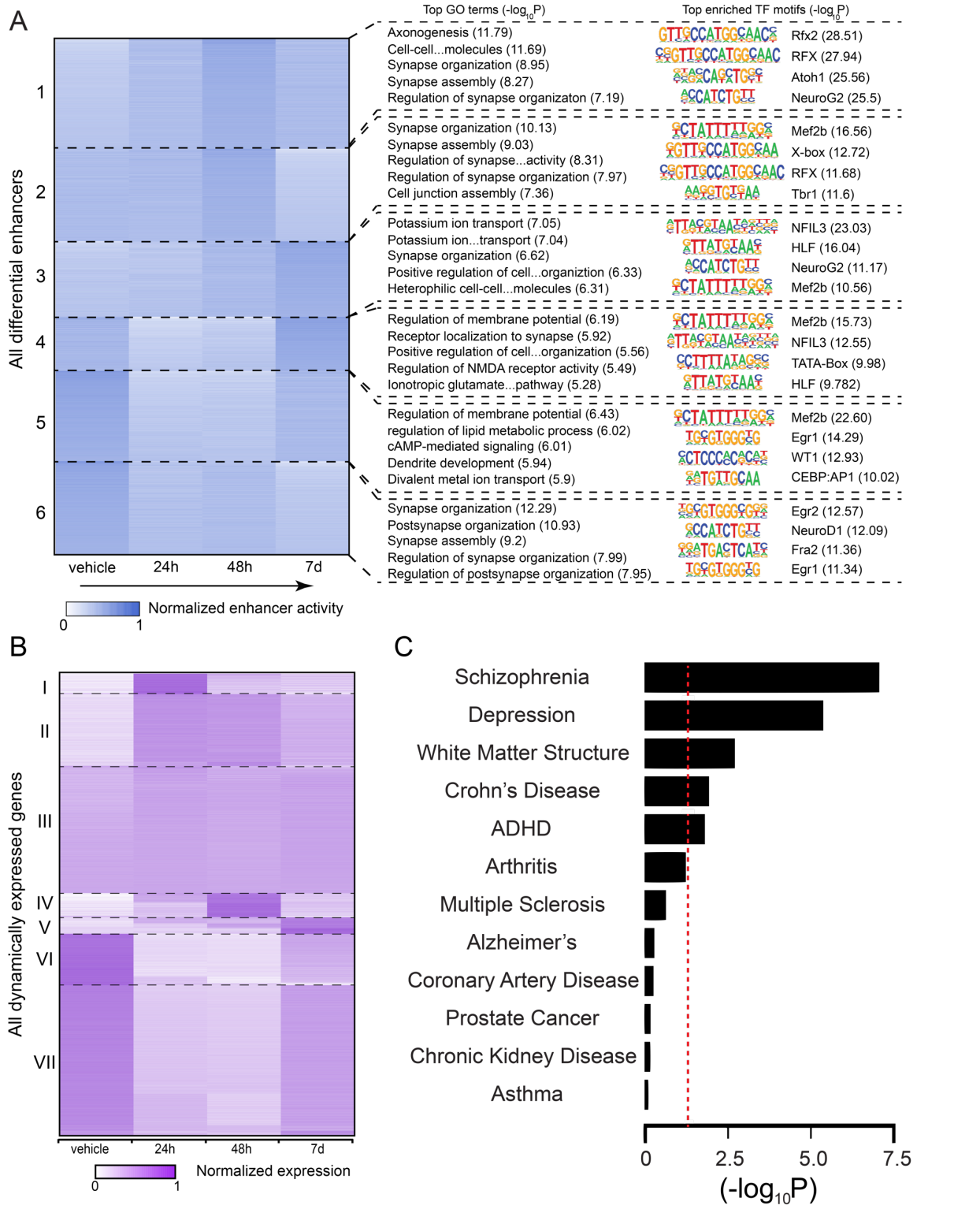

Figure 4-3 Post-acute effects of DOI on frontal cortex epigenomic and transcriptomic variations. (a) Effect of DOI on time-lapse epigenomic variations in neuronal nuclei of

the mouse frontal cortex. K-means clustering of differential enhancers based on normalized H3K27ac signal (n = 3,995; Cluster sizes from top to bottom: 852, 718, 585, 419, 700, and 721). The top-five enriched biological process gene ontology (GO) terms identified by clusterprofiler are listed next to each cluster, and the top four motifs enriched in each cluster are listed next to the GO terms together with their sequences (n = 6 mice per group). (b) K-means clustering of normalized gene expression (n = 13,605). Cluster sizes from top to bottom: 604, 2,120, 3,748, 708, 497, 1,507, 4,421. (c) Significance of overlap between differential H3K27ac peaks and NHGRI-EBI GWAS SNP sets. Significance is calculated using a null distribution and is shown as -log(p-value). The red dotted line denotes $p < 0.05$ cutoff. GWAS sets are ordered according to significance. Frontal cortex samples were collected 24-h, 48-h, and 7-d after a single injection (i.p.) of DOI (2 mg/kg) or vehicle.

To directly interrogate the functional relevance of these findings, we examined the GO terms and TF motifs associated with each of the clusters (Fig. 4-3a). Synapse organization and assembly are enriched in all clusters to various degrees, consistent with the effects of DOI on structural and synaptic plasticity. The top GO terms enriched in clusters 1 and 2 include axonogenesis, synapse organization, and cell junction assembly. Motifs of several TFs importantly involved in neuronal growth and synaptic plasticity, such as Mef2B, NeuroG2, and Atoh1[99], are also highly enriched in these clusters. Clusters 4–6, which experience a decrease upon DOI administration, are associated with GO terms related to regulation of glutamate NMDA receptor activity, regulation of membrane potential, and synapse assembly. Enrichment of two zinc finger-containing transcription factors, Egr1 and Egr2, also occurs in clusters 5 and 6.

Considering previous findings showing that psychedelics lead to an acute increase in glutamate release as well as augmentation of Egr1 and Egr2 transcription in mouse frontal cortex neurons[100], these data suggest a compensatory mechanism as a potential explanation for the enhancer intensity decrease observed in clusters 4–6 upon DOI administration. Clusters 3 and 6 represent enhancers with long-lasting (7 days) DOI-dependent increased and decreased activity, respectively. Cluster 3 shows enrichment in GO terms associated with rhythmic processes. This finding, together with NFIL3, a TF involved in the mammalian circadian oscillatory mechanism[101], as the top enriched TF motif in this cluster, is consistent with medical evidence of long-term effects of psychedelics on sleep-wake cycles. Cluster 6 has synaptic plasticity-related genes heavily enriched.

The transcriptomic variations induced by DOI appear to be much more transient than the epigenomic changes (Fig 4-3b). Based on K-means clustering, only 3.7% (cluster V) of the differentially expressed genes (DEGs, FDR <0.05) present long-term variation (7 days), and only a small number of these genes (n = 574, or 4.2% of the total DEGs) exhibit a pattern of expression that correlates with enhancer dynamics (i.e., with a Spearman's correlation r > 0.4 between enhancer intensity and expression level). These results suggest that the long-term effects of the psychedelic DOI are associated with epigenomic regulations of greater magnitude than changes in transcriptomic dynamics.

To determine if there was a significant overlap of differential H3K27ac peaks with single nucleotide polymorphisms (SNPs) associated with depression that were reported based on genome wide association studies (GWAS), a null distribution was calculated

using the set of all SNPs (Fig. 4-3c). This was repeated for different GWAS sets that are associated with various traits to determine specificity. Of twelve GWAS loci datasets, five (schizophrenia, depression, white matter structure, Crohn's disease and attention deficit hyperactivity disorder [ADHD]) were statistically significant (Fig. 4-3c).

We further investigated our RNA-seq data using weighted correlation network analysis (WGCNA) with the goal of detecting co-expression modules with highly correlated genes to associate these modules with sample traits or experimental groups[102]. Over 14,000 genes with their expression levels throughout all samples in the top 30% were selected, and the generated data matrix was used as the input for WGCNA. After construction of a co-expression network from the matrix, ten modules with densely interconnected genes were detected and the expression pattern in each module was summarized by the module eigengene. The eigengene significance was used for assessing correlation between experimental groups and each module. We performed GO analysis on the genes included in these modules and 5 modules (blue, green, turquoise, yellow, and magenta) showed significantly enriched GO terms (Figure 4-4).

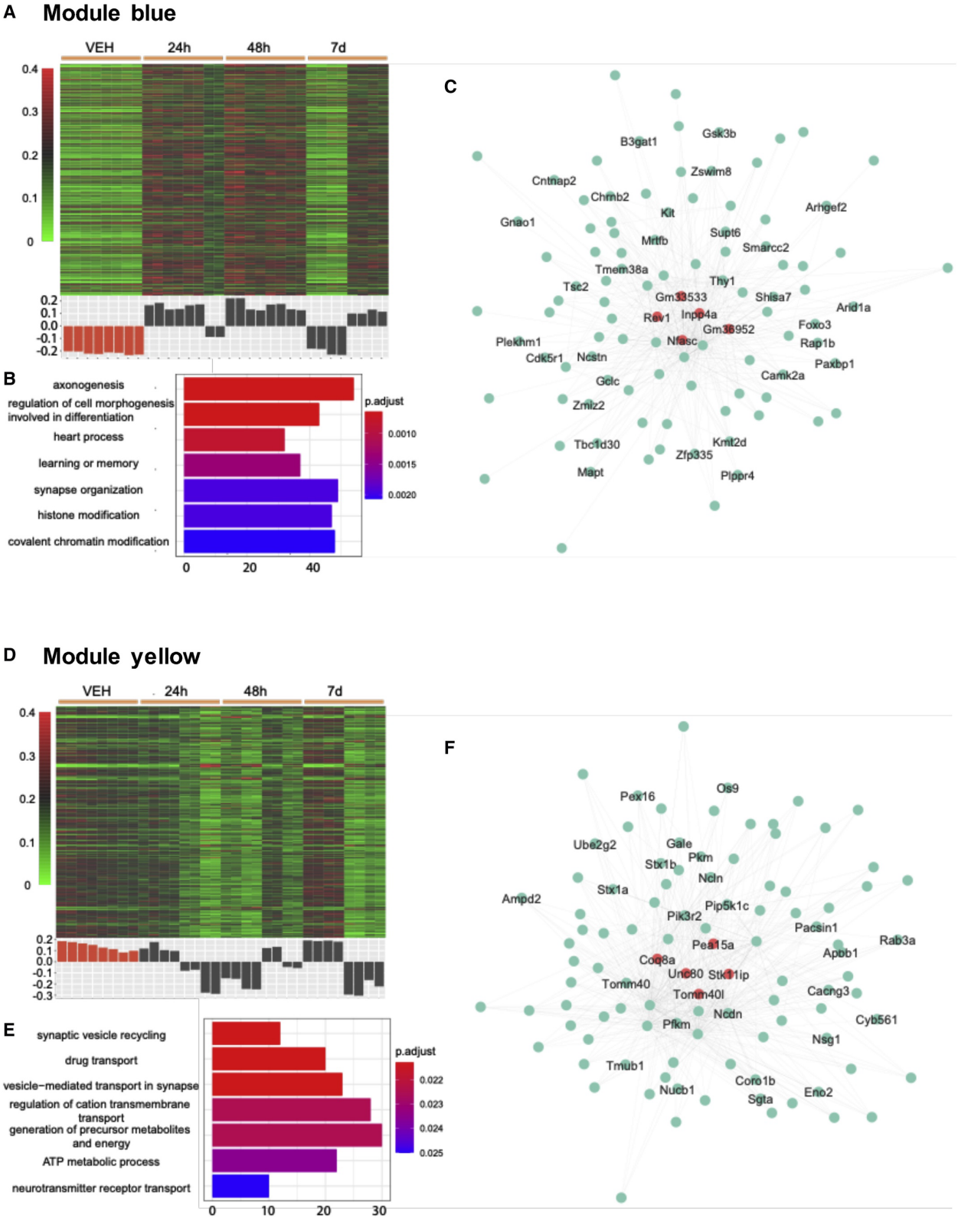
A
Module blue
VEH
24h
48h
7d
B
axonogenesis
regulation of cell morphogenesis involved in differentiation
heart process
learning or memory
synapse organization
histone modification
covalent chromatin modification
p.adjust
C
Gsk3b
B3gat1
Zswim8
Cntnap2
Chmb2
Kit
Arhgef2
Gnao1
Supt6
Mttfb
Smarcc2
Tmem38a
Thy1
Tsc2
Gm38533
Shisa7
Arid1a
Rev1
Inpp4a
Foxo3
Gm38952
Plekhm1
Nfasc
Rap1b
Cdk5r1
Ncstn
Paxbp1
Camk2a
Gclc
Zmiz2
Tbc1d30
Kmt2d
Zfp335
Mapt
Pippr4
D
Module yellow
E
synaptic vesicle recycling
drug transport
vesicle−mediated transport in synapse
regulation of cation transmembrane transport
generation of precursor metabolites and energy
ATP metabolic process
neurotransmitter receptor transport
F
Pex16
Os9
Ube2g2
Gale
Pkm
Stx1b
Ncln
Stx1a
Ampd2
Pip5k1c
Pik3r2
Pacsin1
Pea15a
Rab3a
Apbb1
Coq8a
Unc80
Stk11ip
Tomm40
Tomm40l
Cacng3
Ndn
Pfkm
Cyb561
Nsg1
Tmub1
Nucb1
Coro1b
Sgta
Eno2

Figure 4-4 Two gene co-expression modules (blue and yellow) associated with administration of DOI. (a and d) Heatmap of normalized gene expression profiles in the co-expression module (top). The module eigengene values across samples in four experimental groups (bottom). (b and e) Selected top categories from GO enrichment analysis. (c and f) Visualization of the intramodular connections among the top 100 hub genes in each module. The top 5 genes are in large size and colored orange. The genes involved in the top 15 GO terms are labeled.

Among all 10 modules, module blue showed the strongest positive correlation with group 48 h and the strongest negative correlation with the control group. Compared to the control, the eigengene of module blue is overexpressed in both 24 and 48 h groups before decreasing to the control level on 7 days (Fig. 4-4a). GO analysis of this cluster of 1,933 genes leads to terms including: “axonogenesis” (p.adjust 6.44 × 10−4), “heart process” (p.adjust 8.32 × 10−4), “learning or memory” (p.adjust 1.35 × 10−3), and “histone modification” (p.adjust 1.95 × 10−3) (Fig. 4-4b). The GO term “axonogenesis” is also identified in the cluster 1 of differential enhancers (Fig. 4-3a). Inpp4a, Nfasc, and Cntnap2 are in the top 100 hub genes of module blue (Fig. 4-4c). Previous studies suggested that Inpp4a may play an important role in the development of epilepsy[103]. Epileptic seizures have been reported as a result of psychostimulant drug use, although this adverse effect is less common with psychedelics than with other psychostimulants such as cocaine[104]. Loss of function of Inpp4a also increases severity of asthma, which may explain the mechanism underlying the effect of DOI preventing allergic asthma in a mouse model[105]. Cntnap2 and Nfasc, which are genes previously involved in axonal guidance, synaptogenesis and neuron-glial cell interactions[106-108] as well as implicated

in neurodevelopmental psychiatric conditions such as schizophrenia and autism[109-111], were also identified in differential enhancers from clusters 3 and 6 (Fig. 4-3a), respectively.

The eigengene expression in module yellow shows opposite trend of that in module blue (Fig. 4-4d). Among all 10 modules, module yellow is the only one that showed a positive correlation with the control group (Fig. 4-5j). GO biological process enrichment for the genes in module yellow leads to terms including: “synaptic vesicle recycling” (p.adjust 0.021), “drug transport” (p.adjust 0.021), “vesicle-mediated transport in synapse” (p.adjust 0.021), (Fig. 4-4e). Interestingly, this module was also enriched with genes related to hypoxia-inducible factor 1 (HIF-1) signaling pathway, including Slc2a1, Eno2 (hub gene), Gapdh, Aldoa, Mapk3, Pfkl, and Pfkm (hub gene) (Fig. 4-4f). Since the HIF-1 pathway is a major modulator of hypoxia stress response pathways[112], these findings further support the hypothesis the 5-HT2AR modulate the ventilatory response to hypoxia[113].

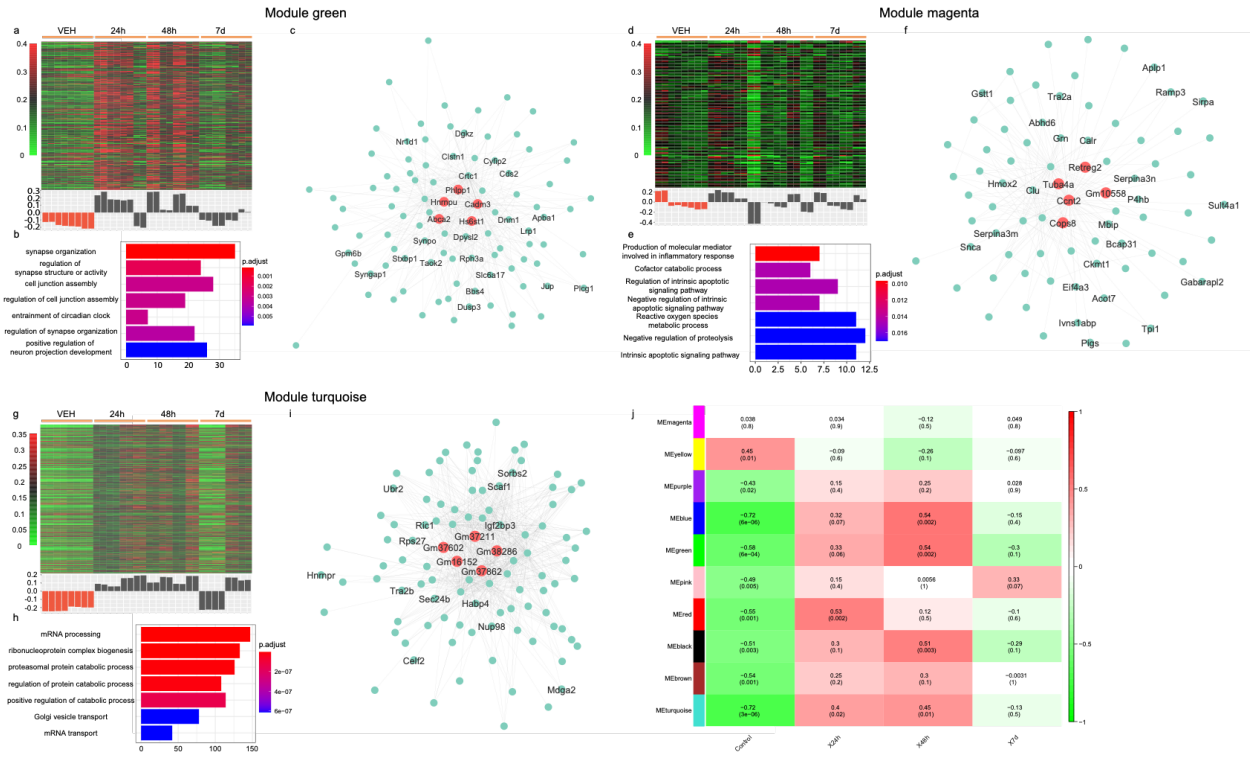

Figure 4-5 Gene co-expression modules (green, magenta, and turquoise) associated with administration of DOI. (a, d, g) Heatmap of normalized gene expression profiles in the co-expression module (top). The module eigengene values across samples in four experimental groups (bottom). (b, e, h) Selected top categories from GO enrichment analysis. (c, f, i) Visualization of the intramodular connections among the top 100 hub genes in each module. The top 5 genes are in large size and colored orange. The genes involved in the top 15 GO terms are labeled. (j) Each row corresponds to the module eigengene, column to the status of DOI administration. The two numbers in each cell represented the corresponding correlation coefficient and p value.

Module magenta's eigengene shows no correlation with any experimental groups (correlation coefficient in the range of −0.12 to 0.05, Fig. 4-5a and j). The genes in module magenta are enriched in terms related to inflammatory response and intrinsic

apoptotic signaling pathways (Fig. 4-5b). Modules green and turquoise have their eigengenes in a similar pattern as that of module blue (Fig. 4-5d, g, and j). Their top GO terms were “synapse organization” (p.adjust 1.89 × 10−4) and “mRNA processing” (p.adjust 2.61 × × 10−17) (Fig. 4-5e and h)

4.5 Conclusions

Our data here support a scheme whereby a single dose of the phenethylamine DOI leads to changes in frontal cortex synaptic plasticity and behavior models of fear extinction. Considering the associative learning processes involved in contextual fear conditioning and extinction, and the effect of DOI on the faster development of the later, it is plausible that activation of 5-HTR2A by this psychedelic accelerates context fear extinction likely in part through alterations in chromatin state at enhancer regions of genes predominantly involved in synapse organization and assembly such as Mef2B, NeuroG2, and Atoh1—a gene network that may be located downstream of the genes showing transient transcription upon acute psychedelic administration. Recent observations have also reported that a putative non-hallucinogenic psychedelic analog referred to as tabernanthalog promotes neural plasticity and reduces rodent behavior models relevant to depression[114].

In conclusion, our study unveils persisting chromatin remodeling events following DOI administration linked to lasting synaptic plasticity and behavioral events. Furthermore, the overlap of the epigenomic markers of the action of the psychedelic DOI with loci associated with schizophrenia reiterate the need for caution in the use of psychedelics in individuals at risk for psychosis.

5 Antipsychotic effect on cell-type-specific epigenomic landscapes in the frontal cortex of schizophrenia subjects

5.1 Project Summary

Schizophrenia results from a combination of genetic and environmental causes. Epigenetic marks allow a dynamic impact of environmental factors, including antipsychotic medications, on access to genes and regulatory elements. However, no study so far has profiled cell-specific genome-wide histone modifications in postmortem brain samples from schizophrenia subjects or the effect of antipsychotic treatment on such epigenetic marks. Here we show the first comprehensive epigenomic characterization of the frontal cortex of 29 individuals with schizophrenia and 29 matched controls. Schizophrenia subjects exhibited thousands of neuronal vs glial epigenetic differences at regions that included several susceptibility genetic loci, such as NRXN1, RGS4 and GRIN3A. Additionally, the analysis of RNA-seq data suggest that the effect of age on the epigenomic landscape is more pronounced in frontal cortex of antipsychotic-treated schizophrenics. This provides important evidence of epigenetic alterations in the frontal cortex of individuals with schizophrenia, and remarks the impact of age and antipsychotic treatment on chromatin organization.

5.2 Introduction

Schizophrenia has traditionally been viewed as a genetic disorder, with heritability rates estimated at ~73%[115,116]. However, previous genome-wide association studies (GWAS) showed a relatively low number of genetic regions associated with

schizophrenia risk – these include 108 loci in the first study[117] that have been expanded to over 270 regions[118,119]. Most of these genetic variants were located in non-coding regions. There were a few exceptions that these variants were located in coding area, however, there is little evidence supporting that coding variants contribute to schizophrenia risk. These findings suggest that genetic factors do not seem to be sufficient to fully explain the molecular causes underlying this severe psychiatric condition.

Twin studies have provided evidence that environmental factors contribute to schizophrenia susceptibility. Thus, it had originally been suggested that monozygotic twins, whose DNA sequences are ~100% identical, have a concordance for schizophrenia of nearly 50%[120,121], yet the most recent work reported a probandwise concordance rate of 33% in monozygotic twins and 7% in dizygotic twins[122]. While such results indicate an important contribution of genetic factors to the etiology of schizophrenia, they argue for a significant role of environmental events in the development of this complex disease. This concept is further supported by epidemiological studies suggesting that prenatal environmental insults, such as maternal infection[123,124] and severe adverse life events[125], increase the risk of schizophrenia in the offspring.

Gene expression is regulated by the ability of the transcriptional machinery to access DNA, which is tightly packed into chromatin. The status of chromatin organization depends on epigenetic factors, such as DNA methylation and histone modifications that primarily occur on amino-terminal tails[126,127]. Hence these epigenetic mechanisms lead to stable changes in gene expression that are mediated via altered

chromatin structure without modification of DNA sequence and remain largely plastic throughout all periods of brain development and aging. It is then tempting to speculate that epigenetic mechanisms mediate, at least in part, the effects of environmental factors on CNS gene activity, and are therefore potentially involved in the pathophysiology of schizophrenia and other mental illnesses.

Supporting this concept, previous studies reported alterations in chromatin structure and accessibility in tissue samples from schizophrenic subjects and controls. Most of these previous reports, however, were focused on DNA methylation differences in peripheral blood[128] and brain[129,130]. Using the assay for transposase-accessible chromatin sequencing (ATAC-seq) in bulk tissue homogenates of postmortem frontal cortex samples, only a few differences in chromatin accessibility were observed between schizophrenia subjects and controls, in contrast to thousands of age-related differential accessible chromatin regions[131]. Although interesting, the ATAC-seq approach does not provide the same chromatin state information as histone modifications mapped by chromatin immunoprecipitation sequencing (ChIP-seq) assays do. Histone modifications, including histone H3 acetylation of lysine 27 (H3K27ac) and histone H3 trimethylation of lysine 4 (H3K4me3) are critically involved in epigenomic regulations; H3K27ac marks active enhancers[48], whereas H3K4me3 marks active promoters[132]. Enhancers are highly dynamic cis-regulatory elements with known involvement in neurodevelopmental processes[133], and the dynamics of promoters are also significantly connected with the genetic risk of certain psychiatric conditions[134]. However, very few studies have been conducted about potential cell-type-specific

genome-wide variations in covalent histone modifications in postmortem human brain samples of individuals with schizophrenia.

As an example, recent work combined fluorescence-activated cell sorting (FACS) of neuronal and non-neuronal cell nuclei with histone ChIP-seq assays in two brain regions (prefrontal cortex and anterior cingulate cortex) from postmortem brain samples of subjects without any known neurological or psychiatric disease[135]. Besides the identification of cell and region-specific histone modification landscapes in this cohort of control subjects, these findings also compared their datasets with previous GWAS of individuals with psychiatric conditions, reporting that strong specific enrichments occurred with schizophrenia and weaker associations with depression in both H3K27ac and H3K4me3 peaks. This correlation was almost exclusively observed in neuronal chromatin, but not in non-neuronal cell nuclei. A more recent investigation conducting H3K4me3 ChIP-seq assays in cortical neurons from schizophrenic subjects and controls has identified rare specific epigenetic variations for a set of non-coding RNA genes that may contribute to the pathogenesis of schizophrenia[136]. However, these previous epigenomic studies in postmortem human brain samples did not address the potential effect of previous exposure to antipsychotics on the regulation of chromatin state. This is particularly relevant considering that repeated administration of antipsychotic medications leads to epigenetic modifications at selected gene regions in mouse[137,138] and postmortem human brain[139,140] samples.

In this study, we combined MOWChIP-seq[12,13] and SMART-seq2[96,97] for low-input profiling of H3K27ac and H3K4me3 histone modifications and transcriptomes, respectively. Here, we present the first dataset with cell type-specific epigenomic and

transcriptomic landscapes in postmortem frontal cortex samples from two cohorts of schizophrenics either previously treated or not with antipsychotic medications and control subjects individually matched by sex and age.

5.3 Materials and Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments including schizophrenia and control pairs were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Postmortem human brain tissue samples

Human brains were obtained during autopsies performed at the Basque Institute of Legal Medicine, Bilbao, Spain. The study was developed in compliance with policies of research and ethical review boards for post-mortem brain studies (Basque Institute of Legal Medicine, Spain). Deaths were subjected to retrospective searching for previous medical diagnosis and treatment using examiner's information and records of hospitals and mental health centers. After searching of antemortem information was fulfilled, 29 subjects (Caucasian) who had met criteria of schizophrenia according to the Diagnostic and Statistical Manual of Mental Disorders (DSMIV) were selected. A toxicological screening for antipsychotics, other drugs and ethanol was performed on blood samples collected at the time of death and, when possible, postmortem brain samples. The toxicological assays were performed at the National Institute of Toxicology, Madrid, Spain, using a variety of standard procedures including radioimmunoassay, enzymatic immunoassay, high-performance liquid chromatography and gas chromatography-mass

spectrometry. Controls (Caucasian) for the present study were chosen among the collected brains on the basis, whenever possible, of the following cumulative criteria: (i) negative medical information on the presence of neuropsychiatric disorders or drug abuse; (ii) appropriate sex, age, postmortem delay (time between death and autopsy), and freezing storage time to match each subject in the schizophrenia group; (iii) sudden and unexpected death (motor vehicle accidents); and (iv) toxicological screening for psychotropic drugs with negative results except for ethanol. Specimens of frontal cortex (Brodmann area 9) were dissected at autopsy (0.5-1 g tissue) on an ice-cooled surface and immediately stored at -80°C until use. The schizophrenia subjects were divided into antipsychotic-free and antipsychotic-treated according to the presence or absence of antipsychotics in blood samples at the time of death. The definitive pairs of antipsychotic-free schizophrenics and respective matched controls are shown in the Table 5-1, and the definitive pairs of atypical antipsychotic-treated schizophrenics and respective matched controls are shown in the Table 5-2. Pairs of schizophrenia and matched controls were processed simultaneously and under the same experimental conditions. Tissue pH values were within a relatively narrow range (control subjects: 6.7 ± 0.08; schizophrenic subjects: 6.6 ± 0.06). Brain samples were also assayed for RIN (RNA integrity number) values using the Agilent 2100 Bioanalyzer (Applied Biosystems), as previously reported (control subjects: 7.87 ± 0.21; schizophrenic subjects: 7.61 ± 0.32).

	Sex (F/M)	Age at death (years)	Postmortem delay (h)	Antipsychotics in blood	Additional drugs in blood
Control 1	M	22	20		(-)
Schizophrenia 1	M	24	45	(-)	(-)
Control 2	M	21	30		ETH (0.34 g/l)
Schizophrenia 2	M	21	24	(-)	(-)
Control 3	M	18	25		(-)

Schizophrenia 3	M	18	18	(-)	(-)
Control 4	M	32	28		ETH (0.68 g/l)
Schizophrenia 4	M	31	14	(-)	(-)
Control 5	F	28	55		(-)
Schizophrenia 5	F	28	22	(-)	(-)
Control 6	M	28	30		(-)
Schizophrenia 6	M	27	24	(-)	(-)
Control 7	M	25	20		cannabis
Schizophrenia 7	M	25	16	(-)	(-)
Control 8	F	66	17		(-)
Schizophrenia 8	F	67	22	(-)	(-)
Control 9	M	34	17		(-)
Schizophrenia 9	M	34	23	(-)	(-)
Control 10	F	51	10		(-)
Schizophrenia 10	F	53	18	(-)	(-)
Control 11	M	33	23		(-)
Schizophrenia 11	M	32	21	(-)	(-)
Control 12	M	44	23		(-)
Schizophrenia 12	M	43	23	(-)	(-)
Control 13	M	49	20		(-)
Schizophrenia 13	M	49	19	(-)	(-)
Control 14	M	71	9		(-)
Schizophrenia 14	M	70	22	(-)	BDZ
Control 15	F	74	20		(-)
Schizophrenia 15	F	74	30	(-)	(-)

Table 5-1 Demographic characteristics of antipsychotic-free schizophrenia subjects and their respective control subjects. Abbreviations: Ethanol in blood is coded as ETH, benzodiazepines (BDZ)

	Sex (F/M)	Age at death (years)	Postmortem delay (h)	Antipsychotics in blood	Additional drugs in blood
Control 16	M	32	20		ETH (2.37 g/l)
Schizophrenia 16	M	32	8	QTP	BDZ
Control 17	M	26	27		(-)
Schizophrenia 17	M	26	39	OLZ	(-)
Control 18	M	36	18		ETH (1.69 g/l)
Schizophrenia 18	M	35	11	CLZ	BDZ
Control 19	M	52	23		(-)
Schizophrenia 19	M	51	28	OLZ, CLZ	(-)
Control 20	M	56	24		(-)
Schizophrenia 20	M	58	24	PAL	(-)
Control 21	F	48	21		(-)
Schizophrenia 21	F	52	31	RIS	(-)
Control 22	M	34	36		ETH (0.74 g/l)
Schizophrenia 22	M	34	21	HAL	BDZ
Control 23	M	42	6		ETH (0.72 g/l)
Schizophrenia 23	M	43	6	PAL	BDZ
Control 24	F	50	10		(-)

Schizophrenia 24	F	50	14	RIS, QTP	BDZ
Control 25	F	60	48		(-)
Schizophrenia 25	F	60	22	CLZ, ASU	BDZ
Control 26	M	54	16		ETH (0.58 g/l)
Schizophrenia 26	M	56	12	OLZ, CLT	(-)
Control 27	M	41	14		(-)
Schizophrenia 27	M	41	11	QTP, CLT	BDZ
Control 28	M	36	12		ETH (1.0 g/l)
Schizophrenia 28	M	36	8	OLA	(-)
Control 29	F	41	22		(-)
Schizophrenia 29	F	42	14	CLZ, SUL, QTP	(-)

Table 5-2 Demographic characteristics of antipsychotic-treated schizophrenia subjects and their respective control subjects. Abbreviations: amisulpiride (ASU), clotiapine (CLT), clozapine (CLZ), haloperidol (HAL), olanzapine (OLZ), quetiapine (QTP), risperidone (RIS), paliperidone (PAL), sulpiride (SUL), and benzodiazepines (BDZ). Ethanol in blood is coded as ETH.

Fabrication of microfluidic devices

The two-layer microfluidic MOWChIP devices were made of polydimethylsiloxane (PDMS). The fabrication of the devices followed multilayer soft lithography with minor optimization. First, two photomasks were prepared with the microscale patterns drawn by software Layout editor and printed on high-resolution transparencies (5080 dpi). Then the masters for fluidic layer (~40 μm) and for control layer (~60 μm) were made of photoresist SU-8 2025 on a 3-inch silicon wafer. After that, 36 g of PDMS mixture (Momentive, cat. no. RTV615) at ratio of 5:1 (reagent A: reagent B) was poured into a Petri dish contains the fluidic layer master to form fluidic layer part of the devices (~6 mm). Similarly, 5 g of PDMS mixture at ratio of 20:1 was poured into another Petri dish contains the control layer master. Both of the Petri dishes with PDMS mixture were cured at 80 °C for 12 min. Fluidic layer was peeled off and aligned to control layer according to the designed align-mark. The two-layer PDMS slab was baked at 80 °C for

another 2 hours. The slab was then cut into several pieces, each piece contains 4 units. Three holes were punch for every unit, and the holes served as inlet, outlet, and sieve valve port later. Finally, all the PDMS pieces, in company with pre-cleaned glass slide, were treated by oxygen plasma and bound with treated glass slide to form MOWChIP chips.

Functionalization of magnetic beads with an antibody

5 µl of Dynabeads Protein A were washed twice with ice-cold IP buffer (20 mM Tris-HCl(pH 8.0), 140 mM NaCl, 1 mM EDTA, 0,5 mM EGTA, 0.1%(w/v) sodium doxycholate, 0.1% SDS, and 1%(v/v) Triton-100X in Milli-Q water). Then, the beads were resuspended in 150 µl of IP buffer. Specific amount of antibody against a target histone modification (0.15 µg of H3K4me3 or 0.5 µg of H3K27ac) was added into the beads suspension. The optimized amount of antibody against specific histone modification was determined by the ChIP-qPCR results. The mixture was incubated at 4 °C overnight on a rotator at 24 r.p.m. Beads coupled with antibody were washed three times with ice-cold IP buffer prior to be used in MOWChIP assays.

Nuclei isolation and sorting via FACS

Nuclei isolation from frozen tissues (never fixed) of postmortem human brain samples was conducted using a published protocol with some modifications. Frontal cortex samples from schizophrenic individuals and individually matched controls were always processed in the same batch. Briefly, all steps were conducted on ice, and all centrifugation was conducted at 4 °C. One piece of brain tissue (~300 mg) was placed in 3 ml of ice-cold nuclei extraction buffer (NEB) [0.32 M sucrose, 5 mM CaCl2, 3 mM

$Mg(Ac)_2$, 0.1 mM EDTA, 10 mM Tris-HCl, and 0.1%(v/v) Triton X-100] with freshly added 30 µl of protease inhibitor cocktail (PIC, Sigma-Aldrich), 3 µl of 100 mM phenylmethylsulfonyl fluoride (PMSF, Sigma-Aldrich) in isopropyl alcohol, 3 µl of 1 M dithiothreitol (DTT, Sigma-Aldrich), and 4.5 µl of recombinant ribonuclease (RNase) inhibitor (2313A,Takara Bio). The tissue was homogenized in tissue grinder (D9063, Sigma-Aldrich). The homogenate was filtered with a 40 µm cell strainer (22-363-547, Thermo Fisher Scientific) and collected in a 15-ml centrifuge tube. The cell suspension was centrifuged at 1000 r.c.f. at 4 °C for 10 min. The supernatant was discarded, and the pellet was resuspended in 0.5 ml of ice-cold NEB with freshly added 5 µl of PIC, 0.5 µl of PMSF, 0.5 µl of DTT, and 0.75 µl of RNase inhibitor. 500 µl of the sample was mixed with 750 µl of 50%(w/v) iodixanol (a mixture of 4 ml of OptiPrepTM gradient (Sigma-Aldrich) and 0.8 ml of diluent [150 mM KCl, 30 mM $MgCl_2$, and 120 mM Tris-HCl]). The mixture was centrifuged at 10,000 r.c.f. at 4 °C for 20 min. Then, the supernatant was removed and 300 µl of 2%(w/v) normal goat serum (50062Z, Life technologies) in Dulbecco's PBS (DPBS, Life technologies) was added to resuspend the nuclei pellet. To label and separate NeuN+ and NeuN- fractions, 6 µl of 2 ng/ml anti-NeuN antibody conjugated with Alexa 488 (MAB377X, Millipore) in DPBS was added into the nuclei suspension. The suspension was mixed well and incubated at 4 °C for 1 hour on a rotator mixer (Labnet). After incubation, the sample was sorted into NeuN+ and NeuNpopulations using a BD FACSARIATM Flow Cytometer (BD Biosciences). 400 µl of sorted nuclei suspension (NeuN+ or NeuN-), containing ~50000 nuclei (for conducting ChIP-seq and input libraries), was added into 600 µl of ice-cold PBS. 200 µl of 1.8 M sucrose solution, 5 µl of 1 M $CaCl_2$, and 3 µl of 1 M $Mg(Ac)_2$ were added into

the mixture. The solution was mixed well and incubated on ice for 15 min. Then, the sample was centrifuged at 1800 RCF at 4 °C for 15 min. The supernatant was discarded, and the pellet was resuspended in 110 µl of PBS with freshly added 1.1 µl of PIC and 1.1 µl of PMSF and stored on ice until use.

Chromatin fragmentation

We used nuclei isolated from frozen mouse brain tissue as example. The extraction of nuclei follows the protocol prepared by Dr. John M. Graham with minor modifications. Then 0.25 µl of protease inhibitor cocktail (PIC) and 0.25 µl of 100 mM Phenylmethanesulfonyl fluoride (PMSF) were freshly added into 25 µl of nuclei suspension which contains 100-10,000 nuclei. The suspension was then mixed with 25 µl of 2x lysis buffer (100 mM tris (pH 8.0), 100 mM NaCl, 30 mM $MgCl_2$, and 4% Triton X-100), mixed by vortex, and incubated at room temperature for 10 min. 2.5 µl of 100 mM $CaCl_2$ and 5.5 µl of 10 U/ µl MNase were added into the mixture, mixed well, and incubated at room temperature for 10 min. 5.5µl of 0.5 M EDTA was added into the mixture, mixed well, and incubated on ice for 10 min. The mixture was centrifuged at 16,100 rcf for 5 min. The supernatant (~55 µl) was then transferred to a new 1.5 ml Eppendorf microcentrifuge tube and placed on ice, for use later.

MOWChIP

Our MOWChIP procedure consists of a number of steps. First, the microfluidic chambers of the device were filled up with IP buffer with the help of syringe pump at a flow rate of 200 µl/min. Then, the antibody coupled magnetic beads were flowed into the device via the combined force provided by pipette and a cylindrical permanent magnet.

After plugging the long tubing assembly (10 cm C-Flex clear tubing connected with 3 cm PFA tubing) containing chromatin sample into the inlet of the device, the sieve valve was partially closed by turn on the pressure control system, which applied 30 psi on the valve. The magnetic beads were manipulated by the magnet and packed against the valve to form a packed bed while injecting the chromatin solution into the device at a flow rate of 1.5 μl/min. Under this flow rate, the immunoprecipitation (IP) part was completed in ~50 min. After that, the long tubing assembly was disconnected with the device.

Oscillatory washing was conducted following the completion of IP. Low salt washing buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X)) was added into two short tubing assembly (6 cm C-Flex clear tubing connected with 3 cm PFA tubing), and then the tubing assemblies were plugged into the inlet and outlet of the microfluidic device unit. The PFA tubing of pressure control system was plugged into the other ends of the short tubing assemblies. Oscillatory washing parameters were set in a pre-written LabVIEW program. Both duration and interval were set as 0.5 s for the alternating pressure pulses operated by the two solenoid valves attached to the same device (In the LabVIEW program, enter 0.5 for both “inlet duration” and “outlet duration”). The number of oscillations was 300, which generated a total washing time of 5 min. The pressure was turned to ~0.65 psi by twisting the knob of pressure regulator. After finishing the washing with low salt washing buffer, low salt washing buffer was replaced by high salt washing buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X), and the process of washing was conducted again with the same washing parameter.

The magnetic IP beads were retained in the microfluidic chamber during oscillatory movement by using permanent magnet. After the accomplishment of oscillatory washing (low salt and high salt washing buffer), non-specifically adsorbed chromatin fragments were removed from the IP beads surface. Magnetic IP beads were then eluted from the microfluidic chamber to a new 1.5 ml Eppendorf microcentrifuge tube with IP buffer at a flow rate of 200 µl/min. The microcentrifuge tube was then placed on a magnetic rack to collect beads. Supernatant was removed and discarded by pipette. 90 µl of elution buffer (10 mM tris-HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA, and 0.03% (wt/v) SDS) and 4 µl of proteinase K (1 mg/ml) were added into the tube to resuspend the beads pellet. The tube was incubated in a multi-therm heat/cool shaker incubator at 65 °C for 2 hours. Phenol chloroform cleanup and ethanol precipitation were performed after incubation to extract and purify ChIP DNA. Finally, ChIP was resuspended in 10 µl of low EDTA TE buffer.

Construction of ChIP-seq libraries

Chromatin fragments were prepared by using micrococcal nuclease (MNase) to digest sorted and concentrated nuclei (NeuN+/NeuN-) following a published protocol. 54 µl of chromatin fragments (from 10,000 nuclei) was used in each ChIP assay. Chromatin immunoprecipitation was carried out using multiplexed MOWChIP assay with anti-H3K4me3 (39159, Active Motif) and anti-H3K27ac (39135, Active Motif) antibody. ChIP-seq libraries were prepared using Accel- NGS 2S Plus DNA Library kit (Swift Biosciences) from the purified immunoprecipitated DNA. The library preparation was conducted without knowledge of the brain sample or the type of histone mark. Minor modification was made to the manufacturer's procedures as detailed below. In the

amplification step, instead of adding 10 µl of low EDTA TE buffer into each reaction, we added the mixture of 7.5 µl of low EDTA TE buffer and 2.5 µl of 20X Evagreen dye to monitor and quantify PCR amplification. The reaction was stopped when the sample's fluorescence intensity increases by 3000 relative fluorescence units (RFU). Then, 50 µl of the mixture after PCR amplification was transferred into an Eppendorf tube and mixed with 37.5 µl of SPRIselect beads. After 5-min incubation at room temperature, the beads were conducted cleanup procedure with 80% ethanol. In the end, the DNA library was eluted from the beads into 7 µl of low EDTA TE buffer.

Construction of RNA-seq libraries

100 µl of sorted nuclei suspension (NeuN+ or NeuN-) from brain tissue, containing ~12,000 nuclei for producing 2 replicate libraries, was used for RNA extraction by using the RNeasy Mini Kit (74104, Qiagen) and RNase-Free DNase Set (79254, Qiagen), following the manufacturer's instruction. Half of the extracted mRNA (from 6,000 nuclei) in 30-µl volume was concentrated by ethanol precipitation and resuspended in 4.6 µl of RNase-free water. mRNA-seq libraries were prepared using Smart-seq230 and a Nextera XT DNA Library Preparation kit (FC-131-1024, Illumina) following the protocol and the manufacturer's instructions with minor modification. Around 2 ng of mRNA (in 4.6 µl of water) was mixed with 2 µl of 100 mM oligo-dT primer and 2 µl of 10 mM dNTP mix. After being denatured at 72 °C for 3 min, the mRNA solution was immediately placed on ice. Then, 11.4 µl of reverse transcript mix [1 µl of SuperScript II reverse transcriptase (200 U/ml), 0.5 µl of RNAse inhibitor (40 U/ml), 4 µl of Superscript II first-strand buffer, 1 µl of DTT (100mM), 4 µl of 5 M Betaine, 0.12 µl of 1 M MgCl2, 0.2 µl of TSO (100 mM), 0.58 µl of nuclease-free water] was mixed with the mRNA solution and

the mixture was incubated at 42 °C for 90 min, followed by 10 cycles of (50 °C for 2 min, 42 °C for 2 min). The reaction was finally inactivated at 70 °C for 15 min. 20 µl of first-strand mixture was then mixed with 25 µl of KAPA HiFi HotStart ReadyMix, 0.5 µl of (100 mM) IS PCR primers, 0.5 µl of Evagreen dye, and 4 µl of nuclease-free water. Generated complementary DNA (cDNA) was amplified by incubated at 98 °C for 1 min, followed by 9-11 cycles of (98 °C 15 s, 67 °C 30 s, 72 °C 6 min). After PCR amplification, 50 µl of PCR mix was purified by using 50 µl of SPRIselect beads. About 600 pg of purified cDNA was used for Nextera XT library preparation. ChIP-seq and RNA-seq library fragment size were measured by using high sensitivity DNA analysis kit (5067-4626, Agilent) on a TapeStation system (2200, Agilent). After this, 18-22 ChIP-seq and RNAseq libraries were randomly pooled together. Around 15 and 11 million reads were allocated to each ChIP-seq and RNA-seq library, respectively. The concentration of each library was examined by a KAPA library quantification kit (KK4809, Kapa Biosystems), and then the quantified libraries were pooled at 10 nM. The libraries were sequenced by Illumina HiSeq 4000 with single-end 50-nt read.

ChIP-seq data processing

Raw ChIP-seq reads and input data, were mapped to human genome (GRCh38) using BWA (0.7.7-r441). Peaks were called using MACS2 (2.1.0.20150420.1) using a p-value cutoff of 0.01 for the narrow marks (H3K4me3 and H3K27ac).

RNA-seq data processing

The human genome (GRCh38) and comprehensive gene annotation were obtained from GENCODE (v29). RNA-seq reads were mapped with STAR (2.7.0f) and quantified with RSEM (v1.3.1) using the default parameters.

Differential analysis for ChIP-seq data

The peaks were called using MACS271. The peaks with p-value<0.05 were taken as input for DiffBind R package. We first created cell-type-specific consensus peak sets using Diffbind for neurons and glia separately. Using the function of dba.peakset in diffbind, “majority-rules” was applied to generate consensus peak sets for both schizophrenia and control groups (i.e., a consensus peak must be present in more than half of the replicate datasets) and then the two consensus peak sets for each group were combined into a master set for analysis. The raw read counts were extracted using the function of dba.count in diffBind, and DESeq2 R package was used to perform the differential peaks analysis based on the TMM normalized reads to identify differential peaks between schizophrenia and control cohort (FDR<0.05). The following covariates were regressed out: age at death, sex, PMD and diagnosis. We annotated enhancers (defined as H3K4me3^{low}+H3K27ac^{high}) to genes using published Hi-C data on neurons and glia when possible and the rest of the enhancers were associated with their nearest genes. We annotated H3K4me3 peaks to genes when they overlapped with the promoter regions. We also generated voom plots using raw binding affinity matrix in H3K4me3 and H3K27ac and gene expression files to validate that low-expressed peaks/genes were filtered before downstream analysis. We generated smoothly decreasing curves fitted to the square root of residual standard deviation by average expression in all cases.

Differential analysis for RNA-seq data

The genes with less than 10 reads in over 70% of the samples were removed before differential analysis. The raw read counts for the rest of genes were taken as input for DESeq2 for differential analysis. The following covariates were regressed out: age at death, sex, PMD and diagnosis. The genes with FDR<0.05 were identified as differential expressed genes.

Taiji pipeline

Active regulatory elements were first identified via the overlap of high confidence peaks from H3K27ac with known gene promoter regions (4kbp upstream and 1kbp downstream of the transcription start sites). The distal H3K27ac peaks were assigned to active promoters using the unsupervised learning method EpiTensor, and assigned as an enhancer-promoter interaction if one locus overlapped with the distal peak and the other locus in the pair overlapped with a known promoter. Putative TF binding motifs were curated from the CIS-BP database. Using FIMO's algorithm. TFs were identified as having binding sites within 150-bp regions centered around H3K27ac peak summits. 58 unique NeuN+ (29 schizophrenia and 29 control) and 58 unique NeuN- (29 schizophrenia and 29 control) network topologies were thus constructed byf orming directed edges between TF and their regulatees, if the TF had a predicted binding site in the gene's promoter or linked enhancer.

EpiSig analysis

To integrate H3K27ac, H3K4me3 and RNA-seq data from two cell types across the postmortem frontal cortex samples from schizophrenia subjects and controls, EpiSig

was employed. This algorithm detects the significant signals from sequencing data in 5kb bins across the whole genome, and then clusters the regions based on the similar epigenomic profiles across all samples.

EpiSig differential enrichment analysis

A hypergeometric test was applied to all EpiSig clusters to assess the enrichment of differential H3K27ac and H3K4me3 peaks and differentially expressed genes. Clusters with FDR < 0.05 were selected and then ranked according to the number of overlapping peaks for each mark. Peaks were then mapped to genes using GREAT with default settings.

Age correlation analysis

Raw expression, pairwise expression and pairwise TF PPR age correlations were calculated using the Pearson R correlation. Significance was assessed by calculating p-values for the Pearson R correlations using the t-distribution with n-2 degrees of freedom for the respective cohort.

5.4 Results and Discussion

Epigenome and transcriptome profiling in neuronal and glial nuclei from frontal cortex of schizophrenia subjects and controls

Frontal cortex is a brain region involved in processes affected in schizophrenia patients, such as perception, cognition and sensorimotor gating[141]. We selected bilateral frontal cortex (Brodmann area 9) gray matter from 58 brain samples (29 schizophrenia and 29 controls). Control subjects were individually matched based on sex and age, and

to a lower degree on postmortem delay (or PMD – time between death and tissue sample collection). Nuclei were FACS-sorted using an anti-NeuN antibody as a marker of neuronal nuclei, and NeuN-positive (NeuN+) and NeuN-negative (NeuN-) nuclei (approximately 60,000 nuclei per NeuN+ or NeuN- sample) were collected for ChIP-seq (10,000 nuclei per library) and RNA-seq (6,000 nuclei per library) (Fig. 5-1a). After library preparation and sequencing, our MOWChIP-seq technology generated high-quality ChIP-seq data with average unique reads of ~11 million and ~14 million on histone modifications H3K27ac and H3K4me3, respectively. These yields were comparable to those in our previous studies using mouse frontal cortex and mammalian tissue culture samples[13,142]. We also generated saturation curves to validate that our sequencing depth is sufficient, and that a further increase in the sequencing depth would not lead to significantly more called peaks. MOWChIP-seq datasets have very low background noise[142] with the fraction of reads in called peaks (FRiP) averagely at 13.2% and 25.2% for our H3K27ac and H3K4me3 profiling, respectively. Using Phantompeakqualtools[143], we also calculated the normalized strand cross-correlation (NSC) and relative strand cross-correlation (RSC) to demonstrate the enrichment of sequencing reads around the histone modification sites. The average NSC was 1.15 and 1.23 for H3K27ac and H3K4me3, respectively; and the average RSC was 3.75 and 2.09 for H3K27ac and H3K4me3, respectively. These NSC and RSC values were higher than the recommended thresholds of 1.05 and 1.0, respectively.

We also compared our QC metrics with those of previously published ChIP-seq data in postmortem human frontal cortex samples. This previous study had an average NSC of 1.33 and 2.81, and RSC of 1.18 and 1.06 for H3K27ac and H3K4me3,

respectively. Thus, our data quality is very comparable to the state-of-the-art. The average Pearson's correlations between replicates were 0.923 and 0.971 for H3K27ac and H3K4me3, respectively, which compare well with those of ENCODE data[144]. Our RNA-seq datasets had an average of uniquely mapped reads of ~ 6.8 million, and the average mapping rate was 82.5%, within the recommended range of 70-90%[145]. Our average GC content was 40.73% and exon percentage was 36.52%.

Using all frontal cortex samples from the 29 schizophrenia subjects and 29 controls, we analyzed the expression of selected neuron- and glia-specific marker genes. Highly significant (median p-value = 6×10^{-7}) pair-wise differences in molecular marker expression were observed for all markers ranging from mature, functional, and synaptic neuron markers to astrocyte, oligodendrocyte and microglial markers (Fig. 5-1b) – confirming neuronal and glial cell-type identities in the NeuN+ and NeuN- nuclei samples, respectively.

To gain insight into the cell type-specific epigenomic changes associated with schizophrenia, we profiled histone marks H3K27ac and H3K4me3 and transcriptomes in neuronal and glial nuclei of the frontal cortex of schizophrenia subjects and controls. MOWChIP-seq data were mapped to the reference genome (Grch38), and significant peaks were called using MACS2[70]. We then performed the overlap analysis to identify consensus peak sets for H3K27ac and H3K4me3 in neuronal and glial nuclei (see Methods section), following "majority-rules" (i.e., consensus peaks must be present in at least half of the datasets). We identified 24257 consensus peaks covering ~78Mb (2.43% of the genome) for H3K27ac in neurons, 23646 consensus peaks covering ~76Mb (2.37% of the genome) for H3K27ac in glia, 36928 consensus peaks covering

~95 Mb (2.96% of the genome) for H3K4me3 in neurons, and 42284 consensus peaks covering ~101Mb (3.14% of the genome) for H3K4me3 in glia. We then investigated the differences in histone modification profiles and gene activity between the schizophrenia and control cohorts. We regressed our covariates including sex, age at death, PMD, and antemortem diagnosis[146]. Our analysis revealed 7750 differential peaks in H3K27ac, 2917 differential peaks in H3K4me3 and 2853 differentially expressed genes (DEGs) in neurons, while 5410, 3451 and 1346 differential peaks/differentially expressed genes were discovered in H3K27ac, H3K4me3 and glia, respectively (Fig. 5-1c). We also defined differential H3K27ac peaks that have no overlap regions with promoters and differential enhancers. We next leveraged the previously identified promoter-anchored chromatin loops in neurons and glia to identify how differential enhancers are linked with genes[147]. We successfully associated 4741 and 3393 cell-type-specific enhancers to 3464 and 2724 genes via enhancer-promoter interactions in neurons and glia, respectively. The rest of the enhancers were associated with their nearest genes. We discovered that the schizophrenia group has varied H3K27ac/H3K4me3 levels and RNA-seq reads compared to controls at various loci involved in processes related to synaptic plasticity and cognitive processes or previously associated with schizophrenia risk4 – these included increased H3K27ac at CACNB2, RGS4 and NRXN1 in neurons, increased H3K4me3 at GRIN2A in neurons, decreased H3K27ac at MACF1 and CPEB4 in neurons, and decreased BDNF mRNA in glia (Figs. 5-1d,e)

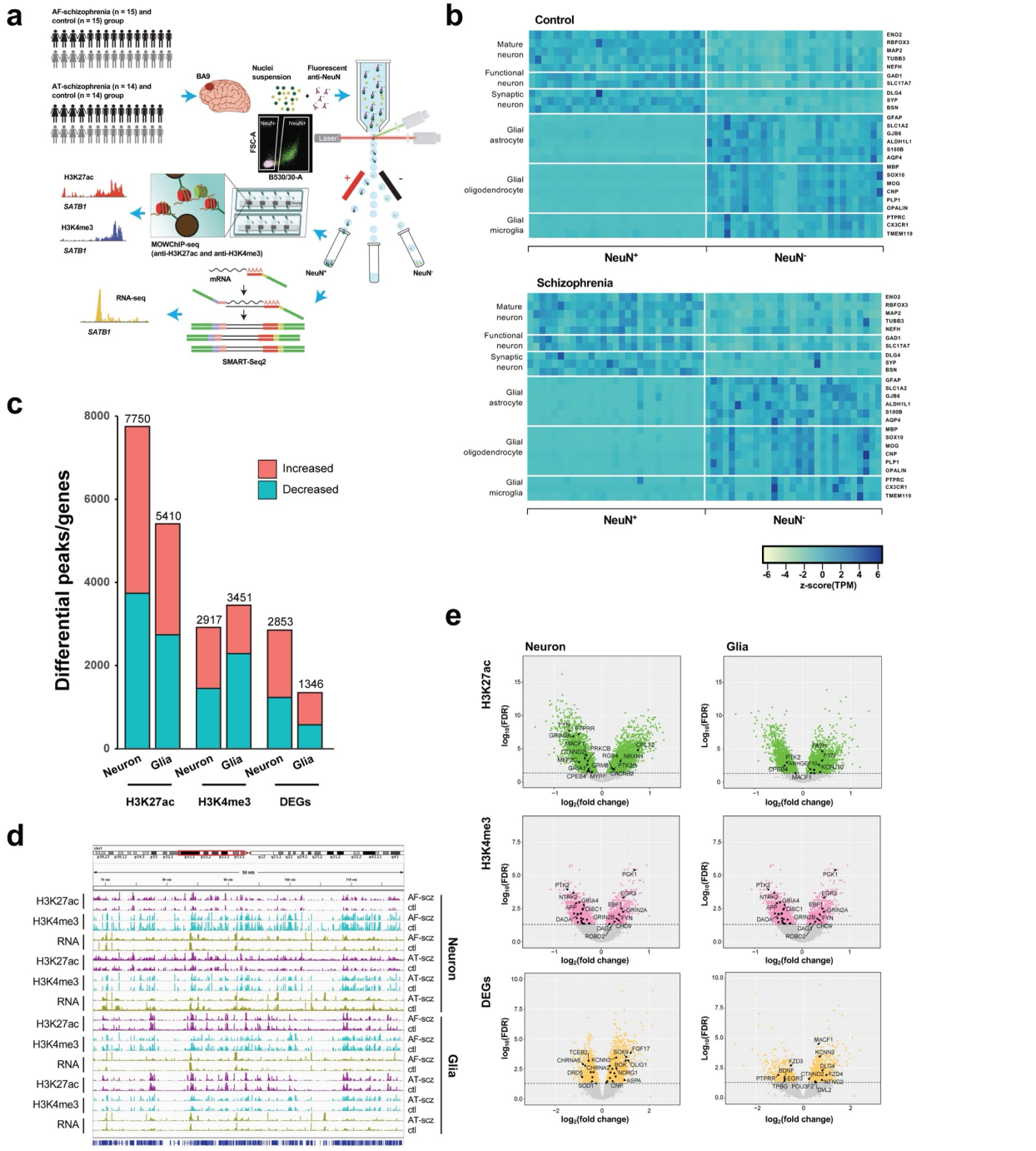

Figure 5-1 Comparison of epigenomic and transcriptomic landscapes in the frontal cortex of schizophrenia subjects and controls. (a) Overview of the experimental design starting from postmortem human frontal cortex samples to generate cell-type-specific H3K27ac, H3K4me3 and RNA profiles. (b) Heatmap of the expression of neuronal and

glial cell markers across all NeuN+ and NeuN- frontal cortex samples from 29 control subjects and 29 schizophrenia subjects. (c) Differential histone modification peaks and DEGs obtained by comparing schizophrenia (n = 29) and controls (n = 29). The differential peaks or DEGs were identified using FDR of less than 0.05. d, Exemplar genomic track view of H3K27ac, H3K4me3 and RNA signals for matched AF-schizophrenia/control and AT-schizophrenia/control pairs in NeuN+ and NeuN- cells. 50Mb region displayed: chr1:68,000,000-118,000,000 (GRCh38). e, Volcano plots showing genes associated with differential H3K27ac and H3K4me3 peaks and DEGs. Candidate genes for schizophrenia or genes involved in significant GO terms are labeled. The horizontal lines indicate FDR of 0.05.

To further explore which differential epigenetic modifications and DEGs were associated with SNP-heritability, we overlapped our differential peak/gene sets with loci previously associated with schizophrenia risk[117]. For neurons, we identified 49 (H3K27ac), 27 (H3K4me3) and 34 (DEGs) overlapping regions which in total mapped to 81 unique genes. These genes were found to be enriched in the GO biological process term “Nervous System Development” (p-value 2.51×10^{-6}) and included the well-known super-enhancer gene RERE[148], along with the genes SATB2 and BCL11B, which are critical in forebrain neuron differentiation. For glia, we found 33 (H3K27ac), 23 (H3K4me3) and 11 (DEGs) overlapping regions which in total mapped to 46 unique genes. Enriched pathways included “Regulation of TP53 Activity through Acetylation” (p-value = 3.48×10^{-4}). Further, we ran functional enrichment analysis for the 19 “double-evidence” overlapping genes from our data set which returned the pathway

"Formation of the beta-catenin: TCF transactivating complex" (p-value = 1.10×10^{-4}) along with two chromatin organization pathways.

To assess agreement with the literature, we compared the DEGs identified in our study with a previous single-nucleus RNA sequencing (snRNA-seq) study in postmortem prefrontal cortex of schizophrenics and controls. Importantly, 694 out of our 2853 neuronal DEGs, and 73 out of our 1346 glial DEGs were also identified in this previous single-cell dissection work. In neurons, several genes encoding metabotropic glutamate receptors (GRM5, GRM7) that are directly associated with schizophrenia risk[149] were found differentially expressed in both studies. We also identified some novel genes, including SOX9, which is associated with neurodevelopmental processes, and POU3F2, which is viewed as a key regulator of gene expression in a schizophrenia-associated gene co-expression module[150].

We also overlapped genes identified from differential H3K27ac/H3K4me3 peaks with DEGs from RNA-seq (Fig. 5-2). We found that 369 and 170 genes were identified in both the DEGs from RNA-seq and differential H3K27ac peaks associated genes in neurons and glia, respectively. Among these, several schizophrenia-associated genes were also detected, including FYN, which is known to be involved in neuronal signal transduction[151], LRRTM3, which plays an important role in excitatory synapse development[152], and GRIN3A, a gene that encodes NMDA receptor subunits in neuronal nuclei[153], and NTNG2, which acts as the axon growth guidance during neuronal development[154].

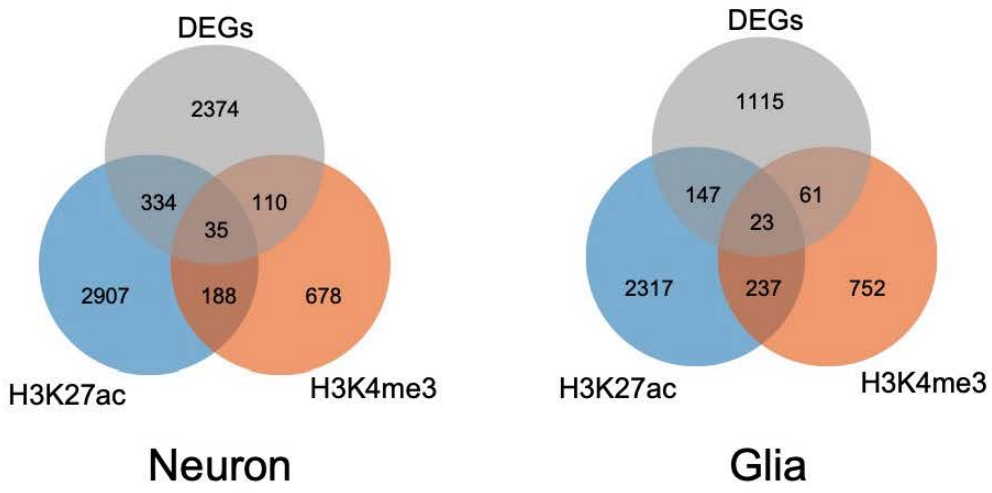

Figure 5-2 Venn diagrams on the relationship among genes associated with differential H3K4me3 or H3K27ac peaks and DEGs.

For integrative analysis of these diverse epigenomic and transcriptomic data, we next employed an unbiased method called EpiSig[155], which combines the epigenomic and transcriptomic dataset into a single analysis to cluster regions with similar epigenomic profiles across all the neuronal and glial nuclei samples. 85,462 signal enriched regions were grouped into 814 epigenomic clusters covering 14.53% of the genome. These clusters were further combined into 6 groups (sections) using the K-means method (Fig. 5-3a). Section I had high coverage in the gene annotations for intron (35%) and intergenic regions (29%) indicating inactive regions. It was also enriched in chromosome X compared to other sections. Section II was annotated as enhancers that are active in neurons but suppressed in glia. A hypergeometric test identified clusters that were significantly enriched in schizophrenia vs control differential histone marks and differentially expressed genes (Fig. 5-3b). The top 5 Section II clusters had schizophrenia positive association (i.e., activity and expression

schizophrenia > controls) in genes enriched in GO terms “Trans-synaptic signaling” (FDR 4.66 x 10^{-2}). Section III was highly enriched in enhancers (average of 21% of all regions in each cluster), and low coverage in intergeneric regions (12%), which is likely associated with active enhancers for both neurons and glia given the high signal strength for H3K27ac. Using the top differentially enriched cluster genes enriched in the GO term, “Amyloid fibril formation” (FDR 5.99 x 10^{-2}) was found to be negatively associated with schizophrenia in neurons, whereas genes enriched in the GO term “Neuron projection development” (FDR 3.19 x 10^{-2}) were positively associated with schizophrenia in glia (Fig. 5-3b). Section IV also had high coverage of enhancers (Fig. 5-3b). However, it had the highest average promoter content with 16% of all gene annotations being promoter regions, further supported by CpG islands showing the highest proportion (20%) in this section; indicating active promoters for both neurons and glia. Both cell types showed enrichment in respiratory electron transport genes that were negatively associated with schizophrenia (Fig. 5-3b). As for the average differential signals across sections, a great variance was observed. For example, the signal of H3K4me3 was higher in schizophrenia subjects compared to controls for neuronal nuclei in section V, while it was lower in schizophrenia subjects for glial samples (Fig. 5-3b). Finally, Section VI was annotated as enhancers that were active in glia but repressed in neurons (Fig. 5-3b).

Transcriptional regulatory processes proceed as a hierarchy of orchestrated events that ultimately modulate the expression of downstream target genes. Using the recently developed Taiji algorithm[156], which allows access to information pertaining to transcriptional cascades deriving from upstream drivers through specific pathway

mechanisms to downstream effects, we integrated epigenomic and transcriptomic data to construct 116 individual transcriptional networks in neuronal and glial nuclei from schizophrenia subjects and controls. We identified active promoters and enhancers using H3K27ac and then predicted transcription factor (TF) binding sites by scanning 1,165 TF motifs linking putative TF binding sites to their targets using EpiTensor[157], an unsupervised method to predict enhancer-promoter associations. TFs were subsequently ranked according to regulatory importance using the Personalized PageRank (PPR) algorithm for each unique network topology[158]. Using the differentially expressed TFs (schizophrenia vs controls FDR < 0.05), TFs were ranked by absolute change in schizophrenia vs control PPR score (Fig. 5-3c – top 10 TFs for each cell type). Of the top 10 neuronal TFs, all were found to be cell-type specific. Using the top 4 TFs, we identified 207 regulatees the were regulated by 3 or more TFs and found they were involved in processes such as “Neurexins and neuroligins” (FDR 2.18×10^{-7}) and “Protein-protein interactions at synapses” (FDR 1.22×10^{-6}) (Fig. 5-3d). Furthermore, all top 10 glia TFs were cell-type specific TFs and the regulatees of the top 4 TFs were enriched in signaling pathways including “RAF/MAP kinase cascade” (FDR 2.80×10^{-2}) and “RHO GTPase cycle” (FDR 3.07×10^{-2}) (Fig. 5-3e).

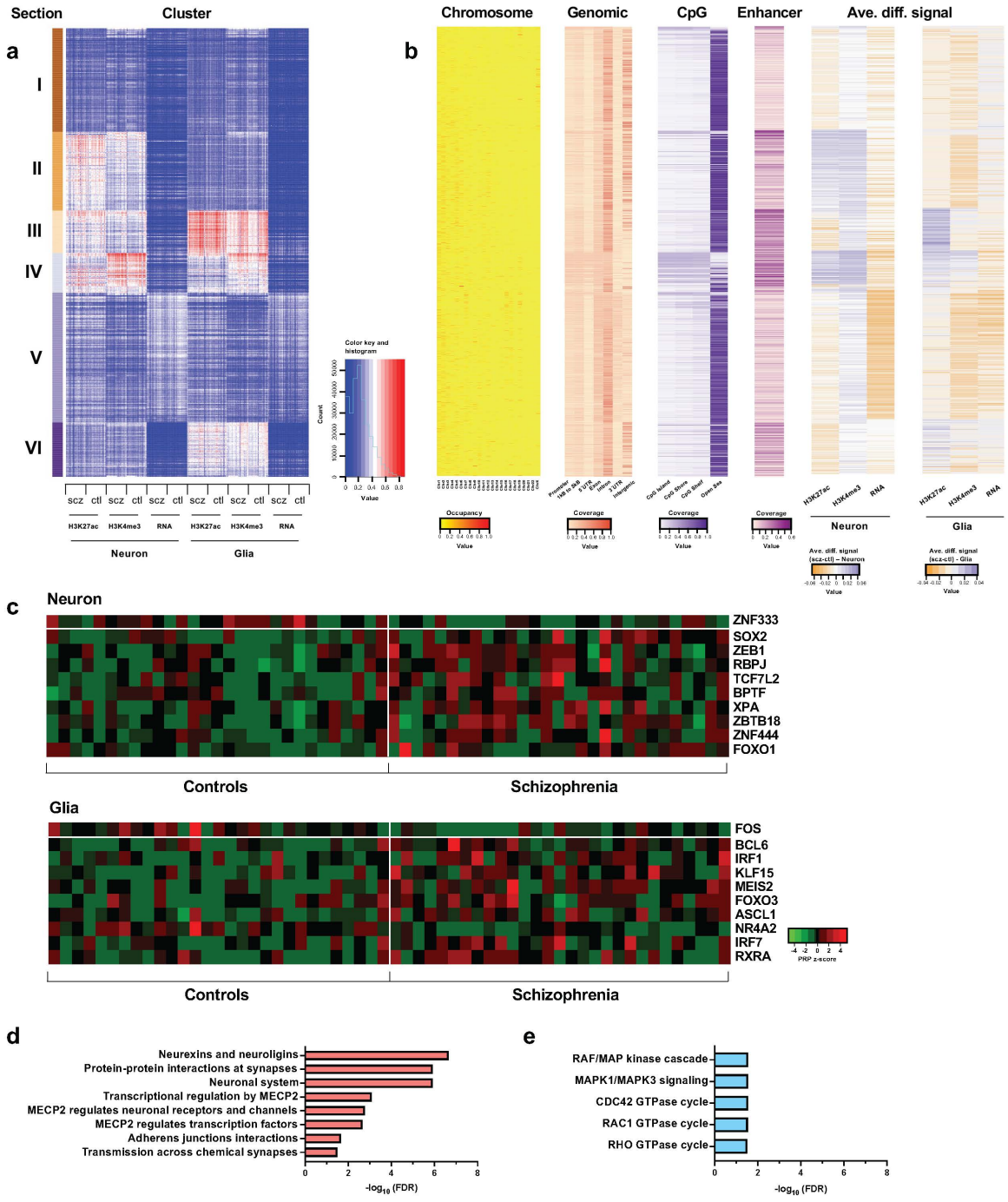

Figure 5-3 Genome-wide multidimensional clusters in the frontal cortex of schizophrenia subjects and controls. (a) Integration analysis using EpiSig. 814 EpiSig clusters across 348 genome-wide sequencing datasets were grouped into 6 sections. The heatmap shows the signal in each EpiSig cluster (row: EpiSig cluster; column:

marker) (b) For each EpiSig cluster, from left to right, the heatmaps are: the region percentage in each chromosome; the genomic annotation; the CpG annotation; the percentage of enhancer; the difference signal between schizophrenia and controls in neuronal and glial nuclei. (c) Heatmap of z-score PPR for top 10 significantly differentially expressed TFs (FDR < 0.05) ranked by absolute change in PPR for neuronal (upper panel) and glial (lower panel) schizophrenia vs control nuclei samples. (d) Overrepresented pathway analysis (FDR < 0.05) for 203 downstream regulatees common to the top 4 schizophrenia vs control neuron-specific TFs (ZNF333, SOX2, ZEB1 and RBPJ). (e) Overrepresented pathway analysis (FDR < 0.05) for 225 downstream regulatees common to the top 4 schizophrenia vs control glial-specific TFs (FOS, BCL6, IRF1 and KLF15).

Alterations in antipsychotic-free but not in antipsychotic-treated schizophrenics

Using preclinical models, it has been suggested that chronic antipsychotic drug administration leads to long lasting changes in frontal gene expression and chromatin organization[138,139], but the epigenomic consequences of antipsychotic treatment in postmortem human brain samples remain largely unexplored. To validate the separation between the antipsychotic-free (AF) schizophrenia and antipsychotic-treated (AT) schizophrenia groups, we first utilized a dimension reduction algorithm – uniform manifold approximation and projection (UMAP) – to visualize the clustering of each sample with TMM normalized binding affinity matrices or gene expression files from MOWChIP-seq and RNA-seq, respectively. The separation between AF-schizophrenia and AT-schizophrenia groups is clearly visible in H3K4me3 and gene expression

profiles for both neurons and glia (Fig. 5-4).

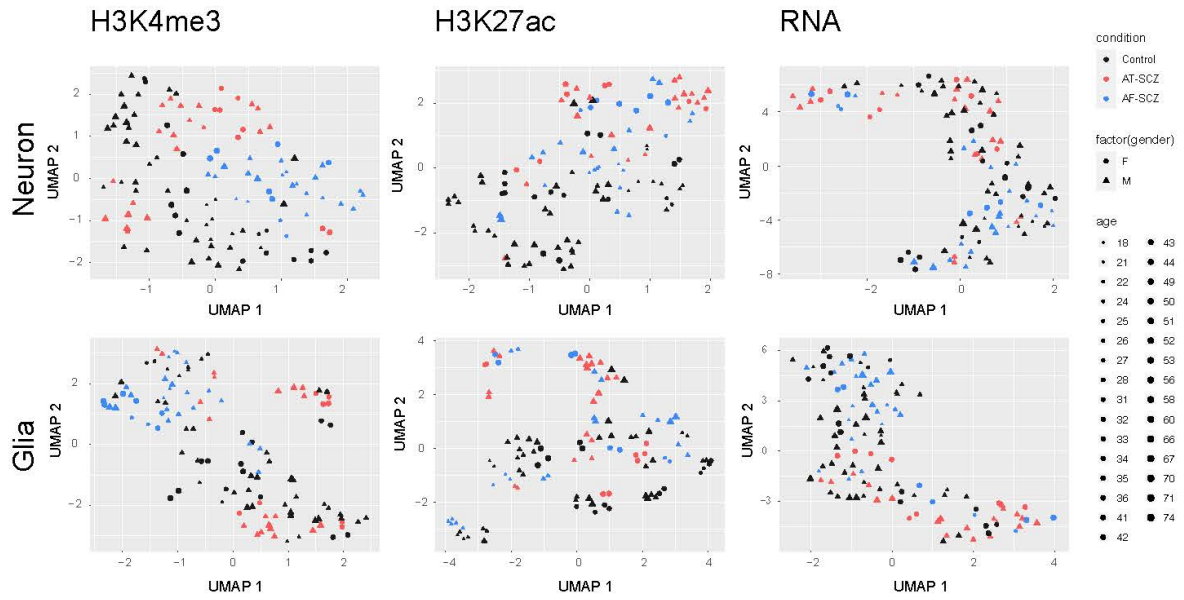

Figure 5-4 UMAP visualization of the feature matrix of differential peaks/genes on H3K4me3, H3K27ac and RNA-seq data among AF-schizophrenia, AT-schizophrenia and control subjects

To further determine the functional relevance of antipsychotic treatment, we aimed to identify the biological pathways dysregulated in the AF-schizophrenia group that were also reversed to control levels in the AT-schizophrenia group as compared to individually matched controls. We calculated the average pairwise difference in PageRank in neuronal nuclei from the AF-schizophrenia/control pair cohort. In conjunction with this, we also calculated the average pairwise difference in PageRank in neuronal nuclei from the AT-schizophrenia/control pair cohort. We then identified those TFs with a difference in these two values greater than 0.5 (Fig. 5-5a). However, when these TFs were further filtered based on a significance cut-off of FDR < 0.05, no significant TFs were identified to be simultaneously changed in the AF-

schizophrenia/control group and not changed in the AT-schizophrenia/control group (Fig. 5-5c). In parallel, AF-schizophrenia/control DEGs were integrated with genes from the AT-schizophrenia/control cohort with no case/control differences in expression resulting in a list of 116 cohort-specific DEGs (Fig. 5-5c). Functional enrichment analysis of these genes resulted in pathways involved in glutamatergic neurotransmission including “Activation of AMPK downstream of NMDARs” (FDR 3.66 x 10^{-3}) (Fig. 5-5g). Structural and functional modifications of dendritic spines are central to brain development and plasticity[159]. Studies from postmortem brains of subjects with neurodevelopmental disorders including schizophrenia demonstrate altered density and morphology of dendritic spines, particularly in the frontal cortex[160,161]. IQGAP scaffold proteins facilitate the formation of complexes that regulate cytoskeletal dynamics including microtubules[162]. Interestingly, another significant pathway restored in the AT-schizophrenia group was “Rho GTPases activate IQGAPs” (FDR 3.66 x 10^{-3}) (Fig. 5-5g). The importance of this pathway was validated by the analysis of the clusters from the EpiSig pipeline. Thus, taking the top 3 clusters ranked for enrichment in H3K27ac and mapping their differential peaks to genes resulted in 166 genes enriched in pathways including “Adherens junctions interactions” (p-value 1.22 x 10^{-4}).

Autophagy has been suggested to play an important role in the pathophysiology of schizophrenia and antipsychotics are known to modulate the process[163]. Notably, the pathways “Aggrephagy” (FDR 3.66 x 10^{-3}) and “Macroautophagy” (FDR 7.71 x 10^{-3}) were significantly enriched (Fig. 5-5g). Expression of the “Macroautophagy” genes ATG16L2, AMBRA1, PRKAB1, TUBA1A, TUBB2A, and TUBA4A was restored in the AT-schizophrenia group (Fig. 5-5i). Ubiquitin B (UBB) expression has been previously

identified as a strong correlate of schizophrenia symptoms[164]. We show a 1.96-fold decrease specific to AF-schizophrenics compared to controls (p-value = 4.0×10^{-2}).

Glial cells modulate and act as effectors in neurodevelopment through a wide range of neuronal-glial cell interactions. Using the same process as above, we identified 2 driver TFs with significant change in PPR between case and control for the AF-schizophrenia cohort and no significant difference between case and control in the AT-schizophrenia group, SOX11 (FDR 1.73×10^{-2}) and MGA (FDR 9.83×10^{-3}) (Fig. 5-5b). 77 downstream DEG regulatees of these two TFs were identified in the AF cohort showing significant regulatory case/control change (Fig. 5-5d). In parallel, 21 cohort-specific DEGs were identified as AF-schizophrenia/control DEG and having no significant difference in expression in the AT-schizophrenia/control cohort (Fig. 5-5d). Functional enrichment analysis of the union of 153 genes included “Post NMDA receptor activation events (FDR 3.65×10^{-2}), and “Long-term potentiation” (FDR 1.95×10^{-2}) (Fig. 5-5h). EpiSig’s analysis did not show glial alterations in the AF-schizophrenia

cohort.

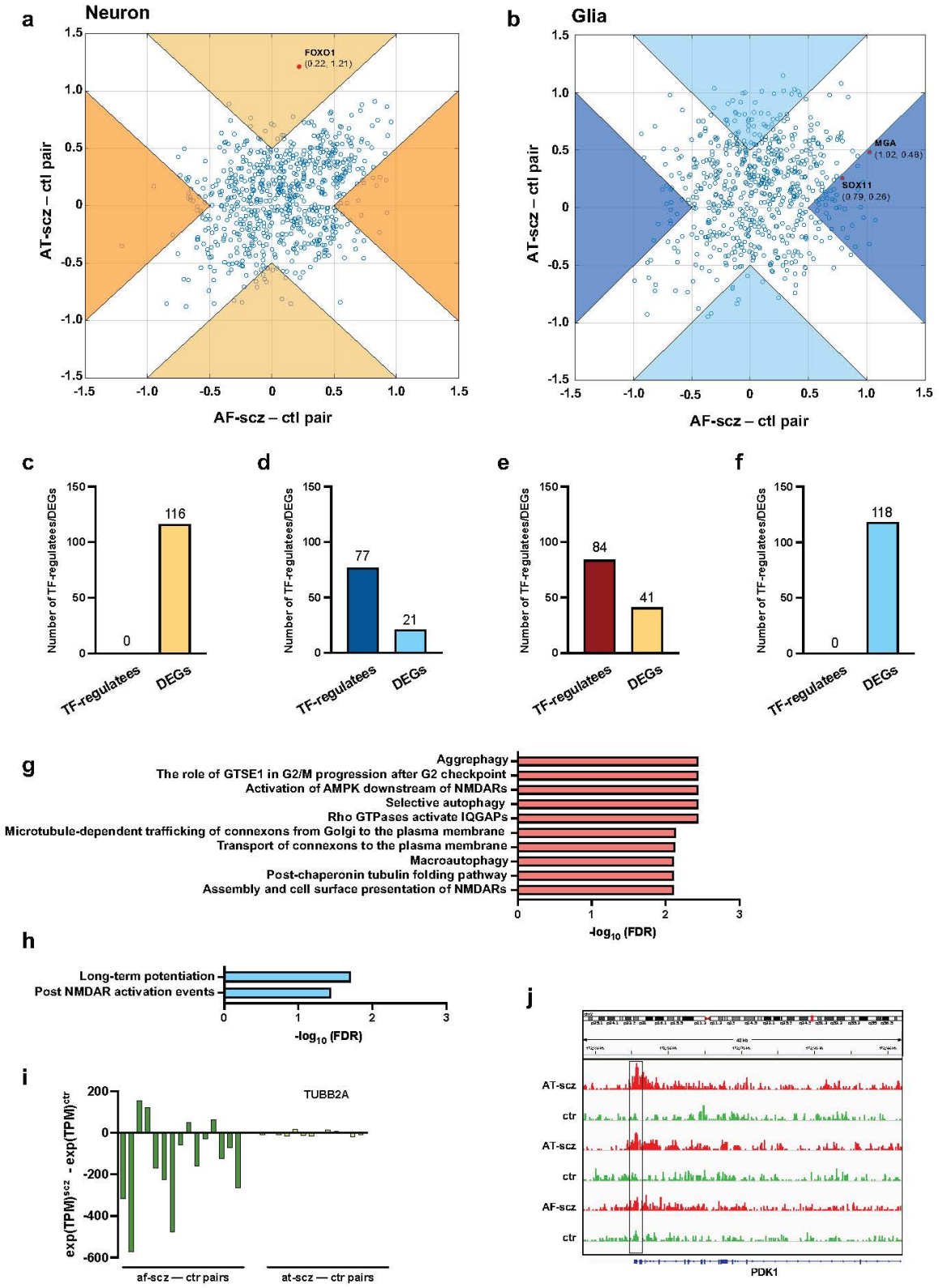
a
Neuron
FOXO1
(0.22, 1.21)
AT-scz – ctl pair
AF-scz – ctl pair
b
Glia
MGA
(1.02, 0.48)
SOX11
(0.79, 0.26)
AT-scz – ctl pair
AF-scz – ctl pair
c
Number of TF-regulatees/DEGs
116
0
TF-regulatees
DEGs
d
77
21
e
84
41
f
118
0
g
Aggrephagy
The role of GTSE1 in G2/M progression after G2 checkpoint
Activation of AMPK downstream of NMDARs
Selective autophagy
Rho GTPases activate IQGAPs
Microtubule-dependent trafficking of connexons from Golgi to the plasma membrane
Transport of connexons to the plasma membrane
Macroautophagy
Post-chaperonin tubulin folding pathway
Assembly and cell surface presentation of NMDARs
-log10 (FDR)
h
Long-term potentiation
Post NMDAR activation events
-log10 (FDR)
i
TUBB2A
exp(TPM)scz - exp(TPM)ctr
af-scz — ctr pairs
at-scz — ctr pairs
j
AT-scz
ctr
AT-scz
ctr
AF-scz
ctr
PDK1

Figure 5-5 Epigenomic alterations affected by antipsychotic treatment. (a) Scatter plot of average pairwise change in PPR ($PPR_{schizophrenia} – PPR_{control}$) for AF vs AT neuronal cohorts. Orange regions show cohort (AF – AT) (*i.e.,* alterations recovered by antipsychotic treatment), whereas beige regions show cohort (AT – AF) > 0.5 (*i.e.,* alterations consequence of antipsychotic treatment). TFs FDR < 0.05 highlighted in red. (b) Scatter plot of average pairwise change in PPR ($PPR_{schizophrenia} – PPR_{control}$) for AF vs AT glial cohorts. Dark blue regions show cohort (AF – AT) *(i.e.,* alterations recovered by antipsychotic treatment), whereas cyan regions show cohort (AT – AF) > 0.5 (*i.e.,* alterations consequence of antipsychotic treatment). TFs FDR < 0.05 highlighted in red. (c) Number of DEG regulatees by TFs, and number of DEGs in neuronal nuclei from AF-schizophrenia/control pairs. (d) Number of DEG regulatees by TFs, and number of DEGs in glial nuclei from AF-schizophrenia/control pairs. (e) Number of DEG regulatees by TFs, and number of DEGs in neuronal nuclei from AT-schizophrenia/control pairs. (f) Number of DEG regulatees by TFs, and number of DEGs in glial nuclei from ATschizophrenia/control pairs. (g) Functional enrichment analysis of union of genes from AFschizophrenia/control pairs in neuronal nuclei. (h) Functional enrichment analysis of union of genes from AF-schizophrenia/control pairs in glial nuclei. (i) Pairwise expression difference (schizophrenia – control) of an exemplar AF-schizophrenia/control cohort DEG (TUBB2A) across all 29 schizophrenia-control pairs in neuronal nuclei. (j) H3K27ac tracks for PDK1 (member of the 84 gene set in e) in neuronal nuclei. Box highlighting the FOXO1 DNA-binding motif in promoter at position chr2: 172,555,706 – 172,555,718 (GRCh38). Two exemplar ATschizophrenia/control

cohort pairs showing differential H3K27ac peak intensity around motif locus and an example AF-schizophrenia/control cohort patient pair with no difference.

We also performed differential analysis with covariates (sex, age at death, PMD, and antemortem diagnosis) regressed out on AF-schizophrenia/control and AT-schizophrenia/control cohorts. On neurons, the results revealed 5342 and 6624 differential peaks in H3K27ac and H3K4me3, respectively, and 534 DEGs between AF-schizophrenia and their controls, while 1165, 10149 and 3760 differential peaks/genes in H3K4me3 and H3K27ac, and DEGs were discovered between AT-schizophrenia and controls. On glia, we identified 1870, 1409, and 161 differential peaks/genes between AF-schizophrenia and controls; 5136, 2733, 2773 differential peaks/genes between AT-schizophrenia and controls, in H3K27ac, H3K4me3, and DEGs, respectively. More differential peaks/genes were detected between AT-schizophrenia and their matched controls than those between AF-schizophrenia and their controls with the exception in neuronal H3K4me3.

Similar to our TF analyses in Fig. 5-5 (see above), we also identified the genes altered in the AF-schizophrenia/control group but not in the AT-schizophrenia/control group using differential analyses of enhancers, promoters, or gene expression. It should be noted that in the differential analyses here, the schizophrenia subjects (whether AF or AT) and their controls were compared at the cohort level, while matched schizophrenia/control pairs were examined individually in the TF-based analyses. At the epigenomic level, we identified 1093 and 1376 genes changed in the AF- but not in AT-schizophrenics by examining differential enhancers and promoters, respectively. These genes were linked to epigenomic features restored to their basal level after treatment. In

glia, there were 419 and 305 recovered genes linked with differential enhancers and promoters, respectively. At the transcriptomic level, 329 DEGs in neurons and 128 in glia were discovered in AF-schizophrenia/control comparison but not in the AT-schizophrenia/control differential analysis. 35 out of the 329 DEGs in neurons overlapped the AT-recovered genes discovered in the epigenomic differential analyses (either by differential enhancer or promoter).

Alterations in antipsychotic-treated but not in antipsychotic-free schizophrenics

We next sorted to characterize those TFs that exhibit regulatory alterations in the AT-schizophrenia/control cohort but not in the AF-schizophrenia/control cohort. Our goal was to identify modifications in pathways that represent a consequence of antipsychotic medication rather that an epigenetic mark of schizophrenia in postmortem human brain. In the same way as above, we identified those TFs with a change in the AT-schizophrenia/control group but not in the AF-schizophrenia/control group (Fig. 5-5a). Further filtering of these TFs based on a significance cut-off of FDR < 0.05 lead to the identification of FOXO1 (FDR 4.89×10^{-2}).

We identified dysregulated AT-schizophrenia/control DEG regulatees of these TFs in neuronal nuclei via analysis of differential edge weights thus obtaining 84 genes (Fig. 5-5e). AT-schizophrenia/control DEGs were intersected with genes from the AF-schizophrenia/control cohort with no case/control differences in expression resulting in a list of 41 cohort-specific DEGs (Fig. 5-5e). Pathway analysis on the union of genes yields the pathway “Regulation of p53 activity through phosphorylation” (FDR 1.13×10^{-2}), including the FOXO1 AT-schizophrenia/control cohort DEG regulatees CCNA1, BLM,

TP53RK, and RBBP8, and the AT-schizophrenia/control cohort-specific DEGs PRKAA1 and TAF15. The p53 regulatory gene PDK1 was also identified as a FOXO1 AT-schizophrenia/control DEG regulate (Fig. 5-5j). p53 is one of the most critical pro-apoptotic genes, and antipsychotics are known to produce complex effects including the activation of both proapoptotic and antiapoptotic signaling pathways[165]. Our data showed that all genes involved in the regulation of p53 presented a significantly lower expression in AT-schizophrenics compared to controls, suggestive of a repressive role for FOXO1 for its 5 regulatees, as FOXO1 had higher PPR in treated schizophrenics and was also 2.38-fold more highly expressed in schizophrenics for the treated-cohort (p-value = 1.07×10^{-2}). Analysis of the H3K4me3 enriched clusters from the EpiSig pipeline for the AT-schizophrenia/control cohort corroborated alterations in pathways related to p53[166].

In glial cells, no TFs showed a significant difference in the AT-schizophrenia/control cohort but not in the AF-schizophrenia/control cohort (Fig. 5-5f). Furthermore, the 118 genes cohort-specific DEGs identified as AT-schizophrenia/control DEGs and having no significant difference in expression in the AF-schizophrenia/control cohort were not significantly enriched in any signaling pathway (Fig. 5-5f). Analysis of the clusters from the EpiSig pipeline remarked the importance of the RHO GTPase pathway on the regulatory alterations observed in AT-schizophrenia subjects. However, examination of differential peaks/genes that were present in the AT-schizophrenia/control group but not in the AF-schizophrenia/control group did not result in meaningful enriched pathways.

We also used differential analyses of enhancers, promoters, and expression to identify the genes altered in the AT-schizophrenia/control group but not in the AF-schizophrenia/control group. In neurons, we found 2153 and 168 treatment-altered genes linked with differential enhancers and promoters, respectively. In glia, we identified 1978 and 557 altered genes linked with differential enhancers and promoters, respectively. At the transcriptomic level, 3555 DEGs in neurons and 2740 in glia were discovered in AT-schizophrenia/control comparison but not in the AF-schizophrenia/control differential analysis. 323 out of the DEGs in neurons overlapped the genes discovered in the epigenomic differential analyses with changes in AT- but not in AF-schizophrenics (299 with differential enhancers and 28 with differential promoters), including GRIA3 that was linked to chronical clozapine treatment[167]. We then performed enriched pathway analysis on the 323 altered genes from neurons identified by both epigenomic and transcriptomic differential analyses. Several notable pathways were associated with ERK/MAPK, which is a signaling pathway affected by antipsychotic agents[168] and involved in antipsychotic treatment resistance[169].

Age differentially affects antipsychotic-treated schizophrenia subjects

In order further to assess the effect of age on gene expression, we first compared transcriptomes of subjects with schizophrenia and the controls to evaluate how these changes correlated with age. Within neurons in the control group, we identified 742 genes that were significantly correlated with age – with most of them (573, or 77.2%) showing decreased expression in older control subjects (Fig. 5-6a). The opposite, however, was observed in neurons from schizophrenia subjects with 18 out of 622

(2.8%) in AF-schizophrenia presenting a negative correlation with age, an effect that was partly reversed in the AT-schizophrenia cohort (85 out of 242 or 35.1%) (Fig. 5-6a).

Our data also demonstrate that within the glial genes correlated with age (1031, 389 and 351 in controls, AF-schizophrenia and AT-schizophrenia, respectively), approximately half (491 or 47.6%) were positively correlated with age in the control group whereas a much higher fraction of genes showed increased expression with age in the schizophrenia group, particularly in the AF-schizophrenia cohort (382 or 98.2% in AF-schizophrenia, and 276 or 78.6% in AT-schizophrenia) (Fig. 5-6b). These results suggest that age differentially affects gene expression in the frontal cortex of AF-schizophrenia vs AT-schizophrenia subjects as compared to age-matched controls. Importantly, this was further confirmed by functional integration of epigenomic and transcriptomic data and the evaluation of how these alterations correlated with age.

Thus, we evaluated pairwise changes in expression between schizophrenia subjects and their age-matched controls, and identified 206 and 310 genes with an absolute Pearson's correlation of > 0.50 in neuronal nuclei from the AF-schizophrenia/control and AT-schizophrenia/control cohorts, respectively (Figs. 5-6c,d). We also found enriched biological processes associated with age, including "Regulation of protein kinase activity" (p-value 6.69×10^{-7}) in the AT-schizophrenia/control group (Fig. 5-6e). Within this gene set, the difference between AT-schizophrenia subjects and control pairs correlated either positively (WNK1) or negatively (SFRP2) with age (Figs. 5-6f,g). Evaluation of pairwise changes in PPR identified 48 TFs with high correlations to age in neuronal nuclei from AT-schizophrenia/control cohorts, whereas this alteration was not observed in the AF-schizophrenia/control group (Figs. 5-6c,d). Pathway

analysis of the neuronal TFs affected by age in the AT-schizophrenia/control cohort led to the top pathway "NGF-simulated transcription" (p-value 8.04 x 10^{-8}), including the TFs EGR2 and ATF2. Hallucinations and delusions typically attenuate with aging[170], which is consistent with the lower PPR difference for EGR2 – a preclinical marker of psychosis-like behavior[100] – that we observed in older subjects (Fig. 5-6h).

In glial nuclei, 147 and 88 genes were identified with AF-schizophrenia/control and AT-schizophrenia/control expression difference vs age correlations of > 0.60, respectively (Figs. 5-6i,j). Enriched pathways in the AT-schizophrenia/control group included: "Degradation of DVL" (p-value 4.04 x 10^{-5}) and "Beta-catenin independent WNT signaling" (p-value 5.06 x 10^{-4}) (Fig. 5-6k). Since dysfunctional WNT signaling is associated with several CNS disorders including Alzheimer's[171], together, these data also suggest that this positive correlation between glial gene differences in AT-schizophrenia subjects/control pairs and age (Fig. 5-6k) may be responsible for some of the negative effects of antipsychotic treatment on cognitive processes. We also identified 11 and 53 TFs correlated with age in the AF-schizophrenia/control and AT-schizophrenia/control cohorts, respectively (Figs. 5-6i,j). However, as in neuronal nuclei, the effect of age became more evident in the AT-schizophrenia/control group with age-related adaptations in glial TF-affected pathways that included "Signaling by NOTCH" (p-value 3.2 x 10^{-4}).

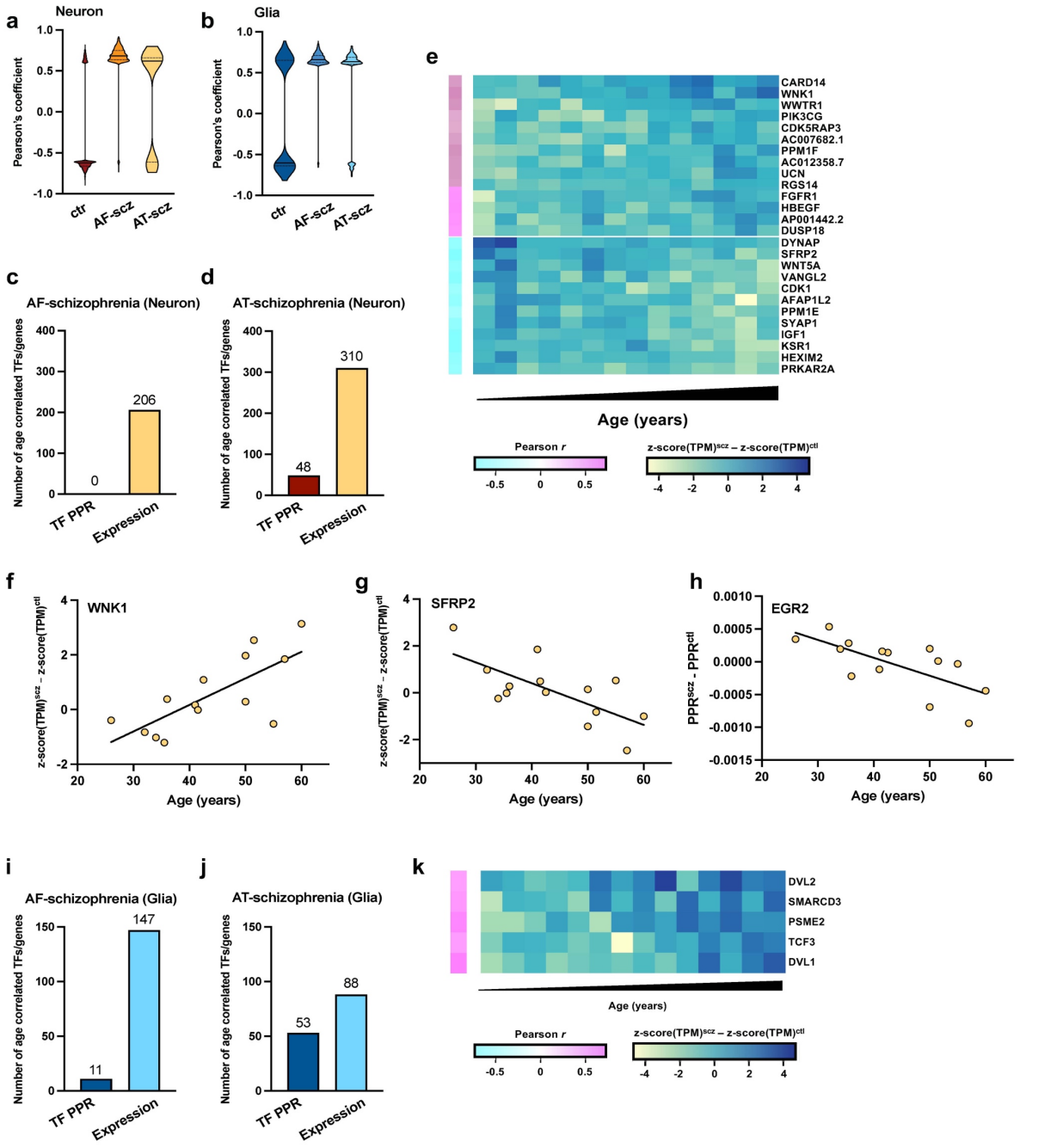

Figure 5-6 Epigenomic effect of age on treated schizophrenia subjects. (a) Violin plots for Pearson's R correlation coefficients of age vs expression for 742, 622 and 242 genes from control, AF-schizophrenia, and AT-schizophrenia neuronal nuclei, respectively. (b) Violin plots for Pearson's R correlation coefficients of age vs

expression for 1031, 389 and 351 genes from control, AF-schizophrenia, and AT-schizophrenia glial nuclei, respectively (c) Number of pairwise TF PPR and pairwise gene expression differences correlated with age in neuronal nuclei from AF-schizophrenia/control pairs. (d) Number of pairwise TF PPR and pairwise gene expression differences correlated with age in neuronal nuclei from AT-schizophrenia/control pairs. (e) Heatmap for the 14 age positively-correlated (schizophrenia – control increase with age) and 12 age negatively-correlated (schizophrenia – control decrease with age) genes of the significant GO term "Regulation of kinase activity" from the AT-schizophrenia/control neuronal cohort. (f) Example gene, WNK1 pairwise expression difference (schizophrenia - control) vs age (Pearson's R = 0.73; p-value = 0.003). (g) Example gene, SFRP2 pairwise expression difference (schizophrenia – control) vs age (Pearson's R = -0.70; p-value = 0.005). (h) Example TF, EGR2 pairwise PPR difference (schizophrenia – control) vs age (Pearson's R = 0.69; p-value = 0.0003 (i) Number of pairwise TF PPR and pairwise gene expression differences correlated with age in glial nuclei from AF-schizophrenia/control pairs. (j) Number of pairwise TF PPR and pairwise gene expression differences correlated with age in glial nuclei from AT-schizophrenia/control pairs. (k) Heatmap for the 5 age positively-correlated (schizophrenia – control increase with age) genes of the significant GO term "Beta-catenin independent WNT signaling" from the ATschizophrenia/control glial cohort.

5.5 Conclusions

Understanding the molecular determinants involved in schizophrenia is critical for devising new treatment strategies and the discovery of the pathogenic mechanisms

underlying this psychiatric condition. In the largest ChIP-seq/RNA-seq study in postmortem brain samples from schizophrenia subjects and controls, our data suggest cell-type specific epigenomic differences in individuals with schizophrenia as well as cellular alterations in signaling pathways potentially involved in either the elimination of schizophrenia-related epigenomic alterations upon antipsychotic drug treatment or the antipsychotic-dependent modulation of alternative epigenetic pathways previously unaffected in the untreated schizophrenia cohort. However, we cannot fully exclude the possibility of previous exposure to antipsychotic medications in the AF-schizophrenia group, and hence that the epigenetic alterations observed exclusively in the AF-schizophrenia group are a consequence of a potential period of decompensation, which typically occurs following voluntary treatment discontinuation[172]. It is also worth noting that our findings were established by examining the average characteristics of entire neuronal and glial fractions. Further studies of individual neuronal and glial cell subtypes may yield additional information on the role of cell-type-subpopulations[173,174].

6 Epigenomic and transcriptomic changes in offspring brain in response to maternal immune activation

6.1 Project Summary

A large body of epidemiological data has suggested that a serious maternal infection during a specific period of pregnancy works as a critical environmental insult that could significantly increase the risk of serval neurodevelopmental and psychiatric disorders in offspring. However, our knowledge of the molecular mechanisms underlying the deleterious effects of this environmental exposures on the developmental trajectory of the fetal brain remains largely unclear. Given the growing impact of the coronavirus disease 2019 (COVID-19), it is vital for us to investigate all of the potential epigenetic or transcriptomic dysregulation that may be involved in the process. In this study, we used influenza virus or vehicle solution to stimulate the pregnant mice and activate their immune system. To examine the prenatal and postnatal effect of maternal immune activation (MIA), we swapped half of the infant mice right after their birth. In that case, each mother would raise both MIA affected baby and MIA unaffected baby, no matter if themselves were exposed to the virus or not. We also applied the same strategy to Mock-only mice as negative control. Overall, this project provides the first dataset with cell-type-specific epigenomic and transcriptomic features in mice frontal cortex after affected by maternal immune activation and suggest that the prenatal effect of MIA is more significant than the postnatal effect from siblings.

6.2 Introduction

The development of brain is a delicate process that is regulated by both genetic and environmental factors; thus it is consequently sensitive to a lot of insults[175-177]. Environmental stimuli during an unfortunate time period could severely change the common developmental trajectory of brain and then cause several neurodevelopmental disorders, such as schizophrenia, autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD)[178-184]. Maternal infection is one of these environmental insults[181,185]. Several early serological studies, like Prenatal Determinants of Schizophrenia study[123,186], examined if the prenatal infection during gestation would increase the risk of schizophrenia in the offspring. Recent epidemiological studies in understanding the etiology of some neurodevelopmental disorders, like schizophrenia, also indicates that prenatal exposure to infection, such as influenza virus, and rubella virus, is associated with an elevated risk of psychotic diseases in the adult offspring.

Animal models have been used to further examine the link between prenatal infection and the development of various neurodevelopmental diseases. Traditionally, the synthetic double-stranded RNA analogue polyinosinic-polycytidylic acid (Poly (I:C))[187,188], the bacterial mimic lipopolysaccharide (LPS)[189], and strains of influenza[190-194] are three commonly used immunogenic agents for inducing maternal immune response during pregnancy in animal models of maternal immune activation (MIA) These animal models were known to generate some abnormalities in behavior and cognition that are similar to the symptoms of schizophrenia and autism[187,195]. While the first two artificial immunogens-based models present some specific advantage, such as time-wise limited and well-defined immune response that allow the onset and intensity

of MIA to be more precisely controlled, influenza-based models are still valuable to the field because they could elicit a complete spectrum of immune responses. Importantly, these rodent models of MIA established the concept that it is the maternal cytokine response, but not any specific pathogen, that works as a risk factor for neurodevelopmental disorders in the offspring[196]. These models also suggested that the key role in the mechanism of MIA is the increased level of pro-inflammatory cytokines such as interleukin-6 (IL-6), IL-17a, tumor necrosis factor (TNF) α, IL-10, and IL-1β in the amniotic fluid of pregnant dams and fetal brain[197-201]. In normal trajectory of brain development, these immune cytokines play an important role throughout neurodevelopment and are also involved in the organization of neural circuits by the regulation of synaptic pruning, plasticity, and transmission. In addition, multiple studies have independently shown that these cytokines might affect epigenetic machinery. For example, it is suggested that IL-6 can activate DNA methyltransferase 1 (DNMT1) and IL-17 inhibits HDAC activity[202]. These findings provide potential interaction between MIA and epigenetic regulation.

Supporting this concept, previous studies reported changes in chromatin structure and accessibility in tissue samples from MIA affected mice offspring. Most of these reports, however, were focused on DNA methylation dysregulation in brain, for example Labouesse et. al., demonstrated the hypermethylation of promoters at GAD1 and GAD2[203]. Using the assays for Western blotting and ChIP-qPCR in bulk tissue homogenates of brain samples, only a few genes showed difference in acetylation patterns of histones at proximal promoter regions[204]. Although interesting, these approaches do not provide the same chromatin state information as histone

modifications profiled by ChIP-seq assays do. Histone modifications, including histone H3 acetylation of lysine 27 (H3K27ac) and histone H3 trimethylation of lysine 4 (H3K4me3) are critically involved in epigenomic regulations; H3K27ac marks active enhancers, whereas H3K4me3 marks active promoters. Enhancers are highly dynamic cis-regulatory elements with known involvement in neurodevelopmental processes[133], and the dynamics of promoters are also significantly connected with the genetic risk of certain psychiatric conditions[134]. However, few studies have been conducted about potential cell-type-specific genome-wide variations in covalent histone modifications in frontal cortex of mice offspring. As an example, a previous investigation conducting H3K4me3 ChIP-seq assays and mRNA profiling by microarray failed to identify any significant epigenetic changes in mature cerebral cortex[205]. This might be because of the use of brain tissue homogenates or the mice model they used, in which they employed Poly (I:C) as the immunogenic agent.

Combining MOWChIP-seq[12,13] and SMART-seq for low-input profiling of H3K27ac and H3K4me3 histone modifications and transcriptomes, respectively, we studied the effect of maternal influenza viral infection, siblings, and cross-fostering on the epigenetic and transcriptomic landscapes in mice adult offspring in order to enhance our understanding about the molecular mechanisms underlying the association between increased risk of neurodevelopmental disorders and maternal viral exposure. We present the first dataset with cell-type-specific epigenomic and transcriptomic features in frontal cortex samples from mice offspring, MIA-treated or Mock-treated. Our aim was to explore the effect of maternal immune activation in the brain of offspring, and to

investigate if these dysregulation are prenatal, postnatal, or the mixture of the two through cross-fostering.

6.3 Materials and Methods

Experimental model and subject details

Experiments were started from pregnant CD-1 dams. The pregnant mice were purchased from Charles River Laboratory by our collaborator, Dr. Javier Gonzalez-Maeso at Virginia Commonwealth University. On GD 9.5, two of these pregnant dams were intranasally injected with influenza virus. After the birth, half of the mice offspring were cross fostered by a mother of opposite experimental condition and weaned once appropriate. Male mice offspring were sacrificed when they are ten-week-old, and their frontal cortices were harvested and snap frozen. All procedures were conducted in accordance with NIH guidelines and were approved by the Virginia Commonwealth University Animal Care and Use Committee. All efforts were made to minimize animal suffering and the number of animals used.

Nuclei isolation and sorting via FACS

Nuclei isolation was conducted using a published. All steps were conducted on ice, and all centrifugation was conducted at 4°C. One piece of mouse frontal cortex tissue (6-10 mg) was placed in 3 mL of ice-cold nuclei extraction buffer (NEB) [0.32 M sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 0.1 mM EDTA, 10 mM tris-HCl, and 0.1%(v/v) Trition X-100] with 30 µL of freshly added protease inhibitor cocktail (PIC, Sigma-Aldrich), 3 µL of 100 mM phenylmethylsulfonyl fluoride (PMSF, Sigma-Aldrich) in Isopropyl alcohol, 3 µL of 1 M dithiothreitol (DTT, Sigma-Aldrich), and 4.5 µL of

ribonuclease (RNase) inhibitor (2313A,Takara Bio). The tissue was homogenized in a tissue grinder (D9063, Sigma-Aldrich). The homogenate was filtered with a 40 μm cell strainer (22-363-547, Thermo Fisher Scientific) and collected in a 15 mL centrifuge tube. The cell suspension was centrifuged at 1000 RCF for 10 min. The supernatant was discarded, and the pellet was resuspended in 0.5 mL of ice-cold NEB with 5 μL of freshly added PIC, 0.5 μL of PMSF, 0.5 μL of DTT, and 0.75 μL of RNase inhibitor. 500 μL of the sample was mixed with 750 μL of 50%(w/v) iodixanol (made by mixing 4 mL of OptiPrep™ gradient (Sigma-Aldrich) and 0.8 mL of diluent [150 mM KCl, 30 mM $MgCl_2$, and 120 mM tris-HCl]). The mixture was centrifuged at 10,000 RCF for 20 min. Then, the supernatant was removed and 300 μL of 2%(w/v) normal goat serum (50062Z, Life technologies) in Dulbecco's PBS (DPBS, Life technologies) was added to resuspend the nuclei pellet. To separate NeuN+ and NeuN- fractions, 6 μL of 2 ng/μl anti-NeuN antibody conjugated with Alexa 488 (MAB377X, Millipore) in DPBS was added into the nuclei suspension. The suspension was mixed well and incubated at 4°C for 1 h on a rotator mixer (Labnet). After incubation, the sample was sorted into NeuN+ and NeuN- fractions using a BD FACSAria™ cell sorter (BD Biosciences). The sorted NeuN+ nuclei were directly used for RNA-seq experiment. 200 μL of sorted NeuN+ nuclei suspension, containing ~26,000 nuclei, was added into 800 μL of ice-cold PBS for ChIP-seq experiment. 200 μL of 1.8 M sucrose solution, 10 μL of 1 M $CaCl_2$, and 3 μL of 1 M $Mg(Ac)_2$ were added into the mixture. The solution was mixed well and incubated on ice for 15 min. Then, the sample was centrifuged at 1800 RCF at 4°C for 15 min. The supernatant was discarded and the pellet was resuspended in 60 μL of PBS with 0.6 μL of freshly added PIC and 0.6 μL of PMSF and stored on ice until use for ChIP-seq.

Chromatin fragmentation

We used nuclei isolated from frozen mouse brain tissue as example. The extraction of nuclei followed the protocol prepared by Dr. John M. Graham with minor modifications. Then 0.25 µl of protease inhibitor cocktail (PIC) and 0.25 µl of 100 mM Phenylmethanesulfonyl fluoride (PMSF) were freshly added into 25 µl of nuclei suspension which contains 100-10,000 nuclei. The suspension was then mixed with 25 µl of 2x lysis buffer (100 mM tris (pH 8.0), 100 mM NaCl, 30 mM $MgCl_2$, and 4% Triton X-100), mixed by vortex, and incubated at room temperature for 10 min. 2.5 µl of 100 mM $CaCl_2$ and 5.5 µl of 10 U/ µl MNase were added into the mixture, mixed well, and incubated at room temperature for 10 min. 5.5µl of 0.5 M EDTA was added into the mixture, mixed well, and incubated on ice for 10 min. The mixture was centrifuged at 16,100 rcf for 5 min. The supernatant (~55 µl) was then removed to a new 1.5 ml Eppendorf microcentrifuge tube and placed on ice, for use later.

MOWChIP

Our MOWChIP procedure consisted of a number of steps. First, the microfluidic chambers of the device were filled up with IP buffer with the help of syringe pump at a flow rate of 200 µl/min. Then, the antibody coupled magnetic beads were flowed into the device via the combined force provided by pipette and a cylindrical permanent magnet. After plugging the long tubing assembly (10 cm C-Flex clear tubing connected with 3 cm PFA tubing) containing chromatin sample into the inlet of the device, the sieve valve was partially closed by turn on the pressure control system, which applied 30 psi on the valve. The magnetic beads were manipulated by the magnet and packed against the valve to form a packed bed while injecting the chromatin solution into the device at a

flow rate of 1.5 µl/min. Under this flow rate, the immunoprecipitation (IP) part was completed in ~50 min. After that, the long tubing assembly was disconnected with the device.

Oscillatory washing was conducted following the completion of IP. Low salt washing buffer (20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X)) was added into two short tubing assembly (6 cm C-Flex clear tubing connected with 3 cm PFA tubing), and then the tubing assemblies were plugged into the inlet and outlet of the microfluidic device unit. The PCA tubing of pressure control system was plugged into the other ends of the short tubing assemblies. Oscillatory washing parameters were set in a pre-written LabVIEW program. Both duration and interval were set as 0.5 s for the alternating pressure pulses operated by the two solenoid valves attached to the same device (In the LabVIEW program, enter 0.5 for both “inlet duration” and “outlet duration”). The number of oscillations was 300, which generated a total washing time of 5 min. The pressure was turned to ~0.65 psi by twisting the knob of pressure regulator. After finishing the washing with low salt washing buffer, low salt washing buffer was replaced by high salt washing buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% (wt/v) SDS, and 1%(v/v) Triton-100X), and the process of washing was conducted again with the same washing parameter. The magnetic IP beads were retained in the microfluidic chamber during oscillatory movement by using permanent magnet. After the accomplishment of oscillatory washing (low salt and high salt washing buffer), non-specifically adsorbed chromatin fragments were removed from the IP beads surface. Magnetic IP beads were then eluted from the microfluidic chamber to a new 1.5 ml Eppendorf microcentrifuge tube with IP buffer at a

flow rate of 200 µl/min. The microcentrifuge tube was then placed on a magnetic rack to collect beads. Supernatant was removed and discarded by pipette. 90 µl of elution buffer (10 mM tris-HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA, and 0.03% (wt/v) SDS) and 4 µl of proteinase K (1 mg/ml) were added into the tube to resuspend the beads pellet. The tube was incubated in a multi-therm heat/cool shaker incubator at 65 °C for 2 hours. Phenol chloroform cleanup and ethanol precipitation were performed after incubation to extract and purify ChIP DNA. Finally, ChIPed DNA was resuspended in 10 µl of low EDTA TE buffer.

Constructure of ChIP-seq libraries

Chromatin fragments were prepared by using micrococcal nuclease (MNase) to digest sorted and concentrated neuronal nuclei following a published protocol. 54 µl of chromatin fragments (from 10,000 nuclei) was used in each ChIP assay. Chromatin immunoprecipitation was carried out using multiplexed MOWChIP assay with anti-H3K4me3 (39159, Active Motif) and anti-H3K27ac (39135, Active Motif) antibody. ChIP-seq libraries were prepared using Accel-NGS 2S Plus DNA Library kit (Swift Biosciences) from the purified immunoprecipitated DNA. The library preparation was conducted without knowledge of the brain sample or the type of histone mark. Minor modifications were made to the manufacturer's procedures as detailed below. In the amplification step, instead of adding 10 µl of low EDTA TE buffer into each reactoin, we added the mixture of 7.5 µl of low EDTA TE buffer and 2.5 µl of 20X Evagreen dye to monitor and quantify PCR amplification. The reaction was stopped when the sample's fluorescence intensity increases by 3000 relative fluorescence units (RFU). Then, 50 µl of the mixture after PCR amplification was transferred into an Eppendorf tube and

mixed with 37.5 µl of SPRIselect beads. After 5-min incubation at room temperature, the beads were conducted cleanup procedure with 80% ethanol. In the end, the DNA library was eluted from the beads into 7 µl of low EDTA TE buffer.

Construction of RNA-seq libraries

100 µl of sorted neuronal nuclei suspension from brain tissue, containing ~12,000 nuclei for producing 2 replicate libraries, was used for RNA extraction by using the RNeasy Mini Kit (74104, Qiagen) and RNase-Free DNase Set (79254, Qiagen), following the manufacturer's instruction. Half of the extracted mRNA (from 6,000 nuclei) in 30-µl volume was concentrated by ethanol precipitation and resuspended in 4.6 µl of RNase-free water. mRNA-seq libraries were prepared using SMART-seq v4 Ultra Low Input RNA kit (Clontech) and a Nextera XT DNA Library Preparation kit (FC-131-1024, Illumina) following the protocol and the manufacturer's instructions with minor modification. Around 600 pg of purified cDNA was used for Nextera XT library preparation. ChIP-seq and RNA-seq library fragment size were measured by using high sensitivity DNA analysis kit (5067-4626, Agilent) on a TapeStation system (2200, Agilent). After this, 18-22 ChIP-seq and RNAseq libraries were randomly pooled together. Around 15 and 11 million reads were allocated to each ChIP-seq and RNA-seq library, respectively. The concentration of each library was examined by a KAPA library quantification kit (KK4809, Kapa Biosystems), and then the quantified libraries were pooled at 10 nM. The libraries were sequenced by Illumina HiSeq 4000 with single-end 50-nt read.

ChIP-seq data processing

Once we got the sequencing data back, we conducted the pre-processing pipeline first. The procedure included read trimming (the sequencing reads were trimmed by Trim Galore! with default settings), filtering low-quality reads (SAMTools -view -bq 10), duplicated reads removal (SAMTools rmdup), and subtracting out the blacklisted regions defined by ENCODE (BEDTools subtract). After the pre-processing, uniquely mapped reads from ChIP and input samples were extended to 250 bp from their original length (50 bp) and stored in BED files. MACS2 was used to finish the peak calling step ($q< 0.05$). For visualization in IGV (Broad Institute), the signal was calculated in 100bp windows over the entire genome and output as a bigWig file. Datasets with poor quality (the number of identified peaks less than 15,000, mapping rate less than 80%, and the value of the fraction of reads in peaks (FRiP) less than 10%) were discarded.

RNA-seq data processing

Sequencing reads were trimmed under default settings using Trim Galore!. The trimmed reads were mapped to GRCm38 genome by hisat2. Mapped bam files were imported into SeqMonk v1.47.1 (Babraham Institute). Datasets with poor quality (percentage of reads aligned to exons < 40%) were discarded. FeatureCounts was used to count reads and DEGs were determined by pairwise comparison using DESeq2 (FDR < 0.05). The threshold of FDR < 0.05 was used to be consistent with that used in differential enhancers.

Differential analysis for ChIP-seq data

The peaks were called using MACS2. The peaks with q-value<0.05 were taken as input for DiffBind R package. We first created a consensus peak sets using Diffbind for

ChIP-seq data at H3K4me3 and H3K27ac separately. Using the function of dba.peakset in diffbind, “majority-rules” was applied to generate consensus peak sets for all the experimental groups (i.e., a consensus peak must be present in more than half of the replicate datasets) and then the seven calculated consensus peak sets, associated with seven conditions, were combined into a master set for analysis. The raw read counts were extracted using the function of dba.count in diffBind, and DESeq2 R package was used to perform the differential peaks analysis based on the TMM normalized reads to identify differential peaks between schizophrenia and control cohort (FDR<0.05). The following covariates were regressed out: batch effect, number of peaks, number of unique reads, and sequencing depth. For H3K27ac data, we considered H3K27ac^{high} regions without any overlap to the promoter regions to be enhancer regions. We mapped enhancers to genes using H3K27ac signal and RNA-seq expression data as per previously published. First, enhancers found in any one of the experimental groups were combined into one master enhancer list. For every H3K27ac sample, RPKM was calculated for each of the enhancers. Similarly, FPKM was calculated for each gene in the RNA-seq data. Next, mouse cortex association data was used to determine the regions in which enhancers would operate. For each topologically associating domain (TAD), we determined what enhancers and genes would fall within the TAD. Then, the Spearman correlation coefficient would be calculated between each enhancer and each gene, and the gene with the highest correlation would be assigned to the enhancer. To determine significance, the enhancer was then correlated against every other gene on the same chromosome to obtain a statistical distribution that could be used in conjunction with the R function pnorm to obtain a p-value. In addition, an empirical p-

value was obtained by calculating what fraction of pairings with other genes had a higher correlation than the current gene. Enhancer-gene pairings were considered significant if the SCC > 0.25 and both p-values < 0.05. We annotated H3K4me3 peaks to genes when they overlapped with the promoter regions. We also checked voom plots using raw binding affinity matrix in H3K4me3 and H3K27ac and gene expression files to validate that low-expressed peaks/genes were filtered before downstream analysis. We generated smoothly decreasing curves fitted to the square root of residual standard deviation by average expression in all cases.

Differential analysis for RNA-seq data

The genes with less than 10 reads in over 70% of the samples were removed before differential analysis. The raw read counts for the rest of genes were taken as input for DESeq2 for differential analysis. The following covariates were regressed out: sequencing depth, batch effect, exon percentage, and number of unique reads. The genes with FDR<0.05 were identified as differential expressed genes.

Taiji pipeline

Active regulatory elements were first identified via the overlap of high confidence peaks from H3K27ac with known gene promoter regions (4kbp upstream and 1kbp downstream of the transcription start sites). The distal H3K27ac peaks were assigned to active promoters using the unsupervised learning method EpiTensor and assigned as an enhancer-promoter interaction if one locus overlapped with the distal peak and the other locus in the pair overlapped with a known promoter. Putative TF binding motifs were curated from the CIS-BP database. Using FIMO's algorithm. TFs were identified

as having binding sites within 150-bp regions centered around H3K27ac peak summits. Network topologies were thus constructed by forming directed edges between TF and their regulatee genes, if the TF had a predicted binding site in the gene's promoter or linked enhancer.

6.4 Results and Discussion

Malfunction of prefrontal cortex region is viewed as the key feature of most main neurodevelopmental disorders, like bipolar disorder, ASDs, ADHD, and schizophrenia. In that case, we chose to examine the epigenetic and transcriptomic features of the frontal cortex of mice offspring involved in this study. There are seven experimental groups of offspring by administering the influenza virus or vehicle solution to isogeneic CD-1 mice on GD 9.5 and swapping the mice offspring between MIA-treated or Mock-treated dams (Fig. 6-1). As shown in the figure, Groups A, B, C, and D followed the common procedure of cross-fostering were designed to study the prenatal and postnatal effect of MIA. Offspring in Groups A and C were born to saline solution treated Mock mothers, while mice in Group B and D were born to influenza virus infected MIA mothers. Within 24 hours of birth, mice in Groups B and C were swapped so that a Mock mother reared mice in Groups A and B and a MIA mother reared mice in Groups C and D. Groups E and F were generate using a similar procedure of cross-fostering without influenza viral infection, while mice in Group G were not involved in cross-fostering of any form. To get insight into the cell-type-specific epigenomic alterations linked to the prenatal and postnatal effect of MIA, we generated profiles of histone modification H3K4me3 and H3K27ac and transcriptomes in neuronal nuclei of the prefrontal cortex of the mice offspring in these seven groups. Nuclei were FACS-sorted

using an anti-NeuN antibody as a marker of neuronal nuclei, and NeuN-positive ($NeuN^+$) nuclei (approximately 60,000 nuclei) were collected for ChIP-seq (~10,000 nuclei per assay) and RNA-seq (~6,000 per assay). We totally generated 324 ($n_{H3K4me3}$ = 115, $n_{H3K27ac}$ = 95 and n_{RNA} = 114) ChIP-seq genome-wide profiles of H3K27ac and H3K4me3 and RNA-seq datasets that passed our stringent quality control test from 60 mice offspring. After library preparation and sequencing, our MOWChIP-seq technology generated high-quality ChIP-seq data with average unique reads of ~7.6 million and ~8.4 million on histone modifications H3K27ac and H3K4me3, respectively. These yields were comparable to those in our previous studies using mouse frontal cortex and mammalian tissue culture samples. MOWChIP-seq datasets have very low background noise with the fraction reads in identified peaks (FRiP) averagely at 21.4% and 41.7% for our H3K27ac and H3K4me3 profiling, respectively. Using Phantompeakqualtools, we also calculated the normalized strand cross-correlation (NSC) and relative strand cross-correlation (RSC) to demonstrate the enrichment of sequencing reads around the histone modification sites. The average NSC was 1.15 and 1.4 for H3K27ac and H3K4me3, respectively; and the average RSC was 4.93 and 2.08 for H3K27ac and H3K4me3, respectively. These values of NSC and RSC were higher than the ENCODE recommended thresholds of 1.05 and 1.0, respectively. Notably, the ChIP-seq datasets within each experimental condition showed high Pearson's correlation coefficient between technical and biological replicates, with the mean correlation coefficients of 0.916 and 0.974 for H3K27ac and H3K4me3, respectively, which compare well with those of ENCODE date. Our RNA-seq datasets had an average of ~10.62 million mapped reads, and the average mapping rate was 93.59%, higher than the

recommended range of 70-90%. Our average GC content was 41.73% and exon percentage was 57.43%.

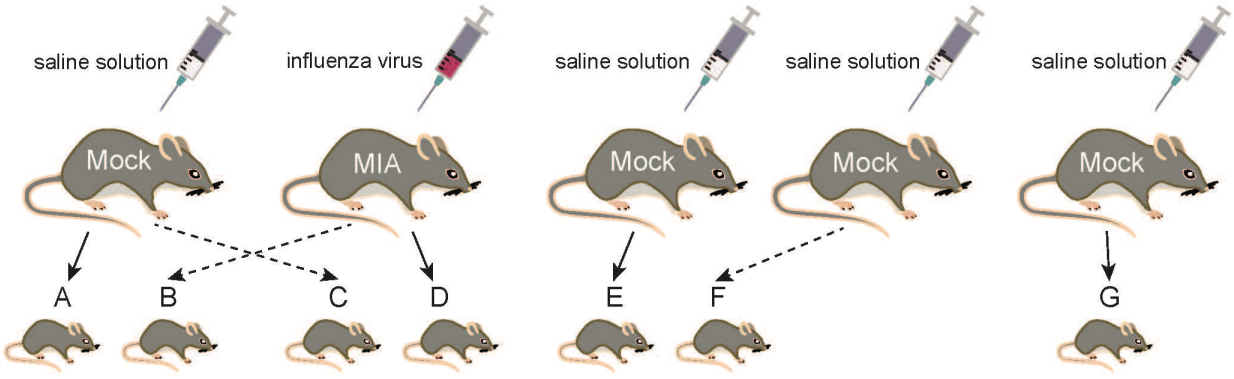

Figure 6-1 Overview of the experimental design and label of each experimental group.

Samples passed the quality control were further processed with our ChIP-seq and RNA-seq downstream pipeline. MOWChIP-seq and RNA-seq data were aligned to the reference genome (GRCm38), and significant ChIP-seq peaks were identified with MACS2. We then conducted the overlap analysis to generate consensus peak sets for H3K27ac and H3K4me3 in neuronal nuclei (see Materials and Methods section), following “majority-rules” that means a consensus peak must be called in at least half of the datasets. We identified 49537 consensus peaks covering ~162Mb (5.95% of the genome) for H3K27ac and 64512 consensus peaks covering ~164Mb (6.02% of the genome) for H3K4me3. Cis regulatory elements such as promoters and enhancers are non-coding DNA elements that play a critical role in the regulation of gene activity. In the brain, delineation of promoters or enhancers is essential for a better understanding of the alterations in gene expression patterns associated with neuronal development, synaptic formation, and dendritic growth. Thus, we predicted active promoters and enhancers by scanning the H3K4me3^{high} regions that intersect with promoters and

scanning the H3K27ac^{high} regions that did not have any overlap with promoters, respectively. H3K4me3 mark is highly enriched at active promoters near transcription start sites and H3K27ac mark distinguishes active enhancers from inactive and poised enhancers. Accordingly, we generated 36480 (covering ~115Mb) and 18008 (covering ~68Mb) putative enhancers and promoters in neurons, respectively.

Epigenetic and transcriptomic alterations caused by prenatal maternal immune activation

Clinical models have suggested that MIA exposure can induce prolonged histone modification changes in the offspring brain[176,206]. To investigate the differences in histone modification features and gene expression patterns between MIA-offspring and Mock-offspring, we performed differential analysis in the comparisons of group D versus group E and group B versus group F. We regressed out co-factors including batch number, NSC, RSC, number of peaks, number of unique reads and sequencing depth. Our analysis revealed 3090 differential enhancers, 770 differential promoters, and 2762 differentially expressed genes (DEGs) between group D and group E, while 1650, 1147, 4461 differential peaks/genes were discovered in enhancers, promoters and DEGs, respectively, between group B and group F (Fig. 6-2a). Unlike even distribution of differential peaks in enhancers or DEGs, we surprisingly detected significant decreases in majority of the differential promoters (87.4% of differential peaks from group D vs group E and 77.6% of differential peaks from group B vs group F) in both sets. We next leveraged the previously reported correlation-based method to identify how differential enhancers are linked with genes. Those enhancers that did not correlated with any genes were associated with their nearest genes. We successfully annotated 3090 and

1650 differential enhancers to 2093 and 1231 genes via enhancer-gene correlation in the comparisons of group D vs group E and group B vs group F. Then, to investigate the prenatal effect of maternal immune activation, we performed the overlap analysis using the differential enhancers/promoters or DEGs identified from the comparison of group D vs group E and group B vs group F. We discovered that MIA affected offspring had commonly dysregulated enhancers/promoters levels and gene expression intensity compared to offspring born to mock mother at various loci involved in biological processes related to immune and stress response, neurotransmission signaling, and neurodevelopmental process or previously associated with SCZ risk – these included enhancers at Htr2a, Robo1, Nrxn1, and Nlgn1, promoters at Nfkb1 and Slc1a2, and DEGs at Grin2a, Grm5, Npy, Cpeb4, and Slc17a7 (Fig. 6-2b).

We also overlapped genes identified from differential enhancers/promoters with DEGs from RNA-seq in the comparison of these two sets of groups. We found that 375 and 428 genes were detected in both the DEGs from RNA-seq and differential enhancers/promoters associated genes in the comparisons of group D vs group E and group B vs group F, respectively (Fig. 6-2c and d). Among these, several neural processes related or psychiatric disorder associated genes were identified, including MACF1, which involved in neurite outgrowth and neuronal migration and associated with several neurodevelopmental and neurodegenerative diseases (SCZ and Parkinson's disease)[207], Cacna1c, which encodes a subunit of a voltage-dependent calcium channel and is associated with many psychiatric disorders such as SCZ and bipolar disorder[208-210], Rgs4, a well-known protein that was found consistently altered in schizophrenia subjects[211,212], and Nfkb1, which involved in immune and/or stress

responses and acted as a regulator of the early response to viral infection[213-215]. Gene Ontology (GO) enrichment analysis of these two sets of overlapped genes further revealed several pathways involved in brain structure including “anatomical structure development” (FDR 1.35 x 10^{-3} in D vs E and 6.51 x 10^{-5} in B vs F). MRI studies from MIA influenced mature rodent and nonhuman primate subjects have demonstrated enlarged ventricles and altered size of some brain regions[216-218]. Also, several neuronal function related GO terms were found enriched, such as “generation of neurons” (FDR 9.8 x 10^{-4} in D vs E and 2.56 x 10^{-3} in B vs F), “neurogenesis” (FDR 3.58 x 10^{-4} in D vs E and 2.82 x 10^{-3} in B vs F), and “regulation of neuron projection development” (FDR 1.08 x 10^{-3} in D vs E and 1.54 x 10^{-4} in B vs F). This supported the finding from previous studies that suggested reduced density of some specific neuron subtypes and decreased long-range functional connectivity between neurons[206,219,220] (Fig. 6-2e).

The abnormalities of neurotransmitter systems are still viewed as the key in the current paradigm of pathophysiology of most major psychiatric diseases. Notably, the GO terms “modulation of chemical synaptic transmission” (FDR 1.44 x 10^{-2} in D vs E and 2.06 x 10^{-2} in B vs F), “regulation of trans-synaptic signaling” (FDR 1.45 x 10^{-2} in D vs E and 2.07 x 10^{-2} in B vs F), “glutamate receptor signaling pathway” (FDR 2.15 x 10^{-2} in D vs E), and “synaptic transmission, glutamatergic” (FDR 5.83 x 10^{-3} in B vs F) were significantly enriched (Fig. 6-2e). Dysregulated neurotransmitter-related signaling pathways identified here might act as the molecular foundation for the effect of MIA on altered behavior and impaired neuronal function. Interestingly, the influence of MIA on the mice offspring in group D and group E showed some difference. Using the 375 overlapped genes identified from the comparison between group D and group E, we

found that several KEGG pathways including “Glutamatergic synapse” (FDR 1.1 x 10^{-3}), “Long-term potentiation” (FDR 1.45 x 10^{-2}), “Dopaminergic synapse” (FDR 2.7 x 10^{-2}) were enriched. However, there is no enriched pathway identified from the 428 overlapped genes from group B vs group F. This might be because the mice in group B were raised in an adopted family which compromises the effect of MIA in some degree.

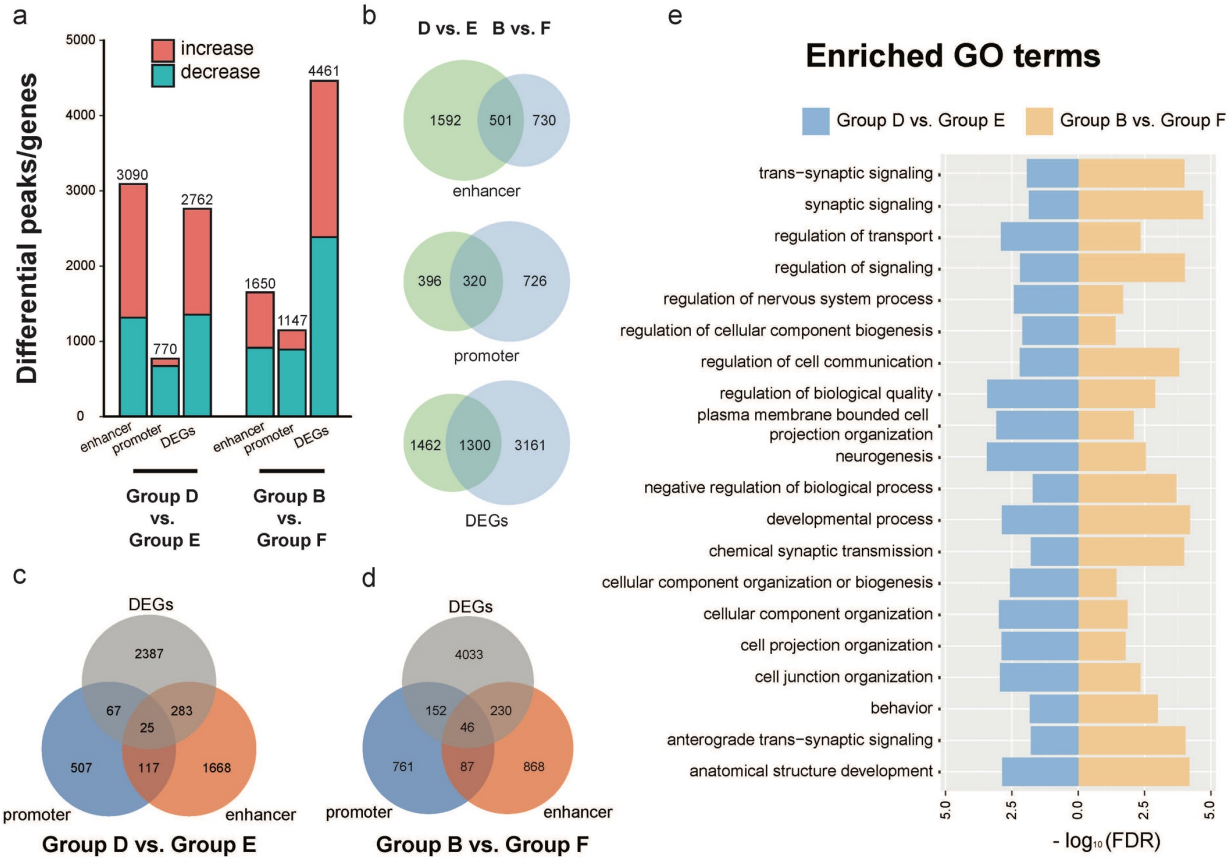

Figure 6-2 Epigenomic alterations induced by maternal immune activation. (a) Differential histone modification peaks and DEGs obtained by comparing the mice from group D with the ones from group E and the mice from group B with the ones from group F. (b) Venn diagrams on the relationship of the genes associated with differential enhancers/promoters and DEGs in the comparison of D vs E and B vs F. (c, d) Venn diagrams on the relationship among genes associated with differential

enhancers/promoters and DEGs in two sets of comparison. (e) Enriched GO terms identified in both comparisons from the overlapped genes

Alterations induced by MIA-treated mice in the adopting family

Studies of animal models of social interactions and stressful early-life experience demonstrated the link between these early life exposures with potential risk of psychiatric disorders like depression and schizophrenia[221-223]. Because of our experimental design, we tested whether these postnatal life experience would cause any epigenetic or transcriptomic variation. To examine if the MIA-affected mice siblings might generate any alteration in histone modification landscapes or the gene expression profiles in the brain of their littermates, we conducted differential analysis in the comparisons of group A versus group E and group C versus group F. Like what we did above, we also regressed out co-factors including batch number, NSC, RSC, number of peaks, number of unique reads and sequencing depth. We identified 4772 differential enhancers, 897 differential promoters, and 2908 differentially expressed genes (DEGs) between group A and group E, while 1180, 957, 2167 differential peaks/genes were discovered in enhancers, promoters and DEGs, respectively, between group C and group F (Fig. 6-3a). Similar to what we saw in the differential promoters from MIA-affected offspring, most of the differential promoters here also significantly decreased compared with their control subjects (89.3% from group A vs. group E and 74.4% from group C vs group F).

Again, we performed the overlap analysis using the differential enhancers/promoters or DEGs identified from the comparison of group A vs group E and group C vs group F. Compared with the mice raised with other Mock-affected

siblings, we detected that the exposure to MIA-affected siblings resulted in dysregulated enhancers/promoters levels and altered gene expression intensity at various loci involved in processes related to immune and stress response, neurotransmission signaling, and neurodevelopmental process or previously associated with SCZ risk – these included enhancers at Gria2, Cacana1c, Nrxn3, and Grm3, promoters at Fyn and Nrxn1, and DEGs at Slc17a7, Grm5, Pvalb, and Mog (Fig. 6-3b). We next overlapped genes identified from differential enhancers/promoters with DEGs from RNA-seq in the comparison of group A with group E and group C with group F. We identified 512 altered genes in both the DEGs from RNA-seq and differential enhancers/promoters associated genes from the comparisons of group A vs group E (Fig. 6-3c). While there were neuronal function and neurotransmitter system related pathways significantly enriched, such as “nervous system development” (FDR 7.79×10^{-7}), “regulation of synapse assembly” (FDR 5.06×10^{-5}), and “regulation of synapse structure or activity” (FDR 2.49×10^{-4}) in the overlapped genes from the comparison of group A and group E (Fig. 6-3e), we did not find any GO terms enriched in the 159 overlapped genes identified from the comparison of group C and group F (Fig. 6-3d). It might be because the mice in group C were not fostered by their own dam, just like what happened to the mice in group B. The extra maternal care from their new dam help to minimize the impact from their MIA-affected siblings, though there is no such study has been conducted.

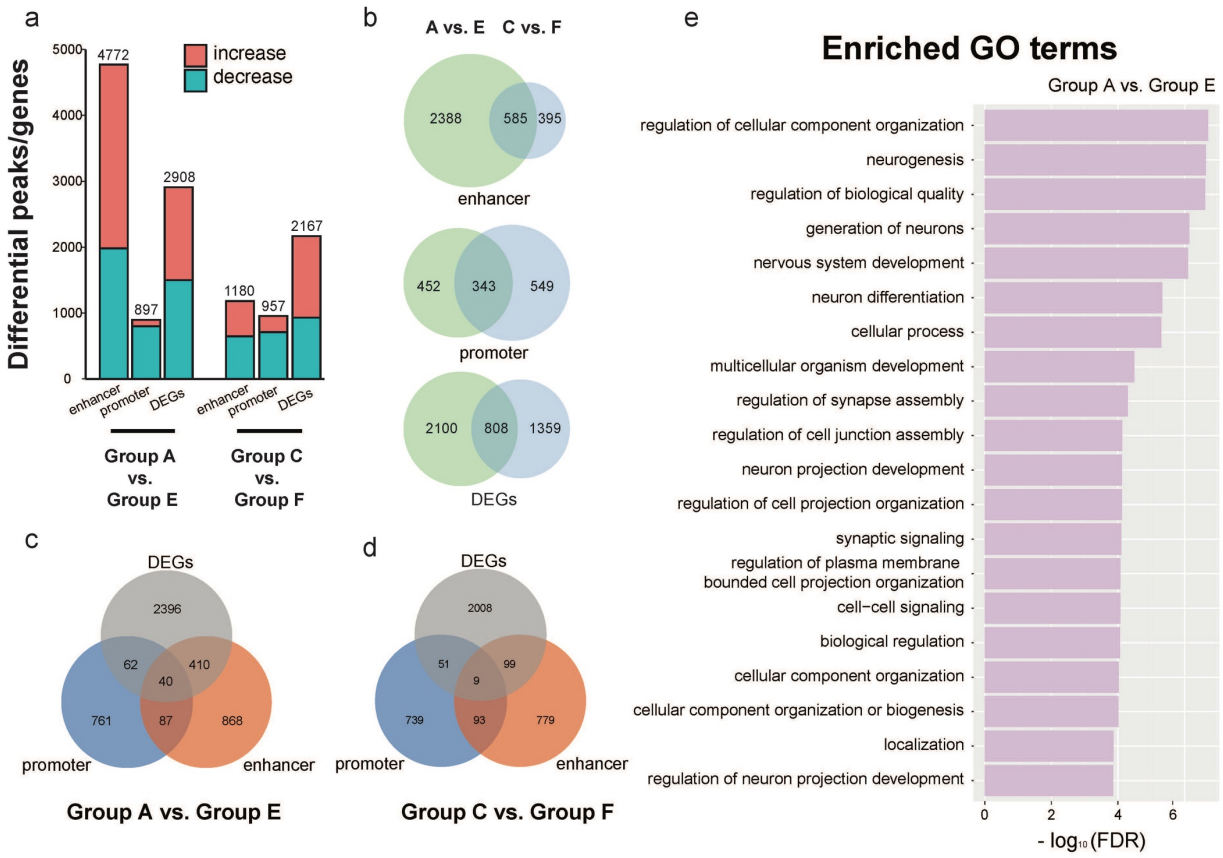

Figure 6-3 Epigenomic alterations induced by MIA-affected siblings. (a) Differential histone modification peaks and DEGs obtained by comparing the mice from group A with the ones from group E and the mice from group C with the ones from group F. (b) Venn diagrams on the relationship of the genes associated with differential enhancers/promoters and DEGs in the comparison of A vs E and C vs F. (c, d) Venn diagrams on the relationship among genes associated with differential enhancers/promoters and DEGs in two sets of comparison. (e) Enriched GO terms identified from the overlapped genes in the comparison of group A and group E

Epigenetic changes caused by cross-fostering

In line with two analysis described above, we hypothesized that the procedure of cross-fostering might also generate some unexpected alterations to the mice. To understand the effect of cross-fostering on the pups, we compared group C with group A and group B with group D. Comparing C to A determined the effect of cross-fostering on Mock-born pups, while B to D reveals the effect of cross-fostering on MIA-born pups. There were 986 differential enhancers, 554 differential promoters, and 510 differentially expressed genes (DEGs) between group C and group A, while 3639, 618, 572 differential peaks/genes were discovered in enhancers, promoters and DEGs, respectively, between group B and group D (Fig. 6-4a). We found that there were substantially fewer differential peaks/genes associated with cross-fostering than the ones induced by prenatal MIA or postnatal MIA-affected siblings, expect for the enhancers in group B and group D. Then, in order to study the common effect of cross-fostering on mice brain development, we performed the overlap analysis using the differential enhancers/promoters or DEGs identified from the comparison of group C vs group A and group B vs group D. We did not identify many commonly dysregulated promoters or genes but found that there were several differential enhancers at various loci involved in neurotransmission and the formation of synaptic communications, like Grm3, Grm7 and Nrxn1 (Fig. 6-4b).

We also did overlap analysis using genes linked to differential enhancers or promoters with DEGs identified from RNA-seq data in the comparison of group C with group A and group B with group D. However, there were still not many overlapped genes detected in either of the comparisons (31 from group C vs group A and 70 from group B vs group D) and no GO term was significantly enriched in these genes (Fig. 6-

4c and d). In that case, we first identified some changes associated with cross-fostering, especially in the features of enhancer when it happened to the MIA-affect mice offspring. However, compared with the effect of MIA and grown up with MIA-affected siblings, cross-fostering might only induce mild epigenetic and transcriptomic changes in the frontal cortex of mice offspring.

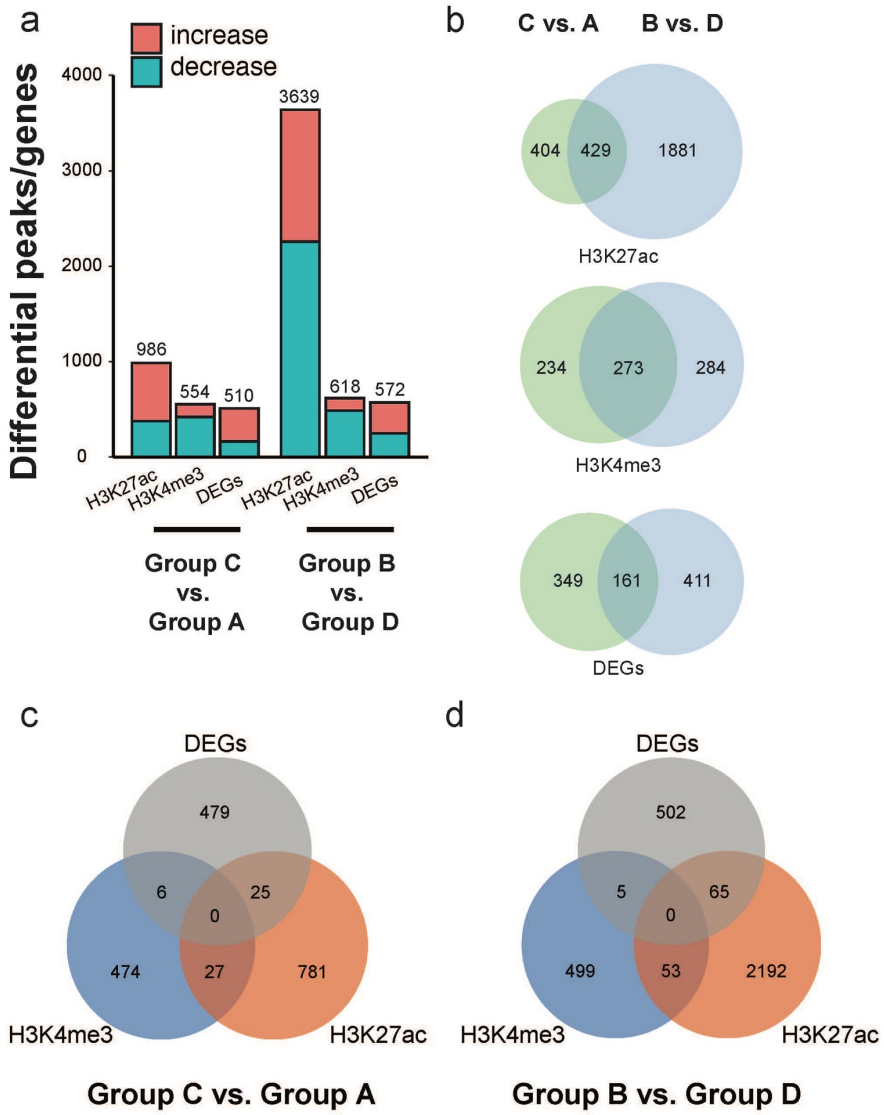

Figure 6-4 Epigenomic alterations induced by cross-fostering. (a) Differential histone modification peaks and DEGs obtained by comparing the mice from group C with the

ones from group A and the mice from group B with the ones from group D. (b) Venn diagrams on the relationship of the genes associated with differential enhancers/promoters and DEGs in the comparison of C vs A and B vs D. (c, d) Venn diagrams on the relationship among genes associated with differential enhancers/promoters and DEGs in two sets of comparison.

Batch variations in the epigenetic alterations induced by prenatal or postnatal MIA effect

We totally generated epigenetic and transcriptomic profiles from three batches of mice in group A and D. The results showed above were the mice from one of these batches, in which they present the most substantial difference compared with their control groups. However, for the other two batches of group A and D, we identified much fewer differential genes no matter associated with enhancers/promoters or from RNA-seq. This agreed with a recent study that demonstrate the variation of the effect of MIA on the offspring[224]. For group D, we identified 940 and 207 differential peaks in enhancer regions, 53 and 780 differential peaks in promoter regions , and 609 and 384 DEGs from the comparisons of group D2 with group E and group D3 with group E, respectively (Fig. 6-5a). Then, we examined the relationship between the genes controlled by differential enhancers/promoters and differentially expressed genes detected from each batch of group D vs group E. We figured that although group D2 and group D3 present less difference than group D1 compared with control group E, several psychiatric disorder related genes were still significantly altered in group D2 and D3, including Robo1 and Gria1 in group D2 and Ctnnd2 in group D3 at enhancer regions, Mef2c in group D2 and Cacna1c in group D3 at promoter regions (Fig. 6-5b).

For the overlap analysis using genes associated with differential enhancers/promoters and DEGs, even fewer overlapped genes were shown up (19 genes from group D2 vs group E and 7 genes from group D3 vs group E) (Fig. 6-5c and d). Similar results were found when we look at the batch variation of the effect of MIA-affected siblings. 31 and 20 genes were overlapped between the genes annotated from differential enhancers/promoters and the genes from DEGs in group A2 vs group E and group A3 vs group E, respectively (Fig. 6-5 g and h). This results suggested that the effect of MIA could vary a lot based on the variation of the condition of pregnant dams.

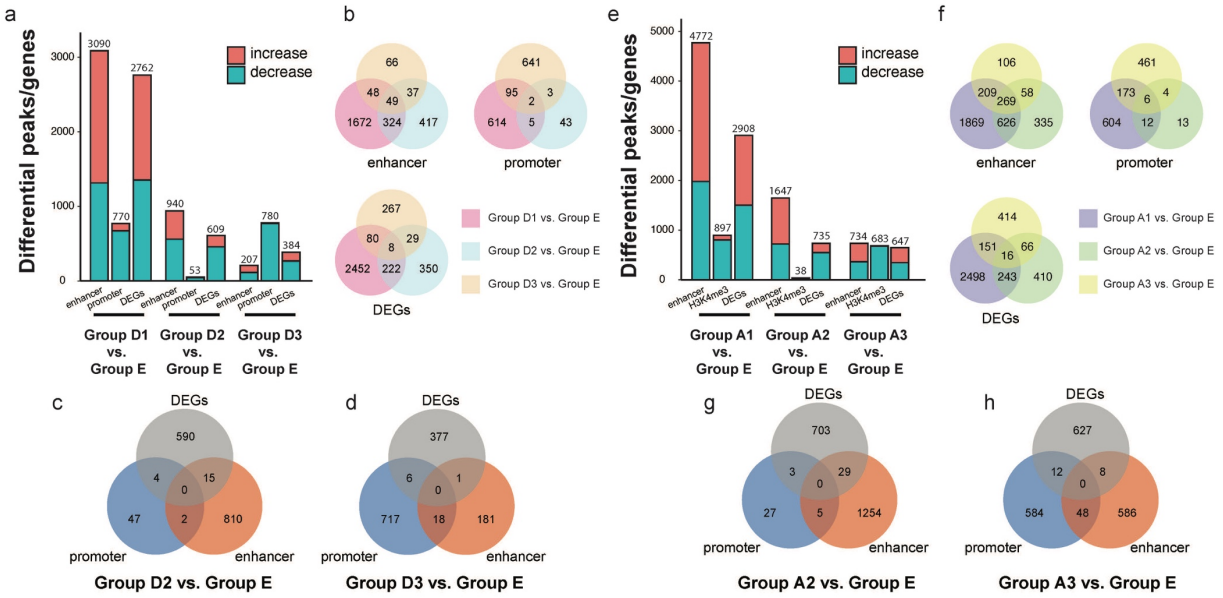

Figure 6-5 Epigenomic and transcriptomic variations from batch effect. (a) Differential histone modification peaks and DEGs obtained by comparing the mice from group D1, D2 and D3 with the ones from group E. (b) Venn diagrams on the relationship of the genes associated with differential enhancers/promoters and DEGs in the comparison of D1, D2, and D3 vs E (c, d) Venn diagrams on the relationship among genes associated with differential enhancers/promoters and DEGs in the comparisons of group D2 with

group E and group D3 with group E. (e) Differential histone modification peaks and DEGs obtained by comparing the mice from group A1, A2 and A3 with the ones from group E. (f) Venn diagrams on the relationship of the genes associated with differential enhancers/promoters and DEGs in the comparison of A1, A2, and A3 vs E. (g, h) Venn diagrams on the relationship among genes associated with differential enhancers/promoters and DEGs in the comparisons of group A2 with group E and group A3 with group E.

Altered transcription factors induced by the prenatal and postnatal effect of MIA

Transcription factors are regulatory proteins that attach to specific DNA sequences and regulate the genes activity. Identification of critical transcription factors (TFs) in the system could enhance our understanding of the mechanisms underlying the specific biological process. Here, we used a recently reported advanced bioinformatic tool, Taiji[156], to integrate our ChIP-seq and RNA-seq data and then construct transcriptional networks to identify key TFs that dysregulated by the prenatal and postnatal effect of MIA. We used H3K4me3 profiles to identify active promoters and H3K27ac profiles to detect active enhancers. Then, we linked these potential transcription factor binding sites to their target TFs using EpiTensor[157], an algorithm for predicting enhancer-promoter interactions, to scan 639 TF motifs. Next, each TF was assigned a PageRank score to indicate its regulatory importance in the transcriptional network by the Personalized PageRank algorithm.

In order to identify the significantly changed transcription factors (p-value < 0.01) caused by prenatal effect of MIA, we performed differential analysis using absolute difference in either group D vs group E or group B vs group F (Fig. 6-6a and b). Several TFs were dysregulated in both comparisons, including increased TFs at EGR1 (a key regulator of synaptic plasticity and neuronal activity)[225], RELA (a gene related to the development of schizophrenia)[226], and decreased TFs at MEIS2 (a co-factor in neurogenesis). We also observed that some TFs were uniquely differentially expressed in group D vs group E, such as STAT2. Previous study about MIA effect has demonstrated elevated level of various cytokines, like interferon-gamma, which plays an important role in the developmental trajectories of brain[227]. Stat2 is known as an antiviral molecule that could modulate the expression of interferon-related genes[228].

We also noticed that transcription factor NR3C1 was among the top 4 decreased TFs in both comparisons of group D vs group E and group A vs group E (Fig. 6-6a and c). NR3C1 was reported as the main transcriptional regulator of the glucocorticoid that involved in most of the steps of the hypothalamic–pituitary–adrenal (HPA) axis[229]. Dysregulated HPA axis has been demonstrated in patients with various psychiatric diseases[230]. Altered expression of NR3C1 has also been associated with the

experience of abnormal maternal care[231].

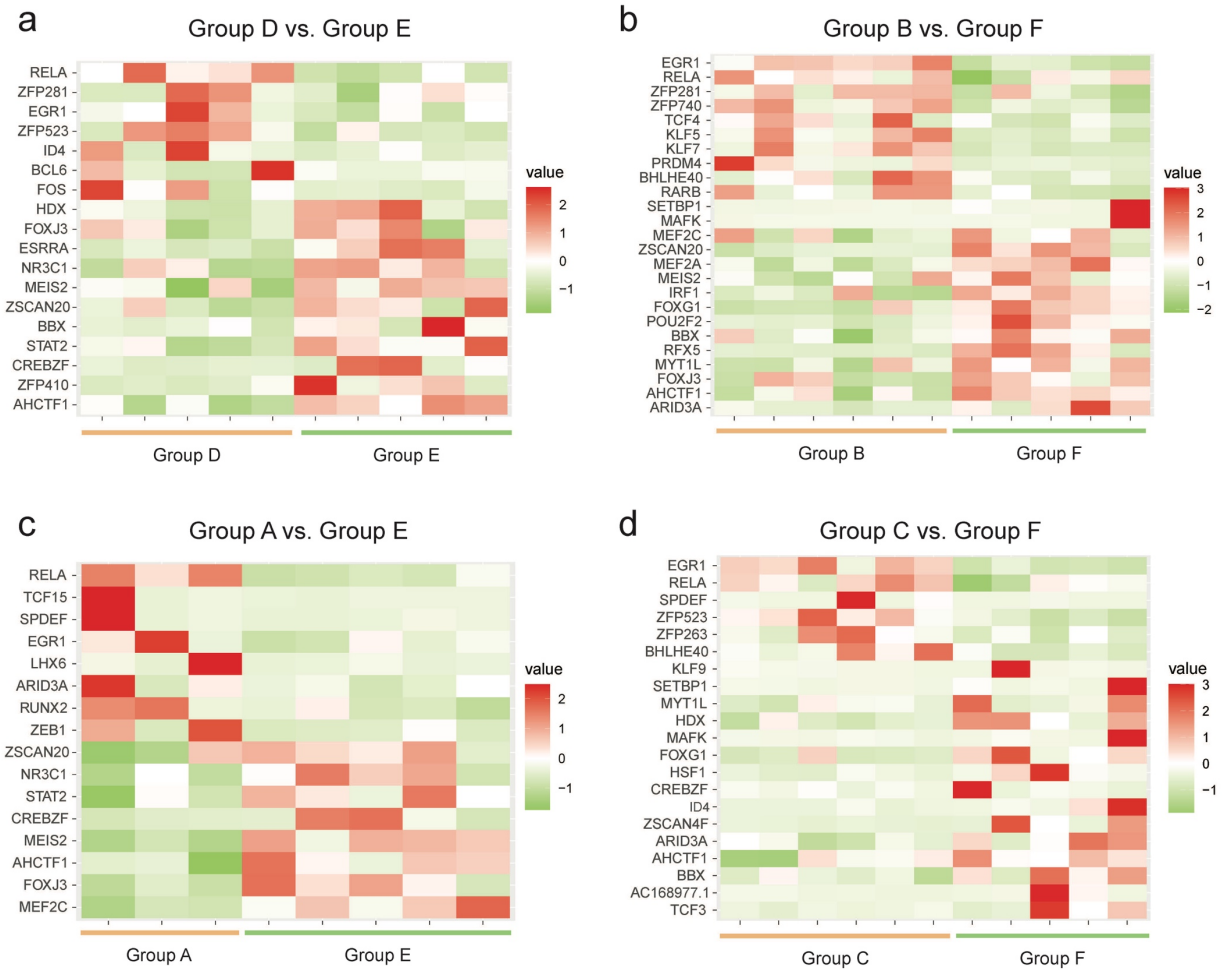

Figure 6-6 Heatmap of z-score personalized pagerank for significantly differentially expressed TFs (p-value < 0.01) in the comparison of (a) group D vs group E, (b) group B vs group F, (c) group A vs group E, and (d) group C vs group F.

6.5 Conclusion

In this study, we combined low-input epigenomic and transcriptomic analysis to define how gene expression and TF regulation vary in the brain of mice directly born from immune activated mom and mice grown up with MIA-affect siblings. While previous

studies failed to detect any significant changes caused by MIA, our data provide, for the first time, evidence suggesting alterations, due to maternal immune response, in covalent histone modifications at different gene regions previously associated with various mental diseases and neurodevelopmental processes. Compared to the effect from MIA mom, MIA-affect siblings induced some weaker epigenomic and transcriptomic changes to their littermates. Since we only examine the changes in neurons from the mice brain, further studies with single-cell based technologies may provide more information on the contribution of various cell-type-subpopulations.

7 Summary and Outlook

In order to get more insight into complex biological diseases like psychiatric disorders, it is vital to conduct multi-omics study and combine the epigenetics and transcriptomic data to detect the molecular target associated with biological process. This could help us to have a more full-scale understanding of the molecular mechanisms involved in the development of diseases and potentially improve our current treatment strategies. In this thesis, the second generation of one of our microfluidic technologies (MOWChIP-seq) was developed. The new MOWChIP-seq protocol enabled us to generate high quality sequencing data with low input and high-throughput. This is especially important when people need to examine the scarce amount of primary tissue samples or want to test homogeneous cell population after cell separation using FACS. Then, we used new MOWChIP-seq protocol to profile genome-wide histone modifications in three mental illness-related biological studies: the effect of psychedelics in mice, schizophrenia, and the effect of maternal immune activation in mice. Together with the histone modifications profiles, we also performed the RNA-seq assays to generate transcriptomic data for each sample.

From the study about examination of alterations induced by single exposure of psychedelic in mice brain, we found that the epigenomic and transcriptomic dynamics following DOI administration are significant. This is the first integrative studies of temporal changes in frontal cortex epigenome and transcriptome after psychedelic administration. Genome-wide epigenetic changes have long been speculated to be the layer of regulation that integrates both genetic and environmental factors[126,127]. Here, we show that the epigenetic landscape can be exogenously reshaped by a drug that

elicits unequivocal psychedelic effects in humans. In this study, we report that, in contrast to the fairly rapid and transient changes in the transcriptome, a large fraction of epigenetic changes in enhancer regions persist for at least 7 days after DOI administration and potentially constitute the molecular basis for the long-lasting effects. It is worth noting that our findings were established by studying the entire neuronal cell population. Further examination of individual cell types (e.g., excitatory and inhibitory neurons) may yield additional information on the roles of specific neuron types. We also studied the points of convergence between genetic loci associated with a number of human disorders and the footprint of DOI in the mouse brain epigenome. The peaks of H3K27ac showed significant overlap with GWAS-discovered variants associated with schizophrenia, depression, and ADHD. Although psychedelics have been shown to have fast-acting and long-lasting therapeutic effect on depression[232], their effects on schizophrenia and ADHD remain unknown even when taking into account some overlap between these three psychiatric disorders[233-235].

In the schizophrenia project, we investigated the epigenomic and transcriptomic changes caused by schizophrenia and the treatment of this disease. In the largest ChIP-seq/RNA-seq study in postmortem brain samples from schizophrenia subjects and controls, our data suggest cell-type specific epigenomic differences in individuals with schizophrenia as well as cellular alterations in signaling pathways potentially involved in either the elimination of schizophrenia-related epigenomic alterations upon antipsychotic drug treatment or the antipsychotic-dependent modulation of alternative epigenetic pathways previously unaffected in the untreated schizophrenia cohort. However, we cannot fully exclude the possibility of previous exposure to antipsychotic

medications in the AF-schizophrenia group, and hence that the epigenetic alterations observed exclusively in the AF-schizophrenia group are a consequence of a potential period of decompensation, which typically occurs following voluntary treatment discontinuation[172]. It is also worth noting that our findings were established by examining the average characteristics of entire neuronal and glial fractions. Further studies of individual neuronal and glial cell subtypes may yield additional information on the role of cell-type-subpopulations[173,174].

We also conducted the research about histone modification and gene expression changes in mice brain due to maternal immune activation. Mouse model with influenza was used in the study. We figured out that the prenatal effect of maternal immune response from mom is much more significant than the postnatal effect of MIA from littermates.

Besides what I have done in the thesis, the future work could focus on three directions. First, microfluidic-based technology could be developed to generate epigenetic information at single cell level. There is no question that approaches for single cell analysis will generate more information with high-resolution and help us to have a better understanding of the biological processes. For example, in the research about schizophrenia, we used low-input technology and found out that both neurons and glia were involved in the development and treatment of the disease. This suggested the newly found association between schizophrenia and glia. It also inspired us that increasing the resolution of technology might enable us to identify more links between the nuclei subtypes and the complex disorder and have a better understanding about the molecular mechanisms underlying the development of schizophrenia. In that case,

developing droplet-based ChIP-seq assay could be an interesting direction for my future study.

Second, the aim of epigenomics is to identify the association between epigenetic features with phenotype. In order to have the whole picture of each single cell, it could be helpful to integrate multi-omic assays together, such as the joint of DNA methylation and gene expression or chromatin interaction and gene expression. The changes of gene expression profile are the result from multiple layers of regulation, like histone modification, DNA methylation, and transcription factors. These multi-omic assays could provide us more insights into how gene activity was altered. This could be very interesting in the study of cell fate, for example what make stem cells finally differentiate as muscle cells.

Third, the functions of a complex tissue are essentially associated with the organization of its resident cell types. In turn, spatial positions of cells in a tissue substantially affect its function. In that case, the technology that could map histone modifications at single-cell level and also keep the spatial information will contribute to the field a lot. For example, brain is such a highly spatial-organized organ. There are six layers in the grey matter of the brain. Some cell types in several specific layers are more vulnerable than the others in a specific condition. Using technology which can preserve spatial information could advance our knowledge of basic neurobiology and provide more insight into the development of various mental disorders.

Reference

1 Schlereth, K. *et al.* The transcriptomic and epigenetic map of vascular quiescence in the continuous lung endothelium. *Elife* **7** (2018). https://doi.org:10.7554/eLife.34423

2 Jia, J. *et al.* An integrated transcriptome and epigenome analysis identifies a novel candidate gene for pancreatic cancer. *BMC Med Genomics* **6**, 33 (2013). https://doi.org:10.1186/1755-8794-6-33

3 Perrino, C. *et al.* Epigenomic and transcriptomic approaches in the post-genomic era: path to novel targets for diagnosis and therapy of the ischaemic heart? Position Paper of the European Society of Cardiology Working Group on Cellular Biology of the Heart. *Cardiovasc Res* **113**, 725-736 (2017). https://doi.org:10.1093/cvr/cvx070

4 Liu, Y. *et al.* Blood monocyte transcriptome and epigenome analyses reveal loci associated with human atherosclerosis. *Nat Commun* **8**, 393 (2017). https://doi.org:10.1038/s41467-017-00517-4

5 Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat Biotechnol* **28**, 1057-1068 (2010). https://doi.org:10.1038/nbt.1685

6 Sadri-Vakili, G. & Cha, J. H. Mechanisms of disease: Histone modifications in Huntington's disease. *Nat Clin Pract Neurol* **2**, 330-338 (2006). https://doi.org:10.1038/ncpneuro0199

7 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831 (2012). https://doi.org:10.1101/gr.136184.111

8 Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680 (2009). https://doi.org:10.1038/nrg2641

9 Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943-949 (2014). https://doi.org:10.1126/science.1256271

10 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 (2015). https://doi.org:10.1038/ncomms7033

11 Ma, S., Hsieh, Y. P., Ma, J. & Lu, C. Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. *Sci Adv* **4**, eaar8187 (2018). https://doi.org:10.1126/sciadv.aar8187

12 Cao, Z., Chen, C., He, B., Tan, K. & Lu, C. A microfluidic device for epigenomic profiling using 100 cells. *Nat Methods* **12**, 959-962 (2015). https://doi.org:10.1038/nmeth.3488

13 Zhu, B. *et al.* MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nat Protoc* **14**, 3366-3394 (2019). https://doi.org:10.1038/s41596-019-0223-x

14 Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annu Rev Genet* **43**, 559-599 (2009). https://doi.org:10.1146/annurev.genet.032608.103928

15 Nichols, D. E. Psychedelics. *Pharmacol Rev* **68**, 264-355 (2016). https://doi.org:10.1124/pr.115.011478

16 Hanks, J. B. & Gonzalez-Maeso, J. Animal models of serotonergic psychedelics. *ACS Chem Neurosci* **4**, 33-42 (2013). https://doi.org:10.1021/cn300138m

17 Tsuang, M. Schizophrenia: genes and environment. *Biol Psychiatry* **47**, 210-220 (2000). https://doi.org:10.1016/s0006-3223(99)00289-9

18 Egger, G., Liang, G., Aparicio, A. & Jones, P. A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429**, 457-463 (2004). https://doi.org:10.1038/nature02625

19 Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692 (2007). https://doi.org:10.1016/j.cell.2007.01.029

20 Ikeuchi, M., Iwase, A. & Sugimoto, K. Control of plant cell differentiation by histone modification and DNA methylation. *Curr Opin Plant Biol* **28**, 60-67 (2015). https://doi.org:10.1016/j.pbi.2015.09.004

21 Verzi, M. P. *et al.* Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev Cell* **19**, 713-726 (2010). https://doi.org:10.1016/j.devcel.2010.10.006

22 Lunyak, V. V. & Rosenfeld, M. G. Epigenetic regulation of stem cell fate. *Hum Mol Genet* **17**, R28-36 (2008). https://doi.org:10.1093/hmg/ddn149

23 Ho, S. M. *et al.* Environmental epigenetics and its implication on disease risk and health outcomes. *ILAR J* **53**, 289-305 (2012). https://doi.org:10.1093/ilar.53.3-4.289

24 Aguilera, O., Fernandez, A. F., Munoz, A. & Fraga, M. F. Epigenetics and environment: a complex relationship. *J Appl Physiol (1985)* **109**, 243-251 (2010). https://doi.org:10.1152/japplphysiol.00068.2010

25 Alegria-Torres, J. A., Baccarelli, A. & Bollati, V. Epigenetics and lifestyle. *Epigenomics* **3**, 267-277 (2011). https://doi.org:10.2217/epi.11.22

26 Torano, E. G., Garcia, M. G., Fernandez-Morera, J. L., Nino-Garcia, P. & Fernandez, A. F. The Impact of External Factors on the Epigenome: In Utero and over Lifetime. *Biomed Res Int* **2016**, 2568635 (2016). https://doi.org:10.1155/2016/2568635

27 Bloom, K. & Joglekar, A. Towards building a chromosome segregation machine. *Nature* **463**, 446-456 (2010). https://doi.org:10.1038/nature08912

28 Oudet, P., Gross-Bellard, M. & Chambon, P. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**, 281-300 (1975). https://doi.org:10.1016/0092-8674(75)90149-x

29 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974). https://doi.org:10.1126/science.184.4139.868

30 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260 (1997). https://doi.org:10.1038/38444

31 Bentley, G. A., Lewit-Bentley, A., Finch, J. T., Podjarny, A. D. & Roth, M. Crystal structure of the nucleosome core particle at 16 A resolution. *J Mol Biol* **176**, 55-75 (1984). https://doi.org:10.1016/0022-2836(84)90382-6

32 Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772-778 (2006). https://doi.org:10.1038/nature04979

33 Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448-453 (2003). https://doi.org:10.1038/nature01411

34 Grigoryev, S. A., Arya, G., Correll, S., Woodcock, C. L. & Schlick, T. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci U S A* **106**, 13317-13322 (2009). https://doi.org:10.1073/pnas.0903280106

35 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**, 786-794 (1964). https://doi.org:10.1073/pnas.51.5.786

36 Barber, C. M. *et al.* The enhancement of histone H4 and H2A serine 1 phosphorylation during mitosis and S-phase is evolutionarily conserved. *Chromosoma* **112**, 360-371 (2004). https://doi.org:10.1007/s00412-004-0281-9

37 Fusauchi, Y. & Iwai, K. Tetrahymena histone H2A. Acetylation in the N-terminal sequence and phosphorylation in the C-terminal sequence. *J Biochem* **95**, 147-154 (1984). https://doi.org:10.1093/oxfordjournals.jbchem.a134578

38 Suka, N., Suka, Y., Carmen, A. A., Wu, J. & Grunstein, M. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Mol Cell* **8**, 473-479 (2001). https://doi.org:10.1016/s1097-2765(01)00301-x

39 Baarends, W. M. *et al.* Increased phosphorylation and dimethylation of XY body histones in the Hr6b-knockout mouse is associated with derepression of the X chromosome. *J Cell Sci* **120**, 1841-1851 (2007). https://doi.org:10.1242/jcs.03451

40 Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-878 (2004). https://doi.org:10.1038/nature02985

41 Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-411 (2002). https://doi.org:10.1038/nature01080

42 Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599 (2000). https://doi.org:10.1038/35020506

43 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007). https://doi.org:10.1016/j.cell.2007.02.005

44 Kebede, A. F., Schneider, R. & Daujat, S. Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *FEBS J* **282**, 1658-1674 (2015). https://doi.org:10.1111/febs.13047

45 Zhang, D. *et al.* Metabolic regulation of gene expression by histone lactylation. *Nature* **574**, 575-580 (2019). https://doi.org:10.1038/s41586-019-1678-1

46 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395 (2011). https://doi.org:10.1038/cr.2011.22

47 Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol* **7**, a025064 (2015). https://doi.org:10.1101/cshperspect.a025064

48 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010). https://doi.org:10.1073/pnas.1016071107

49 Yuen, B. T. & Knoepfler, P. S. Histone H3.3 mutations: a variant path to cancer. *Cancer Cell* **24**, 567-574 (2013). https://doi.org:10.1016/j.ccr.2013.09.015

50 Schilbach, S., Aibara, S., Dienemann, C., Grabbe, F. & Cramer, P. Structure of RNA polymerase II pre-initiation complex at 2.9 A defines initial DNA opening. *Cell* **184**, 4064-4072 e4028 (2021). https://doi.org:10.1016/j.cell.2021.05.012

51 Liu, X., Bushnell, D. A. & Kornberg, R. D. RNA polymerase II transcription: structure and mechanism. *Biochim Biophys Acta* **1829**, 2-8 (2013). https://doi.org:10.1016/j.bbagrm.2012.09.003

52 Paterni, I., Granchi, C., Katzenellenbogen, J. A. & Minutolo, F. Estrogen receptors alpha (ERalpha) and beta (ERbeta): subtype-selective ligands and clinical potential. *Steroids* **90**, 13-29 (2014). https://doi.org:10.1016/j.steroids.2014.06.012

53 Hutchison, C. A., 3rd. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res* **35**, 6227-6237 (2007). https://doi.org:10.1093/nar/gkm688

54 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46 (2010). https://doi.org:10.1038/nrg2626

55 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145 (2008). https://doi.org:10.1038/nbt1486

56 Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* **12**, R112 (2011). https://doi.org:10.1186/gb-2011-12-11-r112

57 Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* **25**, 99-104 (2000). https://doi.org:10.1016/s0968-0004(99)01535-2

58 O'Neill, L. P. & Turner, B. M. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76-82 (2003). https://doi.org:10.1016/s1046-2023(03)00090-2

59 Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J Biol Chem* **290**, 26404-26411 (2015). https://doi.org:10.1074/jbc.R115.651679

60 Adli, M. & Bernstein, B. E. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nat Protoc* **6**, 1656-1668 (2011). https://doi.org:10.1038/nprot.2011.402

61 Shankaranarayanan, P. *et al.* Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat Methods* **8**, 565-567 (2011). https://doi.org:10.1038/nmeth.1626

62 Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-1172 (2015). https://doi.org:10.1038/nbt.3383

63 Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6** (2017). https://doi.org:10.7554/eLife.21856

64 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018). https://doi.org:10.1038/nprot.2018.015

65 Hainer, S. J., Boskovic, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329 e1311 (2019). https://doi.org:10.1016/j.cell.2019.03.014

66 Shin, H., Liu, T., Duan, X., Zhang, Y. & Liu, X. S. Computational methodology for ChIP-seq analysis. *Quant Biol* **1**, 54-70 (2013). https://doi.org:10.1007/s40484-013-0006-2

67 Kukurba, K. R. & Montgomery, S. B. RNA Sequencing and Analysis. *Cold Spring Harb Protoc* **2015**, 951-969 (2015). https://doi.org:10.1101/pdb.top084970

68 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012). https://doi.org:10.1038/nmeth.1923

69 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009). https://doi.org:10.1093/bioinformatics/btp324

70 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137

71 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010). https://doi.org:10.1093/bioinformatics/btp616

72 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014). https://doi.org:10.1186/s13059-014-0550-8

73 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011). https://doi.org:10.1038/nbt.1754

74 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009). https://doi.org:10.1038/nrg2484

75 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635

76 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015). https://doi.org:10.1038/nmeth.3317

77 Gibney, E. R. & Nolan, C. M. Epigenetics and gene expression. *Heredity (Edinb)* **105**, 4-13 (2010). https://doi.org:10.1038/hdy.2010.54

78 Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-254 (2003). https://doi.org:10.1038/ng1089

79 Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048 (2010). https://doi.org:10.1038/nbt1010-1045

80 Rando, O. J. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev* **22**, 148-155 (2012). https://doi.org:10.1016/j.gde.2012.02.013

81 Baumeister, D., Barnes, G., Giaroli, G. & Tracy, D. Classical hallucinogens as antidepressants? A review of pharmacodynamics and putative clinical roles. *Ther Adv Psychopharmacol* **4**, 156-169 (2014). https://doi.org:10.1177/2045125314527985

82 Hesselgrave, N., Troppoli, T. A., Wulff, A. B., Cole, A. B. & Thompson, S. M. Harnessing psilocybin: antidepressant-like behavioral and synaptic actions of

psilocybin are independent of 5-HT2R activation in mice. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2022489118

83 Krishnan, V. & Nestler, E. J. The molecular neurobiology of depression. *Nature* **455**, 894-902 (2008). https://doi.org:10.1038/nature07455

84 Kessler, R. C. *et al.* The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* **289**, 3095-3105 (2003). https://doi.org:10.1001/jama.289.23.3095

85 Duman, R. S. & Aghajanian, G. K. Synaptic dysfunction in depression: potential therapeutic targets. *Science* **338**, 68-72 (2012). https://doi.org:10.1126/science.1222939

86 Insel, T. R. & Wang, P. S. The STAR*D trial: revealing the need for better treatments. *Psychiatr Serv* **60**, 1466-1467 (2009). https://doi.org:10.1176/appi.ps.60.11.1466

10.1176/ps.2009.60.11.1466

87 Carhart-Harris, R. *et al.* Trial of Psilocybin versus Escitalopram for Depression. *N Engl J Med* **384**, 1402-1411 (2021). https://doi.org:10.1056/NEJMoa2032994

88 Davis, A. K. *et al.* Effects of Psilocybin-Assisted Therapy on Major Depressive Disorder: A Randomized Clinical Trial. *JAMA Psychiatry* **78**, 481-489 (2021). https://doi.org:10.1001/jamapsychiatry.2020.3285

89 Griffiths, R. R. *et al.* Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. *J Psychopharmacol* **30**, 1181-1197 (2016). https://doi.org:10.1177/0269881116675513

90 Carhart-Harris, R. L. *et al.* Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *Lancet Psychiatry* **3**, 619-627 (2016). https://doi.org:10.1016/S2215-0366(16)30065-7

91 Cameron, L. P., Benson, C. J., Dunlap, L. E. & Olson, D. E. Effects of N, N-Dimethyltryptamine on Rat Behaviors Relevant to Anxiety and Depression. *ACS Chem Neurosci* **9**, 1582-1590 (2018). https://doi.org:10.1021/acschemneuro.8b00134

92 Zhang, G. *et al.* Stimulation of serotonin 2A receptors facilitates consolidation and extinction of fear memory in C57BL/6J mice. *Neuropharmacology* **64**, 403-413 (2013). https://doi.org:10.1016/j.neuropharm.2012.06.007

93 Gonzalez-Maeso, J. *et al.* Transcriptome fingerprints distinguish hallucinogenic and nonhallucinogenic 5-hydroxytryptamine 2A receptor agonist effects in mouse somatosensory cortex. *J Neurosci* **23**, 8836-8843 (2003).

94 Nichols, C. D. & Sanders-Bush, E. A single dose of lysergic acid diethylamide influences gene expression patterns within the mammalian brain. *Neuropsychopharmacology* **26**, 634-642 (2002). https://doi.org:10.1016/S0893-133X(01)00405-5

95 Ly, C. *et al.* Psychedelics Promote Structural and Functional Neural Plasticity. *Cell Rep* **23**, 3170-3182 (2018). https://doi.org:10.1016/j.celrep.2018.05.022

96 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014). https://doi.org:10.1038/nprot.2014.006

97 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096-1098 (2013). https://doi.org:10.1038/nmeth.2639

98 Nord, A. S. & West, A. E. Neurobiological functions of transcriptional enhancers. *Nat Neurosci* **23**, 5-14 (2020). https://doi.org:10.1038/s41593-019-0538-5

99 Venteo, S. *et al.* Neurog2 Deficiency Uncovers a Critical Period of Cell Fate Plasticity and Vulnerability among Neural-Crest-Derived Somatosensory Progenitors. *Cell Rep* **29**, 2953-2960 e2952 (2019). https://doi.org:10.1016/j.celrep.2019.11.002

100 Gonzalez-Maeso, J. *et al.* Hallucinogens recruit specific cortical 5-HT(2A) receptor-mediated signaling pathways to affect behavior. *Neuron* **53**, 439-452 (2007). https://doi.org:10.1016/j.neuron.2007.01.008

101 Mitsui, S., Yamaguchi, S., Matsuo, T., Ishida, Y. & Okamura, H. Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes Dev* **15**, 995-1006 (2001). https://doi.org:10.1101/gad.873501

102 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). https://doi.org:10.1186/1471-2105-9-559

103 Wang, L., Wang, Y., Duan, C. & Yang, Q. Inositol phosphatase INPP4A inhibits the apoptosis of in vitro neurons with characteristic of intractable epilepsy by reducing intracellular Ca(2+) concentration. *Int J Clin Exp Pathol* **11**, 1999-2007 (2018).

104 Zagnoni, P. G. & Albano, C. Psychostimulants and epilepsy. *Epilepsia* **43 Suppl 2**, 28-31 (2002). https://doi.org:10.1046/j.1528-1157.2002.043s2028.x

105 Nau, F., Jr. *et al.* Serotonin 5-HT(2) receptor activation prevents allergic asthma in a mouse model. *Am J Physiol Lung Cell Mol Physiol* **308**, L191-198 (2015). https://doi.org:10.1152/ajplung.00138.2013

106 Klingseisen, A. *et al.* Oligodendrocyte Neurofascin Independently Regulates Both Myelin Targeting and Sheath Growth in the CNS. *Dev Cell* **51**, 730-744 e736 (2019). https://doi.org:10.1016/j.devcel.2019.10.016

107 McIntyre, J. C., Titlow, W. B. & McClintock, T. S. Axon growth and guidance genes identify nascent, immature, and mature olfactory sensory neurons. *J Neurosci Res* **88**, 3243-3256 (2010). https://doi.org:10.1002/jnr.22497

108 Zonta, B. *et al.* Glial and neuronal isoforms of Neurofascin have distinct roles in the assembly of nodes of Ranvier in the central nervous system. *J Cell Biol* **181**, 1169-1177 (2008). https://doi.org:10.1083/jcb.200712154

109 Rodenas-Cuadrado, P. *et al.* Characterisation of CASPR2 deficiency disorder--a syndrome involving autism, epilepsy and language impairment. *BMC Med Genet* **17**, 8 (2016). https://doi.org:10.1186/s12881-016-0272-8

110 Penagarikano, O. & Geschwind, D. H. What does CNTNAP2 reveal about autism spectrum disorder? *Trends Mol Med* **18**, 156-163 (2012). https://doi.org:10.1016/j.molmed.2012.01.003

111 Friedman, J. I. *et al.* CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Mol Psychiatry* **13**, 261-266 (2008). https://doi.org:10.1038/sj.mp.4002049

112 Majmundar, A. J., Wong, W. J. & Simon, M. C. Hypoxia-inducible factors and the response to hypoxic stress. *Mol Cell* **40**, 294-309 (2010). https://doi.org:10.1016/j.molcel.2010.09.022

113 Anju, T. R. & Paulose, C. S. Amelioration of hypoxia-induced striatal 5-HT(2A) receptor, 5-HT transporter and HIF1 alterations by glucose, oxygen and epinephrine in neonatal rats. *Neurosci Lett* **502**, 129-132 (2011). https://doi.org:10.1016/j.neulet.2011.05.236

114 Cameron, L. P. *et al.* A non-hallucinogenic psychedelic analogue with therapeutic potential. *Nature* **589**, 474-479 (2021). https://doi.org:10.1038/s41586-020-3008-z

115 Sawa, A. & Snyder, S. H. Schizophrenia: diverse approaches to a complex disease. *Science* **296**, 692-695 (2002). https://doi.org:10.1126/science.1070532

116 Freedman, R. Schizophrenia. *N Engl J Med* **349**, 1738-1749 (2003). https://doi.org:10.1056/NEJMra035458

117 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014). https://doi.org:10.1038/nature13595

118 Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022). https://doi.org:10.1038/s41586-022-04434-5

119 Farrell, M. S. *et al.* Evaluating historical candidate genes for schizophrenia. *Mol Psychiatry* **20**, 555-562 (2015). https://doi.org:10.1038/mp.2015.16

120 Hallmayer, J. *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* **68**, 1095-1102 (2011). https://doi.org:10.1001/archgenpsychiatry.2011.76

121 Cardno, A. G. & Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* **97**, 12-17 (2000).

122 Hilker, R. *et al.* Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry* **83**, 492-498 (2018). https://doi.org:10.1016/j.biopsych.2017.08.017

123 Brown, A. S. *et al.* Serologic evidence of prenatal influenza in the etiology of schizophrenia. *Arch Gen Psychiatry* **61**, 774-780 (2004). https://doi.org:10.1001/archpsyc.61.8.774

124 Yudofsky, S. C. Contracting schizophrenia: lessons from the influenza epidemic of 1918-1919. *JAMA* **301**, 324-326 (2009). https://doi.org:10.1001/jama.2008.980

125 Malaspina, D. *et al.* Acute maternal stress in pregnancy and schizophrenia in offspring: a cohort prospective study. *BMC Psychiatry* **8**, 71 (2008). https://doi.org:10.1186/1471-244X-8-71

126 Graff, J. & Tsai, L. H. Histone acetylation: molecular mnemonics on the chromatin. *Nat Rev Neurosci* **14**, 97-111 (2013). https://doi.org:10.1038/nrn3427

127 Bastle, R. M. & Maze, I. Chromatin Regulation in Complex Brain Disorders. *Curr Opin Behav Sci* **25**, 57-65 (2019). https://doi.org:10.1016/j.cobeha.2018.07.004

128 Aberg, K. A. *et al.* Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry* **71**, 255-264 (2014). https://doi.org:10.1001/jamapsychiatry.2013.3730

129 Mendizabal, I. *et al.* Cell type-specific epigenetic links to schizophrenia risk in the brain. *Genome Biol* **20**, 135 (2019). https://doi.org:10.1186/s13059-019-1747-7

130 Jaffe, A. E. *et al.* Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci* **19**, 40-47 (2016). https://doi.org:10.1038/nn.4181

131 Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat Commun* **9**, 3121 (2018). https://doi.org:10.1038/s41467-018-05379-y

132 Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181 (2005). https://doi.org:10.1016/j.cell.2005.01.001

133 Won, H., Huang, J., Opland, C. K., Hartl, C. L. & Geschwind, D. H. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat Commun* **10**, 2396 (2019). https://doi.org:10.1038/s41467-019-10248-3

134 Dincer, A. *et al.* Deciphering H3K4me3 broad domains associated with gene-regulatory networks and conserved epigenomic landscapes in the human brain. *Transl Psychiatry* **5**, e679 (2015). https://doi.org:10.1038/tp.2015.169

135 Girdhar, K. *et al.* Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat Neurosci* **21**, 1126-1136 (2018). https://doi.org:10.1038/s41593-018-0187-0

136 Gusev, F. E. *et al.* Chromatin profiling of cortical neurons identifies individual epigenetic signatures in schizophrenia. *Transl Psychiatry* **9**, 256 (2019). https://doi.org:10.1038/s41398-019-0596-1

137 Ma, S. *et al.* Cell-type-specific brain methylomes profiled via ultralow-input microfluidics. *Nat Biomed Eng* **2**, 183-194 (2018).

138 de la Fuente Revenga, M. *et al.* Chronic clozapine treatment restrains via HDAC2 the performance of mGlu2 receptor agonism in a rodent model of antipsychotic activity. *Neuropsychopharmacology* **44**, 443-454 (2019). https://doi.org:10.1038/s41386-018-0143-4

139 Kurita, M. *et al.* HDAC2 regulates atypical antipsychotic responses through the modulation of mGlu2 promoter activity. *Nat Neurosci* **15**, 1245-1254 (2012). https://doi.org:10.1038/nn.3181

140 Ibi, D. *et al.* Antipsychotic-induced Hdac2 transcription via NF-kappaB leads to synaptic and cognitive side effects. *Nat Neurosci* **20**, 1247-1259 (2017). https://doi.org:10.1038/nn.4616

141 Andreasen, N. C. *et al.* Regional brain abnormalities in schizophrenia measured with magnetic resonance imaging. *JAMA* **272**, 1763-1769 (1994).

142 de la Fuente Revenga, M. *et al.* Prolonged epigenomic and synaptic plasticity alterations following single exposure to a psychedelic in mice. *Cell Rep* **37**, 109836 (2021). https://doi.org:10.1016/j.celrep.2021.109836

143 Marinov, G. K., Kundaje, A., Park, P. J. & Wold, B. J. Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda)* **4**, 209-223 (2014). https://doi.org:10.1534/g3.113.008680

144 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012). https://doi.org:10.1038/nature11247

145 Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol* **17**, 13 (2016). https://doi.org:10.1186/s13059-016-0881-8

146 Sun, W. *et al.* Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* **167**, 1385-1397 e1311 (2016). https://doi.org:10.1016/j.cell.2016.10.031

147 Hu, B. *et al.* Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat Commun* **12**, 3968 (2021). https://doi.org:10.1038/s41467-021-24243-0

148 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013). https://doi.org:10.1016/j.cell.2013.09.053

149 Maj, C., Minelli, A., Giacopuzzi, E., Sacchetti, E. & Gennarelli, M. The Role of Metabotropic Glutamate Receptor Genes in Schizophrenia. *Curr Neuropharmacol* **14**, 540-550 (2016). https://doi.org:10.2174/1570159x13666150514232745

150 Chen, C. *et al.* The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci Transl Med* **10** (2018). https://doi.org:10.1126/scitranslmed.aat8178

151 Ohnuma, T., Kato, H., Arai, H., McKenna, P. J. & Emson, P. C. Expression of Fyn, a non-receptor tyrosine kinase in prefrontal cortex from patients with schizophrenia and its correlation with clinical onset. *Brain Res Mol Brain Res* **112**, 90-94 (2003). https://doi.org:10.1016/s0169-328x(03)00051-2

152 Um, J. W. *et al.* LRRTM3 Regulates Excitatory Synapse Development through Alternative Splicing and Neurexin Binding. *Cell Rep* **14**, 808-822 (2016). https://doi.org:10.1016/j.celrep.2015.12.081

153 Yu, Y. *et al.* Rare loss of function mutations in N-methyl-D-aspartate glutamate receptors and their contributions to schizophrenia susceptibility. *Transl Psychiatry* **8**, 12 (2018). https://doi.org:10.1038/s41398-017-0061-y

154 Eastwood, S. L. & Harrison, P. J. Decreased mRNA expression of netrin-G1 and netrin-G2 in the temporal lobe in schizophrenia and bipolar disorder. *Neuropsychopharmacology* **33**, 933-945 (2008). https://doi.org:10.1038/sj.npp.1301457

155 Ai, R. *et al.* Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. *Nat Commun* **9**, 1921 (2018). https://doi.org:10.1038/s41467-018-04310-9

156 Zhang, K., Wang, M., Zhao, Y. & Wang, W. Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. *Sci Adv* **5**, eaav3262 (2019). https://doi.org:10.1126/sciadv.aav3262

157 Zhu, Y. *et al.* Constructing 3D interaction maps from 1D epigenomes. *Nat Commun* **7**, 10812 (2016). https://doi.org:10.1038/ncomms10812

158 Yu, B. *et al.* Epigenetic landscapes reveal transcription factors that regulate CD8(+) T cell differentiation. *Nat Immunol* **18**, 573-582 (2017). https://doi.org:10.1038/ni.3706

159 Spruston, N. Pyramidal neurons: dendritic structure and synaptic integration. *Nat Rev Neurosci* **9**, 206-221 (2008). https://doi.org:10.1038/nrn2286

160 Glantz, L. A. & Lewis, D. A. Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. *Arch Gen Psychiatry* **57**, 65-73 (2000). https://doi.org:10.1001/archpsyc.57.1.65

161 Black, J. E. *et al.* Pathology of layer V pyramidal neurons in the prefrontal cortex of patients with schizophrenia. *Am J Psychiatry* **161**, 742-744 (2004). https://doi.org:10.1176/appi.ajp.161.4.742

162 Cao, D. *et al.* Signaling Scaffold Protein IQGAP1 Interacts with Microtubule Plus-end Tracking Protein SKAP and Links Dynamic Microtubule Plus-end to Steer Cell Migration. *J Biol Chem* **290**, 23766-23780 (2015). https://doi.org:10.1074/jbc.M115.673517

163 Merenlender-Wagner, A. *et al.* Autophagy has a key role in the pathophysiology of schizophrenia. *Mol Psychiatry* **20**, 126-132 (2015). https://doi.org:10.1038/mp.2013.174

164 Rubio, M. D., Wood, K., Haroutunian, V. & Meador-Woodruff, J. H. Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia. *Neuropsychopharmacology* **38**, 1910-1920 (2013). https://doi.org:10.1038/npp.2013.84

165 Aylon, Y. & Oren, M. Living with p53, dying of p53. *Cell* **130**, 597-600 (2007). https://doi.org:10.1016/j.cell.2007.08.005

166 Endo, Y. *et al.* Regulation of clathrin-mediated endocytosis by p53. *Genes Cells* **13**, 375-386 (2008). https://doi.org:10.1111/j.1365-2443.2008.01172.x

167 O'Connor, J. A., Muly, E. C., Arnold, S. E. & Hemby, S. E. AMPA receptor subunit and splice variant expression in the DLPFC of schizophrenic subjects and rhesus monkeys chronically administered antipsychotic drugs. *Schizophr Res* **90**, 28-40 (2007). https://doi.org:10.1016/j.schres.2006.10.004

168 Cussac, D., Duqueyroix, D., Newman-Tancredi, A. & Millan, M. J. Stimulation by antipsychotic agents of mitogen-activated protein kinase (MAPK) coupled to cloned, human (h)serotonin (5-HT)(1A) receptors. *Psychopharmacology (Berl)* **162**, 168-177 (2002). https://doi.org:10.1007/s00213-002-1043-0

169 Browning, J. L. *et al.* Clozapine and the mitogen-activated protein kinase signal transduction pathway: implications for antipsychotic actions. *Biol Psychiatry* **57**, 617-623 (2005). https://doi.org:10.1016/j.biopsych.2004.12.002

170 Davidson, M. *et al.* Severity of symptoms in chronically institutionalized geriatric schizophrenic patients. *Am J Psychiatry* **152**, 197-207 (1995). https://doi.org:10.1176/ajp.152.2.197

171 Wan, W., Xia, S., Kalionis, B., Liu, L. & Li, Y. The role of Wnt signaling in the development of Alzheimer's disease: a potential therapeutic target? *Biomed Res Int* **2014**, 301575 (2014). https://doi.org:10.1155/2014/301575

172 Liu-Seifert, H., Adams, D. H. & Kinon, B. J. Discontinuation of treatment of schizophrenic patients is driven by poor symptom response: a pooled post-hoc

analysis of four atypical antipsychotic drugs. *BMC Med* **3**, 21 (2005). https://doi.org:10.1186/1741-7015-3-21

173 Lau, S. F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc Natl Acad Sci U S A* **117**, 25800-25809 (2020). https://doi.org:10.1073/pnas.2008762117

174 Nagy, C. *et al.* Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat Neurosci* **23**, 771-781 (2020). https://doi.org:10.1038/s41593-020-0621-y

175 Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol Rev* **20**, 327-348 (2010). https://doi.org:10.1007/s11065-010-9148-4

176 Garay, P. A. & McAllister, A. K. Novel roles for immune molecules in neural development: implications for neurodevelopmental disorders. *Front Synaptic Neurosci* **2**, 136 (2010). https://doi.org:10.3389/fnsyn.2010.00136

177 Bergdolt, L. & Dunaevsky, A. Brain changes in a maternal immune activation model of neurodevelopmental brain disorders. *Prog Neurobiol* **175**, 1-19 (2019). https://doi.org:10.1016/j.pneurobio.2018.12.002

178 Bale, T. L. Epigenetic and transgenerational reprogramming of brain development. *Nat Rev Neurosci* **16**, 332-344 (2015). https://doi.org:10.1038/nrn3818

179 Solek, C. M., Farooqi, N., Verly, M., Lim, T. K. & Ruthazer, E. S. Maternal immune activation in neurodevelopmental disorders. *Dev Dyn* **247**, 588-619 (2018). https://doi.org:10.1002/dvdy.24612

180 Instanes, J. T. *et al.* Attention-Deficit/Hyperactivity Disorder in Offspring of Mothers With Inflammatory and Immune System Diseases. *Biol Psychiatry* **81**, 452-459 (2017). https://doi.org:10.1016/j.biopsych.2015.11.024

181 Estes, M. L. & McAllister, A. K. Maternal immune activation: Implications for neuropsychiatric disorders. *Science* **353**, 772-777 (2016). https://doi.org:10.1126/science.aag3194

182 Careaga, M., Murai, T. & Bauman, M. D. Maternal Immune Activation and Autism Spectrum Disorder: From Rodents to Nonhuman and Human Primates. *Biol Psychiatry* **81**, 391-401 (2017). https://doi.org:10.1016/j.biopsych.2016.10.020

183 Canetta, S. E. & Brown, A. S. Prenatal Infection, Maternal Immune Activation, and Risk for Schizophrenia. *Transl Neurosci* **3**, 320-327 (2012). https://doi.org:10.2478/s13380-012-0045-6

184 Brown, A. S. & Meyer, U. Maternal Immune Activation and Neuropsychiatric Illness: A Translational Research Perspective. *Am J Psychiatry* **175**, 1073-1083 (2018). https://doi.org:10.1176/appi.ajp.2018.17121311

185 Knuesel, I. *et al.* Maternal immune activation and abnormal brain development across CNS disorders. *Nat Rev Neurol* **10**, 643-660 (2014). https://doi.org:10.1038/nrneurol.2014.187

186 Susser, E. S., Schaefer, C. A., Brown, A. S., Begg, M. D. & Wyatt, R. J. The design of the prenatal determinants of schizophrenia study. *Schizophr Bull* **26**, 257-273 (2000). https://doi.org:10.1093/oxfordjournals.schbul.a033451

187 Meyer, U. Prenatal poly(i:C) exposure and other developmental immune activation models in rodent systems. *Biol Psychiatry* **75**, 307-315 (2014). https://doi.org:10.1016/j.biopsych.2013.07.011

188 Reisinger, S. *et al.* The poly(I:C)-induced maternal immune activation model in preclinical neuropsychiatric drug discovery. *Pharmacol Ther* **149**, 213-226 (2015). https://doi.org:10.1016/j.pharmthera.2015.01.001

189 Lin, Y. L., Lin, S. Y. & Wang, S. Prenatal lipopolysaccharide exposure increases anxiety-like behaviors and enhances stress-induced corticosterone responses in adult rats. *Brain Behav Immun* **26**, 459-468 (2012). https://doi.org:10.1016/j.bbi.2011.12.003

190 Fortier, M. E., Luheshi, G. N. & Boksa, P. Effects of prenatal infection on prepulse inhibition in the rat depend on the nature of the infectious agent and the stage of pregnancy. *Behav Brain Res* **181**, 270-277 (2007). https://doi.org:10.1016/j.bbr.2007.04.016

191 Fortier, M. E., Joober, R., Luheshi, G. N. & Boksa, P. Maternal exposure to bacterial endotoxin during pregnancy enhances amphetamine-induced

locomotion and startle responses in adult rat offspring. *J Psychiatr Res* **38**, 335-345 (2004). https://doi.org:10.1016/j.jpsychires.2003.10.001

192 Coyle, P., Tran, N., Fung, J. N., Summers, B. L. & Rofe, A. M. Maternal dietary zinc supplementation prevents aberrant behaviour in an object recognition task in mice offspring exposed to LPS in early pregnancy. *Behav Brain Res* **197**, 210-218 (2009). https://doi.org:10.1016/j.bbr.2008.08.022

193 Engels, G. *et al.* Pregnancy-Related Immune Adaptation Promotes the Emergence of Highly Virulent H1N1 Influenza Virus Strains in Allogenically Pregnant Mice. *Cell Host Microbe* **21**, 321-333 (2017). https://doi.org:10.1016/j.chom.2017.02.020

194 Shi, L., Fatemi, S. H., Sidwell, R. W. & Patterson, P. H. Maternal influenza infection causes marked behavioral and pharmacological changes in the offspring. *J Neurosci* **23**, 297-302 (2003).

195 Harvey, L. & Boksa, P. Prenatal and postnatal animal models of immune activation: relevance to a range of neurodevelopmental disorders. *Dev Neurobiol* **72**, 1335-1348 (2012). https://doi.org:10.1002/dneu.22043

196 Smith, S. E., Li, J., Garbett, K., Mirnics, K. & Patterson, P. H. Maternal immune activation alters fetal brain development through interleukin-6. *J Neurosci* **27**, 10695-10702 (2007). https://doi.org:10.1523/JNEUROSCI.2178-07.2007

197 Choi, G. B. *et al.* The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. *Science* **351**, 933-939 (2016). https://doi.org:10.1126/science.aad0314

198 Garay, P. A., Hsiao, E. Y., Patterson, P. H. & McAllister, A. K. Maternal immune activation causes age- and region-specific changes in brain cytokines in offspring throughout development. *Brain Behav Immun* **31**, 54-68 (2013). https://doi.org:10.1016/j.bbi.2012.07.008

199 Nakamura, J. P. *et al.* The maternal immune activation model uncovers a role for the Arx gene in GABAergic dysfunction in schizophrenia. *Brain Behav Immun* **81**, 161-171 (2019). https://doi.org:10.1016/j.bbi.2019.06.009

200 McColl, E. R. & Piquette-Miller, M. Poly(I:C) alters placental and fetal brain amino acid transport in a rat model of maternal immune activation. *Am J Reprod Immunol* **81**, e13115 (2019). https://doi.org:10.1111/aji.13115

201 Yockey, L. J. & Iwasaki, A. Interferons and Proinflammatory Cytokines in Pregnancy and Fetal Development. *Immunity* **49**, 397-412 (2018). https://doi.org:10.1016/j.immuni.2018.07.017

202 Minakova, E. & Warner, B. B. Maternal immune activation, central nervous system development and behavioral phenotypes. *Birth Defects Res* **110**, 1539-1550 (2018). https://doi.org:10.1002/bdr2.1416

203 Labouesse, M. A., Dong, E., Grayson, D. R., Guidotti, A. & Meyer, U. Maternal immune activation induces GAD1 and GAD2 promoter remodeling in the offspring prefrontal cortex. *Epigenetics* **10**, 1143-1155 (2015). https://doi.org:10.1080/15592294.2015.1114202

204 Tang, B., Jia, H., Kast, R. J. & Thomas, E. A. Epigenetic changes at gene promoters in response to immune activation in utero. *Brain Behav Immun* **30**, 168-175 (2013). https://doi.org:10.1016/j.bbi.2013.01.086

205 Connor, C. M. *et al.* Maternal immune activation alters behavior in adult offspring, with subtle changes in the cortical transcriptome and epigenome. *Schizophr Res* **140**, 175-184 (2012). https://doi.org:10.1016/j.schres.2012.06.037

206 Canetta, S. *et al.* Maternal immune activation leads to selective functional deficits in offspring parvalbumin interneurons. *Mol Psychiatry* **21**, 956-968 (2016). https://doi.org:10.1038/mp.2015.222

207 Moffat, J. J., Ka, M., Jung, E. M., Smith, A. L. & Kim, W. Y. The role of MACF1 in nervous system development and maintenance. *Semin Cell Dev Biol* **69**, 9-17 (2017). https://doi.org:10.1016/j.semcdb.2017.05.020

208 Moon, A. L., Haan, N., Wilkinson, L. S., Thomas, K. L. & Hall, J. CACNA1C: Association With Psychiatric Disorders, Behavior, and Neurogenesis. *Schizophr Bull* **44**, 958-965 (2018). https://doi.org:10.1093/schbul/sby096

209 Rodan, L. H. *et al.* Phenotypic expansion of CACNA1C-associated disorders to include isolated neurological manifestations. *Genet Med* **23**, 1922-1932 (2021). https://doi.org:10.1038/s41436-021-01232-8

210 Dedic, N. *et al.* Cross-disorder risk gene CACNA1C differentially modulates susceptibility to psychiatric disorders during development and adulthood. *Mol Psychiatry* **23**, 533-543 (2018). https://doi.org:10.1038/mp.2017.133

211 Mirnics, K., Middleton, F. A., Stanwood, G. D., Lewis, D. A. & Levitt, P. Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia. *Mol Psychiatry* **6**, 293-301 (2001). https://doi.org:10.1038/sj.mp.4000866

212 Erdely, H. A., Tamminga, C. A., Roberts, R. C. & Vogel, M. W. Regional alterations in RGS4 protein in schizophrenia. *Synapse* **59**, 472-479 (2006). https://doi.org:10.1002/syn.20265

213 Hayden, M. S. & Ghosh, S. NF-kappaB in immunobiology. *Cell Res* **21**, 223-244 (2011). https://doi.org:10.1038/cr.2011.13

214 Hiscott, J., Kwon, H. & Genin, P. Hostile takeovers: viral appropriation of the NF-kappaB pathway. *J Clin Invest* **107**, 143-151 (2001). https://doi.org:10.1172/JCI11918

215 Koo, J. W., Russo, S. J., Ferguson, D., Nestler, E. J. & Duman, R. S. Nuclear factor-kappaB is a critical mediator of stress-impaired neurogenesis and depressive behavior. *Proc Natl Acad Sci U S A* **107**, 2669-2674 (2010). https://doi.org:10.1073/pnas.0910658107

216 Patrich, E., Piontkewitz, Y., Peretz, A., Weiner, I. & Attali, B. Maternal immune activation produces neonatal excitability defects in offspring hippocampal neurons from pregnant rats treated with poly I:C. *Sci Rep* **6**, 19106 (2016). https://doi.org:10.1038/srep19106

217 Fatemi, S. H. *et al.* Maternal infection leads to abnormal gene regulation and brain atrophy in mouse offspring: implications for genesis of neurodevelopmental disorders. *Schizophr Res* **99**, 56-70 (2008). https://doi.org:10.1016/j.schres.2007.11.018

218 da Silveira, V. T. *et al.* Effects of early or late prenatal immune activation in mice on behavioral and neuroanatomical abnormalities relevant to schizophrenia in the adulthood. *Int J Dev Neurosci* **58**, 1-8 (2017). https://doi.org:10.1016/j.ijdevneu.2017.01.009

219 Oh-Nishi, A., Obayashi, S., Sugihara, I., Minamimoto, T. & Suhara, T. Maternal immune activation by polyriboinosinic-polyribocytidilic acid injection produces synaptic dysfunction but not neuronal loss in the hippocampus of juvenile rat offspring. *Brain Res* **1363**, 170-179 (2010). https://doi.org:10.1016/j.brainres.2010.09.054

220 Dickerson, D. D., Wolff, A. R. & Bilkey, D. K. Abnormal long-range neural synchrony in a maternal immune activation animal model of schizophrenia. *J Neurosci* **30**, 12424-12431 (2010). https://doi.org:10.1523/JNEUROSCI.3046-10.2010

221 Suomi, S. J., Collins, M. L., Harlow, H. F. & Ruppenthal, G. C. Effects of maternal and peer separations on young monkeys. *J Child Psychol Psychiatry* **17**, 101-112 (1976). https://doi.org:10.1111/j.1469-7610.1976.tb00382.x

222 Harlow, H. F., Dodsworth, R. O. & Harlow, M. K. Total social isolation in monkeys. *Proc Natl Acad Sci U S A* **54**, 90-97 (1965). https://doi.org:10.1073/pnas.54.1.90

223 Tomalski, P. & Johnson, M. H. The effects of early adversity on the adult and developing brain. *Curr Opin Psychiatry* **23**, 233-238 (2010). https://doi.org:10.1097/YCO.0b013e3283387a8c

224 Mueller, F. S. *et al.* Behavioral, neuroanatomical, and molecular correlates of resilience and susceptibility to maternal immune activation. *Mol Psychiatry* **26**, 396-410 (2021). https://doi.org:10.1038/s41380-020-00952-8

225 Duclot, F. & Kabbaj, M. The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders. *Front Behav Neurosci* **11**, 35 (2017). https://doi.org:10.3389/fnbeh.2017.00035

226 Hashimoto, R. *et al.* Variants of the RELA gene are associated with schizophrenia and their startle responses. *Neuropsychopharmacology* **36**, 1921-1931 (2011). https://doi.org:10.1038/npp.2011.78

227 Warre-Cornish, K. *et al.* Interferon-gamma signaling in human iPSC-derived neurons recapitulates neurodevelopmental disorder phenotypes. *Sci Adv* **6**, eaay9506 (2020). https://doi.org:10.1126/sciadv.aay9506

228 Tsai, M. H., Pai, L. M. & Lee, C. K. Fine-Tuning of Type I Interferon Response by STAT3. *Front Immunol* **10**, 1448 (2019). https://doi.org:10.3389/fimmu.2019.01448

229 Palma-Gudiel, H., Cordova-Palomera, A., Leza, J. C. & Fananas, L. Glucocorticoid receptor gene (NR3C1) methylation processes as mediators of early adversity in stress-related disorders causality: A critical review. *Neurosci Biobehav Rev* **55**, 520-535 (2015). https://doi.org:10.1016/j.neubiorev.2015.05.016

230 Trump, S. *et al.* Prenatal maternal stress and wheeze in children: novel insights into epigenetic regulation. *Sci Rep* **6**, 28616 (2016). https://doi.org:10.1038/srep28616

231 Stroud, L. R. *et al.* Epigenetic Regulation of Placental NR3C1: Mechanism Underlying Prenatal Programming of Infant Neurobehavior by Maternal Smoking? *Child Dev* **87**, 49-60 (2016). https://doi.org:10.1111/cdev.12482

232 Savalia, N. K., Shao, L. X. & Kwan, A. C. A Dendrite-Focused Framework for Understanding the Actions of Ketamine and Psychedelics. *Trends Neurosci* **44**, 260-275 (2021). https://doi.org:10.1016/j.tins.2020.11.008

233 Babinski, D. E., Neely, K. A., Ba, D. M. & Liu, G. Depression and Suicidal Behavior in Young Adult Men and Women With ADHD: Evidence From Claims Data. *J Clin Psychiatry* **81** (2020). https://doi.org:10.4088/JCP.19m13130

234 Grace, A. A. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nat Rev Neurosci* **17**, 524-532 (2016). https://doi.org:10.1038/nrn.2016.57

235 Gough, A. & Morrison, J. Managing the comorbidity of schizophrenia and ADHD. *J Psychiatry Neurosci* **41**, E79-80 (2016).

Publications

Journal Papers

1. **B. Zhu**, J. Saunders, L. B. Naler, J. Gonzalez-Maeso, C. Lu. The effect of maternal immune activation on epigenomics and transcriptomics in the frontal cortex of mice offspring. *In preparation.*

2. **B. Zhu**, R. Ainsworth, Z. Wang, S. Sierra, C. Deng, L. F. Callado, J.J. Meana, W. Wang, C. Lu, J. Gonzalez-Maeso. Antipsychotic-induced epigenomic reorganization in frontal cortex samples from individuals with schizophrenia. *bioRxiv,* (2021); Submitted.

3. **B. Zhu**, YP Hsieh, T. W. Murphy, Q. Zhang, L. B. Naler, C. Lu. MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nature protocols*, 14, 3366-3394 (2019)

4. M. Revenga, **B. Zhu**, C. A. Guevara, L. B. Naler, J. M. Saunders, Z. Zhou, R. Toneatti, S. Sierra, J. T. Wolstenholme, P. M. Beardsley, G. W. Huntley, C. Lu, J. Gonzalez-Maeso. Prolonged epigenetic and synaptic plasticity alterations following single exposure to a psychedelic in mice. *Cell Reports*, 37 (3), 109836 (2021)

5. Z. Liu, L. B. Naler, Y. Zhu, C. Deng, Q. Zhang, **B. Zhu**, Z. Zhou, M. Sarma, A. Murray, H. Xie, C. Lu. Native MOWChIPseq: Genome-wide profiles of key protein bindings reveal functional differences among various brain regions. *NAR Genomics and Bioinformatics*, 4 (2), lqac030 (2022)

6. T. W. Murphy, YP Hsieh, **B. Zhu**, L. B. Naler, C. Lu. Microfluidic Platform for Next-Generation Sequencing Library Preparation with Low-Input Samples. *Analytical chemistry*, 92, 2519-2526 (2020)

7. **B. Zhu**, M. G. Barnes, H. Kim, M. Yuan, H. Ardebili and R. Verduzco. Molecular engineering of step-growth liquid crystal elastomers. *Sensors and Actuators B: Chemical*, 244, 433-440 (2017)

8. A. Agrawal, H. Chen, **B. Zhu**, H. Kim, O. Adetiba, A. Miranda, A. Chipara, P. Ajayan, J. G. Jacot, R. Verduzco. Electromechanically responsive liquid crystal elastomer nanocomposites for active cell culture. *ACS Macro Letters*, 5, 1386–1390 (2016)

9. H. Kim, **B. Zhu**, H. Chen, O. Adetiba, A. Agrawal, P. Ajayan, J. G. Jacot, R. Verduzco. Preparation of Monodomain Liquid Crystal Elastomers and Liquid Crystal Elastomer Nanocomposites. *JoVE*, e53688 (2016)