UNIVERSITY OF CALIFORNIA SAN DIEGO

Systematic Evaluation of ChIP-seq Workflow with Automation

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Yuwei Cao

Committee in charge:

Professor Alon Goren, Chair
Professor Nan Hao, Co-Chair
Professor Heidi Cook-Andersen

2022

The Thesis of Yuwei Cao is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ChIP-seq	Chromatin immunoprecipitation followed by sequencing
H3K27ac	Histone H3 Lysine 27 Acetylation
H3K27me3	Histone H3 Lysine 27 Trimethylation
SETD5	SET Domain Containing 5
CR	Chromatin Regulator
ASD	Autism Spectrum Disorder
FA	Formaldehyde
DSG	Disuccinimidyl Glutarate
IGV	Integrative Genomics Viewer
TSS	Transcription Start Site
PCR	Polymerase Chain Reaction
LB3	Lysis Buffer 3
EBB	Empigen BB Lysis Buffer

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Alon Goren as the chair of my committee and my mentor. He has been very supportive since I joined the lab in the beginning of the pandemic. It is his mentorship and guidance that helped me to learn many invaluable things as a scientist. He has also inspired me to continue my academic career and pursue a PhD degree.

I would also like to acknowledge Dr. Sven Heinz, as his ChIP-seq protocol served as the basis for the development of the automated protocol in my thesis project. He provided many constructive suggestions to improve the efficiency of automation.

I want to thank Professor Nan Hao and Professor Heidi Cook-Andersen for joining my committee and taking time to review my thesis.

I also want to thank the PhD Candidate Lauren Hodge of Benner and Goren Lab, from whom I learned how to conduct ChIP-seq experiments and data analysis.

Lastly, I would like to thank all the members of the Goren Lab for being supportive and encouraging. They have provided a friendly lab environment that taught me the importance and joy of collaborative teamwork.

Material from parts of the Introduction, Results, and Discussion of this master's thesis is currently being prepared for submission for publication. The thesis author is the primary researcher and author of this material.

ABSTRACT OF THE THESIS

Systematic Evaluation of ChIP-seq Workflow with Automation

by

Yuwei Cao

Master of Science in Biology

University of California San Diego, 2022

Professor Alon Goren, Chair
Professor First Nan Hao, Co-Chair

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a technique widely used for systematic *in vivo* mapping of protein-DNA binding and interaction.

However, the experimental process of manual ChIP-seq is laborious and time-consuming. Therefore, to improve efficiency and accelerate the workflow, the goal of my project is to develop an automated ChIP-seq protocol with a robotic liquid handling platform and use it to systematically evaluate the ChIP-seq assay conditions. In the experimental results, we have demonstrated that our automated high-throughput ChIP-seq workflow was reproducible and maintained similar signal-to-noise ratio as the manual workflow, while reducing human errors and total experimental time. The automated ChIP-seq has useful applications, such as antibody titration and evaluation of multiple conditions at once. Besides automation, we assessed crosslinking conditions, comparing formaldehyde (FA) single crosslinking and DSG/FA double crosslinking for the ChIP-seq of chromatin regulator SETD5. We also optimized chromatin shearing conditions in the ChIP-seq workflow, as we tested different sonicators as well as evaluated multiple different lysis buffers for shearing.

INTRODUCTION

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a technique for epigenomic profiling and understanding the interactions between DNA and DNA-associated proteins. It is widely used to characterize and map the genomic localization and binding of histone modifications, histone modifiers, transcription factors, and other DNA binding proteins in regulatory regions (Barski *et al*., 2007; Mikkelsen *et al*., 2007; Johnson *et al*., 2007).

Although the method is highly useful in studying the epigenome, the overall experimental workflow of ChIP-seq remains labor-intensive as it involves multiple steps (Park, 2009). Briefly, the process starts with crosslinking DNA-associated proteins such as transcription factors and histones to the genomic DNA, and then the crosslinked cells are sheared to shorter fragments of chromatin, ideally around 150-300 base pairs (Ren *et al*., 2000; Landt *et al*., 2012). Sheared chromatin is enriched with an antibody specific to the protein of interest to capture the DNA-protein complex. After immunoprecipitation, the ChIP sample is washed with buffers, and the crosslinks are reversed. The ChIP DNA is then purified and used for the downstream library generation for next generation sequencing (Busby *et al*., 2016). As stated, the process is not only manually intensive and time consuming but offers many places for the introduction of mistakes and inconsistency to the results. To overcome these challenges and improve the efficiency of the process, one solution is to incorporate automation into the process using liquid handling robots.

There are existing publications that have reported automated robotic methods for high-throughput ChIP-seq (Garber *et al*., 2012; Aldridge *et al*., 2013; Gasper *et al*., 2014). However, most of them are partially automated. The method from Garber *et al*. only automated the SPRI DNA cleanup steps to analyze the regulatory transcription factors and epigenetic marks (Garber *et al*., 2012). On the other hand, the method from Gasper *et al*. only automated the

immunoprecipitation and washing steps. The later library construction and DNA cleanup steps were still done manually (Gasper *et al*., 2014). Although the method from Aldridge *et al*. automated most of the ChIP-seq workflow, it was based on a longer ChIP-seq workflow that required purification of ChIP-enriched DNA before library preparation. The results also suggested an optimal cell number of 10 million and minimum of 1 million cells (Aldridge *et al*., 2013). We developed our automated ChIP-seq method around a newly optimized ChIP-seq protocol (Texari *et al*., 2021) that is suitable for smaller cell numbers. Because the protocol omits the intermediate DNA purification step after immunoprecipitation and proceeds to library preparation without removing the beads, it not only reduces the workflow time but also allows for preservation of the already low amount of cell inputs which enables rapid, sensitive, and robust generation of high-quality sequencing libraries from smaller cell numbers (Texari *et al*., 2021).

Therefore, to accelerate the ChIP-seq workflow and improve the overall efficiency for the in vivo mapping of protein-DNA binding, with this newly optimized on-bead ChIP-seq protocol (Texari *et al*., 2021), one of the goals of my project is to incorporate a robotic liquid handling platform and develop a robust automated high-throughput ChIP-seq method. The other goal is to systematically evaluate and optimize the experimental conditions for the ChIP-seq workflow. This includes identifying the optimal shearing conditions, lysis buffer selections, crosslinking conditions, and antibody titrations (**Fig. 1**).

In the ChIP-seq experiments, we chose to focus on two key histone modifications, acetylation and trimethylation of lysine 27 on histone H3 (H3K27ac and H3K27me3). H3K27ac is associated with open accessible chromatin as an active mark for gene transcription. On the other hand, H3K27me3 is often linked to compacted chromatin which can lead to repression of

gene transcription (Zhou *et al.*, 2011). We also chose the chromatin regulator SETD5, which is found to be involved in the neurodevelopment of autism spectrum disorder (ASD) (Nisar *et al.*, 2019). A previous study has performed SETD5 ChIP-seq in mice (Sessa *et al.*, 2019), but to our knowledge, there have been no published studies that reported SETD5 ChIP-seq experiments in human cells. Therefore, we are interested in conducting endogenous SETD5 ChIP-seq to study the molecular role of SETD5 as a chromatin regulator in human cells. In addition, we have also included two histone acetyltransferases, HAT1 and EP300. Both are chromatin regulators that function in histone acetylation and regulate transcription by binding transcription factors (Ogryzko *et al.*, 1996; Benson *et al.*, 2007). We performed both automated and manual ChIP-seq with monoclonal antibodies of H3K27ac, H3K27me3, HAT1, and EP300, as well as polyclonal SETD5 antibody in a human breast cancer cell line, HeLa-S3. We used the liquid handling platform Bravo, from Agilent Technologies, as our automated system and included manual samples to compare the reproducibility of automated (Bravo) to the manual experiments. After verifying the feasibility of our ChIP-seq conditions and with the implementation of automation, we also added antibody titration for the SETD5 antibody, aiming to find the amount of antibody that produces successful endogenous SETD5 ChIP-seq. To control for variability, we included two technical replicates for each antibody and condition.

We also evaluated cell crosslinking methods to assess the optimal ChIP assay conditions. Crosslinking is an important initial step in ChIP as it is used to preserve the protein-DNA interaction prior to immunoprecipitation. The conventional way of crosslinking involves a single step of formaldehyde (FA) to create covalent bonds between protein and DNA. The covalent bonds formed are strong yet reversible and do not interfere with most downstream steps. FA easily permeates living cells and reacts very quickly to capture transient functional protein

interactions. In ChIP, shorter incubation time (~10 mins) and lower concentration (1% v/v) of FA is sufficient to establish crosslinking reactions while avoiding the loss of important protein-DNA complexes. However, FA is a small molecule and only captures interactions in the range of 2.3–2.7 Å, which requires the reactive amino acids on proteins to be in very close proximity. The small size of FA limits capturing longer interactions between protein and protein (Sutherland *et al*., 2008). The single-step FA crosslinking does not preserve all the protein-DNA interactions. Proteins that do not crosslink well with DNA include transcription factors that are in hyperdynamic equilibrium with target DNA and transcriptional coactivators that interact with chromatin through protein-protein interactions in a protein complex (Tian *et al*., 2014). Therefore, to overcome the problem, a two-step double crosslinking technique has been developed to combine FA and larger FA-compatible crosslinkers, such as disuccinimidyl glutarate (DSG) to increase the yield of interacting proteins. DSG is less stringent than FA and forms crosslinks that are approximately 7.7 Å, making it a better choice to capture longer protein-protein interactions. This double crosslinking process using DSG first to stabilize large complex of multiple proteins, followed by FA treatment for protein-DNA crosslinking reaction is shown to be a highly effective method for crosslinking transcription factor to promoters or protein complexes and identifying genomic targets for protein-DNA binding within native chromatin environment. DSG/FA double crosslink is shown to yield ChIP reproducible results and robust signal-to-noise ratios for transcriptional factors such as NF-κB and STAT3 (Nowak *et al*., 2005). Therefore, we were interested in whether the traditional FA single crosslinking or the DSG/FA double crosslinking would yield successful ChIP-seq for SETD5, a DNA-associated protein which is thought to function as a chromatin regulator.

Another aspect of the ChIP-seq workflow we evaluated was chromatin shearing. We tested two sonicators, Covaris LE220-plus and Active Motif's PIXUL (Bomsztyk *et al*., 2019). Both machines shear one column at a time, enabling 12 different conditions per 96-well plate. Besides the sonicators, we also compared three lysis buffers, SDS, Empigen BB and LB3 that can be used for lysing the cells for shearing, both for shearing pattern and later ChIP-seq results produced from sheared chromatin.

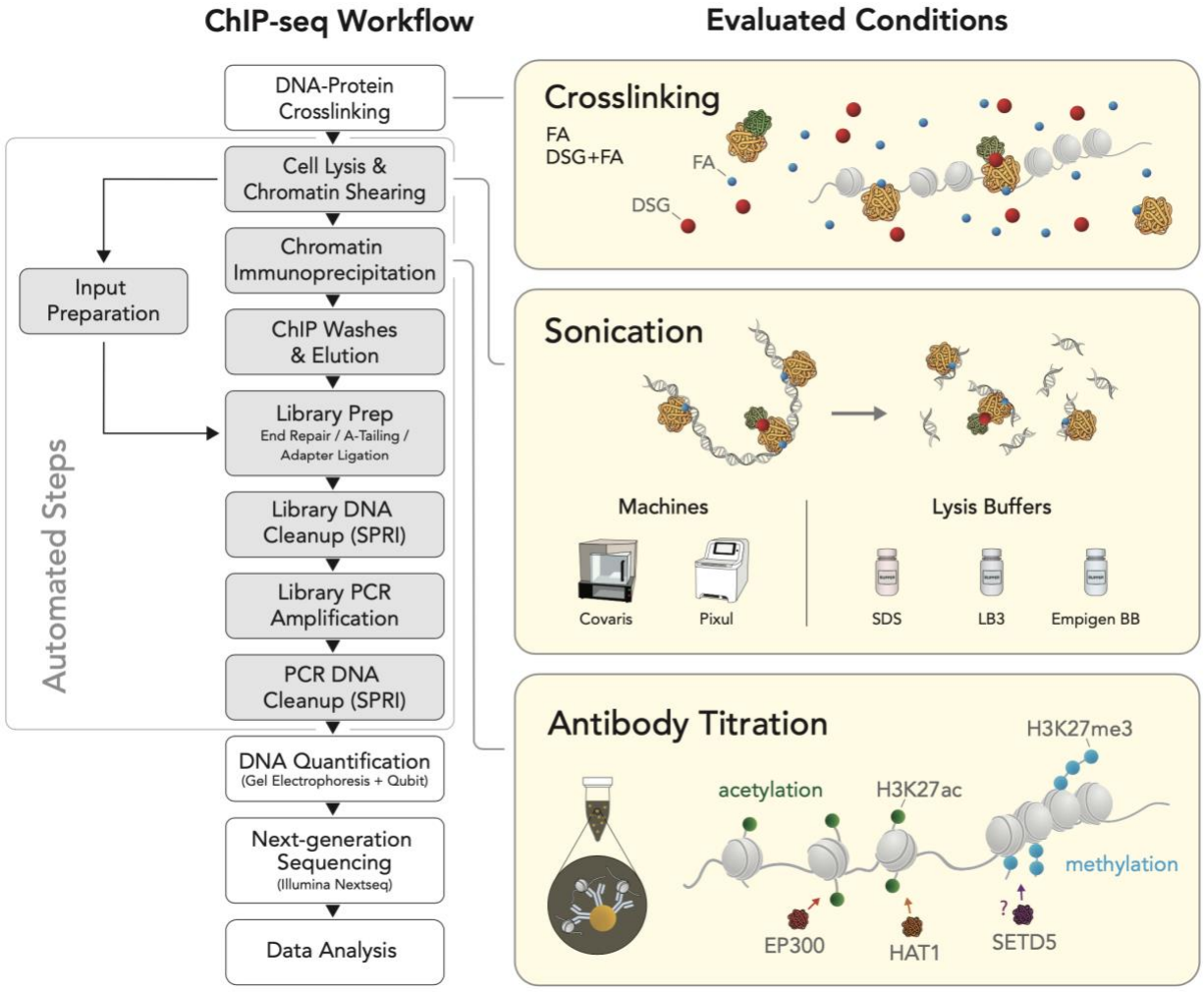

Figure 1.1: Overview of the ChIP-seq workflow with the automated steps highlighted in gray boxes and evaluated conditions in yellow boxes.

RESULTS

Comparison of Shearing Patterns

Prior to the immunoprecipitation step in the ChIP-seq process, the chromatin needs to be sheared to a shorter size, ideally around 150-300 base pairs (Ren *et al*., 2000; Landt *et al*., 2012). Chromatin shearing often utilizes sonicators. We tested two ultrasonic sonicators, Covaris LE220-plus and PIXUL.

We first started to optimize Covaris LE220-plus for chromatin shearing. Covaris uses the Adaptive Focused Acoustics (AFA) Technology to process samples with highly controlled bursts of focused high-frequency acoustic energy. Shearing forces on the LE220-plus can be controlled by optimizing the settings, such as the peak incident power (PIP) (the instantaneous ultrasonic power in Watts), duration of treatment, and duty factor (DF) (% of time that the ultrasound signal is applied to the sample). We used two models of the Covaris sonicators, E220 and LE220-plus. The chromatin sheared with E220 served as a control for our shearing test on LE220-plus because the settings on E220 have already been optimized and shown to work consistently. However, E220 only treats each well individually, which is time-consuming and is only suitable for a smaller number of samples. On the other hand, the LE220-plus treats one column at a time, so in theory it can process samples in 96-well plates much faster and efficiently than E220. To begin, we optimized chromatin shearing on LE220-plus by comparing lysis buffers and changing parameters.

However, the results were not as ideal as we aimed for. As shown in the gel in **Fig. 2.1a**, the chromatin was sheared slightly shorter than the ideal size, which is around 150 bp, and the shearing consistency across the column was poor. After we optimized the peak incident power (PIP) and % duty factor (DF) parameters, the shearing consistency was also improved, and the

size of shearing was approximately 150 bp (**Fig. 2.1b**). Unfortunately, even though the shearing pattern looked optimal and similar to the control (chromatin sheared on E220), when we performed ChIP-seq with the chromatin sheared on LE220-plus, the signal-to-noise ratio was still significantly lower than the control (results not shown).

We then tested Active Motif's PIXUL, a sonicator that utilizes focused ultrasonic waveforms created by ultrasound transducers to shear chromatin. Similar to LE220-plus, PIXUL also sonicates one column at a time, supporting 12 different conditions for a 96-well microplate. In a published study, they compared PIXUL to Covaris LE220-plus and concluded that PIXUL yielded greater shearing consistency and was easier and more cost-effective to use than Covaris LE220-plus (Bomsztyk *et al*., 2019).

Our results agreed with the literature, as the shearing was consistent across the columns. When we compared the shearing pattern with FA single crosslinked and DSG/FA double crosslinked HeLa-S3 cells as well as the two lysis buffers, SDS and LB3, the LB3 lysis buffer produced a more ideal shearing pattern compared to SDS, especially when cells crosslinked with FA only were sheared. However, the chromatin shearing pattern was not ideal for DSG/FA double crosslinked cells as there was still an unsheared upper band present (**Fig. 2.2a**). Therefore, we tested a harsher lysis buffer, Empigen BB, for the shearing of DSG/FA dual crosslinked HeLa-S3 cells. The shearing pattern was clearly improved as a higher percent of the chromatin was sheared to the optimal size of around size of around 150bp, indicated by the darker bands (**Fig. 2.2b**). Up to this point, the shearing tests on PIXUL were done with the same process time of 60 minutes. To futher evaluate the ability of the LB3, we increased the shearing time and the upper unsheared band started to disappear after 108 minutes of shearing (**Fig. 2.2c**).

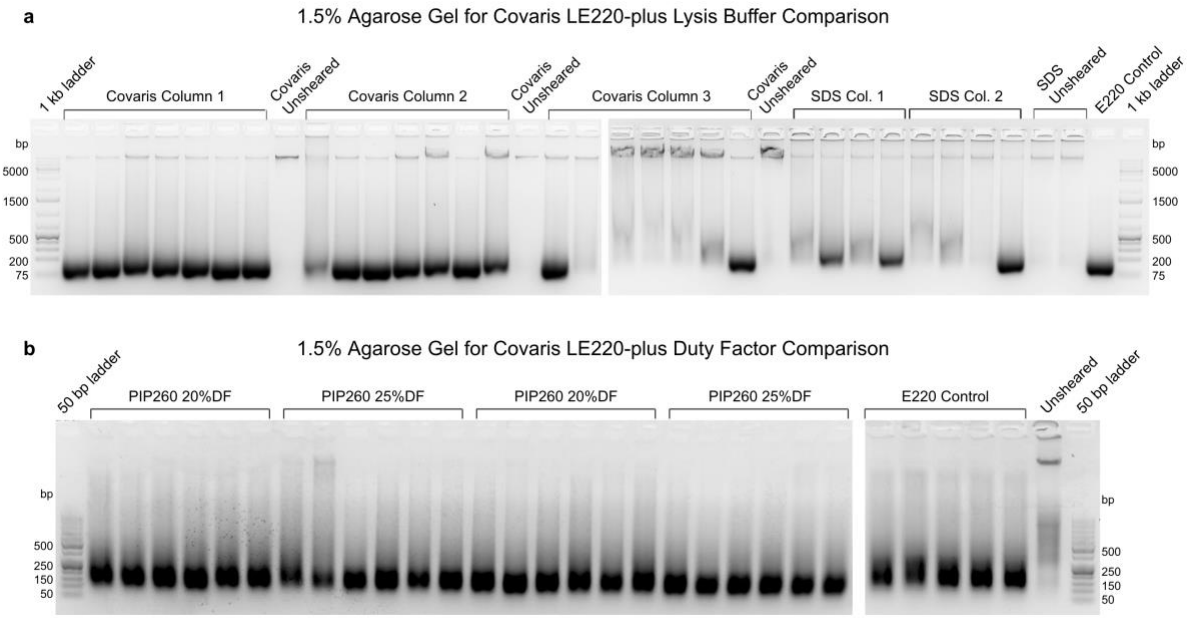

Figure 2.1: Comparison of shearing patterns for Covaris LE220-plus with various shearing conditions. All the gels included an unsheared negative control for each shearing condition. The DNA ladders used are 50bp and 1kb Plus from GeneRuler. **(a)** Agarose gel shows the shearing pattern for chromatin sheared on Covaris LE220-plus comparing Covaris lysis buffer and SDS lysis buffer. Besides the unsheared negative control, chromatin sheared on E220 was run as a positive control. **(b)** Agarose gel shows shearing pattern for chromatin sheared on Covaris LE220-plus with changes in the parameters. The peak incident power (PIP) is the measure of the instantaneous ultrasonic power applied to the sample (range from 2.5 to 500 Watts), and the duty factor (DF) is the % of time that the ultrasound signal is applied to the sample (range from 0.1% to 50%). Chromatin sheared on E220 was also run as a positive control.

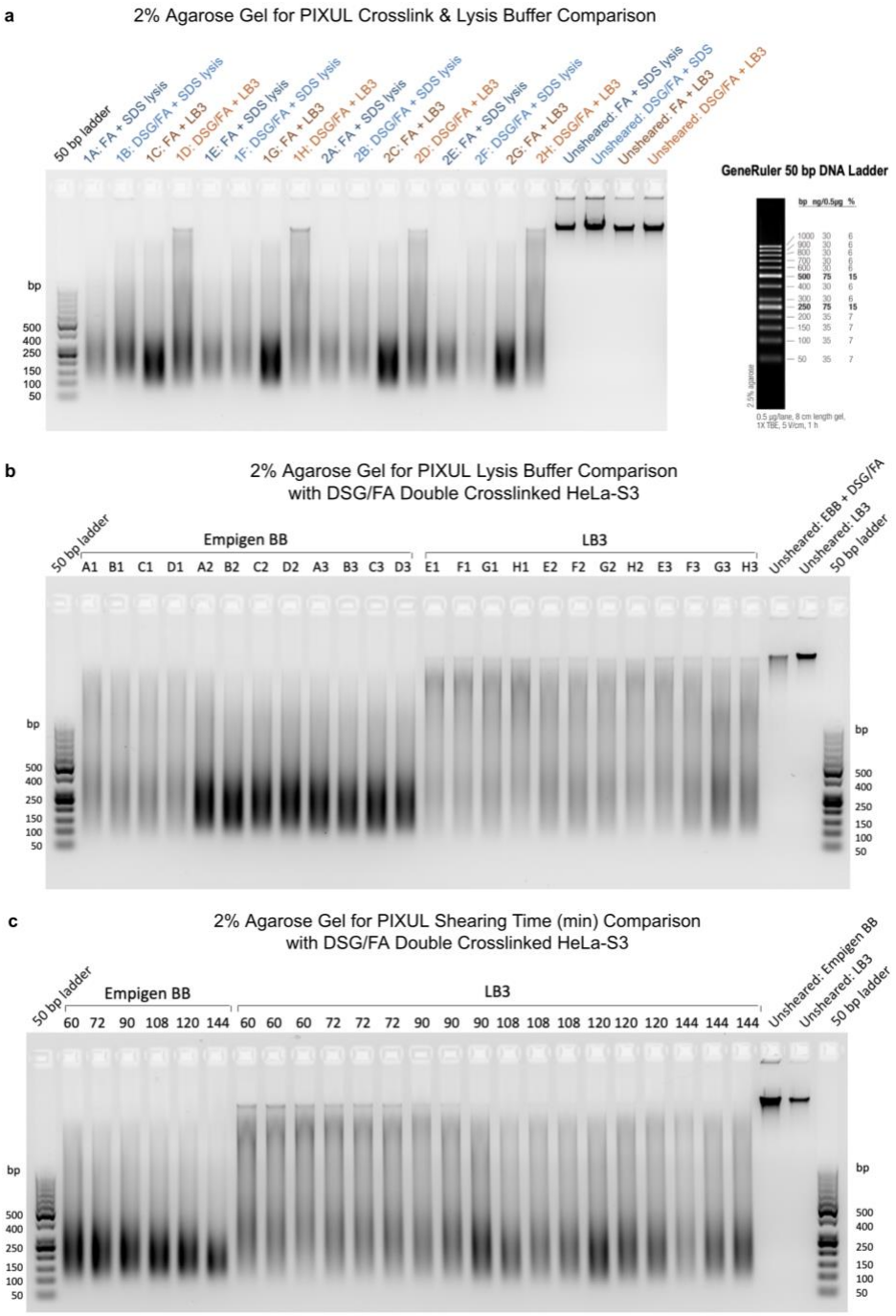

Figure 2.2: Comparison of shearing patterns for PIXUL with various crosslinking and shearing buffer conditions. All the gels included an unsheared negative control for each shearing condition. The DNA ladders used are 50bp from GeneRuler. **(a)** Agarose gel shows shearing pattern for chromatin sheared on PIXUL, comparing formaldehyde (FA) crosslinked HeLa-S3 cells and DSG/FA double crosslinked HeLa-S3 cells, as well as comparing the lysis buffers, SDS and LB3. **(b)** Agarose gel shows shearing pattern for chromatin sheared on PIXUL with double crosslinked HeLa-S3 cells, comparing the lysis buffers, Empigen BB and LB3. **(c)** Agarose gel shows shearing pattern for chromatin sheared on PIXUL with double crosslinked HeLa-S3 cells and Empigen BB or LB3 lysis buffer, comparing various processing time (minutes).

Comparison of Crosslinkers: FA Single Crosslink vs. DSG/FA Double Crosslink

For crosslinking methods, we compared the effects of single crosslinking with formaldehyde only versus double crosslinking with DSG and formaldehyde for capturing interactions of histone modifications, H3K27ac and H3K27me3, and the chromatin regulator SETD5. Our experimental results indicated that there was no obvious difference in both H3K27ac and H3K27me3 ChIP-seq experiments between the automated Bravo workflow and the manual workflow, as the read coverage tracks showed similar peaks (**Fig. 3.1a, b**), and the metagene plot centered around TSS (**Fig. 3.2a**) and scatter plots (**Fig. 3.3a, c**) showed similar signal. However, with SETD5, double crosslinking with DSG/FA had more successful ChIP-seq results than single crosslinking with FA only. The read density tracks of SETD5 ChIP-seq visualized by IGV Genome Browser (Robinson *et al.*, 2011) showed noticeable SETD5 binding enrichment signals or peaks for both manual and Bravo DSG/FA SETD5 ChIP-seq experiments (**Fig. 3.1a**). On the other hand, both manual and Bravo single crosslinked FA only SETD5 ChIP-seq samples had no visible peaks at that specific genomic locus (**Fig. 3.1a**). To further evaluate the genomic localization of SETD5 along the gene bodies, the average ChIP-seq tag density, reads per base pair per gene, across the gene bodies in HeLa-S3 cells was found with Homer for each SETD5 sample (**Fig. 3.2b**). There were many small signals along the gene bodies, but the highest signal was around the transcription start site (TSS) and the signals were only distinctly shown for the DSG/FA double crosslinked ChIP samples (**Fig. 3.2b**). The results were further supported by the scatter plots that show the correlation in the samples for the read counts (log2) at the merged peaks of all the SETD5 ChIP-seq samples across the human genome. The signal-to-noise ratio between the DSG/FA and FA only ChIP-seq samples was shifted towards the

DSG/FA samples for both the Bravo and manual ChIP-seq methods, suggesting that DSG/FA is a better crosslinking condition for SETD5 ChIP-seq (**Fig. 3.3e**).

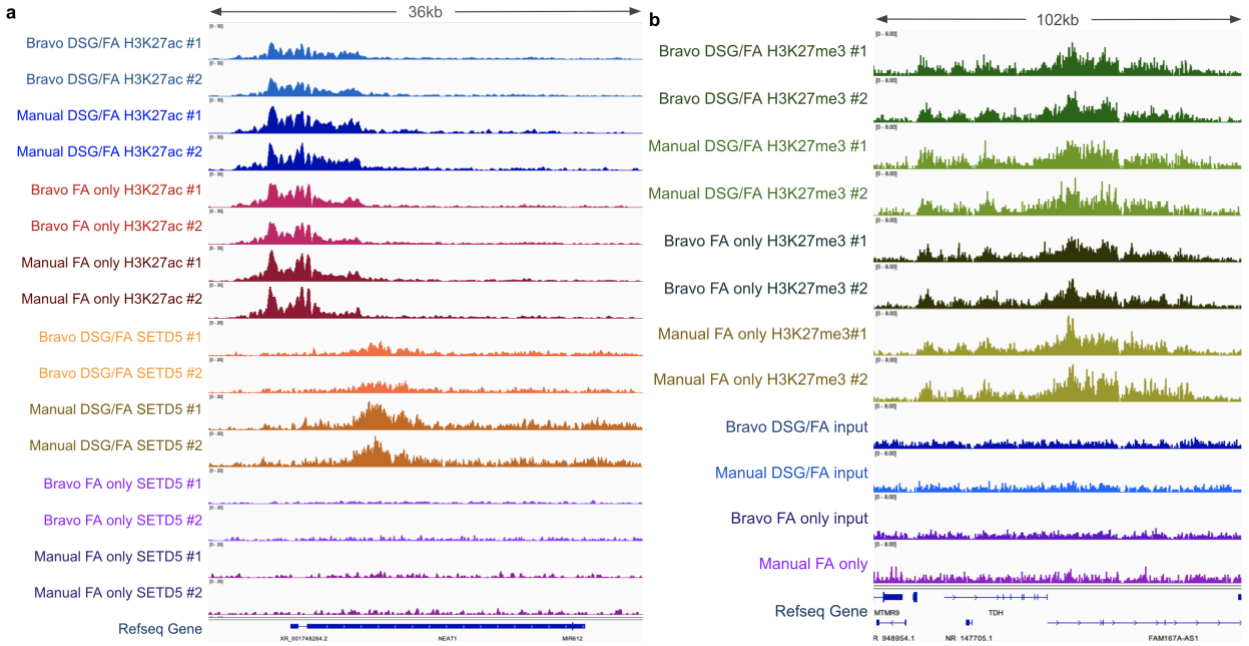

Figure 3.1: IGV Genome Browser tracks of read coverage for H3K27ac, H3K27me3, and SETD5 ChIP-seq experiments that compared DSG/FA double crosslinking and FA only single crosslinking, as well as the automated (Bravo) ChIP-seq workflow and the manual ChIP-seq workflow. **(a)** IGV Genome Browser tracks of read coverage show H3K27ac and SETD5 ChIP-seq binding signals (indicated as peaks) over representative genomic loci, *NEAT1*. The snapshot is taken at a window of 36kb on Chr11:65,414,222-65,450,290 with signal height scale of 0-35 for H3K27ac tracks and signal height scale of 0-20 for SETD5 tracks. Blue tracks show H3K27ac ChIP-seq with DSG/FA double crosslinked cells (Bravo and manual), red tracks show H3K27ac ChIP-seq with FA single crosslinked cells (Bravo and manual), orange tracks show SETD5 ChIP-seq with DSG/FA double crosslinked cells (Bravo and manual), and purple tracks show SETD5 ChIP-seq with FA single crosslinked cells (Bravo and manual). Each condition was done in replicates. **(b)** IGV Genome Browser tracks of read coverage show H3K27me3 ChIP-seq binding signals at a window of 102kb on Chr8:11,319,589-11,422,329 with signal height scale of 0-8. Each condition was done in replicates alongside with a control input that was not immunoprecipitated for each lysis buffer and crosslinking condition pair to indicate background signals.

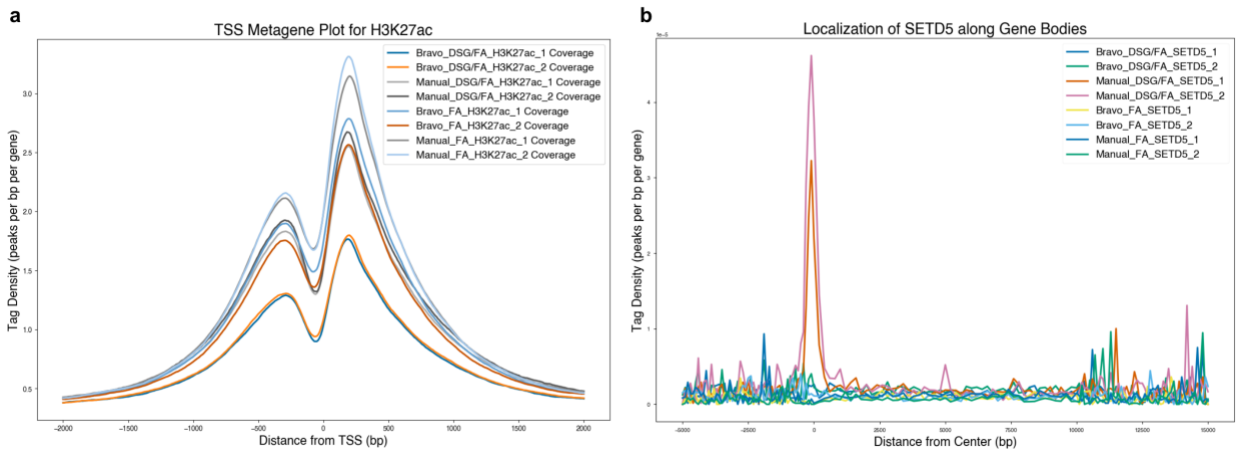

Figure 3.2: Metagene plots for H3K27ac and SETD5 ChIP-seq experiments that compared DSG/FA dual crosslinking and FA only single crosslinking, as well as the automated (Bravo) ChIP-seq workflow and the manual ChIP-seq workflow. **(a)** Metagene plot shows the average H3K27ac ChIP-seq tag density (peaks per base pair per gene) around the transcription start sites (TSS) in HeLa-S3 cells. The lines correspond to all the H3K27ac samples that are also shown in Fig. 3.1a. **(b)** Metagene plot shows the average SETD5 ChIP-seq tag density (peaks per base pair per gene) along the gene bodies in HeLa-S3 cells. The lines correspond to all the SETD5 samples that are also shown in Fig. 3.1a.

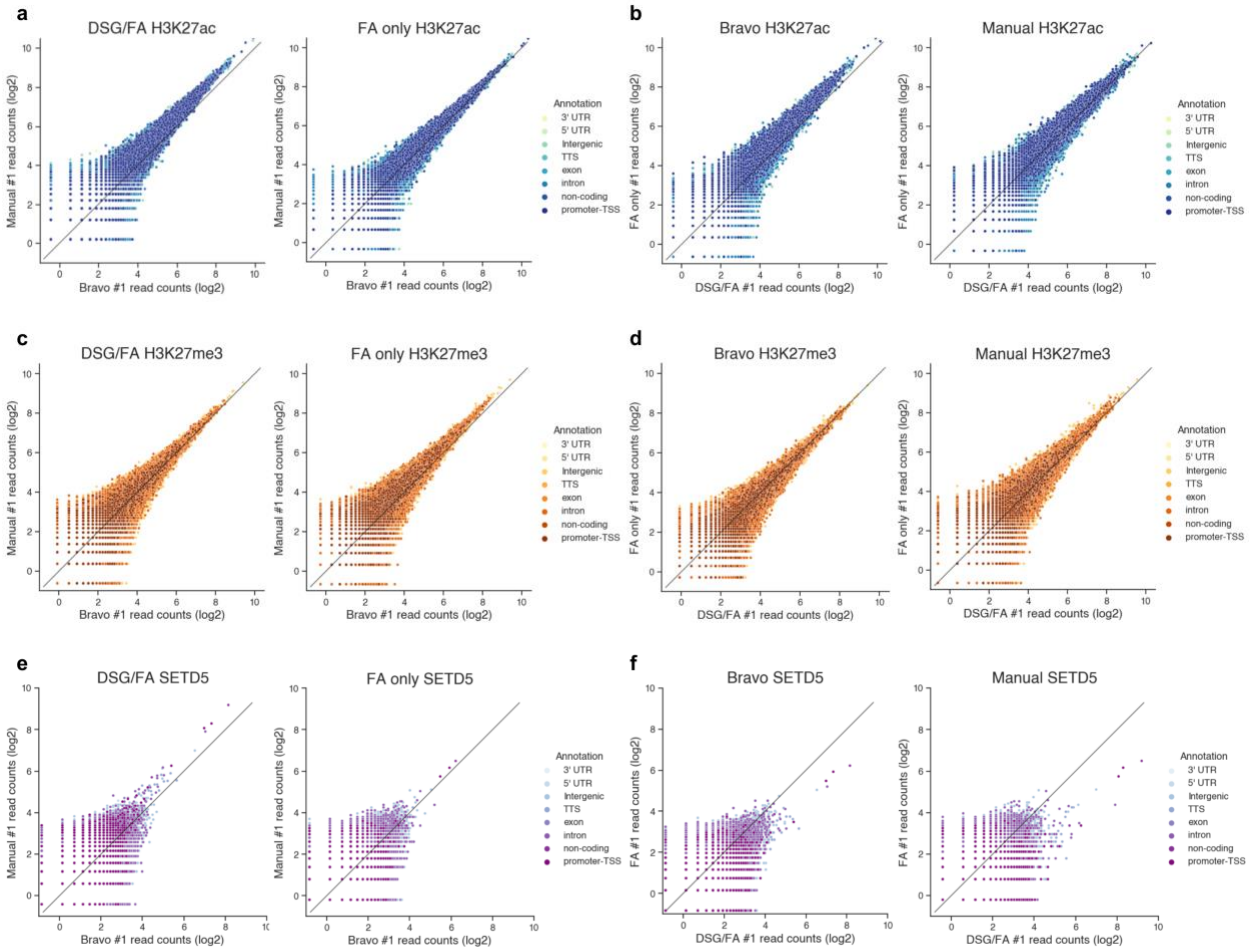

Figure 3.3: Scatterplots for for H3K27ac, H3K27me3, and SETD5 ChIP-seq experiments that compared DSG/FA dual crosslinking and FA only single crosslinking, as well as the automated (Bravo) ChIP-seq workflow and the manual ChIP-seq workflow. **(e)-(j)** Scatterplots show correlation in the samples for the read counts (log2) from 1000 base pairs around the merged peaks of all the H3K27ac, H3K27me3, or SETD5 ChIP-seq samples across the human genome hg38. The color of markers indicates the genomic regions where the peaks fall within. The diagonal y=x line represents slope =1. **e.** Correlation between the Bravo and manual H3K27ac ChIP-seq samples for the two crosslinking methods (DSG/FA versus FA only). **f.** Correlation between DSG/FA and FA only H3K27ac ChIP-seq samples for the two ChIP-seq methods (Bravo versus manual). **g.** Correlation between the Bravo and manual H3K27me3 ChIP-seq samples for the two crosslinking methods (DSG/FA versus FA only). **h.** Correlation between DSG/FA and FA only H3K27me3 ChIP-seq samples for the two ChIP-seq methods (Bravo versus manual). **i.** Correlation between the Bravo and manual SETD5 ChIP-seq samples for the two crosslinking methods (DSG/FA versus FA only). **j.** Correlation between DSG/FA and FA only SETD5 ChIP-seq samples for the two ChIP-seq methods (Bravo versus manual).

Comparison of Lysis Buffers: LB3 vs Empigen BB

From the shearing results with PIXUL, we concluded that LB3 produced a better shearing pattern than SDS lysis buffer (**Fig. 2.2a**), but it is still less ideal with DSG/FA double crosslinking compared to the Empigen BB lysis buffer (**Fig. 2.2b**). Since LB3 had given us successful ChIP-seq results, we wanted to test whether Empigen BB can achieve the same or even more optimal results. We conducted manual and Bravo ChIP-seq experiments with chromatin sheared in the two lysis buffers, LB3 and Empigen BB (EBB). As shown by the visualization of the read coverage tracks in IGV genome browser for H3K27ac ChIP-seq experiments (**Fig. 4.1a**), when comparing the lysis buffers, LB3 samples had significantly higher signals than the Empigen BB samples. The results were further supported by the metagene plot around the transcription start sites (TSS) (**Fig. 4.1b**). The average H3K27ac ChIP-seq tag density (peaks per base pair per gene) around TSS for the LB3 samples had higher average tag density than the Empigen BB samples. Moreover, we found the correlation in the samples by comparing the read counts between two experiments directly at the merged peaks of all the H3K27ac ChIP-seq samples. The signal-to-noise ratio between the LB3 and Empigen BB ChIP-seq samples was shifted towards the LB3 samples for both the Bravo and manual ChIP-seq methods (**Fig. 4.1e**), suggesting that the chromatin sheared in LB3 lysis buffer yielded a better ChIP-seq result than the chromatin sheared in Empigen BB lysis buffer. However, for H3K27me3 ChIP-seq, there was not an obvious difference between the results produced by chromatin sheared in LB3 and Empigen BB. The signal peaks shown in the read density tracks (**Fig.4.2a**) and the scatterplots (**Fig.4.2d**) all suggested similarity in the results.

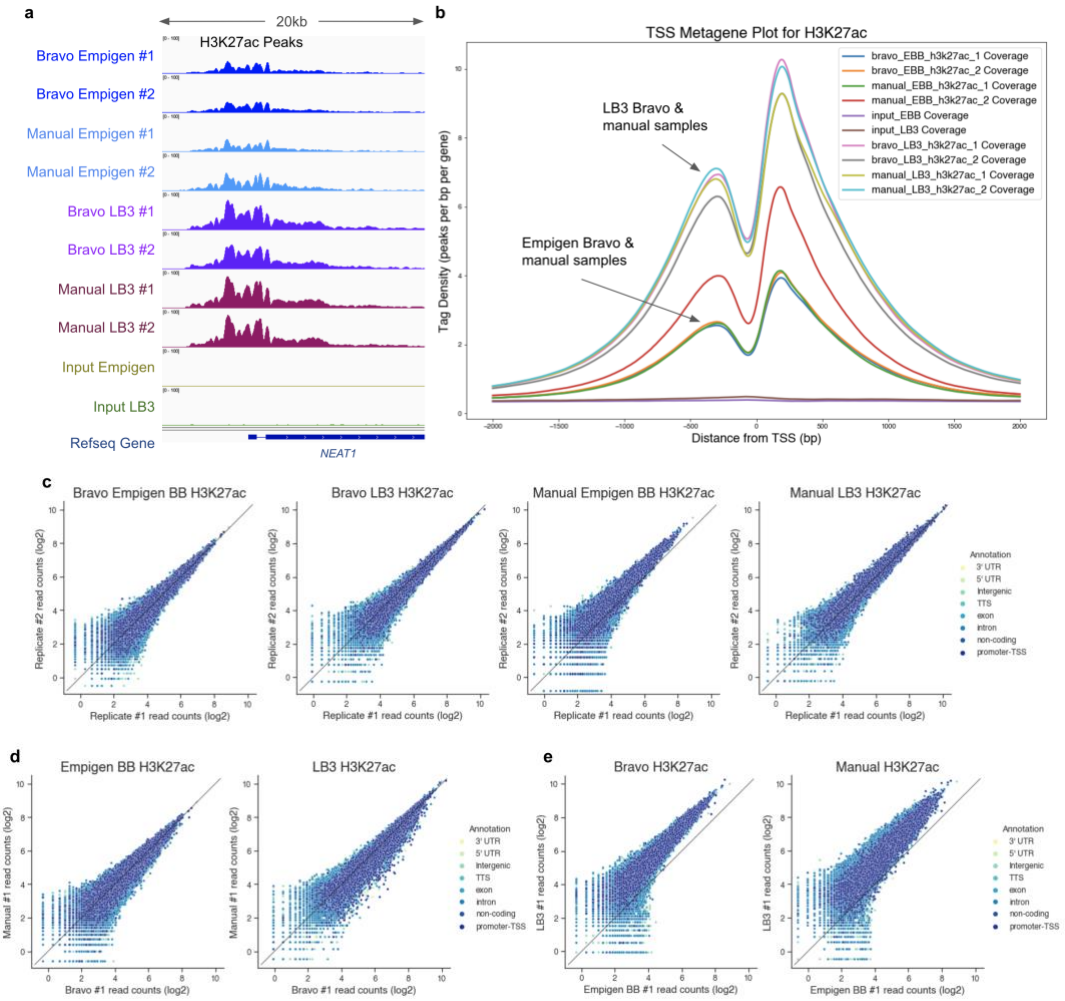

Figure 4.1: H3K27ac ChIP-seq experiments indicated that the automated (Bravo) workflow is very similar to manual workflow, while LB3 performed better as a lysis buffer in sonication than Empigen BB (EBB) lysis buffer. **(a)** IGV Genome Browser tracks of read coverage show H3K27ac ChIP-seq binding signals (indicated as peaks) over representative genomic loci, *NEAT1*. The snapshot is taken at a window of 20kb on Chr11:65,414,222-65,435,022 with signal height scale of 0-100 for all the tracks. Blue tracks show H3K27ac ChIP-seq with Empigen BB buffer (Bravo and manual) and purple tracks show H3K27ac ChIP-seq with LB3 buffer (Bravo and manual). Each condition was done in replicates alongside with a control input that was not immunoprecipitated for each lysis buffer to indicate background signals. **(b)** Metagene plot shows the average H3K27ac ChIP-seq tag density (peaks per base pair per gene) around the transcription start sites (TSS) in HeLa-S3 cells. The lines correspond to all the samples that are also shown in Fig. 1a. **(c)-(e)** Scatterplots show correlation in the samples for the read counts (log2) from 1000 base pairs around the merged peaks of all the H3K27ac ChIP-seq samples across the human genome hg38. The color of markers indicates the genomic regions where the peaks fall within. The diagonal y=x line represents slope =1. **c.** Correlation between the replicates for each method and lysis buffer condition pair. **d.** Correlation between Bravo and manual ChIP-seq samples for the two lysis buffers (LB3 versus Empigen BB). **e.** Correlation between LB3 and Empigen BB ChIP-seq samples for the two ChIP-seq methods (Bravo versus manual).

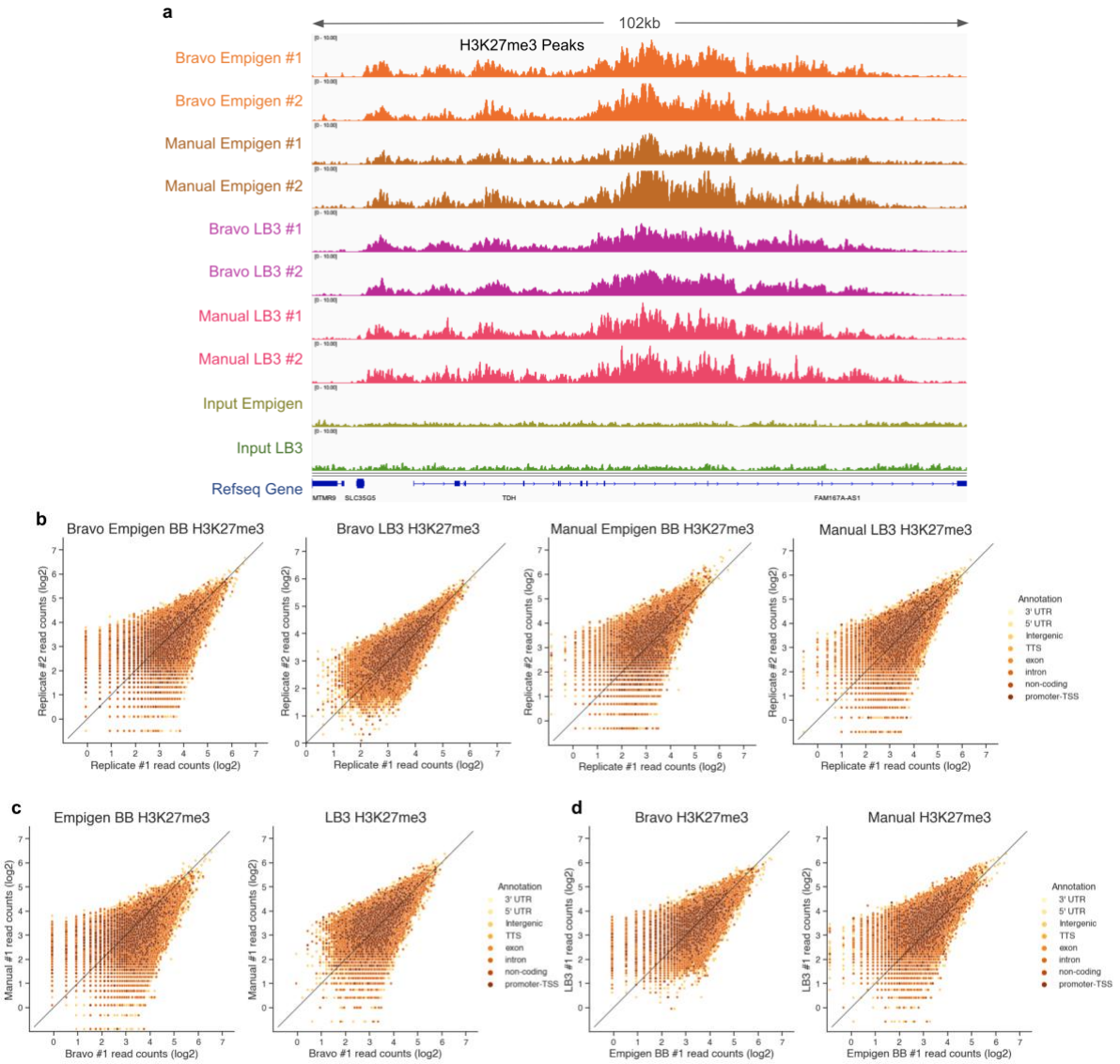

Figure 4.2: H3K27me3 ChIP-seq experiments indicated that the automated (Bravo) workflow is very similar to manual workflow, and LB3 lysis buffer also produced similar results in sonication as Empigen BB (EBB) lysis buffer. **(a)** IGV Genome Browser tracks of read coverage show H3K27me3 ChIP-seq binding signals (indicated as peaks) at a window of 102kb on Chr8:11,319,589-11,422,329with signal height scale of 0-10 for all the tracks. Orange/yellow tracks show H3K27me3 ChIP-seq with Empigen BB buffer (Bravo and manual) and purple/pink tracks show H3K27me3 ChIP-seq with LB3 buffer (Bravo and manual). Each condition was done in replicates alongside with a control input that was not immunoprecipitated for each lysis buffer to indicate background signals. **(b)-(d)** Scatterplots show correlation in the samples for the read counts (log2) from 1000 base pairs around the merged peaks of all the H3K27ac ChIP-seq samples across the human genome hg38. The color of markers indicates the genomic regions where the peaks fall within. The diagonal y=x line represents slope =1. **b.** Correlation between the replicates for each method and lysis buffer condition pair. **c.** Correlation between Bravo and manual ChIP-seq samples for the two lysis buffers (LB3 versus Empigen BB). **d.** Correlation between LB3 and Empigen BB ChIP-seq samples for the two ChIP-seq methods (Bravo versus manual).

Comparison of Automated and Manual ChIP-seq Workflow

The two ChIP-seq experiments comparing the crosslinking conditions and lysis buffers were both performed with Bravo and manual workflows, with the crosslinking comparison experiments done prior to the lysis buffer comparison. As shown by the H3K27ac, H3K27me3, and SETD5 ChIP-seq results in **Fig. 3**, the manual samples had better ChIP-seq signal compared to the automated Bravo in general. We suspected it was due to insufficient ChIP washing, so we made changes to the automated protocol accordingly. Instead of manually moving the plate back and forth on a bar magnet, we washed the ChIP samples by pipetting to thoroughly resuspend the beads in wash buffers.

After the modification of the washing steps in the automated protocol, the results of the following experiment in **Fig. 4.1** and **Fig. 4.2** indicated our automated ChIP-seq workflow produced very similar and consistent results as the manual workflow. As shown by the visualization of the read coverage tracks in the IGV genome browser for H3K27ac ChIP-seq (**Fig. 4.1a**) and H3K27me3 ChIP-seq experiments (**Fig. 4.2a**), the size and pattern of the signals, i.e., peaks, between the automated (Bravo) and manual ChIP-seq samples were similar. The peaks were also nearly indistinguishable between Bravo replicates, suggesting consistency in replicates. The similarity between Bravo and manual samples as well as between the two Bravo replicates were also evident in the metagene plot around the transcription start sites (TSS) (**Fig. 4.1b**). The average H3K27ac and H3K27me3 ChIP-seq tag density (peaks per base pair per gene) around TSS for the Bravo samples were again very similar to the manual samples. Moreover, we found the correlation in the samples by comparing the read counts between two experiments directly at the merged peaks of all the H3K27ac or H3K27me3 ChIP-seq samples. The correlation between the two technical replicates were highly similar, especially for the Bravo

samples (**Fig. 4.1c, Fig. 4.2b**), indicating the high reproducibility of the Bravo automated ChIP-seq workflow. The correlation between Bravo and manual ChIP-seq samples were also very similar for both lysis buffers LB3 and Empigen BB (**Fig. 4.1d, Fig. 4.2c**). This shows that the Bravo ChIP-seq method can work as efficiently as the manual ChIP-seq method.

Antibody Titration with Automated ChIP-seq

Another important yet often overlooked aspect of ChIP is the amount of antibody used for immunoprecipitation. Therefore, we conducted an antibody titration experiment that investigated the effects of varying the amounts of SETD5 antibody on ChIP-seq results. Since we have shown that the automated workflow can perform as well as the manual workflow, the experiment was carried out solely with the automated liquid handling platform Bravo. We compared 4 different amounts of the SETD5 antibody, 0.1 ug, 0.2 ug, 0.5 ug, and 1 ug. Combining results from both the IGV peak visualization tracks (**Fig. 5.1a**) and the metagene plot (**Fig. 5b**), we could identify the optimal ChIP-seq assay conditions for the SETD5 antibody. The genome browser tracks showed 3 conditions that had similar height and shape of peaks, which are (1) chromatin sheared in LB3 lysis buffer and immunoprecipitated with 0.5 ug of SETD5 antibody, (2) chromatin sheared in Empigen BB lysis buffer and immunoprecipitated with 0.5 ug of SETD5 antibody, and (3) chromatin sheared in LB3 lysis buffer and immunoprecipitated with 1 ug of SETD5 antibody (**Fig. 5.1a**). The metagene plot along the gene bodies further supported the results. With the highest overall average ChIP-seq tag density in the metagene plot, the optimal/best condition was obtained using chromatin sheared in LB3 lysis buffer and immunoprecipitated with 1 ug of SETD5 antibody, but 0.5 ug of antibody for chromatin sheared in LB3 lysis buffer was also enough to produce signal in ChIP-seq (**Fig. 5.1b**).

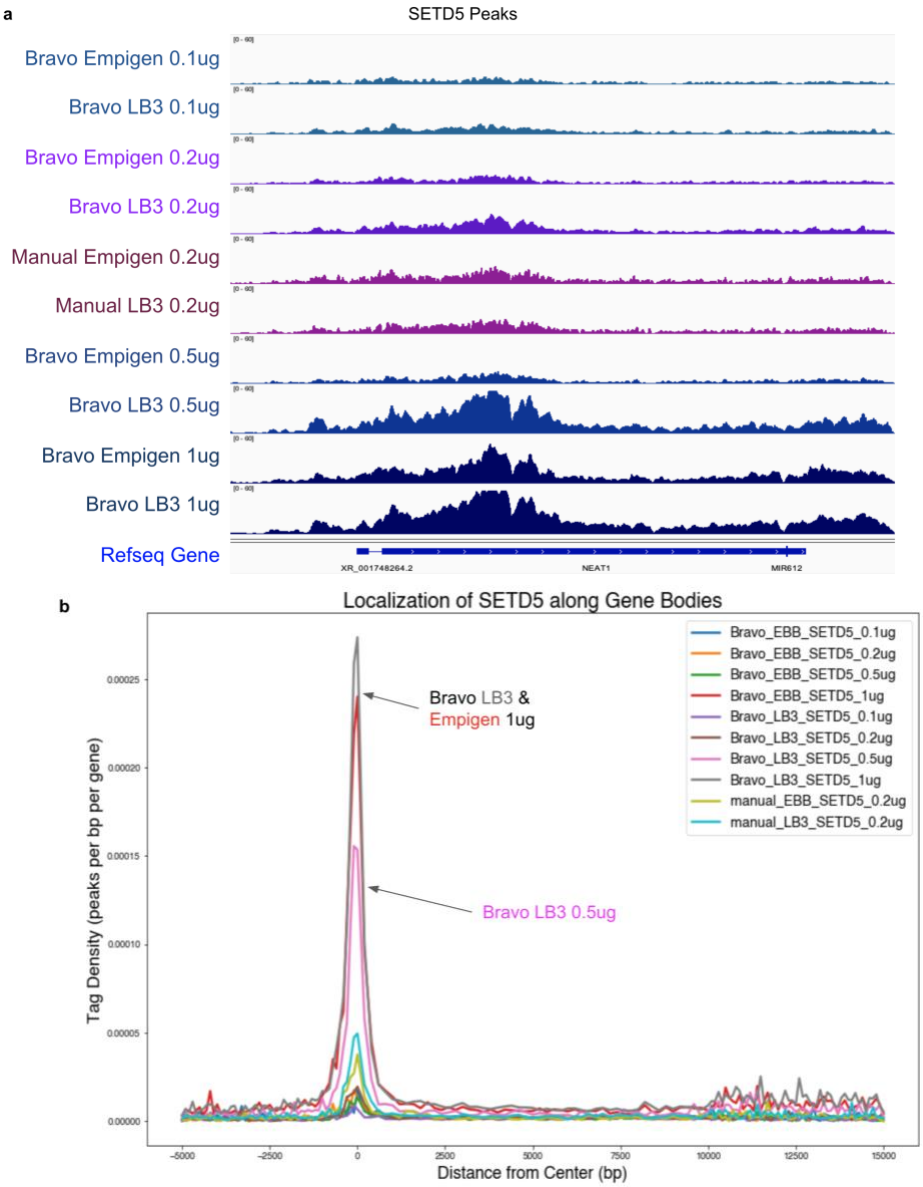

Figure 5.1: SETD5 ChIP-seq experiments indicated that the automated (Bravo) workflow can perform antibody titration along with LB3 vs Empigen BB lysis buffer comparison. **(a)** IGV Genome Browser tracks of read coverage show SETD5 ChIP-seq binding signals (indicated as peaks) over representative genomic loci, *NEAT1*. The snapshot is taken at a window of 36kb on Chr11:65,414,222-65,450,290 with signal height scale of 0-35 for H3K27ac tracks and signal height scale of 0-60 for SETD5 tracks. The tracks show antibody amounts of 0.1 ug, 0.2 ug, 0.5ug, and 1ug. Each condition was done in replicates. **(b)** Metagene plot shows the average SETD5 ChIP-seq tag density (peaks per base pair per gene) along the gene bodies in HeLa-S3 cells. The lines correspond to all the SETD5 samples that are also shown in (a). Arrows highlight three samples with outstanding signal.

In addition, we performed automated ChIP-seq with monoclonal HAT1 and EP300 antibodies and the DSG/FA double crosslinked HeLa-S3 cells sheared in both LB3 and Empigen BB lysis buffers to identify the optimal lysis buffer condition for the antibodies. As shown in the IGV tracks of read coverage, compared to the input samples, there are distinct peaks for the Bravo LB3 and Empigen BB HAT1 as well as the Bravo LB3 EP300 (**Fig. 5.2a**). The metagene plots also indicated that both buffers produced similar and consistent ChIP-seq signal around TSS for HAT1 (**Fig. 5.2b**), but for EP300, only the LB3 sample showed evident signal around TSS, whereas the Empigen BB sample showed no signal (**Fig. 5.2c**). The results suggested that automation (Bravo) successfully performed ChIP-seq with the HAT1 antibody when using chromatin sheared in both the LB3 and Empigen BB lysis buffers. On the other hand, EP300 ChIP-seq was only successful with the LB3 but not with the Empigen BB lysis buffers.

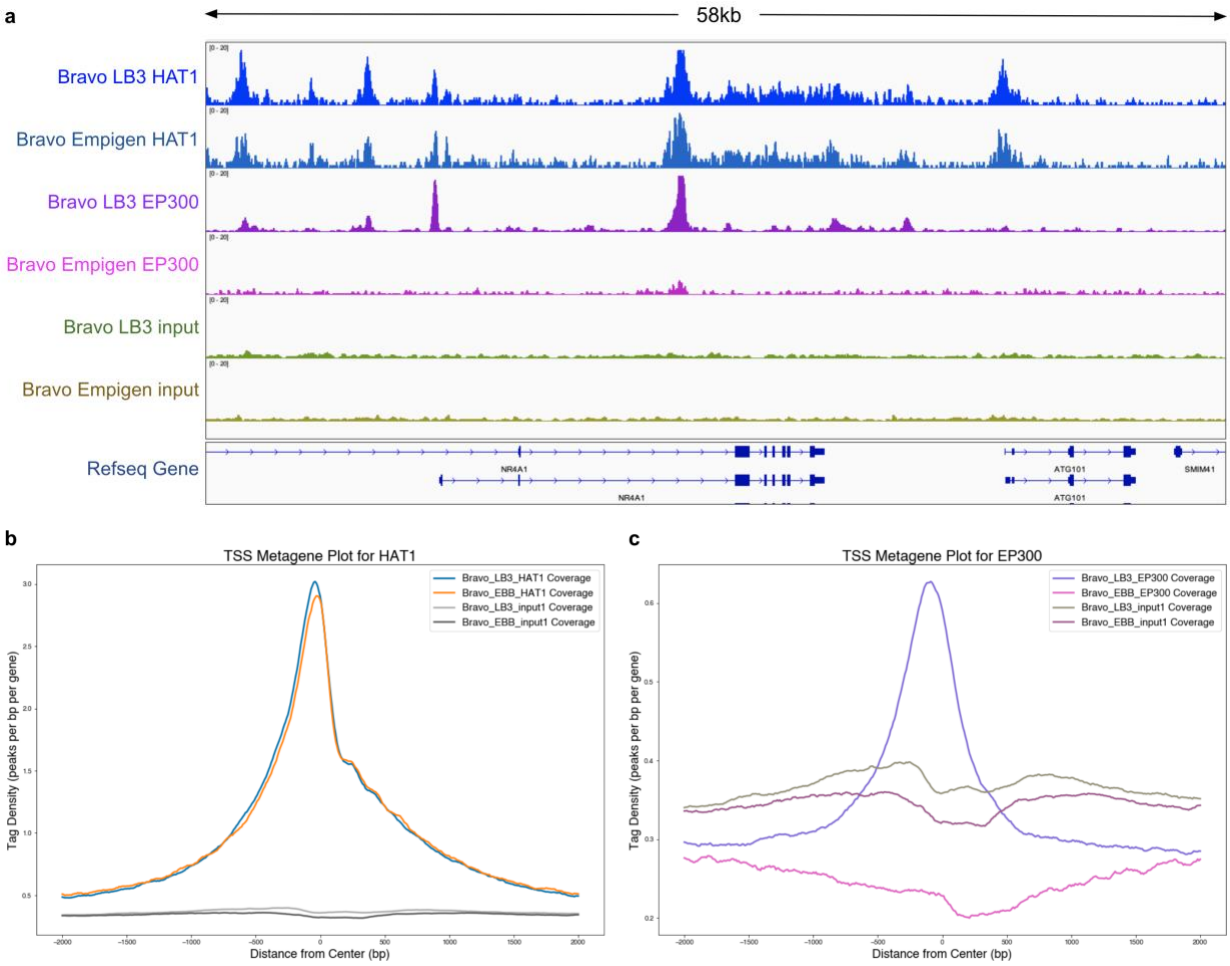

Figure 5.2: HAT1 and EP300 ChIP-seq experiments performed with the automated (Bravo) workflow comparing LB3 and Empigen BB lysis buffers. **(a)** IGV Genome Browser tracks of read coverage show HAT1 and EP300 ChIP-seq binding signals (indicated as peaks) over a window of 58kb on Chr12:52,023,760-52,082,608 with signal height scale of 0-20 for all tracks. A control input that was not immunoprecipitated for each lysis buffer was included to indicate background signals. **(b)-(c)** Metagene plots show the average ChIP-seq tag density (peaks per base pair per gene) over a region 2000 bp downstream and upstream of the TSS in HeLa-S3 cells for HAT1 (b) and EP300 (c). The lines correspond to all the SETD5 samples that are also shown in (a).

DISCUSSION

As shown in the results, we have systematically evaluated several aspects of the ChIP-seq process, including shearing conditions, crosslinking methods, and antibody titration, while comparing the automated workflow to the manual workflow. For chromatin shearing, we tried two sonicators, Covaris LE220-plus and PIXUL. After optimization of the parameters, the shearing on Covaris LE220-plus was consistent but ChIP-seq results were not as ideal. On the other hand, PIXUL is more promising because it not only produced great shearing consistency across the columns, but it was also easier to use as most of the parameters were already optimized by the manufacturer while only needing to change the processing time for different applications. It can also sonicate a wide range of cells, ~10,000 - 5,000,000, per well (Bomsztyk *et al*., 2019).

When comparing lysis buffers, LB3 was shown to work better with cells single crosslinked with FA. When the DSG/FA double crosslinked cells were sonicated for the same amount of time, Empigen BB lysis buffer yielded samples with better shearing patterns, but when ChIP-seq was performed on the same sheared chromatins, LB3 samples had better signal-to-noise ratio. This was evident in ChIP-seq experiments with the chromatin regulators, SETD5 and EP300, in which LB3 produced significantly higher signal-to-noise ratio than Empigen BB. A possible explanation would be that LB3 is a weaker reagent than Empigen BB and requires more sonication time to fully shear the chromatin to ideal size.

We also verified that for proteins, such as chromatin regulator SETD5, that are harder to be crosslinked to DNA, DSG/FA double crosslinking can successfully yield reproducible ChIP-seq results with robust signal-to-noise ratios. This is because FA only forms crosslinks of 2.3–2.7 Å, whereas DSG can form crosslinks of 7.7 Å, enabling DSG to capture longer interactions.

When DSG and FA are combined in the double crosslinking process, DSG first stabilizes the longer protein-protein interactions and then FA crosslinks shorter DNA-protein interactions (Nowak *et al*., 2005; Tian *et al*., 2014).

To improve the efficiency of ChIP-seq, we developed an automated ChIP-seq protocol with a newly optimized ChIP-seq protocol (Texari *et al*., 2021) and the robotic liquid handling platform Bravo. In the comparison of automated and manual ChIP-seq workflow, we have shown that the automated workflow was reproducible and had similar results as the manual workflow. Automation has many advantages. It reduces hands-on time and the probability of human errors while maintaining ChIP signal-to-noise ratio. It increases throughput compared to manual work as it allows for simultaneous processing of up to 96 samples in a standard PCR plate with high consistency between technical replicates. In addition, automated ChIP-seq enables easier evaluation of multiple conditions at once such as lysis buffer comparison and antibody titration. We have shown that our automated ChIP-seq could successfully test new HAT1 and EP300 antibodies and titrate SETD5 antibody to identify the optimal lysis buffer and amount of antibody required for ChIP. The automated ChIP-seq process is powerful in that it will enable faster and more robust study of regulatory elements and transcription factors as well as systematic mapping of protein-DNA binding *in vivo*.

However, there are also limitations of this automated ChIP-seq workflow. Since automation is performed by a robot, it needs to be operated based on a well written program, and the configuration and troubleshooting process can be tedious and time-consuming. In the first few times of running the automation protocol, we might need to pay close attention and watch it run the whole protocol to ensure the process is properly run, but I believe as we run it more frequently, it will be more and more smooth to run the automation and eventually require less

attentive time. Our goal is to make our automated ChIP-seq protocol publicly available with the hope to benefit and improve reproducibility across different labs. It is also possible to increase throughput to use a 384 well plate as Bravo has a head supporting 384 tips. Moreover, our automated protocol used Bravo, but the similar concept can be modified and adapted to any commercially available liquid handling platform.

METHODS

Cell culture preparation

HeLa-S3 (ATCC CCL-2.2™) adherent cells were cultured in supplemented DMEM with GlutaMAX™ (DMEM, 4.5 g/L D-Glucose, 110mg/L sodium pyruvate, 2 mM L-glutamine, 10% FBS, 1% Antibiotic-Antimycotic) in an incubator set to 37°C with 95% Air, 5% CO2. When the cells were grown to confluency, they were harvested by adding 0.25% trypsin.

Crosslink cells

Double crosslinking: The harvested cells were first crosslinked with at least 1 mL of 2mM DSG/PBS solution per 1 million cells at room temperature for 30 min while rotating the tube overhead at 8 rpm. Then the cell pellet was resuspended in 1% formaldehyde/PBS solution per 1 million cells and incubate at room temperature for 10 min while rotating the tube overhead at 8 rpm. The crosslinking reaction was quenched by adding 1/20th volume of 2.625 M glycine and 1/20th volume 10% BSA. The cell pellet was resuspended and stored in 0.5% BSA/PBS.

Single crosslinking: The harvested cells were crosslinked with 1% formaldehyde/PBS solution per 1 million cells and incubatde at room temperature for 10 min while rotating the tube overhead at 8 rpm. The crosslinking reaction was quenched by adding 1/20th volume of 2.625 M

glycine and 1/20th volume 10% BSA. The cell pellet was resuspended and stored in 0.5% BSA/PBS.

Chromatin sonication

Per 1 million crosslinked cell pellets, resuspend in 100µL lysis buffers: SDS lysis buffer [1% SDS, 50 mM Tris-HCl (pH 8.1), 10 mM EDTA, 1× protease inhibitor cocktail], LB3 [10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.1% Deoxycholate, 0.5% Sarkosyl, 1× protease inhibitor cocktail], Empigen BB lysis [0.5% Empigen BB (Sigma), 1% SDS, 50 mM Tris-HCl (pH 7.5), 10 mM EDTA, 1× protease inhibitor cocktail] (Métivier *et al.*, 2003).

For PIXUL, 100µL of lysate was loaded into each well (~10,000 - 5,000,000 cells per well) and the plate was sealed with a PCR plate pressure seal.

For Covaris LE220-plus, 20µL of lysate was loaded into each well and the plate was sealed with an aluminum plate seal.

After the chromatin was sonicated, lysate sheared with SDS was diluted 10-fold with CDB [0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.mM Tris-HCl, pH 8.1, 167mM NaCl, 1 × protease inhibitor cocktail] and lysate sheared with Empigen BB was diluted 2.5-fold in EBB dilution buffer [20 mM Tris/HCl pH 7.4@20°C, 100 mM NaCl, 0.5% Triton X-100, 2 mM EDTA, 1× protease inhibitor cocktail).

Input preparation

Manual workflow: 1% or 10µL of each sonicated lysate was saved as input sample. Per input sample, 55µL 10% Triton X-100, 68µL of Elution Buffer (10mM Tris pH8, 0.5% SDS, 5mM EDTA, 280mM NaCl), and 1µL RNase A were added and incubated in a thermocycler for 15 minutes at 37°C. 1µL of Proteinase K was added to each input sample and incubated at 55°C for 1 hour and then 65°C for 30min. To clean up the input sample, 122µL of Speedbead

mastermix [2µL SpeedBeads 2µL SpeedBeads + 120µL 20% PEG8000/1.5M NaCl (final concentration = 8.5% PEG, 1M NaCl)] was added to each input and incubated at room temperature for 10 minutes. After each sample was washed twice with 80% EtOH and Speedbeads pellet was air-dried, the input sample was eluted with 20µL of TT (0.05% Tween 20, 10mM Tris pH 8.0, warmed to 55°C).

Bravo workflow: Input plate was prepared with 1% or 10µL of each lysate sample, and per lysate sample, the buffer plate was prepared with 55µL 10% Triton X-100, 68µL of Elution Buffer (10mM Tris pH8, 0.5% SDS, 5mM EDTA, 280mM NaCl), and 1µL RNase A. The Buffer Plate was placed on Bravo deck 5 and the Sample Plate was placed on Bravo deck 8. Bravo protocol 1a. Input Preparation was run. Sample plate was incubated in a thermocycler for 15 minutes at 37°C. 1µL of Proteinase K was added to each input sample and incubated at 55°C for 1 hour and then 65°C for 30min. While the sample was incubating, a Speedbeads mastermix plate was prepared with 122µL of speedbead mastermix [2µL SpeedBeads 2µL SpeedBeads + 120µL 20% PEG8000/1.5M NaCl (final concentration = 8.5% PEG, 1M NaCl)] and 20µL of TT (0.05% Tween 20, 10mM Tris pH 8.0, warmed to 55°C) per sample added to each well in the mastermix plate. Ethanol Plate was prepared with freshly made 80% EtOH. After incubation, the sample plate was back on Bravo deck 8, Mastermix Plate on deck 5 and Ethanol Plate on deck 2. Bravo protocol 1b. Input Speedbead Cleanup was run to clean up the input samples. At the end of the protocol, each well should contain 20µL of eluted input in TT buffer.

Chromatin immunoprecipitation

Per each ChIP sample, 10 µL of Dynabeads™ Protein A was washed twice with equal volume of LB3 lysis buffer + 1/9th volume of 10% Triton X-100 for LB3 samples, or EBB dilution buffer for Empigen BB samples. Antibodies used in these experiments are H3K27ac

(Acetyl-Histone H3 (Lys27) (D5E4) XP® Rabbit mAb from Cell Signaling Technology), H3K27me3 (Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb from Cell Signaling Technology), and SETD5 (0.1 mg/mL SETD5 Polyclonal Antibody Catalog # PA5-57318 from Invitrogen).

Manual workflow: 1µL of H3K27ac, 1µL of H3K27me3, 1, 2, 5, or 10 µL of SETD5 antibody was added to each tube containing washed Dynabeads Protein A accordingly. 50-100 µL of sheared chromatin (~500k cells) was aliquoted into each sample tube.

Bravo workflow: Dynabeads Protein A resuspended in lysis buffer and appropriate volume of µL of antibody (same amount as manual workflow) was transferred to Bead/Antibody Plate. PIXUL lysate plate was placed on deck 2 and Dynabead/Antibody Plate on deck 5. Bravo protocol 2. Chromatin+Beads+Antibody Bind was run to aliquot 50-100 µL of sheared chromatin into Beads/Antibody Plate.

The ChIP samples were incubated overnight on HulaMixer™ Sample Mixer at 4°C in the cold room (rotating at 8 rpm).

ChIP Washes

Manual workflow: After overnight incubation, the sample was washed three times with WBI [10mM Tris pH7.5, 0.7% DOC, 1mM EDTA, 250mM LiCl, 1× protease inhibitor cocktail], three times with WBIII [20mM Tris pH7.5, 1% Triton X-100, 2mM EDTA, 150mM NaCl, 0.1% SDS, 1× protease inhibitor cocktail], and twice with TET [10mM Tris pH7.5, 0.2% Tween-20, 1mM EDTA, 1× protease inhibitor cocktail]. Each ChIP sample was eluted in 25µL of TT

Bravo workflow: Wash + Elution Buffer plate was prepared by aliquoting 600µL of WBI/WBIII buffer, 400µL of TET, and 30µL of TT per well. Wash + Elution Buffer Plate was placedon deck 5, ChIP Plate on deck 7 (on ring magnet). Bravo protocol 3. ChIP Wash + Elution

Off Mag was run. At the end of the protocol, each well should contain 25µL of eluted ChIP sample in TT buffer.

Library construction

We used the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (Catalog #E7645L) and NEXTflex® DNA Barcodes - 48 (NOVA-514104) for library construction.

Manual workflow: 2µL of each input supernatant was transferred to a new tube and 23µL of TT was added to make a total volume of 25µL. (1) End Prep: Per each ChIP and input sample, 5µL of End Prep Mastermix (1.5µL of NEBNext Ultra II End Prep Enzyme Mix and 3.5µL of NEBNext Ultra II End Prep Reaction Buffer) was added and incubated for 30min at 20°C, and then 30min at 65°C in a thermal cycler with the lid set to 75°C. (2) Adapter ligation: The ChIP and input samples were recovered from the thermal cycler and 1µL of NEXTflex-DNA-Barcodes-48 Bioo ChIP Adaptor (1:40 diluted 0.625uM) was added to each sample. Per each ChIP and input sample, 15.5µL of Ligation Mastermix [15µL NEBNext Ultra II Ligation Master Mix and 0.5µL of NEBNext Ligation Enhancer] was added and incubated for 15mins at 20°C with the heated lid set to "off." After incubation, ligation reaction was stopped by adding 27µL of ligation stop solution mastermix [4µL 10% SDS + 3µL 0.5M EDTA + 20µL water] and 4.5µL of 5M NaCl per sample.

Bravo workflow: 2µL of each input supernatant was transferred to the ChIP plate and 23µL of TT was added to make a total volume of 25µL. (1) End Prep: Per sample, 5µL of End Prep Mastermix was prepared with 1.5µL of NEBNext Ultra II End Prep Enzyme Mix and 3.5µL of NEBNext Ultra II End Prep Reaction Buffer. Mastermix was added to the Mastermix Plate and the plate was placed on deck 5. The ChIP Plate was placed on deck 8. Bravo protocol 4. End Repair MM Dispense was run to add 5 µL of End Prep Mix to each ChIP and input sample. The

ChIP plate was incubated for 30min at 20°C, and then 30min at 65°C in a thermal cycler with the lid set to 75°C. (2) Adapter ligation: Per sample, 15.5µL of Ligation Mastermix was prepared with 15µL NEBNext Ultra II Ligation Master Mix and 0.5µL of NEBNext Ligation Enhancer. Mastermix was added to the Mastermix Plate and the plate was placed on deck 5. The ChIP plate was recovered from the thermal cycler and 1µL of NEXTflex-DNA-Barcodes-48 Bioo ChIP Adaptor (1:40 diluted 0.625uM) was added to each sample. The ChIP plate was placed back to deck 8. Bravo protocol 5. Adapter Ligation MM Dispense was run to add 15.5µL of Ligation Mastermix to each sample. The ChIP plate was incubated for 15mins at 20°C with the heated lid set to ‘‘off.’’ Per sample, 27µL of ligation stop solution mastermix [4µL 10% SDS + 3µL 0.5M EDTA + 20µL water] and 4.5µL of 5M NaCl was prepared as a Buffer Plate and the plate was placed on deck 5. After incubation, the ChIP Plate was placed back to deck 8 and Bravo protocol 6. Ligation Stop Solution Dispense was run to dispense the stop solution.

Reverse crosslinking and Library cleanup

The samples were reverse crosslinked by adding 1µL of Proteinase K to each sample and incubating for 1 hour (or 30mins for cells crosslinked with FA only) at 55°C and then 30mins at 65°C.

Manual workflow: The supernatant of each sample was separated from Dynabeads Protein A and added to tubes containing 63µL of Speedbeads mastermix [2µL SpeedBeads + 61µL 20% PEG8000/1.5M NaCL (8.6% PEG, 0.8M NaCl)]. After 10 min of incubation at room temperature, each sample was washed twice with 200µL of 80% EtOH. The Speedbeads pellet was air-dried and the sample was eluted with 25µL of TT (0.05% Tween 20, 10mM Tris pH 8.0, warmed to 55°C).

Bravo workflow: Per sample, 63µL of Speedbeads mastermix [2µL SpeedBeads + 61µL 20% PEG8000/1.5M NaCL (8.6% PEG, 0.8M NaCl)] and 30µL of TT (0.05% Tween 20, 10mM Tris pH 8.0, warmed to 55°C) was added to each well in the Buffer Plate. The Buffer Plate was placed on Bravo deck 5. The Ethanol Plate was prepared with 400µL of 80% EtOH per sample and placed on deck 2. The ChIP Plate on Bravo deck 7 (on ring magnet). Bravo protocol 7. Adapter Ligation Cleanup was run. At the end of the protocol, each well should contain 25µL of eluted sample in TT buffer.

PCR Amplification and final cleanup of ChIP and Input Libraries

The PCR mastermix was created by combining 25 µL NEBNext Ultra II Q5 2x Mastermix with 0.25 µL each of 100 µM sequencing platform-compatible forward and reverse primers (Solexa 1GA& Solexa 1GB). The 25.5µL of PCR mastermix was added to each sample either manually or by running Bravo protocol 8. PCR Enrichment MM Dispense with the Mastermix Plate on deck 5 and ChIP Plate on deck 7.

A PCR program was run to amplify the libraries using a thermal cycler with a heated lid set to 105°C: Initial denaturation at 98°C for 30 seconds; 12 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 15 seconds, extension at 72°C for 30 seconds; and final extension at 72°C for 2 minutes.

Manual workflow: 40.5µL of Speedbead mastermix [2µL SpeedBeads + 38.5µL of 20% PEG/2.5M NaCl (final concentration = 8.5% PEG, 1M NaCl)] was added to each sample and incubated at room temperature for 10 minutes. After each sample was washed twice with 80% EtOH and Speedbeads pellet was air-dried, the input sample was eluted with 15µL of TT (0.05% Tween 20, 10mM Tris pH 8.0, warmed to 55°C).

Bravo workflow: Speedbeads mastermix plate was prepared with 40.5µL of Speedbead mastermix [2µL SpeedBeads + 38.5µL of 20% PEG/2.5M NaCl (final concentration = 8.5% PEG, 1M NaCl)] and30µL of TT (0.05% Tween 20, 10mM Tris pH 8.0, warmed to 55°C) per sample added to each well in the mastermix plate. Ethanol Plate was prepared with freshly made 200µL of 80% EtOH. The sample plate was placed on Bravo deck 7 (on ring magnet), Buffer Plate on deck 5 and Ethanol Plate on deck 2. Bravo protocol 9. PCR Enrichment Cleanup was run to clean up the input samples. At the end of the protocol, each well should contain 15µL of eluted sample in TT buffer.

DNA Quantification and Sequencing

DNA size quantification: 5 µL of each library was taken out and mixed with 1 µL of 6X TriTrack DNA Loading Dye to run on a 2% agarose gel pre-stained with GelGreen.

DNA concentration quantification: Qubit HS DNA assay buffer and standards were used to estimate the concentration of the libraries. The libraries were pooled according to the concentration calculations. The pooled libraries were prepared and sequenced with Illumina NextSeq as single reads (64 or 84 bp).

Sequencing data analysis

Sequencing reads were mapped to human reference genome hg38 using Bowtie2 (Langmead & Salzberg, 2012) with default parameters and duplicate reads were removed with SAMtools (Li, *et al.*, 2009). HOMER (Heinz, *et al.*, 2010) was used to make tag directories and bigwig files for visualization on IGV genome browser (Robinson *et al.*, 2011). Peaks were called with HOMER findPeaks program with parameter "-style histone." Metagene plots centered on TSS was made with HOMER annotatePeaks.pl program set in "TSS" mode. Metagene plots along the gene bodies were generated with HOMER makeMetaGeneProfile.pl program set in

"rna" mode. Correlation scatterplots were generated using HOMER annotatePeaks.pl program centered on merged peak file from all samples for each antibody (made with mergePeaks program) to quantify the read counts from each experiment at each of the peaks. Both HOMER programs automatically normalize each tag directory by the total number of mapped reads.

Material from parts of the Introduction, Results, and Discussion of this master's thesis is currently being prepared for submission for publication. The thesis author is the primary researcher and author of this material.

REFERENCES

Aldridge, S., Watt, S., Quail, M. A., Rayner, T., Lukk, M., Bimson, M. F., Gaffney, D., & Odom, D. T. (2013). AHT-ChIP-seq: a completely automated robotic protocol for high-throughput chromatin immunoprecipitation. *Genome biology*, *14*(11), R124. https://doi.org/10.1186/gb-2013-14-11-r124

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–837. https://doi.org/10.1016/j.cell.2007.05.009

Benson, L. J., Phillips, J. A., Gu, Y., Parthun, M. R., Hoffman, C. S., & Annunziato, A. T. (2007). Properties of the type B histone acetyltransferase Hat1: H4 tail interaction, site preference, and involvement in DNA repair. *The Journal of biological chemistry*, *282*(2), 836–842. https://doi.org/10.1074/jbc.M607464200

Bomsztyk, K., Mar, D., Wang, Y., Denisenko, O., Ware, C., Frazar, C. D., Blattler, A., Maxwell, A. D., MacConaghy, B. E., & Matula, T. J. (2019). PIXUL-ChIP: integrated high-throughput sample preparation and analytical platform for epigenetic studies. *Nucleic acids research*, *47*(12), e69. https://doi.org/10.1093/nar/gkz222

Busby, M., Xue, C., Li, C., Farjoun, Y., Gienger, E., Yofe, I., Gladden, A., Epstein, C. B., Cornett, E. M., Rothbart, S. B., Nusbaum, C., & Goren, A. (2016). Systematic comparison of monoclonal versus polyclonal antibodies for mapping histone modifications by ChIP-seq. *Epigenetics & chromatin*, *9*, 49. https://doi.org/10.1186/s13072-016-0100-6

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., Blecher-Gonen, R., Bornstein, C., Amann-Zalcenstein, D., Weiner, A., Friedrich, D., Meldrim, J., Ram, O., Cheng, C., Gnirke, A., Fisher, S., … Amit, I. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Molecular cell*, *47*(5), 810–822. https://doi.org/10.1016/j.molcel.2012.07.030

Gasper, W. C., Marinov, G. K., Pauli-Behn, F., Scott, M. T., Newberry, K., DeSalvo, G., Ou, S., Myers, R. M., Vielmetter, J., & Wold, B. J. (2014). Fully automated high-throughput chromatin immunoprecipitation for ChIP-seq: identifying ChIP-quality p300 monoclonal antibodies. *Scientific reports*, *4*, 5152. https://doi.org/10.1038/srep05152

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Janssens, D.H., Meers, M.P., Wu, S.J., Babaeva, E., Meshinchi, S., Sarthy, J.F., Ahmad, K., Henikoff, S. Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nat Genet* 53, 1586–1596 (2021). https://doi.org/10.1038/s41588-021-00941-9

Janssens DH, Wu SJ, Sarthy JF, Meers MP, Myers CH, Olson JM, Ahmad K, Henikoff S. Automated in situ chromatin profiling efficiently resolves cell types and gene regulatory programs. Epigenetics Chromatin. 2018 Dec 21;11(1):74. doi: 10.1186/s13072-018-0243-8. PMID: 30577869; PMCID: PMC6302505.

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, *316*(5830), 1497–1502. https://doi.org/10.1126/science.1141319

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., Iyer, V. R., … Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research*, *22*(9), 1813–1831. https://doi.org/10.1101/gr.136184.111

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Métivier, R., Penot, G., Hübner, M. R., Reid, G., Brand, H., Kos, M., & Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, *115*(6), 751–763. https://doi.org/10.1016/s0092-8674(03)00934-6

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553–560. https://doi.org/10.1038/nature06008

Nisar, S., Hashem, S., Bhat, A. A., Syed, N., Yadav, S., Azeem, M. W., Uddin, S., Bagga, P., Reddy, R., & Haris, M. (2019). Association of genes with phenotype in autism spectrum disorder. *Aging*, *11*(22), 10742–10770. https://doi.org/10.18632/aging.102473

O'Geen, H., Echipare, L., Farnham, P.J. (2011). Using ChIP-Seq Technology to Generate High-Resolution Profiles of Histone Modifications. In: Tollefsbol, T. (eds) Epigenetics Protocols. Methods in Molecular Biology, vol 791. Humana Press. https://doi.org/10.1007/978-1-61779-316-5_20

Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H., & Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, *87*(5), 953–959. https://doi.org/10.1016/s0092-8674(00)82001-2

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

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., Durham, T., Zhang, X., Donaghey, J., Epstein, C. B., Regev, A., & Bernstein, B. E. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*, *147*(7), 1628–1639. https://doi.org/10.1016/j.cell.2011.09.057

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., & Young, R. A. (2000). Genome-wide location and function of DNA binding proteins. *Science (New York, N.Y.)*, *290*(5500), 2306–2309. https://doi.org/10.1126/science.290.5500.2306

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O. L., He, A., Marra, M., Snyder, M., & Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods*, *4*(8), 651–657. https://doi.org/10.1038/nmeth1068

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Sessa, A., Fagnocchi, L., Mastrototaro, G., Massimino, L., Zaghi, M., Indrigo, M., Cattaneo, S., Martini, D., Gabellini, C., Pucci, C., Fasciani, A., Belli, R., Taverna, S., Andreazzoli, M., Zippo, A., & Broccoli, V. (2019). SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron*, *104*(2), 271–289.e13. https://doi.org/10.1016/j.neuron.2019.07.013

Texari, L., Spann, N. J., Troutman, T. D., Sakai, M., Seidman, J. S., & Heinz, S. (2021). An optimized protocol for rapid, sensitive and robust on-bead ChIP-seq from primary cells. *STAR protocols*, *2*(1), 100358. https://doi.org/10.1016/j.xpro.2021.100358

Zhou, V. W., Goren, A., & Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, *12*(1), 7–18. https://doi.org/10.1038/nrg2905