Development of High Coverage Mammalian Single-Cell Methylation Sequencing Techniques

Development of High Coverage Mammalian Single-cell Methylation Sequencing Techniques

A dissertation presented

by

Yunlong Cao

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

June 2019

Dissertation Advisor: Professor Xiaoliang Sunney Xie Yunlong Cao

Development of High Coverage Mammalian Single-cell Methylation Sequencing Techniques

Abstract

Single-cell methylation studies are essential where cell-to-cell methylation variation play a key role. It is also critical when the cell samples subject to analysis are precious, rare or in minute amounts. Current techniques available to perform mammalian single-cell methylation analysis rely on bisulfite conversion. The DNA damage and loss caused by bisulfite conversion hinders the methylome coverage, and brings great difficulties in data analyzes, rendering high accuracy differential methylation comparison extremely difficult to perform between individual single cells.

In this thesis, the development of a new high coverage methylation sequencing technique is presented, sc-TAPS, for mammalian CpG analysis. sc-TAPS significantly reduces DNA loss and DNA damage, while maintaining a high methyl-conversion rate. Unlike bisulfite conversion, only methylated cytosines will be converted to uracil during sc-TAPS, therefore the DNA complexity remain intact, resulting in a more uniform amplification, higher mapping rate, and less PCR bias.

A highly efficient DNA methylation amplification has been developed to compensate for DNA loss during sc-TAPS. Methylome Replication Loops with Methyl-Transferase (MERLOT) utilizes methyl-transferase DNMT1 for maintenance of the methylation pattern. MERLOT amplifies the DNA template 1.7-fold per cycle. Over 95% methylation maintenance efficiency could be achieved by MERLOT while having a low *de novo* methylated rate of 1.5%. MERLOT combined with sc-TAPS (MERLOT-TAPS), gives by far the highest coverage of all mammalian single-cell whole methylome sequencing techniques known thus far.

The low DNA loss and damage property of sc-TAPS makes it possible to combine with many other techniques previously unable by bisulfite conversion. scATAC-TAPS can simultaneously profile nucleosome occupancy and DNA methylation in single cells. Simultaneous measurement of three-dimensional genome structure and DNA methylation could be achieved by combining dip-C with TAPS for mammalian diploid cells. Finally, scTAPS followed by multiplex PCR enables deep methylation sequencing for over 50 specific gene targets simultaneously.

Together, sc-TAPS is proved to become a new way for single-cell mammalian methylation analysis with high coverage and high accuracy. Many techniques can be applied along with sc-TAPS to build a toolbox, that could overcome a variety of difficulties previously unsolvable by bisulfite conversion.

Acknowledgement

I would like to express my deepest gratitude to my advisor, Professor Xiaoliang Sunney Xie, for his insightful guidance, encouragement, trust, and providing me with the freedom to do the research of my interest. I truly admire his incredible drive and efforts to meet new science and make ground-breaking research. I feel very fortunate and honored to have such a passionate scientist as my advisor in graduate school.

I am deeply grateful to my graduate committee members, Professors Xiaowei Zhuang and Adam Cohen for giving me not only critical and constructive comments on my research projects but also precious advice on the method and attitude to conduct research in general.

Before joining the Xie Group, the bulk of my research experience was in optical physics. Although I changed to single-cell sequencing as my main focus after 2nd year in grad school, the physics mindset helped me transfer extremely smooth. I am very thankful to many people who have patiently taught me how to think like a physicist in college, including Prof. Xiaowei Zhuang, Hazen Babcock, Jeff Moffitt, and Prof. Xin Wan.

My contributions to the single-cell TAPS described in Chapter 2 were made with close collaboration with Bo Zhao. All the bioinformatic analysis was done with the help of Jackie Yang and Yali Bai. Methylome Replication Loops with Methyl-Transferase (MERLOT) described in Chapter 3 was developed with the help of Tianjiao Yuan, Penny Peng, and Leo Song.

There are several people who provided me with the support that was not directly related to particular experiments. I could always count on Alec, Tan, Dong, Lei, Wenlong, and Wenting to help me with my experiments and proposals. I shall never forget the encouragement and support received from Patty. I'm very grateful for all the joy she brought to our lab.

I am blessed with my family. My parents, Xuetao Cao and Jianli Wang, have supported me in every possible way unconditionally. My beloved one, Bella Zhang, has encouraged me to focus on my research and continue my career as a scientist unwaveringly. I cannot imagine doing any of this without the support from them. I humbly dedicate this work to them

Table of Contents

Chapter 1
Introduction

1.1 Mammalian DNA Methylation

DNA methylation is considered to be one of the most important epigenetic mechanism that modifies the function of genes and regulates gene expression by adding a methyl group to DNA nucleotides. In mammalian systems, the most abundant DNA methylation is the addition of a methyl group to the 5th carbon of the cytosine through covalent bonding, which results in 5-methylcytosine. The majority of the methylation events happen in CpG sites where a cytosine nucleotide is followed by a guanine nucleotide in the 5' → 3' direction.

Methylation of a cytosine within a gene can change its expression, frequently resulting in transcriptional silencing or suppression. In mammalian somatic cells, typically ~60% of CpG cytosines are methylated, and a total of 28 million CpG sites exist in the human genome. Mammalian DNA methylation of cytosines within the CpG dinucleotide context is associated with a variety of critical processes, including embryogenesis, genomic imprinting, X-chromosome inactivation, aging, and carcinogenesis.

The addition of methyl groups is controlled by DNA methyltransferases (DNMTs). This enzyme family, which is mainly represented by DNMT1, DNMT3a, and DNMT3b, is required for the maintenance and establishment of DNA methylation patterns. During DNA replication, DNMT1 is responsible for the maintenance of DNA methylation by specifically targeting the hemi-methylated CpG dinucleotides and adding a methyl group to the newly synthesized CpG

site. On the other hand, DNMT3a and 3b are responsible for de novo DNA methylations. The mechanism behind how DNMT3a and 3b choose their targets are mostly unknown.

5-methylcytosine is a very stable modification, and the methylation status is cell-type specific. Different somatic cell types exhibit distinct methylation patterns which would not change during a lifespan. People have long been proposing that CpG methylation defines the differentiation landscape, which refines somatic cells to keep their identities, without changing to other cell types. Recent discoveries have also shown that CpG methylation is highly correlated with genome 3D structure and nucleosome occupation, suggesting methylation may serve as a landscape to guide 3D genome folding and nucleosome occupancy, making these highly variable and dynamic process rather consistent in specific cell types.

1.2 Mammalian DNA Methylation Sequencing

CpG methylation has been shown to be of relevance in many critical biological processes, therefore researchers have developed many ways to probe the methylation status of mammalian systems. Bisulfite sequencing is considered to be the golden standard of mammalian methylation analysis. Treatment of DNA with sodium bisulfite converts cytosine residues to uracil but leaves 5-methylcytosine residues unaffected. This method provides the ability to differentiate unmethylated versus methylated cytosines and offers a single nucleotide resolution map of DNA methylation status.

The major challenge in bisulfite conversion is the degradation and fragmentation of DNA that takes place during this process. The conditions necessary for complete conversions, such as

long incubation times, elevated temperature, and high bisulfite concentration, can lead to the degradation and fragmentation of up to 90% of the DNA in the reaction solution. The degradation occurs as DNA depurination, resulting in random strand breaks. The extensive degradation is problematic and even more so when dealing with a limited amount of starting DNA. Furthermore, the majority of the cytosines that exist in the genome are converted into uracil, which is equivalent to thiamine in a PCR reaction, resulting in DNA complexity reduction.

Recent advances in Enzyme-based or enzyme-assisted conversion of 5-methylcytosine into uracil residues have facilitated new single-base resolution sequencing methods to appear for DNA methylation analysis in the mammalian system. These new sequencing techniques all aimed at solving the DNA low yield problem of the bisulfite sequencing, however each has its own limitations. TET Assisted Pyridine-Borane Sequencing (TAPS) utilizes the TET enzyme, which could oxidize 5-methylcytosine into 5-carboxylcytosine (5caC) with high efficiency. Combined with Pyridine borane, which reduces 5caC into DHU, TAPS could effectively achieve the mapping of 5mC without affecting normal cytosines. However, due to the DNA damage caused by the iron chemistry during TET oxidization reaction and several DNA purification steps involved, DNA poor yield is still prevalent in TAPS and requires 10ng of DNA for starting material.

1.3 Mammalian Single-cell Methylation Sequencing

The capability to perform single-cell methylation studies is essential in studies where cell-to-cell variation and heterogeneity play a key role. It is also critical when the cell samples

subject to analysis are precious or rare or in minute amounts, such as in embryo development or cancer biopsies.

Current techniques available to perform single-cell analysis of DNA methylation rely on bisulfite conversion, such as scRRBS, scWGBS, scNOME-seq, etc. Single cell bisulfite sequencing can be categorized into two categories, reduced representation bisulfite sequencing and post-bisulfite adaptor tagging whole genome bisulfite sequencing. The reduced representation method utilizes restriction enzymes to create PCR priming sites while is also enriched for CpG island region. This method greatly reduces the cost compared to the bisulfite but at the same time confines the covered genomic region to merely 4% of the entire methylome space. The Post-bisulfite adaptor tagging method, on the other hand, uses a high concentration of Klenow fragments (exo-) to multiple rounds of random priming after bisulfite conversion for PCR adaptor addition. This allows maximum recovery of bisulfite converted DNA so it could achieve around 15%~20% methylome coverage. The disadvantage of this method is that due to the multiple rounds of high concentration of Klenow fragment random priming, the amount of primer concatemer accumulates. The mapping rate of the final sequencing library is low due to primer contamination and also to the formation of chimeras at a high rate. The average mapping rate is only 23%, affecting the sequencing cost to an increase of around four-fold compared to standard DNA sequencing. Efforts have been made to reduce the amount of chimera and primer concatemer formed during random priming by cutting down the number of random priming cycles, but this procedure lowers the methylome coverage to 5%.

The DNA damage and loss brought by bisulfite conversion also hinders the methylome coverage, and the highest achievable is on average 15% by scWGBS. The low coverage of single cell methylation sequencing brings great difficulties in population analyzes such as PCA

or t-SNE, and renders high accuracy differential methylation comparison impossible between individual single cells.

In this thesis, we present the development of a new high coverage methylation sequencing technique, sc-TAPS, for mammalian CpG analysis. Combined with other techniques, sc-TAPS can tackle a variety of difficulties and could be applied to several critical biology problems, such as embryogenesis and carcinogenesis.

1.4 Reference

1. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res*. 2013;23(12):2126–2135.

2. Farlik M, Sheffield NC, Nuzzo A, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep*. ;10(8):1386–1397.

3. Smallwood SA, Lee HJ, Angermueller C, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods*. 2014;11(8):817–820.

4. Mulqueen RM, Pokholok D, Norberg SJ, et al. Highly scalable generation of DNA methylation profiles in single cells. *Nat Biotechnol*. 2018;36(5):428–431.

5. Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol*. 2015;16(1):180.

Chapter 2

Single cell TET Assisted Pyridine-Borane Sequencing (sc-TAPS)

Compared to the bisulfite sequencing, TET assisted pyridine borane sequencing (TAPS) shows less DNA damage, higher methylome coverage, and better mapping efficiency, making it a much-improved method for high coverage single-cell level methylation sequencing. However, due to DNA damage caused by Fe(II) during the 5mC oxidation reaction and DNA lost during multiple purification steps, directly performing TAPS at single-cell level results in poor DNA methylome coverage. In this chapter, we present that by utilizing Fe(II) ligand ATP and lambda carrier DNA, DNA damage and loss could be largely reduced. By optimizing the αKG concentration, the amount of PEG added and using NgTET1 as the oxidation enzyme, we demonstrated that TAPS could achieve over 90% methylation conversion rate.

Applying sc-TAPS to GM12878 single cells results in an average of 28% methylome coverage with 3x sequencing depth, nearly three-fold higher compared to previous techniques at the same sequencing depth. The mapping rate is 91.2%, which is three-fold higher than scWGBS, and computation time only takes one-fourth of the time and effort compared to bisulfite sequencing. After analysis of fully methylated and unmethylated lambda DNA spike-in, on average a 90.6% methyl-conversion rate and a 0.48% de novo methylation rate are achieved.

2.1 Background

2.1.1 Mechanism of TAPS

TET assisted pyridine borane sequencing (TAPS) combines the ten-eleven translocation (TET) enzymatic oxidation of 5mC to 5-carboxylcytosine (5caC), with pyridine-borane chemical reduction of 5caC to dihydrouracil (DHU). This procedure successfully converts only methylated and hydroxymethylated cytosines into DHU, which in sequential PCR reaction results in Thiamine. TAPS treatment maintain DNA in its double-stranded form. Compared to bisulfite conversion, the entire reaction is mild and causes far less DNA damage. Furthermore, since normal cytosines are not converted to uracil, the complexity of the input DNA is kept the same as normal DNA, resulting in more uniform amplification, higher mapping rate, and less PCR bias. However, due to DNA damage caused by Fe(II) during TET oxidization reaction and the several DNA purification steps involved, DNA loss is still prevalent in TAPS requiring high amounts of DNA (10 ng) for starting material.

The first step of TAPS utilizes ten-eleven translocation (TET) enzyme for the oxidation of 5mC and 5hmC. TET enzyme belongs to the superfamily of alpha-ketoglutarate-dependent hydroxylases. TET works by using alpha-ketoglutarate (αKG, also known as 2-oxoglutarate, or 2OG) and Fe(II) as co-substrates. By consuming oxygen, TET can incorporate a single oxygen atom from O2 into their substrates. During this conversion, co-substrate αKG will be oxidized into succinate and carbon dioxide (Figure 2.1.1).

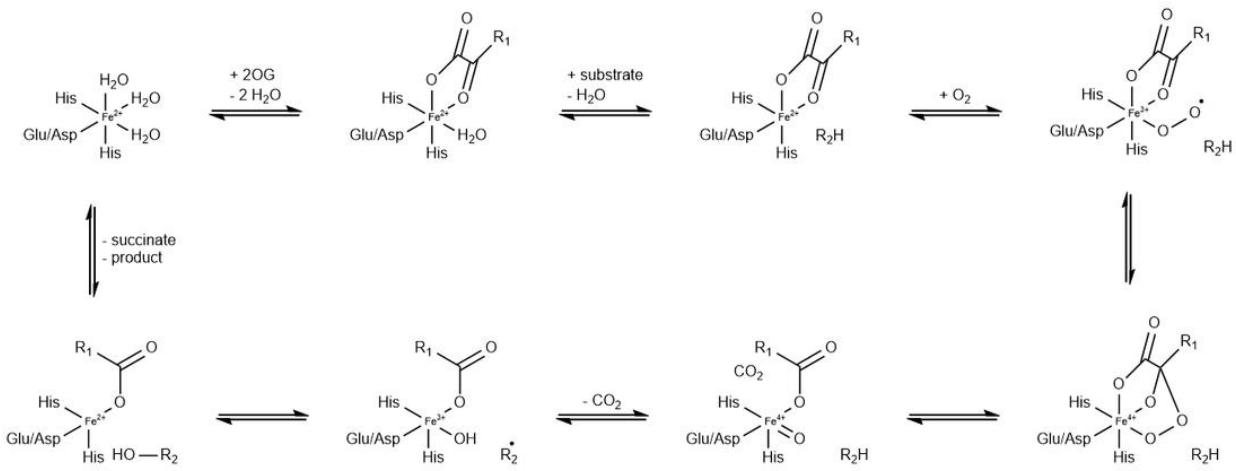

Figure 2.1.1 Mechanism of TET enzyme

akG and Fe(II) serve as co-substrate to transfer a single oxygen atom from O_2 to the substrate. (Wikipededia: Alpha-ketoglutarate-dependent hydroxylases)

Fe(II) is critical during this reaction by serving as a bidentate ligand to αKG, and noncovalently holds the substrate. Subsequently, oxygen molecule binds to Fe(II) and the uncoordinated end of superoxide attacks the keto carbon, resulting in the release of CO2 from αKG and forming a Fe(IV)-oxo intermediate. The intermediate then quickly oxidizes the substrate; in this case, 5-methylcytosine and 5-hydroxylcytosine to 5-carboxylcytosine (Figure 2.1.2).

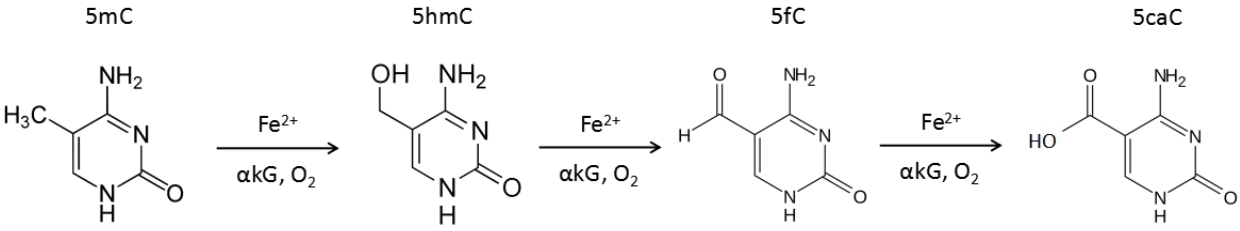

Figure 2.1.2 5mC oxidization pathway by TET

TET oxidizes 5mC to 5caC through 5hmC and 5fC by consuming akG and O_2

After TET oxidization, pyridine borane is used for decarboxylation/deamination reaction to convert 5caC to DHU (Figure 2.1.3). The exact mechanism of the reduction reaction remains to be defined. Reduction efficiency is tightly correlated with buffer pH and salt concentration. 600mM sodium acetate buffer pH 4.3 is used for 5caC reduction to achieve the best conversion efficiency. Normal cytosine will not be affected by pyridine borane reduction, but due to long incubations in high salt condition, low degree of DNA deamination of normal cytosine to uracil has been observed (~ 0.5%). During sequential PCR, DHU is recognized by uracil-recognizing polymerases as Thiamine.

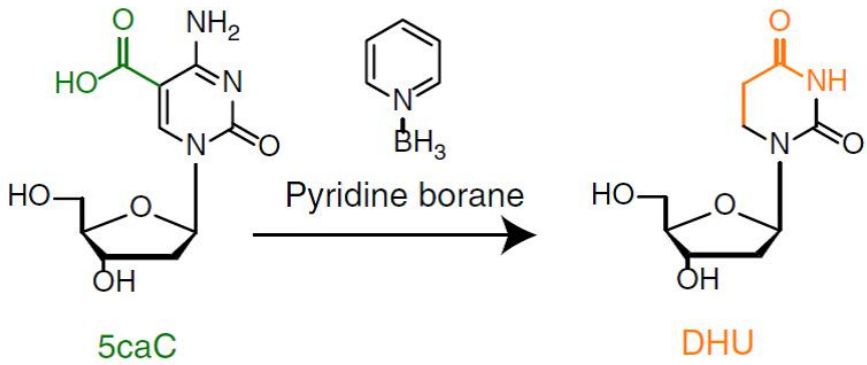

Figure 2.1.3 Pyridine borane reduction of 5caC results in DHU

Pyridine borane reduces both single strand and double strand 5caC at 37 C slowly. The detailed deamination/decarboxylation reaction unknown.

2.1.2 Performance of TAPS

By using CpG fully methylated input DNA and spike-in DNA, the methylation conversion rate could be estimated. De-novo methylation rate of TAPS can also be calculated by using CpG unmethylated DNA as spike-in. Alignment of TAPS requires no additional process compared to standard DNA sequencing, thus exhibits much higher computation speed than bisulfite

sequencing. Cytosine methylation calling is achieved by calling C-T & G-A mutation at every CpG sites (Figure 2.1.4).

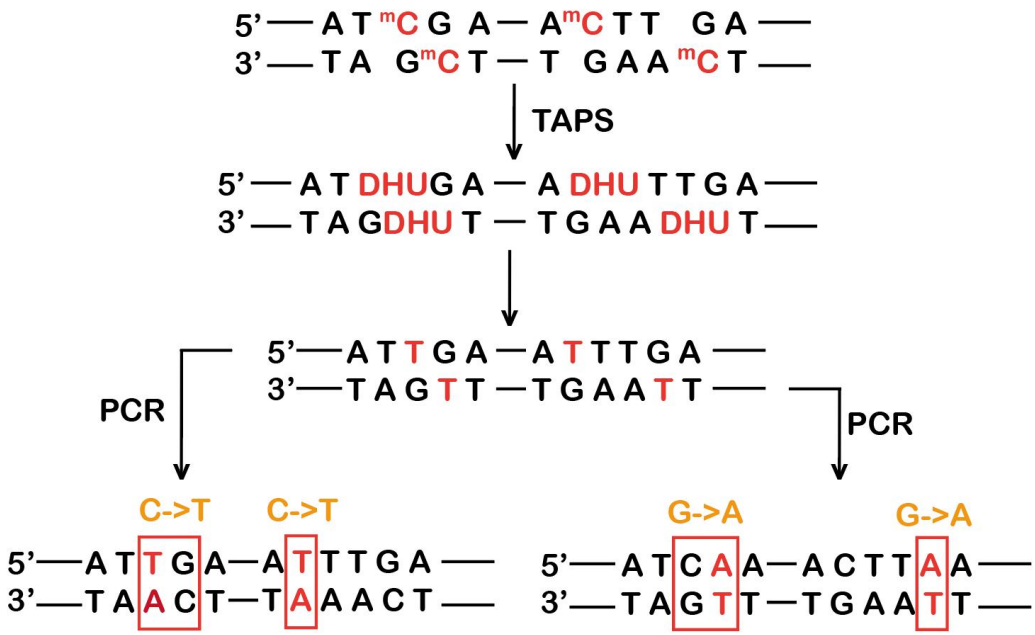

Figure 2.1.4 Cytosine methylation is analyzed by calling C-T or G - A mutation
DHU will be recognized by DNA polymerase as Thiamine, thus introducing C-T mutation or G-A mutation depending on which strand the read origin from.

The TET step of TAPS can be done using a variety of TET protein including but not limited to Human TET2, mouse TET1, and NgTET1. NgTET1 can be expressed and purified from E.coli directly; hence, it is the cheapest and most convenient TET to make. However, according to the original paper's report, mTET1 exhibits higher conversion efficiency (Table 2.1.1). The highest methylation conversion rate achieved by TAPS on bulk DNA is 97% using mTET1 and 84% using NgTET1. The de-novo methylation rate caused by TAPS is 0.2%, which is neglectable.

Conversion rate of 5mC	mTET	NgTET
	pyridine borane	**pyridine borane**
N5mCNN spike-in	97.30%	*
Fully methylated λ-DNA	96.5% ± 0.1	84.1% ± 0.3
2kb unmodified DNA	0.23% ± 0.01	0.26% ± 0.02

Table 2.1.1 TAPS specifications on bulk DNA

TAPS shows promising methyl-conversion rate and neglectable de novo methylation rate. According to the original paper, mTET outperforms NgTET in all aspects.

(Nat Biotechnol. 2019 Apr;37(4):424-429)

In this Chapter, TAPS on single-cell level gDNA is demonstrated to work efficiently using NgTET1 as the oxidation reagent. The reaction condition is optimized to achieve the highest methyl-conversion rate and highest DNA yield.

2.2 Results and Discussion

2.2.1 TAPS on DNA at Single-cell Level using NgTET1

To perform TAPS on gDNA (~10pg) at single-cell level or single cells requires adding PCR priming sites before conversion. Tn5 insertion is used to efficiently fragmentizing and adding adapters to gDNA. Specifically, Nextera Tn5 insertion, or META Tn5 insertion platform is used. Insertion density can be adjusted by Tn5 transposon complex concentration. Because the TET oxidization step in TAPS requires Fe(II) activity, EDTA should be depleted before TAPS. Since the equilibrium constant of Fe(II) - EDTA complex is 14.30, it's challenging to remove EDTA by adding other harmless metal ions, such as Mg and Ca. Thus, for effectively oxidizing 5mC to 5caC, EDTA must be removed from the solution by DNA purification. After testing multiple DNA purification methods, column purification from ZYMO was selected due to high DNA recovery yield.

To measure the methyl-conversion efficiency of TAPS, we used a CpG fully methylated Human somatic gDNA (ThermoFisher). The gDNA has a 97.2% CpG methylated rate that is verified by bisulfite sequencing. Single-cell level (10pg) of CpG methylated gDNA is fragmentized and adaptor-ligated by Tn5 transposon peaked at 200bp. Gap fill is achieved by Q5 polymerase, and column purification is followed by removal of excess EDTA and dNTPs. NgTET1 is chosen for testing as the oxidation enzyme due to its ease of purification and mass productivity.

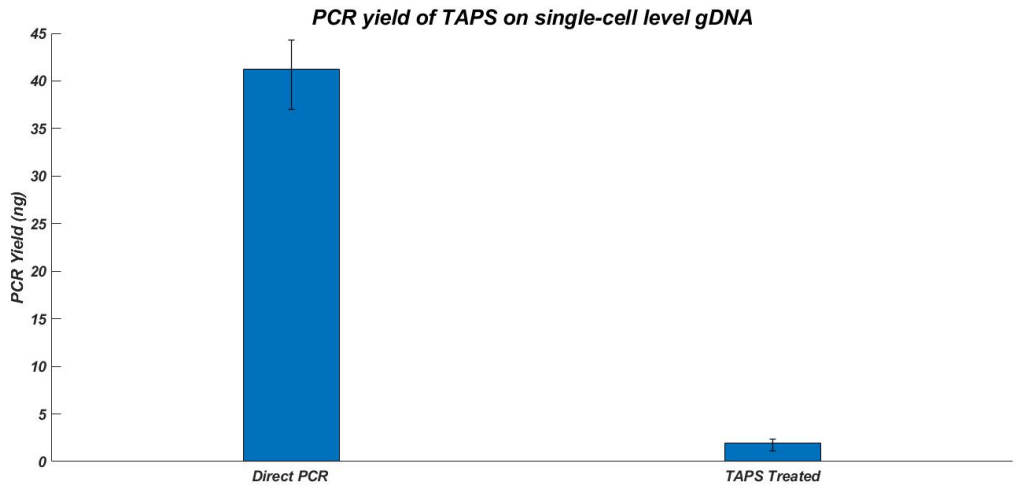

Figure 2.2.1 PCR yield of TAPS on single-cell level gDNA

Only 4.7% of DNA is recovered after TAPS applied on single-cell level gDNA (10pg). 15 PCR cycles were performed on Tn5 inserted gDNA treated or not treated with TAPS, using 2x master mix of Q5U. PCR yield is measure by Qubit after DNA purification and size selection removing primer dimers.

Direct application of TAPS on gDNA at single-cell level results in tremendous DNA loss (Figure 2.2.1). Only 4.7% of DNA is recovered after TAPS, mainly because of low DNA recovery due to three column purification steps. Zymo column purification is only efficient when the input DNA amount is over 10ng, while single-cell level gDNA is only around 10pg. The low DNA recovery of TAPS makes it impossible to directly apply TAPS on single cell methylation sequencing without any modifications.

2.2.2 Reducing DNA Lost by Lambda Carrier DNA

To reduce DNA lost caused by column purification, lambda DNA is added to single-cell level gDNA prior to TAPS. Lambda DNA is sonicated to 200bp but not adapter ligated, thus will not be amplificated during sequential PCR. Up to 100ng of lambda DNA can be added to

the solution without significantly affecting the PCR efficiency of single-cell level Tn5 inserted gDNA. Since the lambda DNA's fragment length is comparable to Tn5 inserted gDNA, DNA lost during column purification, pipetting and liquid transfer will mostly be undertaken by lambda DNA, hence serving as carrier DNA.

The addition of lambda carrier DNA dramatically reduces the DNA lost during TAPS without reducing the methyl-conversion efficiency (Figure 2.2.2). Different amount of lambda DNA is tested for optimization. 20ng of lambda carrier DNA is selected as the most optimal condition, which exhibits on average 80 ~ 85% DNA recovery of column purification after NgTET1 incubation and pyridine borane incubation.

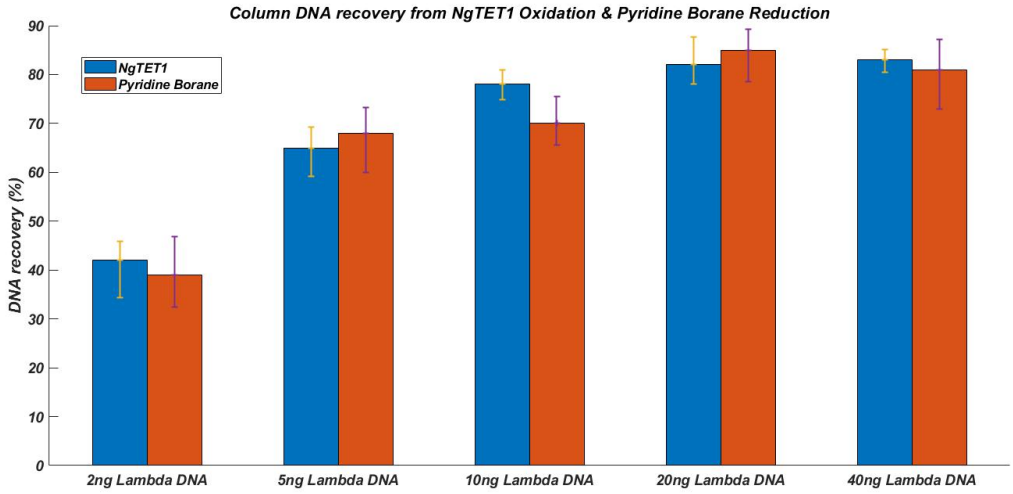

Figure 2.2.2 Column DNA recovery from NgTET1 oxidation & pyridine borane reduction individually

Column DNA recovery measure after NgTET1 incubation and pyridine borane incubation using Qubit. The percentage recovery is calculated by Qubit ratio before and after each step individually. 20ng of lambda carrier DNA is selected to be the most ideal amount of carrier due to high recovery.

2.2.3 Reducing DNA Damage by ATP

Other than DNA loss during column purifications, the DNA damage caused by Fe(II) also reduces the final coverage of TAPS on single-cell gDNA. The highly efficient enzymatic oxidation step in TAPS requires a large pool of free Fe(II) as a co-factor. However, free Fe(II) is highly reactive, and could easily be oxidized to Fe(III) by trace amounts of superoxide in H2O. To maintain accessible Fe(II), L-Ascorbic acid is used as a reducing agent to reduce Fe(III) back to Fe(II). During this oxidation-reduction cycle of Fe(II) and Fe(III), a large number of hydroxyl radicals is formed according to Fenton Reaction (ref). Hydroxyl radicals are considered to be the top DNA damaging reagent, which could result in numerous types of DNA damages, including dsDNA break, ssDNA nicks, and oxidized base. Longer DNA fragments have higher chances of being damaged by hydroxyl radical attacks, so TAPS treated DNA exhibits an uncommon upper size selection feature (Figure 2.2.4)

Multiple common hydroxyl radical scavengers are tested to reduce the DNA damage, including Mannitol, Sorbitol, and glycerol, but none seems to work efficiently. One hypothesis is that hydroxyl radical produced on DNA by DNA bonded Fe(II) will immediately react with DNA, giving hydroxyl radical scavengers no time to absorb the oxidation stress. Noteworthy, other hydroxyl radicals not produced on DNA will react with nearby molecules such as MOPS in picoseconds, and will not damage DNA even without the presence of scavengers.

However, by using Fe(II) ligand ATP, one could effectively reduce the DNA damage during TET oxidation. Unlike EDTA, ATP will preferably bind to Fe(II) rather than Fe(III), and won't cause oxidation state change of Fe ion. The bond is weak so that NgTET1 can still have access to Fe(II) while the formation of hydroxyl radicals by blocking superoxide was largely reduced. By adding 1mM of ATP, longer DNA remain intact after TAPS treatment

(Figure 2.2.4). However, ATP reduces the amount of free Fe(II), so a higher concentration of Fe(II) is needed to reach the same methyl-conversion efficiency. Combined with 20ng of lambda carrier DNA, a final DNA recovery of 60 ~ 70% could be achieved for sc-TAPS.

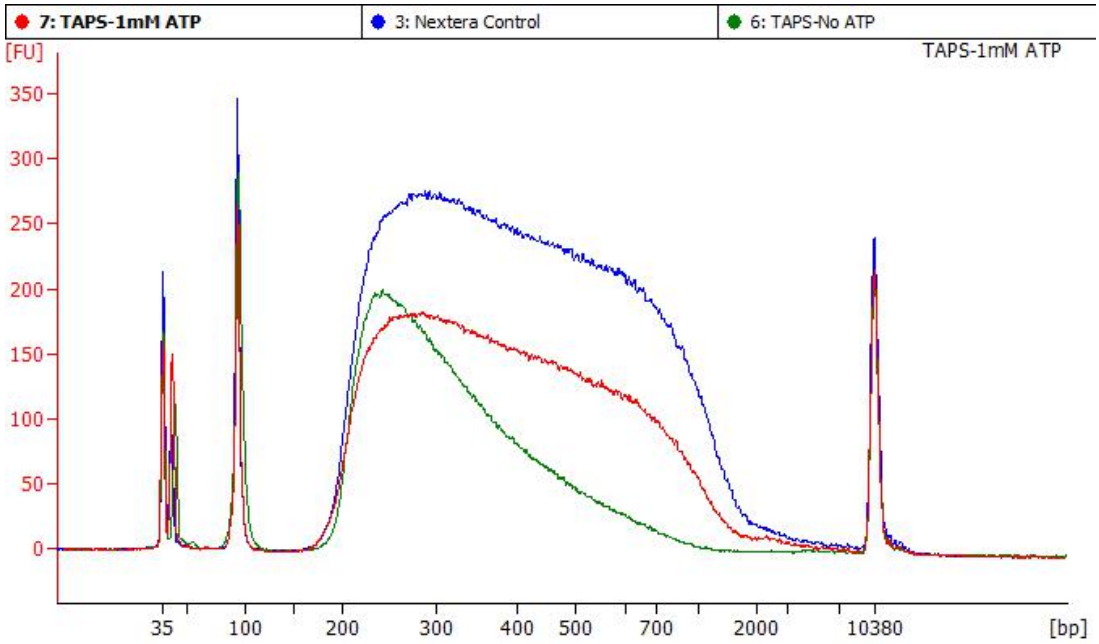

Figure 2.2.4 Bioanalyzer plot of Nextera control DNA library versus TAPS treated DNA library with/without ATP supplyment.

Bioanalzyer result of DNA libraries from High Sensitivity DNA kit. All libraries performed PCR for 14 cycles. Blue: Nextera control, direct PCR using NextEra schematic on 10pg gDNA. Green: TAPS treated 10pg gDNA amplified with NextEra schematic. Red: TAPS treated 10pg gDNA with 1mM ATP supplement, amplified with NextEra schematic.

2.2.4 Optimization of TAPS for Higher Methyl-Conversion Efficiency

The addition of ATP and lambda carrier DNA increases the DNA recovery of single-cell TAPS from 4.7% to over 60% total, but sc-TAPS can only achieve 64% methyl-conversion rate using NgTET1 according to the original protocol (ref). Optimization based on sc-TAPS must be

performed to make sc-TAPS practical to use. Extensive comparative experiments have been conducted aiming for optimization of methyl-conversion rate and DNA recovery. Parameters include buffering reagent selection, buffer pH, concentration of NgTET1, salt, Ascorbic acid, DTT, αKG, Fe(II), ATP & PEG 8k. Among all variables, the concentration of αKG and Fe(II) alters methyl-conversion rate the most (Figure.2.2.5). With the addition of PEG8k, the methyl-conversion rate of sc-TAPS increases by 9.8% when applying 6% PEG8k, probably due to PEG8k serving as a crowding regent, just like in single cell transposon insertion.

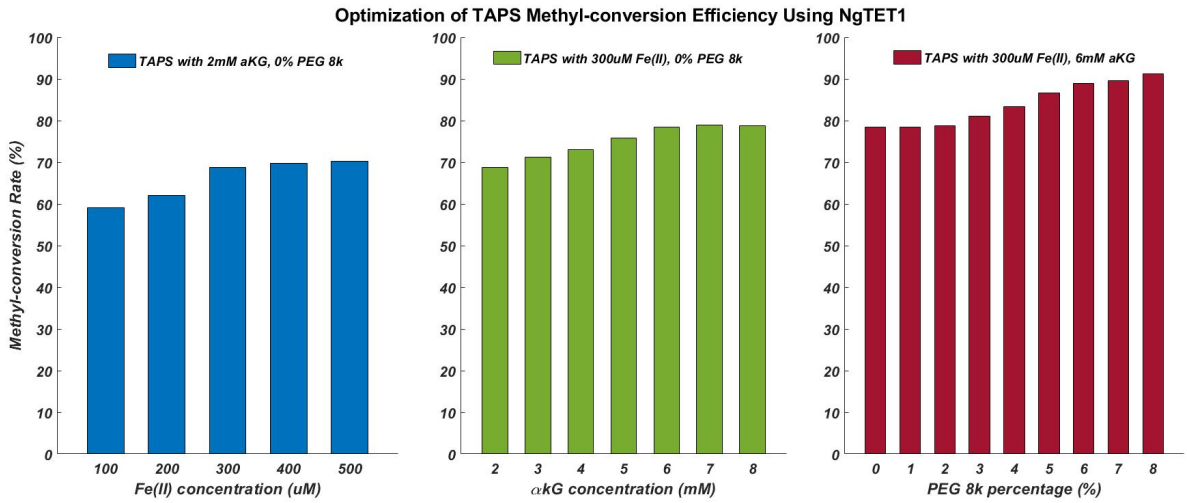

Figure 2.2.5 Optimization of sc-TAPS on Methyl-conversion rate.

Methyl-conversion rate of sc-TAPS is calculated based on the methylation rate of sc-TAPS treated CpG fully methylated Human gDNA after sequencing. Blue: Methyl-conversion rate versus Fe(II) concentration when 1mM ATP is supplied. Green: Methyl-conversion rate versus akG concentration, Red: Methyl-conversion rate versus PEG8K concentration.

However, Increasing Fe(II) and PEG8K concentration also reduce DNA recovery (Figure 2.2.6). The mechanism behind PEG causing DNA loss is unknown. Overall, the optimal sc-TAPS condition is selected as follows:

50mM MOPS pH 6.9, 50mM NaCl, 2mM L-Ascorbic acid, 6mM αKG mono salt, 3mM DTT, 6% PEG8k, 300uM Ammonium iron(II) sulfate, 1mM ATP, 5uM NgTET1

The methyl-conversion rate of sc-TAPS is 90.2%, which is calculated based on the methylation rate of sc-TAPS treated CpG fully methylated Human gDNA. DNA recovery is on average 68.3% for the entire sc-TAPS procedure.

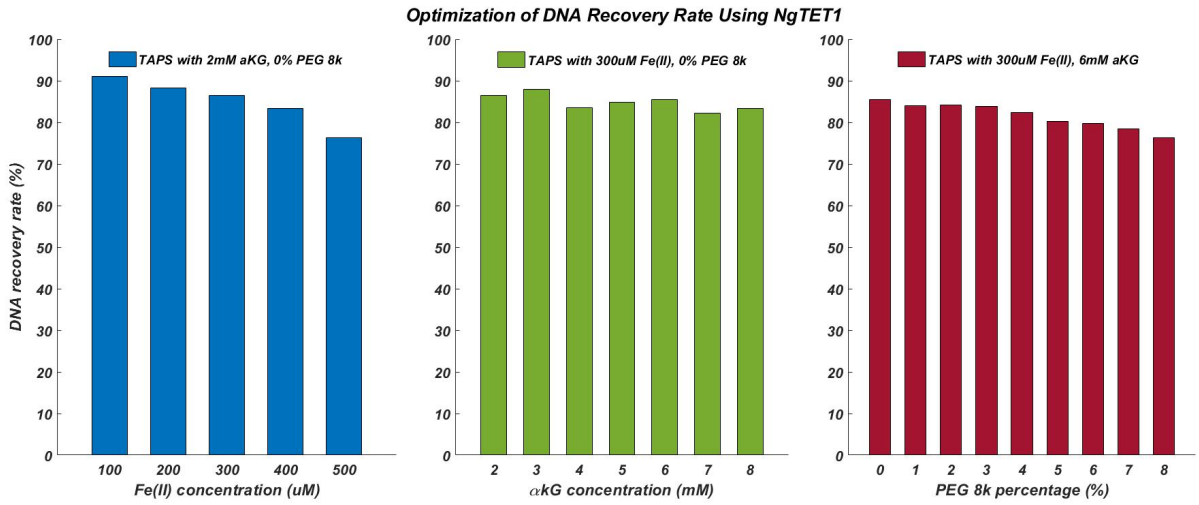

Figure 2.2.6 Optimization of sc-TAPS on DNA Recovery

DNA recovery of sc-TAPS is calculated based Qubit measure after NgTET1 incubation. Blue: DNA recovery versus Fe(II) concentration when 1mM ATP is supplied. Green: DNA recovery versus akG concentration, Red: DNA recovery versus PEG8K concentration.

2.2.5 Performance of sc-TAPS on Single Cells

Eight GM single cells were picked by mouth-pipetting into lysis buffer supplied with protease Q in low-bind 200ul PCR tubes. Tn5 insertion with META schematic is performed on individual cells with peak centered at 250bp. sc-TAPS was applied to all eight single cells for 14 PCR cycles with unique barcodes. The final PCR libraries were pooled together and size-selected with 1.2x AMpure beads. The final library was checked by qubit and FA before

submitting to Illumina X10. All libraries were sequenced 2x 150bp. In total, 82 G of data was received.

Applying sc-TAPS to GM single cells results in an average of 28% methylome coverage with 3x sequencing depth, nearly three-fold higher compared to previous techniques at the depth. The mapping rate is 91.2%, which is three-fold higher than scWGBS, and computation time only takes one-fourth comparing to bisulfite sequencing. After analysis of fully methylated and unmethylated lambda DNA spike-in, on average a 90.6% methyl-conversion rate and a 0.48% de novo methylation rate are achieved.

The performance of sc-TAPS on GM cells demonstrated that sc-TAPS outperforms sc-WGBS in almost every aspect except methyl-conversion rate. The lower methyl-conversion rate compared to bisulfite sequencing might be caused by sequence preference of NgTET1. To solve this problem, three bulk TAPS libraries generated from fully methylated human gDNA were sequenced by x10 with an average depth of 10x. A blacklist was generated by summarizing all the CpGs that were not methylated in all three libraries. After filtering the blacklist, the methyl-conversion rate of sc-TAPS increased to 94.7% calculated from previous samples generated from CpG fully methylated gDNA.

In conclusion, sc-TAPS can achieve higher methyl-conversion rate with higher mapping efficiency and faster computation speed compare to scWGBS. sc-TAPS is valuable to apply to various single-cell scenarios where single-cell methylation pattern is crucial.

2.3 Material and Methods

2.3.1 NgTET purfication

Protocol for NgTet1 Purification

1. Pick a single colony to start 5mL overnight growth.

2. Innoculate 2L culture with overnight culture and grow at 220RPM 37C to A600 ~0.5.

3. Induce with 0.2mM IPTG at 16 C at 180rpm and grow overnight.

4.Harvest cells by centrifugation @ 4000g and freeze pellet in -80 C.

1.Resuspend pellet in 40mL/L Lysis Buffer

20mM HEPES pH, 500mM NaCl, 30mM Imidazole, 1mM DTT

2.Lyse by sonication with 3s on/3s off for 20min at 30% duty cycle.

3. Clarify lysate by centrifugation for 1hr at 35,000g.

4. Add 2.1mL 10% neutralized PEI dropwise to 80mL lysate. Incubate on ice for 15min.

5. Spin the sample at 15,000g for 15min.

6. Add solid ammonium sulfate to 60% saturation. Incubate on ice for >15min.

7. Spin down the sample at 15,000g for 20min.

8. Resuspend pellet in lysis buffer and dialyze overnight in lysis buffer.

9. Filter protein sample and load onto a nickel column.

10.Wash with 5CV lysis buffer and 5CV High Salt Buffer

20mM HEPES pH, 1M NaCl, 1mM DTT

11. Elute protein in 300mM imidazole.

12. Dialyze overnight in the presence of Tev protease into low salt buffer.

20mM HEPES pH, 10mM NaCl, 1mM DTT

13.Load protein onto a tandem Q-SP ion exchange column.

14. Wash with 5CV low salt buffer.

15. Elute on a gradient from 10mM-1M NaCl.

16.Pool fractions and exchange into low salt buffer.

17. Load onto heparin column and elute on a gradient from 10mM-1M NaCl.

18. Pool protein fractions, concentrate and load onto an S75 equilibrated with size buffer.

20mM HEPES pH, 150mM NaCl, 1mM DTT.

19.Pool fractions, concentrate, and store at -80 C in 30% glycerol.

2.3.2 Tn5 insertion of single-cell level DNA and single cells

Single Cell Lysis

1. Prepare Lysis Buffer (2 uL per cell; recipe below for 1 mL):

a) 20 uL 1 M Tris pH 8.0 (Invitrogen 15568025; final: 20 mM)

b) 4 uL 5 M NaCl (Invitrogen AM9760G; final: 20 mM)

c) 15 uL 10% Triton X-100 (Sigma 93443; final: 0.15%)

d) 150 uL 100mM DTT (Sigma 43816; final: 15 mM)

e) 2 uL 0.5 M EDTA (Invitrogen AM9260G; final: 1 mM)

f) 5 uL 100 uM Carrier ssDNA (final: 500 nM)

g) 804uL water

h) Mix and store at -20 C.

2. Prepare 2 X Transposition Buffer (5 uL per cell; recipe below for 1 mL):

a) 20 uL 1 M TAPS pH 8.5 (Boston Bio Products BB-2375) (final: 20 mM)

b) 10 uL 1 M MgCl2 (final: 10 mM)

c) 320 uL 50% PEG 8000 (Hampton Research HR2-535) (final: 16%)

d) 650 uL water

e) Mix well and store at -20 C.

3. Prepare Transposome Removal Buffer (1 uL per cell; recipe below for 1 mL):

a) 120 uL 5 M NaCl (final: 600 mM)

b) 180 uL 0.5 M EDTA (final: 90 mM)

c) 2 uL 10% Triton X-100 (final: 0.02%)

d) 698 uL water

e) Mix and store at -20 C.

Lysis

1. Prepare 7.5 mg/mL Qiagen Protease

a) 1 uL 60 mg/mL Qiagen Protease

b) 7 uL water

Add 2ul lysis buffer per tube

Add 0.5 ul 7.5mg/ml QP

Lysis the cells by running (2.5 uL volume, close tight PCR machine)

a) 50 C for 1 h

b) 65 C for 1 h

b) 70 C for 15 min

c) 4 C forever

Tn5 Insertion

Nextera construct:

Homemade Nextera transposons have one strand of 5′-/Phos/-CTGTCTCTTATACACATCT-3′ and one strand of either 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′ (“P5”) or 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′ (“P7”)

(IDT, purification: PAGE).

Nextera XT (Illumina) can in principle substitute, but has not been tested thoroughly.

Homemade Nextera index primers (IDT, purification: standard desalting; then dissolved in 0.1 X TE to 5 uM and stored at −20 C) are in the format of

5′-CAAGCAGAAGACGGCATACGAGAT-[i7]-GTCTCGTGGGCTCGG-3′ and

5′-AATGATACGGCGACCACCGAGATCTACAC-[i5]-TCGTCGGCAGCGTC-3′.

Their sequences are:

701: CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG

702: CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG

703: CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG

704: CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG

705: CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG

706: CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG

707: CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG

708: CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGG

709: CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGG

710: CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG

711: CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG

712: CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG

501: AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTC

502: AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC

503: AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC

504: AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC

505: AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC

506: AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC

507: AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC

508: AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC

META construct:

In this work, we used META with n = 20 tags:

1. AGAAGCCGTGTGCCGGTCTA

2. ATCGTGCGGACGAGACAGCA

3. AATCCTAGCACCGGTTCGCC

4. ACGTGTTGCAGGTGCACTCG

5. ACACCACACGGCCTAGAGTC

6. TGGACAATCACGCGACCAGC

7. TCATCTAACGCGCACCGTGC

8. TTCGTCGGCTCTCTCGAACC

9. TGGTGGAGCGTGCAGACTCT

10. TATCTTCCTGCGCAGCGGAC

11. CTGACGTGTGAGGCGCTAGA

12. CCATCATCCAACCGGCTTCG

13. CACGAGAAGCCGTCCGCTTA

14. CGTACGTGCAACACTCCGCT

15. CTTGGTCAGGCGAGAAGCAC

16. GGCGTGATCAGTGCGTGGAT

17. GAGCGTTTGGTGACCGCCAT

18. GCCTGCGGTCCATTGACCTA

19. GTAAGCCACTCCAGCGTCAC

20. GATCTGTTGCGCGTCTGGTG

Carrier ssDNA (for use in Lysis Buffer) is the same as in LIANTI

5′TCAGGTTTTCCTGAA-3′

Tn5 transposon is constructed from

5′-/Phos/ CTGTCTCTTATACACATCT-3′, while the other strand was in the form of 5′-[META tag]-AGATGTGTATAAGAGACAG-3′.

Each of the oligos (IDT, purification: PAGE) was dissolved in 0.1 X TE to a final concentration of 100 uM. For each of the n = 20 META tags, two strands were annealed at a final concentration of 5 uM each. The 20 annealed transposons were then pooled with equal volumes. Second, the transposase was purified after expression from the pTXB1-Tn5 plasmid (Addgene). Transposome was assembled at a final concentration of 1.25 uM dimer (2.5 uM monomer), 1:10 diluted (125 nM dimer, or 250 nM monomer), and aliquoted for single uses and store at −80 C.

20-primer Mix (for use in PCR Mix 1) was in the form of 5′-[META tag]AGATGTGTATAAG-3′. Each of the oligos (IDT, purification: standard desalting) was dissolved in 0.1 X TE to a final concentration of 100 uM, and combined with equal volumes (100 uM total, or 5 uM each). Store at −20 C. 40-primer Mix (for use in PCR Mix 2) was in the form of

5′ACACTCTTTCCCTACACGACGCTCTTCCGATCT-[METAtag]AGATGTGTATAAG-3′

for one side of the Illumina adapter, and

5′GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-[METAtag]AGATGTGTATAAG-3′

for the other. Each of the oligos (IDT, purification: PAGE) was dissolved in 0.1 X TE to a final concentration of 50 uM, and combined with equal volumes (50 uM total, or 1.25 uM each). Store at −20 C.

Tn5 insertion using nextera or META, 10ul reaction

For each single cell, add each reagents in low-bind PCR tubes:

A) 2.5ul lysis sample

B) 5ul 2x Trans buffer

C) 2.5ul diluted tn5 complex

55C for 10mins

4 C forever

Stop tn5

Prepare 0.2mg/ml QP

prepare stop buffer

1ul 2x stop buffer

1ul 0.2mg/ml QP (final 100ug/ml)

1. Stop transposition by running (12 uL volume):

a) 50 C for 40 min

b) 70 C for 20 min

c) 4 C forever

Gap fill

1. Make PCR Mix (13 uL per cell)

a) 5 uL Q5 Reaction Buffer (included with Q5)

b) 5 uL Q5 High GC Enhancer (included with Q5)

c) 0.6 uL 10 mM (each) dNTP mix (NEB N0447S)

d) 0.6 uL 100mM MgCl2 (Invitrogen AM9530G)

e) 0.25 uL 20 mg/mL BSA (NEB B9000S)

f) 0.25 uL Q5 (NEB M0491S)

g) 1ul 20ng/ul lambda carrier DNA (sonicated 200-300 bp)

e) 0.3ul H20

Vortex to mix

2. Add 13 uL PCR Mix per tube, avoiding touching the liquid. Vortex and spin down.

Gap fill by running (25 uL volume):

a) 4 C for 3 min (to allow the lid to pre-heat)

b) 65 C for 3 min

c) 4 C store

Clean up

1. Add 200ul binding buffer direct to the pcr tube, mix 10 times, transfer to column (ZYMO DCC)

2. 200ul wash twice

3. Elute in 8.5ul elution buffer (No edta, NEB white cap bottle from EM-seq kit)

2.3.3 scTAPS protocol

Prepare solution:

(Sigma-Aldrich, K2010) α-Ketoglutaric mono salt 16.8mg in 1ml H2O for 100mM

(Sigma-Aldrich, 95209) Ascorbic acid 17.6 mg in 1ml H2O for 100mM

20 mg Fe (Sigma Aldrich, 09719) in 8.5ml H2O for 6mM

2x NgTET1 Buffer	Conc.	Volume/uL	2x Buffer Conc.	Final Buffer condition
MOPs buffer pH=6.9	1M	100	100mM	50mM
aKG salt	100mM	120	12mM	6mM
Ascorbic acid (V.C.)	100mM	40	4mM	2mM
DTT	100mM	60	6mM	3mM
ATP	10mM	200	2mM	1mM
NaCl	5M	12	60mM	30mM
40% PEG 8k		300	12%	6.00%
H2O		168		
Total		1000		

NgTET Reaction

Combine:

1) 10ul 2x NgTET Buffer

2) 1ul NgTET1 (100uM)

Add 11ul NgTET1 mix to 8ul of Tn5 Elution on ice

Add 1ul 6mM Fe(II)

Mix with vortex (not too fast)

All tubes incubate at 34℃, 60min; 4℃,store

Prepare Stop Buffer

Mix 9.5ul protease K with 0.5ul 0.5M EDTA

Add 1ul Stop buffer to each

Mix with vortex (not too fast)

All tubes incubate at 50℃, 30min; 4℃,store

Purify by Zymo column (DCC), Add 200ul binding to each. Elute in 22ul elution buffer (with edta). (Elution buffer incubate for 2 mins)

Borane Reaction

To 21ul NgTET1 elution

Add 3ul of 10M Pyridine Borane

6ul 3M Sodium Acetate

Mix with vortex (not too fast)

37 C for 16 hours, stop at RT forever

Purify by Zymo column (DCC), Add 200ul binding to each. Elute in 20ul elution buffer

Library prep

Nextera construct

Combine the following:

19.2ul DNA elute

20ul Q5U 2x master mix

0.8 ul 100mM nextera index

Amplify by running (40 uL volume):

a) 4 C for 3 min (to allow the lid to pre-heat)

c) 98 C for 20 s

d) 14 cycles of 98 C for 10 s, 62 C for 30 s, 72 C for 1 min

e) 72 C for 5 min

f) 4 C forever

1.2 x Beads size selection using AMpure Beads.

META construct:

Combine the following:

META 20-primer Mix, 0.8 uL

19.2ul DNA elute

20ul Q5U 2x master mix

incubation at

72 C for 3 min,

98 C for 20 s,

12 cycles of [98 C for 10 s, 65 C for 1 min,72 C for 2 min],

and 65 C for 5 min.

The amplification product is purified at this step and the sequencing libraries were prepared by two additional PCR steps. In the first PCR step, previous primers were removed by the addition of 0.5 uL 20 U/uL ExoI (NEB M0293S) and incubation at 37 C for 30 min, 72 C for 20 min. White precipitates might form at this step or at the following steps. PCR was performed by the addition of 9.5 uL PCR Mix 2 (2 uL Q5 reaction buffer (NEB), 2 uL Q5 high GC enhancer (NEB), 3 uL 50 uM (total) META 40-primer Mix, 0.2 uL 10 mM (each) dNTP mix, 2.2 uL water, 0.1 uL Q5 (NEB M0491S)) and incubation at 98 C for 30 s, 2 cycles of 98 C for 10 s + 65

C for 1 min + 72 C for 2 min, and 65 C for 5 min. In the second PCR step, primers were similarly removed by the addition of 0.5 uL 20 U/uL ExoI (NEB M0293S) and incubation at 37 C for 30 min, 72 C for 20 min. PCR was similarly performed by the addition of 2.5 uL NEB Index Primer (NEB E7335S, E7500S, E7710S, E7730S) and 7 uL PCR Mix 3 (2 uL Q5 reaction buffer (NEB), 2 uL Q5 high GC enhancer (NEB), 2.5 uL NEB Universal Primer, 0.2 uL 10 mM (each) dNTP mix, 0.2 uL water, 0.1 uL Q5 (NEB M0491S)) and incubation at 98 C for 30 s, 2 or more cycles of 98 C for 10 s + 65 C for 1 min + 72 C for 2 min, and 65 C for 5 min. Libraries could be pooled at this step or at any step afterwards. 1.2x AMpure beads is used for size selection

2.4 Reference

1. Liu Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. Nat Biotechnol. 2019 Apr;37(4):424-429.

2. Xia B, Han D, Lu X, et al. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat Methods*. 2015;12(11):1047–1050.

3. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev*. 2016;30(7):733–750.

4. Shen L, Song CX, He C, Zhang Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu Rev Biochem*. 2014;83:585–614.

5. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*. 2013;502(7472):472–479.

6. Shen L, Wu H, Diep D, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*. 2013;153(3):692–706.

7. Shen L, Zhang Y. Enzymatic analysis of Tet proteins: key enzymes in the metabolism of DNA methylation. *Methods Enzymol*. 2012;512:93–105.

8. Wu H, Wu X, Shen L, Zhang Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnol*. 2014;32(12):1231–1240.

9. Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev*. 2011;25(23):2436–2452.

10. Li E, Zhang Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol*. ;6(5):a019133.

11. Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 2011;333(6047):1300–1303.

12. Xu Y, Xu C, Kato A, et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell*. 2012;151(6):1200–1213.

13. Xu Y, Wu F, Tan L, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell*. 2011;42(4):451–464.

14. Arand J, Spieler D, Karius T, et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet*. ;8(6):e1002750.

15. He YF, Li BZ, Li Z, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011;333(6047):1303–1307.

Chapter 3

Methylome Replication Loops with Methyl-Transferase (MERLOT)

Although the sc-TAPS method significantly reduces DNA loss caused by the purification procedure and DNA damage, still an average of 40% DNA will be unavoidably lost during the TAPS treatment. Instead of further optimization of DNA loss reduction, perhaps amplifying the DNA before TAPS treatment could also solve the problem of DNA loss by compensation. However, DNA amplification through PCR reaction would erase all methylation modifications since DNA polymerase does not exhibit methyl-transfer activity, therefore this would not be a possible solution.

In this Chapter, we present Methylome Replication Loops with Methyl-Transferase (MERLOT) as a highly efficient DNA methylation amplification method. MERLOT utilizes methyl-transferase DNMT1 for methylation pattern maintenance. Combined with PCR reaction, MERLOT could amplify the DNA template 1.7-fold per cycle. Over 95% efficiency could be achieved by MERLOT when methylating hemi-methylated double-stranded DNA templates, while having a rather low de novo methylated rate of 1.5% on un-modified CpG sites.

MERLOT followed by sc-TAPS gives by far the highest coverage of all mammalian single-cell whole methylome sequencing techniques. scMERLOT-TAPS on GM12878 single cells results in 37.3% genome coverage with one round of MERLOT. More rounds of MERLOT may bring even higher genome coverage, but have not yet been tested.

3.1 Background

3.1.1 DNMT1 Maintains Methylation Pattern During DNA Replication

Standard PCR methods will not incorporate methylation marks into newly synthesized DNA template since DNA polymerase does not have methyl-transferase activity. Hence PCR amplification of any DNA will result in erasing methylation information. However, during the replication of mammalian somatic cells, DNA methylation pattern will be inherited in the two daughter cells. The critical player behind the methylome replication has found to be the methyl-transferase DNMT1. DNMT1 is highly transcribed during S phase and is responsible for methylating the newly generated hemimethylated sites on daughter DNA strands. DNMT1 specifically recognizes double-stranded CpG sites and is 30-fold more specific toward hemi-methylated CpG sites over non-methylated CpG sites based on in vitro experiments. This property makes DNMT1 the ideal methyl-transferase method to maintain the methylome pattern unaltered during DNA replication.

In this Chapter, we demonstrate that DNMT1 could be used with combination of PCR cycles to amplify DNA while maintaining methylation status intact.

3.2 Results and Discussion

3.2.1 DNMT1 Methylates Hemi-methylated Templates with High Efficiency

To test the performance of DNMT1 in vitro and to develop an optimal reaction buffer for the amplification and methylation reactions described herein, synthetic dsDNA containing a methylation sensitive restriction cutting site is used (Figure 3.2.1). The 87bp dsDNA contains a 6bp Cla1 motif, 5'-ATCGAT-3', which is CpG methylation sensitive. The 87bp dsDNA will be cleaved into two fragments once incubated with Cla1 restriction enzyme if the CpG site is unmethylated. If the CpG site is methylated, the dsDNA will not be cleaved. When the template is hemi-methylated, Cla1's cleavage rate is largely reduced but will still result in fragmentation if the reaction time is long enough to reach saturation. By running electrophoresis on a high sensitive DNA bioanalyzer after Cla1 incubation, the percentage of intact dsDNA template is calculated, which indicates the percentage of fully methylated template in the sample (Figure 3.2.1).

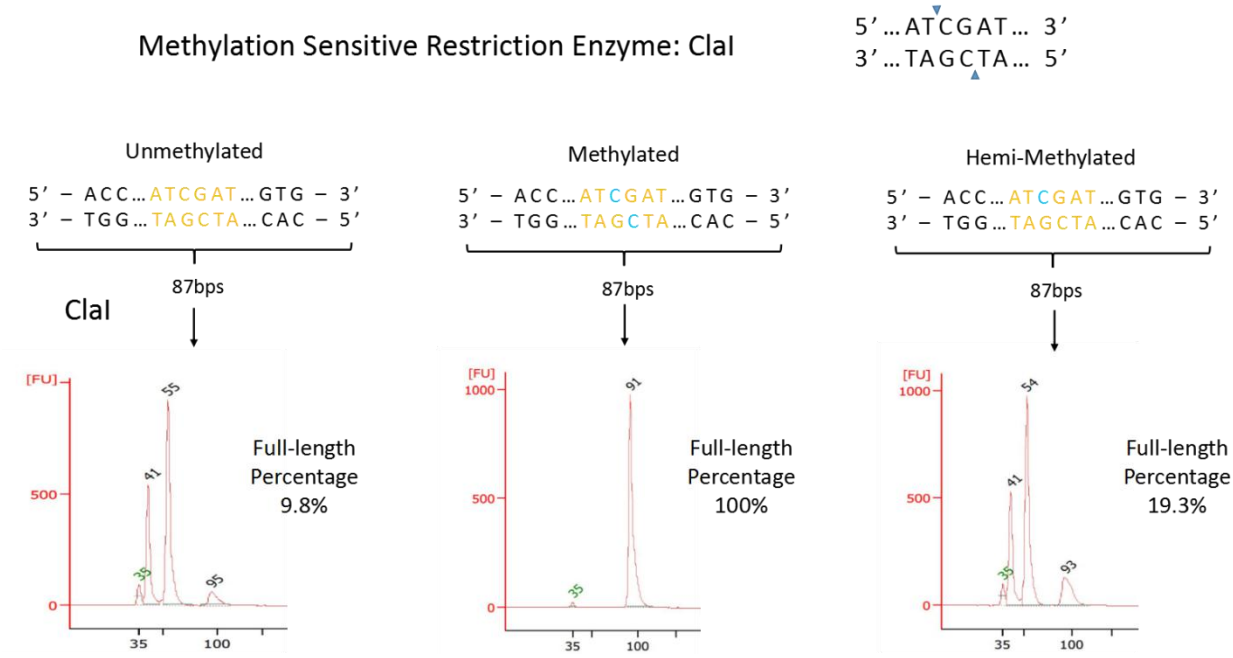

Figure 3.2.1 Cla1 methylation-sensitive restriction enzyme assay

An 87bp synthetic oligo is made containing a 6bp Cla1 methylation-sensitive restriction site. The dsDNA template is annealed from synthetic ssDNA containing or not containing a methylated CpG site. The methylation status of the dsDNA templated could be revealed by incubation with Cla1 followed by electrophoresis.

To probe the methyl-transferase efficiency of DNMT1, hemi-methylated dsDNA template is incubated with DNMT1 in a homemade methyl-transferase buffer (MT buffer) followed by Cla1 cleavage and electrophoresis (Figure 3.2.2). EDTA is needed to chelate all excess Mg ion in the solution. Presence of Mg will cause DNMT1 binding DNA in a dispersive fashion rather than processive. DNMT1 could achieve near 100% methyl-transferase efficiency in a 15ul reaction containing 2ul 2U/ul DNMT1 (NEB) or 0.2ul 0.5uM DNMT1 (Creative Biomart).

To estimate the de novo methylation rate of DNMT1 on unmethylated CpG sites, unmethylated dsDNA template is incubated with DNMT1 in a homemade buffer (please see methods) followed by Clai cleavage and electrophoresis (Figure 3.2.2). DNMT1 did not show

any visible de novo effects. The residue peak of 87bp is mainly due to template mismatch that is caused by imperfect oligo synthesis.

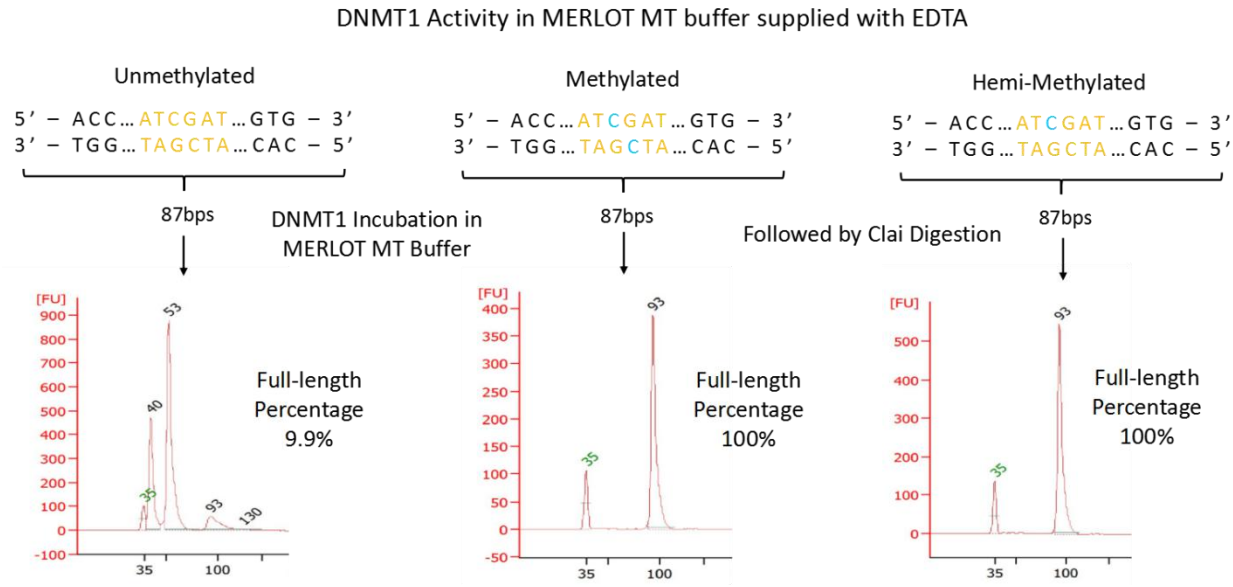

Figure 3.2.2 DNMT1 incubation on 87bp Cla1 dsDNA template with different methylation status
DNMT1 can successfully methylate hemi-methylated dsDNA with high efficiency while having no detectable *de novo* methylation rate.

3.2.2 Methylome Replication Loops with Methyl-Transferase

DNMT1's robust performance can be achieved when reacting in MT buffer, but in order to amplify the methylation status, a robust polymerase extension reaction is also needed. The ideal buffer condition for polymerase extension should not conflict with MT buffer, and the switching between PCR buffer and MT buffer should not be difficult. Unlike DNMT1, nearly all DNA polymerases require Mg concentration to be at least 2mM to maintain processivity in a PCR reaction. This buffer conflict between DNMT1 and DNA polymerase could be solved by Mg-EDTA chelation reaction. EDTA is added to the buffer to chelate all the Mg excess when DNMT1 incubation is needed, while Mg is supplied to 2mM when a PCR reaction is performed. DNMT1 is not a thermo-stable protein, therefore the heat denaturation step during a PCR

reaction would heat inactivate DNMT1. Thus fresh DNMT1 must be added for each methyl-transfer step.

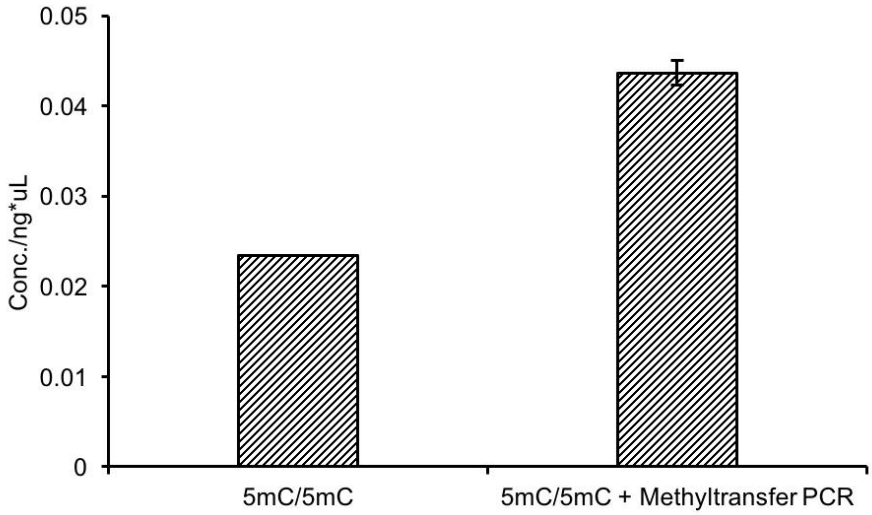

Figure 3.2.3 MERLOT PCR amplification yield on synthetic oligo

One round of MERLOT is performed on 500pg methylated Cla1 dsDNA template. PCR yield was measured by high sensitivity qubit after purification. MERLOT shows a promising 1.7-fold amplification of the dsDNA template for 1 round.

By combining the polymerase extension and DNMT1 methyl-transfer reaction, one can achieve the replication of the methylation status of the original template. The reaction is abbreviated as MERLOT, which stands for Methylome Replication Loops with Methyl-Transferase. 1 Round of MERLOT on the 87 bp methylated template results in 1.7-fold increase in dsDNA amount (Figure 3.2.3) and a 96.6% methyl-transfer efficiency of DNMT1 (Figure 3.2.4). 2 Rounds of MERLOT on the 87 bp methylated template results in 95.4% full-length template, demonstrating the success of the buffer used and the strategy of switching the chelation reaction.

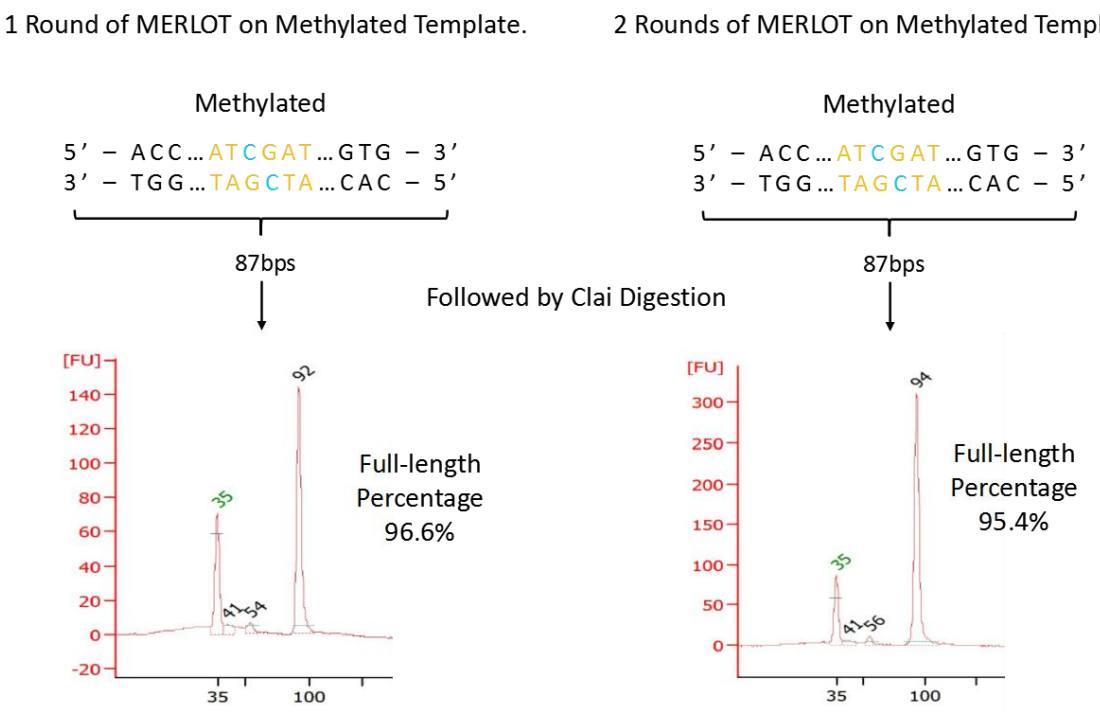

Figure 3.2.4 MERLOT performed on synthetic oligos

1~2 rounds of MERLOT is performed on 87bp Cla1 dsDNA template showing high methyl-transfer efficiency for each round.

3.2.3 MERLOT on gDNA at Single-Cell Level

To test whether MERLOT is viable on gDNA at the single-cell level, bisulfite sequencing was performed on MERLOT amplified, 10 pgs of fully methylated Hela gDNA. PCR adaptors were inserted through Tn5 insertion, as described in previous chapters. MERLOT was performed 1 round, and a control in the absence of MERLOT was also included. Among all the reads that uniquely aligned to the human genome, 98.7% of the cytosine in CpG context are methylated for the control while 93.60% of the cytosine in CpG context is methylated for 1 round MERLOT amplified gDNA (Figure 3.2.5). This indicates a methyl-transfer efficiency of 95.1% during DNMT1 incubation.

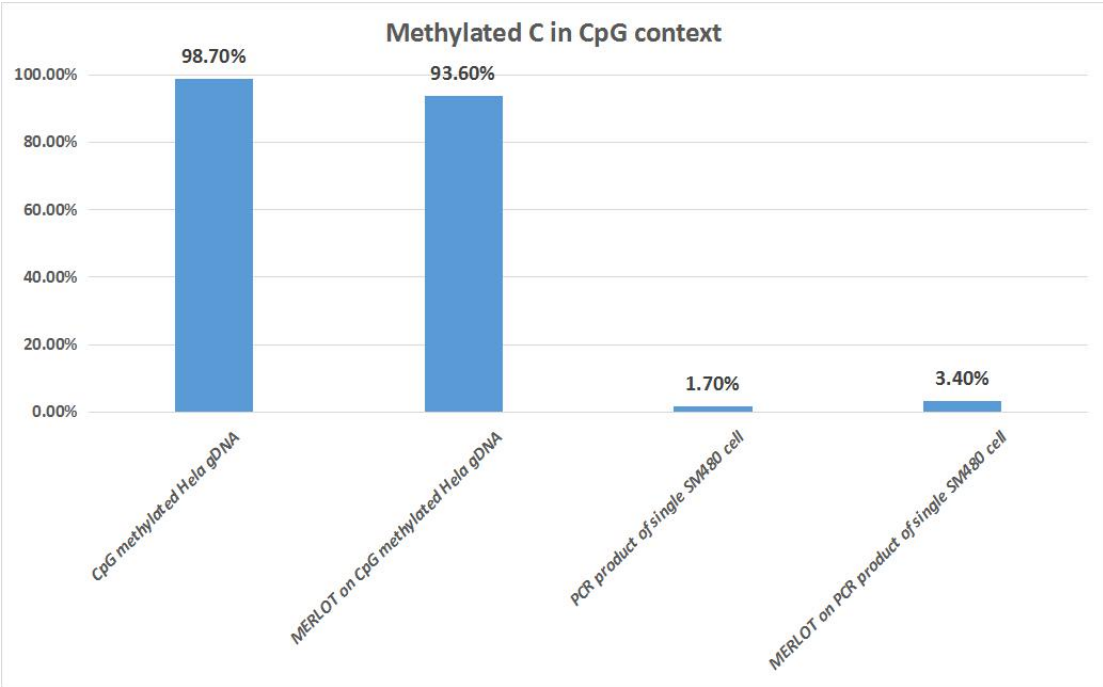

Figure 3.2.5 Percentage of methylated cytosine of bisulfite sequencing results.
Percentage of methylated cytosine is calculated from all uniquely aligned reads of bisulfite sequencing results. Bisulfite sequencing is performed on MERLOT treated gDNA at single-cell level.

To infer the de-novo methylation rate of DNMT1, bisulfite sequencing was performed on 10 pgs of PCR libraries constructed from a single SM480 cell. The library went through 16 rounds of PCR to ensure that all the methylation modifications are erased. The bisulfite sequencing results indicate a de-novo methylation rate of 1.7% across the whole genome.

3.2.4 MERLOT-TAPS on Single Cells

Three GM12787 single cells were picked by mouth-pipette into lysis buffer supplied with protease Q in low-bind 200ul PCR tubes. Tn5 insertion with META schematic is performed on individual cells with peak centered at 250bp. 1 round of MERLOT is applied directly to the Tn5 inserted product with purification. sc-TAPS was applied to all three single cells for 13 PCR

cycles with unique barcodes. The final PCR libraries were pooled together and size-selected with 1.2x AMpure beads. The final library product was checked by qubit and FA before submitting to Illumina X10. All libraries were sequenced 2x 150bp. In total, 31 G of data was received.

Applying MERLOT-TAPS to single cells results in an average of 37.3% methylome coverage with 3x sequencing depth. After analysis of fully methylated and unmethylated lambda DNA spike-in, an average of 86.3% methyl-conversion rate and 1.4% de novo methylation rate was achieved.

In conclusion, MERLOT followed by sc-TAPS gives by far the highest coverage of all mammalian single-cell whole methylome sequencing techniques known thus far. scMERLOT-TAPS on GM12878 single cells results in 37.3% genome coverage with one round of MERLOT. Although the methyl-conversion rate is reduced due to imperfect methyl-transferase efficiency of DNMT1, further bioinformatic analysis can be applied for correction (data not shown). More rounds of MERLOT may bring even higher genome coverage, but remain to be tested.

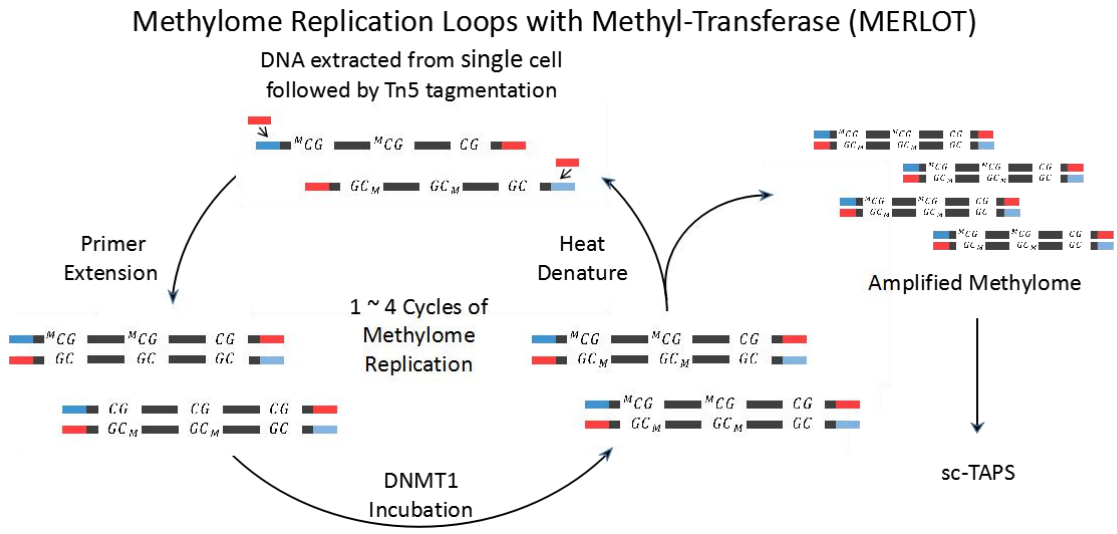

Figure 3.2.6 Schematic of MERLOT-TAPS on single cells

Multiple rounds of MERLOT subjected by sc-TAPS can be applied to Tn5 inserted single cell gDNA.

3.3 Material and Methods

3.3.1 Cla1 Methylation-Sensitive Restriction Assay

The DNA sequence of 87bp Cla1 dsDNA template is:

Upper Strand: 5'-ACC TGT GAC TGA GAC ATC TGA AGG TGC AAT CAG GTG TCA GTC TTA AAG GAT CGA TAA GGA AGC GGA AGT AGT GGT CTC GTC GTA GTG-3'

Lower Strand:5'-CAC TAC GAC GAG ACC ACT ACT TCC GCT TCC TTA TCG ATC CTT TAA GAC TGA CAC CTG ATT GCA CCT TCA GAT GTC TCA GTC ACA GGT-3'.

dsDNA were made by annealing two ssDNA in 1x DNA duplex buffer (IDT) with the following thermo cycle:

95 C for 5 mins

- 0.1C/s till 20 C

20 C forever

Cla1 restriction enzyme cutting is performed in 1x Cutsmart buffer (NEB) at 37 C for 3 hours. The final product was purified by ZYMO column at a 1:5 ratio binding, and eluted in 6 ul elution buffer. 1ul of elution was loaded to high sensitivity DNA bioanalyzer.

3.3.2 MERLOT on Cla1 Template

The methyl-transfer reaction buffer condition is: 1x MT buffer supplied with 0.15ul 160uM SAM, 0.15ul 100ug/ml BSA, 0.3ul 200mM EDTA. 1x buffer condition for MT Buffer includes 20mM Tris-HCL, 1mM DTT, 5% Glycerol, pH 7.8 at 25 °C.

MERLOT polymerase extension performed on Cla1 template uses 0.1ul 2U/ul DeepVent exo- in a 10ul reaction volume. The buffer condition is 1x MERLOT PCR Buffer supplied with 0.2ul dNTP and 0.2ul 100mM MgSO4. The buffer switching is achieved by chelating Mg2+ with EDTA. 1x MERLOT PCR Buffer includes 20mM Tris-HCL, 1mM DTT, 5% Glycerol, pH 7.8 at 25 °C. The thermal cycle for polymerase extension of 87bp dsDNA template is 94 °C for 2 minutes, 58 °C for 60 secs, and 72 °C for 3 minutes. The forward primer is 5'-ACC TGT GAC TGA GAC ATC TG-3'. The reverse primer is 5'-CAC TAC GAC GAG ACC ACT AC-3'.

3.3.3 MERLOT on gDNA at Single-Cell Level or Single Cells

Single-cell Lysis

A single cell is sorted by FACs or mouth pipet into 2.5ul Lysis buffer. The lysis buffer contains: 1.825ul H2O, 0.05ul 1M TE buffer pH 8.0, 0.05ul 1M KCL, 0.375ul 0.1M DTT,

0.075ul 10% Triton X-100, 0.125 20mg/ml Protease Q (Qiagen). The lysis reaction happens in the following thermo-cycle: 50 °C for 20mins, 75 °C for 20 mins, 80 °C for 5 mins. After lysis, dsDNA is released from the single cell.

Tn5 Tagmentation

A Tn5 transposon complex is prepared for fragmentation as follows. Mix 1ul of purified 5uM Tn5 Protein with 1ul 5uM Tn5 dsDNA. Incubate at 25 °C for 45 mins. Add 98ul of Tn5 Dilution Buffer to the 2ul Transposon mix to achieve 0.05uM Tn5 complex. The Tn5 Dilution Buffer includes 10ul 1M TE Buffer pH 8.0, 4ul 0.5M NaCl, and 84ul H2O. The Tn5 dsDNA includes upper strand 5′-CAT TAC GAG CGA GAT GTG TAT AAG AGA CAG-3′ and lower strand 5′-Phos-CTG TCT CTT ATA CAC ATC invdT-3′. To the 2.5ul cell lysis, add 1.5ul of 0.05uM Tn5 transposon complex, 1ul of 5x Tn5 Insertion Buffer. 1x Buffer Condition for Tn5 Insertion Buffer includes 10mM Tris-HCl, 5mM MgCl2 at pH 7.8 at 25 °C. Incubate the 5ul reaction at 50 °C for 10 mins. Add 1ul 1mg/ul ProteaseQ (Qiagen) to the reaction and incubate at the following thermo-cycle: 50 °C for 20 mins and 70 °C for 30 mins. The resulting dsDNA should be 500bp – 1000bp in length with 30bp DNA priming sites added on both ends leaving a 9bp gap on the 3′ end.

Gap repair and 1st Round of PCR reaction

To fill in the 9bp gap, Deep vent exo- polymerase with strand displacement activity is used to repair the gap. After gap repair, the DNA fragments are heat denatured. The resulting ssDNA with complimentary ends form a stem-loop structure. To amplify the stem-loop structures, a single primer PCR reaction with step-down annealing temperature is performed. To the 6ul

reaction, add 1ul of 10x PCR Buffer, 0.2ul dNTP, 0.1ul 2U/ul DeepVent exo-, 0.2ul 1uM 30bp ssDNA primer having the sequence 5′-CAT TAC GAG CGA GAT GTG TAT AAG AGA CAG-3′, 0.2ul 100mM MgSO4, and 2.3ul H2O. 1x PCR Buffer condition includes 20mM Tris-HCL, 10mM KCL, 0.1% Triton X-100, pH 7.8 at 25 °C. The following thermo-cycle is performed on the 10ul reaction: 72 °C for 10 minutes, 95 °C for 3 minutes, 68 °C for 60 secs, 67 °C for 60 secs, 66 °C for 60 secs, 65 °C for 60 secs, 64 °C for 60 secs, 63 °C for 60 secs, 62 °C for 60 secs, 61 °C for 60 secs, 60 °C for 60 secs, 59 °C for 60 secs, 58 °C for 60 secs, and 72 °C for 3 minutes. This results in hemi-methylated dsDNA fragments.

1st round of DNMT1 methyl-transfer reaction

The resulting extension products are then incubated with DNMT. EDTA is added to chelate Mg2+. To the 10ul reaction, add 1.5ul of 10x Methyl-Transfer (MT) Buffer, 0.15ul 160uM SAM, 0.15ul 100ug/ml BSA, 0.3ul 200mM EDTA, 2ul 2U/ul DNMT1, 0.9ul H20. 1x buffer condition for MERLOT MT Buffer includes 20mM Tris-HCL, 1mM DTT, 5% Glycerol, pH 7.8 at 25 °C. The 15ul reaction is incubated at 37 °C for 3 hours. This results in methylated dsDNA fragments and a complete single round of amplification, i.e. single primer extension, and methylation.

Multiple rounds of amplification and methylation on demand

A further 1 to 4 rounds of amplification and methylation can be performed based on demand. The thermo-cycle is the same as 1st round of amplification and methylation. The following reagents should be added for the 2nd, 3rd, 4th round of amplification and methylation respectively.

2nd round of amplification, PCR

To the 15ul reaction, add 1ul of 10x PCR Buffer, 0.2ul dNTP, 0.1ul 2U/ul DeepVent exo$^-$, 0.2ul 1uM 30bp ssDNA primer, 0.55ul 100mM $MgSO_4$, and 2.95ul H_2O.

2nd round of DNMT1 Methyl-transfer reaction

To the 20ul reaction, add 1ul of 10x Methyl-Transfer (MT) Buffer, 0.25ul 160uM SAM, 0.1ul 100ug/ml BSA, 0.325ul 100mM EDTA, 2ul 2U/ul DNMT1, 0.15ul 100mM DTT, and 1.175ul H_20.

3nd round of amplification, PCR

To the 25ul reaction, add 1ul of 10x PCR Buffer, 0.2ul dNTP, 0.1ul 2U/ul DeepVent exo$^-$, 0.2ul 1uM 30bp ssDNA primer, 0.85ul 100mM $MgSO_4$, and 2.65ul H_2O.

3nd round of DNMT1 Methyl-transfer reaction

To the 30ul reaction, add 1ul of 10x MERLOT Methyl-Transfer (MT) Buffer, 0.35ul 160uM SAM, 0.1ul 100ug/ml BSA, 0.475ul 200mM EDTA, 2ul 2U/ul DNMT1, 0.25ul 100mM DTT, and 0.825ul H_20.

4nd round of PCR

To the 35ul reaction, add 1ul of 10x PCR Buffer, 0.2ul dNTP, 0.1ul 2U/ul DeepVent exo$^-$, 0.2ul 1uM 30bp ssDNA primer, 1.15ul 100mM $MgSO_4$, and 2.35ul H_2O.

4nd round of DNMT1 Methyl-transfer reaction

To the 40ul reaction, add 1ul of 10x Methyl-Transfer (MT) Buffer, 0.45ul 160uM SAM, 0.1ul 100ug/ml BSA, 0.625ul 200mM EDTA, 2ul 2U/ul DNMT1, 0.35ul 100mM DTT, and 0.475ul H_20.

5nd round of PCR

To the 45ul reaction, add 1ul of 10x PCR Buffer, 0.2ul dNTP, 0.1ul 2U/ul DeepVent exo$^-$, 0.2ul 1uM 30bp ssDNA primer, 1.45ul 100mM $MgSO_4$, and 2.05ul H_2O.

5nd round of DNMT1 Methyl-transfer reaction

To the 50ul reaction, add 1ul of 10x Methyl-Transfer (MT) Buffer, 0.55ul 160uM SAM, 0.1ul 100ug/ml BSA, 0.775ul 100mM EDTA, 2ul 2U/ul DNMT1, 0.45ul 100mM DTT, and 0.125ul H_20.

The amplified dsDNA that has been fully methylated can be purified using ZYmo column and elute in H2O. The final product can be subjected to sc-TAPS for methylation sequencing.

3.4 Reference

1. Zhang G, Estève PO, Chin HG, et al. Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation. *Nucleic Acids Res*. 2015;43(12):6112–6124.

2. Wang KY, Chen CC, Tsai SF, Shen CJ. Epigenetic Enhancement of the Post-replicative DNA Mismatch Repair of Mammalian Genomes by a Hemi-mCpG-Np95-Dnmt1 Axis. *Sci Rep*. 2016;6:37490.

3. Ma HS, Wang EL, Xu WF, et al. Overexpression of DNA (Cytosine-5)-Methyltransferase 1 (DNMT1) And DNA (Cytosine-5)-Methyltransferase 3A (DNMT3A) Is Associated with Aggressive Behavior and Hypermethylation of Tumor Suppressor Genes in Human Pituitary Adenomas. *Med Sci Monit*. 2018;24:4841–4850.

4. Wu F, Wu Q, Li D, et al. Oct4 regulates DNA methyltransferase 1 transcription by direct binding of the regulatory element. *Cell Mol Biol Lett*. 2018;23:39.

5. Qin W, Wolf P, Liu N, et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res*. 2015;25(8):911–929.

6. Elliott EN, Sheaffer KL, Kaestner KH. The 'de novo' DNA methyltransferase Dnmt3b compensates the Dnmt1-deficient intestinal epithelium. *Elife*. 2016;5:e12975.

7. Bewick AJ, Sanchez Z, Mckinney EC, Moore AJ, Moore PJ, Schmitz RJ. Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, Oncopeltus fasciatus. *Epigenetics Chromatin*. 2019;12(1):6.

8. Yarychkivska O, Tavana O, Gu W, Bestor TH. Independent functions of DNMT1 and USP7 at replication foci. *Epigenetics Chromatin*. 2018;11(1):9.

9. Peters SL, Hlady RA, Opavska J, et al. Essential role for Dnmt1 in the prevention and maintenance of MYC-induced T-cell lymphomas. *Mol Cell Biol*. 2013;33(21):4321–4333.

10. Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J Biol Chem*. 2014;289(7):4106–4115.

11. Berkyurek AC, Suetake I, Arita K, et al. The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J Biol Chem*. 2014;289(1):379–386.

12. Londoño Gentile T, Lu C, Lodato PM, et al. DNMT1 is regulated by ATP-citrate lyase and maintains methylation patterns during adipocyte differentiation. *Mol Cell Biol*. 2013;33(19):3864–3878