UNIVERSITY OF CALIFORNIA SAN DIEGO

Optimization and Application of State-of-art Techniques for Studying Epigenetic Changes

A Thesis submitted in partial satisfaction of the

requirements for the degree Master of Science

in

Chemistry

by

Jia Xu

Committee in charge:

Professor Wei Wang, Chair
Professor Pieter Dorrestein
Professor Alexis Komor

2019

The Thesis of Jia Xu is approved, and is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2019

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Professor Wei Wang at University of California San Diego. The door to Prof. Wang office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work but steered me in the right the direction whenever he thought I needed it.

I would also like to thank my committee members, Professor Pieter Dorrestein and Professor Alexis Komor, for their advice and participation. I would also like to give thanks to all the professors who gave me lectures throughout my studies at UCSD

I would also like to acknowledge Dr. Zhao Chen and Dr. Ying Zhao in our lab at University of California San Diego as the second reader of this thesis, and I am gratefully indebted to their very valuable comments on this thesis.

Finally, I must express my very profound gratitude to my parents and to my friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

ABSTRACT OF THE THESIS

Optimization and Application of State-of-art Techniques for Studying Epigenetic Changes

by

Jia Xu

Master of Science in Chemistry

University of California San Diego, 2019

Professor Wei Wang, Chair

I have been working on two projects through my master thesis. The first project described in Chapter 1 is focusing on studying the regulatory mechanisms of specific methylated DNA motifs using optimized CAPTURE technique. DNA motifs are the sequences that are presumed to have biological functions in gene regulatory, however,

the regulatory compositions of most motifs remain unknown. CAPTURE is a perfect tool that we can optimize and apply on our interest methylated DNA motifs.

Epigenetic mechanisms of gene regulation such as DNA methylation and chromatin modification are playing an important role in the development of pathology have been demonstrated, but the epigenetic regulation of gout disease is unclear. The second project in Chapter 2, we modified and applied Atac-seq and RNA-seq to study the difference of open chromatin pattern and gene expression respectively between gout patients and healthy individuals.

Chapter 1

Overall Introduction

Epigenetic is a study of heritable changes in phenotype without the changes in the underlying DNA sequence. Epigenetic changes, including DNA methylation and histone modification, are identified to influence gene expression. DNA methylation and histone modification can regulate gene expression by controlling the binding of regulatory proteins to specific locus or altering DNA accessibility and chromatin structures, respectively. Many researches demonstrated that epigenetics play a crucial role in many human disorders and fatal diseases.

For my master thesis, I was working on understanding regulatory mechanisms by which gene expression can be altered through epigenetic changes using state-of-art techniques. The first project is to study the fundamental mechanisms of DNA methylation of DNA motifs using CAPTURE technique and the second project is to study the difference of gene expression and chromatin pattern between gout patients and healthy individuals and identify potential druggable targets.

Chapter 2

2.1 Introduction

Cis-regulatory elements (CREs) are regions of non-coding DNA sequence which play crucial roles in regulating gene expression. Enhancers and promoters are important CREs. A research showed that the mutation in the core enhancer/promoter II region of hepatitis B virus (HBV) genome in immunosuppressed patients will cause severe liver disease[1]. Another paper indicated that colorectal cancer patients are overexpress a histone lysine methyltransferase, which is regulated by an enhancer[2]. More and more cases are showing that CREs are relative to gene regulatory. From established data, only a small percentage genome is responsible for coding proteins, so the majority of genome are non-coding DNA sequence which may be identified as CREs. Until now, a lot of CREs are still waiting for us to recognize and study. Enhancers and promoters contain motifs which are the specific sequences that are conjectured to have biological functions by interacting with proteins, RNAs and long-range DNAs[34] (Figure 1.1.1). CREs functions in gene expression by recruiting the regulatory proteins such as DNA methyltransferases (DNMFs) which lead to epigenetic modifications (such as methylation) of DNA motifs[5]. A recent study demonstrated that methylated CREs are responsible for recruitment of transcription factors such as Kruppel-like factor 4 (KLF4) which are functions as cancer driven gene[6].

Chromatin immunoprecipitation coupled to high throughput sequencing (ChIP-seq) is a powerful technique to study CREs[7]. Antibodies specific to the target protein are used to immune-precipitate the DNA-protein complex. This technique can start with a small

number of cells. However, the success of ChIP-seq experiments heavily relies on the availability and specificity of antibodies for enriching the target protein[8]. Most of the currently commercially available antibodies fall in the protein groups that their functions are well studied. Therefore, it limits our capacity to utilize ChIP-seq to investigate proteins less understood. Targeted bisulfite sequencing is a tool to study locus specific DNA methylation[9]. The advantage of this technique is that it has low cost by focusing on specific region rather than whole genome. The limitation of this technique is that it cannot provide any information on the interacting protein on the targeted DNA region[10].

Recently, a new technique, CRISPR affinity purification in situ of regulatory elements (CAPTURE) has been developed to study DNA-binding proteins based on Cas9/CRISPR system[8] (Figure 1.1.2). In CRISPR system, single guide RNA (sgRNA) direct Cas9 to bind a target genomic sequence followed by Cas9 cutting. CAPTURE uses deactivated Cas9 (dCas9) and engineered an N-terminal FLAG and biotin-acceptor-site (FB)-tag on it, which makes the enzyme incapable to cut the sequence after binding. In cells, with the presence of biotinylated dCas9 by biotin ligase BirA and sequence-specific sgRNA, the macromolecular on dCas9-binding specific locus can be isolated by high-affinity streptavidin purification. The purified protein, DNA and RNA complexes can be identified and analyzed individually using mass-spectrometry-based proteomics, high throughput sequencing respectively. The DNA methylation level can also be determined by analyzing the purified DNA. This technique is locus specific and can be used to study multiple elements of one DNA region simultaneously.

Previous study from our lab identified DNA motifs are relative to the regulation of DNA methylation level through analyzing whole-genome bisulfite sequencing (WGBS)

data in 34 cell lines[5]. However, the mechanisms of how these motifs regulate DNA methylation level and protein binding remain puzzled and attractive. CAPTURE is a perfect technique for validating our computational results and profiling the complete composition of the interacting complexes with the methylated DNA motif simultaneously.

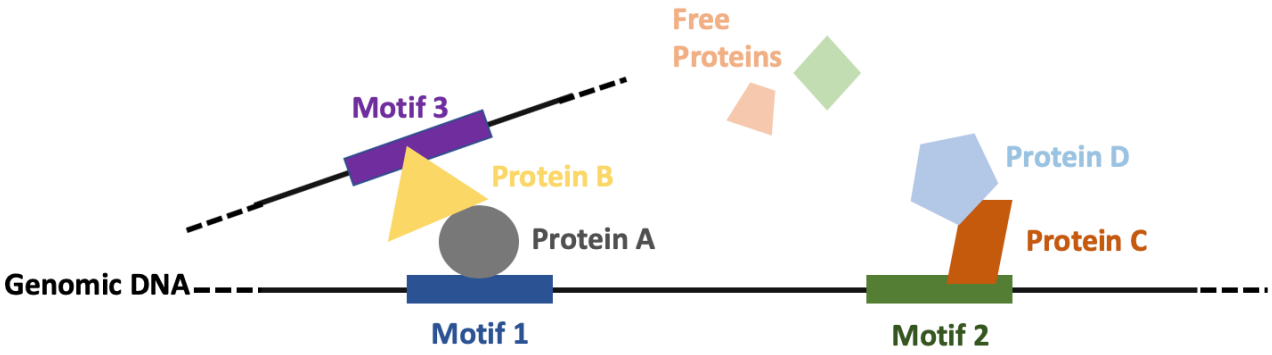

Figure 1.1.1 Schematic of Protein Recruitment by Motifs

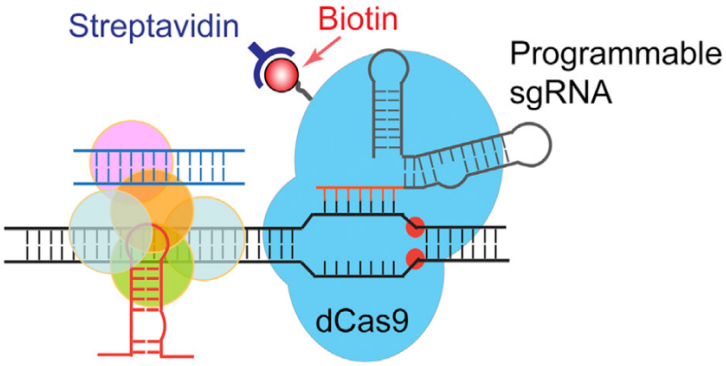

Figure 1.1.2 Schematic of CAPTURE System (Adapted from "In Situ Capture of Chromatin Interactions by Biotinylated dCas9" by X. Liu, 2017, Cell, 170(5), p1. Copyright [2017] by Elsevier Inc.)

2.2 Results

2.2.1 Establishment of Stable Cell Line

We choose the K562 cell line as it is one of the 34 cell lines from the methylation motif analysis. The main components of CAPTURE are dCas9, biotin ligase BirA and sgRNA (Figure 1.2.1). Thus, the first step is to generate stable K562 cells expressing dCas9 and BirA proteins. The function of BirA is to ligate biotin to dCas9 protein to form dCas9 complexes (Figure 1.2.2).

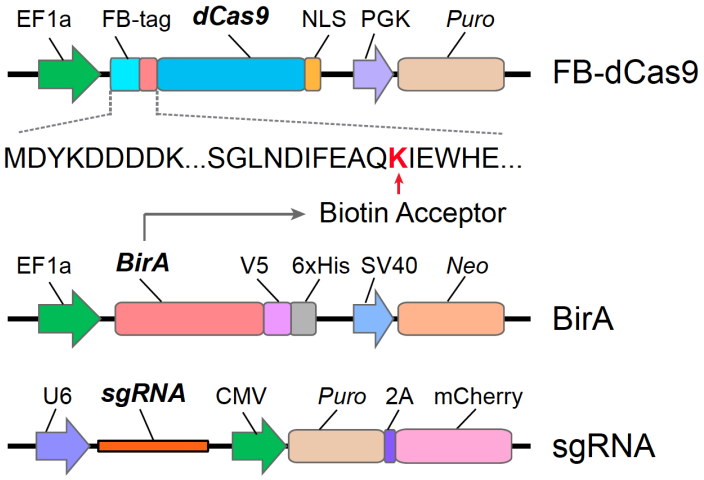

Figure 1.2.1 Three Main Components of the CAPTURE System: FB-tagged dCas9, a biotin ligase BirA, and target-specific sgRNA (Adapted from "In Situ Capture of Chromatin Interactions by Biotinylated dCas9" by X. Liu, 2017, Cell, 170(5), p3. Copyright [2017] by Elsevier Inc.)

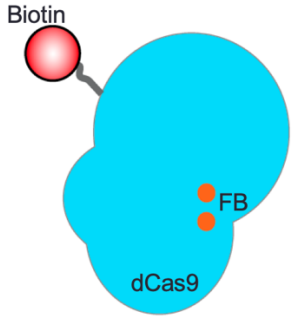

Figure 1.2.2 dCas9 Complexes in CAPTURE System

2.2.2 Transfection of Plasmids

Liu et al. 2017 used Nucleofection method to deliver the pEF1a-FB-dCas9 (Addgene #100547) and pEF1a-BirA-V5 (Addgene #100548) plasmid containing FLAG tagged dCas9 and V5 tagged BirA gene respectively (Figure 1.2.3) followed by puromycin and G418 double-selection for plasmid-integrated cells expressing both proteins. Since our lab only has Gene Pulser Xcel™ Electroporation System (Bio-Rad), we tested electroporation conditions by transfecting K562 cells with the plasmids that contains green fluorescent protein (GFP) sequence and followed by fluorescence-activated cell sorter (FACS) to quantify the transfection efficiency. We tested 5 electroporation conditions comparing with one control which was the cells without transfection (Table 1.2.1).

The 1st condition (320V, 200uF) has the highest live cells percentage and highest GFP positive percentage according to Table 1.2.1, so we applied this condition on transfecting K562 with pEF1a-FB-dCas9 and pEF1a-BirA-V5 plasmids and followed by cell selection with puromycin and G418 antibiotics. Then western blot assay is used for testing protein expression. The western blot results (Figure 1.2.4) showed that K562 cells only express BirA but not dCas9 protein. The CAPTURE system cannot be established without dCas9.

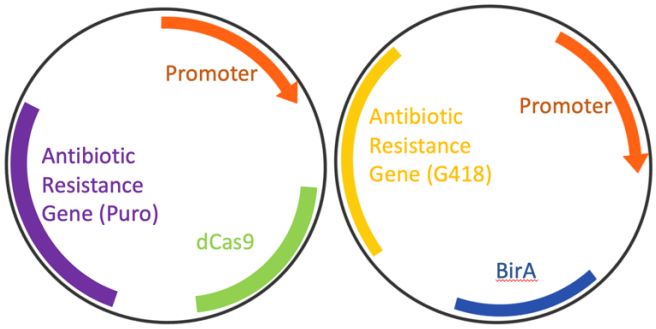

Figure 1.2.3 Main Functional Regions of pEF1a-FB-dCas9 and pEF1a-BirA-V5 Plasmid

Table 1.2.1 Detailed Results of Electroporation Conditions

Condition	Voltage (V)	Capacitance (uF)	Cell Number(M)	Plasmid (ug)	DPBS Volume (uL)	GFP%	Live cells%
1	320	200	20	20	400	48	16.3
2	200	950	20	20	400	14	21
3	270	950	20	20	400	47	1.7
4	320	950	20	20	400	0	1.53
5	350	950	20	20	400	0	2.16
6	0	0	20	0	0	0	47.3

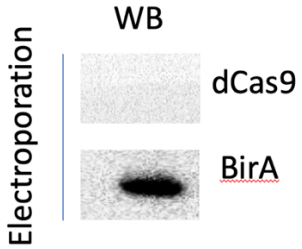

Figure 1.2.4 Western Blot Result of Electro-transfection

2.2.3 Lentiviral Transduction

Study shows that transfecting cells with viral vectors has high efficiency than electro-transfection[11]. Since an applicable condition for electroporation on K562 cells is critical, we decided to transduce viral vectors containing dCas9 and BirA sequences into K562 cells. Compared with other viral vectors, lentiviral vector has more advantages on lentiviral transduction. Therefore, we constructed a lentiviral plasmid, pLenti-dCas9-BirA, that contains both dCas9 and BirA sequence, lentivirus backbone and puromycin antibiotic resistance gene and the total length of the plasmid is 14,000 base pair (Figure 1.2.5).

Before transduction, the lentivirus needed to be produced and purified. We use 293T human embryonic kidney (HEK) cells to package the lentivirus with pLenti-dCas9-BirA plasmid and other two helping plasmids. The lentivirus was packaged and harvested at three different concentrations: ⅓ x 10cm dish, 1 x 10cm dish, 3 x 10cm dishes. 0.25 million K562 cells were infected by these three different concentrations of lentivirus separately and followed by cell selection. We also transfected 293T cells using the packaged lentivirus as a positive control because they could be transducted with lentivirus more easily than K562 cells. However, a fatal disadvantage of viral transduction was ignored and result in the failure of the transfection. We did not aware this problem until we tested the transfection efficiency by western blot assay and the results was summarized in Table 1.2.2. Table 1.2.2 showed that in the control cell line 293T, dCas9 and BirA proteins were detected by FLAG and V5 antibody respectively. On the contrary, in K562 cells, dCas9 protein was not detected. Another information was given by this result was that no matter how we increase the concentration of the lentivirus, dCas9 was not expressed or expressed in low concentration by K562 so that we cannot detected it using the FLAG antibody.

In conclusion, the length of the viral plasmid is a potential condition of transduction. The longer the length of the plasmid, the more significant decrease in the transduction efficiency. 14,000 bp was an extreme large size for a viral plasmid, which made the transduction efficiency extremely low and caused our failure in this step.

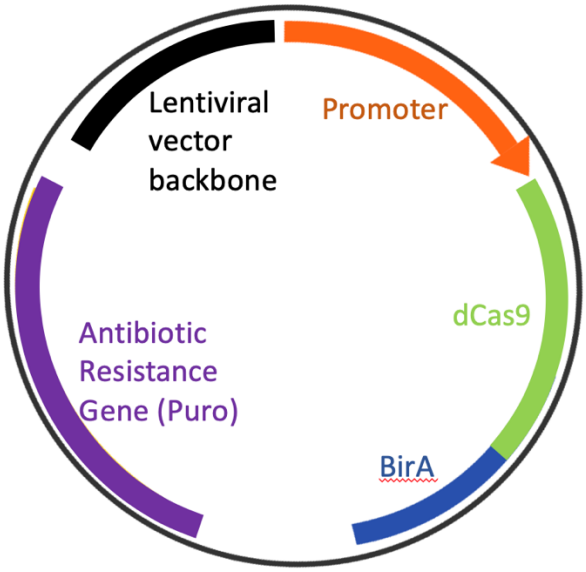

Figure 1.2.5 Main Functional Regions of pLenti-dCas9-BirA Plasmid (14kb)

Table 1.2.2 Summary of Lentivirus Transduction Results

Cell Type	Protein Detected	⅓ x 10cm Dish of Lentivirus	1 x 10cm Dish of Lentivirus	3 x 10cm Dish of Lentivirus
K562	**dCas9**	×	×	×
	BirA	✓	✓	✓
293xT	dCas9	✓	✓	✓
	BirA	✓	✓	✓

2.2.4 Puromycin Antibiotic Test

To select the cells that contains dCas9 protein, the concentration of puromycin antibiotic is tested. The cells are tested by live-cell dye, Cell-Titer Blue. Viable cells can convert Cell-Titer Blue dye into reddish color, in the other words, the redder the medium of the cell, the more live cells. According to this characteristic of this dye, we can visually identify the result (Figure 1.2.6). The figure showed that the K562 wild types with 2ug/mL puromycin were eliminated since there was no color change but the

transfected K562 cells are still alive because they possessed the resistance to puromycin antibiotics.

The concentration of G418 antibiotic used for K562 cells was 600ug/ul as described by Liu et al. 2017.

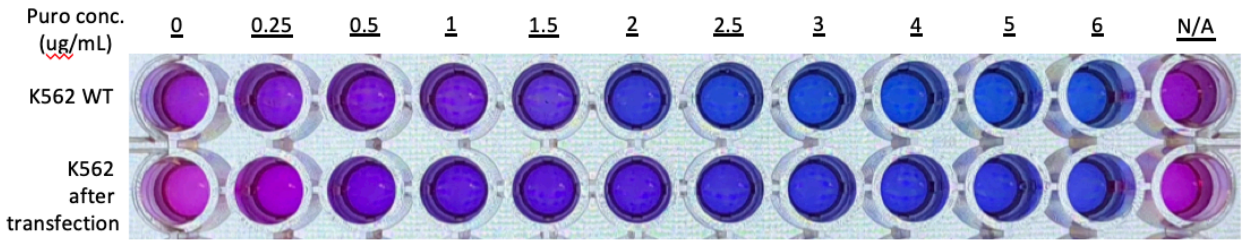

Figure 1.2.6 Picture of Color Changes of Cells in Different Puromycin Concentration After 4 h Cell-Titer Blue treatment.

2.2.5 Gene expression test

Western blot is an assay for detecting specific proteins that expressed by the cells. However, the specific antibody is required for recognizing the targeted protein, in our experiment, we used FLAG antibody for recognizing dCas9 protein. Since FLAG is a relatively low sensitive antibody comparing with V5 antibody which recognizes BirA protein, the dCas9 could not be detected by western blot assay was reasonable. The question that if the dCas9 sequence was incorporated into the genomic DNA could be identified by testing the gene expression of the transfected K562 cells. We quantified gene expression using real-time polymerase chain reaction (RT-PCR). There was no big difference of Ct value between electro-transfected K562 samples and wild type samples which means that the dCas9 sequence was not incorporated into the genomic DNA of electro-transfected K562 cells.

2.2.6 Combination of electroporation and lentiviral transduction

The failure of transfecting pEF1a-FB-dCas9 and pEF1a-BirA-V5 plasmids using electroporation and lentiviral transduction methods stopped us in CAPTURE system construction. From the conclusion we obtained in unsuccessful lentiviral transduction, we conjecture that if we decrease the size of pLenti-dCas9 plasmid, the lenti-transfection efficiency will significantly increase. In the process of construction pLenti-dCas9-BirA plasmid, we constructed a lentiviral plasmid only contains dCas9 sequence, which we named it pLenti-dCas9 (Figure 1.2.7). pLenti-dCas9 plasmid contained puromycin antibiotic resistance gene and the length was 12kb.

Then we packaged this shorter lentiviral plasmid into virus and infected the K562 cells which contained BirA G418 antibiotic resistance gene, which were established by electro-transduction (Figure 1.2.8).

After transfection, the cell lysis was tested by western blot. Figure 1.2.9 was the western blot result which showed that transfected K562 cells expressed dCas9 and BirA proteins while the K562 wild type did not. We also tested the gene expression of the transduced cells using RT-PCR. The Ct values of the transduced cells are smaller than the wild type which means that the dCas9 sequence was incorporated into the genomic DNA.

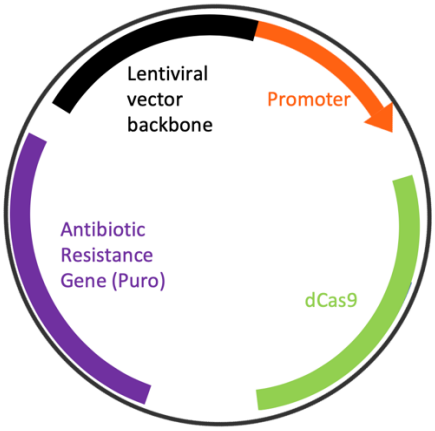

Figure 1.2.7 Main Functional Regions of pLenti-dCas9 Plasmid (12kb)

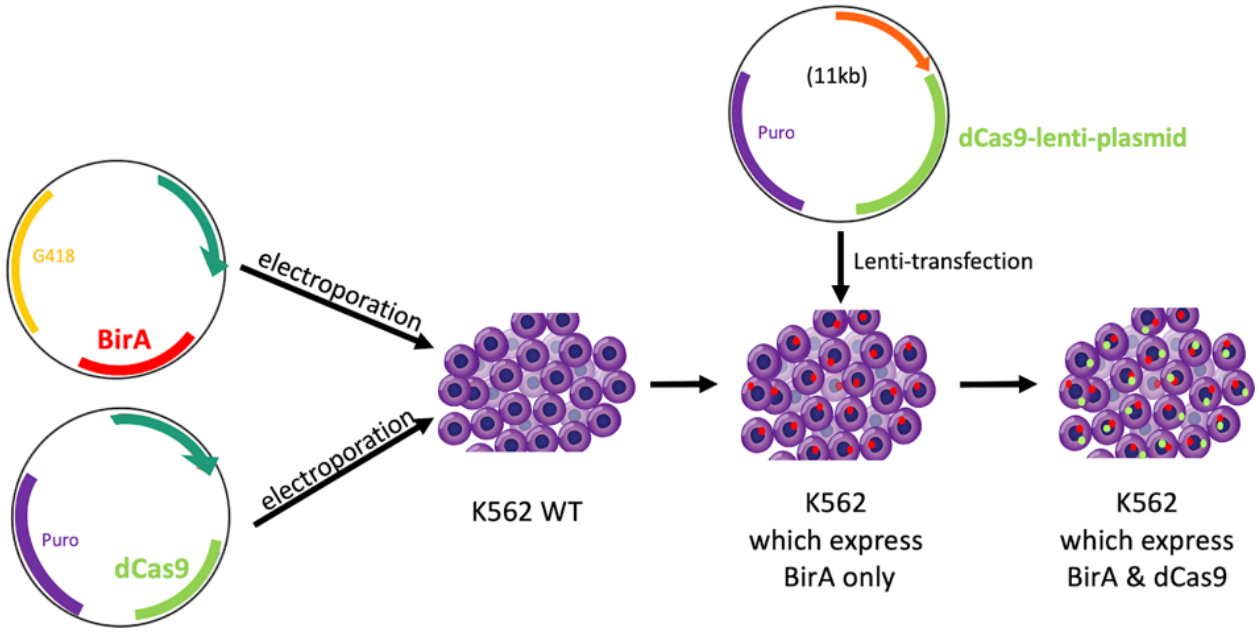

Figure 1.2.8 Schematic of the Establishment of K562 Cell Lines Expressing dCas9 and BirA Proteins.

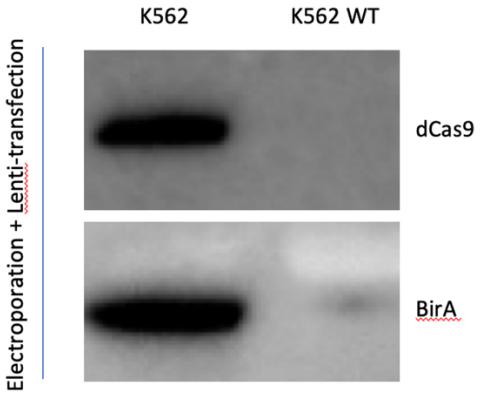

Figure 1.2.9 Western Blot Result of K562 Cells Expressing dCas9 and BirA Proteins.

2.2.7 gRNA Transfection

sgRNA is another core component in CAPTURE system. dCas9-biotin complex could not bind with the targeted DNA loci without the help of designed sgRNA. To test the feasibility of the system, we transfected pSLQ1651-sgTelomere plasmids contained telomere-targeting sgRNA into the K562 cells expressing dCas9 and BirA (Figure 1.2.10). The pSLQ1651-sgTelomere plasmids was transduced into K562 cells by lentivirus because we want long-term expression of gRNA.

After sgRNA transduction, K562 cells expressed green fluorescent protein (GFP) simultaneously since pSLQ1651-sgTelomere plasmid contained GFP sequence. Therefore, the cells that contains sgRNA coupled with GFP expression can be selected from the wild type by fluorescence-activated cell sorter (FACS) (Figure 1.2.11). To have the best quality cells for the following experiments, we selected the top 20% of GFP positive cells from whole population as described by Liu et al. 2017 (Figure 1.2.12).

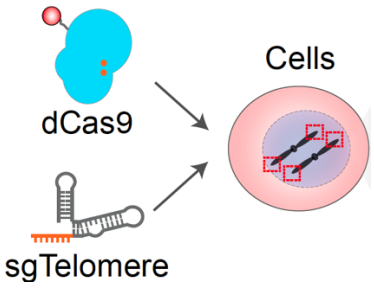

Figure 1.2.10 Schematic of dCas9-biotin Complex Capture of Human Telomeres (Adapted from "In Situ Capture of Chromatin Interactions by Biotinylated dCas9" by X. Liu, 2017, Cell, 170(5), p3. Copyright [2017] by Elsevier Inc.)

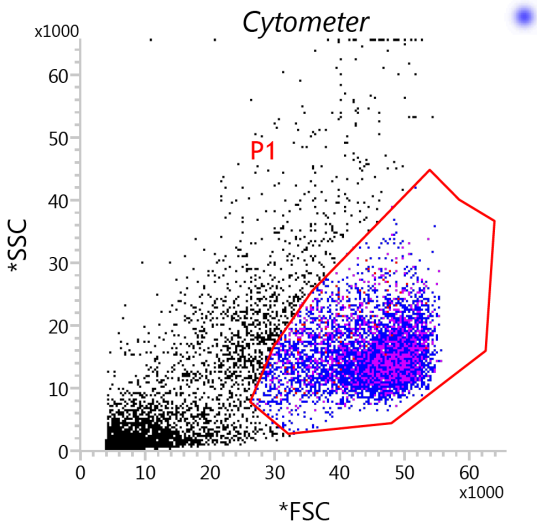

Figure 1.2.11 Gating and Sorting the K562 Cells Contained sgRNA from the Whole Population.

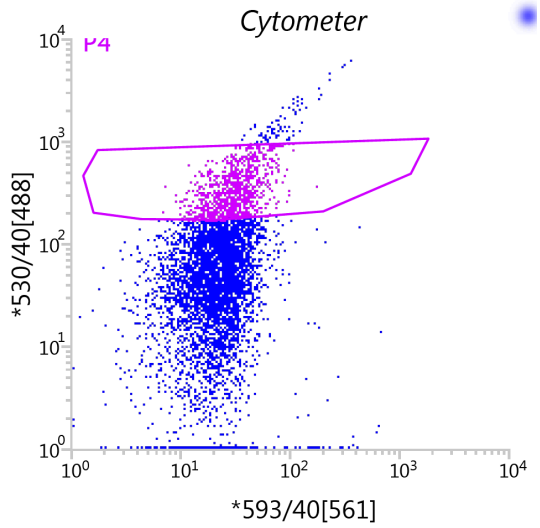

Figure 1.2.12 Gating and Sorting the Top 20% of GFP Positive K562 Cells

2.2.8 dCas9 Affinity Purification

After we generated stable K562 cells expressing dCas9 and BirA accompanied with telomere-targeted sgRNA, next step was to confirm that the dCas9-biotin complexes bind with DNA can be purified using streptavidin beads which had strong binding with biotin.

Firstly, we obtained the chromatin after the cross-linking and cell lysis as described by Liu et al. 2017. Then the chromatins were sheared into fragments (~500 bp) at moderate condition for following steps.

The difficulty in this step was to find a moderate condition to shear chromatin while maintaining the binding of dCas9 complexes for following streptavidin beads purification. We tried 12 conditions on the same samples (Table 1.2.3). Each sample contained 10 million cells and are prepared using same protocol. After chromatin shearing, DNAs are purified using Quick-DNA Kit (Zymo Research). The size of the DNA fragments was analyzed by Tapestation bioanalyzer. The tapestation result showed that most of the DNA are fragmented into 200 bp which are too short for streptavidin beads binding (Figure S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, Appendix).

Table 1.2.3 Summary of 12 Chromatin Shearing Conditions

Test	Peak Incident Power (PIP) (Watt)	Cycle/Burst	Duty Factor	Duration (min)	Volumn (uL)
2	75	200	20%	20	1000
3	75	200	20%	30	
4	75	500	26%	40	
5	75	300	20%	20	
6	75	400	20%	20	
7	75	500	20%	20	
8	75	200	15%	20	
9	75	300	15%	20	
10	75	400	15%	20	
11	75	300	10%	20	
12	75	400	10%	20	

2.3 Methods and Materials

2.3.1 Agarose Gel Electrophoresis

1% agarose gel was prepared by following steps: 2 g of agarose powder was weighed and mixed with 200mL 1 x TAE buffer in a 500mL glass flask. The mixture was microwaved for 2 minutes. The flask was taken out and swirled to mix and then microwaved for 30 seconds. The agarose solution was cooled down to about 70℃ before poured into gel tray with the well comb in place (avoid bubble). The gel was solidified at room temperature after 20-30 minutes.

The agarose gel was placed into the gel box and the gel box was filled with 1 x TAE buffer until the gel is covered. The 6 x DNA gel loading dye (Thermo Fisher) and 1Kb Plus DNA Ladder (Thermo Fisher) were prepared as described in the manufacturer's instructions protocol. 6 x DNA gel loading dye was added to the DNA sample (>1ug). 20ul visible samples and DNA ladder were loaded to the well side by side carefully. The gel was run at 120 V until the dye line was approximately 80% of the way down the gel depending on the gel length. After the electrodes were disconnected, the gel was removed from the gel box. The DNA fragments can be visualized under UV light. The size of the DNA can be measured by comparing with the DNA ladder as described in the manufacturer's instructions.

2.3.2 Bacterial Transformation

The GC10 competent cells were thawed on ice for 5 mins. The cells were mixed by tapping gently. 1 ul plasmid (~100ng DNA) was added into a new tube with 20 ul

competent cells. The tube was placed on ice for 30 minutes after tapping. The tube was placed into a 42℃ water bath for 45 seconds to heat shock the cells. The tube was put back to ice for 2 minutes. 500 ul LB was added to the tube (without ampicillin) and was shook in 37℃ incubator for 30 minutes. 100 ul of the mixture was added to a new agar plates (containing ampicillin) and incubated at 37℃ or 30℃ overnight.

2.3.3 Inoculating Liquid Bacterial Culture

Liquid LB was prepared and autoclaved. After LB was cooled down to room temperature, 100mg/mL ampicillin was added it. A single colony was selected using a sterile pipette tip from the LB agar plate. The tip was dropped into the liquid LB. The cap of the culture was loosely covered. The bacterial culture was shaken overnight at 37℃. After incubation, the culture is characterized by a cloudy haze. The bacterial culture could be stored at 4℃ for few days.

2.3.4 LB Agar Plate

7.4 g of pre-mixed LB-agar powder were measured and dissolved in 200mL sterile water in a 500mL flask. The flask was covered with aluminum foil and taped with autoclave tape. The gel mix was placed in the autoclave and the autoclave was set to 121℃, 20 psi, 30 minutes. The tape was black after autoclave and the gel mix was cooled down to 60℃ at room temperature. Then the ampicillin (100mg/mL) was added to the gel mix and well mixed by swirling the flask. With the flame on, 10mL gel mix was added to each sterilized culture plate. The plates were swirled to remove bubbles and solidified for

1 hour at room temperature. After cooled down and solidified, the plates were stored in a plastic bag at 4℃.

2.3.5 Electroporation

1 x 10^7 cells were counted and collected in a new tube and spun down at 450 x g for 5 minutes at room temperature. The cells were resuspended in DPBS (3x volume of the original growth medium) and spun down at 450 x g for 5 minutes at room temperature. The cells were resuspended in 400ul DPBS. pEF1a-FB-dCas9/pEF1a-BirA-V5 plasmids were added into the mixture and transferred to Gene Pulser®/MicroPulser™ Electroporation Cuvettes with 0.1/0.2/0.4 cm gap. The cuvette was put in the chamber and the voltage and capacitance of the Gene Pulser Xcel™ Electroporation System (Bio-Rad) was set (this step did not need to be processed in the cell culture hood). After the pulse sound of the machine was stopped, the cuvette was taken out and returned back to the hood. The transfected cells were taken out from the cuvette and put in 6 cm culture plate with 5 mL RPMI-1640 medium (with 10% FBS and 1% Pen/Strep) and put back to 37℃ incubator with 5% CO_2. After the cells were recovered for 1-2 days, the cells can be selected by antibiotics.

2.3.6 Lentiviral Plasmid Construction

The dCas9 sequence and BirA sequence was cut from pEF1a-FB-dCas9 and pEF1a-BirA-V5 plasmids respectively by restriction enzymes. The sequence was checked by size using 1% Agarose Gel Electrophoresis. The dCas9 sequence was inserted into pLenti-backbone (named pLenti-dCas9) and the sequence was checked by

sequencing (Eton Bioscience). pLenti-dCas9 plasmid was cut by restriction enzyme at the end of the dCas9 sequence. Then BirA sequence was inserted into the open region of pLenti-dCas9. The recombinant plasmid was transformed into GC10 competent cells (Sigma) and cultured on agar plates. The agar plates were incubated at 30℃ incubator for 20 hours. Single colonies were picked and inoculated individually in liquid bacterial culture and shake overnight in 37℃ incubator. The bacterial DNA of each single colony were extracted using ZymoPURE Plasmid Miniprep Kit (Zymo Research) and checked by Eton Bioscience sequencing. The plasmids with correct sequencing were combined together and transformed into GC10 competent cells and inoculated in large volume of liquid bacterial culture. The plasmid DNA was extracted by ZymoPURE Plasmid Maxiprep Kit (Zymo Research) and named pLenti-dCas9-BirA.

2.3.7 Lentivirus Packaging and Infection

5 x 10^6 293T cells were seeded in a 10cm plate for the confluency 50-70% the next day. The culture medium (DMEM with 10% FBS) of 293T cells was change into 10mL serum-free medium (opti-MEM, Gibco). In a 1.5 mL tube with 1.2mL opti-MEM, 6ug lentiviral plasmid was mixed with 4ug psPAX2 (Addgene #12260) plasmid and 2ug pMD2G (Addgene #12259) plasmid (3:2:1). 36ul X-tremeGENE^{TM} HP DNA transfection reagent (Roche) was added to the tube with opti-MEM and plasmid mixture then gently mixed by pipette. The mixture was incubated for 15-20 minutes at room temperature. The mixture was transferred into 293T plate and mixed by tilting the plate back and forth. The plate was returned to 37℃ incubator with 5% CO2. 6-8 hours post transfection, the opti-MEM was changed into DMEM with 10% FBS. 72 hours post transfection, the supernatant

that contains the virus was harvested. The virus was filtered by 0.45 uM syringe filter into a tube. The clarified supernatant was combined with ⅓ volume of Lenti-X Concentrator and mixed by gently inverting the tube. The mixture was incubated at 4℃ for 30 minutes to overnight. The virus pellet was visible after centrifuged at 1,500 x g for 45 minutes at 4℃. The supernatant was removed carefully, and the pellet was resuspended in 200ul RPMI-1640 medium.

24 hours before infection, 2.5 x 10^5 K562 cells are plated in each 12-well plate with 1mL RPMI-1640 medium with 10% FBS. 200ul virus were added to the medium with 8ug/mL of polybrene. 24 hours post infection, the medium containing polybrene and virus was changed back to RPMI-1640 medium with 10% FBS and then the cells are returned to the incubator.

2.3.8 Puromycin Concentration Test

Transfected K562 and K562 wild-type cells were counted using hemocytometer after added same volume of trypan blue dye. 60,000 cells were seeded with 100ul RPMI-1640 medium in a 96-well plate and puromycin antibiotic was added into each well with different concentration. The negative control (two types of the cells without puromycin) and the background control (medium only) were added to the plate as well. The plate was incubated at 37℃ with 5% CO_2. After two days, 20 ul warmed Cell-Titer Blue Reagent (Promega) was added to each well that contained the samples and returned back to the 37℃ incubator. After 4 hours, the color change of each well with different puromycin concentration was visible.

2.3.9 cDNA Reverse Transcription

Nuclease-free H_2O was added to 1ug RNA to make 10ul final volume. RT master mix was prepared and pipette mixed: 2ul 10x RT buffer (Thermo Fisher), 10x RT Random Primers, 0.8 ul 100mM dNTP Mix, 1 ul MultiScribe™ Reverse Transcriptase, 1 ul RNase Inhibitor, and 3.2 ul nuclease-free H_2O (as described in the manufacturer's instructions of High-Capacity cDNA Reverse Transcription Kit). 10 ul RT master mix was added to the 10 ul RNA sample. The mixture was pipette mixed and performed in PCR machine with following program: 25℃ for 10 minutes, 37℃ for 120 minutes, 85℃ for 5 minutes. 200 ul nuclease-free H_2O was added to the cDNA and stored at -20℃ for the future gene expression test.

2.3.10 Gene Expression Test Using RT-PCR

9ul diluted cDNA obtained from reverse transcription method was mixed with 10 ul 2x PowerUp™ SYBR™ Green Master Mix (Thermo Fisher) and 1 ul 5 uM primer mix (targeting dCas9, BirA or GAPDH) and amplified in qPCR machine with following program: 1 cycle of 94℃ for 2 minutes, 20 cycles of 94℃ for 15 seconds, 50℃ for 1 minutes. We used GAPDH as an internal control because is often stably and constitutively expressed at high levels in K562 cells.

2.3.11 Chromatin Shearing

0.05 x 10^9 K562 cells expressing dCas9 and BirA transfected with sequencing-specific sgRNA were harvested and cross-linked with 2% formaldehyde for 10 min rotation and quenched with 0.25M glycine for 5 min rotation. Cells were washed twice

with PBS and then lysed with 2mL of cell lysis buffer ((25mMTris-HCl, 85mMKCl, 0.1% Triton X-100, pH 7.4, freshly added 1mMDTT and 1:200 proteinase inhibitor cocktail (Roche)) and rotated for 15min at 4℃. Cell lysate were centrifuged at 2,300 g Cell lysates were centrifuged at 2,300 x g for 5 min at 4℃ to isolate the nuclei. The nuclei were resuspended in 1 mL nuclear lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 4% SDS, pH 7.4, freshly added 1mMDTT and 1:200 proteinase inhibitor cocktail) and incubated for 10 min at room temperature. Nuclei suspension was then mixed with 3 mL of 8 M urea buffer and centrifuged at 16,100 x g for 25 min at room temperature. Nuclei pellets were then resuspended in 1 mL nuclear lysis buffer and mixed with 15 mL of 8 M urea buffer, and centrifuged at 16,100 x g for 25 min at room temperature. The samples were washed twice more in 1 mL nuclear lysis buffer and mixed with 3 mL of 8 M urea buffer, followed by centrifugation at 16,100 x g for 25 min at room temperature. Pelleted chromatin was then washed twice with 1 mL cell lysis buffer. Chromatin pellet was resuspended in 1 mL of IP binding buffer without NaCl (20 mM Tris-HCl, 1 mM EDTA, 0.1% NP-40, 10% glycerol, pH 7.5, freshly added proteinase inhibitor) and aliquoted into Eppendorf tubes. Chromatin suspension was then sonicated into an average size 500bp on the Covaris M200 Focused-ultrasonicator at 4℃. Fragmented chromatin was centrifuged at 16,100 x g for 25 min at 4℃. Supernatant was combined and final concentration 150mMNaCl was added to the sheared chromatin. The DNA in 200 ul of sheared chromatin was extracted using Quick-DNA Plus™ Kits (Zymo Research) and analyzed by Agilent 4200 Tapestation bioanalyzer (UCSD IGM). To prepare the streptavidin beads for affinity purification, 10ul of streptavidin agarose slurry (Life Technologies) was washed 3 times in 0.5mL of IP binding buffer and added to the rest soluble chromatin. After overnight incubation at 4℃,

streptavidin beads were collected by centrifugation at 800 x g for 3 min at 4℃. The beads were then washed 5 times with 0.5 mL of IP binding buffer with NaCl (20 mM Tris-HCl, 1 mM EDTA, 0.1% NP-40, 10% glycerol, 150 mM NaCl, pH 7.5, freshly added proteinase inhibitor) and resuspended in 20 ul of 2 x XT sample loading buffer (Bio-Rad) containing 5% 2-mercaptoethanol followed by incubation at 98℃ for 10 min. The proteins were separated by SDS-PAGE and analyzed by western blot.

2.4 Discussion

In summary, a K562 cell line that stably expressed dCas9 and BirA proteins was established and confirmed by western blot. However, the Biotin-dCas9 complex system was not confirmed by streptavidin beads affinity purification, so if the Biotin-dCas9 would bind to the specific locus with the help of gRNA was not verified. The following steps for the confirmation of the CAPTURE system was necessary. In the future, first thing we need to do is to test the optimal sonication condition for chromatin shearing. From the current results we obtained, every condition was exceeded. To achieve the 500 base pairs fragment size, the duty factor and the cycle per burst should be highly decreased. The duration time of sonication should also be shortened.

Chapter 3

3.1 Introduction

Gout is a common disease that more than 8 million people or 4% of American adults are suffering from it[12]. Gout is a form of inflammatory arthritis caused by high levels of uric acid in the blood[12]. Acute gout attacks are characterized by rapidly intense pain in the affected joint followed by swelling, redness and inflammation. Recent studies showed that human genomes can offer clues to major challenge in understanding human disease. Human genome has complex architecture. Chromosome consist of packed nucleosome, which is naked DNA wrapped around histones. The highly ordered architecture contains various functional elements, including most well-studied promoter and gene body regions, and long-range regulatory elements. These components function in gene expression and epigenetic regulation. Researchers reported abnormal gene expression on a couple gouty cases. For example, the gene expression of IL-1β causes cartilage destruction[13]. However, the epigenetic regulation is unknown.

Gout is an inflammatory disease which are caused by a disorder of the immune system. The cells of the immune systems can be categorized as lymphocytes (T-cells, B-cells etc.) and monocytes, neutrophils. These cells are released into the blood and transferred into inflamed locations[14]. In humans, lymphocytes and monocytes make up the majority of peripheral blood mononuclear cell (PBMC) population. In other words, blood contains the cells that are related to gout disease. Moreover, blood is easily accessible. Therefore, blood of gout patients is a great bio-specimen for us to better understand the development and pathogenesis of gout disease.

Previous research in our lab (done by Dr. Ying Zhao and Dr. Miao Zeng) identified the difference of DNA methylation and gene pathway between gout patients and healthy individuals by analyzing cryopreserved PBMC from blood. Patients and healthy individuals are clearly divided into two groups by analyzing the results of 5000 gene locus (Figure 2.1.1).

The dynamic landscape of open chromatin is a part of epigenetic regulation, because this region is accessible to various regulator components (Figure 2.1.2). Assay for Transposase Accessible Chromatin with high throughput sequencing (Atac-seq) is a powerful technique for studying epigenetic regulation in the context of disease development and cell differentiation. For example, the specific chromatin accessibility signature of naïve B cells from Systemic Lupus Erythematosus (SLE) patients was identified by using Atac-seq [15]. This assay uses hyperactive Tn5 transposase to simultaneously cut and ligate adaptors for sequencing at the open chromatin region, which is accessible for Tn5. With next generation sequencing, we can profile the open chromatin pattern. Although this technique is fairly mature, optimal results can only be achieved with assays tailored toward our biospecimen. In this project we optimized several steps for the cryopreserved PBMC samples from 13 gout patients and 16 healthy individuals.

A large portion of transcripts relative to gout disease remained unknown. RNA-seq is a powerful technique for profiling transcriptomes of gout[16]. RNA-seq is focusing on the RNA extracted from large cell population or tissue. Previous research in our lab (done by Dr. Ying Zhao and Dr. Miao Zeng) showed that there is no big difference between patients and healthy individuals according to the bulk RNA-seq data. However, this result cannot

represent that there is no variance in RNA expression between the patients and healthy individuals because the result of the bulk RNA-seq is an average of thousands to millions of individual cells transcriptomes presents in each sample, the potentially significant difference between cells may be blurred[17]. To discover the mechanisms of RNA of cells individually, we decided to apply single cell RNA-seq on the biospecimen of patients and healthy individuals.

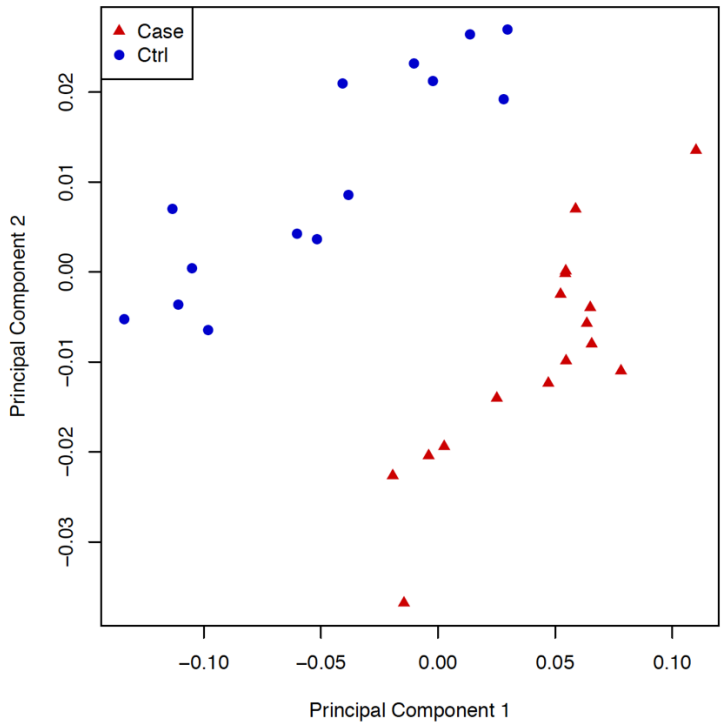

Figure 2.1.1 Significant Difference in DNA Methylation of Patients (Case) and Healthy Individuals (Ctrl)

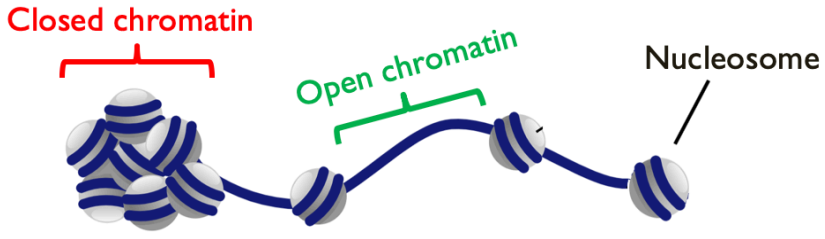

Figure 2.1.2 Schematic of Chromatin

3.2 Results

3.2.1 Cell Sorting

The biospecimen we analyzed were cryopreserved PBMCs from bio-bank. The cells would die during the processing and the viability of cells in each sample varies so we need to eliminate the dead cells which would affect our results. We used fluorescent dye Calcein AM to label the live PBMC after thawed them at 37℃ from liquid nitrogen. After the cell was labeled with Calcein AM, we sorted the live cells from the whole population using fluorescence-activated cell sorter (FACSDiva Version 8.0) (Figure 2.2.1). Figure 2.2.1 is the result of one of the samples we sorted. It took us 2 hours for preparing and sorting 2 samples.

The recovery rate varied from 60% to 70% and RNA integrity number (RIN) was 8.8 (full mark is 10), which means that the RNA quality of the cells was good, so we can process to next steps with sorted cells.

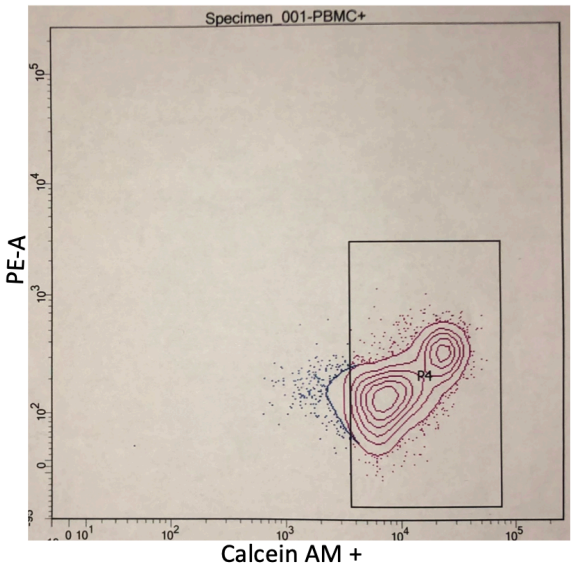

Figure 2.2.1 Calcein AM Positive Gated Cells from Whole Population

3.2.2 Atac-seq

For this assay, the first step was to lyse the sorted cells to get the nuclei. In the nuclei, there were closed and open chromatin region. Transposase Tn5 accessed and cut the open region into fragments. After amplified and sequenced the DNA fragments, we profiled the open chromatin pattern associated with gout disease (Figure 2.2.2). Therefore, the most crucial step in Atac-seq was the nuclei preparation. We set up three major quality controls steps to achieve good nuclei preparation: Tn5 accessibility test, Tn5 quantity, and mitochondrial DNA (mtDNA) contamination.

According to the analyzing the library sequencing result (Figure 2.2.3), the gout patients and healthy individuals PBMC samples cannot be separated. We are proposing applying single cell Atac-seq on these samples because the bulk Atac-seq focused on a large population of cells instead of individual cells. The individual cells may give more information than bulk.

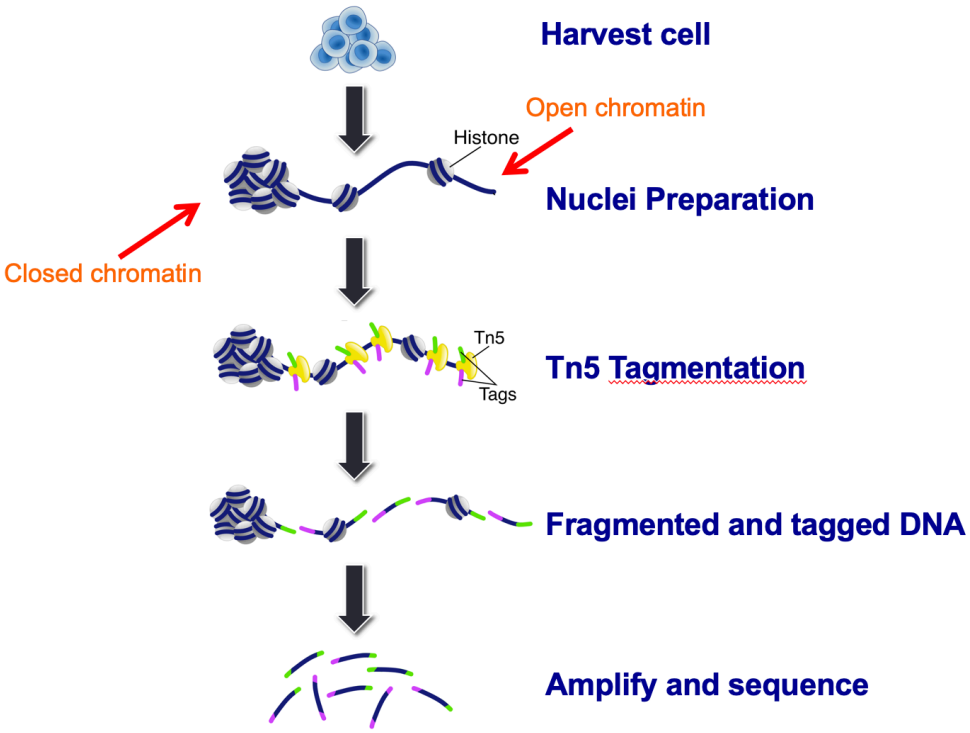

Figure 2.2.2 Schematic of Atac-seq

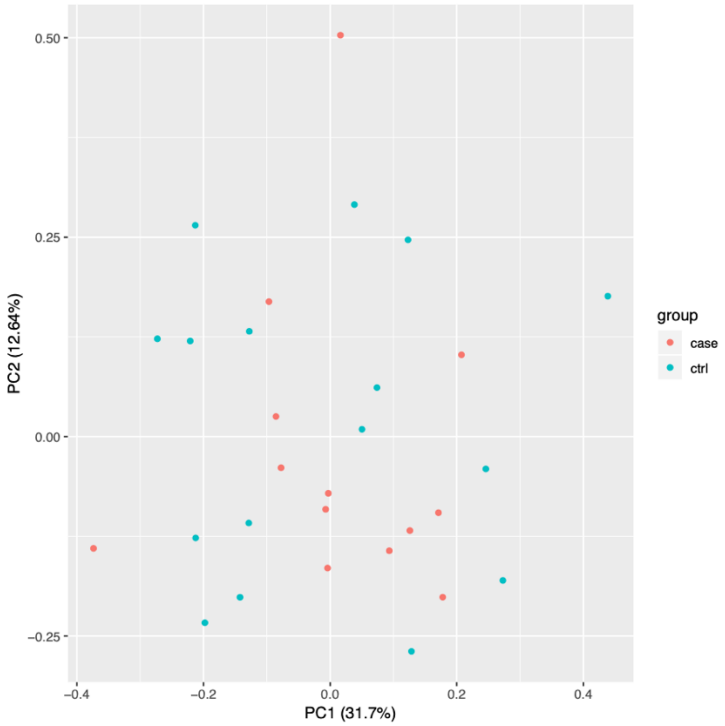

Figure 2.2.3 Analysis Result of cDNA Sequencing Library

3.2.3 Tn5 Accessibility Test

Tn5 had to get into the nuclei to do the tagmentation on nuclear DNA (Figure 2.2.3). Intact cell membrane excluded Tn5 from access to nuclei, so the cell membrane had to be lysed. Nuclei morphology also affect the Tn5 accessibility. High quality isolated nuclei would be oval shape with smooth membrane, which means that the nuclear DNAs were not damaged. To check the cell membrane intact or not, we stained the lysed cells using trypan blue. The cells with the lysed membrane were stained blue and visually checked under a microscope. We expected the majority of the cells (>90%) were blue. DAPI staining solution was a fluorescent stain and bound with DNA. After adding the DAPI, we checked the nuclei morphology under fluorescence microscope. The lysed cells would release strong fluorescent given by undamaged nuclear DNA. Another requirement was that the nuclei could not cluster together. Clustered nuclei also made Tn5 hard to enter which can be directly checked under a microscope after cell lysis (Figure 2.2.4 as an example). We tested 16 lysis buffer recipe, 2 lysis time, and 2 wash buffer recipes (Table 2.2.1). By screening the nuclei condition and dead cell percentage, we narrowed down to four different lysis conditions and proceed further for the next step.

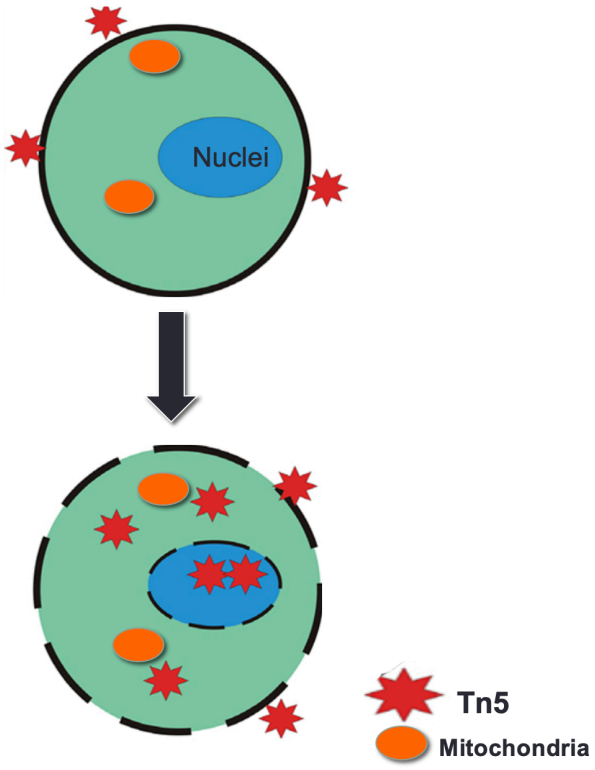

Figure 2.2.4 Schematic of Tn5 Accessibility

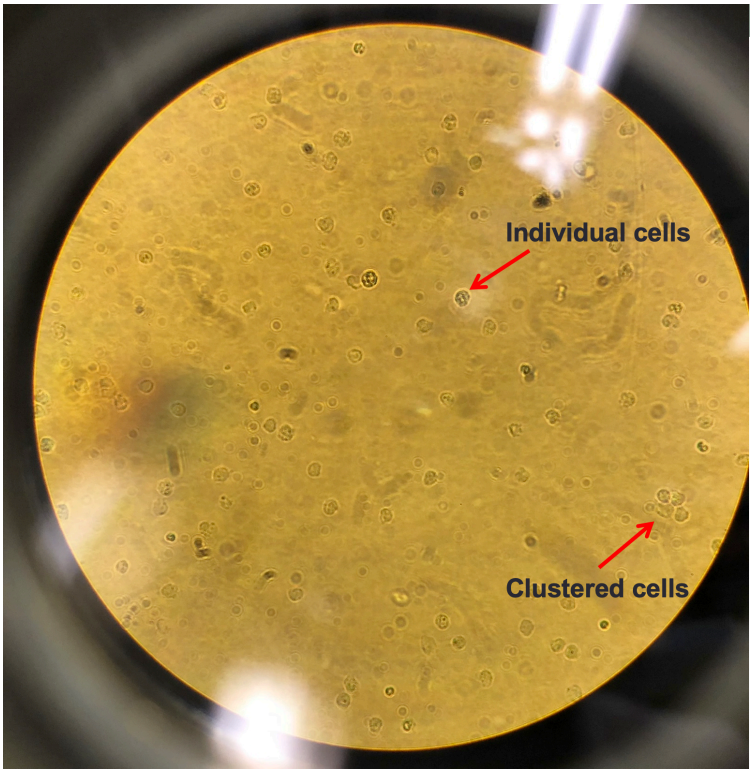

Figure 2.2.5 Example of Individual Cells vs. Clustered Cells Under a Microscope

Table 2.2.1 Results of Nuclei Conditions Treated by 16 Cell Lysis Conditions

Condition	Lysis Buffer	Lysis Time	Wash Buffer	Nuclei condition
1	0.1% NP-40	5 sec	+ 0.1% Tween 20	✗
2		3 min		✗
3	0.1% IGEPAL			✗
4	0.05% IGEPAL			✓
5	0.15% IGEPAL			✗
6	0.05% NP-40	5 sec		✗
7	0.075% NP-40			✗
8	0.15% NP-40			✗
9	0.1% NP-40 + 0.1% Tween 20			✗
10	0.1% NP-40 + 0.1%Tween 20 + 0.01% Digitonin	3 min		✓
11		5 sec		✗
12	0.1% Tween 20	3 min		✗
13		5 sec		✗
14	0.05% Tween 20	3 min	No detergent added	✓
15	0.075% Tween 20			✓
16	0.15% Tween 20			✗

3.2.4 Tn5 Quantity Test

The second QC was Tn5 quantity. Too many or insufficient Tn5 would cause overcutting or undercutting of nuclear DNA. During wash and centrifuge steps, there might be cell loss, so counting the cells after nuclei preparation was crucial for calculating the amount of Tn5 for the right cell number. The cutting efficiency was measured by tapestation bioanalyzer (Agilent) result. Figure 2.2.5 was a previous example of overcut DNA when added excess Tn5. This tapestation result showed that there is no nucleosome ladder pattern and most of the fragments are very small. The tapestation result of an appropriate tagmentation showed obvious nucleosome ladder pattern (Figure 2.2.6). The first and the most abundant peak were mononucleosome, followed by di-nucleosome and so on. In the nucleosome-free open chromatin regions, many molecules of Tn5 kicked in

and chopped the DNA into small pieces; around nucleosome-occupied regions, in other words, close chromatin regions, Tn5 can only access the linker regions. Therefore, in a normal Atac-seq library, we expected to see a sharp peak at the <100 base pair (bp) region, which was the open chromatin region, and a peak at ~200 bp region, which were mono-nucleosomes, and other larger peaks, which were multi-nucleosomes. We tried four lysis conditions that passed the previous QC and 50k and 100k cell numbers for each condition and the samples were analyzed by tapestation bioanalyzer. By looking at the tapestation results, all of the conditions were good for the following sequencing step (Table 2.2.2).

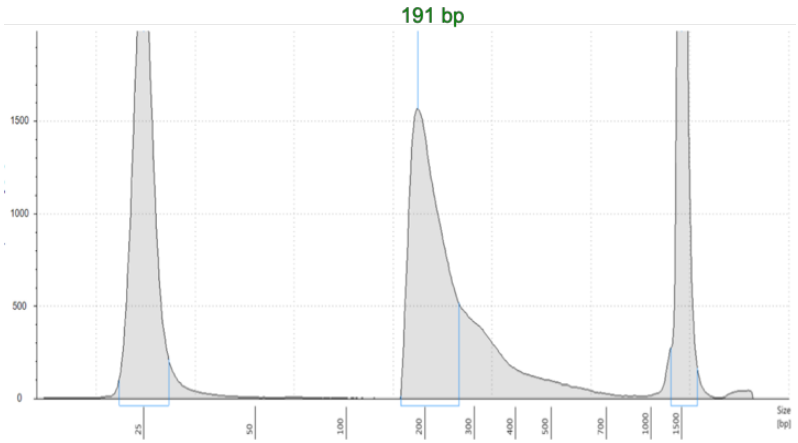

Figure 2.2.6 Tapestation Result of Overcut DNA (Example)

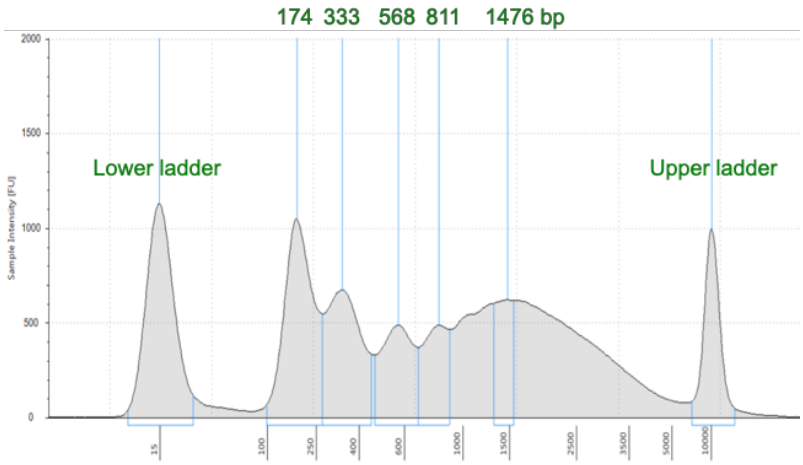

Figure 2.2.7 Tapestation Result of a Good DNA Tagmentation (Example)

Table 2.2.2 Summary of Tapestation Result of The Tested Lysis Conditions

Samples	Lysis Buffer	Wash Buffer	Cell Number	Nucleosome Ladder Pattern
1	0.1% NP-40 + 0.1%Tween 20 + 0.01% Digitonin	+ 0.1% Tween 20	50k	✓
2			100k	✓
3	0.05% IGPEL		50k	✓
4			100k	✓
5	0.05% Tween 20	No detergent added	50k	✓
6			100k	✓
7	0.075% Tween 20		50k	✓
8			100k	✓

3.2.5 MtDNA Contamination

The last QC for nuclei preparation is to determine how severe the mtDNA contamination was. MtDNA was nucleosome free similar to the open chromatin region of nuclear DNA. To avoid the mtDNA contamination, the mitochondrial membrane cannot be destroyed, so the Tn5 cannot access the mtDNA. The mtDNA percentage was analyzed using the data of sequencing result. We aligned the sequencing data with the human reference genome and obtained the percentage of the mitochondrial DNA that was in our sequencing results.

We repeated twice with the rest of the lysis conditions and the results of them are listed in Table 2.2.3. The duplicate meant how many copies of the same fragments were sequences, the lower the duplicate rate the better. The lysis condition with 0.05% IGPEL lysis buffer had the most mitochondria DNA reads, so it was eliminated. The table also showed that the more input cells, the less duplicate rates which means the higher the coverage of the reads. The first lysis condition (0.1% NP-40 + 0.1% Tween 20 + 0.01% Digitonin, 3min lysis time, wash buffer with 0.1% Tween 20) was the best lysis condition

which it had the least mtDNA reads and duplicate rate, so we applied this condition on all the clinical samples we collected.

Table 2.2.3 Results of Lysis Conditions

Lysis Condition	Lysis Buffer	Lysis Time	Wash Buffer	Cell Number	Total Reads	Mapped Reads	MtDNA Reads	Duplicate Rate
1	0.1% NP-40 + 0.1% Tween 20 + 0.01% Digitonin	3 min	+ 0.1% Tween 20	50k	17M	94%	**1%**	**4%**
2				100k	16M	94%	**1%**	**3%**
3	0.05% IGPEL			50k	14M	95%	59%	49%
4				100k	18M	95%	56%	49%
5	0.005% Tween 20		No detergent added	50k	24M	94%	1%	11%
6				100k	32M	95%	1%	8%
7	0.075% Tween 20			50k	15M	92%	22%	25%
8				100k	17M	92%	4%	11%

3.2.6 ATAC-seq Results

After aligned the sequencing library of gout patients and healthy individuals on human reference genome, the distributions of the DNA fragment size were analyzed and shown in the Figure S12 and S13. All the distributions showed that the fragments from open chromatin region had the highest peak followed by mono nucleosome, di-nucleosome…and so on, which were consistent with what we expected. Then PCA analysis was used to check if the gout patients and the control healthy samples could be separated based on the open chromatin region identified from the sequencing data (Figure 2.2.8). From Figure 2.2.8, there was no clear separation between two groups. Bulk ATAC-seq could not distinguish the heterogeneity within the PBMC population. From the result of bulk RNA-seq, two groups could not be separated based on the gene expression of the same PBMC population data. Therefore, we used the single cell RNA-seq to study the difference gene expression between two groups on the individual cell level.

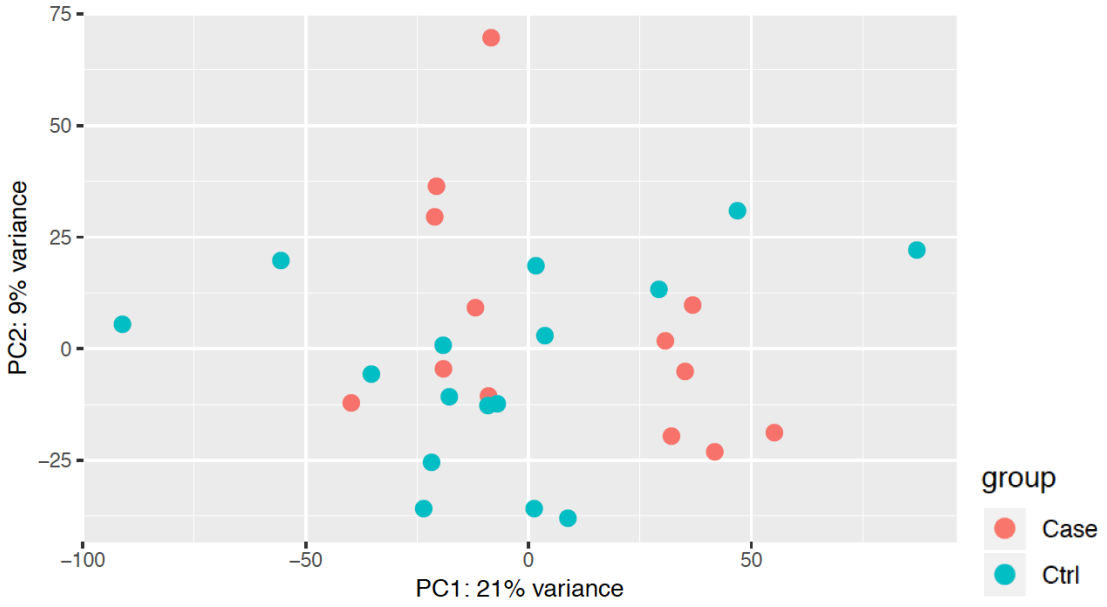

Figure 2.2.8 PCA Analysis Result Based on the Open Chromatin Region of Gout Patients and Healthy Individuals

3.2.7 Single Cell RNA-seq

For this assay, the first step is to isolate the single cells from the sample, then extract the RNA from the single cells and reverse transcribed into cDNA. The cDNAs were amplified by polymerase chain reaction (PCR) before sequencing (Figure 2.2.9). The most important step in the technique is the single cell isolation. This step required high-quality cells because the dead cells would result in increasing the amount of ambient RNA and low- quality cells might contain degraded RNA, so that the good and valid transcripts profiling can be generated from each single cell. Drop-seq is a type of single cell RNA-seq based on droplet microfluidics with easy equipment installation and we adopt drop-seq method[18] on our samples using homemade chip.

The quality of amplified cDNA which is cDNA library should be checked before sequencing since the library sequencing is money-consuming. A good library tapestation result only has one broad peak at around 500 bp, which means that the tagmentation on cDNA was sufficient, so the sample is ready for library sequencing.

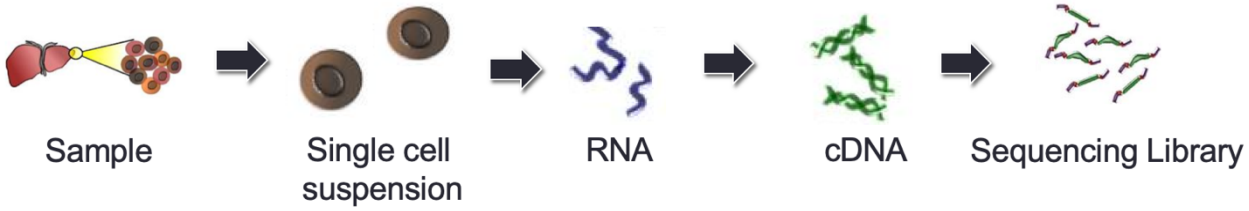

Figure 2.2.9 Schematic of Single Cell RNA-seq

3.2.8 Drop-seq Results

A good tapestation result of drop-seq sequencing libraries should have a broad peak at around 450 base pairs. We analyzed a pair of samples which case 1 was a gout patient PBMC sample and control 5 was a matched healthy individual sample. The tapestation results of this pair were looks good and can be further analyzed (Figure 2.2.12 and Figure 2.2.13). We then use knee plots to select the qualified cells (Figure 2.2.14). The dash line in the knee plots indicates the knee threshold. The cells on the left side of the dash line were the qualified cells. 471 cells in case 1 sample and 964 cells in control 5 sample were qualified for the following analysis. The qualified cells were clustered and annotated with known marker from each sub population in PBMC. Four clusters and eight clusters were found in case 1 and control 5 respectively (Figure 2.2.15). Case 1 was missing four clusters: FCGR3A+ monocyte cells, dendritic cells, memory CD4 T cells, and

B cells. This result was preliminary because the selected population of the patient case was much smaller than the healthy individual. The rest of the paired samples would be analyzed for a solid conclusion.

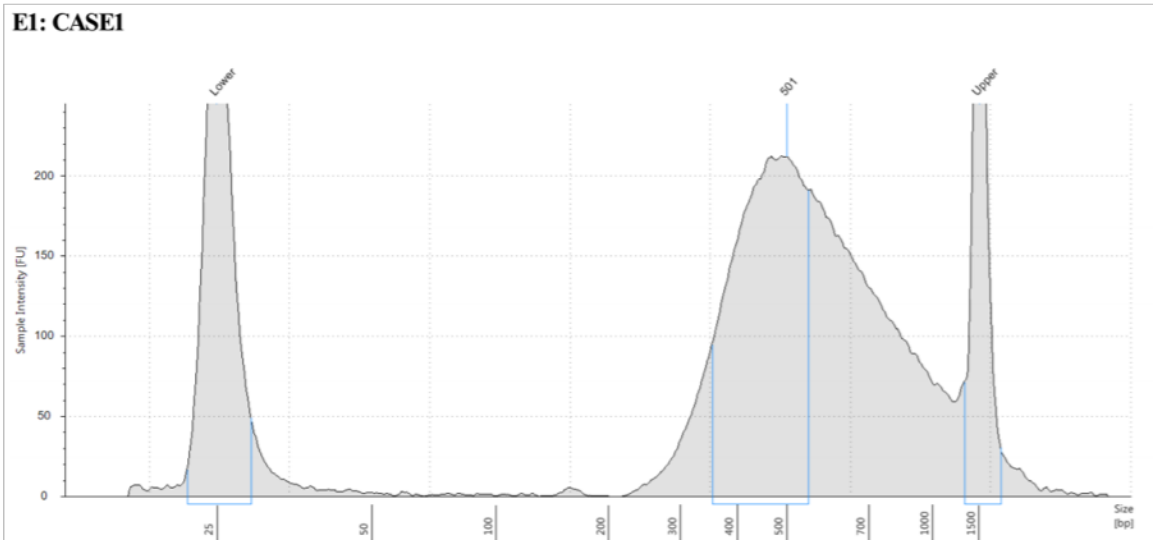

Figure 2.2.10 Tapestation Result of Sequencing Library of Gout Patient “Case 1”

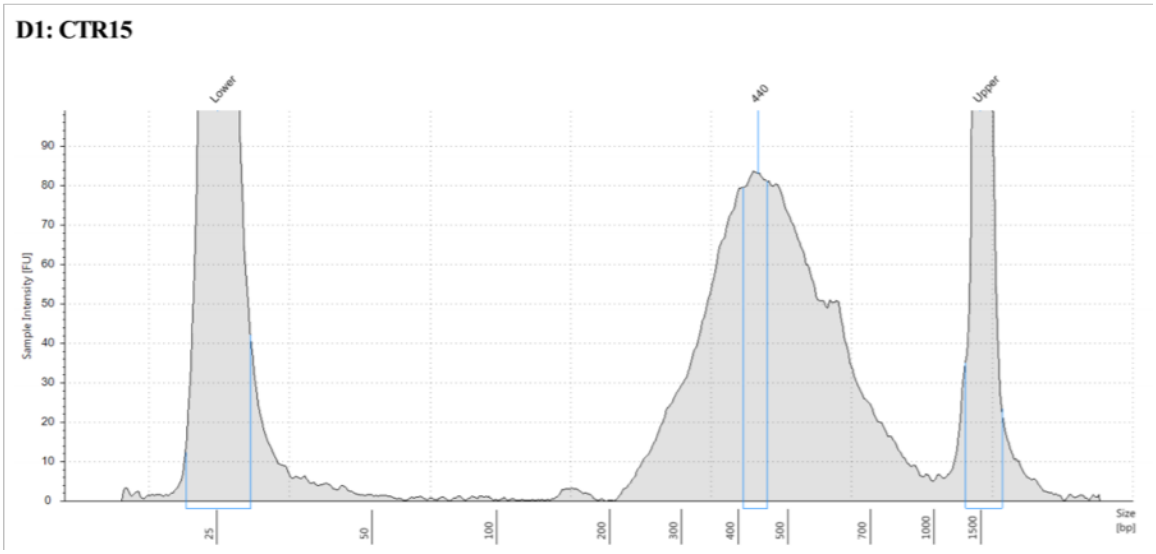

Figure 2.2.11 Tapestation Result of Sequencing Library of Healthy Individual “Control 5”

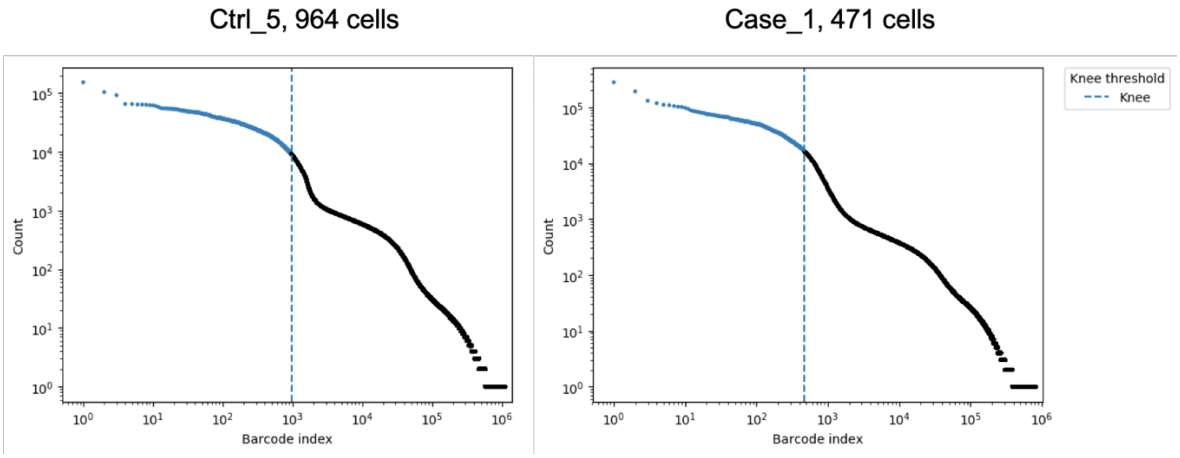

Figure 2.2.12 Qualified Cells

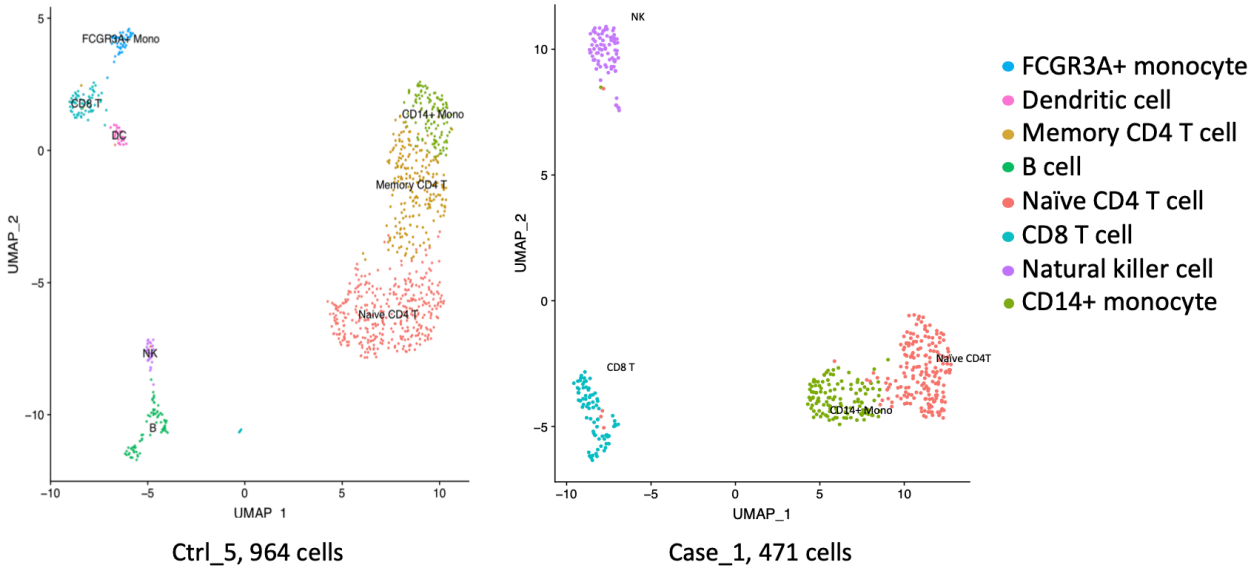

Figure 2.2.13 Clustering Results of Control 5 and Case 1 samples

3.3 Methods and Materials

3.3.1 Cell Sorting with Calcein AM Dye

10 mL warmed RPMI-1640 medium was added to a new 15 mL conical tube. PBMC was thawed in 37℃ water bath until a small piece of ice remained in the cryogenic storage vial. Thawed PBMC was added into 10 mL warmed medium and centrifuged at 450 x g for 5 minutes at room temperature. The supernatant was decanted without disturbing the cell pellet. The cell pellet was resuspended in 1 mL FACS buffer (2% FBS, 5mM EDTA, and PBS) and centrifuged at 450 x g for 5 minutes at room temperature. The cell pellet was resuspended in 600 ul FACS buffer. 100 ul cell mixture was taken out to another vial with 500ul FACS buffer. 10 ul 1 uM Calcein AM working solution (0.5 ul 1mM Calcein AM stock solution was added to 500 ul DPBS) was added to the rest of 500 ul cells in FACS buffer. The mixture was incubated in the dark at room temperature for 20 minutes. The mixture was centrifuged at 450 x g for 5 minutes at room temperature. The cell pellet was resuspended in 1 mL FACS buffer and proceeded on a cell sorter. The sorter was gated to Calcein AM positive cells with 85-micron nozzle and 4-way purity.

3.3.2 Atac-seq

Resuspension buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl2, and nuclease-free H2O) was prepared and stored long-term at room temperature. 100,000 cells were collected in each epi tube by centrifuged at 450 x g for 5 minutes at room temperature. The cell pellet was washed with 100 ul cold FACS buffer (2% FBS, 5mM EDTA, and PBS) and centrifuged at 450 x g for 5 minutes at 4℃. The supernatant was

discarded. The cell pellet was resuspended by 50 ul cold lysis buffer (final 0.1% NP-40, final 0.1% Tween 20 and final 1% Digitonin in resuspension buffer) by gently pipetted up and down three times. The mixture was incubated on ice for three minutes. Then 1 mL cold wash buffer (final 0.1% Tween 20 in resuspension buffer) was added to the tube. The wash buffer was mixed by gently inverting tube 3 times. The mixture was centrifuged at 500 x g for 10 minutes at 4℃. The supernatant was discarded, and the nuclei pellet was resuspended, and pipet mixed in 50 ul transposition reaction mixture (as described in the original atac-seq method). The mixture was incubated at 37℃ for 30 minutes on thermomixer at 1,000 rpm. DNA fragments were purified using DNA Clean & Concentrator-5 kit (Zymo Research) and eluted in 10ul elution buffer.

10 ul purified transposed DNA was combined with 10ul nuclease-free H2O, 2.5 ul custom Nextera PCR primer 1, 2.5 ul custom Nextera PCR primer 2 and 25 ul NEBNext High-Fidelity 2x PCR master mix. The mixture was amplified in PCR machine with following program: 1 cycle of 72℃ for 5 min and 98℃ for 30 sec, 5 cycles of 98℃ for 10 sec, 63℃ for 30 sec, 72℃ for 1 min (as described in the original atac-seq method). Then part of the sample was proceed using qPCR to determine the cycle number for the following PCR. 5 ul of partially-amplified DNA was combined with 3.85 ul nuclease-free H2O, 0.5 ul custom Nextera PCR primer 1, 0.5 ul custom Nextera PCR primer 2, 0.15 ul 100 x SYBR Green and 5 ul NEBNext High-Fidelity 2x PCR master mix. The mixture was amplified in PCR machine with following program: 1 cycle of 98℃ for 30 sec, 20 cycles of 98℃ for 10 sec, 63℃ for 30 sec, 72℃ for 1 min. The number of additional PCR cycles needed for each sample are determined by the number of cycles needed to reach ⅓ of the maximum R on the plot of R vs Cycle Number. The rest 45 ul sample was amplified

by PCR with the following program: 1 cycle of 98℃ for 30 sec, N cycles (determined by qPCR) of 98℃ for 10 sec, 63℃ for 30 sec, 72℃ for 1 min.

The amplified DNA fragments were purified using NA Clean & Concentrator-5 kit and eluted in 20 ul nuclease-free H2O. The sample was then ready for library sequencing if the tapestation result was good.

3.4 Discussion

In summary, we have finished patient recruitment. There are 16 gout patients and 13 matched healthy individuals. We optimized Atac-seq assay and generated ATAC-seq library for sequencing. The data showed that there is no difference between two groups based on the open chromatin profiling which may be caused by the heterogeneity of the mixed population within peripheral blood mononuclear cells (PBMC). Previously, we also did bulk RNA-seq on the same samples and we obtained the same result: no difference of gene expression between two groups, which may also be caused by the heterogeneity of PBMC. Then we decided to use single cell RNA-seq to study the gene expression on the individual cell level. In-house drop-seq platform was used for single cell RNA-seq. The preliminary result from drop-seq showed that there is fewer cell type in one gout patient samples, because the selected population of the gout patient sample was much smaller than the healthy individual sample. Therefore, the rest of the samples of the cohort to get a solid conclusion.

Chapter 4

Overall Discussion and Future Plans

In our view, CAPTURE technique will be crucial to develop application on different cell lines to study the cis regulatory elements. The high locus specificity of this technique provides us an easier way to study multiple motifs simultaneously by only designing sequencing-specific sgRNA. However, the limitation of this technique is the establishment of stable cell lines. K562 cells have high capacity in accepting foreign particles, which makes it easy to be modified. Despite that, it is not easy to construct K562 cell that stably expressing dCas9 and BirA, thus, establishment of stable cell lines in other fragile cell lines is a big challenge. The major part of my work is to establish the CAPTURE system in our lab. I have established stable dCas9 and BirA expressions in K562 cell line which is a fundamental step in this project. The future plan is to firstly test the sonication conditions for chromatin shearing for streptavidin beads affinity purification. If the enrichment of dCas9 and telomere-binding protein was verified using western blot, the system was worked in K562 cells. Then with different sequence-targeted gRNA transfection, the locus-specific DNA binding proteins or RNAs on motifs can be identified using MS-based proteomics or next generation sequencing techniques respectively.

Since the DNA methylation data of these samples identified significant difference in gout patients and healthy individuals, we speculate that the transcriptome profiling is different between the parties. However, in the preliminary study of peripheral blood mononuclear cells (PBMC) samples from gout patients and healthy individuals, the RNA sequencing data obtained using bulk RNA-seq technique did not reveal any difference of gene expression between two groups. The limitation of the bulk RNA-seq is that it profiles

the transcriptomes from a large population, which may conceal the information from individual cells. PBMC was a mixed population which contained heterogeneity. Therefore, we access the transcriptome of a mixed population at individual cell level using single cell RNA-seq technique. Moreover, it helped to identify rare cell population. The result from single cell RNA-seq showed that there was a fewer cell types in one gout patient sample. However, this result was preliminary because the selected cell population of the gout patient sample was much smaller than the healthy individual sample. The rest of the samples of the cohort will be analyzed to provide a solid conclusion.

We also applied optimized atac-seq technique on the PBMC samples from gout patients and healthy individuals. The analyzed results showed there was no significant difference in open chromatin regions between gout patients and matched healthy controls. Therefore, we will apply the single cell ATAC-seq technique on the same samples to focus on the open chromatin patterns on the individual cells rather than the large population. Optimization and application of single cell atac-seq technique is another part of the future plan of this project.

Appendix

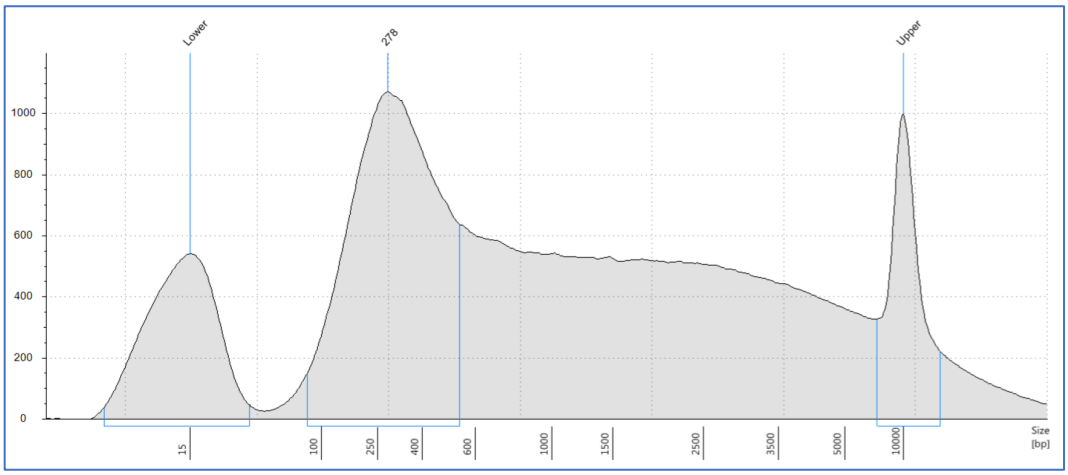

Figure S1 Tapestation Result of Sonication Condition 1

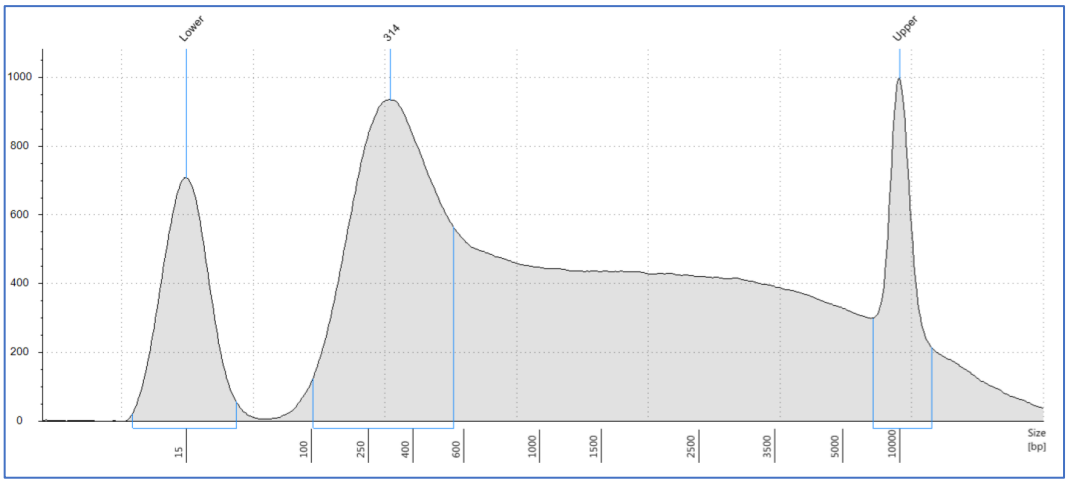

Figure S2 Tapestation Result of Sonication Condition 2

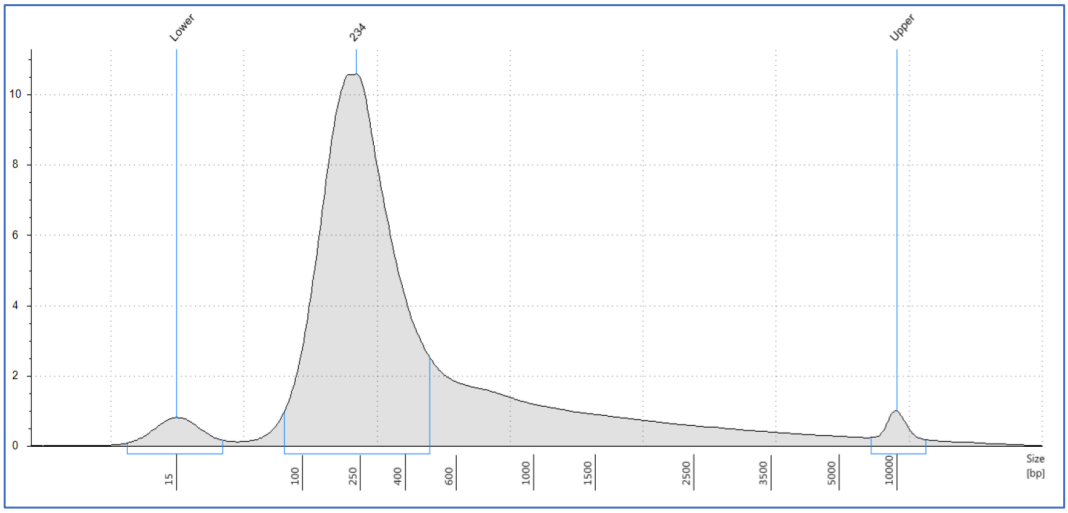

Figure S3 Tapestation Result of Sonication Condition 3

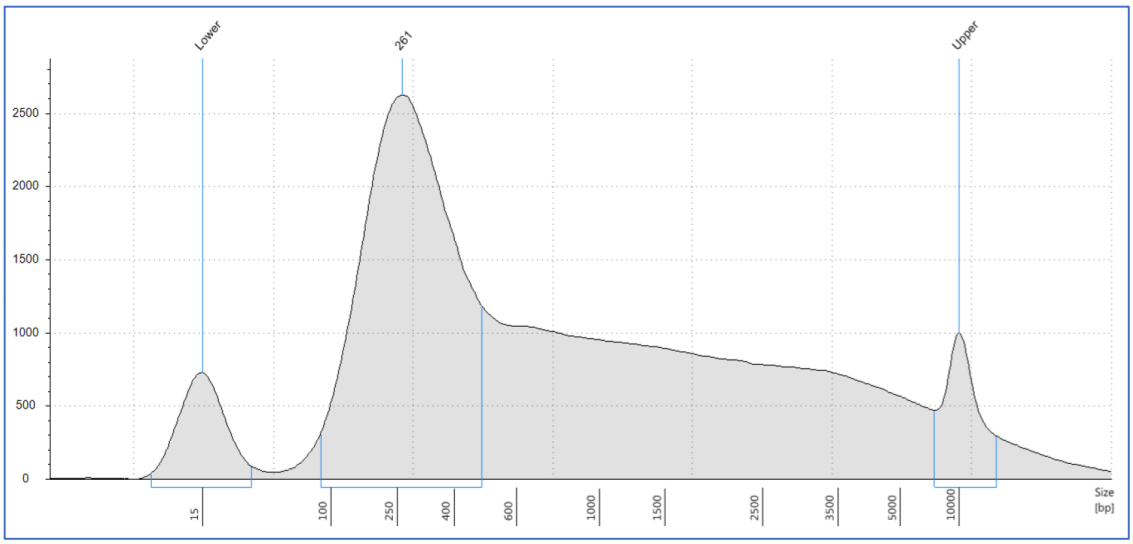

Figure S4 Tapestation Result of Sonication Condition 4

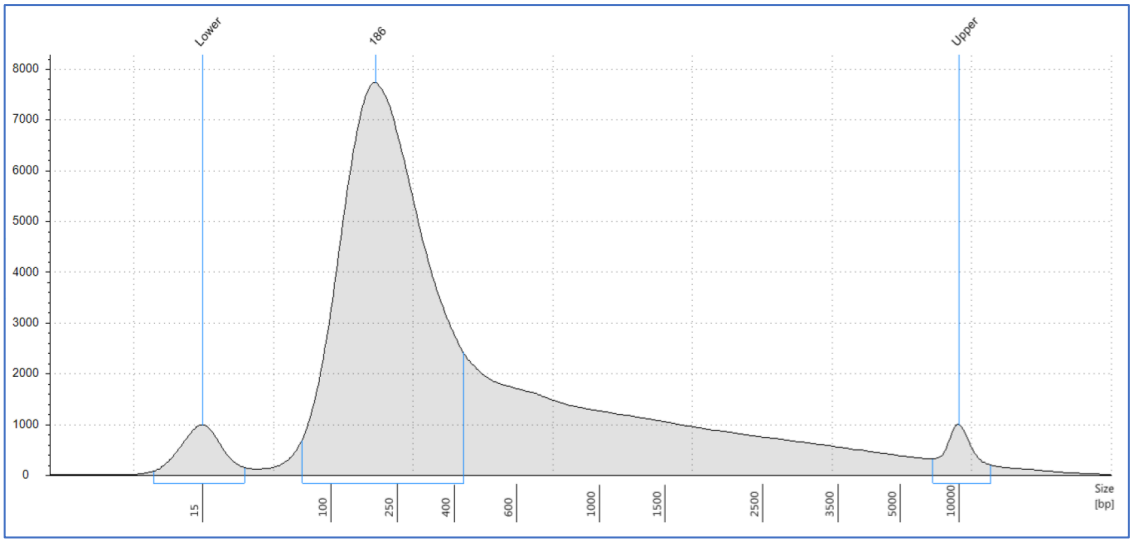

Figure S5 Tapestation Result of Sonication Condition 5

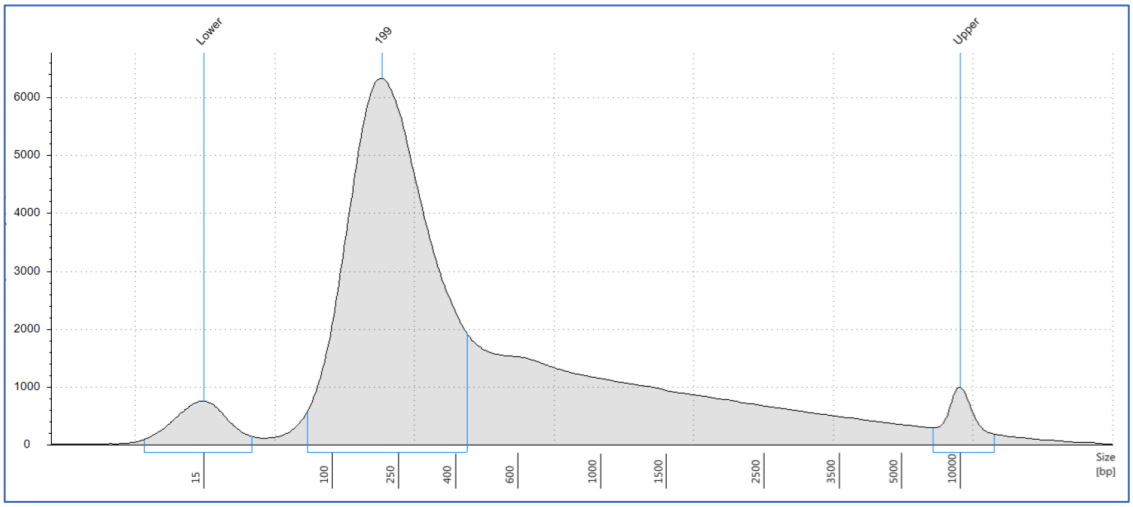

Figure S6 Tapestation Result of Sonication Condition 6

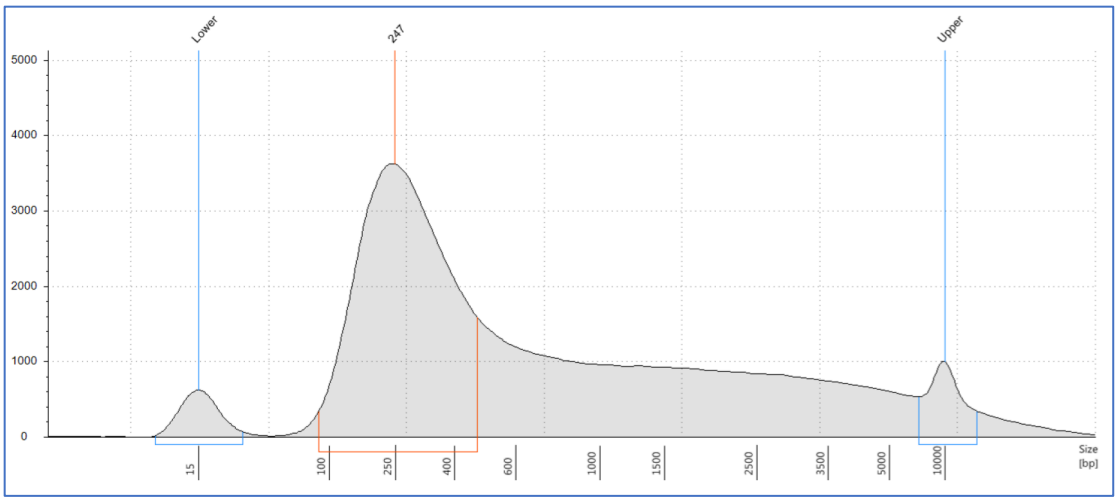

Figure S7 Tapestation Result of Sonication Condition 7

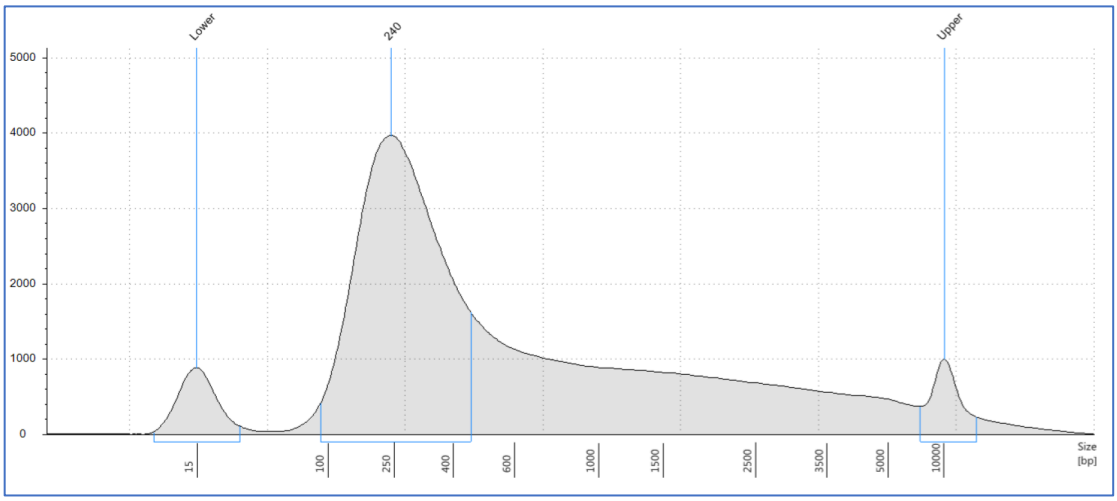

Figure S8 Tapestation Result of Sonication Condition 8

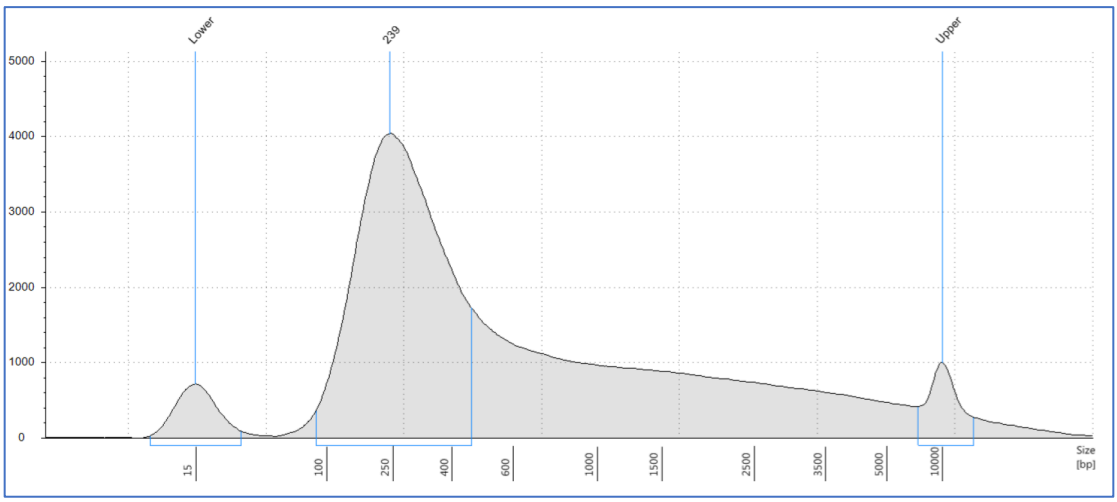

Figure S9 Tapestation Result of Sonication Condition 9

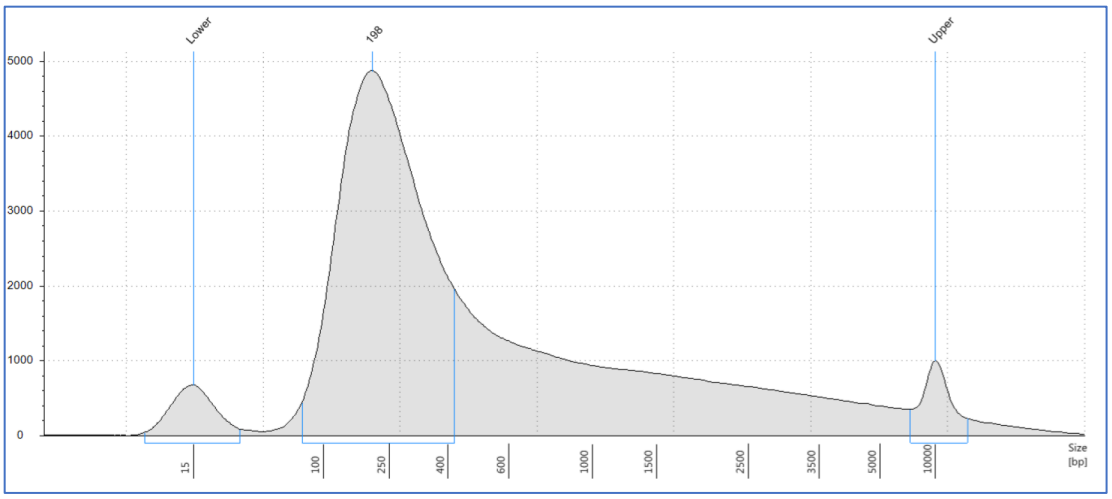

Figure S10 Tapestation Result of Sonication Condition 10

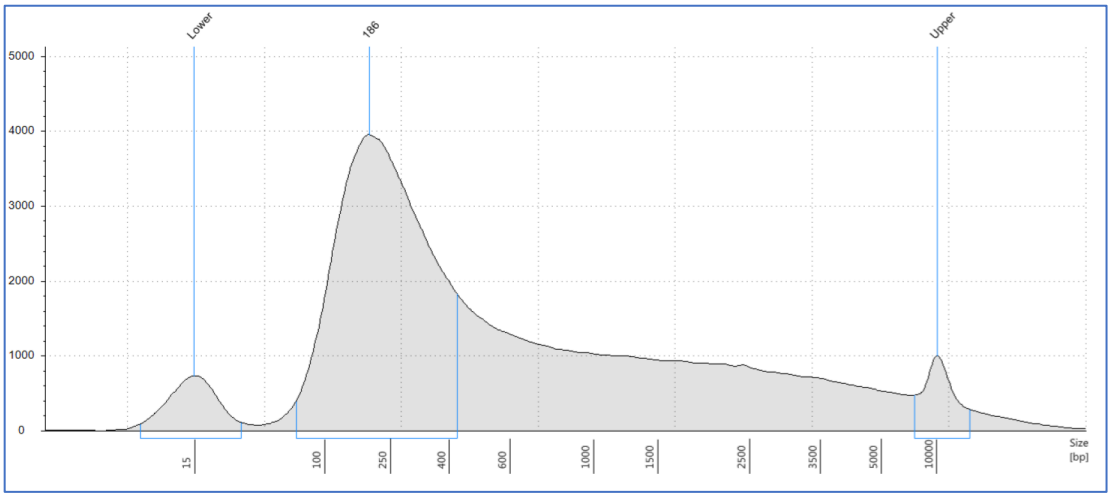

Figure S11 Tapestation Result of Sonication Condition 11

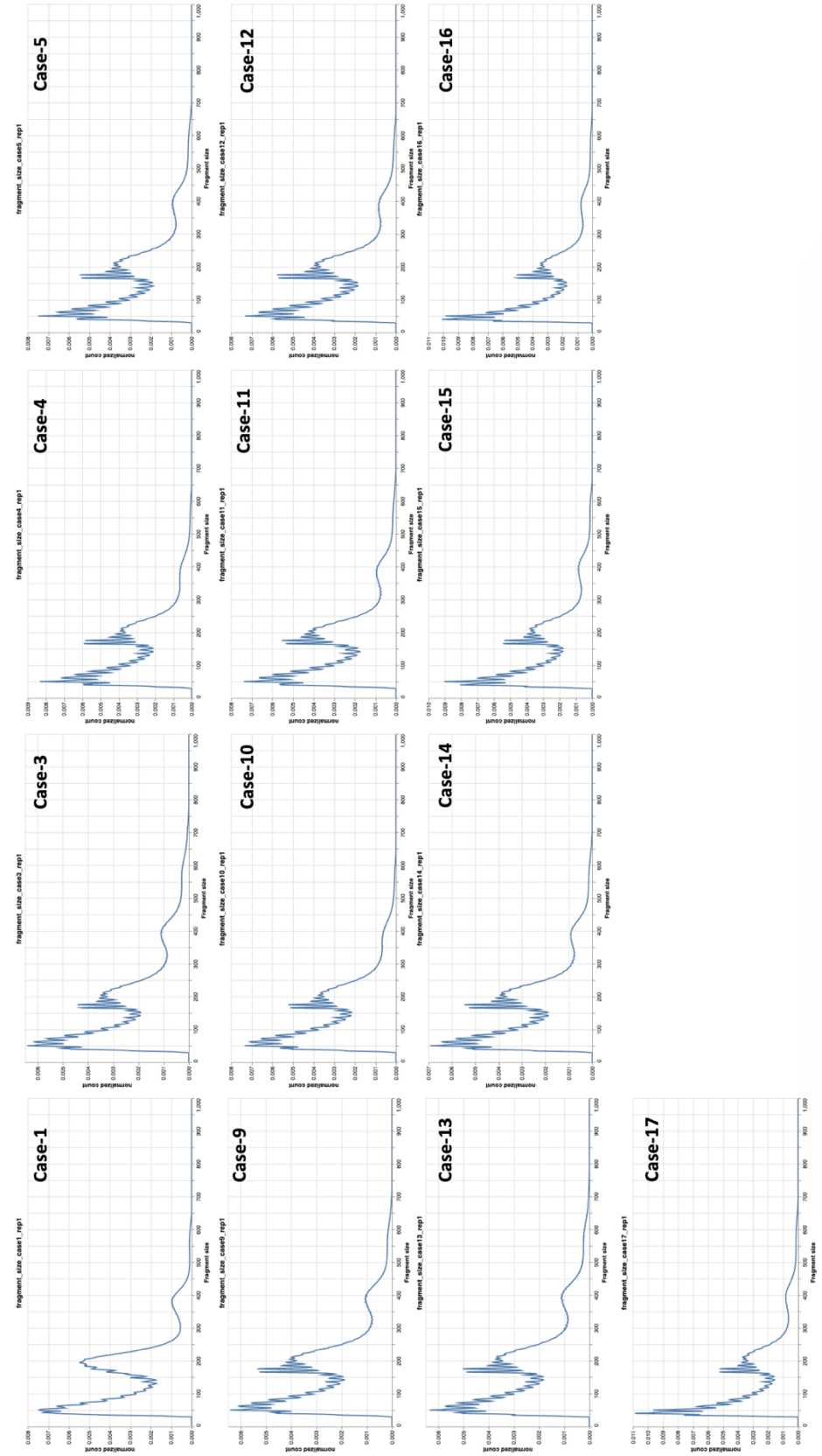

Figure S12 Distribution of Fragment Size of Gout Patients Samples

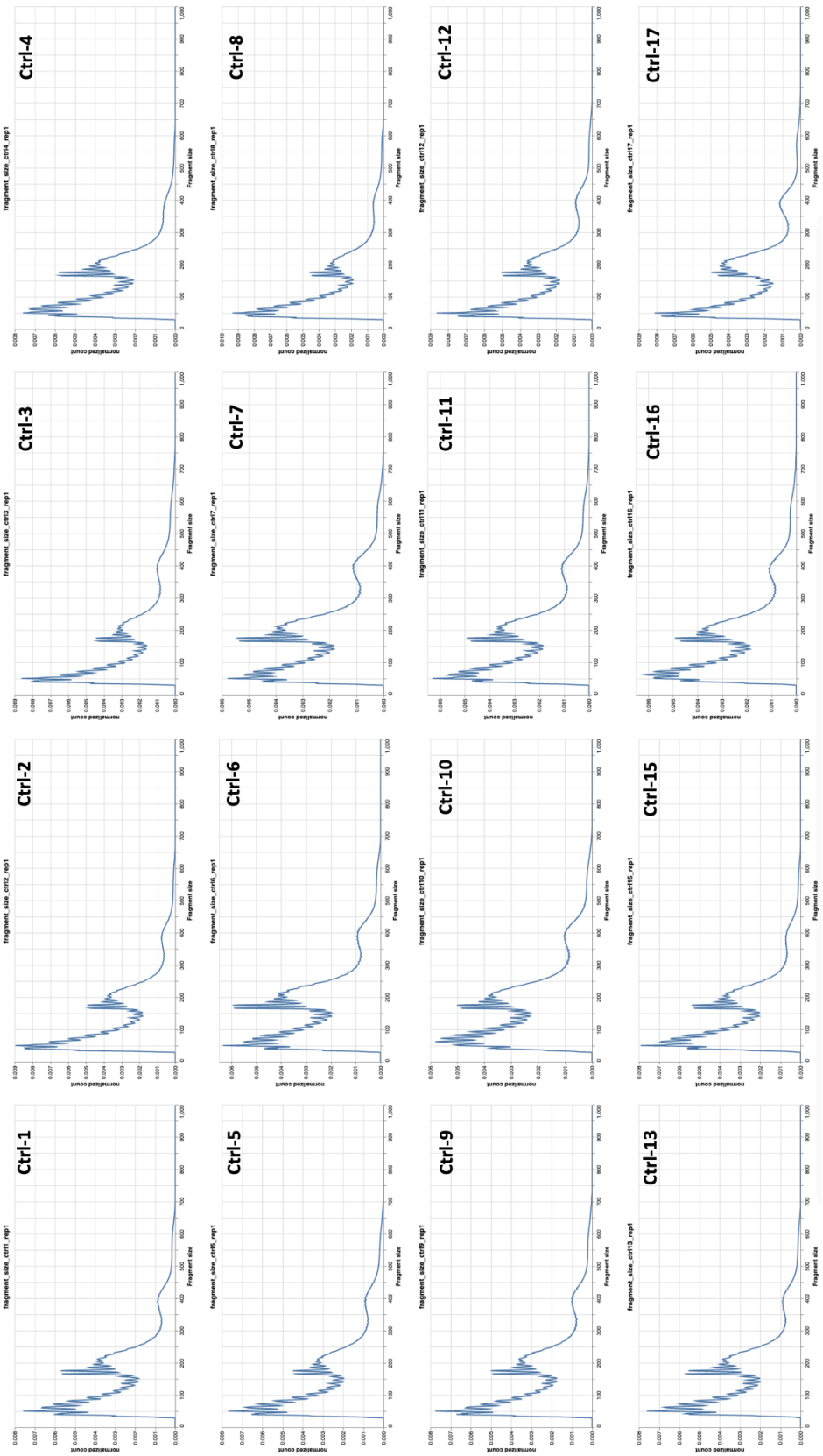

Figure S13 Distribution of Fragment Size of Healthy Individual Samples

References

1. Günther, S., Piwon, N., Iwanska, A., Schilling, R., Meisel, H., & Will, H. (1996). Type, prevalence, and significance of core promoter/enhancer II mutations in hepatitis B viruses from immunosuppressed patients with severe liver disease. Journal of virology, 70(12), 8318-8331.

2. Ougolkov, A. V., Bilim, V. N., & Billadeau, D. D. (2008). Regulation of pancreatic tumor cell proliferation and chemoresistance by the histone methyltransferase enhancer of zeste homologue 2. Clinical cancer research, 14(21), 6790-6796.

3. Siepel, A., & Arbiza, L. (2014). Cis-regulatory elements and human evolution. Current opinion in genetics & development, 29, 81-89.

4. Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics, 13(1), 59.

5. Wang, M., Zhang, K., Ngo, V., Liu, C., Fan, S., Whitaker, J. W., Chen, Y., Ai, R., Chen, Z., Wang, J., Zheng, L, & Zheng, L. (2019). Identification of DNA motifs that regulate DNA methylation. BioRxiv, 573352.

6. Wang, M., Zhang, K., Ngo, V., Liu, C., Fan, S., Whitaker, J. W., Chen, Y., Ai, R., Chen, Z., Wang, J., Zheng, L, & Zheng, L. (2019). Identification of DNA motifs that regulate DNA methylation. BioRxiv, 573352.

7. Wan, J., Su, Y., Song, Q., Tung, B., Oyinlade, O., Liu, S., Ying, M., Ming, GL., Song, H., Qian, J., Xia, S., & Zhu, H. (2017). Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. Elife, 6, e20068.

8. Tsui, C., Inouye, C., Levy, M., Lu, A., Florens, L., Washburn, M. P., & Tjian, R. (2018). dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. Proceedings of the National Academy of Sciences, 115(12), E2734-E2741.

9. Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., Dickerson, K. E., Xie, S., Hon, G. C., Xuan, Z., Zhang, M. Q., Shao, Z., & Xu, J. (2017). In situ capture of chromatin interactions by biotinylated dCas9. Cell, 170(5), 1028-1043.

10. Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. Genomics, 92(5), 255-264.

11. Wojdacz, T. K., Møller, T. H., Thestrup, B. B., Kristensen, L. S., & Hansen, L. L. (2010). Limitations and advantages of MS-HRM and bisulfite sequencing for single locus methylation studies. Expert review of molecular diagnostics, 10(5), 575-580.

12. Heller, L., Jaroszeski, M. J., Coppola, D., Pottinger, C., Gilbert, R., & Heller, R. (2000). Electrically mediated plasmid DNA delivery to hepatocellular carcinomas in vivo. Gene Therapy, 7(10), 826.

13. What is Gout? Retrieved from https://www.arthritis.org/about-arthritis/types/gout/what-is-gout.php

14. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics, 10(1), 57.

15. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics, 10(1), 57.

16. Olsen, T. K., & Baryawno, N. (2018). Introduction to Single-Cell RNA Sequencing. Current protocols in molecular biology, 122(1), e57.

17. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., McCarroll, S. A. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5), 1202-1214.