MICROFLUIDICS FOR LOW INPUT EPIGENOMIC ANALYSIS AND ITS APPLICATION TO BRAIN NEUROSCIENCE

CHENGYU DENG

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
Rafael V. Davalos
Aaron S. Goldstein
Rong Tong

December 7, 2020
Blacksburg, Virginia

Keywords: Microfluidics, Chromatin immunoprecipitation, next generation sequencing, histone modifications

MICROFLUIDICS FOR LOW INPUT EPIGENOMIC ANALYSIS AND ITS APPLICATION TO BRIAN NEUROSCIENCE

CHENGYU DENG

ABSTRACT

The epigenome carries dynamic information that controls gene expression and maintains cell identity during both disease and normal development. The inherent plasticity of the epigenome paves new avenues for developing diagnostic and therapeutic tools for human diseases. Microfluidic technology has improved the sensitivity and resolution of epigenomic analysis due to its outstanding ability to manipulate nanoliter-scale liquid volumes. In this thesis, I report three projects focusing on low-input, cell-type-specific and spatially resolved histone modification profiling on microfluidic platforms. First, I applied Microfluidic Oscillatory Washing-based Chromatin Immunoprecipitation followed by sequencing (MOWChIP-seq) to study the effect of culture dimensionality, hypoxia stress and bacterium infection on histone modification landscapes of brain tumor cells. I identified differentially marked regions between different culture conditions. Second, I adapted indexed ChIPmentation and introduced mu-CM, a low-input microfluidic device capable of performing 8 assays in parallel on different histone marks using as few as 20 cells in less than 7 hours. Last, I investigated the spatially resolved epigenome and transcriptome of neuronal and glial cells from coronal sections of adult mouse neocortex. I applied unsupervised clustering to identify distinct spatial patterns in neocortex epigenome and transcriptome that were associated with central nervous system development. I demonstrated that our method is well suited for scarce samples, such as biopsy samples from patients in the context of precision medicine.

MICROFLUIDICS FOR LOW INPUT EPIGENOMIC ANALYSIS AND ITS APPLICATION TO BRIAN NEUROSCIENCE

CHENGYU DENG

GENERAL AUDIENCE ABSTRACT

Epigenetic is the study of alternations in organisms not caused by alternation of the genetic codes. Epigenetic information plays pivotal role during growth, aging and disease. Epigenetic information is dynamic and modifiable, and thus serves as an ideal target for various diagnostic and therapeutic strategies of human diseases. Microfluidics is a technology that manipulates liquids with extremely small volumes in miniaturized devices. Microfluidics has improved the sensitivity and resolution of epigenetic analysis. In this thesis, I report three projects focusing on low-input, cell-type-specific and spatially resolved histone modification profiling on microfluidic platforms. Histone modification is one type of epigenetic information and regulates gene expression. First, we studied the influence of culture condition and bacterium infection on histone modification profile of brain tumor cells. Second, we introduced mu-CM, combining a low-input microfluidic device with indexed ChIPmentation and is capable of performing 8 assays in parallel using as few as 20 cells. Last, we investigated spatial variations in the epigenome and transcriptome across adult mouse neocortex, the outer layer of brain involving in higher-order function, such as cognition. I identified distinct spatial patterns responsible for central nervous system development using machine learning algorithm. Our method is well suited for studying scarce samples, such as cells populations isolated from patients in the context of precision medicine.

ACKNOWLEDGEMENTS

The completion of this dissertation would not have been accomplished without the support from my teachers, family and friends. I would like to express my deep gratitude to all of them for their love and patience.

First of all, I would like to thank my advisor and head of my committee, Dr. Chang Lu, for his guidance and support. Dr. Lu is the one that leads me into the wonderful field of microfluidics and epigenomics. I greatly appreciate all his contribution of time, concerns, and financial support to make my PhD journal a rich experience.

I would also like to thank the rest of my committee members, Dr. Rafael Davalos, Dr. Aaron Goldstein and Dr. Rong Tong for their insightful feedbacks and suggestions. Their support during my prelim, dissertation and job hunting have been priceless. I would like to thank Dr. Scott Verbridge and his graduate student Megan Cox. Their knowledge in 3D tumor model inspired me for my exploration in epigenomics. I would also like to thank Dr. Daphne Yao and her graduate student Ya Xiao. Their expertise in machine learning greatly helped us better understand our sequencing data.

I am very grateful to the department of Chemical Engineering at Virginia Tech. The staffs have offered their help and support for numerous times. The faculty members have provided lectures, comments and help along my PhD.

I also want to thank my current and former labmates, Zhenning Cao, Chen Sun, Sai Ma, Yan Zhu, Travis Murphy, Mimosa Sharma, Yuan-pang Hsieh, Qiang Zhang, Lynette Naler, Bohan Zhu, Zhengzhi Liu, Zirui Zhou. Their support and help in my research and personal life made my PhD life enjoyable.

I would like to thank my family for their continuous love and support. Especially, I want to thank my parents Yaping Deng and Xiaorong Cheng, who I always love and admire. They have always been there for me, and their love has been a forever inspiration to me.

Last but not least, I want to thank my partner, Yao Xiao for his unconditional love and help. He always listened attentively and made me feel I can talk openly about my frustration, anxiety and fear. He always did his best to support my career. With all the love and encourage, every step of the way is a wonderful adventure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SYMBOL	*DESCRIPTION*
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ATAC-seq	Assay for transposase accessible chromatin using sequencing
CGI	CpG island
ChIP-seq	Chromatin immunoprecipitation followed by next-generation sequencing
CpG site	5'-C-phosphate-G-3
CRISPR	Clustered regularly interspaced short palindromic repeats
Ct	Cycle threshold
CUT&RUN	Cleavage under target & release using nuclease
CUT&Tag	Cleavage under targets and tagmentation
DE	Differentially expressed
DI water	Deionized water
DNase-seq	DNase I hypersensitive sites sequencing
dNTP	Nucleotide triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ENCODE	Encyclopedia of DNA Elements
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
FISH	Fluorescence in situ hybridization
FPKM	Fragments per kilobase of transcript per million aligned reads
GO	Gene ontology
H3K4me1	Histone 3 lysine 4 mono-methylation

H3K4me3	Histone 3 lysine 4 tri-methylation
H3K27me3	Histone 3 lysine 27 tri-methylation
HDAC	Histone deacetylase
HOTAIR	HOX antisense intergenic RNA
HOTTIP	HOXA transcript at the distal tip
LAD	Lamina associated domains
lncRNA	Long non-coding RNA
LOCK	Large organized chromatin K9-modification
MBD1	Methyl-CpG-binding domain protein 1
MECP2	Methyl-CpG-binding protein
MeDIP-seq	Methylated DNA immunoprecipitation coupled with sequencing
miRNA	Micro RNA
MNase	Micrococcal Nuclease
MNase-seq	Micrococcal nuclease digestion with sequencing
MOWChIP	Microfluidic oscillatory washing-based ChIP
mRNA-seq	mRNA sequencing
NGS	Next generation sequencing
NTCs	Non template controls
pA	Protein A
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDMS	Polydimethylsiloxane
PEG	Polyethylene glycol
PIC	Protease inhibitor cocktail
PMSF	Phenylmethylsulfonyl fluoride
PRC	Polycomb repressive complex
qPCR	Quantitative PCR

RRBS	Reduced representation bisulfite sequencing
SBS	Sequencing-by-synthesis
SDS	Sodium dodecyl sulfate
TF	Transcription factor
Tn5	Transposase
WGBS	Whole genome bisulfite sequencing

Chapter 1 - Overview

The research scope of this thesis focuses on studying the epigenome of low-abundance samples using microfluidics. The small dimensions of microfluidic devices have made them extremely suitable for low-input low efficient epigenomic assays. This dissertation presents three projects including developing ultralow-input microfluidic devices for ChIP-seq based on indexed ChIPmentation and applying low-input microfluidic assays to study the influence of culture condition on epigenomic dynamics and spatially resolved epigenomic and transcriptomic variations in mammalian brain.

Epigenetics investigates how cell phenotype can change while the DNA sequence remains unchanged. The genome consists of the entire information and is the blueprint of life while the epigenome determines how the blueprint is interpreted in each cell. The epigenome is dynamic and can be changed by genetic alternations, environments or random factors. Hence epigenetics stand at the interface of the genome, environment and phenotypes, and provides a window to investigate the influence of the environment on phenotypes and to gain insight into pathogenesis(*1*).

Histone modification is a major type of epigenomic regulation and refers to the post-translational chemical modification of N-terminal tail of histone protein and histone variant modification. Histone modification controls gene expression by recruiting transcription factors, activating transcriptional enhancers, and interacting with other epigenetic machinery, such as DNA methylation. Hundreds of histone modifications have been identified, yet their functions are not fully understood(*2, 3*).

Abnormal alternation of epigenome has been associated with various diseases including cancer. It has been found that epigenetic changes cause or precede with many human diseases, like Beckwith-Wiedemann syndrome. On the other hand, epigenetics stands at the interface of cell phenotypes and their microenvironments. Hence profiling the epigenome under different

microenvironment stresses provides a better understanding of the pathogenesis of diseases and enable the observation of variable responses to novel drugs or therapies. We applied Microfluidic-oscillatory-washing-based ChIP-seq (MOWChIP-seq) and successfully profiled histone 3 lysine 4 tri-methylation (H3K4me3) histone mark from 1,000 brain tumor cells under normoxic/hypoxic and 2D/3D culture. We identified differentially marked regions between different culture conditions (Chapter 3).

Conventional gold-standard method to profile histone modifications is chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq). The key principle is immunoprecipitation of specific histone modifications and the DNA fragment bound to them. ChIP-seq includes several main steps: (i) Extract chromatin from cell and fragment it to small pieces; (ii) Use antibody that specifically bind to histone modification of our interest; (iii) The antibody is coupled with magnetic beads for easy separation of our target; (iv) After antibody-antigen binding and separation, purify DNA from the mixture; (v) Sequence DNA with Illumina sequencer machine and analysis the histone modification profiles.

Major limitations of current ChIP-seq method are the large number of cells required and the time-consuming process. Low antibody binding efficiency and material loss during purification and sample transferring are two reasons for high input amount. We developed mu-CM (microfluidic ChIPmentation), a low-input method by combining a tagmentation-based indexing step (i.e. ChIPmentation) with a simple and multiplexed microfluidic device. Our method permitted processing 8 ChIP assays in one run with as few as 20 cells per assay (Chapter 4).

Low-input microfluidic epigenomic analysis also expands the scope of information we can obtain from primary tissues with small number of cells. In mammalian brain, the neocortex is the part essential for higher functions such as sensory perception, motor and cognitive behaviors. Advancing our understanding of its structure and function requires identification of

novel regulatory elements. Histone 3 lysine 27 tri-methylation (H3K27me3) is catalyzed by Polycomb repressive complexes (PRCs) and inhibits gene expression by impeding the recruitment of RNA polymerase II. We applied low-input ChIP-seq and RNA-seq (MOWChIP-seq and SMART-seq2) and generated spatially resolved H3K27me3 and gene expression landscapes in coronal slices of adult mouse neocortex. We applied unsupervised clustering (k-means) to identify the spatial patterns in epigenome and transcriptome of mouse neocortex, and we identified distinct spatial patterns of H3K27me3 signals that were associated particularly with central nervous system development. Our data also revealed profound different spatial patterns between neurons and glia (Chapter 5).

This thesis has three projects focusing on applying microfluidics to profile histone modification landscapes. First, we used MOWChIP-seq to study the effect of culture condition on H3K4me3 profile of brain tumor cells (Chapter 3). Second, we adapted ChIPmentation-based method and designed a simple microfluidic device for low-input ChIP named mu-CM (Chapter 4). Last, we used MOWChIP-seq and SMART-seq2 to reveal the spatially resolved H3K27me3 and gene expression profiles in adult mouse neocortex (Chapter 5). The focus of each chapter is summarized briefly:

CHAPTER 3 - Effects of Culture Condition on Epigenomic Profiles of Brain Tumor Cells

The epigenome is dynamic and constantly influenced by environmental exposure and random chance. Understanding how epigenome is altered during diseases provides insights to developing diagnostic and therapeutic strategies. Here we applied a three-dimensional cell culturee based on type I collagen hydrogels to model tumor. We profiled H3K4me3 of U251 MG glioblastoma cells in 2D cultures and 3D cultures under both normoxic and hypoxic conditions. We compared H3K4me3 profiles between different culture conditions and found

significant changes due to culture dimensionality and hypoxic stress. Our work provides the first step toward analyzing the spatiotemporal epigenome variations of cancer cell phenotypes.

CHAPTER 4 – Multiplexed and Ultralow-input ChIP-seq Enabled by Tagmentation-based Indexing and Facile Microfluidics

Extensive efforts have been made to improve the sensitivity, resolution and throughput of epigenome profiling. However, there is still a great demand for methods combining low-input, high-throughput and fast speed. We designed mu-CM (microfluidic ChIPmentation), a simple microfluidic device to facilitate simultaneous processing of multiple cell samples. The microfluidic ChIP process was combined with ChIPmentation to index each sample with unique sequencing adaptors. Our method permitted processing 8 assays in one run with as few as 20 cells per assay and the entire protocol can be finished in 7 h (compared to 1 million cells and 2~3 days required by the conventional method). Our technology is well suited for low-abundance tissue types and high-throughput research projects.

CHAPTER 5 – Spatially Resolved Epigenomic And Transcriptomic Variations In Adult Mouse Neocortex

The mammalian neocortex is essential for sensory perception and cognitive behaviors. Genome-wide assays, such as mRNA-seq and ChIP-seq, facilitate the discovery of novel candidate genes and regulatory elements in brain functions, but they lack spatial information in tissue. H3K27me3 is a repressive histone mark and involves in cell type-specific gene silencing. We sliced mouse neocortex into around 19 left and right coronal sections and performed MOWChIP-seq and SMART-seq2 to generated spatially resolved and cell-type-specific epigenomic/transcriptomic profiles of neuronal/glial cells. We observed distinct differences in H3K27me3 and gene expression profiles between neurons and glia. Also, in

contrast to glia, neurons had greater spatially restricted variation of H3K27me3 signal. To identify H3K27me3 peaks that have similar spatial variation patterns, we clustered peaks into 20 groups by k-means clustering and obtained top enriched GO terms for each cluster. We identified a few neuron clusters that were highly enriched for GO terms related with central nervous system development.

In conclusion, microfluidics has been applied to facilitate high-sensitivity chemical and biological analysis and detection since its inception. In recent years, microfluidics started to be increasingly used in conjunction with next-generation sequencing and genome-wide analysis. A significant fraction of these works included epigenomic assays. The future of microfluidic epigenomic assays will mainly focus on developing massively parallel and highly integrated systems and expanding the repertoire of microfluidic epigenomic analyses. In the next chapter, the background knowledge was described including microfluidic techniques, the development of DNA sequencing technology, and different types of epigenomic modifications. In addition, recent technical advancement of sequencing-based microfluidic epigenomic assays was also reviewed with an emphasis on methods for profiling histone modifications.

Chapter 2 - Microfluidics for Epigenomic Analysis

2.1 Introduction

Epigenetics is the study of heritable phenotype changes that do not involve alternations of DNA sequence. Epigenome is the body of genome-wide chemical structures, nucleic acids, and proteins that manipulates the DNA(*4*). DNA wrapping around histone proteins forms nucleosomes, which are then further coiled up to form chromatin(*5*). In contrast to the relatively static genome, the epigenome is highly dynamic and can be altered by environmental influences, lifestyle, and aging. For example, monozygotic twins have indistinguishable epigenetic signatures at birth, yet they have increasingly significant epigenetic differences over their lifespans(*6*). This can cause one twin to become more susceptible to a disease than the other. For instance, schizophrenia only occurs in both twins about 50% of the time(*7, 8*).

In the past two decades, the resolution of epigenomic profiling has evolved from locus-specific to genome-wide. Since the completion of human genome projects in 2003, next-generation sequencing (NGS) has gradually become the most commonly used technique in epigenomic studies. It took $3 billion, ten years, and the cooperation between hundreds of researchers from different countries to complete the human genome project. In contrast, nowadays the same information can be obtained by NGS at a few thousand dollars in one day by one technician. Despite that, NGS still suffers from short read length and sample amplification bias. A new generation of sequencing, including PacBio SMRT-seq and Oxford Nanopore, that is real-time and does not require amplification has been under development(*9, 10*).

The fundamental principle of epigenomic analysis is to isolate DNA fragments that have specific epigenetic marks and convert those DNA fragments into a sequencer-readable format. The first step usually requires antibodies to capture DNA fragments of interest while the second step involves several enzymatic reactions. The entire procedure takes multiple steps

and DNA loss is inevitable during each step. Hence epigenomic analysis generally requires a large number of cells as input materials (> 10^6 cells), which limits its application to low-abundance samples.

Microfluidic technology manipulates micro- or nano-liter scale liquid volumes in miniaturized devices. It has shown great advantages in performing chemical or biological assays as tiny reaction volume increases efficiency and reduce starting materials and reaction time. Numerous efforts have been made to adapt epigenomic analysis to microfluidic platforms with improved resolution and throughput(*11*).

In this chapter, I reviewed the history of developing DNA sequencing technology, different aspects of epigenome, and how epigenomic analysis has been greatly accelerated by new sequencing techniques. The rest part has then focused on introducing microfluidics technology and summarizing the recent advancement on microfluidics for epigenomic analysis.

2.2 Microfluidic technology

Microfluidics refers to technology that manipulate the small amount of fluids (10^{-9} to 10^{-18} liters) at dimensions of tens of micrometers, which has presented tremendous promise for biomedical research. Low volumes of fluids are moved, mixed, separated or otherwise processed to achieve multiplexing, automation and high-throughput analysis in miniaturized microfluidic devices. Advantages of microfluidic technologies, including substantially reducing the sample volume and cost, improving sensitivity, precise control of fluids and rapid sample processing, have provided unprecedented opportunities to revolutionize the way modern biology is performed.

Microfluidics involves manipulate fluids at the submillimeter length scale. The difference of fluid phenomena between microscale and macroscale is considerable. Gravity that dominates at the macroscale becomes relative negligible at microscale. Instead surface

tension and capillary forces are dominant at the microscale, which are widely used in microfluidics for passively pumping, precisely patterning surfaces, filtering, generating droplets and so on. Hence, low-Reynolds-number flow dominates at the microscale. Moreover, the surface-to-volume at the microscale is typically larger than that at macroscale. This helps to reduce sample volumes, and thus is suitable for low-input biological, biochemical and pharmaceutical reactions.

There are two main forms in the field of microfluidics based on the way liquid is handled and manipulated: continuous-flow microfluidics and droplet-based microfluidics. In continuous-flow microfluidic devices, liquid is delivered continuously by external means. Mixing and separation of the single or multiple phases flow are the major tasks. Due to the dominant laminar flow in microfluidic devices, mixing and separation tend to require large length scales. Typical mixing/separation methods includes passive (without external energy) and active (in the presence of external energy) techniques.

Droplet-based microfluidics combines the science of emulsion with microfluidics, and manipulates individual droplets, rather than continuous streams of liquid. Droplet-based microfluidics has numerous advantages, such as minimal reagent requirement, fast reaction rates and the capacity for multiplexing. Those advantages make droplet microfluidics an ideal candidate for high throughput single cell analysis.

Microfluidic devices are mostly fabricated by soft lithography using polydimethylsiloxane (PDMS). Soft lithography involves replicating a topographically defined (typically in photoresist) structure on a silicon master onto a soft elastomer. Replications can be used repeatedly to reduce the cost. Hence soft lithography enables rapid, inexpensive, and simple fabrication of microfluidic devices. Typically, a microfluidic device can be fabricated based on the following steps: (1) CAD or other software are used to design microfluidic devices and the design is printed onto a high-resolution transparency (~5000 dpi). This transparent film

is used as a photomask in 1:1 contact photolithography. (2) Photoresist is used to create a mask on the surface of a silicon wafer and serves as a mold for PDSM molding. (3) Degased liquid PDMS prepolymer is cast onto the master and cured at 65 ℃ for around 1 hour. (4) The PDMS replica is then peeled off the master and bond (following plasma oxidation) to a glass to form sealed microfluidic channels.

Microfluidics provides flexible platforms for manipulating liquids at pico/nanoliter scale in miniaturized devices. Furthermore, it has shown advantages in performing chemical and biological analysis. Compared to conventional tube-based and manual assays, microfluidic assays are often rapid, automated, and scalable for high-throughput operations. In addition, the small dimensions of microfluidics are amenable to low-input assays and integrating distinct steps to reduce the sample loss. Altogether, this has earned microfluidics the nickname "lab-on-a-chip".

Over the past decade, the development and commercialization of microfluidic technologies have led to new revolutions in biology research related to single-cell analysis and high-sensitivity molecular detections(*59*). Microfluidics has found widespread applications to cell biology(*60-66*), drug research(*67-70*), microbiology(*71, 72*), immunology(*73, 74*), and physiology(*75*).

2.3 Next-generation sequencing

Introduction of NGS

The nucleotide is the basic building block of DNA, and there are four types of nucleotides: adenine (A), cytosine (C), thymine(T), guanine(G). DNA is composed of two poly-nucleotide chains that coil around forming the well-known double helix structure. DNA sequencing is the process of determining the order of arrangement of nucleotides for a given DNA fragment.

Sanger Sequencing is the first generation of sequencing developed in 1977 and has dominated the industry for nearly two and a half decades and led to the accomplishment of Human Genome Project in 2003. It had greatly accelerated the identification of the repertoire of genes. However, the throughput and cost to sequence the entire human genome using Sanger sequencer has remained a major limitation for its wide application.

Second or next generation sequencing was developed in 2005, and it greatly outperformed Sanger sequencing in throughput, cost and speed. Over the past decade, there has been a fundamental shift from Sanger sequencing to NGS platform for genome-wide analysis. Compared to Sanger sequencing, NGS costs less and yield higher throughput. Thus, NGS has been extensively applied to biological and medical research over the past decade. There are multiple NGS platform in the marketplace, such as Roche/454 Life Science, Illumina/Solexa, Applied Biosystem/Solid. All NGS platforms share a common technological feature: massively parallel sequencing of clonally amplified or single DNA molecule that is spatially separated in a flow cell. In this way, hundreds of megabases to gigabases of nucleotide-sequence output can be generated in a single instrument run. Nowadays the market of NGS has almost been dominated by Illumina/Solexa sequencing. The latest machine Nextseq 2000 has a maximum output of 300 Gb, and a maximum 1 billion reads per run and the sequencing can be finished in less than 24 hours.

The key technology of NGS is sequencing-by-synthesis (SBS). In SBS, a standard PCR reaction is carried out, and each type of deoxynucleotides (dNTPs) is labeled with a characteristic fluorophore. During elongation, the incorporation of each dNTP is recorded based on the detected fluorescent signals. NGS is short-read sequencing, so the read length is generally shorter than Sanger sequencing (35 ~ 150 bp for NGS; 500 ~ 800 bp for Sanger). In NGS, DNA molecules are immobilized on the surface of a miniaturized flow cell, which enables easy buffer exchange by streaming reagents into the flow cell. Each molecule is

replicated thousands of times so that the fluorescent signal can be detected. Millions of DNA molecules are processed in parallel on the surface of one flow cell, thus NGS is also called massively parallel sequencing. After sequencing reads have been generated, short DNA reads are aligned to established reference genome or assembled *de novo*. The DNA sequence is then replaced with the genomic coordinates.

Despite of the advantages mentioned above, one major drawback of NGS is that its error rate (0.1~15%) is much higher than Sanger sequencing, making it hard to distinguish true genetic variation from false positive. Besides, the short reads of NGS generally provides less informations compared to long reads sequencing, which poses challenges for *de novo* assembly. The major problem of the short-reads alignment is whether the reads are uniquely aligned to the reference genome. An increasing number of alignment tools have been developed specifically for rapid alignment of large volume of short reads allowing mismatches or gaps.

One promising new technology is single-molecule real-time sequencing, including SMRT sequencing from Pacific Biosciences and Nanopore sequencing from Oxford Nanopore Technologies. The advantage of these single-molecule sequencing methods includes ultra-long reads (> 100 kb), detection of base modifications, and simple sample preparation.

Nanopore uses a distinct cylindrical nanopore protein embedded in an electrically charged membrane. An enzyme protein guided DNA through the nanopore and, as the DNA passed through, the nucleotide sequence is determined by measuring the ionic current blockade. Each nucleotide has its own characteristic electronic signature due to its different atomic structure, even methylated cytosine or adenosine, allowing the nucleotide to be sequenced in real-time[1-22]. Unmethylated CpGs can be quantified using electro-optical nanopore to determine hypomethylation levels. Moreover, DNA bound proteins, such as MBD1, MeCP2, GCD, methylcytosine antibody or TFs, can also be detected using solid-state nanopores. Similarly, modified bases, such as uracil or 8-oxoguanine, selectively labeled with biotin-dNTP,

streptavidin or ferrocene⊂ cucurbit (Fc⊂CB) can be discriminated by nanopores. In fact, microRNA detection has been reported that used PEG-labeled probe hybridizing with target miRNA to generate ionic current signatures. And several groups have reported translocating nucleosomal dsDNA to study nucleosome stability and histone variants. Nanopore sequencing has already demonstrated usefulness as a low-cost, same-day method for molecularly classifying tumors by their genetic and epigenetic profiles.

Except for aforementioned nanopore sequencing, SMRT sequencing is another single-molecule sequencing technology. It determines DNA sequence by detecting the fluorescence pulse during incorporation of fluorescently labeled nucleotides by DNA polymerase. Additionally, 5mC, 5hmC and other DNA modifications were discriminated by polymerase kinetic signature, including the arrival time and duration of the pulse. SMRT sequencing generates longer reads while nanopore sequencing has better overall data quality. The two sequencing methods have similar performance in identifying transcript, analyzing simple sequence repeat, and predicting long non-coding RNA.

Library preparation for NGS

During library preparation, input DNA samples are constructed into sequencer-compatible libraries. Both ends of input DNA molecules are ligated with adapters, which are short DNA oligos (~150 bp) containing unique indexes and primer binding sites. It contains sequence complementary to flow cell oligos, enabling the immobilization of input DNA molecules on flow cell surface. It also contains primer binding sites required to initialize the sequencing-by-synthesis process. Unique index sequences allow the sequencing of multiple samples in the same run, which greatly facilitates multiplexing and increases throughput.

Conventional library preparation methods include several steps: fragmentation, end repair, dA-tailing, ligation and PCR amplification (Fig. 2.2). RNA samples are first converted

into cDNA by reverse transcription. Long DNA fragments are cut into small pieces (< 500 bp) using either physical (i.e., acoustic shearing and sonication) or enzymatic (i.e., DNase I digestion) methods. The Illumina sequencing platform requires short input fragments because DNA is bridge amplified on the flow cell. Shorter DNA is amplified more efficiently than longer one. After fragmentation, end repair is performed to form blunt ends. Next the 3' ends are dA-tailed, which is necessary for ligation of adapters. After ligation, a few cycles of PCR amplification are performed. This step can be omitted if the amount of input material is large. The entire process generally takes a few hours and requires at least 1~10 pg of input DNA. An alternative method is tagmentation, a one-step reaction using tn5 transposase enzyme. Tn5 is able to simultaneously cut input DNA and insert adapters into both ends of the DNA. The advantages of tagmentation includes short processing time and simple hands-on operation.

2.4 Epigenomics

Epigenetics is the study of heritable changes in phenotype without changes in genotype. Epigenetics explains why cells inside a human body share the same DNA sequence while each cell type has vastly different morphology and function with others. It also explains why genetically identical twins show different phenotypes(*12*).

Epigenome determines how the genome is interpreted, or which gene is expressed, and which is inhibited. The genome is stable throughput the life span while epigenome is in dynamic change due to developmental or environmental factors, such as aging, dietary, or stochasticity. Hence epigenetics stands at the interface of genetics, environment, and phenotypes.

The epigenome is regulated by writer, eraser, and reader enzymes, which add, remove and recognize specific epigenetic marks, respectively. Epigenetic regulation of gene expression consists of three major steps. First, writer proteins add an epigenetic mark which serves as a

signal to downstream regulatory activity. Second, reader proteins bind to the epigenetic mark and interpret its encoded messages. Last, the interpreted message is responded, for example DNA structure is changed or transcription factors are recruited.

Epigenomic regulatory mechanisms include DNA methylation, histone modification, higher-order chromatin organizations, and regulations by non-coding RNAs. These mechanisms are described in detail below.

DNA methylation

In eukaryotic cells, DNA methylation usually occurs at the carbon-5 position of cytosine, known as 5-methylcytosine or 5-mC. It is well studied as a repressive epigenetic mark, when occurring in the promoter region of a gene(*13, 14*). DNA methylation is mostly found in CpG sites (a cytosine nucleotide followed by a guanine nucleotide separated by a phosphate) while rarely found in CpG islands (CpG dense region), especially those located in the promoter regions of actively expressed genes. It has been observed that global patterns of DNA methylation changes rapidly during fertilization and embryonic progression. After that it remains stable throughout life.

DNA methylation plays important roles during cell differentiation and development. For example, genomic imprinting (genes are expressed in an allele-specific manner) relies on appropriate establishment of DNA methylation. Abnormal imprinting causes several fatal diseases, such as Angelman syndrome.

Histone modification

In eukaryotic cells, DNA is stored in a highly compacted manner. Approximately 150 bp DNA wrapping around histone proteins forms the nucleosome and the “beads-on-a-string” structure. Histone proteins consist of five subunits: H2A, H2B, H3, H4 and H1. A histone

protein is formed of two H2A-H2B dimmers, a H3-H4 tetramer and a linker protein H1. Packing of nucleosomes forms chromatin and further condensation of chromatin forms chromosomes.

The amino-terminal tails of histone subunits are subject to post-translational modifications, such as methylation, phosphorylation, and acetylation. For example, H3K4me3 represents tri-methylation of H3 at lysine 4.

Histone modifications play pivotal roles in activating and inhibiting transcription. It affects the open/closed state of chromatin and provides binding domains for transcription factors. Both position and type of histone modification determine whether it activates or represses gene expression(*5*). Disruption of histone modifications has been associated with various diseases, including cancer. Hence it has become new solutions for drug development and has obtained promising clinical results. Besides, taking advantage of CRISPR/dCas9 fusion system, histone marks can be locus-specifically and reversibly modified, which allows understanding of the exact roles of histone modifications for therapeutic applications(*15, 16*).

3D chromatin structure

As mentioned before, DNA inside the nucleus is stored in a highly packed manner as chromatin. The three-dimensional architecture of chromatin controls accessibility of DNA to transcription factors and regulates enhancer-promoter interactions. Euchromatin is the chromatin that is accessible/open due to certain histone modifications and low nucleosome density. In contrast, heterochromatin is the chromatin that is inaccessible/closed owing to histone modifications and high nucleosome density.

One of the well-studied chromatin domain is large organized chromatin lysine modifications (LOCKs), which refers to the domain enriched for heterochromatin histone modifications, such as H3K9me2(*17*). Another heterochromatin domain is lamina-associated

domains (LADs). It has been observed that LADs were involved in organizing chromosomes inside the nucleus and inhibiting gene expressions(*18*).

Non-coding RNA

Non-coding RNA are RNA transcripts that are not involved in protein encoding. Non-coding RNA includes miRNA, siRNA, piRNA and long non-coding RNA (lncRNA, > 200 bp). MicroRNAs (miRNAs) are responsible for the regulation of over 60% of protein-coding genes. Multiple miRNAs have been shown to have oncogenic or tumor suppressive effects and are being used as potential therapeutic targets. Thus, miRNAs also make enticing targets for better understanding disease pathology and have become a recent focus in microfluidics. Also it has been found that the expression of lncRNA was different in each cell type and changed during development or diseases(*19*).

Functions of lncRNA include regulating chromatin structure by both *cis*- and *trans*-acting mechanisms and regulating nearby gene expression. For example, HOTTIP (HOXA distal transcript antisense RNA) is a *cis*-acting lncRNA that deposits H3K4me3 histone mark and activates HOXA genes. HOTAIR is a lncRNA that works *in trans* to repress HOXD expression and work as a scaffold for PRC2 that modifies H3K27me3(*20*).

Taking advantages of high-throughput sequencing, a large amount of lncRNA have been recognized while the function of most of them remains unclear. In the future, we expect novel computational and experimental methods will accelerate complete lncRNA annotation and functional characterization.

Epigenetics in cancer and epigenetic therapy

For many years, we have believed that the initiation and progression of cancer were only due to genetic mutations, yet more recently, a prevalence of aberrant epigenetic

modifications have also been found(*21-25*). Now, the epigenome has a more widespread role and many researchers argue that epigenetic disruption, often connected with genetic mutations, is the central driving force of cancer(*26*). In fact, many of the genes mutated in cancers are genes that directly modify the epigenome, such as DNA methyltransferase genes that control DNA methylation(*27*).

Almost every cancer has abnormal specific or widespread DNA methylation levels(*28*). For instance, CpG (cytosine and guanine separated by one phosphate group) island (CGI) hypermethylation and genome-wide hypomethylation frequently occur in cancer cells. CGIs are thousand base-pair long regions of the genome that are rich in CpGs, have little to no methylation, and often are associated with gene promoters(*29*). Hypermethylation of a CGI in the promoter of a tumor suppressor gene often silences the function of that gene. For example, the BRCA1 gene is inactivated in breast and ovarian carcinomas by CGI hypermethylation(*30*). BRCA1 is part of the protein complex that repairs DNA when it is damaged. Its dysfunction leads to damage building up which can cause cancerous mutations. Meanwhile, global hypomethylation, which mainly affects repetitive regions of the genome and gene promoters that have few CpG dinucleotides, is common in all types of cancer and results in genomic instability(*28*). Furthermore, hypomethylation of gene promoters in cancer frequently activates oncogenes, metastasis-related genes, and drug-resistance genes.

Similarly, histone modification aberrations in cancer are also genome-wide and can be tumor-type specific(*31*). For example, global loss of the histone modifications H4K16ac and H4K20me3 is a common hallmark of human cancers(*32*). A recent study shows that mutations in the H3F3A gene, which encodes the histone variant H3.3, are associated with pediatric brain cancers. H3.3 mutation disrupts post-translational modification leading to a loss of the histone mark at lysine 27(*33*). Lastly, non-coding RNAs that regulate the epigenome are also expressed aberrantly in cancer. It has been observed in different cancers that a long noncoding RNA,

HOTAIR, is overexpressed and can be used as a prognostic biomarker. HOTAIR interacts with both PRC2 and LSD1 complexes leading to silencing of tumor suppressor genes(*34*).

Interestingly, these epigenomic aberrations are also hallmarks of aging, though it is unclear whether modifications caused by aging contribute to cancer(*35*). Overall, understanding the epigenome has prognostic and diagnostic significance for both cancer and aging.

The reversible nature of the epigenome has opened new avenues for precision medicine, which leverages the genetic and epigenetic information of patients to determine personalized courses of treatment(*36-38*). For example, the use of RG108, a DNA methyltransferase inhibitor, showed demethylation and re-expression of the tumor suppressor gene, p16, in a human cancer cell line(*39*). As such, epigenetic patterns are useful in assessing various chemotherapeutic targets and responses(*38*).

In fact, several epigenome-modifying drugs have been FDA approved as treatments for cancers(*40-43*). One of these drugs is used in the treatment of peripheral T-cell lymphomas which are an aggressive subgroup of non-Hodgkin lymphomas with a high rate of relapse(*43*). As hypoacetylation frequently occurred in cancers, Romidepsin, one type of histone deacetylase (HDAC) inhibitors that prevented acetyl groups on histones from being removed, was clinically tested and found to be suitable for treatments in relapsed patients.

State of the art of ChIP-seq

The first step of studying epigenomics is profiling its genome-wide landscapes. Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) is the gold standard technique to profile the presence of histone modification and transcription factors (TF) on a genome-wide scale. ChIP-seq includes several steps: (1) Cell lysis and chromatin fragmentation; (2) Chromatin fragments with target histone mark or TF are captured by specific

antibody; (3) The antibody is pre-coated on magnetic beads, thus chromatin of interest can be easily separated by separating beads from liquid solution using magnet; (4) Washing using salt and surfactant is conducted to remove non-specific adsorption on antibody and beads surface; (5) DNA molecules are released by proteinase K digestion and is purified by ethanol precipitation or SPRIbeads cleanup; (6) Purified ChIP DNA is converted into sequencing-ready libraries by adapter ligation and PCR amplification; (7) ChIP libraries are sequenced using a Illumina sequencer; (8) Sequencing reads are aligned back to reference genome and regions enriched with sequencing reads are identified as target histone mark/TF binding regions.

As described above, the entire ChIP-seq procedure involves multiple steps which are not only time-consuming (~2 days), but also cause a considerable amount of material loss. Moreover, there are only two copies of DNA in each diploid cell. As a result, ChIP-seq assay usually requires more than 1 million cells as starting materials. This has hindered the wide application of ChIP-seq when sample size is small, such as biopsy sample.

Over the past decades, numerous efforts have been made to improve the sensitivity and resolution of ChIP-seq. Two of the most commonly used low-input strategies are microfluidic technology and *in situ* (inside nuclei) immunocleavage by either pA-Mnase or pA-Tn5.

Our lab had developed a series of microfluidic devices for ultra-low-input, multiplexed and automated ChIP-seq, including MOWChIP-seq(*44*), SurfaceChIP-seq(*45*), LIFE-ChIP-seq(*46*) and mu-CM(*47*). The two core elements of MOWChIP-seq (Microfluidic Oscillatory Washing-based ChIP-seq) are packed beads bed in microfluidic chambers and oscillatory washing. The reaction chamber contains a partially closed on-chip valve which allows only solution to flow through, thus beads are packed tightly against the valve. Sheared chromatin was flowed through the tightly packed bed of antibody-coated beads. High surface-to-volume ratio and short diffusion length (< 2 μm) facilitated efficient capture of target chromatin. After ChIP, oscillatory washing is performed by applying alternative pressure pulse in chamber inlet

and outlet. The back and forth moving of beads in washing buffer efficiently removes adsorption of non-specific chromatin. Even with the low cell input, MOWChIP was capable of collecting ChIP-DNA within the theoretical limits. MOWChIP-seq has successfully profiled H3K4me3 and H3K27ac from as few as 100 GM12878 cells.

Based on the similar concept of MOWChIP-seq, LIFE-ChIP-seq (Microfluidic Low-Input Fluidized-Bed Enabled ChIP-seq) was developed for automated and parallel analysis of histone marks. The device contains 4 reactions chambers connected to common inlets through a series of splits, which enables running 4 ChIP assays in parallel on different histone marks with as few as 50 cells per assay. SurfaceChIP-seq replaces immunomagnetic beads with antibody-coated microfluidic channels. In this way, the design and operation of the microfluidic device are greatly simplified. The device contains 8 straight channels (40 mm × 1 mm × 60 μm), and antibody is hybridized in the channel surface. The channel is sequentially filled with chromatin solution, washing buffer and elusion buffer for ChIP, washing and releasing DNA respectively. SurfaceChIP-seq is able to profile histone mark with as low as 30 cells per assay.

in situ immunocleavage is another low-input strategy, and the most representative method is CUT&RUN, and it outperforms conventional protocol in resolution and single-to-noise ratio as it only solubilizes target chromatin fragments. CUT&RUN (Cleavage Under Targets and Release Using Nuclease), a *in situ* immunocleavage method, has significantly increased the sensitivity of tube-based profiling of histone marks and transcription factors. In CUT&RUN, cells are permeabilized to let in antibody and the protein A-Micrococcal Nuclease (pA/Mnase) fusion protein. Compared to traditional methods that fragmentizes the entire genome, in CUT&RUN only chromatin fragments bound by antibody are cut by Mnase and released into solution. As a result, only targeted chromatin enters into solution, and the rest remains insoluble. Hence it dramatically reduces non-specific backgrounds and requires ~10

fold less sequencing depth in contrast to traditional method. It is able to profile ~100 cells for abundant histone marks and ~1000 cells for transcription factors. Similar to CUT&RUN, CUT&Tag (Cleavage Under Targets and Tagmentation) substitutes pA/Mnase with protein A-Tn5 transposase fusion protein. The difference between Mnase and Tn5 transposase is that Tn5 not only cut DNA, but also insert adapters to both ends of the DNA molecules. This eliminates tedious library preparations, saves times, and minimizes sample loss.

Over the recent years, ChIP-seq assays have expanded our understanding of histone modifications and transcription factor binding during development and disease progression. However, like other epigenomic assays, ChIP-seq is a bulk assay, which generates population-averaged signals. Cell-to-cell variations appear in both normal and disease cells, especially in cancer cells. Hence, a reliable single-cell method that examines cell heterogeneity is urgently needed. Single cell data provides a sensitive way to classify cell types and track cell lineages. It reveals the intrinsic cell-to-cell variations, thus has wide applications, such as cell classification, or identifying drug resistant tumor cell subpopulations.

Recently there have been several reports on single-cell ChIP-seq that reveal epigenetic heterogeneity with coverage improved from 1,000 reads to 100,000 unique reads per cell (Drop-ChIP(*48*), scChIC-seq(*49*), scCUT&Tag(*50*), uliCUT&RUN(*51*), scChIP-seq(*52*), ACT-seq(*53*), sc-itChIP-seq(*54*), coBATCH(*55*)). Current single-cell ChIP-seq techniques can be roughly divided into two main categories: droplet microfluidic based methods and *in situ* immunocleavage methods.

Droplet microfluidic methods have been widely used in single cell assays, including scRNA-seq, scATAC-seq and scWGS (whole genome sequencing). The basic idea is to encapsulate individual cells into droplets and perform reactions, including barcoding, in droplets. Droplets serve as miniaturized reactors. The biggest advantage of droplet microfluidics is the high throughput and automation, as hundreds of thousands of droplets can

be easily generated and manipulated together. The first reported single-cell ChIP-seq was developed based on droplet microfluidics, named Drop-ChIP(*48*). Drop-ChIP encapsulates single-cells into droplets, where they are lysed and their chromatin is digested. The chromatin, still contained within the droplet, is mixed with barcodes that have unique DNA sequence for each cell. This allows the identity of each cell to be determined after being pooled for downstream ChIP and sequencing. Drop-ChIP is able to successfully delineate three subpopulations within mouse embryonic stem cells based on various chromatin signatures. However, it only yielded ~1,000 unique reads per cell, which was too sparse compared to other types of single-cell data. Recently, another similar droplet microfluidic platform for scChIP-seq was reported with up to 10,000 unique reads per cell(*56*). Besides, a commercial droplet microfluidic platform was developed by 10x Genomics and has been used by many labs in different field of biology and immunology due to its streamlined workflow(*57, 58*).

Unlike droplet microfluidics, several on-tube scChIP-seq methods based on *in situ* immunocleavage have been reported recently including scChIC-seq(*49*), with larger than 10,000 unique reads per cell, which greatly outperforms Drop-ChIP. Among them, scChIC-seq has been reported to produce 100,000 unique reads on average, which is the highest number obtained to date.

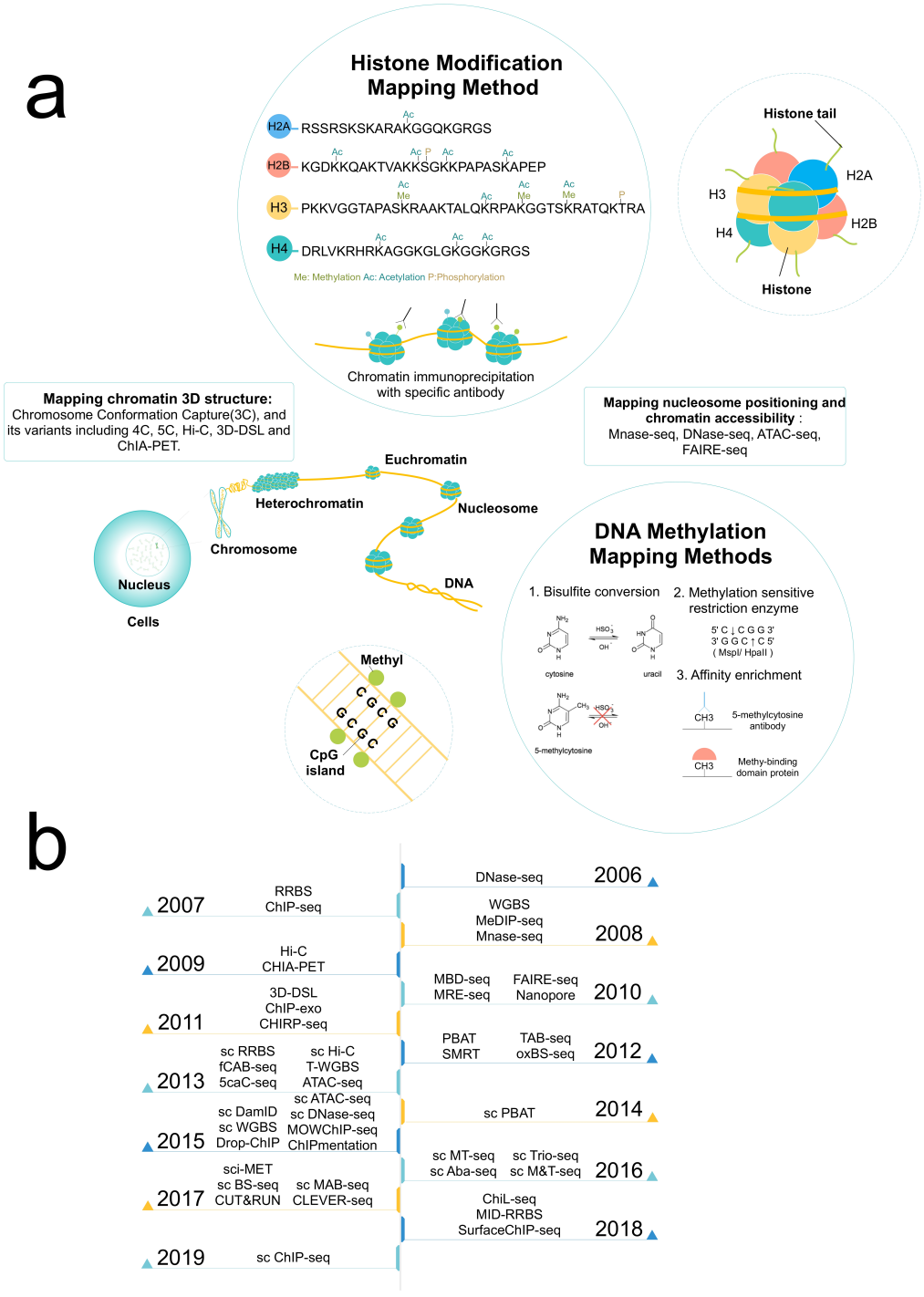


Figure 2.1 Summary of epigenetic mechanisms and various genome-wide mapping methods.

a. Overview of various epigenetic modifications and profiling methods. **b.** The timeline on development of NGS-based epigenomic mapping toolkits over the years. "sc" refers to "single cell". References: ChIP-seq(*76-78*); RRBS(*79*); MeDIP-seq(*80, 81*); DNase-seq(*82*); Mnase-seq(*83*); WGBS(BS-seq/ MethylC-seq)(*84, 85*); Hi-C(*86*); CHIA-PET(*87*); MBD-seq(*88, 89*); MRE-seq(*90*); FAIRE-seq(*91*); Nanopore(*92-94*); ChIP-exo(*95*); 3D-DSL(*96*); CHIRP-seq(*97,*

98); PBAT(*99*); oxBS-seq(*100*); TAB-seq(*101*); SMRT(*102*); sc RRBS(*103-105*); fCAB-seq(*106*); 5caC-seq(*107*); sc Hi-C(*108, 109*); ATAC-seq(*110*); T-WGBS(*111*); sc PBAT(*112*); sc ChIP-seq (Drop-ChIP)(*113*); scDamID(*114*); sc Dnase-seq(*115*); sc ATAC-seq(*116, 117*); sc WGBS(*118*); MOWChIP-seq(*44*); ChIPmentation(*119*); sc Aba-seq(*120*); sc MT-seq(*121*); sc M&T-seq(*122*); sc Trio-seq(*123*); CLEVER-seq(*124*); sc MAB-seq(*125*); sci-MET(*126*); CUT&RUN(*127, 128*); sc BS-seq(*129*); MID-RRBS(*130*); SurfaceChIP-seq(*130*); ChiL-seq(*131*); sc ChIP-seq(*56*).

2.5 Sequencing-related microfluidic technology for epigenomic analysis

Microfluidic technology has been widely used to perform biological and chemical analysis due to its unique advantages, including small volume and high reaction efficiency. It has greatly improved the performance of many epigenomic analyses in terms of resolution, throughput and cost. Microfluidic technology for sequencing-related analysis is reviewed in fours subsections below.

Microfluidics for mapping histone modification and DNA-binding protein

Several microfluidic devices have been reported to reduce the necessary input materials of ChIP-seq. The basic strategy is to increase the efficiency of immunoprecipitation of target chromatin by decreasing diffusion length and increasing surface-to-volume-ratio. Shen et al. designed a dead-end rectangular ring-chambers inside which reagents circulate to perform ChIP from 1,000 cells(*132*). Our lab has developed a series of devices for low-input ChIP-seq. Cao et al. designed MOWChIP (Microfluidic Oscillatory Washing based ChIP) that is capable of performing ChIP-seq from only 100 cells(*44*). The device contains a half-closed valve that stopps beads while allows solution to flow through. The densely packed beads facilitate fast and complete adsorption of target chromatin fragments. After ChIP, intermittent pressure pulse is applied at the inlet and outlet of reaction chambers. As a result, beads are moved back and forth vigorously and non-specific adsorption on beads are removed. LIFE-ChIP (Low-Input Fluidized-bed Enabled ChIP) inherited the basic idea of MOWChIP-seq and added more inlets to allow 4 reactions running in parallel(*46*). To further decrease the lower limit of ChIP-seq,

Ma et al. developed SurfaceChIP-seq that required as few as 30 cells for genome-wide histone modification profiling(*133*). Traditionally, the antibodies are coated on magnetic beads, while with SurfaceChIP the antibodies are coated on microfluidic channel surface, which greatly simplifies the operation. Another similar device named FloChIP was reported that profiles histone marks from 500 cells(*134*). Antibody is immobilized on channel surface through biotin-streptavidin interactions. It adapts on-chip tagmentation and designs an interconnected device allowing sequential ChIP. Mu-ChIP was recently developed that combined ChIPmentation based indexing with in-parallel microfluidic assays and was able to profile histone modification with only 20 cells. Details are reported on Chapter 4. PnP (Plug and Play)-ChIP-seq is a fully automated and integrated platform that can be operated by novices of microfluidics and ChIP. After loading samples, the further steps are fully automated on a specially designed Fluidigm C1 controller. It enables 24 parallel ChIPs from as few as 100 cells.

Microfluidics for DNA methylation profiling

The principal method of determining DNA methylation is by conducting bisulfite conversion followed by sequencing, known as Whole-Genome Bisulfite Sequencing (WGBS) and sometimes referred to as BS-seq or methylC-seq. Sodium bisulfite treatment converts un-methylated cytosines to uracil while methylated cytosine remains intact. After sequencing, the methylated cytosines are determined in single nucleotide resolution by comparing the sequence of bisulfite treated and untreated samples. There are many microfluidic devices designed for detection and quantification of methylated DNA in a locus-specific manner. Here we focus on microfluidics for genome-wide DNA methylation profiling. Our lab has developed a diffusion-based device for low-input RRBS called MID-RRBS(*130*). RRBS (Reduced Representation Bisulfite Sequencing) uses restriction enzymes, such as MspI, to cut the highly methylated fragments which mainly locate at CpG islands and promoters. The fragmented DNA is then

bisulfite-treated and sequenced. This reduces cost and time compared to whole genome bisulfite sequencing. The highlight of the MID-RRBS design is the diffusion-based reagent loading and buffer exchange. The device contains a chamber and a pair of on-chip valve. Closed valves separated the chamber into reaction channels and loading channels. Reagents in the loading channel diffused into reaction channel through the gaps of valves driven by concentration differences. In contrast, the starting DNA has a low diffusivity due to its substantial sizes, thus it mainly remains inside the reaction channel. MID-RRBS reduces the amount of input DNA from several hundred nanograms to that of a single cell. The results in high bisulfite conversion rate and high coverage of CpGs (2.0-2.6 million CpGs with over 1x coverage and 1.38 million CpGs with over 10x coverage from 0.3 ng input DNA). It profiles methylome of antipsychotic clozapine treated mice frontal cortex and identifies distinct patterns between treated and control mice including key regulators in mental disorders.

Affinity enrichment assay and methylation sensitive restriction enzyme digestion are two alternative methods for bisulfite conversion. They avoid drawbacks of bisulfite treatment, such as severe DNA damage and long treatment time. Affinity enrichment assay utilizes antibody and proteins to selectively enrich methylated DNA. Its basic concept is similar to ChIP-seq. Our lab has developed a microfluidic device that performed MeDIP-seq (Methylated DNA immunoprecipitation followed by next-generation sequencing) from as low as 0.5 ng of DNA(*135*). Device design and experimental procedure are mostly adapted from MOWChIP-seq. Immunoprecipitation of DNA fragments containing 5mC is performed at the surface of tightly packed bead bed in a microfluidic chamber. Small diffusion length inside the chamber greatly increases efficiency of immunoprecipitation, thus it enables profiling methylomes using only 0.5 ng DNA compared to 50 ng DNA required by conventional methods. The basic idea of using restriction enzymes to profile methylation status is that methylation-sensitive restriction enzymes cut fragments where methylation occurs, while its methylation-insensitive

isoschizomer cuts fragments regardless of the methylation status. Location of methylation can be obtained by comparing the results of the two restriction enzymes.

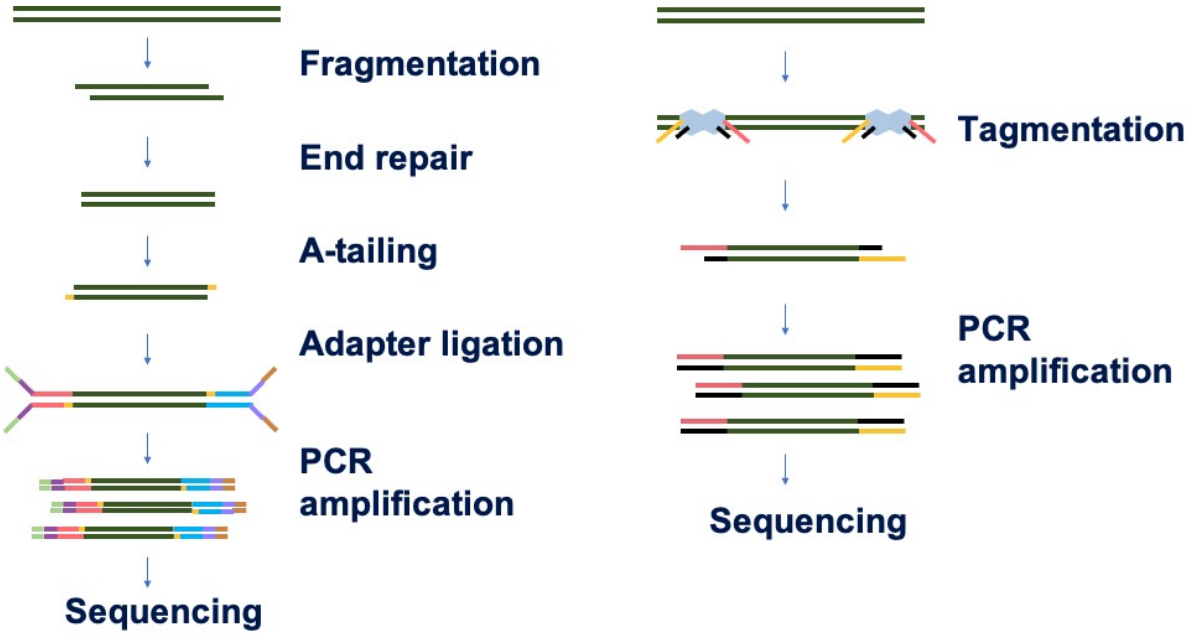


Figure 2.2 NGS library preparation procedure.
Conventional library preparation steps (left) and Tagmentation (right).

Microfluidic for mapping chromatin hierarchy

Chromatin accessibility assays map the open and/or closed regions of the genome. Typical methods are DNase-seq, MNase-seq and ATAC-seq. Among them, DNase-seq and ATAC-seq directly map accessible regions while Mnase-seq maps regions occupied by nucleosome(*136*). MNase digests DNA fragments not protected by nucleosomes or transcription factors. With proper titration of digestion conditions, mono-nucleosomes are obtained and sequenced. Hence, Mnase-seq can be used to locate inaccessible regions. In contrast, DNase cuts nucleosome-free regions known as DNase I hypersensitive sites (DHSs) into small pieces, which are then extracted and sequenced. Thus, it directly maps active chromatin regions. ATAC-seq is the newest method and has become the most popular method

to study chromatin structure due to its simplicity and versatility. It utilizes hyperactive Tn5 transposases to simultaneously cut DNA and insert sequencing adapters into fragment ends. This one-step reaction greatly simplifies and speeds up the procedure. Both nucleosome-free DNA and DNA bound by histones are cut and sequenced. In this way, ATAC-seq provides comprehensive information, including both open chromatin and nucleosome positioning. However, Tn5 transposase has the preference to cut open chromatin region. As a result, nucleosomes that are away from open chromatin may not be mapped as efficiently as those closed to active sites. Hi-C is a variant of Chromosome Conformation Capture (3C) technology and is the preferred method to map genome-wide higher-order chromatin structure. Folding and looping of chromatin bring DNA fragments near to each other, even if they are thousands of base pair away. Hi-C profiles the spatial organization of chromatin by measuring the genomic coordinates of DNA fragments that physically close together. In Hi-C, *in vivo* chromatin structure is kept by formaldehyde crosslinking, then chromatin is fragmented by enzyme digestion, the interacted fragments are ligated forming a single DNA fragment and sequenced.

Xu *et al* reported a droplet microfluidic device that performed Mnase-seq from 2,500 cells(*137*). Cells were encapsulated in droplets with reaction buffer for cell lysis and enzyme digestion. The author demonstrated that this platform not only automats the Mnase-seq preparation procedure, but also could be potentially incorporated with other epigenomic assays, such as ChIP-seq.

Besides bulk experiments, several groups have reported microfluidic devices for single cell chromatin accessibility assays. The first microfluidic-based single cell chromatin accessibility assay was reported at 2015(*138*). It integrated ATAC-seq into a commercial microfluidic platform, Fluidigm C1 Auto Prep, which was specially designed for single cell assays. Bulk cells loaded into the platform were individually captured in the integrated fluidic

circuit (IFC). Once an IFC was occupied by a cell, the following cells would travel through the by-pass channels to the next available IFC. Captured cells were lysed under the flowing lysis buffer in the chamber. Tn5 and tagmentation buffer were then flowed into each IFC to conduct tagmentation. After that, tagmentation was stopped by adding EDTA and DNA fragments were amplified directly on chip. In this way, lysis, tagmentation and PCR were carried out sequentially in micro-chambers. DNA loss was minimized as reactions were performed in a "one-pot" manner. The amplified samples were collected and prepared into indexed libraries for sequencing. This scATAC-seq revealed cell-to-cell variations in chromatin accessibility and its associated transcription factors and *cis*-regulatory elements.

The throughput of C1 platform was limited to several hundred cells per run. To improve the throughput, two recent droplet microfluidic platforms for scATAC-seq have been reported. The first used a commercial microfluidic system, Chromium platform from 10x Genomics(*139*). In this approach, cells were first transposed in bulk before individually loading into droplets. Each droplet also contained a gel bead with unique barcode. About 100,000 unique gel beads were used to enable barcoding of tens of thousands of cells. The other droplet-based platform was named 'droplet single-cell assay for transposase-accessible chromatin using sequencing' (dscATAC-seq)(*140*). Taking advantage of droplet manipulation technology, thousands of single cells can be rapidly and uniformly encapsulated into individual droplets. The throughput was further increased by combining it with combinatorial indexing (dsciATAC-seq). Chromatin accessibility of over 100, 000 cells were measured to reveal cell-type-specific *cis*- and *trans*- regulatory network. In dsciATAC-seq, each cell was lysed and cut by uniquely barcoded Tn5 transposase before loading into droplet microfluidic. This step added the first layer of barcode. Then transposed cells and PCR barcode beads were co-encapsulated into droplets. A droplet contains multiple cells and beads in order to increase throughput. In this way a second layer of barcode was added. Each cell was identified by the combination of

the first and second barcodes. The use of droplet microfluidics for combinatorial indexing increased the number of theoretical barcode combinations, allowing higher throughput of cells.

Except scATAC-seq, single cell multi-omics analysis revealing both chromatin accessibility and transcriptome of the same cell has attracted increasing interests. Droplet-based single-nucleus chromatin accessibility and mRNA expression sequencing (SNARE-seq) is a method to obtain a cell's transcriptome and its chromatin accessibility(*141*). In this method, nuclei are permeabilized to allow tn5 transposase cut chromatin *in vivo*. The chromatin fragments remains intact once tn5 transposase is not striped away from the fragments. Then nuclei and barcode beads are co-encapsulated into droplets. Oligonucleotide on bead surface contains unique barcode and complementary sequence to capture mRNA and transposed chromatin fragments. In this way, mRNA and open chromatin from the same nuclei are labeled with the same barcode. This joint profiling of transcriptome and accessible chromatin region reveals the connection between transcriptional regulation and its outcomes.

Microfluidic for mapping non-coding RNA

Much like DNA methylation and histone modifications, non-coding RNAs are involved in gene regulations, and defects in them have also been implicated in cancer. Methods to investigate long-noncoding RNA includes ChIRP (Chromatin isolation by RNA purification), CHART (Capture Hybridization Analysis of RNA Targets), RAP (RNA antisense purification), RIP (RNA Immunoprecipitation), CLIP (cross-linking and immunoprecipitation), and RNA pull-down(*142*). Target lncRNA fragments are enriched either by probe hybridization or antibody-antigen affinity binding. On the other hand, microRNA-seq is very similar to RNA-seq. After total RNA extraction, an additional size selection will be performed to enrich only miRNA. Then miRNA is reverse transcribed into cDNA and prepared into libraries for sequencing.

A microfluidic flow cytometry device was developed for miRNA detection using locked nucleic acid flow fluorescent *in situ* hybridization (LNA-Flow FISH)(*143*). The device, which utilized their previously published chamber designs and microfluidic FISH methods, contains 10 individual holding chambers with a 3-way channel for hydrodynamic focus and laser detection(*13, 143, 144*). The device was coated with surface adhesives before fixed cells were loaded and allowed to bind to the surface. The cells were incubated with fluorescently tagged CD69 antibody, permeabilized, and incubated with miR155-specific LNA probe. miR155 was chosen for its role in T-cell activation, as the cells were obtained from a stimulated T-cell model cell line. The miRNA, hybridized to the LNA probe, was then amplified through rolling circle amplification (RCA) along with the fluorophore FITC, resulting in circular amplicons for robust signal detection. Finally, the cells were incubated with Hoechst to stain the nucleus blue and flowed through the fluorescence detection channel.

A SERS-based microfluidic device was developed using silver coated silicon-PDMS chambers for detection of miR-222, a miRNA involved in multiple cancers(*145*). The device focused on two methods. The one-step method used fluorescently labeled miRNA that then bound to thiol-capped ssDNA probes immobilized on silver nanoparticles. After hybridization, solid surface enhanced Raman scattering spectra was obtained to quantify the miR-222 level. In contrast, the two-step method only used half of the probe on the nanoparticles. The miRNA bound to the first half of the probe before the second half of the probe, tagged with a Raman label, was then hybridized and quantified. The device could selectively detect miR-222, even in the presence of another miRNA, and had a limit of detection of 0.55nM (one-step) and 1.51nM (two-step). Altogether, this makes it possible to detect miR-222 as a means of early diagnosis for cancer.

In addition to the previously mentioned academic platforms, multiple commercial platforms have been used in miRNA analysis. µParaflo Technology from LC Science is an

array-based microfluidic platform for miRNA expression profiling that performs highly parallel RT-PCR to examine a large number of miRNAs(*146-152*). This platform enabled massive custom DNA and RNA probes synthesized on the chamber surface in a photo-programmable manner(*153*). Other microfluidic platforms for miRNA qPCR analysis included the BioMark from Fluidigm(*154-156*), and the TaqMan array microfluidic card(*157-161*). Furthermore, it was also reported that single-nuclei RNA-seq performed in the C1 platform from Fluidigm has been used in identifying differential miRNA expression(*162*).

2.6 Conclusions

Epigenomic mapping of tissue samples generates critical insights into genome-wide regulations of gene activities and expressions during normal development and disease processes. Epigenomic profiling using a low number of cells produced by patient and mouse samples presents new challenges to biotechnologists. Over the past decades, the development and commercialization of microfluidic technologies have led to new revolutions in genomic and epigenomic research related to single-cell analysis and high-sensitivity molecular detections. With NGS cost decreasing and potential of precision medicine looming large, the processing of a large number of patient samples is becoming more important than ever and microfluidics has the chance to play a significant role in this process. Furthermore, since microfluidics dramatically reduces costs of reactions through reduction of reagents, these epigenomic tests can be made available to a wide population of patients. Although microfluidics is an ideal solution for low-input and high-throughput epigenomic analysis, the relatively high implementation barrier of microfluidic devices/assays still prevents its wide application. Microfluidic researchers will need to take it upon themselves to implement the microfluidic tools, often under collaboration with scientists and clinicians, to make major biomedical discoveries.

Chapter 3 - Effects of Culture Condition on Epigenomic Profiles of Brain Tumor Cells

3.1 Introduction

Personalized cancer medicine offers the promise of more effective treatments that are tailored to an individual's own dynamic cancer phenotype. Meanwhile, tissue-engineering approaches to modeling tumors may complement these advances by providing powerful new strategies for understanding the adaptation dynamics occurring during treatment. However, in both of these areas, new tools will be required to gain a full picture of the genetic and epigenetic regulators of phenotype dynamics occurring in the small populations of cells that drive resistance. In this study, we perform epigenomic analysis of brain tumor cells that are collected from microengineered three-dimensional tumor models, overcoming the challenges associated with the small numbers of cells contained within these microtissue niches, in this case collecting ~1000 cells per sample. Specifically, we use a high-resolution epigenomic analysis method known as microfluidic-oscillatory-washing-based chromatin immunoprecipitation with sequencing (MOWChIP-seq) to analyze histone modification patterns (H3K4me3). We identified gene loci that are associated with the H3K4me3 modification, which is generally a mark of active transcription. We compared H3K4me3 profiles in standard 2D cultures and 3D cultures based on type I collagen hydrogels, under both normoxic and hypoxic conditions. We found that culture dimensionality drastically impacted the H3k4me3 profile and resulted in differential modifications in response to hypoxic stress. Differentially H3K4me3-marked regions under the culture conditions used in this study have important implications for gene expression differences that have been previously observed. In total, our work illustrates a direct connection between cell culture or tissue niche condition and genome-wide alterations in histone modifications, providing the first steps toward analyzing the spatiotemporal variations in epigenetic regulation of cancer cell phenotypes. This study, to the best of our knowledge,

also represents the first time broad-spectrum epigenomic analysis has been applied to small cell samples collected from engineered microtissues.

Despite tens of billions of dollars of investment in the United States alone into the development of new therapies, cancer is still a leading cause of death worldwide.(*163-165*) A substantial contributing factor to this statistic is an incomplete understanding of the mechanisms of cancer initiation and development. However, for the mechanisms that are well understood, a lack of efficient model platforms that can accurately predict human therapeutic responses is hindering treatment development. One area where improvement is needed is a greater appreciation and understanding of the highly variable and dynamic abnormalities that may be present in any one type of cancer or even within a single tumor.(*166, 167*) Heterogeneity can occur in bulk tumor tissue, circulating tumor cells, disseminated/metastasized tumor cells, and cancer stem cells. Phenotypic variations usually happen at the single cell level and include alterations in protein expression, treatment response, and growth rate.(*168, 169*) This cellular heterogeneity can impede the selection of effective treatments, as treatments that only target a specific cell phenotype or that induce initial significant reductions in primary tumor mass can leave rare aggressive tumor cells behind, causing tumor recurrence and drug resistance. This is the underlying rationale for growing efforts in personalized medicine. The current standard of therapy is to treat the average profile of selected biopsied tumor cells, however, cancer is an extremely complex disease and there is an emerging realization that not taking a complete account of the molecular, phenotypic, and genetic subgroups present within a patient's cancer will lead to the selection of therapies that are unlikely to yield more than incremental benefits.(*166, 167, 170, 171*) Additionally, most targeted therapeutic strategies focus on genetic markers and mutations, but emerging research has shown that epigenetic abnormalities also play a role in regulating cancer progression.(*172*) The development and implementation of high-resolution analysis of the genetic and epigenetic

profile of a patient's specific cancer is needed in order to improve the efficacy of current therapies.

Epigenetic alterations, which regulate the activation or silencing of cancer-associated genes, have been shown to be an important mechanism driving the development of phenotypic variability within a tumor. Histone modifications in particular have the potential to play a major role in regulating cell phenotype. Histone modifications include acetylation, methylation, and phosphorylation of histone proteins and can impact transcription, DNA replication, DNA repair, and genomic stability(*173-175*) Next-generation sequencing (NGS) is a valuable tool that can be employed to analyze the profile of these histone modifications among cancer subgroups. Although NGS can be used for the targeted sequencing of well characterized cancer associated genes, whole-exome or whole-genome sequencing can be utilized to discover disease-causing gene variants in genetically heterogeneous or rare diseases.(*176*) Epigenome profiling in particular could provide a better understanding of the pathogenesis of tumors and enable the observation of variable responses to anti-cancer drugs.(*177*)

The role of epigenetics in drug resistance and tumor progression has been elucidated in a number of studies. Cellular heterogeneity in histone modification levels has been linked to the prognosis of prostate cancer.(*178, 179*) Variations in specific histone modifications have also been linked to the development of temozolomide (TMZ) resistance in glioblastoma multiforme (GBM) tumors.(*180*) Epigenetics have also been used to identify a distinct subpopulation of breast cancer cells that promote collective invasion, suggesting that invasion, metastasis, and patient outcome may be influenced by symbiotic relationships among epigenetically distinct tumor cell subpopulations.(*181*) Components of the tumor microenvironment, such as the extracellular matrix (ECM), can also influence histone modification dynamics. Work initiated by Li et al. demonstrated that the phosphorylation state and cellular localization of histone deacetylase 4 (HDAC4) is influenced by matrix mechanics

and is capable of regulating gene expression associated with fibroblast-myofibroblast transitions, establishing a link between ECM mechanics and epigenetic regulation.(*182*)

Better *in vitro* cancer models are also needed, in conjunction with NGS, to better understand how epigenetic alterations could influence tumor heterogeneity and the unique response of each patient to various treatments. Traditionally, cancer therapies are first screened in 2D model platforms before being translated to mouse models and then ultimately to patients. However, 2D screening platforms commonly select ineffective therapeutics to move on to the next phase of clinical trials, as evidenced by the mere 9.6% success rate for a therapy to pass through all three phases.(*183*) Another issue with the 2D screening platforms is that they could also be screening out potentially effective therapies. Therefore, model platforms that more accurately recapitulate the *in vivo* environment are needed. Altered cell phenotypes, genetic mechanisms, and responses to therapeutics for monolayer versus 3D systems have been widely observed.(*184-186*) The use of 3D platforms may provide a more accurate model of *in vivo* physiology and cell response to therapeutics than traditional 2D culture methods due to the presence of an ECM mimic and physiologically relevant cell-cell and cell-ECM interactions in the 3D culture models. The combination of 3D model platforms that better mimic the tumor microenvironment than their 2D counterparts, with NGS analysis, could greatly enhance the ability to analyze tumor responses to therapies in a controllable and customizable manner, benefiting the advancement of personalized cancer care. Epigenomic analysis of 3D *in vitro* cultures in comparison to 2D cultures could also help to uncover the mechanisms by which the microenvironment epigenetically regulates cell phenotype and therapeutic response. However, the impact of a 3D culture environment on the epigenetic response to microenvironmental stresses has yet to be explored due to significant challenges in obtaining the cell sample sizes necessary for epigenetic analysis.

The aim of this study was to characterize the global H3K4me3 profiles of U-251 MG Glioblastoma cells (U251) cultured under various microenvironmental stresses, which included 2D, 3D, normoxic, and hypoxic culture conditions. U251 cells were selected as a model of a highly aggressive cancer that has particularly few treatment options. However, research aimed at GBM heterogeneity has largely been focused on the role of clonal evolution and cancer stem cells and little is known about how microenvironmental stresses could be regulating the development of tumor subpopulations.(*187*) Culture dimensionality has a significant impact on cell phenotype. The introduction of an ECM mimic provides a vastly different environment than traditional 2D cultures for cell-cell and cell-ECM interactions. Along with culture dimensionality, hypoxia has also been shown to drive phenotypic changes, particularly in the context of cancer, where tumor cells under hypoxic conditions have been shown to have increased angiogenic potential and therapeutic resistance.(*188*) The histone modification profile was obtained using our recently invented highly sensitive microfluidic-oscillatory-washing based chromatin immunoprecipitation (MOWChIP) platform that reduces nonspecific adsorption to obtain high yields of highly enriched DNA, enabling the collection and analysis of approximately 180 picograms of ChIP DNA from 1,000 cells, where traditional ChIP protocols were only able to obtain tens of picograms of DNA from 10,000 cells.(*189*) MOWChIP was followed by NGS using Illumina sequencing. We will analyze the implications of the disparities we observed between culture conditions in the global H3K4me3 profile of U251 cells. We will then discuss the significance of this study in advancement of methods to improve personalized cancer care efficiency. The data and discussion presented in this chapter had been published in ACS Biomat Sci Eng 2019(*190*).

3.2 Material and methods

Cell culture

U251-MG human glioblastoma cells (Sigma-Aldrich, St. Louis, MO) were cultured in Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine serum (FBS) (Life Technologies, Grand Island, NY), 1% penicillin/streptomycin. (Lonza), and 0.1% non-essential amino acids (VWR, Radnor, PA). The cells were maintained in a 37˚C humidified incubator with 5% CO_2.

Preparation of collagen stock

The collagen stock was prepared as published previously.(*191*) Briefly, type I collagen stock was prepared from Sprague Dawley rat tails (BioreclamationIVT, Baltimore, MD). Tendons were resected from rat tails and added to 200 ml of 0.1% sterile acetic acid per gram of tendon. After at least 48 hours at 4˚C and occasional agitation (twice a day), the collagen solution (still containing bits of tissue) was centrifuged for 45 minutes at 30,000 x g at 4˚C in an Avanti J-25 centrifuge (Beckman Coulter, Brea, CA). The clear supernatant collagen solution was collected and the pellet was discarded. The collagen solution was then lyophilized and stored at -20˚C. The lyophilized collagen was re-suspended in 0.1% acetic acid at a stock concentration of 10 mg/ml. To re-dissolve the lyophilized material, the suspension was vortexed and stored at 4˚C for 48 hours to ensure that the collagen was completely dissolved (i.e. uniform solution with no visible collagen masses).

Formation of cellularized collagen hydrogels

Cellularized collagen scaffolds were formed in (poly)-dimethylsiloxane (PDMS) wells (Sylgard® 184 Silicone Elastomer Kit, Dow Corning, Midland, MI) of 10 mm diameter and 1 mm thickness. The PDMS wells were surface treated with 1% poly (ethylenimine) (PEI) for 10 minutes and 0.1% glutaraldehyde for 20 minutes (Fisher Scientific, Pittsburg, PA) to crosslink the collagen to the PDMS. Excess glutaraldehyde was rinsed off of the wells with DI water. Hydrogels with a final concentration of 5 mg/ml collagen were created from the 10 mg/ml stock solution. U251s were seeded within the collagen hydrogels at a density of 1

million cells/ml. The hydrogels were prepared by transferring the desired volume of collagen stock (50% of the final hydrogel volume) into a centrifuge tube kept on ice. Ice cold 10X DMEM (10% of final solution) and 1N NaOH (2% of final solution) were added to and mixed thoroughly with the collagen stock. In a separate centrifuge tube, U251s were added to 1X DMEM (volume needed to complete the remaining 50% of the final hydrogel volume). The U251 cell suspension was then added to the uncrosslinked collagen solution and slowly mixed on ice until cells were homogenously dispersed in the collagen solution. 10X DMEM and 1X DMEM were purchased from Sigma-Aldrich (St. Louis, MO) and NaOH was purchased from Fisher Scientific (Pittsburg, PA). Using a syringe, the neutralized collagen solution was transferred to the PDMS wells and allowed to polymerize at 37˚C at 5% CO_2 for 30 minutes. After collagen was crosslinked, sufficient media was added to the wells to completely submerge the collagen scaffolds.

Cell sample collection

The day after U251s were seeded in the collagen hydrogels, media was refreshed and well plates or flasks were placed in an incubator at normoxic (20% O_2) or hypoxic (1% O_2) conditions. Samples were then incubated for 72 hours under these controlled conditions. Cells in flasks were trypsinized, rinsed once in ice cold PBS, then re-suspended in ice cold PBS at 1,000 cell/µl. Cells seeded in collagen hydrogels were obtained by digesting the collagen in a solution of 0.5% collagenase (Thermo Fisher, Waltham, MA) and 1% FBS in Hanks Buffered Salt Solution (HBSS) (Lonza). Collagen scaffolds were submerged in collagenase solution and incubated at 37˚C and 5% CO_2 for 2 hours. Digested collagen solution containing released cells was collected, rinsed once in ice cold PBS, and re-suspended in ice cold PBS at 1,000 cell/µl. 1 µl of phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich, St. Louis, MO) and 1 µl of protease inhibitor cocktail (1X concentration) (Sigma-Aldrich, St. Louis, MO) were added to a 100 µl aliquot of cell suspension for each culture condition.

MOWChIP-Seq

The custom microfluidic chip for cellular analysis was fabricated and prepared as previously described.(*189*) The microfluidic chip was mounted on an inverted microscope (IX 71, Olympus). A data acquisition card (NI SCB-68, National Instruments) and a LabVIEW (LabVIEW 2012, National Instruments) program were employed to control the switching of the solenoid valve. 10 µl cell suspension containing 1,000 U251 cells was mixed with 10 µl of 2X lysis buffer (4% Triton X, 100 mM Tris pH 7.5, 100 mM NaCl and 30 mM MgCl2) and incubated at room temperature for 10 min. 1 µl of 0.1 M $CaCl_2$ and 2.5 µl of 10 U/µl MNase (88216,ThermoFisher) were rapidly mixed with the sample and incubated at room temperature for 10 min. 2.22 µl of 0.5M EDTA (pH 8) was added and incubated on ice for 10 min. The sample was centrifuged at 16,100xg at 4°C for 5 min. Supernatant was transferred to a new microcentrifuge tube and stored on ice. Superparamagnetic Dynabeads Protein A (2.8 µm, 30 mg/ml, 1001D, Invitrogen) were used. 150 µg beads were mixed with 0.75 µg anti-H3K4me3 antibody (ab8505, Abcam) in 150 µl IP buffer (20 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% (w/v)sodium deoxycholate, 0.1% SDS, 1% (v/v)Triton X-100) at 4 °C on a rotator mixer at 24 r.p.m. for 2 hours. Chromatin immunoprecipitation and oscillatory washing was conducted as previously described.(*189*) Beads were suspended in 50 µl of freshly made elution buffer (1 mg/ml proteinase K, 10 mM Tris-HCl pH 7.4, 50 mM NaCl, 10 mM EDTA pH 8.0, and 0.03% SDS) and incubated on a thermocycler at 60 °C for 1 hour. DNA was purified using SPRIselect beads (Beckman Coulter) following the manufacturer's instructions. Quantitative PCR (qPCR) was performed for selecting high quality ChIP samples. Samples with relative fold enrichment of positive loci VEGF over 15 were used for sequencing library construction. All ChIP-seq libraries were constructed using Accel-NGS 2S plus DNA library kit (Swift Bioscience) following the manufacturer's instructions with minor modifications. During library preparation, ChIP DNA was end repaired and sequentially ligated

with Illumina adapters. PCR amplification was used to increase the yield of indexed libraries. EvaGreen dye (1x, Biotium) was added to avoid PCR over-amplification. The fragment size of the library was determined by a TapeStation and quantified by KAPA qPCR library quantification kit. A volume of 1 μl of each diluted library were pooled together at a final concentration of 10 nM for sequencing by Illumina HiSeq 4000 with single-end 50 nucleotide read.

ChIP-seq data analysis

Sequencing reads were processed by AQUAS pipeline. Raw reads were mapped to hg19 genome. Duplicates and low-quality mapped reads were filtered out. Filtered reads were used for narrow peak calling with MACS2 ($q < 0.05$). Cross-correlation scores and IDR (irreproducible discovery rate) were calculated. Signal tracks were generated and visualized on IGV (Integrative Genomics Viewer 3.0). Genome-wide differentially marked regions between sample groups were identified by DiffBind.(*192*) The analysis was executed using DESeq2 and a p-value of 0.05 was set as a threshold. RPKM (reads per kilobase million) fold was used as a scoring method. The Venn diagram was plotted by InteractivVenn.(*193*) Functional enrichment GO (Gene Ontology) terms were obtained by PANTHER overrepresentation tests.(*194*) Pathway analysis was conducted using ConsensusPathDB. ChIP peak annotation was obtained by ChIPseeker with default settings. (*195, 196*)

3.3 Discussion

Culture dimensionality and oxygen status impact H3K4me3 modifications

The use of hydrogels for 3D tissue mimics are the most widely used due to their capacity to provide a dynamic microenvironment for modeling cell processes. Type I collagen is a popular ECM mimic because it is derived from native ECM and provides a bioactive microenvironment for cells.(*197*) In order to examine the impact of culture dimensionality on

the modifications of H3K4me3 across the genome of U251 GBM cells, we cultured the cells in a 2D monolayer on conventional tissue culture polystyrene in a cell flask or seeded them within type I collagen hydrogels for 72 hours. To analyze the impact of hypoxia on U251 H3K4me3 modifications, the cells were also cultured under normoxic (20% O_2) or hypoxic (1% O_2) conditions. Furthermore, our 3D hydrogels had a depth of 1 mm, within the diffusion limits of oxygen, and U251s were seeded at a density of 1 million cells/ml to ensure the cells maintained within the 3D cultures were exposed to uniform oxygen concentrations.(*198*) A MOWChIP-seq was then performed using 1,000 cell sample sets to obtain a genome-wide profile of H3K4me3 for each culture condition (Figure 3.1). Each 3D scaffold only contains approximately 37,000 cells, far below the cell numbers needed for traditional ChIP protocols. The use of an MOWChIP platform enabled us to use a small 1,000 cell sample size and NGS was performed using Illumina Solexa technology.(*189*)

Table 3.1 Raw data of cell numbers collected from digested collagen hydrogel scaffolds. 12 scaffolds were digested using collagenase, after which the cell suspensions were combined and re-suspended in 1 ml PBS prior to counting.

Culture Condition	Hemocytometer Quadrant #1	Hemocytometer Quadrant #2	Hemocytometer Quadrant #3	Hemocytometer Quadrant #4	Total Cell Count
3D Normoxia	39	29	52	37	392,500
3D Hypoxia	45	54	41	39	447,500

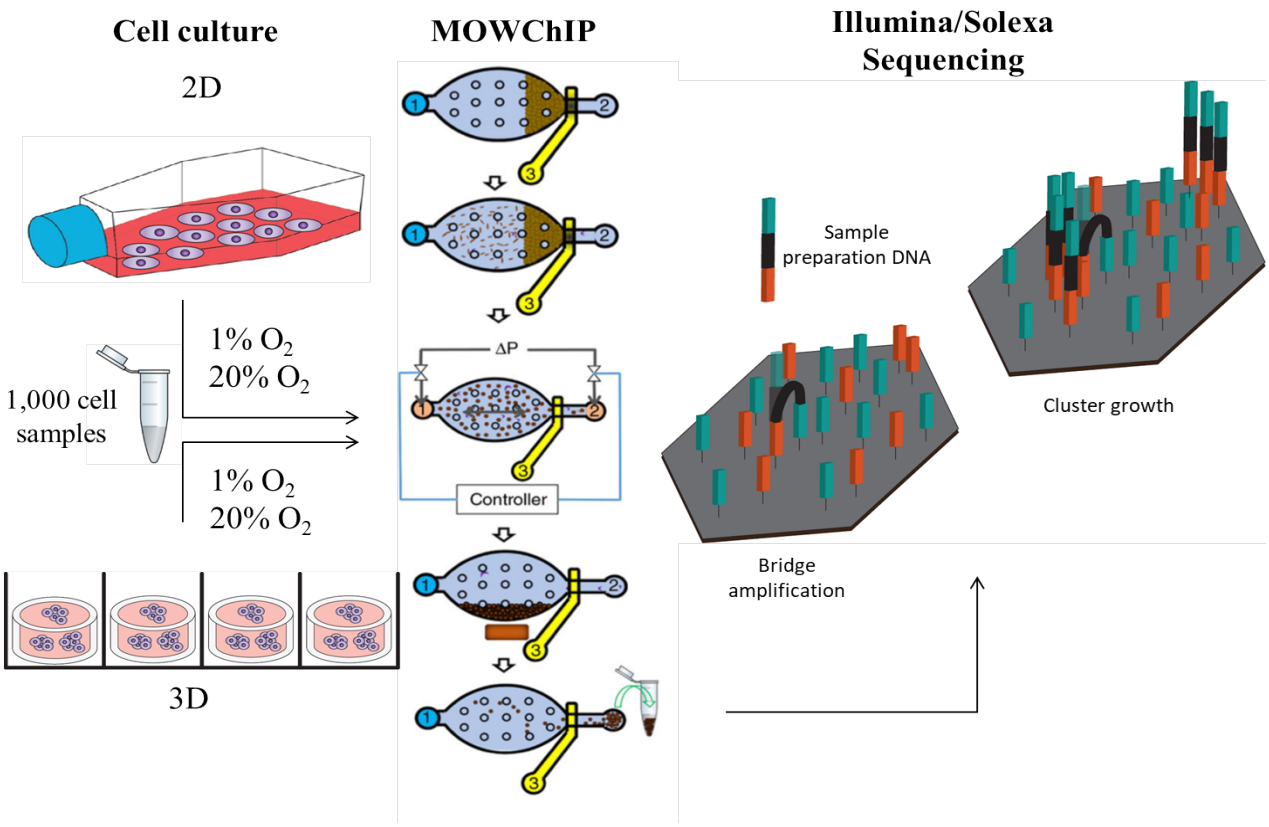


Figure 3.1 Schematic of sample culture and processing.
U251s were cultured under 2D or 3D conditions in normoxic (20% O_2) or hypoxic (1% O_2) oxygen conditions, 1,000 cell sample sizes were collected and processed using a MOWChIP platform, figure adapted with permission from Nature(*189*), followed by Illumina/Solexa sequencing technology

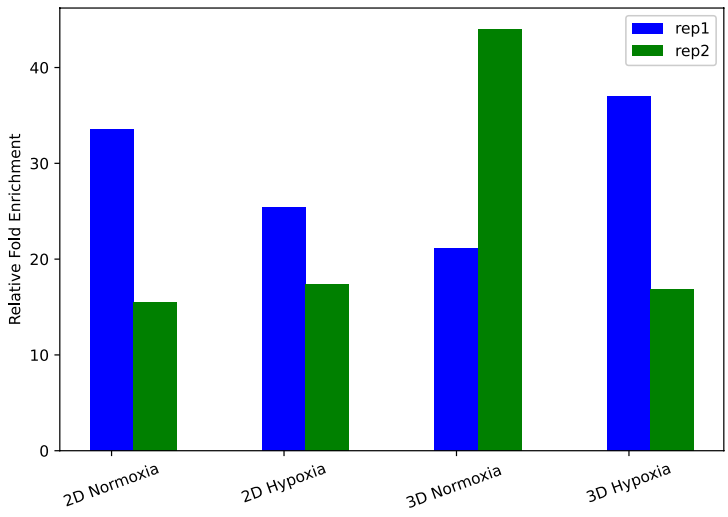


Figure 3.2 Relative fold enrichment at C9orf3 locus calculated using qPCR results.

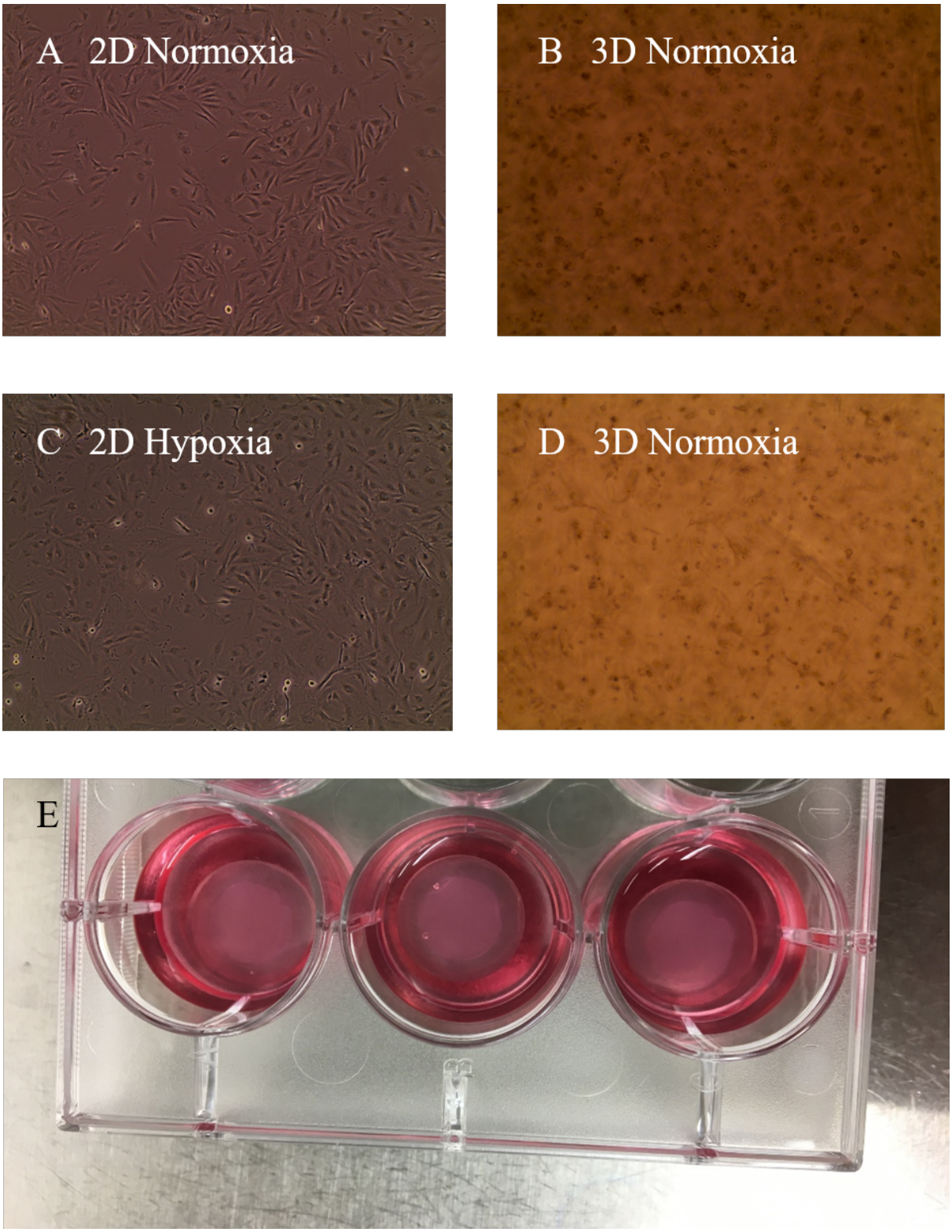


Figure 3.3 Brightfield images of U251 cells under different culture conditions. Brightfield images of U251 cells under (A) 2D and (B) 3D normoxic and (C) 2D and (D) 3D hypoxic culture conditions. (E) Image of 3D scaffolds in well plate.

We observed extensive changes in H3K4me3 profiles when culture conditions such as dimensionality or oxygen level changed, as shown by the differential enrichment at the hypoxia induced CA9 gene locus under hypoxic versus normoxic conditions (Figure 3.5a). In order to determine the impact of hypoxia and culture dimensionality on H3K4me3 profiles, differential enrichment analysis of MOWChIP-seq peak data was performed. In general, genome wide H3K4me3 varied much more substantially due to the alteration of culture dimensionality (3D vs. 2D) than alterations in oxygen status. 11,303 differentially marked regions were discovered under normoxic conditions and 11,863 under hypoxic conditions between 3D and 2D culture sets, compared to the 1,000 and 1,246 differentially marked regions identified under varying oxygen levels in 2D and 3D culture, respectively (Figure 3.5b). The majority of these differentially marked regions (11,371/11,863 or 96% for 2D versus 3D culture under hypoxia, or 11,003/11,303 or 97% for 2D versus 3D culture under normoxia) experienced a decrease in H3K4me3 enrichment when the culture dimensionality changed from 2D to 3D (supplementary info). The genes associated with these differentially marked regions were also analyzed (Figure 3.5b). The majority of genes impacted by oxygen status (94%, 1,180 out of 1,257 genes as

determined by the sum of compartments labeled A divided by the sum of compartments labeled B in Figure 3.5c) were unique to either 2D or 3D culture conditions. In contrast, only 42% (3,796 out of 9,049 as determined by the sum of compartments labeled C divided by the sum of compartments labeled D in Figure 3.5c) of the genes that were impacted by culture dimensionality were unique to either normoxic or hypoxic conditions. Principal component analysis of sequencing data from 2D and 3D samples cultured under normoxic and hypoxic conditions shows a clustering of samples within each treatment group indicating the reliability of the replicates to produce independent and consistent H3K4me3 profiles (Figure 3.5d).

Table 3.2 Summary of H3K4me3 MOWChIP-seq data on 1,000 U251 cells.
Raw data were analyzed using AQUAS pipeline.

Samples	Raw Reads (Million)	Peaks Number	Alignment Rate, %	Redundant Rate, %	FRiP, %	NSC	RSC	Reproducibility test
2D H1	17	65,213	91.34	34	20.61	1.141	2.245	pass
2D H2	13.3	89,318	91.49	23	16.57	1.100	2.003	
2D N1	20.4	62,776	95.37	37	20.25	1.144	2.965	pass
2D N2	15	67,496	94.96	20	20.13	1.098	2.066	
3D H1	15.8	44,526	97.5	21	7.84	1.034	1.602	pass
3D H2	11.4	44,538	96.7	19	10.99	1.054	1.793	
3D N1	27.5	28,160	97.5	20	6.95	1.032	2.111	pass
3D N2	20.2	33,179	97.4	19	8.60	1.036	1.853	

Table 3.3 Differential binding analysis of H3K4me3 across genome.
Differential binding analysis of MOWChIP-seq peak data by DiffBind, P-value = 0.05 used as threshold, ↑ indicates increased binding affinity in hypoxia or 3D samples, ↓ indicates decreased binding affinity in hypoxia or 3D samples.

	2D Hypoxia vs 2D Normoxia		**3D Hypoxia vs 3D Normoxia**		**3D Hypoxia vs 2D Hypoxia**		**3D Normoxia vs 2D Normoxia**	
Number of Differentially Binding Sites	1000		1246		11863		11303	
	318↑	682↓	399↑	847↓	492↑	11371↓	300↑	11003↓

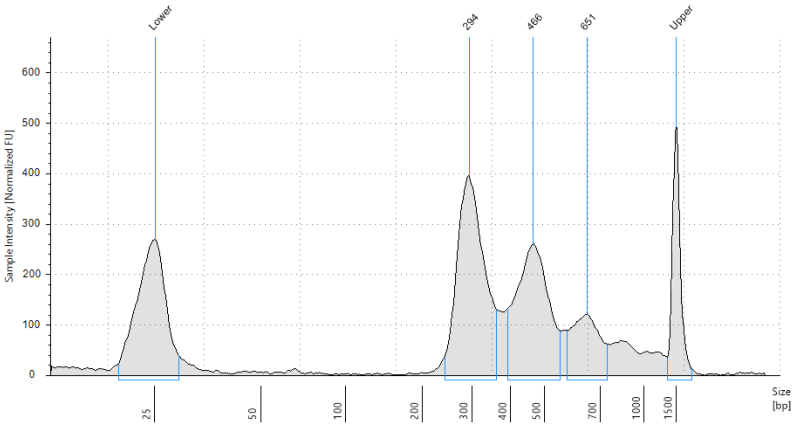
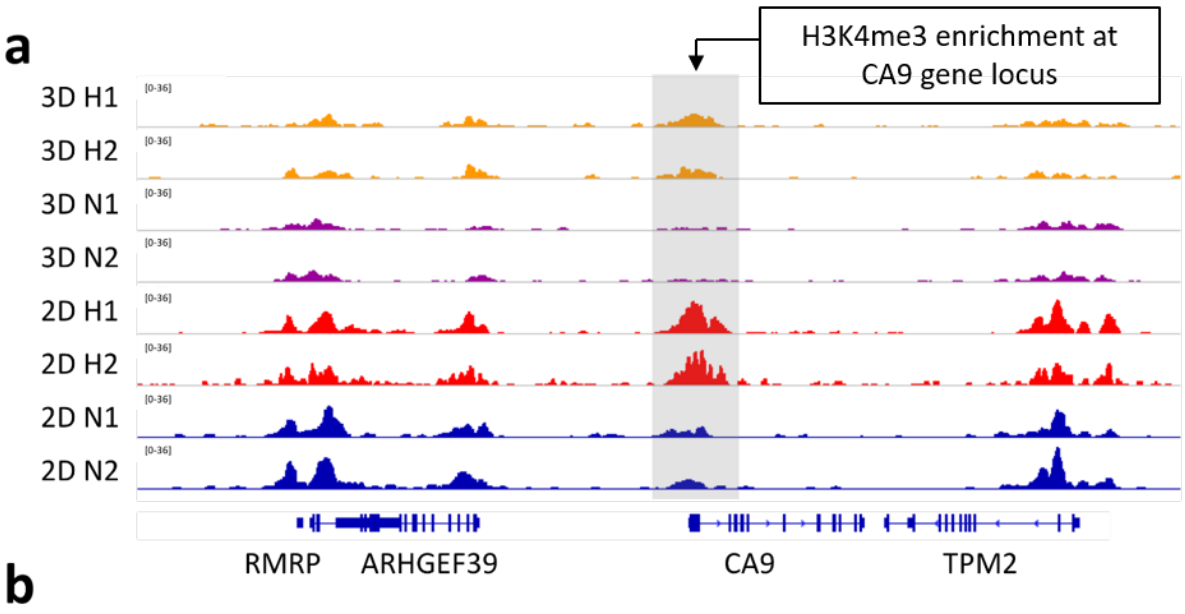


Figure 3.4 Fragment size of a ChIP-seq library with characteristic "nucleosome ladder".
Mnase preferentially cut DNA not binding to histone proteins, resulting in the characteristic nucleosome ladder.



b

	2D Hypoxia vs 2D Normoxia	3D Hypoxia vs 3D Normoxia	3D Hypoxia vs 2D Hypoxia	3D Normoxia vs 2D Normoxia
Number of Differential Marked Regions	1000	1246	11863	11303
Number of related genes	565	769	7277	7025

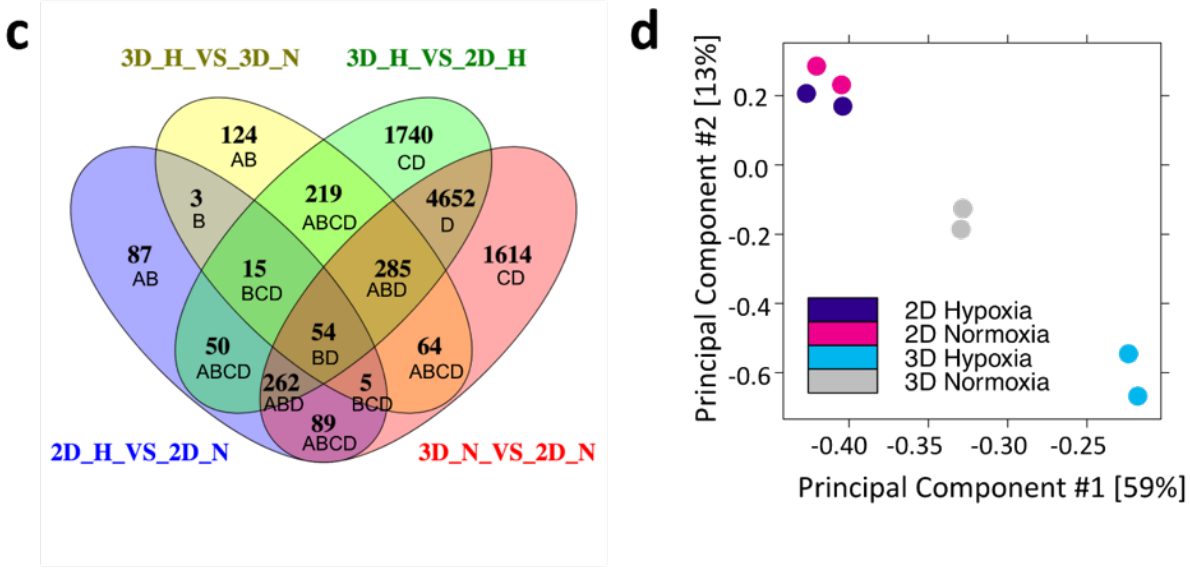
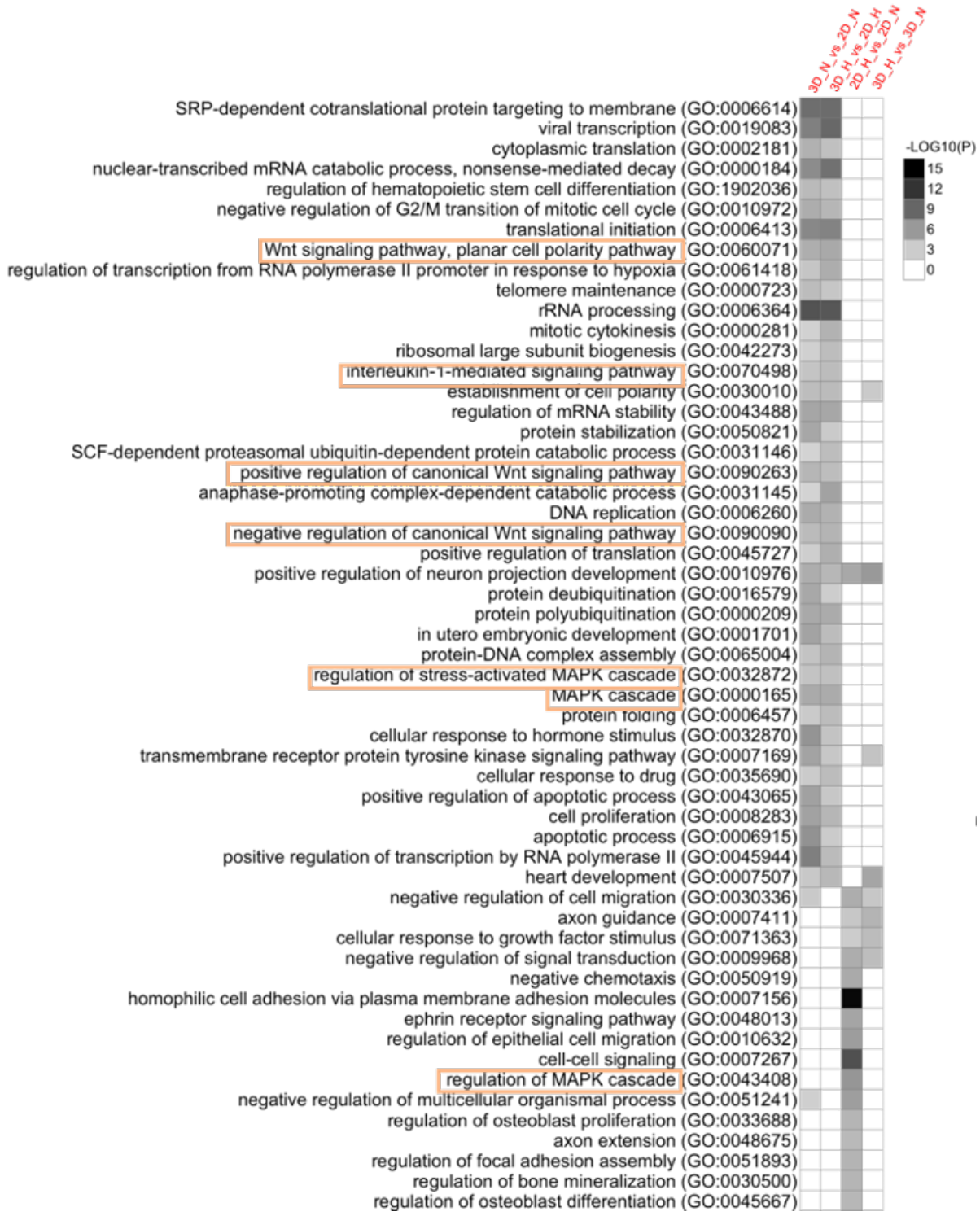


Figure 3.5 Peak analysis of H3K4me3 across the genome.
a. Normalized H3K4me3 MOWChIP-seq signals at CA9 gene locus. **b.** Differential enrichment analysis of MOWChIP-seq peak data by DiffBind, P-value = 0.05 used as threshold, differentially marked regions were identified using edgeR; **c.** The Venn diagram displays the genes where differential H3K4me3 enrichment occurred between changes in culture dimensionality and oxygen status. A denotes the genes where differential binding occurred when oxygen status was changed but was specific to 3D or 2D cultures. B denotes the all the gene sets where H3K4me3 enrichment differed when oxygen status was changed. C indicates the genes where differential binding occurred when culture dimensionality was changed but was unique to either hypoxic or normoxic environments. D indicates that all of the gene sets

where differential binding occurred when culture dimensionality changed; **d.** Principal Component Analysis shows group clustering based on differentially marked regions

Culture dimensionality impacts the H3K4me3 patterns on WNT, IL-1, and MAPK pathways and protein synthesis pathways

Functional enrichment analysis was performed to find out the enrichment of GO (Gene Ontology) terms, which represent biological pathways in which a significant number of genes within those pathways had differences in H3K4me3 enrichment between culture conditions. (Figure 3.6) Following the trends seen previously, the majority of the biological pathways that had changes in H3K4me3 enrichment were a result of altering culture dimensionality as opposed to oxygen status (121 vs 46 GO terms). It is also interesting to note that only 5 of the 41 pathways that had differences in enrichment under hypoxic vs normoxic conditions overlapped between 2D and 3D culture sets. There were 39 GO terms that had alterations in H3K4me3 enrichment when culture dimensionality was changed that overlapped between hypoxic and normoxic conditions. These processes are likely to be the ones solely impacted by culture dimensionality. Among these pathways include alterations in the regulation of Wnt signaling, IL-1-mediated signaling, and the MAPK cascade which have been highlighted in Figure 3.6. We also found certain pathways had decreased H3K4me3 marking under 3D culture conditions compared to 2D culture under both hypoxia and normoxia. Those pathways included G-protein coupled receptor signaling pathways, mRNA splicing via spliceosome, rRNA processing and positive regulation of transcription by RNA polymerase II; the majority of which regulate protein synthesis (Figure 3.7).


3D_N_vs_2D_N
3D_H_vs_2D_H
2D_H_vs_2D_N
3D_H_vs_3D_N
-LOG10(P)
15
12
9
6
3
0
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)
viral transcription (GO:0019083)
cytoplasmic translation (GO:0002181)
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)
regulation of hematopoietic stem cell differentiation (GO:1902036)
negative regulation of G2/M transition of mitotic cell cycle (GO:0010972)
translational initiation (GO:0006413)
Wnt signaling pathway, planar cell polarity pathway (GO:0060071)
regulation of transcription from RNA polymerase II promoter in response to hypoxia (GO:0061418)
telomere maintenance (GO:0000723)
rRNA processing (GO:0006364)
mitotic cytokinesis (GO:0000281)
ribosomal large subunit biogenesis (GO:0042273)
interleukin-1-mediated signaling pathway (GO:0070498)
establishment of cell polarity (GO:0030010)
regulation of mRNA stability (GO:0043488)
protein stabilization (GO:0050821)
SCF-dependent proteasomal ubiquitin-dependent protein catabolic process (GO:0031146)
positive regulation of canonical Wnt signaling pathway (GO:0090263)
anaphase-promoting complex-dependent catabolic process (GO:0031145)
DNA replication (GO:0006260)
negative regulation of canonical Wnt signaling pathway (GO:0090090)
positive regulation of translation (GO:0045727)
positive regulation of neuron projection development (GO:0010976)
protein deubiquitination (GO:0016579)
protein polyubiquitination (GO:0000209)
in utero embryonic development (GO:0001701)
protein-DNA complex assembly (GO:0065004)
regulation of stress-activated MAPK cascade (GO:0032872)
MAPK cascade (GO:0000165)
protein folding (GO:0006457)
cellular response to hormone stimulus (GO:0032870)
transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)
cellular response to drug (GO:0035690)
positive regulation of apoptotic process (GO:0043065)
cell proliferation (GO:0008283)
apoptotic process (GO:0006915)
positive regulation of transcription by RNA polymerase II (GO:0045944)
heart development (GO:0007507)
negative regulation of cell migration (GO:0030336)
axon guidance (GO:0007411)
cellular response to growth factor stimulus (GO:0071363)
negative regulation of signal transduction (GO:0009968)
negative chemotaxis (GO:0050919)
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)
ephrin receptor signaling pathway (GO:0048013)
regulation of epithelial cell migration (GO:0010632)
cell-cell signaling (GO:0007267)
regulation of MAPK cascade (GO:0043408)
negative regulation of multicellular organismal process (GO:0051241)
regulation of osteoblast proliferation (GO:0033688)
axon extension (GO:0048675)
regulation of focal adhesion assembly (GO:0051893)
regulation of bone mineralization (GO:0030500)
regulation of osteoblast differentiation (GO:0045667)

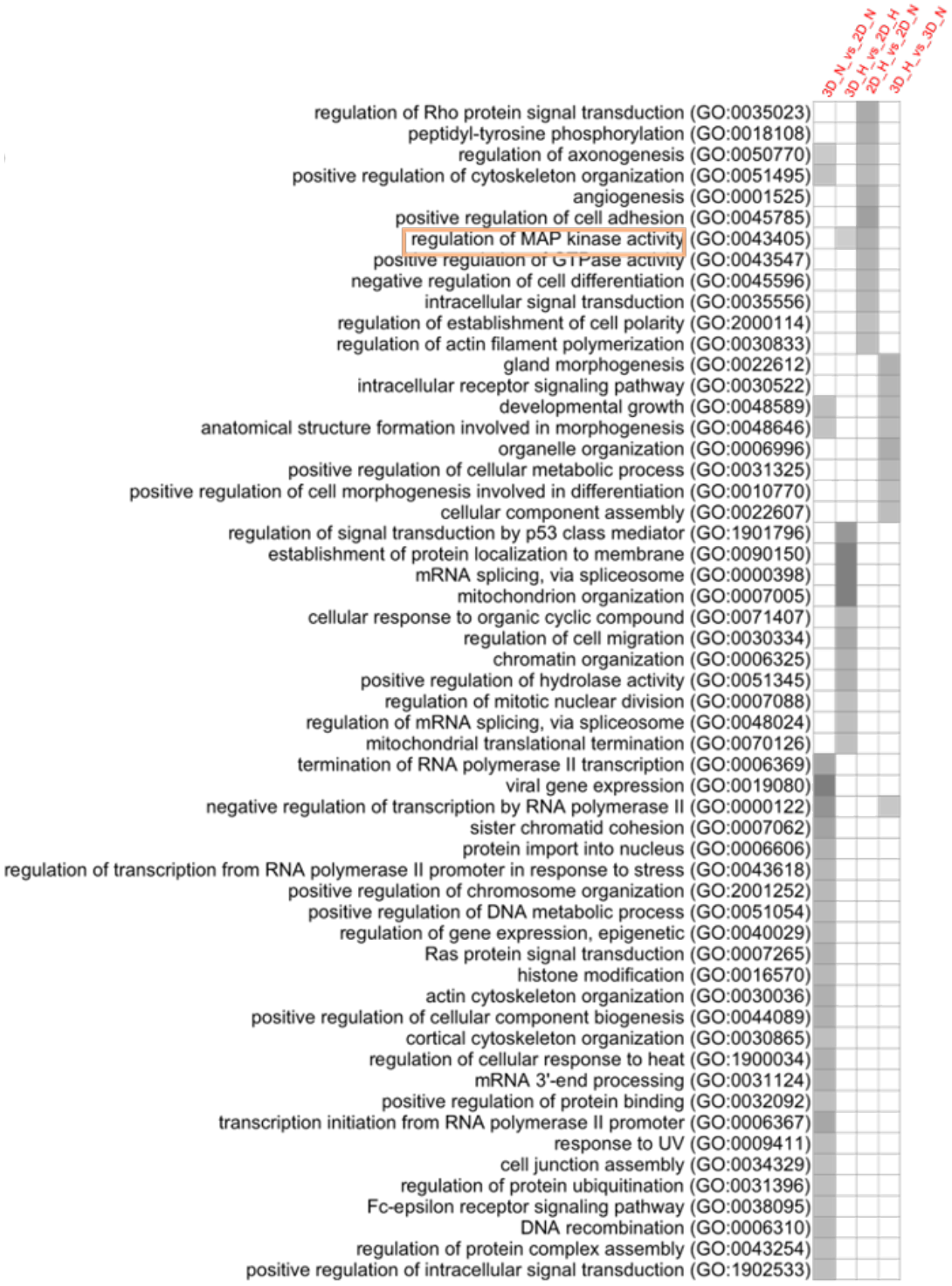
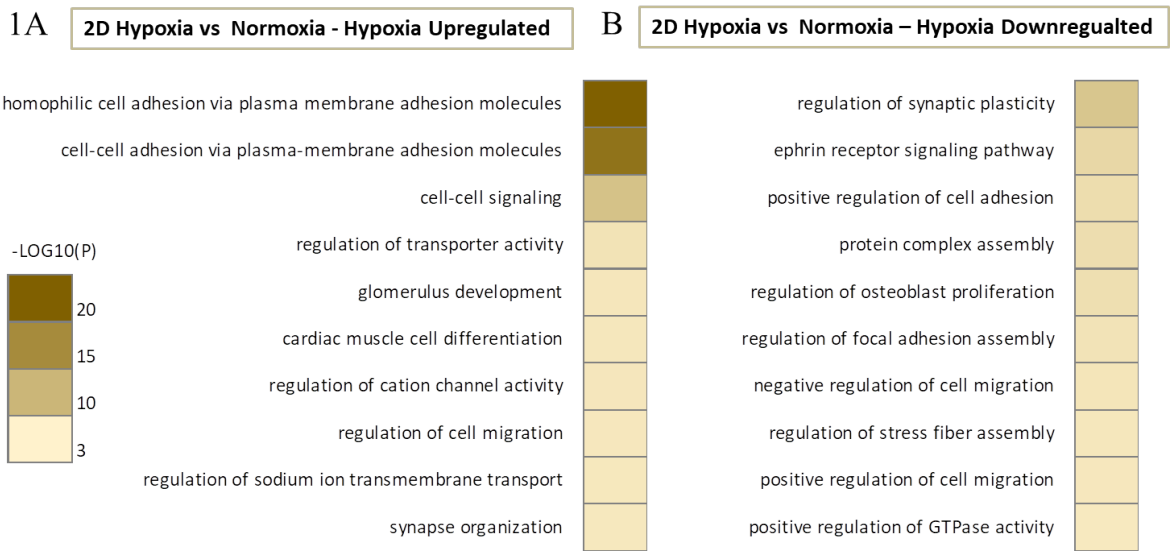


Figure 3.6 GO term analysis of H3K4me3 differential binding on biological pathways. Functional enrichment analysis obtained by PANTER Overrepresentation Tests on the sets of differentially marked regions using a GO database released February 2, 2018. LOG10(P)= -3.5 was used as a cutoff value. Heatmap shows the biological pathways that had differential H3K4me3 binding between the culture conditions indicated above each column.

Culture dimensionality impacts the distribution of H3K4me3 within gene structures

To examine the broader characteristics of the H3K4me3 profile, the location of the histone modification across the entire genome and within gene structures was analyzed. The 3D cultured cells under both normoxic and hypoxic conditions showed a greater concentration of H3K4me3 in gene promoter regions, whereas in 2D cultured cells there was a more uniform distribution of H3K4me3 between the promoter and intergenic regions (Figure 3.8a). Differences between 2D and 3D cultured cells were also observed in analyzing the H3K4me3 enrichment around the transcription start sites (TSS). There was a greater enrichment of H3K4me3 near TSS in 3D cultured cells under both normoxic and hypoxic conditions (Figure 3.8b).



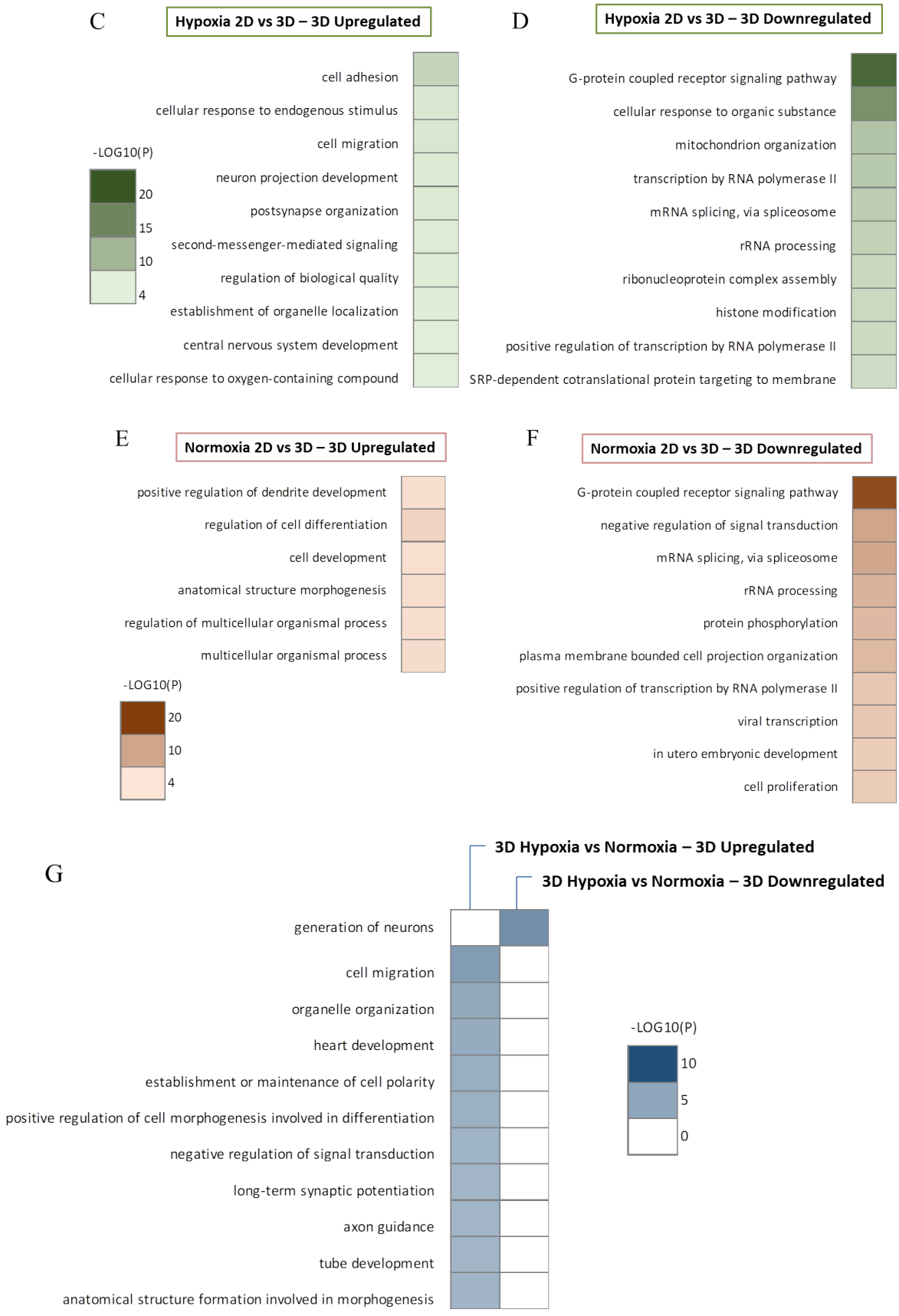


Figure 3.7 Pathway enrichment analysis.
Heatmaps of functional enrichment obtained by a PANTER Overrepresentation Test on the sets of differential binding sites using GO Ontology database released February 2, 2018, top

GO terms with lowest P-value were selected for (a) 2D hypoxia vs 2D normoxia, hypoxia upregulated and (b) downregulated, (c) 2D hypoxia vs 3D hypoxia, 3D upregulated and (d) 3D downregulated, (e) 2D normoxia vs 3D normoxia, 3D upregulated and (f) 3D downregulated, (g) 3D hypoxia and normoxia upregulated and downregulated pathways.

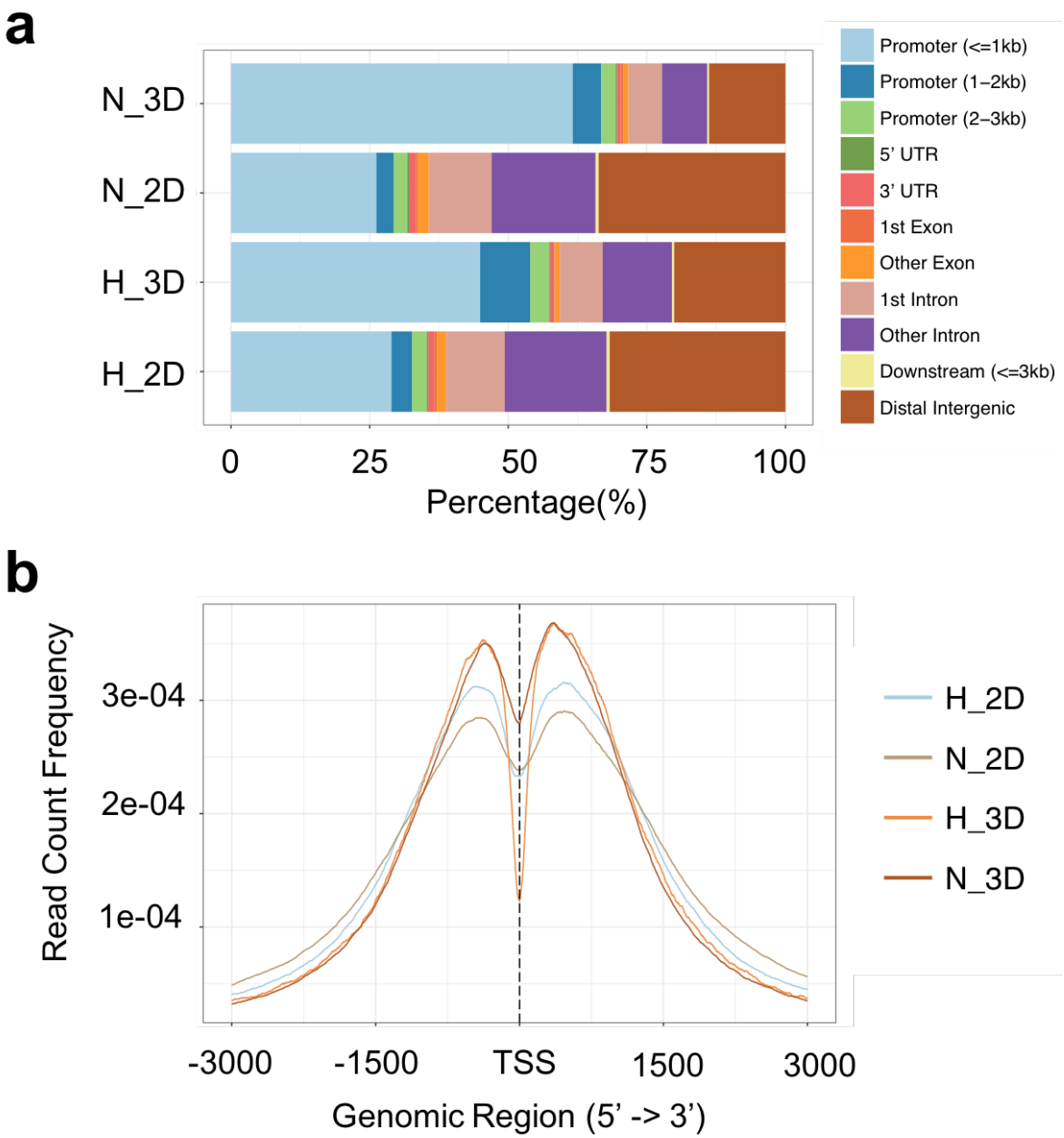


Figure 3.8 Analysis of H3K4me3 distribution along gene structure.
a. genomic annotation among MOWChIP-seq data. The 3D cultured cells showed a greater concentration of H3K4me3 in gene promoter regions, whereas in 2D cultured cells there was a more uniform distribution of H3K4me3 between the promoter and intergenic; **b.** average profiles of ChIP peaks among different culture conditions, replicates were analyzed and pooled using AQUAS pipeline. There was a greater enrichment of H3K4me3 near TSS in 3D cultured cells under both normoxic and hypoxic conditions

Hypoxia changes epigenomic and transcriptomic response to bacterial infection

Bacteria have emerged recently as profound regulators of the Tumor Microenvironment (TME)(*199*). Yet analyzing the complex interactions of tumor-resident bacteria within specific components of a dynamic TME remains challenging in vivo due to confounding multi-factorial interactions, and in vitro host-microbe interaction models typically lack vascular and immune cell components critical to in vivo pathophysiology. This knowledge gap is especially evident when determining why some commensal bacteria in rare instances acquire a malignant influence and act to exacerbate cancer progression. To challenge these technical and conceptual limitations, we harnessed tissue engineering and microfluidics to analyze complex processes at the tumor-microbe-immune system axis which may ultimately reveal tumor-specific microbiome cues that should be considered as emerging hallmarks of cancer as well as novel targets for therapy.

We will specifically work to answer the question of how the non-migratory commensal gram-negative oral bacterium Fusobacterium nucleatum is able to accelerate colorectal cancer (CRC) through enhancement of inflammatory signaling, cell growth acceleration, and metastasis. We hypothesize that these 'oncomicrobes' actively hijack intrinsic host cell migration programs for their own survival and dissemination, which in turn leads to adverse reactions by the host in the form of cancer modulation. F. nucleatum is an ideal test microbe for this work because while it is nonmotile, it nevertheless successfully disseminates from the mouth to CRC tissue by mechanisms that remain poorly understood. We have recently discovered that this microbe directly drives cancer spread by amplifying existing tumor pro-migratory pathways(*200*) (reported as a Science Signaling cover article and highlighted in Scientific American). While prior research from ourselves and others suggests that viable intracellular F. nucleatum dominate this process in CRC, our more recent data suggest that the mechanism(s) of invasion and pathway regulation are not conserved within phenotypically

similar consensus molecular subtypes (CMS) - a critical finding that will guide the way to novel mechanisms and targets for therapy. The experience of the PIs with 3D tumor models(*201-203*), low-input genome-wide analysis(*44, 130, 204*), cancer-inducing bacteria(*205*), and colon cancer immunology(*206*) uniquely positions our team to uncover F. nucleatum-CRC interactions that enhance metastasis through TME niche-dependent mechanisms.

We have found that hypoxia condition extensively decreased the intensity of H3K27ac signals across the genome (Figure 3.9a), in agreement with the previous literature(*207*). H3K27ac is an important activating mark that predicts enhancers, thus extensive suppression of H3K27ac signal suggests significant changes in the gene regulation. Interestingly, the RNA-seq data showed significant but not lopsided changes in gene expression (HF has 1847 upregulated and almost equal number of downregulated genes (1567) compared to NF) (Figure 3.9b). Overall various infection and oxygen conditions created distinguishable epigenomic profiles (Figure 3.9c), with the ones under normoxia largely having higher correlation with the standard profiles of the cell line (ENCODE data). Finally, the infection by E. coli and Fuse created significant difference in a number of biological processes, including rRNA maturation and mRNA localization (Figure 3.9d).

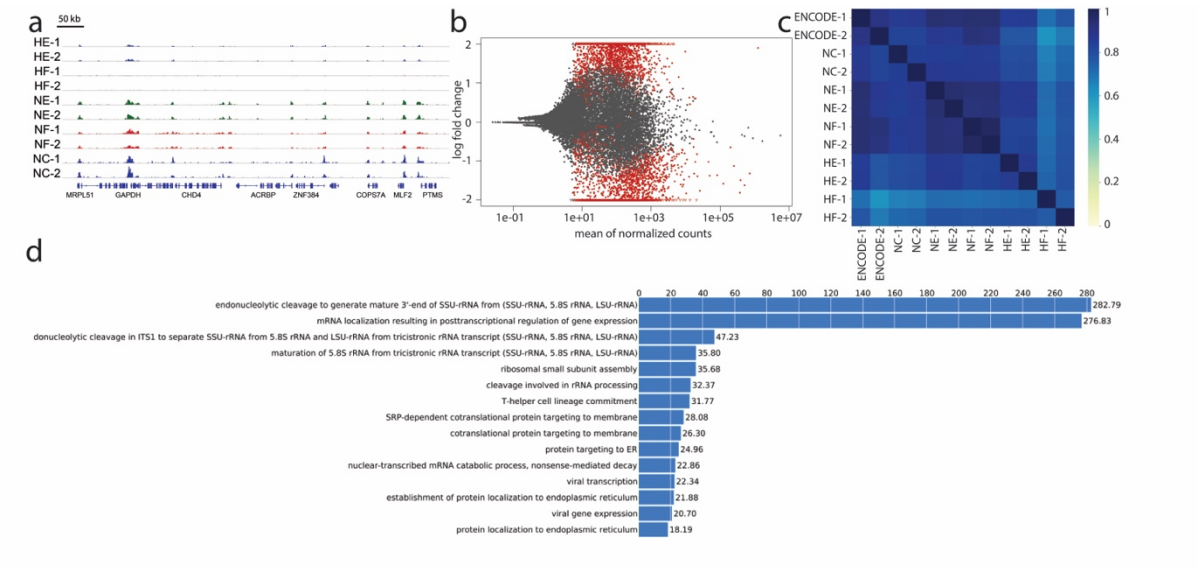


Figure 3.9 Epigenomic and transcriptomic variations upon Fuse and E. coli infection under normoxia and hypoxia.
a. Genome browser tracks on H3K27ac in 10,000 cells under various conditions obtained by MOWChIP-seq. HE: E. coli infection under hypoxia, HF: Fusobacteria infection under hypoxia, NE: E. coli infection under normoxia, NF: Fusobacteria infection under hypoxia, NC: normoxia control. **b.** MA plot that shows expression variations of genes between HF and NF reflected by RNA-seq data. **c.** Heat map that shows the genome-wide correlations among H3K27ac data under various conditions. **d.** Gene ontology (GO) enrichment analysis that shows the biological processes differentially involved between HF and HE. Top 15 GO terms are included.

To the best of our knowledge, this is the first study that combines an *in vitro* 3D micro-tissue culture model with ChIP-seq to analyze the global histone modifications of U251 glioblastoma cells under varying culture conditions. A major barrier to overcome has been the small number of cells typically contained within these platforms. While such platforms provide the power to analyze dynamic responses to complex spatiotemporal gradients, there may only be tens to hundreds of cells within localized niches. For this study we have used MOWChIP-seq to perform epigenomic analysis on 1,000 cell samples characteristic of the number of cells contained within our micro-tissues.

For this study, we elected to direct our analysis towards epigenetic alterations in the U251 glioblastoma multiforme cell line. As mentioned previously, GBM is a highly heterogeneous cancer with few effective treatment options. GBM is characterized by tumor

recurrence of aggressive cell phenotypes and it has been shown to develop epigenetic mechanisms of therapeutic resistance.(*180, 208*) Obtaining a better understanding of how the microenvironment may be facilitating the generation and support of the highly migratory and aggressive tumor cell phenotype that recurs after treatment, as well as how cues from the microenvironment may be shaping the development of chemoresistance, could elucidate potential new treatment targets. It has been shown that pediatric gliomas are biologically distinct from adult gliomas.(*209*) The observation that certain pediatric gliomas harbor very few recurrent genetic events and instead are often regulated by epigenetic events such as CpG hypermethylation and post-translational histone modifications further underscores the importance of understanding the epigenetic events occurring within a tumor and throughout its progression in order to develop effective treatment strategies.(*210*) Not only could studies such as this elucidate the mechanisms behind GBM cellular heterogeneity and the development of therapeutic resistance overall, but could also help to better differentiate these processes in pediatric versus adult glioblastomas.

We used traditional 2D culture methods along with our 3D micro-tissue models based on type I collagen as a widely used physiological tissue mimic to assess the impact of 3D culture on the H3K4me3 profile. The impact of hypoxia on the global H3K4me3 histone profile was also assessed by culturing the U251s under normoxic (20% O_2) and hypoxic (1% O_2) conditions. Many studies have also elucidated the significant impact of the microenvironment on cell behavior, particularly in the context of cancer.(*185, 211, 212*) For example, breast cancer progression is not only dependent on genetic changes, but alterations in the surrounding ECM can drive tumor cell dissemination as well.(*213*) The ability of the microenvironment to drive cell phenotype alterations in cases such as this warrants further investigation into the potential epigenetic mechanisms behind this regulation of cell behavior. The presence of hypoxia is another microenvironmental factor that alters tumor cell phenotype and drives

heterogeneity within a tumor.(*214, 215*) Our culture system enabled us to expose U251 GBM cells to these microenvironmental stresses and analyze how histone modifications are impacted to help answer questions that have arisen surrounding the role of epigenetics in tumor initiation and progression. The use of small cell samples in this study is also an important first step in the ability to analyze cells within hypoxic vs normoxic regions of an *in vitro* or *in vivo* tumor or any other subpopulations of tumor cells that spontaneously form during tumor development. The epigenetic profile obtained through this analysis could be used to enhance our understanding of the mechanisms driving tumor development as well as to develop personalized therapeutic strategies.

Table 3.4 H3K27ac ChIP-seq data summary

	Alignment rate, %	Unique reads, %	Peak number
HE1	97.48	74.89	5,911
HE2	97.45	71.88	4,655
HF1	95.1	83.36	533
HF2	95.08	80.25	540
HN1	97.73	83.21	131
HN2	97.80	77.95	82
NE1	97.27	69.7	16,650
NE2	97.69	73.55	17,100
NF1	97.84	70.78	20,260
NF2	97.92	76.87	19,933
NN1	96.09	29.01	9,863
NN2	96.68	24.66	10,682

Table 3.5 RNA-seq data summary.

	Alignment rate, %	Unique reads, %	Exon rate, %
HE1	82.16	6.41	~50
HE2	64.64	7.39	~50
HF1	94.47	19.54	~90
HF2	94.19	20.82	~90
HN1	95.16	21.82	~90
HN2	95.31	18.03	~90
NE1	91.93	11.84	~55
NE2	91.99	13	~55
NF1	85.06	7.51	~50
NF2	77.80	6.31	~50
NN1	95.76	32.05	~90
NN2	95.72	30.02	~90

Using MOWChIP-seq peak data analysis, we observed that both dimensionality and oxygen status impacted H3K4me3 enrichment across the respective genomes. The variation in H3K4me3 between cells cultured in 2D versus 3D was approximately ten-fold greater than the differential enrichment that occurred between normoxic and hypoxic conditions in the same culture platform. This may indicate that certain microenvironmental factors such as the culture dimension, substrate, or ECM composition that cells are exposed to may have a greater impact on cell phenotype than other factors such as oxygen concentration. Interestingly, this is consistent with previous work which analyzed global gene expression changes resulting from alterations in culture dimensionality.(*216*) It was also observed that, overall, the intensity of H3K4me3 peaks decreased under 3D culture conditions compared to 2D. The decrease in global H3K4me3 levels under 3D culture conditions could indicate a decrease in overall gene expression. This global reduction in H3K4me3 levels in 3D culture could be a result of

differences in cell shape or demethylase activity or both between 2D and 3D culture conditions, which would agree with the increase in demethylase expression, such as KDM5B, that has been observed in glioma tissues.(*217*) Previous work has also shown that cell shape can regulate global histone modifications. In the study by Le Beyec et al. cell rounding was a sufficient stimulus to induce histone deacetylation and chromatin condensation in the absence of biochemical signals transduced from the ECM.(*218*)

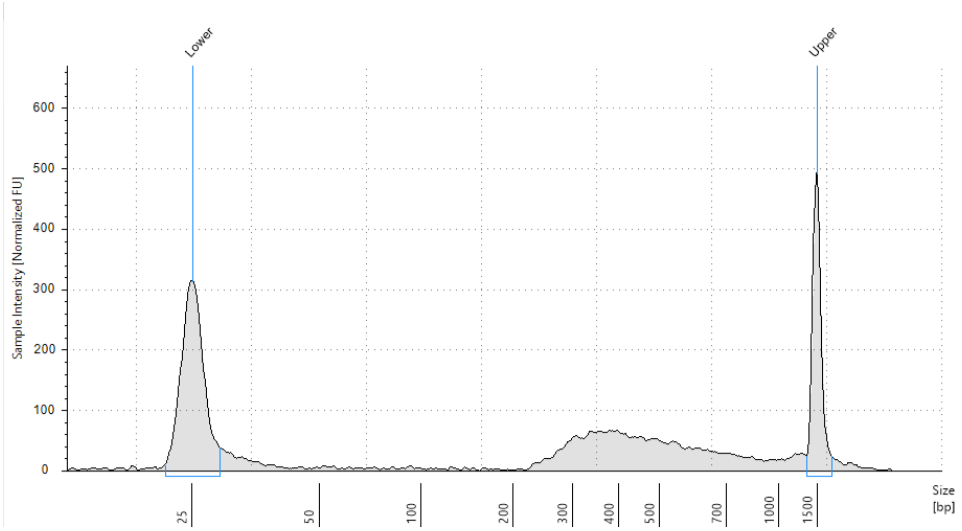


Figure 3.10 Fragment size of an RNA-seq library prepared using Tn5 transposase. Optimized tagmentation reaction results in a fragment length between 200 and 700 bp.

Understanding what biological pathways are involved in the H3K4me3 differential enrichment is the next step in linking culture dimensionality and oxygen concentration to variations in cell phenotype. Following what was observed regarding the number of differentially marked regions, a greater number of signaling pathways were impacted by culture dimensionality than oxygen concentration. GO term analysis identified the pathways that had the greatest changes in H3K4me3 when culture conditions were manipulated. The lack of overlapping terms between 2D and 3D culture sets when oxygen status was altered illustrates varying epigenetic responses to hypoxia depending on culture condition, potentially signifying that culture dimensionality alters what histone modifications occur in response to other

environmental stresses. We do not believe that the variations observed between the 2D and 3D culture sets was due to disparities in oxygen concentration between the culture conditions. Although the oxygen diffusion rates within the 3D scaffold could slightly alter the timescale over which the cells would be exposed to changes in oxygen concentration, uniform oxygen conditions in the 3D cultures should be present within an hour.(*198*) Previous work that has analyzed the downstream gene expression profiles of cells cultured under similar conditions found that the majority of hypoxia-associated genes were unique to either 2D or 3D cultures.(*216*)

The GO term analysis also elucidated important signaling pathways that had differential H3K4me3 enrichment when culture dimensionality was altered under both hypoxic and normoxic conditions. These pathways included the Wnt signaling pathway, which is involved in cell proliferation and migration, and the MAPK cascade which regulated cell cycle entry and proliferation. Importantly, both signaling pathways play a major role in many cancers and alterations in histone modifications on these pathways between culture dimensions may have significant impacts on cell behavior and cancer progression. Furthermore, we found that the IL-1 mediated signaling pathway had variations in H3K4me3 enrichment when culture dimensionality was changed, which agrees with previous reports that inflammatory signaling is altered between 2D and 3D culture conditions.(*216*) The alterations in inflammatory signaling in the 2D and 3D cultures could have implications in the cellular response to hypoxia and regulation of angiogenesis.(*219*) Our GO term analysis also showed that H3K4me3 levels increased in 2D cultures, regardless of oxygen status, on the mRNA splicing via spliceosome, rRNA processing, and positive regulation of transcription by RNA polymerase II biological pathways. Each of these pathways involve protein synthesis and are regulators of gene expression, which are pathways that are characteristically upregulated in traditional 2D *in vitro* cultures compared to cells *in vivo*.(*220*) Interestingly, this over-proliferative phenotype of 2D

cells is thought to be a major cause of the failure of these traditional cultures to accurately predict patient responses to cancer therapeutics.(*221*) Through our differential enrichment and GO term analysis, we have shown that the epigenetic response to microenvironmental stresses is highly dependent on the culture dimensionality and the H3K4me3 differential enrichment patterns on oxygenation-independent biological pathways agree with important phenotypic differences (e.g. protein synthesis and gene expression) that have been observed between 2D and 3D cultures. In order to accurately analyze the role of epigenetics in cancer progression, physiologically relevant *in vitro* models are essential. Our work has shown that epigenetic responses are not necessarily translatable between 2D and 3D culture models and the H3K4me3 profiles we have demonstrated suggest that 3D culture platforms may more accurately mimic an *in vivo*-like epigenome. However, further analysis of epigenetics *in vitro* and *in vivo* is needed in order to validate current *in vitro* culture models, which was not ultimately the goal of this present study.

We next analyzed the distribution of H3K4me3 peaks along the gene structure. Previous work has noted that lower levels of histone modifications lead to a poorer prognosis in cases of prostate cancer.(*179*) This was thought to potentially be an indication of a re-distribution of histone modifications from most histones throughout the genome, to a limited number of genes that confer a more aggressive phenotype to the tumor. However the decrease in global histone modifications could also indicate a switch from a euchromatic to a heterochromatic state, that protects the cells against genotoxic stress by limiting DNA exposure.(*178, 179*) There have also been studies that indicate that the location of the modification impacts whether gene activation or repression is induced. H3K4 methylation in both the coding and regulatory regions is associated with active transcription. Similarly, H3K9 methylation in the coding region is associated with active transcription, however H3K9 methylation in the regulatory region will lead to transcriptional repression.(*222, 223*) Not only are the specific genes where

histone modifications are occurring important, but so is the location of the histone modifications along that gene. Our analysis shows that culturing cells within 3D platforms results in a greater concentration of H3K4me3 on the gene promoter regions compared with cells cultured in 2D. The enrichment profile we observed in 3D could potentially impact both gene transcription and chromatin configuration as H3K4me3 concentrated at promoter regions has been associated with gene activation, but a reduction in overall H3K4me3 and concentration at specific sites could also alter the conformation of the chromatin, negatively regulating gene transcription but protecting the DNA from toxic stresses. We also observed a greater concentration of H3K4me3 around TSS in 3D versus 2D cultured cells. Previous work has shown that an increase in H3K4me3 distribution frequency near TSS was positively correlated with gene expression.(*224*) However, future downstream gene expression analysis is needed in order to better understand the implications of the enrichment profiles we have observed in this study.

There is currently no consensus on what extent genetic or epigenetic mechanisms are driving intratumoral cancer cell heterogeneity.(*225*) Overall this study has shown that histone modifications are strongly regulated by culture conditions and that the ability to engineer and tune culture microenvironments will greatly contribute to furthering our understanding of how the tumor microenvironment could drive cell phenotypic alterations. During tumor progression towards an invasive metastatic state, the ECM and the rest of the tumor microenvironment also change along with the developing tumor.(*211, 226*) Our observation that histone modification profiles are strongly regulated by culture context could indicate that alterations in the ECM during tumor progression may be inducing epigenetic alterations that contribute to the gene expression variations occurring during tumor progression. This process would require downstream gene expression analysis of the induced epigenetic changes within the tumor cells to validate but could be an overlooked mechanism of tumor development. It will also be

important to ensure the scaffold composition accurately recapitulates the physiological system of interest. Along with the epigenomic differences between 2D and 3D cultures presented in this work, transcriptomic variations have been found when culture dimensionality and 3D scaffold composition were altered.(*227*) These discoveries emphasize the need to investigate specific cell-ECM interactions in future work, where perhaps the use of a more brain-specific matrix material, like hyaluronan, would have resulted in epigenomic profiles that more closely represent what is seen *in vivo* than those we characterized with the use of collagen hydrogels. Along with the significant variations in histone modification profiles observed between 2D and 3D culture conditions, we also show profound changes in epigenetic stress response as a function of culture context, which could have significant implications for therapeutic responses. It is known that epigenetics can play a role in therapeutic resistance, therefore it is important to take into account the epigenetic profile of cancer cell subpopulations and the impacts of varying microenvironmental conditions on this profile in order to administer effective therapies.(*180*) The capabilities of the high-resolution MOWChIP platform, which has been used to analyze as few as 100 cells, will also be a great asset in analyzing the epigenetic heterogeneity present within cell samples from the same culture.(*189*) We have shown it is possible to analyze small numbers of cells obtained from *in vitro* model tissue samples, which is a promising start to the analysis of the small cancer cell subpopulations that exist *in vivo*, such as cancer stem cells or circulating cancer cells, which are a major contributor to disease progression and therapeutic resistance. However, the ability to generate and maintain cellular heterogeneity within our 3D culture models will be important to validate in future studies. The integration of engineered *in vitro* tumor mimics with high resolution epigenetic analysis that was utilized in this study will help to further the understanding of how microenvironmental epigenetic regulation of cell phenotype works in concert with genetic variations to generate

cellular heterogeneity within a tumor, and the characterization of these processes will enable more effective personalized therapeutic strategies to be developed.

3.4 Conclusion

The current study helps to elucidate how microenvironmental signals can impact histone modification profiles across a human genome. The use of an engineered 3D micro-tissue model in concert with traditional 2D culture methods and varying oxygen concentrations enabled us to analyze the impact of culture dimensionality and oxygen status on H3K4me3 enrichment. Our findings illustrated that microenvironmental stresses directly impact histone modification profiles. Culture dimensionality is an essential regulator of the histone modifications that occur in response to microenvironmental signals and the global methylation profile. The H3K4me3 biological pathway binding profile also suggests that the phenotype of cells cultured in the 3D platforms more closely resembles what is seen *in vivo* compared with the 2D counterparts. This was also the first study that has analyzed small sample sizes of cells culture in 3D micro-tissue models using a high-resolution ChIP-seq platform. The methods employed in this work could be a key tool to understanding how the tumor microenvironment regulates tumor cell phenotype, particularly characterizing the formation of heterogeneous tumor cell subpopulations. Utilizing sensitive cell characterization assays such as this could help to improve the efficacy of personalized cancer therapy strategies.

Chapter 4 - Multiplexed and Ultralow-Input ChIP-seq Enabled by Tagmentation-Based Indexing And Facile Microfluidics

4.1 Introduction

Epigenome constitutes an important layer that regulates gene expression and dynamics during development and diseases. Extensive efforts have been made to develop epigenome profiling methods using a low number of cells and with high throughput. Chromatin immunoprecipitation (ChIP) is the most important approach for profiling genome-wide epigenetic changes such as histone modifications. In this report, we demonstrate microfluidic ChIPmentation (mu-CM), a microfluidic technology that enables profiling cell samples that individually do not generate enough ChIP DNA for sequencing library preparation. We used a simple microfluidic device to allow 8 samples to be processed simultaneously. The samples were indexed differently using a tagmentation-based approach (ChIPmentation) and then merged for library preparation. Histone modification profile for each individual sample was obtained by demultiplexing the sequencing reads based on the indexes. Our technology allowed profiling 20 cells and is well suited for cell-type-specific studies using low-abundance tissues. Molecular biology within cells is not only affected by DNA sequences (genomics) but also by epigenomic regulations. Epigenomic regulatory mechanisms include DNA methylation, histone modification, higher-order chromatin organizations and regulations by non-coding RNAs(*11*). These mechanisms do not directly change DNA sequence but strongly affect gene expression and cell state. Histone modifications, or post-translational modifications of histones, play pivotal roles in activating and inhibiting transcription(*228*). Histone modifications define the "open/closed" state of chromatin and provide binding domains for transcription factors(*229*). Deregulation of histone modifications has been associated with various diseases, including cancer(*26*). Hence specific histone marks involved in disease development have

become new targets for drug development and these studies have yielded promising clinical results(*230*).

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is the gold standard technique to profile genome-wide histone modifications(*231*). Major limitations of ChIP-seq include the large number of cells required (~10^6 cells) and time-consuming and complex process (~3 days). Numerous efforts have been made in the past years to reduce sample size to hundreds to tens of cells using ChIP or ChIP-free methods (iChIP(*232*), MOWChIP-seq(*44, 233, 234*), μChIP-seq(*235*), STAR ChIP-seq(*236, 237*), CUT&RUN(*128, 238-241*), TCL(*242*), SurfaceChIP-seq(*133*), LIFE-ChIP-seq(*46*)). More recently, there have been several reports on single-cell methods that reveal epigenetic heterogeneity with 1,000 to 17,000 unique reads per cell (Drop-ChIP(*48*), scChIC-seq(*49*), scCUT&Tag(*50*), uliCUT&RUN(*51*), scChIP-seq(*52*), ACT-seq(*53*), sc-itChIP-seq(*54*), coBATCH(*55*)). Although single-cell technologies are extremely informative for deconvoluting cellular heterogeneity, their data quality is generally compromised due to the lossy single-cell barcoding process. For example, a low-input technology like SurfaceChIP-seq produced 3 million unique reads on H3K4me3 using 30 cells(*133*). This is 100 times more unique reads than pooling 30 single-cell data sets when the single-cell technology yields only 1000 unique reads per cell(*48*). The low limit of the existing low-input technologies is often determined by the amount of ChIP DNA required for library preparation. For example, in our previous SurfaceChIP-seq work, the sequencing library preparation kit required a minimum of 10 pg DNA and our ChIP step collected about 10-15% of the genomic DNA. This means that we needed to start with at least 30 cells (containing roughly 5-6 pg genomic DNA per cells) in order to prepare the sequencing library. In this work, we aim to further lower the number of cells required for each sample by conducting multiplexing and combining multiple samples, taking advantage of tagmentation-based indexing and a simple microfluidic system.

In this work, we use a simple microfluidic device to facilitate simultaneous processing of multiple cell samples. The microfluidic ChIP process is combined with a tagmentation-based step (i.e. ChIPmentation(*243*)) to index each sample with unique sequencing adaptors. The produced ChIP DNA from various samples is subsequently pooled and used to produce one sequencing library. The sequencing data can be demultiplexed based on the indexes to produce ChIP-seq data sets on each sample. Because the ChIP DNA for library preparation was pooled from multiple samples, our approach allows each sample to contain less cells than the amount dictated by the library preparation requirement. Our method, referred to as microfluidic ChIPmentation or mu-CM, permits processing 8 assays in one run with as few as 20 cells per assay and the entire protocol can be finished in 7 h. Our technology provides new opportunities for the study of histone modifications in low-abundance samples in the context of precision medicine. The data and discussion presented in this chapter had been published in Anal Chem 2020(*47*).

4.2 Material and methods

Fabrication of the microfluidic device

Microfluidic device was fabricated by soft lithography using poly(dimethylsiloxane) (PDMS) (RTV615, Momentive). Design was drawn up using LayoutEditor (juspertor GmbH) and laser-plotted on photomasks (10,000 dpi). Features in the photomask were molded onto a silicon wafer (University Wafers) by photolithography. SU-8 2025 photoresist (MicroChem) was spun at 500 rmp for 10 s and then at 1500 rmp for 20 s to achieve a thickness of 60 μm. PDMS was produced by mixing 40 g of A and 4 g of B and poured onto the mold in a petri dish. PDMS was degassed using a vacuum pump at 60 mTorr for 1 h, and baked at 80 ℃ for 1 h to cure. Cured PDMS was peeled off from the mold and inlet/outlet holes were punched. The PDMS was finally bonded to a clean glass slide after air plasma treatment (PDC-32G, Harrick Plasma) of both surfaces, and baked at 80 ℃ for 1 h to achieve strong bonding.

Setup of the microfluidic device

The common outlet of the microfluidic chamber was connected with an infusion/withdrawal syringe pump (OEM Pump Half Case Module, Chemyx Inc) via perfluoroalkoxy alkane (PFA) tubing (1622L, ID: 0.02 in. and OD: 0.0625 in., IDEX Health & Science). A LabVIEW (LabVIEW 2019, National Instruments) program was used to change the flow rate and control the on/off of the syringe pump. Small-volume (< 3 μl) solutions, including chromatin and tn5, were directly added into inlet reservoirs. Large-volume solutions, including bead suspension, tagmentation buffer and washing buffer, were loaded into pipette tips attached to inlet reservoirs (as shown in Figure 4.1c). Solutions were slowly drawn into all the chambers simultaneously by withdrawal of the syringe pump at 1.5 μl/min for small-volume solutions and 10 μl/min for large-volume solutions at the outlet. A cold pack and a hot plate (TC-124, Warner) were utilized to maintain the temperature of 4 ℃ and 37 ℃ for certain steps, respectively.

Cell culture

GM12878 was obtained from Coriell Institute for Medical Research. Cells were cultured in RPMI 1640 media (30-2001, ATCC) supplemented with 15% FBS (16000044, Life Technologies Corporation), 10,000 U/mL Penicillin-Streptomycin (15140122, Life Technologies Corporation) in an incubator maintaining 5% CO_2 and 37 ℃. Mycoplasma contamination was tested every 6 months using Universal Mycoplasma Detection Kit (30-1012K, ATCC).

Cell lysis and MNase digestion

Cells (20-1000) were suspended in 1 μl PBS (14190136, Life Technologies Corporation) and mixed with 1 μl lysis buffer [4% Triton X-100, 100 mM Tris (pH 7.5), 100 mM NaCl, and 30 mM $MgCl_2$] containing freshly added 1% PIC (P8340-5ML, Sigma-Aldrich) and 1 mM PMSF (P7626-1G, Sigma-Aldrich). After 10-min incubation at room temperature, 0.2 μl of 500 mM

$CaCl_2$ and 0.2 μl of 100U/μl MNase (88216, Life Technologies Corporation) were added followed by another 10 min incubation. The reaction was stopped by adding 0.222 μl of 0.5M EDTA and incubated on ice for 10 min. Chromatin solution was stored on ice before ChIP assay.

Preparation of magnetic beads

372 μg Dynabeads Protein A (10001D, Life Technologies Corporation) (contained in 12.4 μl of bead suspension) were washed twice with 200 μl IP buffer (20 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% (w/v) sodium deoxycholate, 0.1% SDS, 1% (v/v) Triton X-100, 0.1% (v/v) Tween 20). Beads were suspended in 600 μl of IP buffer and mixed with 2 μg of antibody (H3K4me3: ab8580, Abcam; H3K4me1: 39498, Active Motif) at 4 ℃ on a rotator mixer at 24 r.p.m. for 1 h. After coating, beads were washed twice with 200 μl IP buffer and suspended in 120 μl IP buffer. These functionalized IP beads were used in one run using the 8-unit device.

Assembly of indexed transposome complexes

We followed a previous protocol for assembly of indexed transposome complexes(*244*). Briefly, to prepare the T5/T7 transposon stock solution, 200 µl single-strand T5 or T7 transposon (IDT, standard desalting) at a concentration of 100 µM was annealed with 200 µl of 100 µM pMENTS, a 5'-phosphorylated 19-bp mosaic end complementary oligonucleotide, in the annealing buffer (10 mM Tris-HCl, 1 mM EDTA, 25 mM NaCl, pH 8.0). 95 ℃ for 5 min followed by a slow ramp to 25 ℃ at -0.1℃/sec was used for the annealing. T5 and T7 transposons have sequences that are complementary to P5 and P7 oligos in the Illumina sequencing platform, respectively, while T7 transposon contains an 8-bp index that is unique for each of 8 units/assays conducted in our mu-CM platform. 2 µl of EZ-Tn5 (TNP92110, Lucigen) was diluted with 8 µl of dilution buffer (50% glycerol, 50 mM Tris, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.1% NP-40) to 600 nM. 5 µl of the diluted tn5 was mixed

with 5 µl of 600 nM annealed T5 or T7 transposon (created from the stock solutions by dilution) and incubated at 37 ℃ for 1 h. Assembled T5 and T7 transposomes (10 µl each) were mixed finally to form 20 µl dimers and were stored at -20 ℃ till use. 8 µl was used in each experiment using the 8-unit device (with 1 µl added to each unit).

Microfluidic ChIPmentation

The microfluidic chambers were primed with IP buffer to prevent adsorption of protein on PDMS. A long permanent magnet (BZ082, 3" × 1/2" × 1/8", K&J Magnetics) was attached to the glass side of the device using double-sided tape and antibody-coated IP beads were then loaded into the chambers under magnetic force and formed a fluidized bead bed. 2.5 μl of Mnase-digested chromatin was directly added into the inlet reservoirs. The chromatin solution flowed through the bead beds in the chambers by withdrawal of the syringe pump at a flow rate of 1.5 μl/min at the outlet. The ChIP step took around 40 min till the reservoirs were essentially dry. Washing buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% (v/v) Triton X-100, 0.1% (v/v) Tween 20) was flowed through the bead beds at a withdrawal flow rate of 10 μl/min for 5 min.

After washing, 10 mM Tris and Tagmentation buffer (10mM Tris, pH 8.0, 5 mM $MgCl_2$, 35% (v/v) dimethylformamide) were then sequentially loaded for conditioning. 1 μl fully assembled indexed tn5 was added directly to the inlet reservoir and flowed through beads bed at an outlet withdrawal rate of 1.5 μl/min for 20 min. The device was put on a hot plate (TC-124, Warner) to reach 37 ℃. RIPA LS buffer (10 mM Tris, pH 8.0, 140 mM NaCl, 1mM EDTA, 0.1% SDS, 0.1% DOC, 1% Triton x-100) was loaded at 10 μl/min outlet withdrawal rate for 5 min to stop tagmentation and the device was placed on a cold pack. Finally, the beads were flowed out of the 8 chambers under a flow rate of 100 μl/min and collected into one Eppendorf tube.

Purification of ChIP DNA

Beads were suspended in a mixture of 198 μl Elution buffer (10mM Tris, pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.4% SDS) and 2 μl of 20 mg/mL Proteinase K (P2308-100MG, Sigma-Aldrich), and incubated at 65 ℃ for 1 h. DNA was purified by phenol extraction and ethanol precipitation, and eluted in IDTE buffer (10 mM Tris, pH 8.0, 0.1 mM EDTA).

PCR Amplification

20 μl ChIP DNA was mixed with 1.5 μl of P5 and P7 PCR primers (each at 25 μM) (IDT, standard desalting), 25 μl activated polymerase (KK2601, Kapa Biosystems) (Pre-heated at 98 ℃ for 30 sec), and 2.5 μl Evagreen (31000-T, Biotium). The qPCR was performed with the following program: 72 ℃ for 5 min; 98 ℃ for 30 s; n cycles of 98 ℃ for 10 s, 63 ℃ for 30 s; and 72 ℃ for 30 s; and a final extension at 72 ℃ for 1 min(*243*). The amplification was stopped once the increase in the relative fluorescence units (RFUs) reached ~500. The library was purified using 0.85× SPRIbeads (B23317, Beckman Coulter) and eluted in 7 μl IDTE buffer.

Data analysis

Sequencing reads were trimmed by Trim Galore and aligned to hg19 genome using Bowtie2(*245*) with default setting. Duplicates, low-quality mapped reads, and reads in blacklist regions were removed before calling peaks with MACS2(*246*) using the following parameters: --nomodel --extsize 100 -q 0.0001. Pearson correlation was calculated using deepTools(*247*) with default parameters.

4.3 Discussion

A multiplexed microfluidic device combing indexed ChIPmentation

We designed a very simple multiplexed microfluidic device made of PDMS/glass slide that has no valving system. Our device consists of 8 bell-shaped reaction chambers (each with a volume of ~ 180 nl) with each having an individual inlet and connected to a common outlet through a series of splits (Figure 4.1). The operation only required an infusion/withdrawal

syringe pump for liquid manipulation at the common outlet and avoided complex pneumatic control and valving system. The operation associated with each individual unit involved several steps: (i) Cells were lysed and chromatin was fragmentized with MNase in a tube with each tube containing 2.5 μl at the end of the step; (ii) A pipette tip containing 15 μl of antibody-coated magnetic beads was inserted into the inlet reservoir. The beads were pulled into the microfluidic chamber under the magnetic force generated by a long permanent magnet adhered to the glass slide side of the device using double-sided tape (Figure 4.1c i). After the beads formed a fluidized bed in the chamber(*46*), the pipette tip was removed from the inlet and the bulk of the solution was pipetted out with roughly 2 μl solution left in the reservoir; (iii) 2.5 μl chromatin solution was added into each inlet reservoir. The chromatin solution was pulled through the bead bed by a syringe pump connected to the common outlet that withdrew at a flow rate of 1.5 μl/min for 20-40 min until the reservoirs were mostly empty. In this step, ChIP occurred during which the chromatin fragments that contained targeted histone modification bound to its antibody on the IP bead surface. The small volume of the chamber and the small diffusion length within the bead bed facilitate efficient immunoprecipitation of target chromatin fragments; (iv) Washing was then conducted by plugging a pipette tip containing 150 μl washing buffer into the inlet reservoir and withdrawing solutions from the common outlet at a flow rate of 10 μl/min for 5 min (Figure 4.1c ii). The washing removed nonspecific binding on the bead surface. (v) Next, the pipette tip containing the leftover washing buffer was then replaced by one containing 150 μl tagmentation buffer and the withdrawing was performed again (10 μl/min for 5 min). The pipette tip was then removed together with the remaining buffer inside. (vi) 1 μl of indexed tn5 transposase complexes (unique for each unit and sample) was loaded into the reservoir and then the entire device (including the thin magnet) was carefully placed on a hot plate that was set at 37 °C. The transposase complex solution was pulled through the bead bed by a syringe pump connected to the common outlet that withdrew

at a flow rate of 1.5 μl/min until the reservoirs were mostly empty. The process took roughly 10-20 min (the tagmentation requires at least 10 min). During this step, Tn5 transposase complex cut and tagged the chromatin fragments on the bead surface with adapters that contained an index that was unique for an individual unit). After this step, the ChIPed chromatin fragments in each chamber/unit were tagged by unique indexes. (vii) After tagmentation, the magnet was removed and the magnetic beads were flushed out of the chambers and pooled together for DNA purification and PCR amplification to prepare one library from the pooled samples. The library was sequenced at a depth of 15 million total reads with single-ended 50 bp.

Profiling histone modification from ultralow number of cells

Using our method, we profiled two histone marks (H3K4me3 and H3K4me1) with various number of GM12878 cells (1000 cells per unit for H3K4me1, 100/40/20 cells per unit for H3K4me3) (Supplementary Table S2). Our mu-CM data show high correlation among units, with an average Pearson's correlation of 0.90 for H3K4me1 with 1,000 cells per unit, 0.89, 0.67 for H3K4me3 with 100 and 40 cells per unit, respectively (Figure 4.4a and Figure 4.4b). The average Pearson's correlation with ENCODE data sets were 0.75 and 0.70 for H3K4me3 with 100 and 40 cells, and 0.82 for H3K4me1 with 1000 cells, respectively (Figure 4.4a and Figure 4.4b). These correlations were generally lower than what we obtained using other low-input methods that did not involve ChIPmentation. For example, the Pearson's correlation between MOWChIP-seq 100 cell data set and the ENCODE data was 0.82. All data showed high fraction of reads in peak (FRiP) (in the range of 8% to 28%) that is comparable to that of ENCODE data.

Our indexing using ChIPmentation provided an approach to examine a very low number of cells. In our previous works, we needed at least 30-100 cells to start with in order to generate enough ChIP DNA to prepare a sequencing library(*44, 133*). With the tagmentation-based indexing, we produced enough ChIP DNA from 8 samples of 20 cells each that were indexed differently. By demultiplexing the sequencing data based on the indexes, we were able to produce ChIP-seq data using as few as 20 cells (Figure 4.5). Among the 8 samples processed in one run in our device, only sample 8 presented fairly low quality (having an average correlation of 0.57 with the other 7 samples, Figure 4.5b). The other 7 data sets were strongly correlated (average correlation ~ 0.75). For the 7 good data sets, each has an average of 3,886 peaks, and 500,000 unique reads. The decrease in the peak number and the correlation among data sets was likely due to the fact that 20-cell samples lost even larger fraction of their chromatin fragments to surface adsorption than samples containing higher cell numbers. It is also worth noting that there is a small probability for the assay to fail (1/8 in the case of Figure 4.5a) with 20 cells per sample. Nevertheless, the results shw that our mu-CM technology allows probing cell samples that do not provide enough ChIP DNA for library preparation by itself.

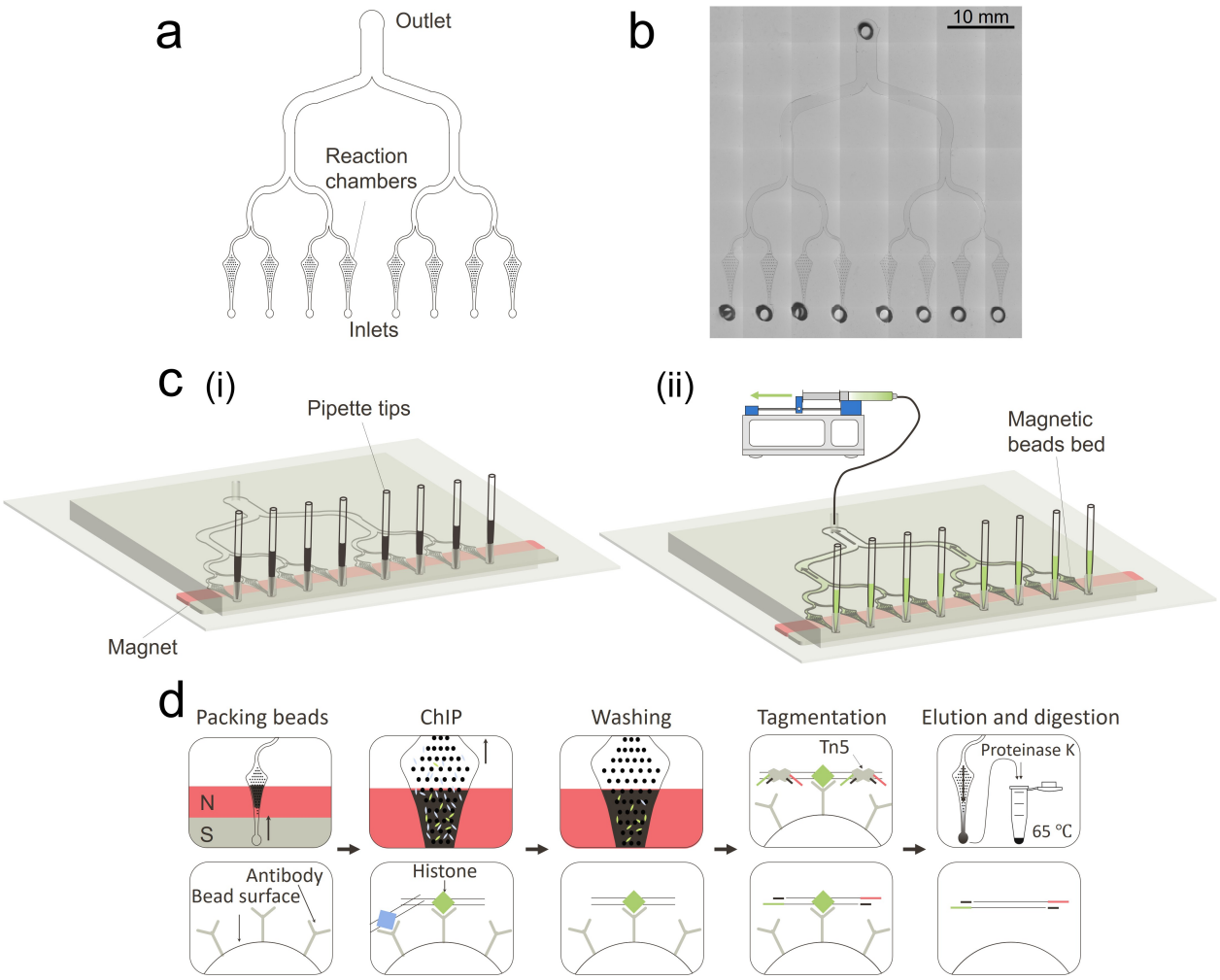


Figure 4.1 Overview of microfluidic ChIPmentation device and operation.
a. Schematic of the microfluidic device. The device contains 8 reaction chambers. Each chamber has an individual inlet, which is connected to a common outlet through a series of splits. Each chamber has supporting pillars to prevent collapse. **b.** Microscopic image of the microfluidic device. The stitched image is created by OLYMPUS CellSens. **c.** The device operation. Two major steps are shown: (i) Packing of magnetic IP beads. Pipette tips containing magnetic bead suspensions are inserted into the inlet reservoirs. A long magnet is placed under the device to retain the beads. Magnetic beads are drawn into the chambers under magnetic force and form a fluidized bead bed; (ii) Solutions are loaded into the chambers by withdrawal using a syringe pump at the outlet. Solutions slowly flow through the bead bed. **d.** Flow chart of the steps and corresponding molecular biology.

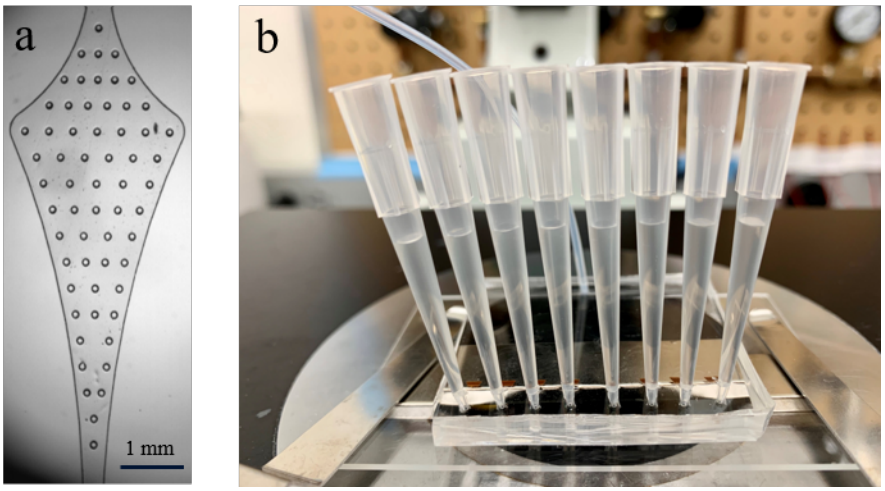


Figure 4.2 Bell-shape reaction chambers and the setup of the device.
a, the bell-shape created velocity gradient that generated enough shear force to efficiently remove non-specific binding. The chamber included micropillar with a diameter of 100 μm. **b**, Setup of the device: 200 μl pipette tips containing reagents were inserted into the inlet reservoir of each reaction chamber.

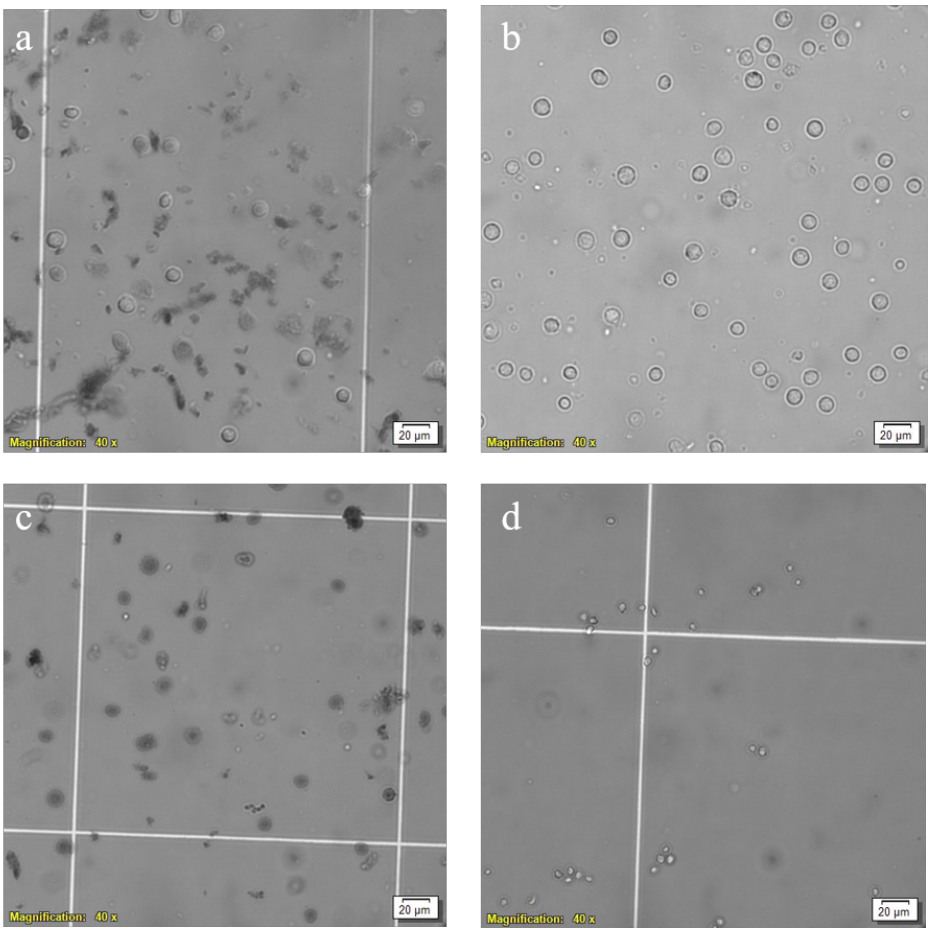


Figure 4.3 Microscopic images of GM129878 cells before and after lysis.
A, before cell lysis with trypan blue; b, before cell lysis without trypan blue; c, after cell lysis with trypan blue; d, after cell lysis without trypan blue.

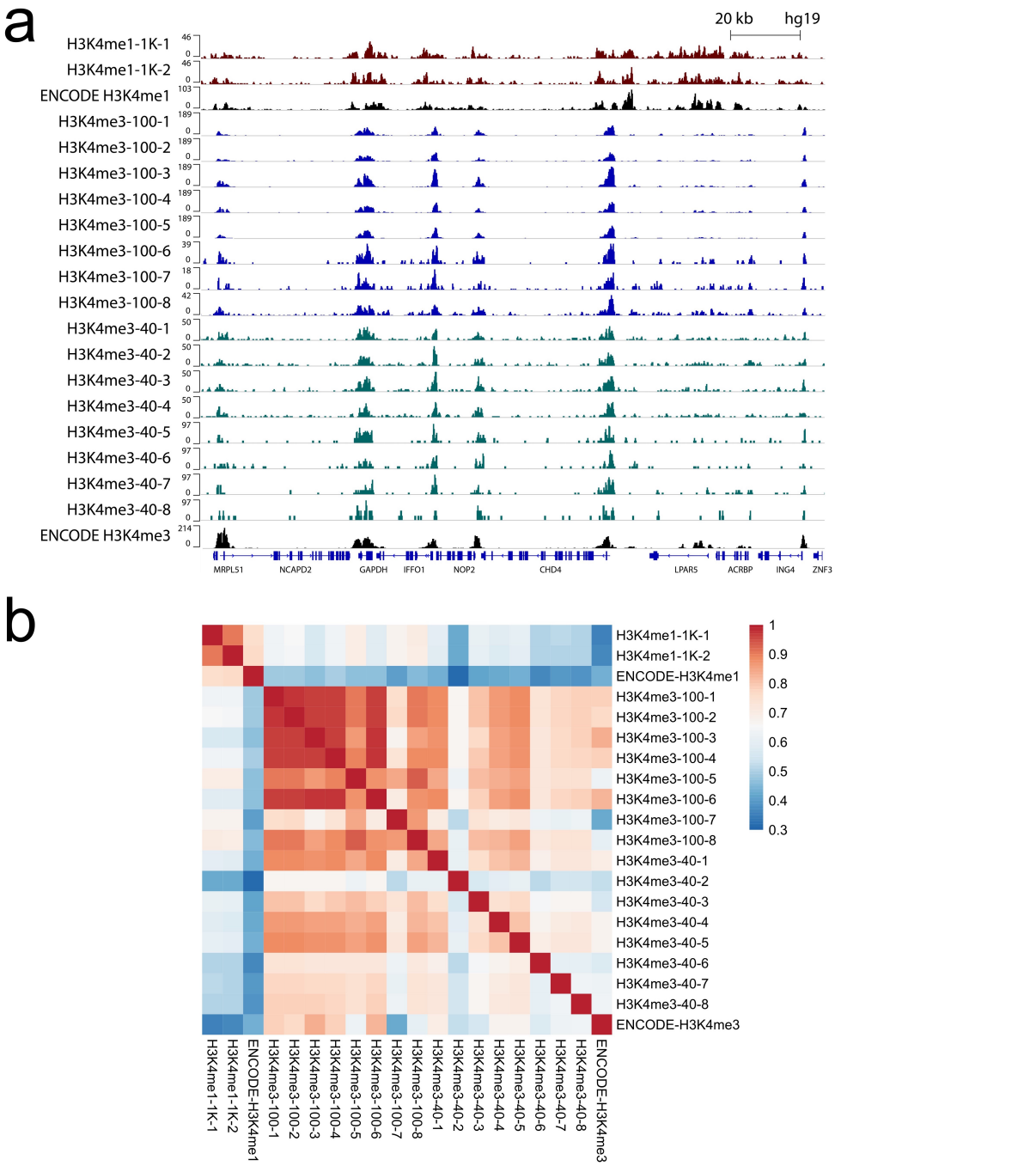


Figure 4.4 Mu-CM data using 1000, 100, and 40 GM12878 cells per unit.
a. Normalized H3K4me3 and H3K4me1 signals at house-keeping gene *GAPDH* using data generated with various sample sizes. ENCODE H3K4me3 data set (GSM733708) and H3K4me1 data set (GSM733772) are used for comparison. **b.** Pearson's correlations calculated using normalized signal in promoter regions among our ChIP-seq data sets (1k H3K4me1, 100/40 cells H3K4me3) and ENCODE data. Promoter regions are defined as 2kb upstream and downstream from the transcription start sites.

Table 4.1 Summary of mu-CM data.

Sample	Raw Reads (Million)	Alignment Rate, %	Number of Peaks	Redundancy Rate, %
H3K4me3-100-1	8.5	96.61	20,476	42.67
H3K4me3-100-2	8.7	96.35	20,235	44.98
H3K4me3-100-3	11.9	97.05	19,502	60.45
H3K4me3-100-4	8.8	96.69	18,705	52.4
H3K4me3-100-5	6	96.74	17,925	21.23
H3K4me3-100-6	14	97.28	22,522	46.33
H3K4me3-100-7	10.1	97.23	7,501	25.38
H3K4me3-100-8	10.7	96.95	20,698	34.93
H3K4me3-40-1	9.9	97.07	3,616	86.04
H3K4me3-40-2	4.1	95.64	965	90.17
H3K4me3-40-3	10.7	96.95	5,631	84.54
H3K4me3-40-4	7.3	96.99	7,913	80.97
H3K4me3-40-5	7	96.87	7,504	78.57
H3K4me3-40-6	6.1	97.25	1,498	83.39
H3K4me3-40-7	4.2	97.18	1,757	82.6
H3K4me3-40-8	5.6	97.03	2,445	86.1
H3K4me3-20-1	1.9	96.83	3,933	89.58
H3K4me3-20-2	3.5	97.13	4,437	90.41
H3K4me3-20-3	8.5	97.12	3,871	91.05
H3K4me3-20-4	4.7	96.11	4,384	90.74
H3K4me3-20-5	5.8	97.33	3,623	89.94
H3K4me3-20-6	6.5	97.12	2,397	90.6
H3K4me3-20-7	8.7	97.28	4,555	91.21
H3K4me3-20-8	4.4	97.37	611	88.5
H3K4me1-1K-1	7.6	97.94	25,355	57.52
H3K4me1-1K-2	7.8	97.87	24,402	59.08

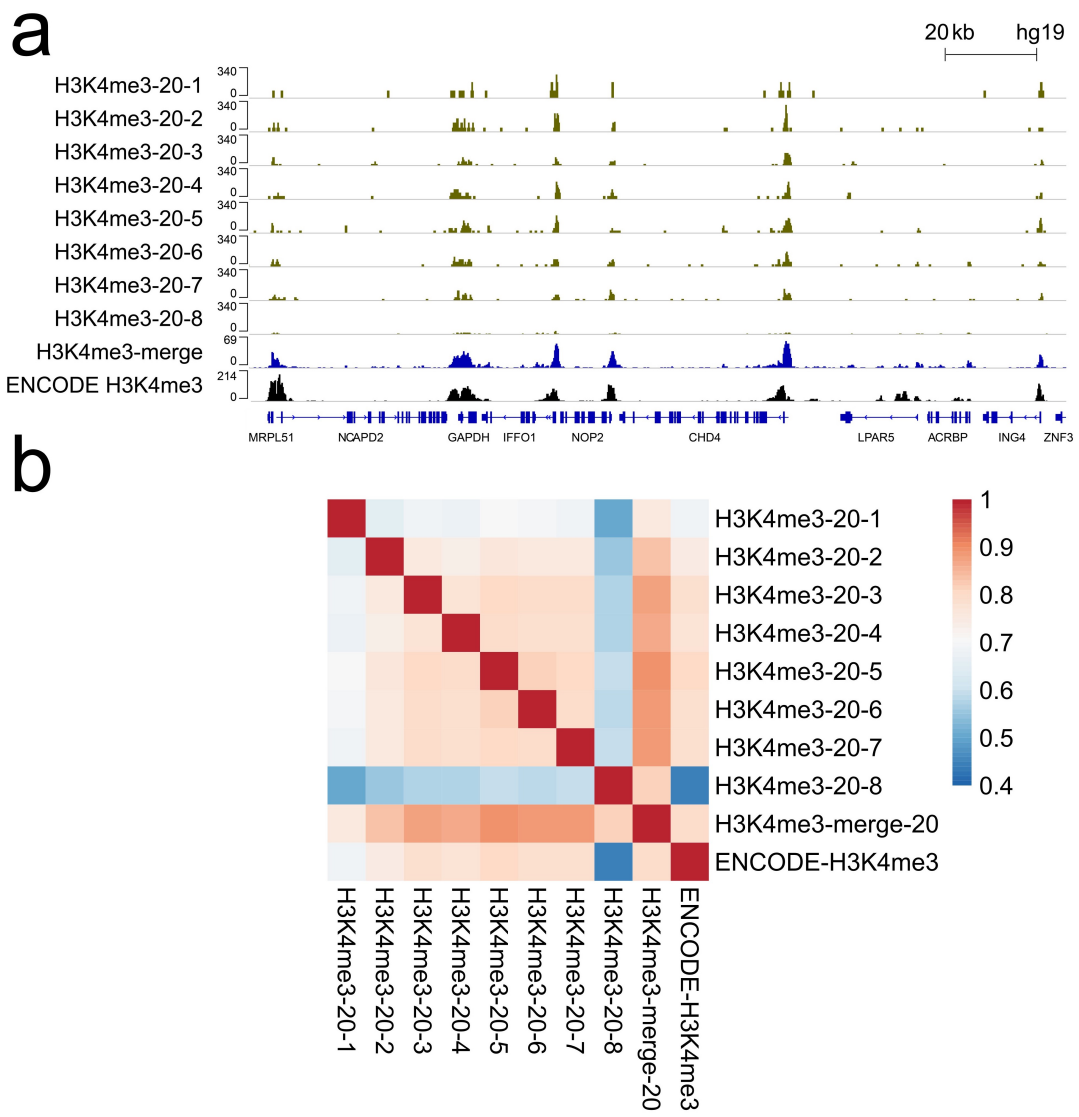


Figure 4.5 Mu-CM data using 20 cells per unit.
a. Normalized H3K4me3 signals at house-keeping gene *GAPDH*. 8 replicates of 20 cell data were also merged together and labeled as 'H3K4me3-merge'. Merged data was in excellent agreement with ENCODE data. **b.** Pearson correlations in promoter regions among the data sets and ENCODE data.

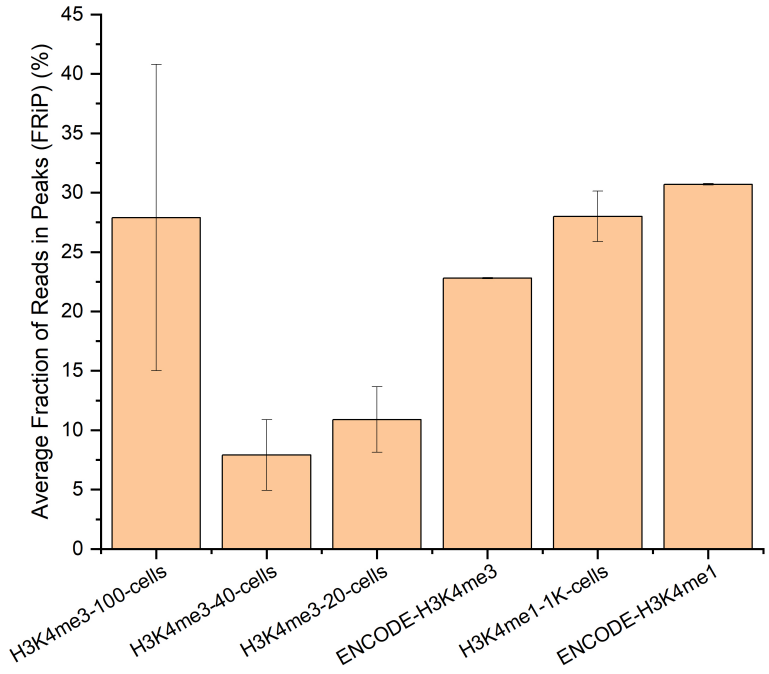


Figure 4.6 Average fraction of reads in peak (FRiP) calculated using featureCounts(*248*)

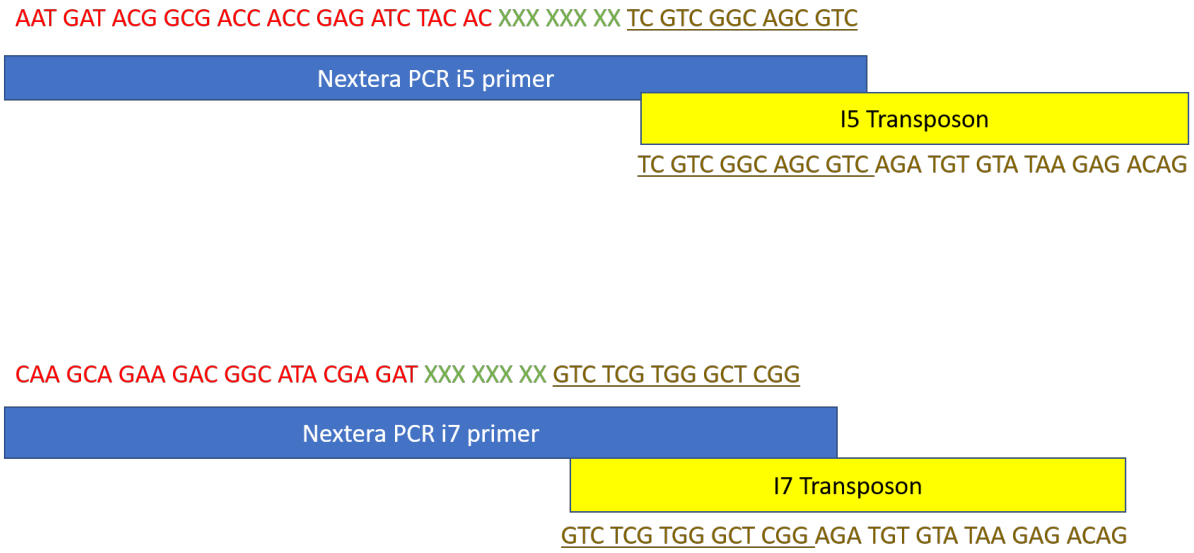


Figure 4.7 Design of transposon in Nextera DNA library prep kit.
i5 and i7 barcode were introduced during PCR amplification. The kit provides pre-assembled tn5 transposome complex.

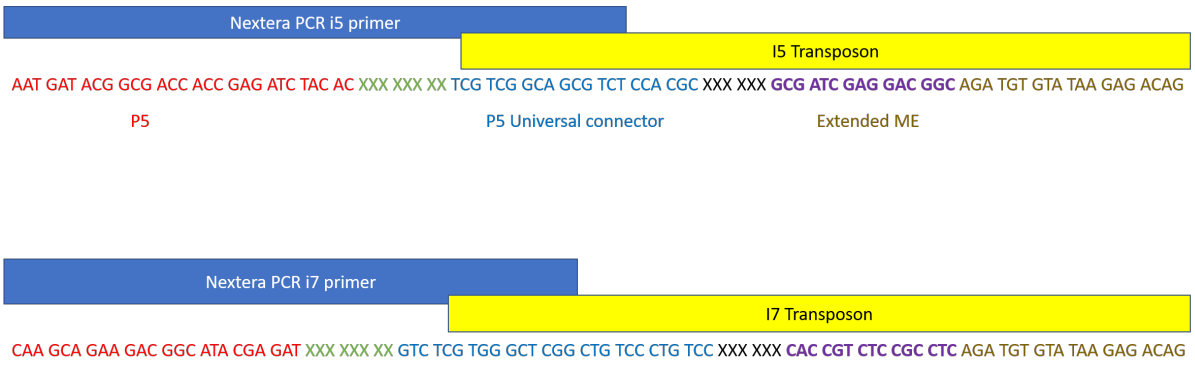


Figure 4.8 Design of full-length transposon used in combinatorial indexing.
The combinatorial indexing was performed by introducing one round of barcode during tagmentation, and another round of barcode during PCR amplification(*244*). In this way, a large number of cells can be uniquely barcoded with a small number of 1st and 2nd barcodes. Additional index reading primer will be needed during sequencing. This strategy had been widely used in single cell sequencing(*117*).

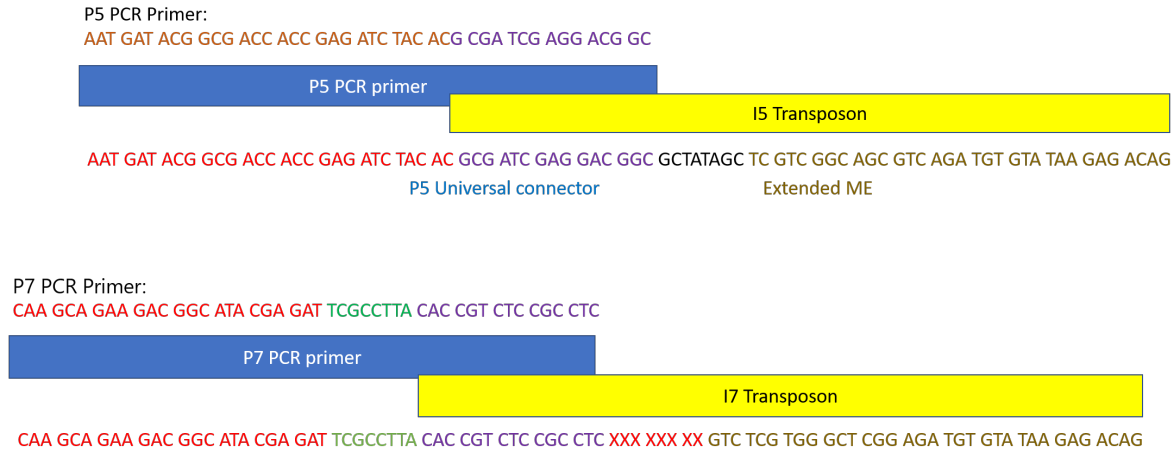


Figure 4.9 First version of transposon for home-made tn5 transposase.
The design of transposon was adopted from the transposon designed for combinatorial indexing. We only keep barcode in i7 transposon as we only need single barcode for each reaction. However, the corresponding PCR primer are prone to form dimmer during amplification, which inhibited the amplification of libraries.

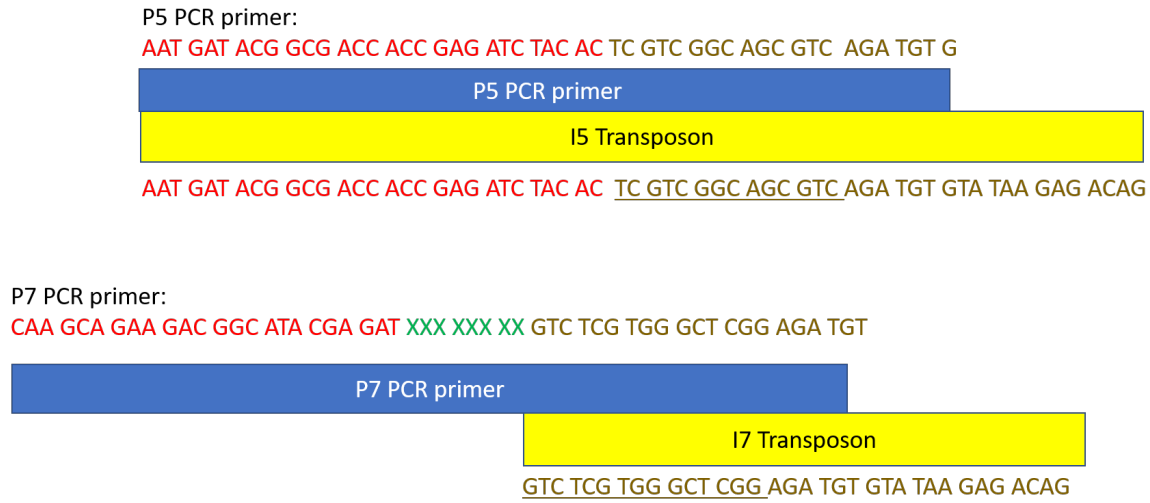


Figure 4.10 Second version of transposon for home-made tn5 transposase.
In this version, unnecessary universal connectors were removed. Barcode was added during amplification, which didn't allow us to pool samples right after tagmentation.

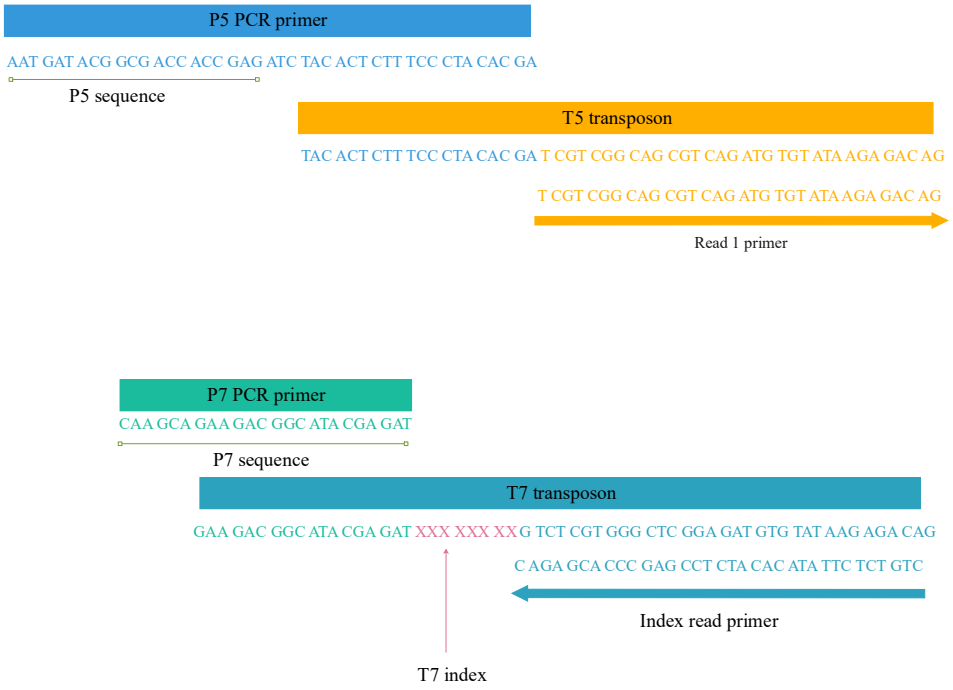


Figure 4.11 Final design of T5/T7 transposons and PCR primers.
T5 universal transposon has sequence for illumina sequencing read 1 primer binding. T7 transposon contains index primer binding site and 8-bp T7 index sequence. P5/P7 flow cell oligo binding sequence are introduced during PCR amplification (Sequence of oligonucleotides is listed on Table S4). We use Truseq PCR amplify primer, which didn't give rise to primer dimmers. Barcode was added during tagmentation, enabling barcoding and pooling.

4.4 Conclusion

Here we demonstrate that mu-CM is an ultralow-input microfluidic platform for profiling histone modifications using as few as 20 cells. Our simple and multiplexed device requires minimal ancillary control system, permitting easy implementation in most biology labs. It greatly saves labor and can be performed by novices with little experience on microfluidics. Moreover, integrating ChIPmentation-based indexing enables pooling of ChIP DNA from multiple samples for library preparation and dramatically reduces the needed input cell number for each sample. Taken together, mu-CM is suitable for studies based on animal models or patients that generate low quantity of tissues.

Chapter 5 - Spatially Resolved Epigenome and Transcriptome in Adult Mouse Neocortex

5.1 Introduction

In mammalian brain, neocortex is the part essential for higher functions such as sensory perception, motor and cognitive behaviors. Advancing our understanding of its structure and function requires identification of novel regulatory elements. H3K27me3 is a repressive histone mark and involves in cell type-specific gene silencing. Here, we applied a low-input epigenomic profiling method developed in our lab called MOWChIP-seq and generated spatially resolved H3K27me3 landscapes in adult mouse neocortex. We identified distinct spatial patterns of H3K27me3 signals that are associated particularly with central nervous system development. Our data also revealed profound different spatial patterns between neuron and glia.

Neurons and glia are two major types of cells in the nervous system, though they are differentiated from the common progenitor cells including radial glial cells and polydendrocytes. Neurons store, transmit and analyze information while glia provides mechanical and metabolic support for neurons. Neurons can be divided into three major categories: receptors, interneurons, and motor neurons. Receptors receive and process information, such as sensation and perception. Interneurons receive and send signals between cells and constitute the bulk of the human nervous system. Motor neurons send signals to the muscles and glands of the body. While glia can be classified intro microglia that has immune functions, and macroglia (astrocyte, oligodendrocytes, polydendrocytes, and perineuronal cells) that supports the function of neurons. In the past century, scientists have spent great efforts in establishing neuron classification system. Recent technology advancement in single-cell sequencing has paved new avenues for identifying neuron cell types. However, there are unresolved differences between taxonomy based on epigenome and transcriptome. Hence, the

defined cell types could include several linkages with various properties, such as a specific gene expression or open chromatin region.

Neocortex is a set of layers of the mammalian cerebral cortex. It is the evolutionary youngest part of the brain and is responsible for sophisticated control, analysis and integration of information such as sensory or motor activity. Adult mouse cortex consists of around 14 million neurons and 12 million glia(*249*). Neocortex is one of the most complex structure in mammalian brain containing a variety of neuron subtypes with different shapes and spatial arrangements. The neocortex is functionally divided into different regions, which is a result of genomic and epigenomic regulation. In neocortex, there are six layers including L1, L2/3, L4, L5, L6, L7(L6b). Neocortex is dominated by principal neurons (62-85%), which are mostly excitatory neurons using glutamate as a neurontransmitter and have long axons transmitting information over long distances from one region to another(*250, 251*). All layers except for layer 1 have a much higher percent of principal neurons than interneurons. However, interneurons have a much greater variety in cortical neuron types compared with principal neurons.

Tri-methylation of histone 3 at lysine 27 (H3K27me3) is known as a repressive histone mark and is considered to involve in cell-type-specific gene silencing during development. H3K27me3 is deposited and maintained by EZH2 (enhancer of zeste homolog 2) subunit of the Polycomb Repressive Complex 2 (PRC2). Studies have found H3K27me3 is a characteristic marker of silencers, one type of *cis* regulatory elements that binds transcription factors called repressor and inhibits transcription of a gene. In embryonic stem (ES) cells, H3K27me3 often co-occurs with tri-methylation of histone 3 at lysine 4 (H3K4me3) at the 5' ends of genes within CpG islands, known as bivalent domains. Bivalent domains play important role in cell differentiation. Bivalent domains are used to poise the expression of key developmental genes (genes encoding developmental regulators) for lineage-specific activation

or repression during differentiation(*252*). Upon cell lineage specification, the developmental genes that need activation will retain H3K4me3 and remove H3K27me3 while other developmental genes not required for that cell lineage will lose H3K4me3 and maintain inhibited by H3K27me3. Recently with technological advances in epigenome editing tools, histone modification can be either removed or deposited at target region depending on effector fused on cas9 protein. Several proof-of-concept studies have demonstrated editing of H3K27me3 resulted in changes of targeted gene transcription, which opened avenues to study the causal relationship between histone modification and phenotypes(*253*). One of the reasons we decide to look at H3K27me3 is that previous data from our lab has demonstrated that compared to other commonly studied histone mark, including H3K4me3 and H3K27ac, H3K27me3 showed the greatest tissue-type-specific variations between prefrontal cortex and cerebellum of mouse brain(*133*).

Neocortex is functionally divided into many regions under the regulation of different gene expression and epigenome. Spatially restricted gene expression and relevant epigenomic modifications are fundamental during neuron axonogenesis and neuron fate commitment. Identifying regulatory elements or genetic regulators that work in a defined region of the neocortex is important to understand the process of central nervous system development. Microscopy based imaging method such as mRNA in situ hybridization are widely used to study the spatially restricted gene expression patterns in various model organisms and tissues. However, those methods are locus-specific and only investigate a few genes per sample. A few strategies have been developed to examine the spatial patterns of the entire transcriptome. The power of multiplexing can be increased by taking advantage of super-resolution microscopy(*254*). Alternatively, the spectral barcoding can be replaced by sequential barcoding(*255*). Besides, *in situ* sequencing combines FISH with SOLiD sequencing and is able to generate RNA-seq libraries and fluorescent tissue images simultaneously(*256*).

Similarly, slide-seq was reported that utilized DNA-barcoded beads and single-cell RNA sequencing to identify the location and cell type of each transcript(*257*). Besides, a commercial spatial transcriptome platform was introduced by 10x Genomics and has been applied to study the progression of various diseases(*258-260*). Previously "RNA Tomography" was reported that performed low-input RNA-seq on cryosectioned slices of zebrafish embryo in three body axis and generated spatially resolved transcriptomics in 3D using mathematical image reconstruction(*261*). "RNA tomography" benefited from the technology advancement in low-input RNA sequencing. As a result, high quality transcriptome was profiled from scarce RNA extracted from thin tissue slices. Similarly, using the low-input ChIP-seq method developed in our lab(*44*), we were able to generate spatially resolved H3K27me3 landscapes in micro-dissected mouse neocortex.

Here we applied low-input methods including MOWChIP-seq and SMART-seq2(*262*) to coronal slices of adult mouse neocortex and generated spatially-resolved and cell-type-specific epigenomic/transcriptomic profiles of neuronal/glial cells. In our approach, we sliced the mouse brain into coronal sections, isolated left and right neocortex from each section, extract NeuN+ (neuronal) and NeuN- (glial) factions from each dissected tissue, and generated ChIP-seq libraries from 4,000 nuclei per assay. We observed distinct spatially restricted H3K27me3 signals in neurons while glia had relatively uniform signals across the entire neocortex. We identified H3K27me3 peaks with similar spatial patterns using K-means clustering and found peaks with spatial variations tend to involve in functions such as central nervous system development. We compared clustering results using H3K27me3 signals and gene expression level and observed substantial differences between H3K27me3 landscape and transcriptome. Our data provides a comprehensive spatial map of H3K27me3 in adult mouse neocortex and serves as a powerful resource to identify target regions marked by H3K27me3 with a particular spatial pattern.

5.2 Methods and materials

Fabrication of microfluidic devices

Microfluidic devices were fabricated using Polydimethylsiloxane (PDMS) by multi-layer soft lithography. LayoutEditor (juspertor GmbH) was used to draw pattern in micrometer scale and the designed pattern was printed out on high resolution transparencies (10, 000 dots per inch) (Fineline imaging Inc). Master molds were fabricated using photolithography. ~3 g of negative photoresist SU8-2025 was poured onto pre-cleaned 76-mm silicon wafer (University Wafers), and the wafer was spun using the following program to obtain an even coating of photoresist with a thickness of 60 um (fluidic master): 500 r.p.m. for 10 s followed by 2,500 r.p.m for 30 s. To obtain a thickness of 40 um (control master), the following program was used: 500 r.p.m. for 10 s, followed by 2,500 r.p.m. for 30 s. PDMS prepolymer (RTV615, Momentive) for fluidic layer was mix with a mass ratio of base/crosslinker = 40g:4g. Mixed PDMS was poured onto the master in a petri dish to yield ~ 6 mm thickness. For control layer, 5g of base was mixed with 0.25 g of crosslinker and PDMS was spin coated onto control master using the two-step program: 500 r.p.m. for 10 s and 1,100 r.p.m. for 30 s. The petri dish containing the fluidic master and PDMS was vacuumed for 1 h using pump at ~60 mTorr. PDMS was pre-baked at 80 ℃ for 15min. The fluidic PDMS was then peeled off and put on top of control PDMS layer. Align marks were added during device design to enable accurate alignment of two layers. Aligned PDMS was post baked for 1 h to strength the thermal bonding. PDMS was then peeled off and was punched for 1.5mm inlet and outlet holes. The PDMS structure was bonded to a clean glass slide by plasma binding (PDC-32G, Harrick Plasma). The bonded device was baked at 80 ℃ to increase binding strength.

Mouse strain and brain dissection

Eight-week-old C57BL/6J male mice were obtained from Envigo and were sacrificed after two weeks of accommodation in the animal facility with 12-hour cycles of light/dark and

food/water. Compressed CO_2 following cervical dislocation and decapitation were used to sacrifice the mice. Mouse brain was isolated and sectioned in the coronal plane at an interval of 1 mm on an adult mouse brain slicer matrix (BSMAS005-1, Zivic). White matter was discard and only grey matter was kept. Left and right neocortex were micro-dissected from each slice and immediately frozen on dry ice. The dissected tissues were stored at -80 ℃ before used for nuclei extraction. The Institutional Animal Care and Use Committee (IACUC) at Virginia Tech approved this study, and IACUC guidelines were carefully followed.

Nuclei extraction

Nuclei extraction from dissected brain tissue was performed using previously described protocol(*263*) with minor modifications. All steps were conducted on ice. The centrifuge was pre-cool to 4 ℃. For buffer used in each step, 1% PIC, 0.1 mM PMSF and 1mM dithiothreitol (DTT) were added in fresh. Brain tissue was thawed on ice and transferred to a glass tissue homogenizer (Sigma, d9036-1SET) using 4 ml of ice-cold nuclei extraction buffer [0.32 M sucrose, 5mM $CaCl_2$, 2 mM $Mg(Ac)_2$, 0.1 mM EDTA, 10 mM tris-HCl, and 0.1% Triton X-100, freshly added 1% PIC, 0.1 mM PMSF, 1 mM DTT, and 0.06 U/μl ribonuclease (RNase) (Thermo Fisher, EN0531)if mRNA extraction will be performed]. The tissue was homogenized with 10 times of slow douncing a loose pestle and 15 times douncing using a tight pestle (D9063, Sigma-Aldrich). A 40-μm cell strainer (22363547, Thermo Fisher Scientific) was used to filter out large debris. The sample was then transferred into a 15-ml centrifuge tube and centrifuged at 1000 rcf for 10 min at 4 ℃. Supernatant was discarded and the pellet was resuspended using 0.5 ml of ice-cold nuclei extraction buffer. 0.75 ml of 50% iodixanol (D1556, Sigma-Aldrich, diluent was prepared as following: 150 mM KCl, 30 mM $MgCl_2$, and 120 mM tris-HCl, pH 7.8) was added. The mixture was gently mix by inverting the tube up and down 100 times. The solution was centrifuged at 10, 000 rcf for 20 min. Supernatant was carefully removed and the pellet incubated with 10 μl of 2% normal goat serum (50062Z, Life

Technologies) on ice. The pellet was then resuspended and transferred into a new Eppendorf tube with an addition of 250 µl of 2% normal goat serum. 4 µl of 2 ng/µl Anti-NeuN antibody conjugated with Alexa 488 (MAB377X, EMD Millipore) was mixed with the suspension. The reaction was incubated at 4 ℃ for 30 min on a rotator. The nuclei solution was sorted using FACS (BS FACSAria, BD Bioscience) to separate neuronal and glial cells.

Chromatin fragmentation

In order to obtain desirable volume of chromatin solution, the nuclei suspension after sorting was concentrated to have a small volume. To avoid breaking up nuclei during centrifugation, sucrose was used. The nuclei solution (8,000 per samples. Each sample has two technical replicates.) was topped up to 1 ml using Dulbecco's PBS (DPBS; 14190144, Life Technologies). 200 µl of 1.8M Sucrose, 10 µl of 1M CaCl2, and 3 µl of 1M Mg(Ac)2 was added. Samples were mixed gently by inverting the tube up and down and incubated for 15 min on ice. The solution was centrifuged at 3,000 rpm for 15 min at 4 ℃. Supernatant was carefully removed, and the pellet was resuspended with 10 µl DPBS with freshly added PIC and PMSF. 10 µl of 2x lysis buffer [4% Triton X-100, 100 mM tris (pH 7.5), 100 mM NaCl, and 30 mM $MgCl_2$] was added to the sample. The sample was incubated at room temperature for 10 min. The sample was then mixed with 1 µl of 0.1 M $CaCl_2$ and 0.4 µl of 100 U/µl Mnase (88216, Thermo Fisher Scientific) and incubated at room temperature for 10 min. 2.22 µl of 0.5 M EDTA (pH 8) was added to quench the rection. The sample was vortexed to mix and incubated on ice for 10 min. The sample was centrifuged at 16,100 g for 5min. The supernatant was transferred to a new tube and 60 µl of IP buffer [20 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% (w/v)] sodium deoxycholate, 0.1% SDS, 1% (v/v) Triton X-100, 0.1% (v/v) Tween 20) was added to obtain a total volume of ~80 ul. For each replicate experiments, 40 µl of chromatin solution was used. The chromatin was stored on ice before loaded into microfluidic device.

Antibody coating on bead surface

Antibody was conjugated with magnetic beads to facilitate easy separation of enriched target fragments. Beads solution were vortexed thoroughly before use. 4 μl (120 μg of beads) of protein A beads suspension (Thermo Fisher, 10001D) was washed twice with 150 μl of IP buffer, and resuspended with 200 μl IP buffer. The bead solution was mixed with H3K27me3 antibody (0.4 μl per reaction, Active motif, 39155). The solution was incubated on a rotator at 24 r.p.m for 1 h. Antibody-coated beads were gently washed twice with 150 μl of IP buffer and resuspended with 5 μl of IP buffer.

MOWChIP-seq

The on-chip micromechanical valve was connected to compressed air outlet controlled by a solenoid valve (18801003-12v, ASCO Scientific). The on/off switch of the solenoid valves were controlled by a LabVIEW (LabVIEW 2012, National Instruments) program. A pressure of 40 p.s.i was applied to the control channels to close the on-chip valve. Closed valve will stop the flow of solid particles in the fluidic chamber while liquid is allowed to pass through. Solutions were introduced into the microfluidic chamber via perfluoroalkoxy alkane (PFA) high-purity tubing (1622L, ID: 0.02 in. and OD: 0.0625 in., IDEX Health and Science) and a infuse syringe pump. Microfluidic chamber was rinsed with IP buffer to prevent large molecules adsorbed on PDMS surface. Air bubbles inside chambers was squeezed out by close the pneumatic valves while infusing solution into the chamber. 5 μl antibody coated beads were loaded into a microfluidic chamber via magnetic force generated by a magnet. 40 μl of chromatin solution was infused into the chamber via pressure-driven flow provided by the syringe pump at a flow rate of 1.5 μl/min. Beads were packed against the on-chip valve under the driven of magnet. Complete loading the chromatin solution takes 40~50 min. The inlet and outlet were connected to a soft tube containing ~ 15 μl washing buffer and ~ 1 p.s.i intermittent

pressure was applied to the inlet and outlet for 5 min. This makes beads move back and forth inside the chamber and efficiently remove non-specific binding of fragments on beads. Washing was performed with low salt and high salt washing buffer separately [20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.1% SDS, 1% (v/v) Triton X-100, 0.1% (v/v) Tween 20, 500 mM NaCl for low salt buffer, 750 mM NaCl for high salt buffer]. After washing, a 200-µl pipette tip was inserted into the outlet and the inlet is infused with IP buffer at a flow rate of 200 µl/min for 1 min to flash out beads. Beads solution in the pipette tip was transferred into a new Eppendorf tube. The tube was placed on a magnetic rack and the supernatant was removed. 195 µl of elution buffer [10mM Tris, pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.4% SDS] and 5 µl of proteinase K (P2308-100MG, Sigma-Aldrich) was added to resuspend the beads. The sample was incubated at 65 ℃ for 1h. The sample was added 200 µl of phenol: Chloroform: Isoamyl Alcohol 25:24:1 (Sigma, P2069) and vortex vigorously to mix and centrifuged at maximum speed for 5 min. Supernatant was carefully transferred to a new Eppendorf tube without any phenol carryovers. 0.5 µl of glycogen (Millipore, 361507), 100 µl of ammonium acetate and 600 µl pure ice-cold ethanol was added to the sample sequentially and mixed by vortexing. The sample was incubated at -80 ℃ for 45 min and centrifuged at 4 ℃ at full speed for 15 min. Supernatant was removed without disturbing the pellet and 500 µl ice-cold 70% ethanol was added followed by a 5-min full speed centrifuge. Supernatant was removed and the pellet was air dried to allow ethanol residuals fully evaporate. The pellet was eluted with 12 µl of low EDTA TE buffer and stored at -20 ℃. The quality of ChIP DNA was assessed by qPCR. A positive primer (Active motif, 71019) and a negative primer (Active motif, 71011) were used. 0.5 µl ChIP DNA was used for each locus. ChIP enrichment was calculated using the following equation:

$$Relative\ Fold\ Enrichment = \frac{Ct_{input} - Ct_{positive}}{Ct_{input} - Ct_{negative}}$$

ChIP-seq libraries were prepared using Accel-NGS 2S Plus DNA Library kit (Swift Biosciences). 6 μl of ChIP DNA was used for library construction followed the manufacture's instruction with minor modifications. The volume of all reagents was decreased to one fifth of the original volume. 0.5 μl EvaGreen dye (1x, Biotium) was added in order to monitor the amplification in real time. Amplification was stopped when the fluorescence signal intensity increased about 1,000. After amplification, the reaction was cleaned up using 0.75 x SPRIbeads to remove adapter dimmers.

mRNA-seq

The workspace was cleaned by RNaseZap before RNA extraction to avoid sample loss. 50 μl of FACS-sorted nuclei from neocortex slices was mixed with 0.5 μl RNase inhibitor and was used for total RNA extraction by the RNeasy Mini kit (QIAGEN) following the manufacture's instructions. 80 μl Deoxyribonuclease (DNase) treatment was added to remove genomic DNA contamination. mRNA libraries were prepared with off-shelf reagents following the SMART-Seq2 reported previously(*264*) with small modifications. After 14 cycles of reverse transcription, cDNA was purified using x0.75 SPRIbeads (Beckman Coulter). cDNA was either stored at -20 ℃ or directly used for subsequent tagmentation. Home-made tn5 and self-designed transposons were used to assemble tn5 transposome. Tn5 production was followed by protocol described previously(*265*). The sequence of self-designed P5 and P7 transposons were listed in the appendix. Assembling of tn5 transposome was performed following previous described protocol(*244*). Briefly, T5/T7 transposon stock solution was prepared by annealing 200 μl single-strand T5 or T7 transposon (IDT, standard desalting) at a concentration of 100 μM with 200 μl of 100 μM pMENTS, a 5'-phosphorylated 19-bp mosaic end complementary oligonucleotide, in the annealing buffer (10 mM Tris-HCl, 1 mM EDTA, 25 mM NaCl, pH 8.0). The following annealing program was used: 95 ℃ for 5 min followed by a slow ramp to 25 ℃ at -0.1℃/sec. T5 and T7 transposons have sequences that are complementary to P5 and

P7 oligos in the Illumina sequencing platform, respectively, while T7 transposon contains an unique index for sample multiplexing purpose(Appendix Table S4). 20 μl of home-made tn5 was mixed 20 μl of 50 μM annealed T5 or T7 transposon and incubated at 37 ℃ for 1 h. Assembled T5 and T7 transposomes (40 μl each) were mixed finally to form 80 μl dimers and were stored at -20 ℃ till use. 8 μl assembled tn5 was mixed with 1 μl cDNA (~500 pg) and 1 μl of tagmentation buffer (TNP92110, Lucigen). The reaction mixture was incubated at 37 ℃ for 1 h with heated lid off. 1μl of 10 x stop buffer (TNP92110, Lucigen) was added to the sample and mixed well. Two rounds of 1x beads cleanup were performed to recover tagmented cDNA and fully remove transposon dimmers. The sample was eluted with 10 μl low EDTA TE buffer. 25 μl KAPA HiFi hotstart readymix (KAPA, KR0370) was heated at 98 ℃ for 30 sec. The activated polymerase was mixed with 10 μl tagmented cDNA and 10 μl Nuclease-free water and incubated at 72 ℃ for 5 min. 1.5 μl of 25 μM P5 primer, 1.5 μl of 25 μM P7 primer and 2.5 μl Evagreen were mixed with the sample. The following amplification program was used: 98 ℃ for 30 s; n cycles of 98 ℃ for 10 s, 63 ℃ for 30 s; and 72 ℃ for 30 s; and a final extension at 72 ℃ for 1 min. Amplified cDNA libraries purified with 0.8x SPRIbeads and resuspended in 10 μl low EDTA TE buffer.

Library quality control and sequencing

Before sending libraries to sequencing facility for sequencing, the quality of each libraries was assessed by measuring its size by Tapestation (Aiglent, High Sensitivity DNA ScreenTape) and its concentration by KAPA qPCR Library quantification kit. ~ 20 libraries were pooled for sequencing by Illumina HiSeq 4000 with single-end 50-nt read.

ChIP-seq data analysis

Quality of raw sequencing data was assessed using fastqc. Adapter sequence was removed by Trim Galore! with default setting. Trimmed reads were aligned to mm10 reference genome

using bowtie2(*245*). Output sam files were converted and sorted to bam files using samtools(*266*). Unmapped reads, multi-mapped reads and duplicate reads were removed using sambamba(*267*). Bam files were indexed using samtools and were converted to bed files using bedtools(*268*). Blacklist regions were subtracted from bed files. Broad peak calling was performed using epic2(*269*) with the following parameters: -gn mm10 -fs 250 -g 4 -fdr 1e-20. For genome track visualization, bam files were normalized with a 10 bp bin size using bamCoverage from DeepTools(*270*) using the following parameters: --binSize 10 –numberOfProcessors max—normalizeUsing RPGC –effectiveGenomeSize 2652783500 –ignoreForNormalization chrM –extendReads 350. The output bigwig files were opened in IGV for visualization. Input bam files was normalized by bamCoverage using the same setting. To obtain consensus peak, we devided the neocortex into three groups: anterior, middle and posterior, according to the genome-wide heatmap. We randomly picked three raw data from each group and merge the three raw data and call peaks. Three called peaks were then merged to obtain consensus peak sets. We obtained a score-in-peak matrix by calculating the average normalized read counts in each peak region using multiBigwigSummary. Sample score was subtracted with input score for input normalization. After input normalization, the score-in-peak matrix was used for k-means clustering using L2 normalization. The optimized number of clusters was determined by Elbow method. Go analysis was performed using GREAT(*271*). ChIP-seq score-in-peak in promoter regions were extracted. Promoter regions were defined as ±2 kb from transcription start sites. Corresponding gene expression level was calculated and was used for computing Pearson correlation coefficients between promoter H3K27me3 peaks and mRNA signals.

$$\begin{aligned} &Input\ normalized\ signal \\ &= ChIP\left(\frac{total\ number\ of\ uniquely\ mapped\ reads * fragment\ length}{effective\ geonome\ size}\right) \\ &- Input\left(\frac{total\ number\ of\ uniquely\ mapped\ reads * fragment\ length}{effective\ geonome\ size}\right) \end{aligned}$$

mRNA-seq data analysis

Raw reads were trimmed by Trim Galore! with default setting. The trimmed reads were mapped to mm10 reference genome using hisat2 with default setting. Exon rate was calculated using SeqMonk. Differential expression analysis was performed using salmon in mapping-based mode(*272*). Salmon index for decoy-aware mm10 transcriptome was built. After transcript counting, the quant.sf file was processed using Deseq2(*273*) with custom R script. 43 high-confidence genes were picked from H3K27me3 clustering. We first obtained the overlapped GO terms between neuronal cluster 0 of left and right neocortex from mouse 1. Genes involved in these overlapped GO terms are defined as high-confidence genes.

5.3 Results

ChIP-seq "tomography" for spatially resolved H3K27me3 profiles

Different regions of neocortex have specific functions such as sensory perception, generation of motor command, or vision. In order to investigate the spatial variations of H3K27me3 in the rostral/caudal neuraxis, we dissected brain from ten-week-old mouse, sliced it into ~ 20 coronal sections with 1 μm thickness in a slicer, and isolated left and right neocortex from each section (Figure 5.1a). We then extracted nuclei from each individual tissue and sorted out neuronal and glial fractions using NeuN labeling and fluorescence-activated cell sorting (FACS). We generated ~320 ChIP-seq libraries (2 biological replicates and 2 technical replicates) using 4,000 sorted nuclei per assay. We observed 8,000 ~ 12,000 H3K27me3 peaks per sample and high correlation between technical replicates and biological replicates, demonstrating that we were able to reproducibly generate high-quality data. Left and right neocortex had similar H3K27me3 landscapes. The correlation between left and right were around 0.9, and glia had a slightly lower left-and-right correlation in contrast to neurons (Figure 5.6). We observed substantial differences in the epigenome between neuron and glia (Figure

5.1a). Neurons exhibited several spatial variation patterns while glia had relatively consistent profiles across the entire neocortex. To identify H3K27me3 peaks with similar spatial variation, we applied K-means clustering and partitioned all H3K27me3 peaks into 12 clusters (Figure 5.1b). Besides, we calculated Pearson correlation coefficient between tissues separated by a various number of tissues (Figure 5.1d). As expected, adjacent sections were more highly correlated than slices that were spaced far apart. The results showed that nearest tissues had similar epigenomic profiles and differences in epigenome became increasingly significant when the distance increased between two tissues. Also, our data demonstrated that neurons had a higher degree of spatial variations in contrast to glia, which was likely due to the fact that neurons have a greater variety of shapes and spatial arrangements in contrast to glia (*251*).

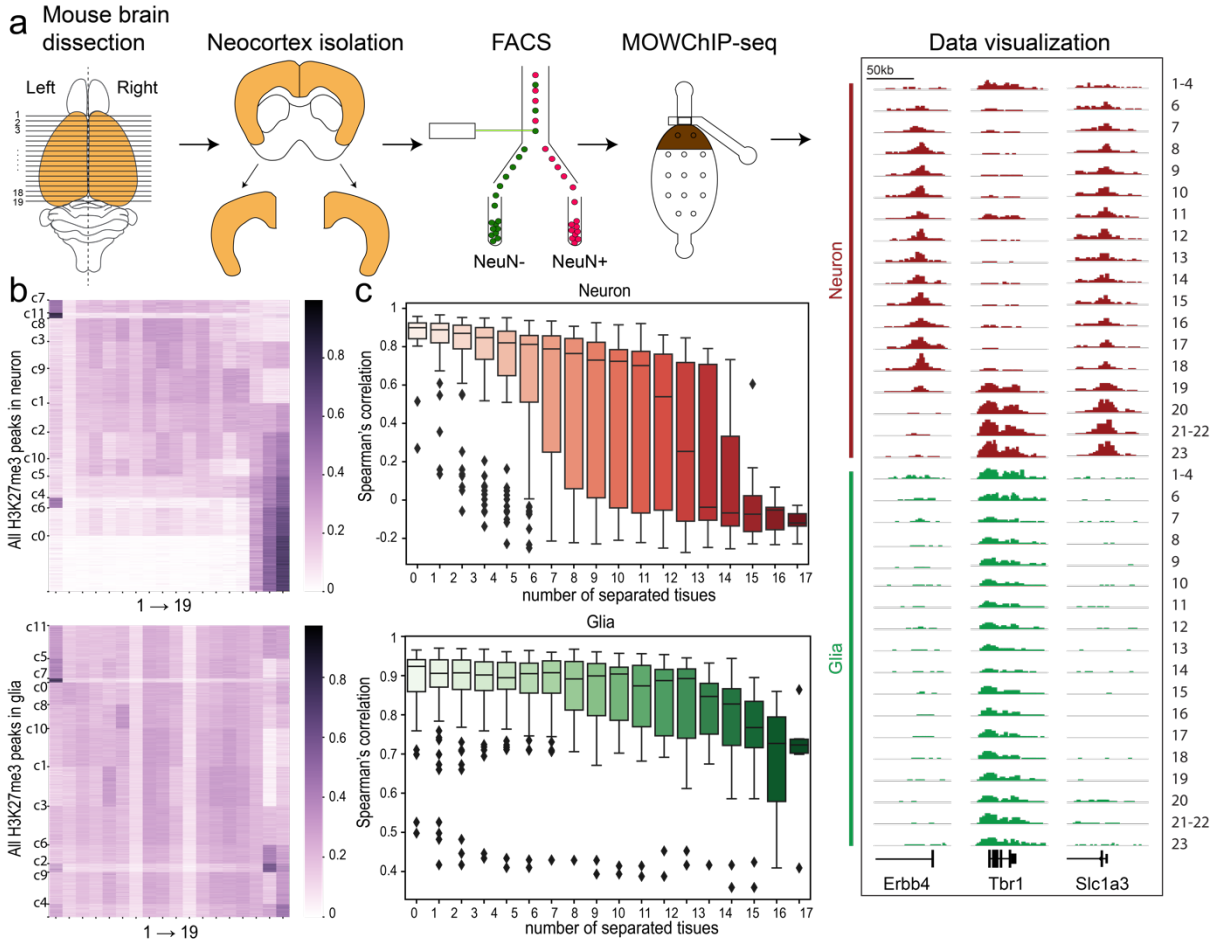


Figure 5.1 Profiling spatial variations of H3K27me3 in adult mouse neocortex.

a, Schematic experiment workflow and H3K27me3 ChIP-seq signals in brain related genes. Neocortex was isolated from ~20 coronal sections and was separated into left and right part (~40 tissues in total). Sorted neuronal and glial cells from each piece of tissue were used for low-input ChIP-seq (MOWChIP-seq, 4,000 nuclei per reaction). Normalized H3K27me3 signals in neuronal/glial cells were shown. Brain related genes were selected to represent spatial or cell-type-specific pattens. **b**, K-means clustering of all H3K27me3 peaks in neuronal/glial cells from mouse 1 right neocortex sections. L2 norm and Euclidean were used. The number of k was determined by Elbow method. Neuronal cells have clusters with distinct spatial patterns while no clear pattern shows in glia clusters. **c**, Boxplot of Spearman's correlation between biological replicates at the same location (number of separated tissues = 0), or between two tissues separated by a various number of tissues.

Different regions in neocortex has different H3K27me3 profiles

After manually examining the data in the genome browser, we identified several distinct spatial variation patterns in neurons. In order to systematically investigate the pattern of H3K27me3 in neocortex, we first calculated the genome-wide Pearson correlation for each pair of data (Figure 5.3, bin size = 10bp). As expected, the correlation between neurons and glia was relatively small, demonstrating different cell types had distinct H3K27me3 profiles. We observed three blocks of contiguous sections in neurons that were highly correlated among each other. However, glia didn't have such spatial variations as all glia data were strongly correlated with each other. Based on the spatial patterns of neurons, the neocortex can be roughly divided into three parts: anterior, middle, and posterior neocortex. Interestingly, we observed the posterior neurons had slightly higher correlation with glia then neurons. One possible explanation is that posterior neocortex had a distinct type of neurons that had similar H3K27me3 profiles with glia at certain genomic regions.

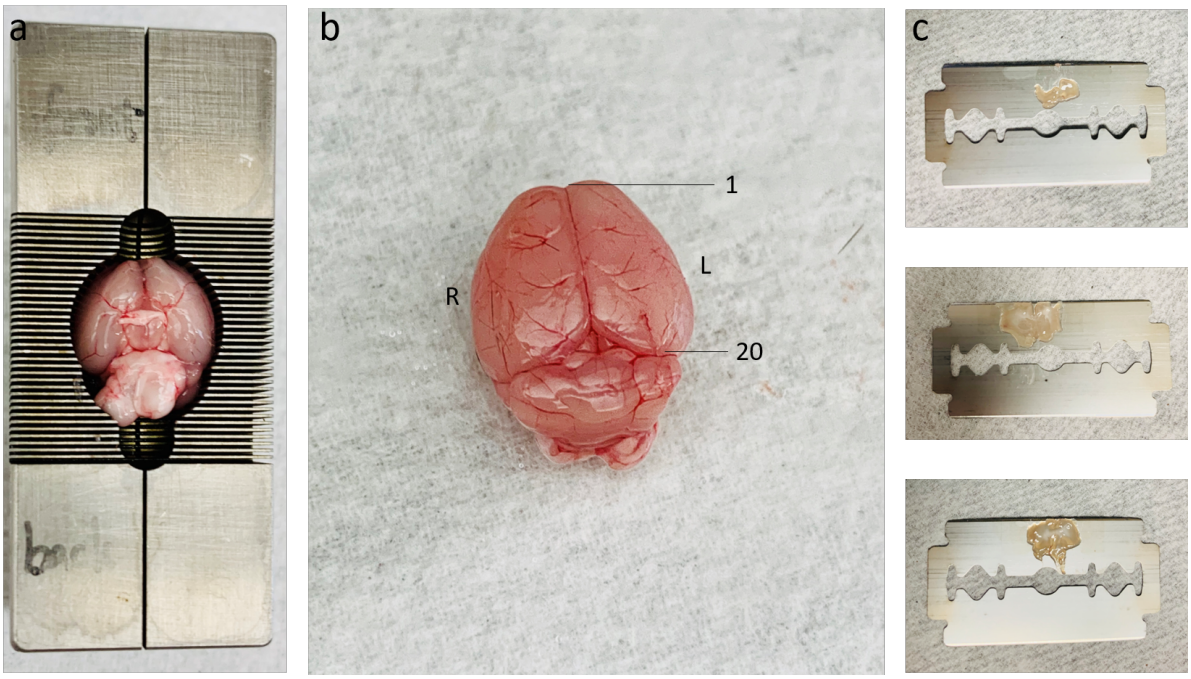


Figure 5.2 Mouse dissection.
a, mouse brain placed in a slicer. **b**, mouse brain before dissection. L and R notes for left and right. Sections from anterior to posterior was noted from 1 to 20. **c**, coronal sections of the brain.

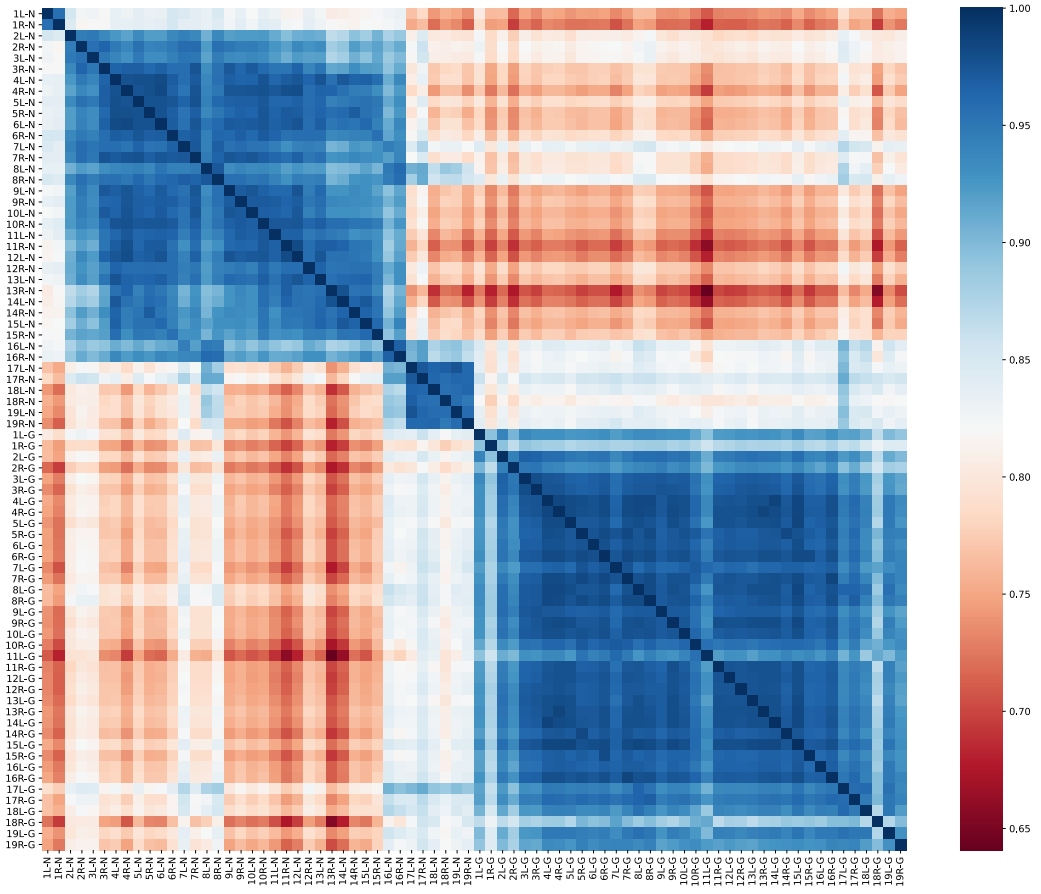


Figure 5.3 Genome-wide correlation of neuronal/glial H3K27me3 ChIP-seq data.

Data from two technical replicates and two biological replicates were combined.

Previous studies have pointed out that neuron-glia ratios vary throughout the nervous system and are important to the cortical dynamics(*251*). Thus, we calculated the neuron/glia ratios in each neocortex sections by FACS (Figure 5.5). We found a consistent variation pattern between biological replicates: neuron/glia ratio increased significantly in posterior neocortex. It's possible that the unique H3K27me3 profiles in posterior neocortex could be associated with variation of neuron/glia ratio. One hypothesis is that neuron/glial ratio is determined by the size of neuron as large neuron should require support from a larger number of glial cells(*274*). Glia is believed to vary little in in morphology and cell mass, hence a larger neuronal size results in smaller neuronal/glia ratio. By this hypothesis, posterior neocortex would have a large fraction of neurons with small size, which should have distinct functions and epigenome profiles.

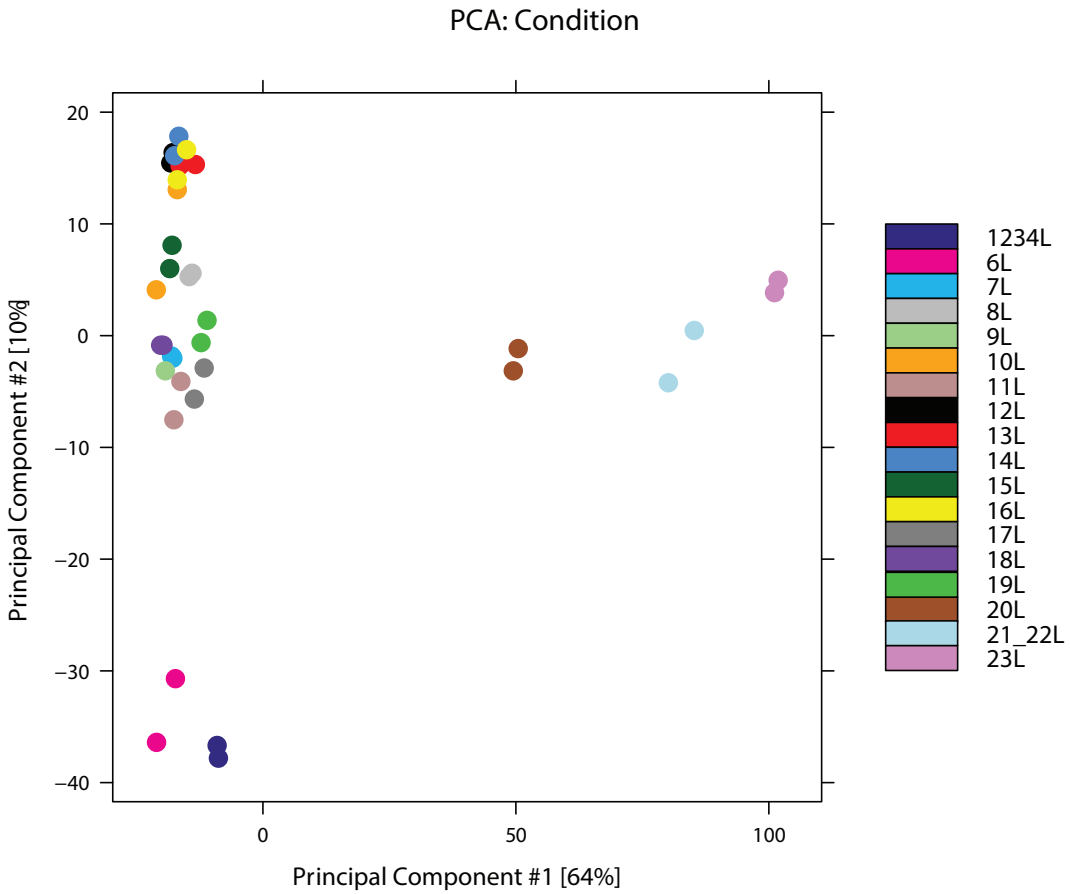


Figure 5.4 Principal component analysis of H3K27me3 data from mouse 1 left neurons.

K-means clustering identified H3K27me3 peaks with distinct spatial variation

To better understanding the spatial structure of H3K27me3 in mouse neocortex, we applied unsupervised clustering to identify the peaks that drive these spatial variations. We performed Gene ontology (GO) analysis for each cluster using GREAT. We compared k-means clustering with k-medoid clustering and found the result of k-means clustering is more biological meaningful (Figure 5.8). Clusters partitioned by k-means were enriched for genes critical for the development of central nervous system, such as central nervous system vasculogenesis, central nervous system neuron development, and lateral motor neuron migration. In contrast, k-medoid clustering failed to group together peaks involved in brain related biological functions. Hence all clustering results discussed below were generated using k-means clustering.

Based on the enriched GO terms in each cluster we were able to identify the a few clusters with different functions. For example, cluster 0 (c0) were highly enriched for memory, glutamate receptor signaling pathway, axonogenesis and transmission of nerve pulse. Glutamate is an important excitatory neurotransmitter and plays a key role in learning and memory(*275*). Thus cluster 0 was likely involved in the establishment and maintenance of memory. In this way we annotated the major function of each cluster based on its top enriched GO terms. We assigned annotations for each cluster as following (Figure 5.9) : memory (c0), cell fate commitment (c1), embryonic organ development (c2, c3, c8), cell migration in brain (c4), pattern specification (c5), skeletal system development (c6), spinal cord development (c7), negative regulation of transcription (c9), epidermis morphogenesis (c10), central nervous system development (c11). These annotations showed that H3K27me3 peaks with similar spatial variation tend to involve in the same biological activity, and many peaks that had brain related functions exhibited unique spatially structured signals.

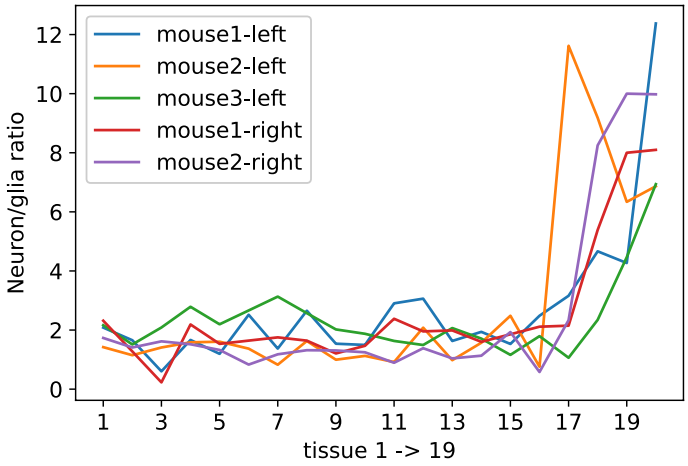


Figure 5.5 Neuron/glia ratio across left and right neocortex from different biological replicates.
This neuron/glia ratio in grey matter from each neocortex slices were determined by FACS.

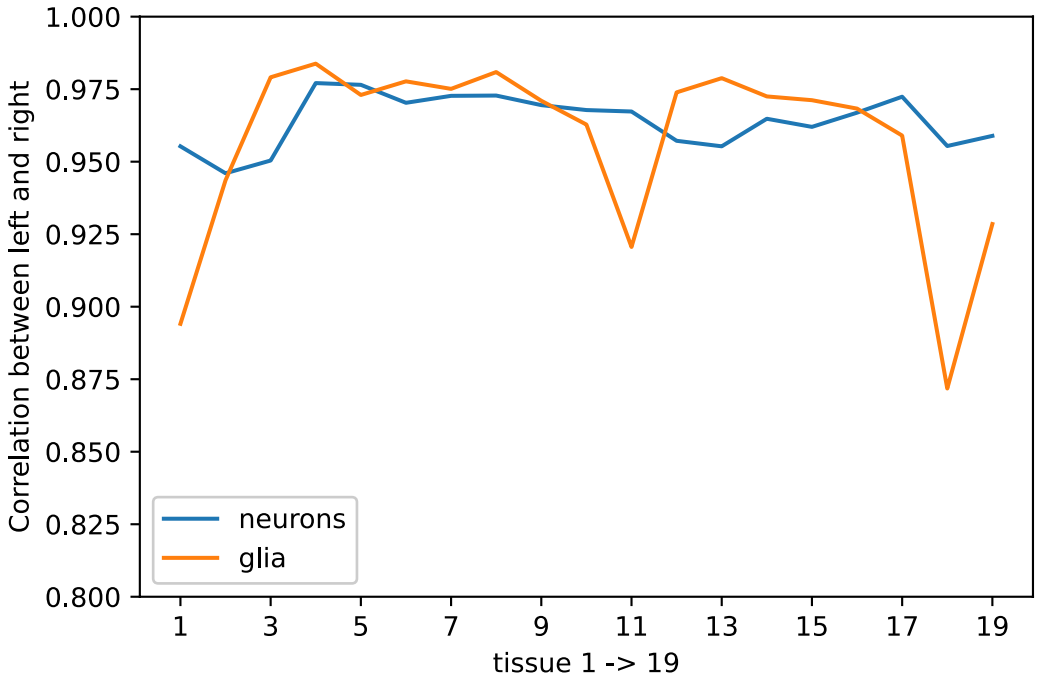


Figure 5.6 Pearson correlation between left and right sections across entire neocortex.

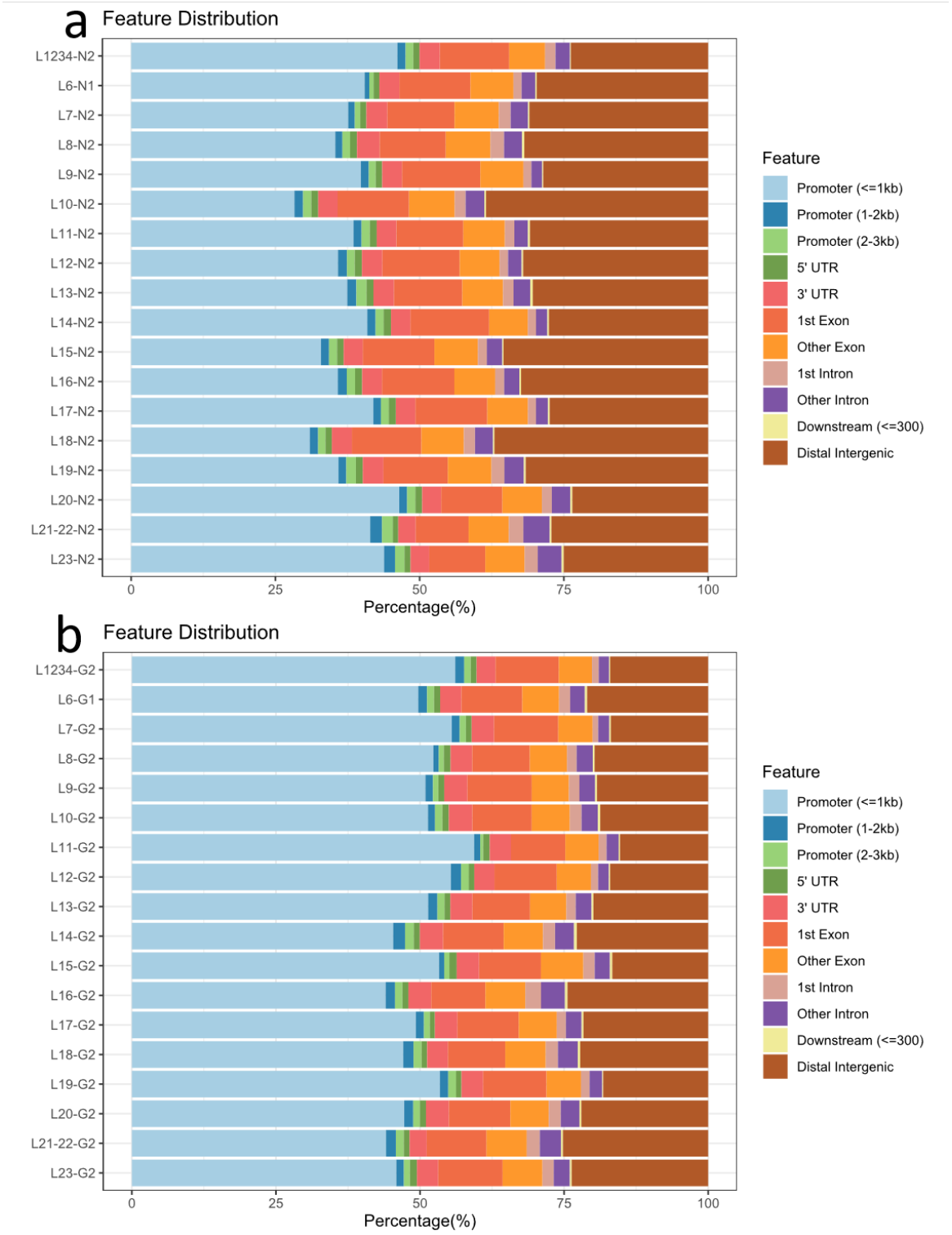


Figure 5.7 Genome-wide annotation for H3K27me3 peaks in neurons and glia.
a, neurons from left neocortex; **b**, glia from left neocortex.

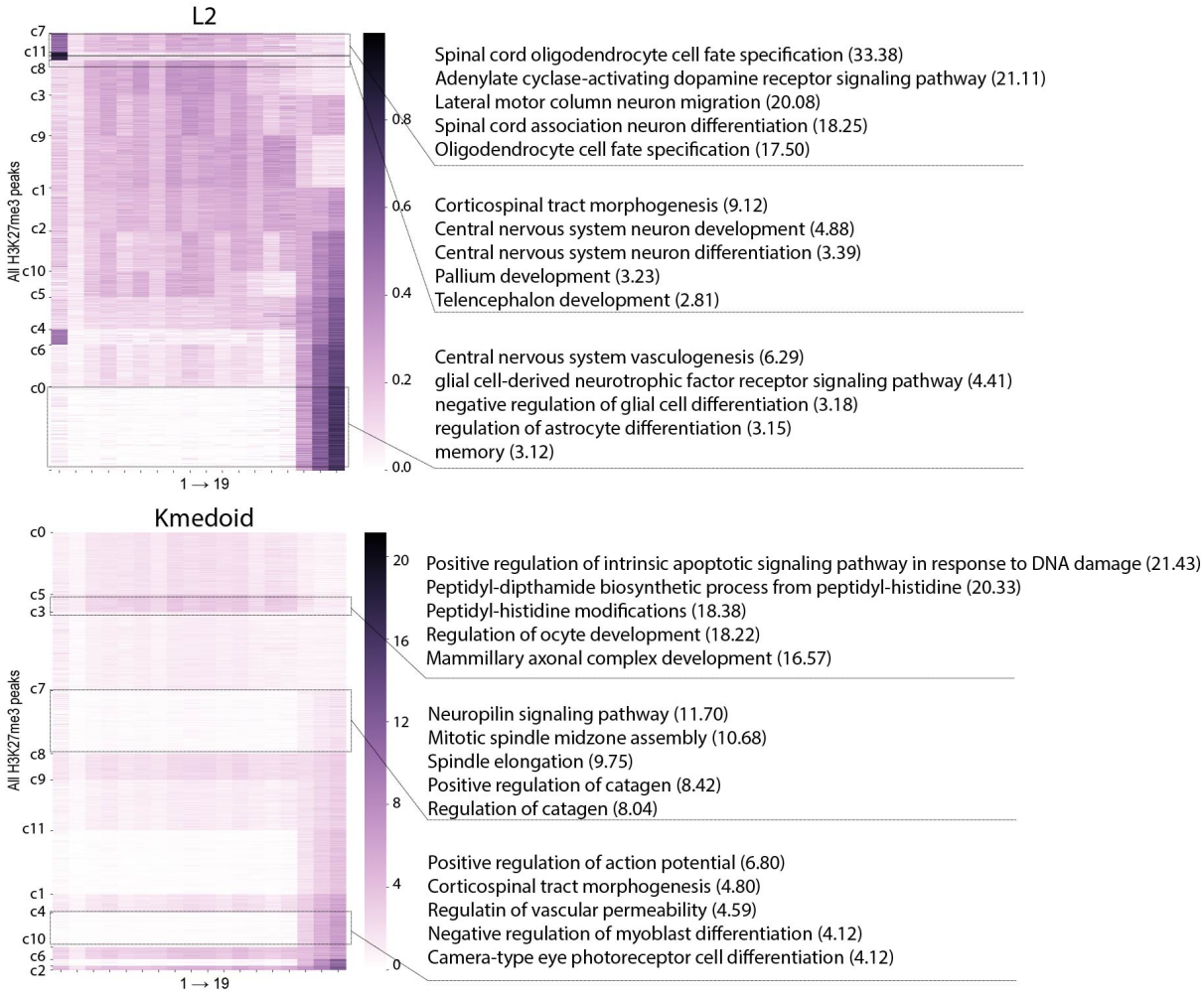


Figure 5.8 Optimization of K-means clustering.
Clustering heatmap using H3K27me3 peaks in neuronal cells from mouse 1 right neocortex. Two types of data preprocessing method were compared: L2 norm and Kmedoid. Go term analysis for representative clusters was performed using GREAT. Relevant ontologies with the highest Binomial fold enrichment were listed. Full list of ontologies for each cluster can be found on supplementary tables. Clustering result using L2 normalized data have more brain related biological process terms compared to Kmedoid.

We performed k-means clustering with L2 normalization for all datasets (Figure 5.10). The number of k was determined by Elbow method. We observed a universal consistency between left and right in both neurons and glia. Two biological replicates showed similar clustering results demonstrating that our data is reproducible and robust. As expected, all neuron data had clusters with distinct spatially restricted pattern while clusters in glia didn't show any spatial specific signals.

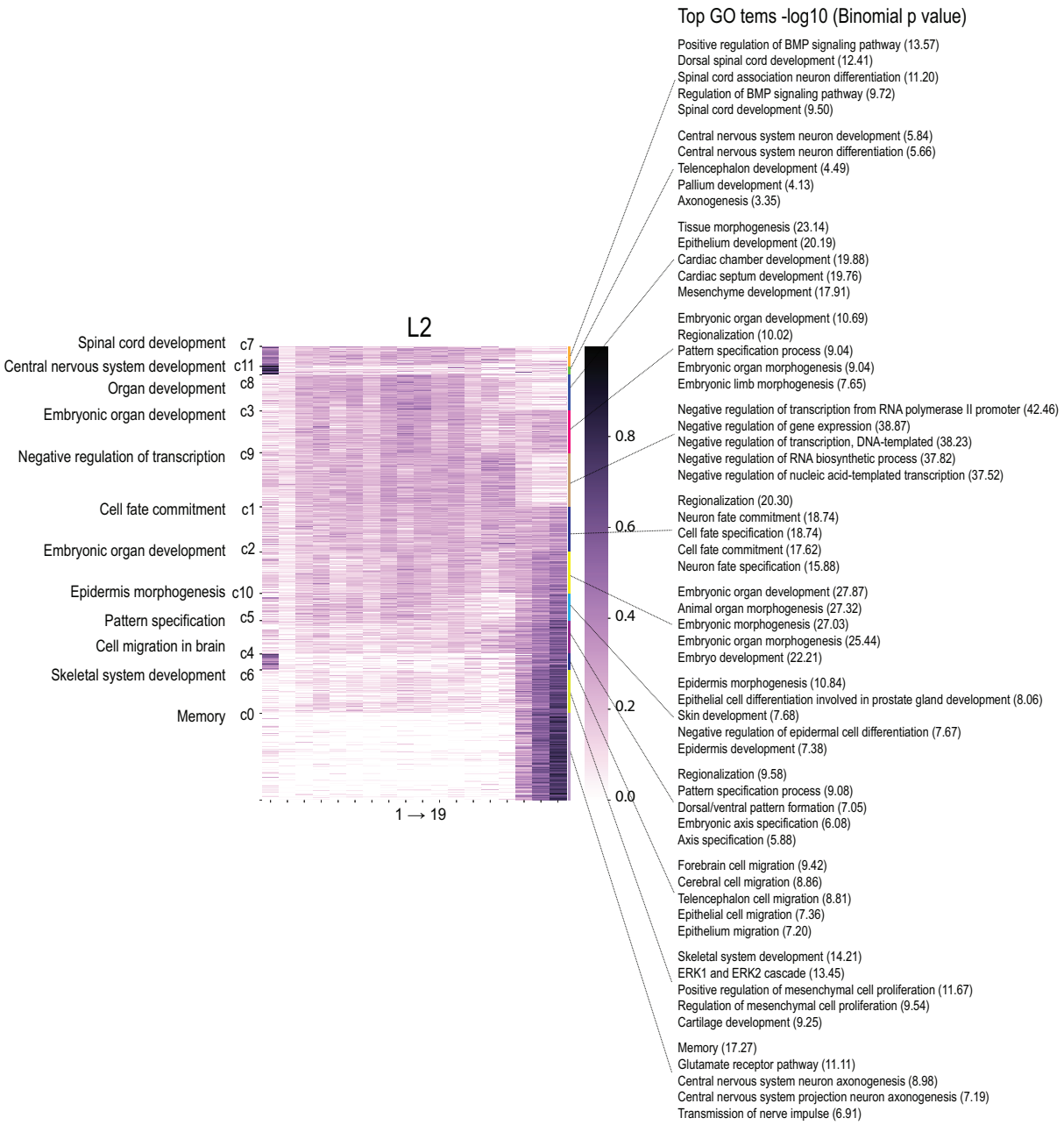


Figure 5.9 Annotation of H3K27me3 clusters in neurons from right neocortex.
K-means clustering with L2 normalization was performed. Go term analysis for representative clusters was performed using GREAT. Relevant ontologies with the highest -log10 (Binomial p value) were listed.

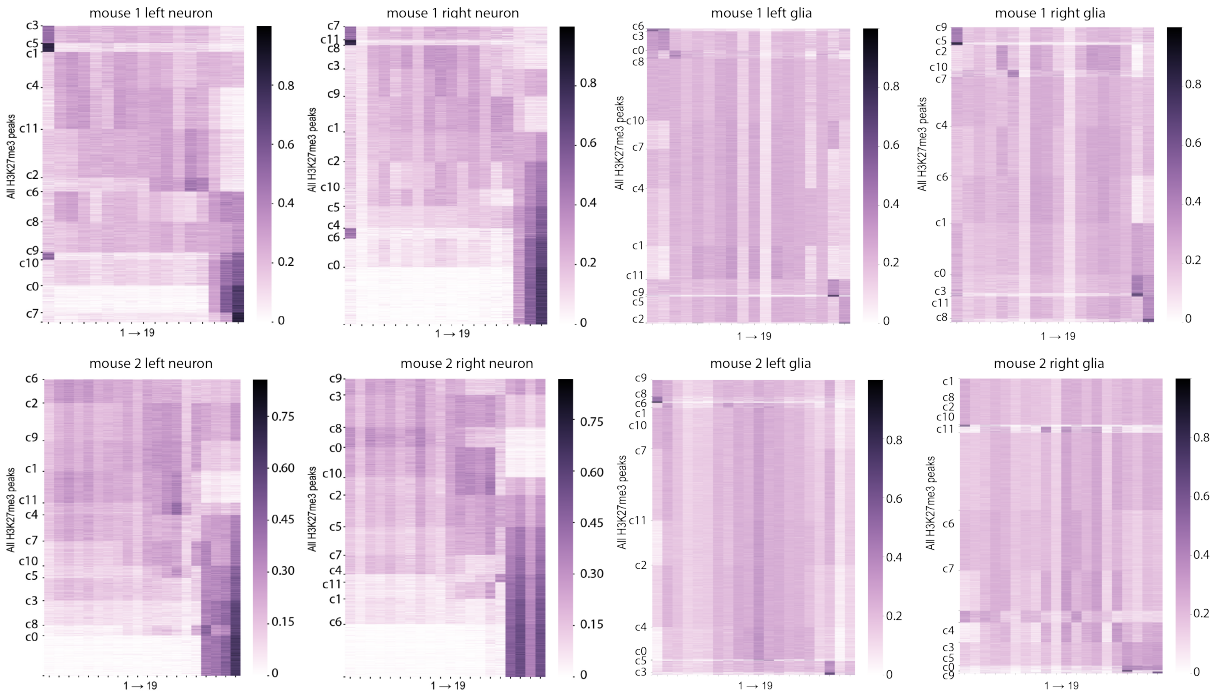


Figure 5.10 K-means clustering using L2 norm.
Clustering results are similar between left and right in both neuron and glia. Mouse 1 and mouse 2 show consistent clustering profiles.

Comparison between H3K27me3 and gene expression tomography

To better understand the spatial variations in neocortex, we generated mRNA-seq tomography by applying SMART-seq2 to neuronal cells extracted from left neocortex slices. We found genes where its expression level was in inverse relationship with H3K27me3 signals (Figure 5.11c), such as Ptn, known as a key neuromodulator for central nervous system(*276*). However, there were also many genes having spatial expression patterns different from the variation of nearby H3K27me3 signals. To further understand gene expression variations across entire neocortex, we performed k-means clustering of all genes (Figure 5.11a). We observed two large clusters of genes with a uniform distribution of expression level across the space. Those genes showed minimal spatially restricted expression, which also means they were not cell-type-specific genes. Among those genes, we found many house-keeping genes, such as GAPDH, PGK1, PPIA and RPL13A. Except for these two large clusters, the rest 10 clusters all showed different spatial patterns to certain extend. Several clusters had genes

expressed in a highly spatial-specific manner, such as cluster 9. Based on assignment of marker genes to these clusters, we identified the dominant neuron cell types in some neocortex slice. For example, cluster 4 had MGE-derived marker genes (Pvalb and Reln), CEG-derived marker (Prox1) and pan-inhibitory marker (Slc6a1).

Interesting most spatially restricted gene expression occurred in anterior and posterior neocortex. The two regions also had the most drastic differences in H3K27me3 profiles. However, transcriptomic spatial variation was much more complex and dynamic compared with H3K27me3 profiles. For examples, we selected 43 high-confidence genes from H3K27me3 cluster 0 and plotted heatmaps using expression level and nearby H3K27me3 signal intensity (Figure 5.11b). H3K27me3 peaks related with the 43 genes showed similar pattern as they were from the same cluster. In contrast, the gene expression heatmap contained a few different variation patterns. Although these 43 genes were under the same H3K27me3 regulation, they didn't express the same level of transcription activity. This demonstrated that transcriptome and H3K27me3 are ultimately two layers of information, and H3K27me3 mark has to work with other epigenomic modifications, such as DNA methylation or other histone mark to regulate gene expression.

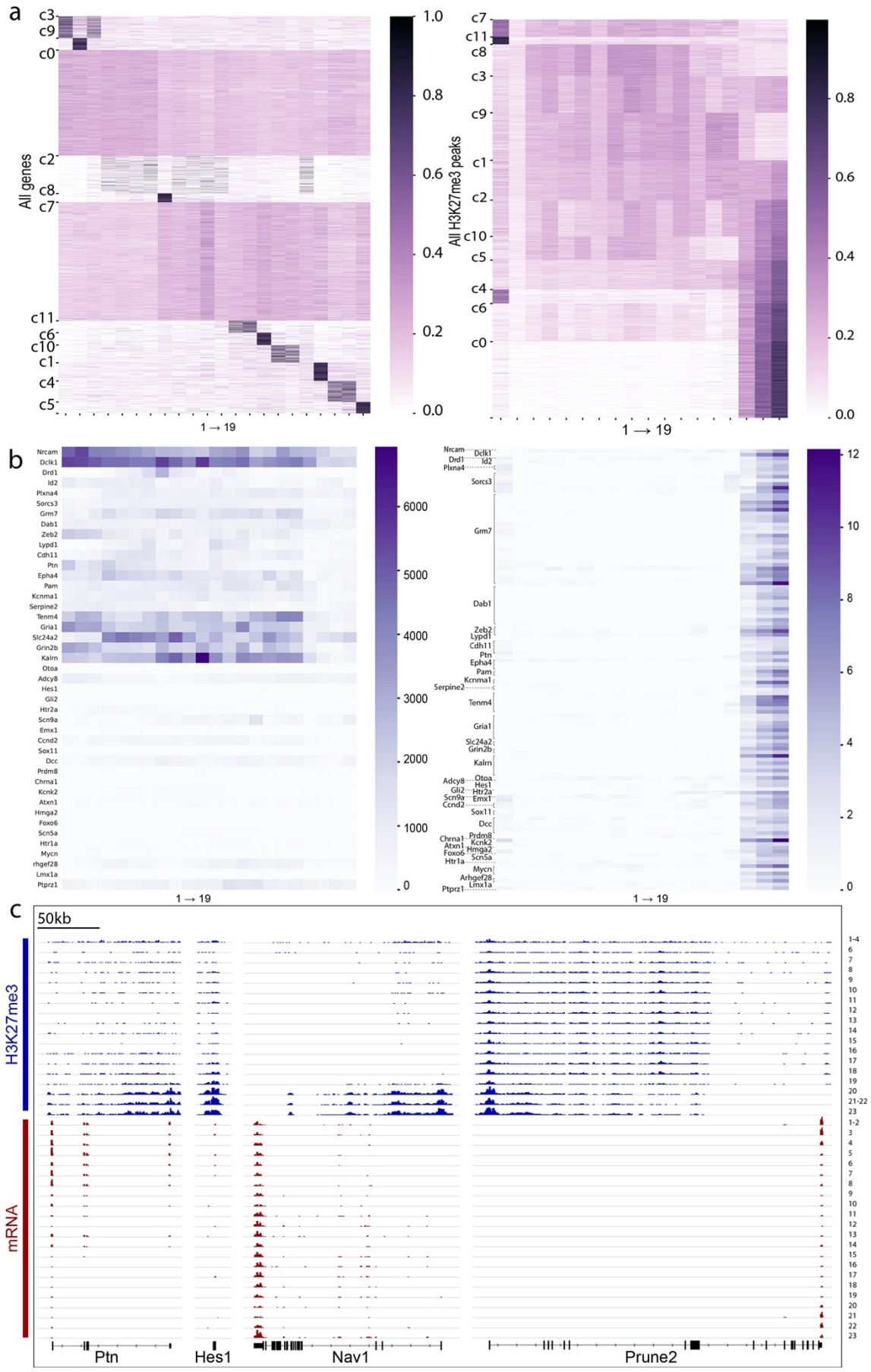


Figure 5.11 Comparison between H3K27me3 ChIP-seq and mRNA-seq.

a, Comparison of K-means clustering of ChIP-seq and mRNA-seq using neurons from right neocortex. **b,** 43 genes related with brain functions were selected using GO term results. ChIP-seq and mRNA-seq signals over those genes were plotted. Among the 43 genes, a small group of genes show an inverse relationship with H3K27me3 signals while the majority of genes don't. **c,** Normalized ChIP-seq and mRNA-seq data in genome browser tracks. Brain related genes were selected to represent the relationship between epigenome and transcriptome.

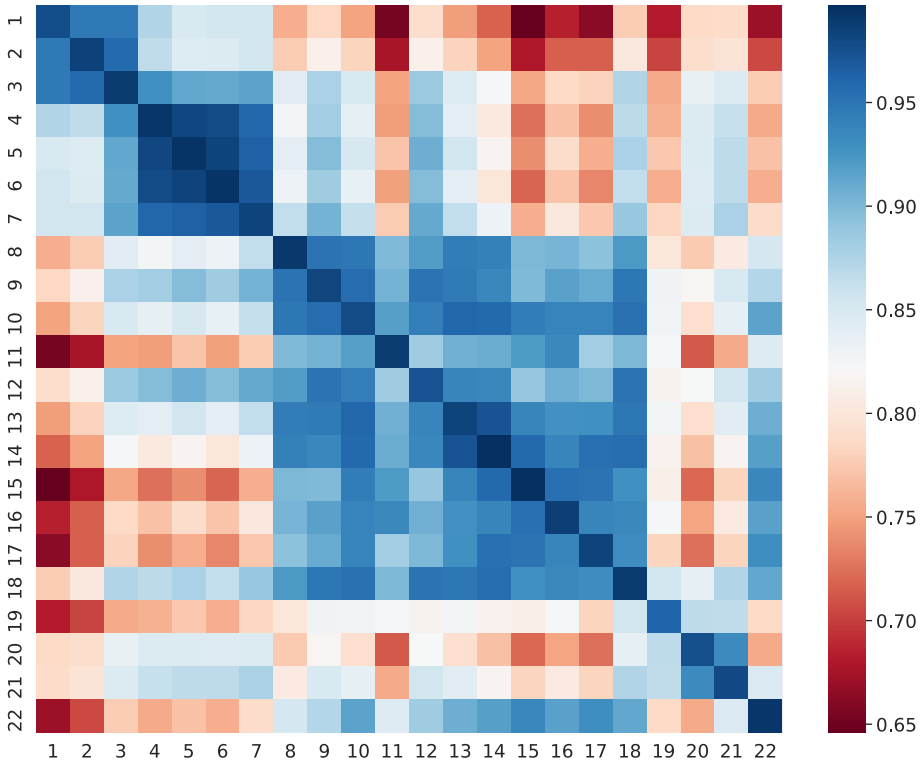


Figure 5.12 Pearson correlation heatmap of mRNA-seq of neurons in right neocortex.

5.4 Discussion

We generated H3K27me3 tomography by profiling genome-wide H3K27me3 landscapes of neuronal and glial cells extract from ~20 slices of adult mouse neocortex using low-input ChIP-seq. We observed cell-type-specific epigenomic differences between neurons and glia. More importantly, we observed three blocks of neocortex that had distinct H3K27me3 profiles in neurons but universal profiles in glia. We identified 8,000 ~ 11,000 H3K27me3 peaks per data and found 40~50% peaks showed spatial restricted signals. We applied unsupervised clustering to group peaks with similar spatial variation of signals and found that

peaks exhibiting spatial-specific signal distribution tend to involve in brain functions. Our data set can also be used as a reference for selecting candidate peaks for future locus-specific studies.

To further understand this spatial pattern, we continued to generate transcriptomic tomography of neurons from right neocortex slices. Most genes with spatial specific expression were particularly expressed in anterior or posterior neocortex, which was in agreement with our H3K27me3 data. Interestingly, transcriptome showed much more complex variation patterns compared with H3K27me3. Genes with similar H3K27me3 signals can have very different expression profiles. The difference is probably because H3K27me3 is only part of the epigenomic regulatory network and has to work with other elements such as other histone marks.

The resolution of our method can be improved by dissecting thinner slices using cryosection, though extracting small number of nuclei by centrifuged based approach may become new problems. On the other hand, as spatial transcriptomics technology continues to evolve, we expect spatial epigenomics based on similar strategy will be developed in the near future. However, our method can be used as a simple and straight-forward way to investigate spatial epigenome and to search for candidate regulators. We expect our epigenomic tomography be applied to both normal and diseased brain and facilitate genome-wide comparison and discovery of epigenomic mechanisms underlying neuro-functional disease progression.

Chapter 6 - Summary and Outlook

Microfluidic technology has greatly improved the performance of genomic and epigenomic analysis including sensitivity, specificity and throughput. It provides unique capabilities for processing low-input (including single-cell) samples due to its dramatic sample size reduction and implementing automated and high-throughput operations due to its extraordinary scalability. In this thesis, a microfluidic device combining indexed ChIPmentation for profiling ultralow-input histone modification was described. Besides, another two projects were discussed that applied low-input microfluidic protocol, MOWChIP-seq, and studied culture conditions on epigenomic dynamics and spatial variation in epigenome of mouse neocortex. More applications of microfluidics in the field of epigenomics are expected in the near future. The future work will potentially focus on three aspects.

Firstly, droplet microfluidics will continue to be applied in single cell technology to increase the number of cells processed per assay. Single cell sequencing generally requires thousands of cells in order to obtain meaningful results from downstream analysis. Droplet microfluidics is able to process large number of cells in a much faster manner compared to 96-well plates. Two commonly used strategies for high-throughput single cell sequencing are droplet microfluidics and combinatorial indexing. The major advantage of droplet microfluidics is the capability to easily scale up the throughput to tens, hundreds of thousands of cells. Currently the market of droplet microfluidics has been dominated by 10x Genomics, and the cost to prepare one single-cell library is a few thousand dollars. I expect more commercial systems coming up in the near future with increased throughput and decreased price.

Secondly, commercial microfluidic platforms that can be adapted in most biology laboratories will greatly expand the application of microfluidics. Currently, the requirement of setting up complex system for fabrication and operation have prevented many labs from using

microfluidics. An easy solution with detailed instructions and requiring little external equipment are needed. In the future microfluidic devices, all necessary reagents are pre-stored on the device and operations is fully programed. Within a few hours after loading samples, results will be automatically sent to user's smart phones. Such highly integrated and automated microfluidic systems will greatly benefit researchers who is unwilling or unable to use microfluidics technology by themselves due to its high entry barriers. Since most epigenomic profiling requires multiple steps (on both the treatment of genomic DNA and sequencing library preparation), how to seamlessly assemble these steps into an efficient yet manageable microfluidic process will be a new challenge for the community. New strategies and practices will be needed.

Lastly, others epigenomic assays, such as mapping non-coding RNA, chromatin organization, or DNA methylation variants from low number of cells will be adapted to microfluidic devices in the near future. Decreasing input cell number and eventually going down to single cell resolution is the future directions of all epigenomic assays. Due to the unique advantages of microfluidic devices handling small sample volumes, there is no doubt that microfluidics will continue to provide solutions to low-input epigenomic assays. These solutions will be designed to suit the needs of specific epigenomic assays whose working principles can be drastically different.

In summary, the future of microfluidic epigenomic technology is anticipated to include high throughput single cell sequencing, highly integrated commercial platforms, and a broad repertoire of on-chip epigenomic assays.

APPENDIX

Table S1. Transposon and primer sequences

Oligo	Sequence (5'-3')
T5_Universal	TAC ACT CTT TCC CTA CAC GAT CGT CGG CAG CGT CAG ATG TGT ATA AGA GAC AG
T7.1_TAAGGCGA	GAA GAC GGC ATA CGA GAT TCG CCT TAG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.2_AGGCAGAA	GAA GAC GGC ATA CGA GAT TTC TGC CTG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.3_TCCTGAGC	GAA GAC GGC ATA CGA GAT GCT CAG GAG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.4_GGACTCCT	GAA GAC GGC ATA CGA GAT AGG AGT CCG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.5_TAGGCATG	GAA GAC GGC ATA CGA GAT CAT GCC TAG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.6_CTCTCTAC	GAA GAC GGC ATA CGA GAT GTA GAG AGG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.7_CAGAGAGG	GAA GAC GGC ATA CGA GAT CCT CTC TGG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.8_GCTACGCT	GAA GAC GGC ATA CGA GAT AGC GTA GCG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.9_CGAGGCTG	GAA GAC GGC ATA CGA GAT CAG CCT CGG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.10_AAGAGGCA	GAA GAC GGC ATA CGA GAT TGC CTC TTG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.11_GTAGAGGA	GAA GAC GGC ATA CGA GAT TCC TCT ACG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.12_GTCGTGAT	GAA GAC GGC ATA CGA GAT ATC ACG ACG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.13_ACCACTGT	GAA GAC GGC ATA CGA GAT ACA GTG GTG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.14_TGGATCTG	GAA GAC GGC ATA CGA GAT CAG ATC CAG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
T7.15_CCGTTTGT	GAA GAC GGC ATA CGA GAT ACA AAC GGG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
P7.16_TGCTGGGT	GAA GAC GGC ATA CGA GAT ACC CAG CAG TCT CGT GGG CTC GGA GAT GTG TAT AAG AGA CAG
pMENTS	/5Phos/CT GTC TCT TAT ACA CAT CT
P5_PCR_primer	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GA
P7_PCR_primer	CAA GCA GAA GAC GGC ATA CGA GAT

REFERENCES

1. R. Margueron, D. Reinberg, Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* **11**, 285-296 (2010).
2. X. Dong, Z. Weng, The correlation between histone modifications and gene expression. *Epigenomics* **5**, 113-116 (2013).
3. T. Suganuma, J. L. Workman, Crosstalk among histone modifications. *Cell* **135**, 604-607 (2008).
4. A. A. Soshnev, S. Z. Josefowicz, C. D. Allis, Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. *Mol Cell* **62**, 681-694 (2016).
5. C. D. Allis, T. Jenuwein, The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500 (2016).
6. M. F. Fraga *et al.*, Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* **102**, 10604-10609 (2005).
7. J. E. Castillo-Fernandez, T. D. Spector, J. T. Bell, Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med* **6**, 60 (2014).
8. I. I. Gottesman, L. Erlenmeyer-Kimling, Family and twin strategies as a head start in defining prodromes and endophenotypes for hypothetical early-interventions in schizophrenia. *Schizophrenia Research* **51**, 93-102 (2001).
9. N. Rusk, Cheap third-generation sequencing. *Nature Methods* **6**, 244-244 (2009).
10. D. J. Munroe, T. J. Harris, Third-generation sequencing fireworks at Marco Island. *Nature biotechnology* **28**, 426-428 (2010).
11. C. Deng, L. B. Naler, C. Lu, Microfluidic epigenomic mapping technologies for precision medicine. *Lab Chip* **19**, 2630-2650 (2019).
12. L. Liu, Y. Li, T. O. Tollefsbol, Gene-environment interactions and epigenetic basis of human diseases. *Curr Issues Mol Biol* **10**, 25-36 (2008).
13. K. D. Hansen *et al.*, Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* **43**, 768-775 (2011).
14. M. M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476 (2008).
15. P. I. Thakore *et al.*, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015).
16. I. B. Hilton *et al.*, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517 (2015).
17. B. Wen, H. Wu, Y. Shinkai, R. A. Irizarry, A. P. Feinberg, Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* **41**, 246-250 (2009).
18. B. van Steensel, A. S. Belmont, Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791 (2017).
19. S. Geisler, J. Coller, RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* **14**, 699-712 (2013).
20. R. A. Gupta *et al.*, Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071-1076 (2010).
21. I. Versteege *et al.*, Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206 (1998).
22. T. J. Pugh *et al.*, Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**, 106-110 (2012).
23. J. H. Zhang *et al.*, A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329-334 (2012).

24. R. S. Lee *et al*., A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *J Clin Invest* **122**, 2983-2988 (2012).
25. S. C. Mack *et al*., Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445-450 (2014).
26. A. P. Feinberg, The Key Role of Epigenetics in Human Disease Prevention and Mitigation. *New Engl J Med* **378**, 1323-1334 (2018).
27. A. P. Feinberg, M. A. Koldobskiy, A. Gondor, Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet* **17**, 284-299 (2016).
28. M. Ehrlich, DNA hypomethylation in cancer cells. *Epigenomics-Uk* **1**, 239-259 (2009).
29. A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Gene Dev* **25**, 1010-1022 (2011).
30. M. Esteller *et al*., Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *Jnci-J Natl Cancer I* **92**, 564-569 (2000).
31. Y. Chervona, M. Costa, Histone modifications and cancer: biomarkers of prognosis? *Am J Cancer Res* **2**, 589-597 (2012).
32. M. F. Fraga *et al*., Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature Genetics* **37**, 391-400 (2005).
33. B. T. K. Yuen, P. S. Knoepfler, Histone H3.3 Mutations: A Variant Path to Cancer. *Cancer Cell* **24**, 567-574 (2013).
34. M. Hajjari, A. Salavaty, HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biol Med* **12**, 1-9 (2015).
35. L. Zane, V. Sharma, T. Misteli, Common features of chromatin in aging and cancer: cause or coincidence? *Trends Cell Biol* **24**, 686-694 (2014).
36. S. Heerboth *et al*., Use of Epigenetic Drugs in Disease: An Overview. *Genet Epigenetics* **6**, 9-19 (2014).
37. C. A. Hamm, F. F. Costa, Epigenomes as therapeutic targets. *Pharmacol Therapeut* **151**, 72-86 (2015).
38. W. Yan, J. G. Herman, M. Guo, Epigenome-based personalized medicine in human cancer. *Epigenomics-Uk* **8**, 119-133 (2016).
39. B. Brueckner *et al*., Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. *Cancer Res* **65**, 6305-6311 (2005).
40. E. Kaminskas, A. T. Farrell, Y. C. Wang, R. Sridhara, R. Pazdur, FDA drug approval summary: Azacitidine (5-azacytidine, Vidaza((TM))) for injectable suspension. *Oncologist* **10**, 176-182 (2005).
41. B. S. Mann *et al*., Vorinostat for treatment of cutaneous manifestations of advanced primary cutaneous T-cell lymphoma. *Clin Cancer Res* **13**, 2318-2322 (2007).
42. F. P. S. Santos, H. Kantarjian, G. Garcia-Manero, J. P. Issa, F. Ravandi, Decitabine in the treatment of myelodysplastic syndromes. *Expert Rev Anticanc* **10**, 9-22 (2010).
43. S. P. Iyer, F. F. Foss, Romidepsin for the Treatment of Peripheral T-Cell Lymphoma. *Oncologist* **20**, 1084-1091 (2015).
44. Z. N. Cao, C. Y. Chen, B. He, K. Tan, C. Lu, A microfluidic device for epigenomic profiling using 100 cells. *Nat Methods* **12**, 959-962 (2015).
45. S. Ma, Y. P. Hsieh, J. Ma, C. Lu, Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. *Science Advances* **4**, (2018).
46. T. W. Murphy, Y. P. Hsieh, S. Ma, Y. Zhu, C. Lu, Microfluidic Low-Input Fluidized-Bed Enabled ChIP-seq Device for Automated and Parallel Analysis of Histone Modifications. *Anal Chem* **90**, 7666-7674 (2018).

47. C. Deng *et al*., Multiplexed and Ultralow-Input ChIP-seq Enabled by Tagmentation-Based Indexing and Facile Microfluidics. *Anal Chem*, (2020).
48. A. Rotem *et al*., Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-1172 (2015).
49. W. L. Ku *et al*., Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325 (2019).
50. H. S. Kaya-Okur *et al*., CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1-10 (2019).
51. S. J. Hainer, A. Boskovic, K. N. McCannell, O. J. Rando, T. G. Fazzio, Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329 (2019).
52. K. Grosselin *et al*., High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060-1066 (2019).
53. B. Carter *et al*., Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat Commun* **10**, 1-5 (2019).
54. S. S. Ai *et al*., Profiling chromatin states using single-cell itChIP-seq. *Nat Cell Biol* **21**, 1164-1172 (2019).
55. Q. H. Wang *et al*., CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol Cell* **76**, 206-216 (2019).
56. K. Grosselin *et al*., High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nature genetics* **51**, 1060 (2019).
57. G. S. Kinker *et al*., Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* **52**, 1208-1218 (2020).
58. M. R. Corces *et al*., Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nature Genetics*, 1-11 (2020).
59. L. R. Volpatti, A. K. Yetisen, Commercialization of microfluidic devices. *Trends Biotechnol* **32**, 347-350 (2014).
60. S. M. Kim, S. H. Lee, K. Y. Suh, Cell research with physically modified microfluidic channels: A review. *Lab Chip* **8**, 1015-1023 (2008).
61. J. El-Ali, P. K. Sorger, K. F. Jensen, Cells on chips. *Nature* **442**, 403-411 (2006).
62. H. Andersson, A. van den Berg, Microfluidic devices for cellomics: a review. *Sensor Actuat B-Chem* **92**, 315-325 (2003).
63. G. B. Salieb-Beugelaar, G. Simone, A. Arora, A. Philippi, A. Manz, Latest Developments in Microfluidic Cell Biology and Analysis Systems. *Anal Chem* **82**, 4848-4864 (2010).
64. L. Mazutis *et al*., Single-cell analysis and sorting using droplet-based microfluidics. *Nature Protocols* **8**, 870-891 (2013).
65. T. W. Murphy, Q. Zhang, L. B. Naler, S. Ma, C. Lu, Recent advances in the use of microfluidic technologies for single cell analysis. *Analyst* **143**, 60-80 (2018).
66. S. Ma, T. W. Murphy, C. Lu, Microfluidics for genome-wide studies involving next generation sequencing. *Biomicrofluidics* **11**, 021501 (2017).
67. P. S. Dittrich, A. Manz, Lab-on-a-chip: microfluidics in drug discovery. *Nat Rev Drug Discov* **5**, 210-218 (2006).
68. M. H. Wu, S. B. Huang, G. B. Lee, Microfluidic cell culture systems for drug research. *Lab Chip* **10**, 939-956 (2010).
69. J. Pihl, M. Karlsson, D. T. Chiu, Microfluidic technologies in drug discovery. *Drug Discov Today* **10**, 1377-1383 (2005).
70. C. Kleinstreuer, J. Li, J. Koo, Microfluidics of nano-drug delivery. *Int J Heat Mass Tran* **51**, 5590-5597 (2008).

71. S. Saleh-Lakha, J. T. Trevors, Perspective: Microfluidic applications in microbiology. *J Microbiol Meth* **82**, 108-111 (2010).
72. D. B. Weibel, W. R. DiLuzio, G. M. Whitesides, Microfabrication meets microbiology. *Nat Rev Microbiol* **5**, 209-218 (2007).
73. A. H. C. Ng, U. Uddayasankar, A. R. Wheeler, Immunoassays in microfluidic systems. *Anal Bioanal Chem* **397**, 991-1007 (2010).
74. S. Baratchi *et al*., Immunology on chip: Promises and opportunities. *Biotechnol Adv* **32**, 333-346 (2014).
75. T. M. Squires, S. R. Quake, Microfluidics: Fluid physics at the nanoliter scale. *Rev Mod Phys* **77**, 977-1026 (2005).
76. B. E. Bernstein *et al*., Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181 (2005).
77. P. J. Park, ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680 (2009).
78. P. Collas, The Current State of Chromatin Immunoprecipitation. *Mol Biotechnol* **45**, 87-100 (2010).
79. A. Meissner *et al*., Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* **33**, 5868-5877 (2005).
80. T. A. Down *et al*., A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* **26**, 779-785 (2008).
81. F. V. Jacinto, E. Ballestar, M. Esteller, Methyl-DNA immunoprecipitation (MeDIP): Hunting down the DNA methylome. *Biotechniques* **44**, 35-43 (2008).
82. G. E. Crawford *et al*., Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome research* **16**, 123-131 (2006).
83. D. E. Schones *et al*., Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-898 (2008).
84. S. J. Cokus *et al*., Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215-219 (2008).
85. P. W. Laird, Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* **11**, 191-203 (2010).
86. E. Lieberman-Aiden *et al*., Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289-293 (2009).
87. M. J. Fullwood *et al*., An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462**, 58 (2009).
88. N. Li *et al*., Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods* **52**, 203-212 (2010).
89. D. Serre, B. H. Lee, A. H. Ting, MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res* **38**, 391-399 (2010).
90. A. K. Maunakea *et al*., Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253-257 (2010).
91. K. J. Gaulton *et al*., A map of open chromatin in human pancreatic islets. *Nat Genet* **42**, 255-259 (2010).
92. I. M. Derrington *et al*., Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A* **107**, 16060-16065 (2010).
93. M. Jain *et al*., Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **36**, 338-345 (2018).
94. D. R. Garalde *et al*., Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201-206 (2018).

95. H. S. Rhee, B. F. Pugh, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408-1419 (2011).
96. O. Harismendy *et al*., 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* **470**, 264-268 (2011).
97. C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, H. Y. Chang, Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**, 667-678 (2011).
98. C. Chu, J. Quinn, H. Y. Chang, Chromatin isolation by RNA purification (ChIRP). *J Vis Exp*, e3912 (2012).
99. F. Miura, Y. Enomoto, R. Dairiki, T. Ito, Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* **40**, e136 (2012).
100. M. J. Booth *et al*., Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336**, 934-937 (2012).
101. M. Yu *et al*., Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc* **7**, 2159-2170 (2012).
102. J. Eid *et al*., Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009).
103. H. S. Guo *et al*., Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Research* **23**, 2126-2135 (2013).
104. I. C. Macaulay, C. P. Ponting, T. Voet, Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet* **33**, 155-168 (2017).
105. H. S. Guo *et al*., Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nature Protocols* **10**, 645-659 (2015).
106. C. X. Song *et al*., Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678-691 (2013).
107. L. Shen *et al*., Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692-706 (2013).
108. T. Nagano *et al*., Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
109. V. Ramani *et al*., Massively multiplex single-cell Hi-C. *Nat Methods* **14**, 263-266 (2017).
110. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
111. Q. Wang *et al*., Tagmentation-based whole-genome bisulfite sequencing. *Nature Protocols* **8**, 2022-2032 (2013).
112. S. A. Smallwood *et al*., Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* **11**, 817-820 (2014).
113. A. Rotem *et al*., Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature Biotechnology* **33**, 1165-U1191 (2015).
114. J. Kind *et al*., Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-147 (2015).
115. W. Jin *et al*., Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142-146 (2015).
116. J. D. Buenostro *et al*., Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-U264 (2015).
117. D. A. Cusanovich *et al*., Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914 (2015).

118. M. Farlik *et al*., Single-Cell DNA Methylome Sequencing and Bioinformatic Inference of Epigenomic Cell-State Dynamics. *Cell Rep* **10**, 1386-1397 (2015).
119. C. Schmidl, A. F. Rendeiro, N. C. Sheffield, C. J. N. m. Bock, ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. **12**, 963 (2015).
120. D. Mooijman, S. S. Dey, J. C. Boisset, N. Crosetto, A. van Oudenaarden, Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat Biotechnol* **34**, 852–856 (2016).
121. Y. Hu *et al*., Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol* **17**, 88 (2016).
122. C. Angermueller *et al*., Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* **13**, 229-232 (2016).
123. Y. Hou *et al*., Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* **26**, 304-319 (2016).
124. C. Zhu *et al*., Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* **20**, 720-731. e725 (2017).
125. X. Wu, A. Inoue, T. Suzuki, Y. Zhang, Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. *Genes Dev* **31**, 511-523 (2017).
126. R. M. Mulqueen *et al*., Highly scalable generation of DNA methylation profiles in single cells. *Nat Biotechnol* **36**, 428-431 (2018).
127. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature protocols* **13**, 1006 (2018).
128. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
129. S. J. Clark *et al*., Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). **12**, 534 (2017).
130. S. Ma *et al*., Cell-type-specific brain methylomes profiled via ultralow-input microfluidics. *Nat Biomed Eng* **2**, 183-194 (2018).
131. A. Harada *et al*., A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat Cell Biol* **21**, 287-296 (2019).
132. J. Shen *et al*., H3K4me3 epigenomic landscape derived from ChIP-Seq of 1,000 mouse early embryonic cells. *Cell Res* **25**, 143-147 (2015).
133. S. Ma, Y. P. Hsieh, J. Ma, C. Lu, Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. *Science Advances* **4**, eaar8187 (2018).
134. R. Dainese *et al*., A parallelized, automated platform enabling individual or sequential ChIP of histone marks and transcription factors. *Proc Natl Acad Sci U S A* **117**, 13828-13838 (2020).
135. Y. Zhu, Z. Cao, C. Lu, Microfluidic MeDIP-seq for low-input methylomic analysis of mammary tumorigenesis in mice. *Analyst* **144**, 1904-1915 (2019).
136. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220 (2019).
137. Y. Xu *et al*., A droplet microfluidic platform for efficient enzymatic chromatin digestion enables robust determination of nucleosome positioning. *Lab Chip* **18**, 2583-2592 (2018).
138. J. D. Buenostro *et al*., Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490 (2015).
139. A. T. Satpathy *et al*., Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936 (2019).

140. C. A. Lareau *et al.*, Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat Biotechnol* **37**, 916-924 (2019).
141. S. Chen, B. B. Lake, K. Zhang, High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* **37**, 1452-1457 (2019).
142. M. Cao, J. Zhao, G. Hu, Genome-wide methods for investigating long noncoding RNAs. *Biomed Pharmacother* **111**, 395-401 (2019).
143. M. Wu, M. Piccini, C.-Y. Koh, K. S. Lam, A. K. Singh, Single cell microRNA analysis using microfluidic flow cytometry. *Plos One* **8**, e55044 (2013).
144. M. Y. Wu *et al.*, Microfluidically-unified cell culture, sample preparation, imaging and flow cytometry for measurement of cell signaling pathways with single cell resolution. *Lab Chip* **12**, 2823-2831 (2012).
145. C. Novara *et al.*, SERS-active metal-dielectric nanostructures integrated in microfluidic devices for label-free quantitative detection of miRNA. *Faraday Discuss* **205**, 271-289 (2017).
146. X. Zhou *et al.*, in *Next-Generation MicroRNA Expression Profiling Technology*. (Springer, 2012), vol. 822, pp. 153-182.
147. Y. Chen, J. A. Gelfond, L. M. McManus, P. K. Shireman, Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *Bmc Genomics* **10**, 407 (2009).
148. A. I. Pogue, W. J. Lukiw, Up-regulated Pro-inflammatory MicroRNAs (miRNAs) in Alzheimer's disease (AD) and Age-Related Macular Degeneration (AMD). *Cell Mol Neurobiol* **38**, 1021-1031 (2018).
149. C. Wei *et al.*, Identification of the Differentially Expressed microRNAs Involved in Cleft Palate Induced by Retinoic Acid (RA) in Mouse Model. *J Hard Tissue Biol* **27**, 243-249 (2018).
150. W. Zhou *et al.*, microRNA-183 is involved in the differentiation and regeneration of Notch signaling-prohibited hair cells from mouse cochlea. *Molecular medicine reports* **18**, 1253-1262 (2018).
151. B. Lazar *et al.*, Comparison of the MicroRNA Expression Profiles of Male and Female Avian Primordial Germ Cell Lines. *Stem Cells Int* **2018**, 1780679 (2018).
152. F. Mar-Aguilar *et al.*, Identification and characterization of microRNAs from Entamoeba histolytica HM1-IMSS. *Plos One* **8**, e68202 (2013).
153. J. D. Tian *et al.*, Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050-1054 (2004).
154. P. Takousis, Analysis of Micro-RNA Expression by qPCR on a Microfluidics Platform for Alzheimer's Disease Biomarker Discovery. *Methods Mol Biol* **1750**, 283-292 (2018).
155. J. S. Jang *et al.*, Quantitative miRNA expression analysis using fluidigm microfluidics dynamic arrays. *Bmc Genomics* **12**, 144 (2011).
156. M. Kara *et al.*, Differential expressions of cancer-associated genes and their regulatory miRNAs in colorectal carcinoma. *Gene* **567**, 81-86 (2015).
157. M. Mensah *et al.*, MicroRNA based liquid biopsy: the experience of the plasma miRNA signature classifier (MSC) for lung cancer screening. *JoVE (Journal of Visualized Experiments)*, e56326 (2017).
158. S. Rizzo *et al.*, Analysis of miRNA expression profile induced by short term starvation in breast cancer cells treated with doxorubicin. *Oncotarget* **8**, 71924-71932 (2017).
159. Z. Kanaan *et al.*, A Plasma MicroRNA Panel for Detection of Colorectal Adenomas A Step Toward More Precise Screening for Colorectal Cancer. *Annals of Surgery* **258**, 400-408 (2013).

160. N. Joly-Tonetti, J. Viñuelas, O. Gandrillon, J. Lamartine, Differential miRNA expression profiles in proliferating or differentiated keratinocytes in response to gamma irradiation. *Bmc Genomics* **14**, 184 (2013).
161. Z. Kanaan *et al*., Plasma miR-21: a potential diagnostic marker of colorectal cancer. *Annals of surgery* **256**, 544-551 (2012).
162. W. Zeng *et al*., Single-nucleus RNA-seq of differentiating human myoblasts reveals the extent of fate heterogeneity. *Nucleic Acids Res* **44**, e158 (2016).
163. M. Håkanson, E. Cukierman, M. Charnley, Miniaturized pre-clinical cancer models as research and diagnostic tools. *Advanced drug delivery reviews* **69**, 52-66 (2014).
164. J. A. DiMasi, R. W. Hansen, H. G. Grabowski, The price of innovation: new estimates of drug development costs. *Journal of health economics* **22**, 151-185 (2003).
165. M. C. Cox, L. M. Reese, L. R. Bickford, S. S. Verbridge, Toward the broad adoption of 3D tumor models in the cancer drug pipeline. *ACS Biomaterials Science & Engineering* **1**, 877-894 (2015).
166. J. Rodon *et al*., Challenges in initiating and conducting personalized cancer therapy trials: perspectives from WINTHER, a Worldwide Innovative Network (WIN) Consortium trial. *Annals of Oncology* **26**, 1791-1798 (2015).
167. M. Gerlinger *et al*., Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England journal of medicine* **366**, 883-892 (2012).
168. M. Qian, D. C. Wang, H. Chen, Y. Cheng, in *Seminars in cell & developmental biology*. (Elsevier, 2017), vol. 64, pp. 143-149.
169. N. E. Navin, J. Hicks, Tracing the tumor lineage. *Molecular oncology* **4**, 267-283 (2010).
170. G. Ciriello *et al*., Emerging landscape of oncogenic signatures across human cancers. *Nature genetics* **45**, 1127 (2013).
171. J. Munoz, C. Swanton, R. Kurzrock, Molecular profiling and the reclassification of cancer: divide and conquer. *Am Soc Clin Oncol Educ Book* **2013**, 127-134 (2013).
172. P. A. Jones, S. B. Baylin, The epigenomics of cancer. *Cell* **128**, 683-692 (2007).
173. B. M. Turner, Reading signals on the nucleosome with a new nomenclature for modified histones. *Nature structural & molecular biology* **12**, 110 (2005).
174. M. D. Shahbazian, M. Grunstein, Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* **76**, 75-100 (2007).
175. Y. Z. Kim, Altered histone modifications in gliomas. *Brain tumor research and treatment* **2**, 7-21 (2014).
176. Y. B. Khotskaya, G. B. Mills, K. R. Mills Shaw, Next-generation sequencing and result interpretation in clinical oncology: challenges of personalized cancer therapy. *Annual review of medicine* **68**, 113-125 (2017).
177. M. Farlik *et al*., Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell reports* **10**, 1386-1397 (2015).
178. S. Kurdistani, Histone modifications as markers of cancer prognosis: a cellular view. *British journal of cancer* **97**, 1 (2007).
179. D. B. Seligson *et al*., Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* **435**, 1262 (2005).
180. G. J. Kitange *et al*., Inhibition of histone deacetylation potentiates the evolution of acquired temozolomide resistance linked to MGMT upregulation in glioblastoma xenografts. *Clinical Cancer Research*, clincanres. 0560.2012 (2012).
181. J. M. Westcott *et al*., An epigenetically distinct breast cancer cell subpopulation promotes collective invasion. *The Journal of clinical investigation* **125**, 1927-1943 (2015).

182. Y. Li, C. B. Tang, K. A. Kilian, Matrix Mechanics Influence Fibroblast–Myofibroblast Transition by Directing the Localization of Histone Deacetylase 4. *Cellular and Molecular Bioengineering* **10**, 405-415 (2017).
183. A. Mullard. (Nature Publishing Group, 2016).
184. R. M. Sutherland, J. A. McCredie, W. R. Inch, Growth of multicell spheroids in tissue culture as a model of nodular carcinomas. *Journal of the National Cancer Institute* **46**, 113-120 (1971).
185. D. Theodorescu, C. Sheehan, R. Kerbel, TGF-β gene expression depends on tissue architecture. *In Vitro Cellular & Developmental Biology-Animal* **29**, 105-108 (1993).
186. B. S. Croix *et al.*, Impact of the cyclin–dependent kinase inhibitor p27Kip1 on resistance of tumor cells to anticancer agents. *Nature medicine* **2**, 1204-1210 (1996).
187. R. Bonavia, W. K. Cavenee, F. B. Furnari, Heterogeneity maintenance in glioblastoma: a social network. *Cancer research*, (2011).
188. A. L. Harris, Hypoxia—a key regulatory factor in tumour growth. *Nature Reviews Cancer* **2**, 38 (2002).
189. Z. Cao, C. Chen, B. He, K. Tan, C. Lu, A microfluidic device for epigenomic profiling using 100 cells. *Nature methods* **12**, 959 (2015).
190. M. C. Cox, C. Deng, L. Naler, C. Lu, S. S. Verbridge, Effects of culture condition on epigenomic profiles of brain tumor cells. *ACS Biomaterials Science & Engineering* **5**, 1544-1552 (2019).
191. V. L. Cross *et al.*, Dense type I collagen matrices that support cellular remodeling and microfabrication for studies of tumor angiogenesis and vasculogenesis in vitro. *Biomaterials* **31**, 8596-8607 (2010).
192. C. S. Ross-Innes *et al.*, Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389 (2012).
193. H. Heberle, G. V. Meirelles, F. R. da Silva, G. P. Telles, R. Minghim, InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC bioinformatics* **16**, 169 (2015).
194. P. D. Thomas *et al.*, PANTHER: a library of protein families and subfamilies indexed by function. *Genome research* **13**, 2129-2141 (2003).
195. A. Kamburov, C. Wierling, H. Lehrach, R. Herwig, ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic acids research* **37**, D623-D628 (2008).
196. G. Yu, L.-G. Wang, Q.-Y. He, ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).
197. H. Page, P. Flood, E. G. Reynaud, Three-dimensional tissue cultures: current trends and beyond. *Cell and tissue research* **352**, 123-131 (2013).
198. S. M. Ehsan, S. C. George, Nonsteady state oxygen transport in engineered tissue: implications for design. *Tissue Engineering Part A* **19**, 1433-1442 (2013).
199. I. Cho, M. J. Blaser, The human microbiome: at the interface of health and disease. *Nat Rev Genet* **13**, 260-270 (2012).
200. M. A. Casasanta *et al.*, Fusobacterium nucleatum host-cell binding and invasion induces IL-8 and CXCL1 secretion that drives colorectal cancer cell migration. *Sci Signal* **13**, (2020).
201. S. S. Verbridge *et al.*, Oxygen-controlled three-dimensional cultures to analyze tumor angiogenesis. *Tissue Eng Part A* **16**, 2133-2141 (2010).
202. S. S. Verbridge *et al.*, Physicochemical regulation of endothelial sprouting in a 3D microfluidic angiogenesis model. *J Biomed Mater Res A* **101**, 2948-2956 (2013).
203. Y. Hosseini, M. Agah, S. S. Verbridge, Endothelial cell sensing, restructuring, and invasion in collagen hydrogel structures. *Integr Biol (Camb)* **7**, 1432-1441 (2015).

204. S. Ma, Y.-P. Hsieh, J. Ma, C. J. S. a. Lu, Low-input and multiplexed microfluidic assay reveals epigenomic variation across cerebellum and prefrontal cortex. **4**, eaar8187 (2018).
205. M. A. Casasanta *et al.*, A chemical and biological toolbox for Type Vd secretion: Characterization of the phospholipase A1 autotransporter FplA from Fusobacterium nucleatum. *J Biol Chem* **292**, 20240-20254 (2017).
206. Y. Zhang, C. Lee, S. Geng, L. W. Li, Enhanced tumor immune surveillance through neutrophil reprogramming due to Tollip deficiency. *Jci Insight* **4**, (2019).
207. X. Gao *et al.*, Acetate functions as an epigenetic metabolite to promote lipid synthesis under hypoxia. *Nat Commun* **7**, (2016).
208. A. Sottoriva *et al.*, Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences* **110**, 4009-4014 (2013).
209. C. Jones, S. J. Baker, Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma. *Nature reviews Cancer* **14**, 651 (2014).
210. S. Mack *et al.*, Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445 (2014).
211. P. Lu, V. M. Weaver, Z. Werb, The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol* **196**, 395-406 (2012).
212. V. M. Weaver *et al.*, β4 integrin-dependent formation of polarized three-dimensional architecture confers resistance to apoptosis in normal and malignant mammary epithelium. *Cancer cell* **2**, 205-216 (2002).
213. K.-V. Nguyen-Ngoc *et al.*, ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. *Proceedings of the National Academy of Sciences*, 201212834 (2012).
214. H. Axelson, E. Fredlund, M. Ovenberger, G. Landberg, S. Påhlman, in *Seminars in cell & developmental biology*. (Elsevier, 2005), vol. 16, pp. 554-563.
215. L. Y. Joanne *et al.*, Heterogeneous vascular dependence of tumor cell populations. *The American journal of pathology* **158**, 1325-1334 (2001).
216. P. DelNero *et al.*, 3D culture broadly regulates tumor cell hypoxia response and angiogenesis via pro-inflammatory pathways. *Biomaterials* **55**, 110-118 (2015).
217. B. Dai *et al.*, Overexpressed KDM5B is associated with the progression of glioma and promotes glioma cell growth via downregulating p21. *Biochemical and biophysical research communications* **454**, 221-227 (2014).
218. J. Le Beyec *et al.*, Cell shape regulates global histone acetylation in human mammary epithelial cells. *Experimental cell research* **313**, 3066-3075 (2007).
219. E. Voronov, Y. Carmi, R. N. Apte, The role IL-1 in tumor-mediated angiogenesis. *Frontiers in physiology* **5**, 114 (2014).
220. A. Birgersdotter, R. Sandberg, I. Ernberg, in *Seminars in cancer biology*. (Elsevier, 2005), vol. 15, pp. 405-412.
221. R. Edmondson, J. J. Broglie, A. F. Adcock, L. Yang, Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay and drug development technologies* **12**, 207-218 (2014).
222. S. Lee *et al.*, Multi-dimensional histone methylations for coordinated regulation of gene expression under hypoxia. *Nucleic acids research* **45**, 11643-11657 (2017).
223. L. Qin *et al.*, Systemic LPS causes chronic neuroinflammation and progressive neurodegeneration. *Glia* **55**, 453-462 (2007).
224. A. Barski *et al.*, High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837 (2007).

225. A. A. Alizadeh *et al*., Toward understanding and exploiting tumor heterogeneity. *Nature medicine* **21**, 846 (2015).
226. S. M. Pupa, S. Ménard, S. Forti, E. Tagliabue, New insights into the role of extracellular matrix during tumor onset and progression. *Journal of cellular physiology* **192**, 259-267 (2002).
227. H. Tekin *et al*., Effects of 3D culturing conditions on the transcriptomic profile of stem-cell-derived neurons. *Nature Biomedical Engineering*, 1 (2018).
228. M. Lawrence, S. Daujat, R. Schneider, Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet* **32**, 42-56 (2016).
229. K. Hyun, J. Jeon, K. Park, J. Kim, Writing, erasing and reading histone lysine methylations. *Exp Mol Med* **49**, e324 (2017).
230. P. A. Jones, J. P. J. Issa, S. Baylin, Targeting the cancer epigenome for therapy. *Nat Rev Genet* **17**, 630-641 (2016).
231. S. G. Landt *et al*., ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* **22**, 1813-1831 (2012).
232. D. Lara-Astiaso *et al*., Chromatin state dynamics during blood formation. *Science* **345**, 943-949 (2014).
233. B. H. Zhu *et al*., MOWChIP-seq for low-input and multiplexed profiling of genome-wide histone modifications. *Nat Protoc* **14**, 3366-3394 (2019).
234. M. C. Cox, C. Y. Deng, L. B. Naler, C. Lu, S. S. Verbridge, Effects of Culture Condition on Epigenomic Profiles of Brain Tumor Cells. *Acs Biomater Sci Eng* **5**, 1544-1552 (2019).
235. J. A. Dahl *et al*., Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552 (2016).
236. B. J. Zhang *et al*., Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557 (2016).
237. H. Zheng *et al*., Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Mol Cell* **63**, 1066-1079 (2016).
238. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019 (2018).
239. D. H. Janssens *et al*., Automated in situ chromatin profiling efficiently resolves cell types and gene regulatory programs. *Epigenet Chromatin* **11**, 1-14 (2018).
240. M. P. Meers, T. D. Bryson, J. G. Henikoff, S. Henikoff, Improved CUT&RUN chromatin profiling tools. *Elife* **8**, e46314 (2019).
241. M. Schmid, T. Durussel, U. K. Laemmli, ChIC and ChEC: genomic mapping of chromatin proteins. *Mol Cell* **16**, 147-157 (2004).
242. M. A. Zarnegar, F. Reinitz, A. M. Newman, M. F. Clarke, Targeted chromatin ligation, a robust epigenetic profiling technique for small cell numbers. *Nucleic Acids Res* **45**, e153-e153 (2017).
243. C. Schmidl, A. F. Rendeiro, N. C. Sheffield, C. Bock, ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* **12**, 963-965 (2015).
244. S. Amini *et al*., Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat Genet* **46**, 1343-1349 (2014).
245. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
246. Y. Zhang *et al*., Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, 1-9 (2008).
247. F. Ramirez *et al*., deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-W165 (2016).

248. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
249. S. Herculano-Houzel, B. Mota, R. Lent, Cellular scaling rules for rodent brains. *Proc Natl Acad Sci U S A* **103**, 12138-12143 (2006).
250. M. T. Kirkcaldie *et al*., Neurofilament triplet proteins are restricted to a subset of neurons in the rat neocortex. *J Chem Neuroanat* **24**, 163-171 (2002).
251. M. T. Kirkcaldie, Neocortex. 10.1016/B1978-1010-1012-369497-369493.310004-369494 (2012).
252. X. Liu *et al*., Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562 (2016).
253. M. Albert *et al*., Epigenome profiling and editing of neocortical progenitor cells during development. *EMBO J* **36**, 2642-2658 (2017).
254. E. Lubeck, L. Cai, Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods* **9**, 743-748 (2012).
255. E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* **11**, 360-361 (2014).
256. J. H. Lee *et al*., Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363 (2014).
257. S. G. Rodriques *et al*., Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
258. S. Maniatis *et al*., Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364**, 89-93 (2019).
259. E. Berglund *et al*., Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* **9**, 2419 (2018).
260. K. Thrane, H. Eriksson, J. Maaskola, J. Hansson, J. Lundeberg, Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. *Cancer Res* **78**, 5970-5979 (2018).
261. J. P. Junker *et al*., Genome-wide RNA Tomography in the zebrafish embryo. *Cell* **159**, 662-675 (2014).
262. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).
263. Y. Jiang, A. Matevossian, H. S. Huang, J. Straubhaar, S. Akbarian, Isolation of neuronal chromatin from brain tissue. *BMC Neurosci* **9**, 42 (2008).
264. S. Picelli *et al*., Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).
265. S. Picelli *et al*., Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033-2040 (2014).
266. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).
267. A. Tarasov, A. J. Vilella, E. Cuppen, I. J. Nijman, P. Prins, Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034 (2015).
268. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
269. E. B. Stovner, P. Saetrom, epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392-4393 (2019).
270. F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).
271. C. Y. McLean *et al*., GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).

272. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
273. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
274. S. Herculano-Houzel, The glia/neuron ratio: how it varies uniformly across brain structures and species and what that means for brain physiology and evolution. *Glia* **62**, 1377-1391 (2014).
275. S. Peng, Y. Zhang, J. Zhang, H. Wang, B. Ren, Glutamate receptors and signal transduction in learning and memory. *Mol Biol Rep* **38**, 453-460 (2011).
276. C. González-Castillo, D. Ortuño-Sahagún, C. Guzmán-Brambila, M. Pallàs, A. E. Rojas-Mayorquín, Pleiotrophin as a central nervous system neuromodulator, evidences from the hippocampus. *Frontiers in cellular neuroscience* **8**, 443 (2015).

PUBLICATIONS

(1) Zhu, B., Wang, Z., Ainsworth R., Sierra, S., **Deng, C**., Callado, L.F., Meana, J.J., Wang, W., Lu, C., Gonzalez-Maeso, J., Antipsychotic effect on cell-specific histone modification landscapes in the frontal cortex of schizophrenic subjects, In preparation.

(2) **Deng, C**., Xiao, Y., Liu, Z., Zhu, B., Zhou, Z., Naler, L.B., Yao, D., Lu, C. Spatially resolved epigenomics in mouse neocortex cluster sensory perception related loci, In preparation

(3) **Deng, C**., Murphy, T.W., Zhang, Q., Naler, L.B., Xu, A., Lu, C., Multiplexed and ultralow-input ChIP-seq enabled by tagmentation-based indexing and facile microfluidics. ***Analytical Chemistry***. DOI: 10.1021/acs.analchem.0c02550

(4) Cox, M.C.*, **Deng, C.*(co-first author)**, Naler, L.B., Lu, C. and Verbridge, S.S., 2019. Effects of culture condition on epigenomic profiles of brain tumor cells. ***ACS biomaterials science & engineering***, 5(3), pp.1544-1552. DOI: 10.1021/acsbiomaterials.9b00161

(5) **Deng, C**., Naler, L.B. and Lu, C., 2019. Microfluidic epigenomic mapping technologies for precision medicine. ***Lab on a Chip***, 19(16), pp.2630-2650. DOI: 10.1039/C9LC00407F

(6) Hao, J., **Deng, C**., Wang, X., Hu, J., 2014. Study of dispersion character of mercerized pulp. ***BioResources***, 9(3), 4336-4342.