**VILNIUS UNIVERSITY**

**LIFE SCIENCES CENTRE**

ZAINAB SHOUR

# Synthetic Modification of Mouse Genome Fragments

**Master's Thesis**

Genetics Study Programme

**Supervisor**

PhD, Migle Tomkuviene

Vilnius 2024

Table of Contents

# LIST OF ABBREVIATIONS

5caC – 5-carboxylcytosine

5fC – 5-formylcytosine

hmC – 5-hydroxymethylcytosine

5hmC – 5-hydroxymethylcytosine

mC – 5-methylcytosine

5mC – 5-methylcytosine

6mA – 6-methyladenine

oct – Octadiyn- modification group

by – Butynyl- modification group

hy – Hexynylazide modification group

hyC – 5-(6-azidohex-2-ynyl)cytosine

octC – 5-octadiynyllcytosine

byC – butynylcytosine

MTase – DNA methyltransferase

SAM – S-adenosyl-L-methionine

MBD – methyl-CpG-binding domain

VILNIUS UNIVERSITY
LIFE SCIENCES CENTER

Zainab Shour

# Synthetic Modification of Mouse Genome Fragments

Master Thesis

## SANTRAUKA

DNR yra aprūpinta įvairiomis epigenetinėmis modifikacijomis, kurios nulemia genų raiškos kaitą vystymosi metu ir vėliau. Norint papildyti dabartines žinias apie tai, kaip genomo metilinimas nustatomas ir palaikomas vystymosi, senėjimo ir ligų metu, būtina stebėti metiltransferazės aktyvumą. Vienas iš būdų tai padaryti yra citozino bazių modifikavimas naudojant inžinerintas MTazes, gebanbčias pernešti sintetines grupes nuo atitinkamų sintetinių kofaktorių ant DNR. Šios sintetinės modifikacijos tada gali būti detektuojamos naudojant nanoporų sekoskaitos metodą.

Modifikacijų detekcijos patikrai buvo atrinkti pelių genai dalyvaujantys pelių neurogenezėje embriono vystymosi metu. Šio darbo buvo siekiama įvesti sintetines modifikacijas į pelių genų fragmentus. Tam atlikti buvo užsibrėžti trys tikslai: optimizuoti pelės genų fragmentų PGR sąlygas; atrinkti PGR produktus pagal išeigą ir kokybę ir įvesti sintetines bei natūralias modifikacijas; paruošti PGR produktus nanoporų sekoskaitai ir modifikacijų detekcijai.

PGR produktų išeiga ir specifiškumas buvo geriausi amplifikuojant pasirinktas trumpesnes genų promotorių sritis. Optimizuotų reakcijų metu PGR produktai buvo įsotinti natūraliomis (mC ir hmC) arba sintetinėmis modifikacijomis arba abejomis kartu. Galiausiai, sintetinės modifikacijos buvo tikrinamos nanoporų sekoskaitos metodu įvairiomis sąlygomis. Pastebėta, kad siekiant efektyviau nustatyti sintetines DNR modifikacijas ir jas atskirti nuo natūralių, reikia papildomai optimizuoti modifikavimo reakcijas ir patobulinti sintetines DNR modifikacijas nustatantį nanoporų modelį.

VILNIUS UNIVERSITY
LIFE SCIENCES CENTER

Zainab Shour

# Synthetic Modification of Mouse Genome Fragments

Master Thesis

SUMMARY

DNA is furnished with a variety of epigenetic modifications conferring an alternation in gene expression during development and beyond. To enhance and supplement current knowledge on how genomic methylation is established and maintained during development, senescence, and disease, it is essential to track methyltransferase activity. Therefore, one method to employ is the modification of Cytosine bases using engineered MTases capable of attaining synthetic groups from their corresponding synthetic cofactors. Further, detection of said modifications by nanopore sequencing would be advantageous for more extensive, higher-throughput research.

Therefore, this project aimed at synthetically modifying murine gene fragments to train and test nanopore detection of synthetic cytosine modifications. The genes selected are of murine origin with some implications in mouse neurogenesis during embryonic development. Three objectives were set in place to be met. Firstly, the optimization of PCR conditions for the gene fragments. Secondly, modifying the selected PCR products, having the highest quality and yield, with synthetic and natural modifications. And finally, nanopore sequencing of selected modified fragments with direct modification detection.

Results revealed high and mostly specific PCR product yields when amplifying the selected gene promoter regions. With regards to optimizing the synthetic modifications, it was achieved to produce selected fragments with either natural (mC and hmC) or synthetic modifications, and all in combination. Finally, detection of the synthetic modifications was validated upon nanopore sequencing with varying degrees of base-calling confidence and reads coverage. We drew from these results that further optimizations would be required for the PCR and modification reactions, in addition to further training of the nanopore sequencing algorithm for more consolidated identification of synthetic modifications.

# INTRODUCTION

DNA sequences carry information for the development, growth, and functioning of organisms. The information is dictated by the nucleotide sequence inherited through reproduction. In addition to the sequence, epigenetic modifications also dictate the expression of this information (Tate & Bird, 1993). While different modifications encompass DNA, RNA, and Proteins across all life forms, the most widely explored epigenetic DNA modification in mammals is 5-methylcytosine (5mC) . For such a modification to be generated, cofactor S-Adenosyl-L-methionine (SAM) donates -$CH_3$ group to a DNA Methyltransferase (MTase), which in turn deposits the group on the fifth position of cytosine ring in a specific DNA sequence. Generally, methylation in mammals is predominantly occurrent at CpG dinucleotides of somatic cells (Menezo et al., 2020).

Put simply, enzyme families upholding their responsibility in DNA methylation include the writers, for establishing methylation; erasers, for removing methylation; and readers, for recognizing methylation and ultimately influencing gene expression (Moore et al., 2013). Ultimately, the role of DNA methylation has established itself as vital for embryonic mammalian development. Such a function is due to the change in gene expression conferred by the modifications. Some of the most highlighted roles of mC include its general association with repression of transposons and genes when located in promoter regions, while it's associated with active transcription when found within gene bodies (Hellman & Chess, 2007; Lorincz et al., 2004). Their prevalence hinders the binding of transcription factors and hence the transcription of the respective gene body, thus silencing genes in a cell and tissue-type specific manner (S.-M. Yang et al., 2013).

There are various methods that could be employed in order to unravel the methylation pattern in regions of interest in elucidating the mechanisms of gene expression. While each method has its own utility depending on the objective to be achieved, the methods would fall short in addressing the minute contributions of MTase modifications due to the overlapping interplay amongst them (Okano et al., 1999).

Therefore, advancements Next-Generation Sequencing (NGS) have shown promising employability in the detection of mC or 5-hydroxymethylcytosine (hmC) (Tse et al., 2021). However, natural modifications are known to be poor reporters and are difficult to further amend with reporters (Tomkuvienė et al., 2019). Therefore, synthetic modifications are a prospective solution at harnessing the biotechnological potential of MTase activity tracking (Stankevičius et al., 2022). Therefore,

engineered MTases with compatible synthetic cofactors are harnessed for the deposition of synthetic groups onto Cytosine. Other than the applicability of such synthetic groups in being further derivatized with various techniques of marked tracking *in vivo,* they also have the potential of being directly identified by third-generation sequencing, such as nanopore (Barros-Silva et al., 2018).

Therefore, this project aimed at synthetically modifying murine gene fragments to train and test nanopore detection of synthetic cytosine modifications. The genes selected have implications in mouse neurogenesis during embryonic development (*FAT1*; *Sfi1*; *H1fnt*; *TMEM267*; *Gm26917).*

# AIMS AND OBJECTIVES

Project aim:

To prepare PCR-amplified mouse genome fragments with natural and synthetic modifications for training and testing the nanopore sequencing with direct modification detection.

Project objectives:

1- Optimize the reaction conditions for PCR amplification of murine gene fragments: *FAT1*; *Sfi1*; *H1fnt*; *TMEM267*; *Gm26917*.
2- Utilize the optimized PCR products for the optimization of several modification reactions, including: Butynyl-, Octadiyn-, Hexynazide-, Hydroxymethylation and Methylation.
3- Validate the direct detection of modifications by library preparation and nanopore sequencing of the prepared modified samples.

# 1. LITERATURE REVIEW

DNA modifications are one of the main epigenetic mechanisms which regulate gene expression in both plants and animals. DNA methylation indicates the attachment of a methyl (-$CH_3$) group at, most often, 5'-carbon of the pyrimidine ring(Boyes & Bird, 1991). In addition to methylation of cytosine residue, other modifications such as oxidation of methylated cytosine (mC) to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (5-fC), 5-carboxylcytosine (5-caC), and methylation of adenine (A) to N6-methyladenine (6-mA), have been identified as participants of epigenetic pathways.(Montero et al., 1992)

## 1.1. Methylation and other DNA modifications

Methylation of the cytosine base plays such a vital role as the most abundantly modified base of the genome to an extent of being deemed the fifth base (Montero et al., 1992). Since the covalent attachment of methyl on the cytosine of CpG sites is the most abundant DNA modification in the vertebrate genome, it plays a pivotal role in directing development and gene regulation. Notable exceptions to this pervasive genomic methylation are gene regulatory regions such as CpG-rich promoter elements, also known as CpG islands, and active enhancers that are often associated with little or no DNA methylation.

Given that epigenetic regulation is a complex mechanism, it is essential for cells to have numerous machinery involved in the modulation and regulation of said modifications; the enzymes in question are categorized into: "writers", "readers", and "erasers" (Nicholson et al., 2015).

### Writers

DNA methylation is mediated by the DNA methyltransferases (DNMT) family, primarily focusing DNMT1, DNMT3A, and DNMT3B. DNMT1 is deemed as a maintenance methyltransferase which faithfully propagates methylation patterns across DNA during replication. Its fidelity relies on copying said pattern from the parent strand to the newly synthesized daughter strand; hence, exhibiting a strong affiliation to hemi-methylated DNA (Nicholson et al., 2015).

On the other hand, DNMT3A and DNMT3B are deemed as *de novo* DNA MTases. They are responsible for the establishment of DNA methylation patterns anew. While those MTases were believed to have redundant functionalities, researchers proved otherwise. Knockout of either one of the corresponding genes led to embryonic lethality, establishing the vital role they separately play in embryogenesis (Liao et al., 2015). It has been revealed that loss of DNMT3B in establishing *de novo* methylation leads to an impairment in placental labyrinth and eventually mid-gestation embryonic

lethality (Andrews et al., 2023). However, DNMT3A differentiates its functionality since its knockout leads to embryonic lethality at different stages of development than the aforementioned counterpart. However, when capitalizing on the vitality of DNMT1 functioning in proper regulation of DNA methylation at imprinted genes, it has been revealed that partial mutations or severe loss of function in the MTase in question has an inevitable influence on abnormal development, even to the extent of embryonic lethality (Zwier et al., 2012).

Readers

The recognition of epigenetic modifications is mediated by the "readers." Since DNA methylation plays a role in gene expression (to be discussed further on), reader proteins mediate this expressional change by recognizing mC sites. This recognition is conferred by methyl-CpG-binding proteins (MBPs) which initiate the silencing of the region in question by recruiting factors involved in chromatin compaction and directly blocking transcription factor binding. It has recently been revealed that there are proteins that can recognize other forms of modifications, hmC and 5fC. Methyl-CpG-binding domain protein 3 (MBD3), Methyl-CpG-binding domain protein 2 (MeCP2) have been reported to preferentially bind to hmC as compared to 5mC.(Frauer et al., 2011; Mellén et al., 2012; Yildirim et al., 2011)

Erasers

While DNA methylation was once thought to be semi-permanent, due to dilution during DNA replication, quite the opposite was revealed with the "erasers" (Nicholson et al., 2015). Erasers' functionality encompasses modifying the epigenome by yielding other cytosine modifications, influencing gene expression, and active removal of epigenetic marks by the ten-eleven translocation (TET) family of proteins. (A. Y. Yang et al., 2016)

TET proteins (TET1, TET2, and TET3) are able to oxidize mC to hmC, as well as further oxidation producing, 5fC and 5caC (Figure 1.1). TET1 is involved in the erasure of imprinting marks in primordial germ cells . While, TET2 has pleiotropic roles during hematopoiesis through, at least in part, modulation of DNA methylation. TET3, meanwhile, is essential for the active erasure of paternal DNA methylation marks in the zygote, an important reprogramming event during preimplantation development. (Rasmussen & Helin, 2016) Generally, TET enzymes have shown to be vital for the completion of embryonic genome activation.(Arand et al., 2021)

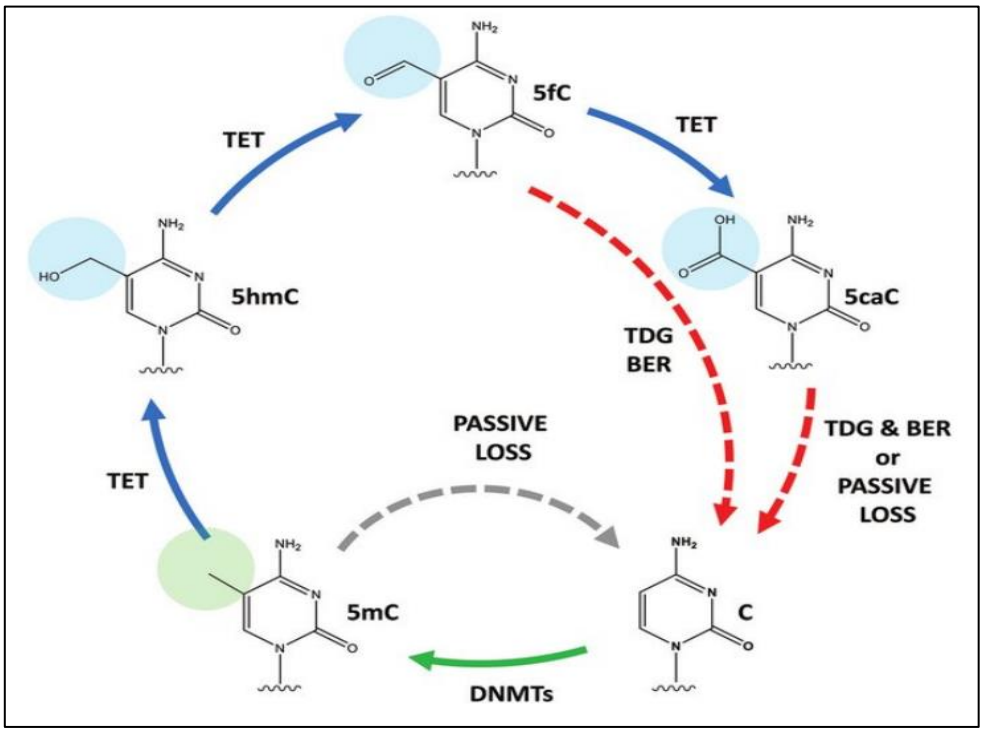


**Figure 1.1. Methylation-demethylation cycle cytosine in DNA.** *(Ravichandran et al., 2017)*

The successive oxidation of mC, while being on one part responsible for the production of unmodified cytosine, implicates yielding intermediate products of epigenetic modifications.

TET, BER, and MTases in DNA Modification Formation

As aforementioned, CpG methylation is dictated by DNMTs utilizing S-adenosyl-L-methionine (SAM) as a cofactor. Reiterating that there are two pathways highlighted with regards to the production of CpG methylation patterns: (1) De Novo Establishment; and (2) Maintenance.

De novo establishment of the methylation pattern will be highlighted to be dependent upon the DNMT3 family of enzymes, as aforementioned. DNMT3A and DNMT3B function together in developing the initial pattern of methylation during the earliest stages of embryogenesis. However, while they are of somewhat similar structure, they have distinct preferentiality of function. DNMT3A is critical for methylation of single-copy genes and/or regions undefined by “long stretches of CpGs” to be methylated and is believed to be ubiquitously expressed. On the other hand, DNMT3B is not as ubiquitously expressed as its counterpart in differentiated cells, but rather targets highly repetitive and

long stretches of CpGs; thus, believed to play a more vital role of *de novo* methylation during embryogenesis. One additional enzyme belonging to the family which is believed to regulate *de novo* methylation during earlier phases of development is DNMT3L; in X chromosome inactivation, as well as imprinted and retrotransposon loci. However, it's worthy to mention that DNMT3L is not catalytically active in establishing *de novo* methylation itself but rather regulates the mechanism by interacting with both DNMT3A and DNMT3B, stimulating their enzymatic activity (Veland et al., 2019).

However, the maintenance of the methylation pattern is more dependent upon DNMT1. During DNA replication, the newly synthesized daughter strand has to maintain the methylation pattern as the parent strand in order to consolidate the phenotypic and biological expression. Through facilitated cooperation with Ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1) protein that recognizes hemi-methylated CpGs, DNMT1 preferentially binds to the region in question and restores the methylation pattern on the newly synthesized daughter strand (Newkirk & An, 2020).

Nevertheless, demethylation can take on two trajectories. The first entails passive demethylation of DNA during replication. In this case, DNA methylation is lost on the DNA strand opposite to that which is hemi-methylated, thus inducing a loss of methylation. However, when considering the second possible trajectory, it entails the active demethylation by the TET/TDG(thymine–DNA–glycosylase)/BER (base excision repair)-dependent pathway. mC could be subject to oxidation by TET to produce, as aforementioned, hmC, 5fC, and 5caC. TDG component of the BER pathway excises 5fC and 5caC, thereby yielding an abasic site, it's repaired by BER, and restoring the unmodified form of cytosine. (Drohat & Coey, 2016; Wu & Zhang, 2014)

A noteworthy phenomenon entails the spontaneous hydrolytic deamination of mC into Thymine (T). This results in a base mismatch that could be repaired by redeeming the original C or substituting the Guanine (G) across from T into Adenine (A). Thus, inducing a point mutation from C:G to T:A. (Cooper et al., 2010; Yebra & Bhagwat, 1995).

### 1.2. Role of DNA Methylation in Gene Expression

Epigenetic modifications have proven to not only enable cells to inherit the genetic code, but also in further altering the expression of the encoded information (Gibney & Nolan, 2010). Generally, epigenetic modifications induce changes in gene activity, whether it be expression or inhibition. While gene expression entails a delicate multi-level meshwork of complex processes, a specific set of

mechanisms as proceeded with below. There are generally three mechanisms most postulated on the influence of methylation in gene expression:

1- DNA methylation of promoter regions affecting Transcription Factor (TF) binding
2- Recruitment of Methyl-CpG binding proteins and remodeling chromatin
3- DNA methylation within gene body impacting transcription.

Regulation of Gene Expression - DNA methylation of promoter regions affecting TF binding

It is believed that around 70% of the promoters of the human genome contain CpG islands (Saxonov et al., 2006). The postulation is that methylation of the islands prevents the binding of the transcription factors onto gene promoters, inducing a transcriptional repression, and thus, silencing gene expression. It's worth mentioning that CpG methylation could sustain gene repression, even beyond the binding site of the transcriptional factors (Tate & Bird, 1993).

However, the disruption of transcriptional factor binding by methylated CpGs isn't the only method of gene expression control. It is suggested that DNA methylation by exploring the mechanism of action between DNMTs and transcription factors. An interaction between the aforementioned parties is believed to induce a site-specific methylation at promoter regions. This site-specific methylation would either further influence the action of transcriptional machinery or alter the chromatin structure, essentially influencing gene expression either way (Hervouet et al., 2010).

Regulation of chromatin structure by recruitment of methyl CpG binding proteins

Another mechanism in which DNA methylation regulates the repression of gene expression is through the recruitment of MeCPs. One of the first identified of the proteins is MeCP1 which is able to mediate the repression through directing the deacetylation of histones to areas of methylated DNA (Boyes & Bird, 1991). On the other hand, MeCP2 preferentially binds mCpG and through an interaction, recruits chromatin-modifying complex mSin3A and histone deacetylase, thus, uniting their repressive functions on gene expression (Nan et al., 1998). Further identification of four additional proteins with the same methyl-CpG-binding domain (MBD) led to the implication of those protein in playing a critical role of mediating the interaction between DNA methylation, histone modifications, and chromatin restructuring (Cross et al., 1997; Hendrich & Bird, 1998).

In eukaryotic cells, DNA and histone proteins form chromatin, and it is in this context that transcription takes place. The core histones, being highly conserved basic proteins with globular domains

around which the DNA is wrapped with relatively unstructured flexible ‘tails’ that protrude from the nucleosome. The tails are subject to a variety of post-translational modifications (PTMs). PTMs directly affect the structure of chromatin, regulating its higher order conformation and thus acting in regulating transcription, in a cis manner/effect. Moreover, attract certain effector proteins to the chromatin in a trans manner/effect. Therefore, the chromatin structure can modulate gene expression at a distance through looping, nuclear matrix association and nucleosome positioning (Skinner, 2011).

DNA Methylation Within Gene Body Impacting Transcription

Generally, DNA methylation is associated with gene transcriptional repression. However, this is believed to more likely be true in promoter methylated regions (Bommarito & Fry, 2019). While there is some evidence to suggest that intragenic methylation in also associated with transcriptional repression (Lorincz et al., 2004), there is also some counter evidence that it’s associated with gene activation. An example of the former is considering X-inactivation. Both inactive and active X chromosomes have dissimilar methylation patterns. However, looking deeper into the said patterns, inactivated X chromosome is characterized by promoter hypermethylation, while activated X chromosome is characterized by dense methylation within its gene bodies (Hellman & Chess, 2007). Implications of an increase in gene expression given gene-body methylation isn’t restricted to X chromosomal activation but also within a range of cell and tissue types (Aran et al., 2011; Jjingo et al., 2012; Pfeifer & Rauch, 2009).

The functions of the various epigenetic marks and factors are distinct. DNA methylation has a role in early development to help establish early cell lineages (e.g., stem cells) and can regulate the activity of promoters and general genome regions (e.g., repeat elements) (Kazazian, 2004).

### 1.3. Role of DNA methylation in mouse development

Considering the implications of methylation in the activation and repression of genes, its implications become apparent with regards to mammalian embryonic development. The emanation of tissue heterogeneity of a developing organism is dependent upon amenable methylation patterns during development. Some of the most widely explored mechanisms of DNA-methylation-based silencing include X-inactivation, imprinted genes, and germline specific genes (Hellman & Chess, 2007).

Genomic Imprinting

In the two developing germlines, genomic imprinting will confer the establishment of differential methylation patterns withstanding the proceeding reprogramming that takes place during early embryogenesis (further explained) (Bajrami & Spiroski, 2016). There are around 20 genomic regions known as imprinting control regions (ICRs) in the mouse and human; since the said regions withstand reprogramming, they enforce a monoallelic expression(Arnaud, 2010).

Embryogenesis

However, during embryogenesis, the DNA of cells in a developing organism undergo "several waves of DNA methylation and demethylation". The first demethylation event occurring after fertilization, within the maternal genome, is implicated with a passive, replication-dependent manner. However, the paternal genome would be subject to a rapid and active demethylation process prior to DNA replication and cellular division (Mayer et al., 2000).

Neurogenesis

It has been revealed that DNA methylation in neurons occurs within regions of low CpG densities rather than CpG islands; contrary to revelations in exploring genomic stability, genomic imprinting, and X-inactivation (Guo et al., 2011). In the central nervous system of the embryonic mouse, *Dnmt1*, the counterpart of DNMT1 found in humans, is ubiquitously expressed in both proliferating neural progenitor cells and differentiated neurons (Goto et al., 1994).

A mutation or abnormality of functioning any of the mice *Dnmts* could lead to serious developmental abnormalities and/or embryonic/postnatal lethality (Li et al., 1992; Okano et al., 1999). Specifically, a knockout or deletion of *Dnmt1* could lead to a degree of hypomethylation that confers a non-repressed gene, leading to premature glial differentiation (G. Fan et al., 2005). With regards to *Dnmt3a*, mice experiencing a deletion of the said gene would survive birth but suffer from postnatal neurodevelopmental impairment (Nguyen et al., 2007).

## 1.4. Methylation Modulation Implication in Genes of Murine Development and Neurogenesis

*FAT1*

*FAT1* atypical cadherin is a member of the cadherin superfamily. It's one of four members whose role was explored in mouse embryonic neurogenesis. It plays a crucial part in the development of the

vertebrate neuroepithelium, while its knock-out in mice led to the manifestation of severe neurological phenotypes, highlighting the crucial role of its expression (Ahmed, 2015).

*Sfi1*

*Sfi1* centrin binding protein gene's expressional change had led to postulation on its involvement in neurodevelopmental disorders of mice, including copy-number variation in autism disorder (Segurado et al., 2005); it is known to be upregulated in the young brain while downregulated in the older (Trent et al., 2014). The differential methylation pattern in *Sfi1*'s promoter region influencing neurogenesis, revealed an implication in its involvement in neuronal development and differentiation (Saenz-de-Juano et al., 2019).

*H1fnt*

*H1fnt* is H1 histone family, member N, testis-specific, acting as an essential gene product required for proper DNA condensation and cell reconstruction during the elongation phase of spermiogenesis (Lin et al., 2000). Previous investigations on H1 linker histones explored their DA regulatory role in the context of methylation. Results revealed, through a triple knockout, that histone H1 acts as a silencer of gene expression by two mechanisms. One of which involves a direct interaction and recruitment of DNMT1 and DNMT3B. Another of which involves the inhibition of SET7/9 (a methyltransferase which could catalyze the methylation of a variety of proteins) binding and methylation of H3K4 (Daks et al., 2022; S.-M. Yang et al., 2013).

Additionally, other investigations took to investigating the elimination of the many subtypes of H1 histone genes. An elimination of a single subtype of H1 genes didn't noticeably hinder mouse development, suggesting an upregulation in the remaining subtypes compensates for a lack thereof. However, the concomitant elimination of three different subtypes exhibited embryonic lethality, suggesting the vital role the H1 family plays in mammalian development (Y. Fan et al., 2003). However, there is a discrepancy on detailing the involvement of the said family in its involvement of regulating neurogenesis of mice (Kishi et al., 2012).

*TMEM267*

*TMEM267* encodes a transmembrane protein prominently known for its expression in the medial ganglionic eminence. It has been revealed that the methylation pattern at the 5' region of the gene plays a powerful role in the maintenance of intestinal homeostasis by controlled expression of Aldh1a1 when

considering epigenetic aberrations and their relatedness to major psychiatric diseases through gut and microbiome alternations (Nohesara et al., 2023).

*Gm26917*

*Gm26917* is a lncRNA believed to be regulated by FoxM1 and acts as a competing endogenous RNA source which accelerates the apoptosis of muscle satellite cells (Chen et al., 2018). However, while FoxM1 has already proven to play a vital role in the regulation of neural progenitor fate during spinal cord regeneration, this does not directly correlate with the inclusivity of *Gm26917* in such a neurogenesis mechanism. Rather, *Gm26917* exhibited to have an interaction network with seven Autism Spectrum Disorder-related genes, encouraging further exploration on its mechanisms to the nervous system (Mizuno et al., 2020).

### 1.5. DNA Methylation detection

Since it's already been mentioned that DNA modifications elicit changes in expression that could precede the development of various pathological conditions, it's vital the modification status of said DNA is profiled to reveal the underlying mechanisms of disease and development.

One of the many implications in unravelling the methylation pattern within an organism is to reveal its involvement in a myriad of biological processes. Some of which include fear-related memory explored in rats, revealing that the behavior was dependent on *de novo* methylation (Bali et al., 2011). Another of which pertains the discovery of 353 epigenetic markers (CpG methylation) related to the biological clock and its role in tissue aging (Horvath, 2013). Moreover, it has already been marked how epigenetic modifications are flexible under environmental conditions including diet, stress, toxicity, etc. One prominent retrospectively explored example is the winter famine of WWII where regions and populations facing a severity in food deprivation (i.e., nutritional) exhibited a higher susceptibility to diseases such as schizophrenia, obesity, etc (Kim, 2005).

Additionally, methylation has become a targeted mechanism of exploration of disease development in human beings ranging from cancers, autoimmune disorders, neurological disorders, and even disease severity (Aberg et al., 2012). Considering the influence of changes in methylation patterns, it has already been implicated in diseases cancer patients the considerable effect of hypermethylation in promoter regions of tumor suppressor genes inducing the inactivation in numerous tumor suppressor functions (Kim, 2005). Thus, further solidifying the significance in unravelling the modification patterns through a number of methods.

There are various techniques and methods effective enough for the detection of DNA methylation. methods include, but are not restricted to, conversion-based methods, anti-body-based assays, third generation sequencing, and so on. Given the variety of DNA methylation detection, it's safe to say that each have their own advantages and/or disadvantages depending on the suitability on choice of method and the intended purpose of application.

Conversion-Based Methods

With Bisulfite sequencing, Cytosine is converted to uracil from being treated with sodium bisulfite, whereas methylated cytosines will not be affected. However, bisulfite sequencing is known to damage DNA. While bisulfite sequencing is a gold standard of revealing methylations status at a single-base resolution, it's still a rather expensive and crude method as it results in extensive DNA degradation and obstructed genomic mapping of sequencing reads due to the conversion of canonical cytosines into thymine. Taking that into consideration, whole-genome bisulfite sequencing amplifies the discrepancy of read mapping between the original genomic sequence to be profiled and that of the conversion's product. The quantitative assessment of the DNA strand's degree of methylation is measured by determining the ratio of T and C. Thus, providing a simple reaction protocol assuring reproducibility and accurate measurement (Delaney et al., 2015)

Pyrosequencing is another method that depends on the bisulfite conversion of DNA. Once treated and converted, the method would entail the quantitative measurement of NDA methylation levels in a sequencing-by-synthesis method, monitoring the real-time released proportional light signal of the incorporated nucleotide (Delaney et al., 2015).

Another method which depends on the bisulfite conversion of DNA is Methylation-specific (MS)-PCR. The method relies on the said conversion, and then PCR amplification of the DAN by utilizing two sets of independent primers. One pair would be specifically designed to amplify the methylated sequence, while the other would amplify the bisulfite-converted unmethylated sequence. One the amplicons produce the expected sizes, this is indicative of the methylation status of the original DNA sample (Huang et al., 2013).

Antibody-based assays

Moreover, with Microarray hybridization, it was one of the very first applied technologies upscaling the DNA methylation studies onto a genome-wide level. While there are many types of genomic microarrays, the offerings are dependent upon the designed and selected microarray.

Considering, Methylated DNA immunoprecipitation (MeDIP), this this method, the DNA fragments are immunoprecipitated via an antibody (AB) raised against 5-methylcytosine; to improve the binding affinity the DNA strands are subject to denaturation and rendered single-stranded (Jacinto et al., 2008). The methylated DNA bound to the AB complex is separated from the rest of the DNA fragments, leaving the enriched methylated DNA to be purified. This method is relatively known to be an unbiased approach for the detection of methylated DNA (Borgel et al., 2012). However, a limitation of the method is the typical dependency on the enrichment of modified cytosines, which may underrepresent regions of lower modification densities (Mohn et al., 2009). And also is has limited resolution, which is not single-base.

However, Human MethylationEPIC BeadChip is an array ideal for epigenome-wide association studies, various experimental biological applications, and even tumor profiling. While Illumina offers a murine-specific version of the array, Mouse Methylation BeadChip. On the other hand, HpaII tiny fragment enrichment by ligation-mediated PCR (HELP) is an assay that serves to have a more positive representation of hypomethylated regions, overcoming the technical drawback of requiring hypermethylated regions pronouncing detection (Oda & Greally, 2009).

Breaking away from more conventional methods of methylation pattern profiling, a relatively new and alternative method which enables profiling on an epigenome-wide level is Tethered Oligonucleotide-Primed sequencing (TOP-seq). The method relies on the covalent tagging of individual unmodified CpG sites, followed by in situ non-homologous priming of the DNA strand synthesized from a tethered oligonucleotide by a DNA polymerase. This enables the direct production of adjoining regions for their sequencing and precise genomic mapping (Staševskij et al., 2017).

Direct detection by third-generation sequencing

While conversion-based next generation sequencing methods are more widely adopted, third generation sequencing analysis allows for the detection of DNA modifications without previous chemical conversion. The following methods are an attempt at overcoming the drawbacks of conversion-based ones given the inconsistency in conversion efficiency and, at times, inaccurate alignments of altered sequences to that of the reference (Barros-Silva et al., 2018).

SMRT sequencing is a method of assessing modified-cytosine bases without any preparatory chemical/enzymatic requirements; be it conversions and/or PCR amplifications (Tse et al., 2021). Sequencing by synthesis doesn’t detect modified-cytosine, perse, but rather the accumulating influence

of mCs on the kinetic signal changes of the DNA polymerase (Zhang et al., 2024). However, some sources site a lower sensitivity of mC, undermining the actual level of modified bases (Clark et al., 2013).

Oxford Nanopore Technologies (ONT) overcomes a myriad of the drawbacks innate to indirect mC detection methods. The process entails the passage of single DNA molecules through a voltage-based nanopore sensor. The voltage-based nanopore sensor then detects the background ionic current alterations of each base in the traversed DNA molecule.(Rang et al., 2018) Differentiation between modified and unmodified bases is achieved by the distinctive patterns of electric currents registered by the sequencer (Y. Liu et al., 2021).

The differences in detection occur after nanopore read base-calling and alignment in comparison to an *in silico* reference or the background pattern from an unmodified control sample.; as well as pre-trained learning models can be used. Applying the aforementioned principle into the question of detecting synthetic base modifications, ONT already allows for a commercially available algorithm, subject to machine learning and training to detect any DNA modifications of interest (Q. Liu et al., 2019).

ONT has attempted to improve the nanopore chemistry several times as a means of increasing the accuracy of the single-molecule, long-read sequencing. ONT releases new versions of flow cells, taking into consideration an improvement in sequencing accuracy, simplified workflow and rapid sample preparation protocols. Integrating advancements as well in base calling algorithms, generation of long reads, and the collection of much larger datasets (Mostafa, 2024).

### 1.6. Synthetic DNA Modifications:

The available spectrum of synthetic DNA modifications

As previously mentioned, the modification patterns of DNA confer an expressional change characterized by the combinatorial influence of said modifications. In order to linearize the link between distinct modifications and their elicited effect, DNA could be furnished with reporter tags that enable the visualization and isolation of distinct regions at specific loci of interest.

Reconsidering that MTases catalyze the transfer of a methyl group from the cofactor SAM. While it’s involved in numerous essential biochemical processes in living organisms, the transferred methyl group is characteristic of poor reactivity which impedes detection (Lukinavičius et al., 2007) A particular approach of overcoming such quality, is based on the replacement of the original sulfonium-bound methyl group of SAM with an extended transferrable sidechain, yielding an SAM analog. The

replacement of SAM by an aliphatic carbon chain includes a myriad of synthetic analogs, including an allylic system, a propargylic system, or an aromatic ring. Such development of sidechain-activated SAM analogs gave rise to a new approach of labeling deemed methyltransferase-directed Transfer of Activated groups (mTAG) (reviwed in Tomkuvienė et al., 2022).

However, in order to overcome the compatibility issue between natural MTases and the synthetic transferrable groups of the corresponding cofactor analogs, natural MTases are engineered to accommodate the said transferrable groups, giving rise an engineered Mtase (eMTase)(Tomkuvienė et al., 2019). A few of eMTases have been developed, capable of depositing the synthetic groups based on the natural MTase's originally recognized sequence. Per example, eM.HhaI, with a recognition sequence of GCGC, has exhibited compatibility with the following propargyl-based analogs (and the transferrable group): AdoButyn (Butynyl-), Ado-6-amine (Hexynamine-), Ado-6-azide (Hexynazide-).

Biotechnological Applications of Synthetic modifications

Synthetically modifying Cytosine prompts several biotechnological applications that aid in overcoming the innate shortcomings of natural modifications. Such shortcomings could be noted in the fact that DNMTs are in fact segregated in their functionality; however, the interplay amongst them remains heavily elusive. Additionally, naturally modified bases are unfortunately poor reporters, especially in DNA sequencing-based methods, and do not have the capacity to be appended with detectable derivatives (Tomkuvienė et al., 2019). Therefore, there are various applications for synthetic modifications in serving to further linearize the mechanisms of natural modifications and their regulation.

One application exploits the advancements in individual fluorophore optical detection that enables a determination of the physical location of fluorescently labeled target sites. This renders a nick-free optical map (OM) displaying the distribution of target sequences within the genome. Moreover, OM of mTAG, utilized for covalent tagging of DNA and RNA, has demonstrated potential utility in virus or macro-satellite genotyping. Another application exploited the innate blocked activity of MTase at mC modified bases by engineering M.SssI in biotin-labeling unmodified bases in a two-step process. This application is followed by NGS, rendering the "unmethylome" (Kriukienė et al., 2013). This application was further amended with the attachment of a DNA oligonucleotide, eventually deemed TOP-seq, which was validated in human tissue and cancer cell line whole-genome epigenetic profiling (Staševskij et al., 2017).

Moreover, M.SssI-directed tagging of hmC enabled the selective covalent capture of hmC-modified CG sites, through the utilization of cysteamine and succeeded amine-selective biotin labeling (Liutkevičiūtė et al., 2011). Similarly, this tagging method was applied to demonstrate photochemical biosensor for hmC quantitation as a proof of principle by M.HhaI-directed labeling (Z. Yang et al., 2016).

Applications of Synthetic modifications *in vivo*

However, in vivo applications of such techniques are also essential in transposing their utility. Stankevicius et al. replaced one of the three natural MTases by sterically engineering a biorthogonal MTase-Cofactor pair capable of covalently and selectively tagging respective targets in live mouse embryonic stem cells (mESC). This was followed by TOP-seq read-out of the genomic positions with the chemical tags. The data revealed eMTase transalkylation rates, with Ado-6-azide cofactor, was almost synonymous with that of wtMTase. Additionally, showing high efficiency in transferring propargylic groups (inclusive of AdoButyn cofactor) (Stankevičius et al., 2022)

Ahigh-resolution genome-wide unmethylome profiling was produced by MTase-TOP-seq, Stankevicius et al. moved towards taking a more detailed look at individual loci and weigh them up in comparison to native mC patterns. The range of validated genes, which showed moderately modified regions, include *Sfi1* and *H1fnt* – the genes implicated in murine neurodevelopment. The findings were also validated using Bisulfite sequencing (BS-seq). The aforementioned genes exhibited 30%–40% methylation coverage having good positional resemblance to that of TOP-seq data, nevertheless with a significant deviation within the promoter regions (Stankevičius et al., 2022). Therefore, assessing these regions and beyond by direct nanopore sequencing and modification detection is of our further scientific interest.

# 2. MATERIALS AND METHODS

## 2.1. Fragment amplification for modification experiments:

PCR reactions were performed using purified mESC genomic DNA (153ng/ul) as a template and either Phusion™ Plus DNA Polymerase or DreamTaq DNA Polymerase (5 U/µL) reaction kits (*Thermo Fisher Scientific*) according to manufactures' protocols. Each gene had only one forward (Fw) primer and two reverse (Rv) ones to produce different product lengths. Therefore, primers were designed to enable the amplification of either the target promoter (Rv1) regions of the genes or the target promoters along with downstream regions (Rv2) (Table 2.1.). As an exception, *Gm26917* had one reverse primer. The general trend proceeded with utilizing Phusion™ Plus DNA Polymerase for the amplification of long Rv2 fragments and DreamTaq DNA Polymerase for short Rv1 fragments.

**Table 2.1.** Gene fragments of interest, their forward primer, their two reverse primers, and the yielded fragment size of respective primer

| Gene | Forward Primer (Fw) | Reverse Primers (Rv) | PCR Product Size (bp) |
|---|---|---|---|
| ***FAT1*** | CACTGCCCATGTATTCATTG | **1:** GCAGTAACATAGCGAATGAC | 1129 |
| | | **2:** CATAGCCCACTGCTGTC | 5386 |
| ***Sfi1*** | GTACAAACGCAACGTGAG | **1:** CCCACTAGTTTTCACTGTTC | 826 |
| | | **2:** GCCATCATATGGTTGCTG | 7851 |
| ***H1fnt*** | CTTGATCTGGCACTTGAC | **1:** CTCAAGAAGGAGCTGG | 455 |
| | | **2:** CTGCAAGAACTTCAGAGC | 8726 |
| ***TMEM267*** | GCAGTCAGTGTTCTTAACC | **1:** CACAAAGAGACCATGTCTC | 1297 |
| | | **2:** CAGGAAACGAGTATCATCC | 7365 |
| ***Gm*** | GAAGCATCTGTGATTAAGAAGG | **-:** GATTTGAACTCCTGACCTTC | 6002 |

Thermocylic conditions were varied according to Phusion Plus (Table 2.2) per the manufacturer's protocol. Melting temperatures ($T_m$) based on the primers were calculated using Thermo Scientific Tm

Calculator (*Tm Calculator - LT*, n.d.) by selecting Phusion Polymerase specific option. While the annealing temperatures ($T_A$) were calculated based on subtraction of 5 °C from Tm.

**Table 2.2. Phusion Plus Cyclic Conditions of Gene Set**

| **Cycle Step** | **Temperature** | | | **Time** | **Cycles** |
|---|---|---|---|---|---|
| Initial Denaturation | 98 °C | | | 30 s | 1 |
| Denaturation | 98 °C | | | 5 – 10 s | 30 cycles |
| Annealing ($T_A$) (Tm - 5°C) | Gene | Tm | Gradient | 30 s | |
| | *FAT1* | Rv1: 59.9 °C<br>Rv2: 59.9 °C | 55 °C | | |
| | *Sfi1* | Rv1: 59.5 °C<br>Rv2: 59.5 °C | 54.5 °C | | |
| | *H1fnt* | Rv1: 58.5 °C<br>Rv2: 58.5 °C | 53.5 °C | | |
| | *TMEM267* | Rv1: 58.9 °C<br>Rv2: 58.9 °C | 53.9 °C | | |
| | *Gm26917* | Rv: 59.9 °C | 54.9 °C | | |
| Extension | 72°C | | | 15-30 s/kb | |
| Final Extension | 72°C | | | 5 – 10 min | 1 |

Similarly, DreamTaq cyclic conditions (Table 2.3) were per manufacturer's protocol, with $T_m$ and $T_A$ calculated based on selection of DreamTaq polymerase specific option.

**Table 2.3. DreamTaq Cyclic Conditions of Gene Set (continued in page 26)**

| **Cycle Step** | **Temperature** | | | **Time** | **Cycles** |
|---|---|---|---|---|---|
| Initial Denaturation | 95°C | | | 3 min | 1 |
| Denaturation | 95°C | | | 30 s | 30 |
| Annealing (Ta) (Tm - 5°C) | Gene | Tm | Gradient | 30 s | |
| | *FAT1* | Rv1: 59.9 °C<br>Rv2: 59.9 °C | 55 °C | | |
| | *Sfi1* | Rv1: 59.5 °C<br>Rv2: 59.5 °C | 54.5 °C | | |
| | *H1fnt* | Rv1: 58.5 °C<br>Rv2: 58.5 °C | 53.5 °C | | |
| | *TMEM267* | Rv1: 58.9 °C<br>Rv2: 58.9 °C | 53.9 °C | | |
| | *Gm26917* | Rv: 59.9 °C | 54.9 °C | | |

| | | | |
|---|---|---|---|
| Extension | 72°C | 1 min | |
| Final Extension | 72°C | 5 – 15 min | 1 |

2.2. PCR Product Validation by Restriction Analysis:

DNA digestion using restriction enzymes (Table 2.4) was performed according to manufacturer recommendations. The amount of purified PCR product utilized for restriction validation was ~50 ng, equal to that of the non-cleaved control sample of each gene. While the minimum amount of modified product restriction validation was ~30 ng.

**Table 2.4. Restriction Enzymes Utilized** with their respective buffer, objective in PCR or Modification Product Validation, respective recognition sequence, and gene set applied to

| **Enzyme Object** | **Restriction Enzyme** | **Recognition Site** | **PCR Product** |
|---|---|---|---|
| **PCR Product Validation** | LguI (5 U/µL)<br>10X Buffer Tango | 5'-GCTCTTC-3' | *FAT1*_2; *Gm26917* |
| | BamHI (10 U/µL)<br>10X Buffer BamHI | 5'-GGATCC-3' | *Sfi1*_1; *Sfi1*_2 |
| | XmaJI (AvrII) (10 U/µL)<br>10X Buffer Tango | 5'-CCTAGG-3' | *Sfi1*_1; *Sfi1*_2 |
| **Modification Validation** | Hin6I (HinP1I) (10 U/µL)<br>10X Buffer Tango | 5’-GCGC-3’ | *FAT1*_1; *FAT1*_2; *Sfi1*_1; *Gm26917*; *H1Fnt*_1 |
| | HpaII (10 U/µL)<br>10X Buffer Tango | 5’-CCGG-3’ | *FAT1*_2; *H1Fnt*_1 |
| | TaiI (MaeII) (10 U/µL)<br>10X Buffer R | 5’-ACGT-3’ | *Fat1*_1; *Sfi1*_1; *H1Fnt*_1 |

Validation of PCR Products in Gel Electrophoresis

1% Agarose gel electrophoresis in 1X TBE buffer (90mM Tris-base (Fisher Scientific, 10667243), 90mM boric acid (Carl Roth, 5935.2), 2mM EDTA (Carl Roth, 8040.2), pH=8 regulated with NaOH (Honeywell, 38215). The gel electrophoresis was set to run for a standard duration of 35 - 45 minutes at 150V with 350A. Samples whose corresponding restriction buffer was transparent were dyed with Orange DNA Loading Dye (6X) (Fisher Scientific, R0631) prior to gel loading. Band sizes were evaluated with loaded O'GeneRuler (Fisher Scientific, SM1163).

## 2.3. DNA Modifications

### Synthetic Modification

The modification reactions included the deposition of the synthetic groups with their corresponding cofactors, Butyn group along with AdoButyn cofactor (70mM); Octadiyn group along with AdoOctadiyn cofactor (42mM); Hexynazide group along with Ado-6-Azide cofactor (42mM). The types of buffers were attempted correspondingly with the synthetic methyltransferases, CutSmart Buffer along with eM.HhaIQN (109 µM); and M.SssI Buffer along with eM.SssIQN (11.3 µM).

Preparation of Synthetically Modified Samples

To prepare synthetically-modified samples at GCGC target sites, the purified PCR products were supplied with eM.HhaIQN (109 µM) MTase (Final Concentration $C_f$ = 1% of final reaction volume), LuxS (754 mM) ($C_f$ = 1% of final reaction volume), MTAN (478 mM) ($C_f$ = 1% of final reaction volume), CutSmart Buffer, and Milli-Q water up to total reaction volume. The supplied cofactor was supplied according to the desired modification outcome AdoButyn (70mM) ($C_f$ = 0.5 % of final reaction volume), AdoOctadiyn (42mM) ($C_f$ = 1% of final reaction volume), or Ado-6-Azide (42mM) ($C_f$ = 1% of final reaction volume), for the deposition of either Butyn, Octadiyn, or Hexynazide, respectively. Samples were incubated for a period of 3 hours at 37 °C. Incubation time was followed by deactivation for a period of 20 minutes at 65 °C.

hmC Modifications

In the case of sole hmC modification on the target fragment, CpG Methyltransferase (M.SssI) (Thermo Scientific™, EM0821) was utilized, 10X M.SssI buffer, 10% of total reaction volume of Formaldehyde (diluted to 1:100 ratio with Milli-Q water from stock concentration) and Milli-Q water up

to the total reaction volume. The samples were incubated at room temperature in the dark in a range of at least 2 to overnight incubation and deactivated at 65 °C for 20 minutes.

mC Modifications

In the case of sole mC modification on the target fragment, CpG Methyltransferase (M.SssI) was utilized along with SAM cofactor ($C_f$ = 2% of final reaction volume) according to manufacturer recommendations (Thermo Scientific, EM0821), and Milli-Q water up to total reaction volume. Samples were incubated for 15 minutes at 37 °C according to manufacturer protocol. Incubation time was followed by deactivation for a period of 20 minutes at 65 °C.

In the case of mC modification following hmC modification, SAM cofactor was added ($C_f$ = 2% of final reaction volume) into the reaction mixture. Samples were incubated for a period of 20 minutes or up to 1 hour at 37 °C. Incubation time was followed by deactivation for a period of 20 minutes at 65 °C.

Proteinase K treatment

Proteinase K treatment was performed in a reaction of 1% Proteinase K, along with 0.1% SDS, and Milli-Q water up to 100ul. The incubation period was 1 hour at 55 °C followed by 20 minute period of deactivation at 65 °C. Following Proteinase K treatment, reaction mixtures were subject to Zymo DNA Clean and Concentrator the Zymo-Spin IIC Fast-Spin column. The products were regularly eluted in 15 µl of Milli-Q water.

Library preparation and Sample Sequencing

Library preparation of the samples was conducted according to manufacturer's protocol (Ligation Sequencing Amplicons - Native Barcoding Kit 24 V14 (SQK-NBD114.24). Briefly, at least 30 ng of each of the following samples:

(1) *FAT1*_1;

(2) *FAT*:mC

(3) *FAT*:hmC

(4) *FAT*:hmC:mC

(5) *FAT*:byC:hmC:mC

(6) *FAT*:octC

(7) *FAT*:octC:hmC:mC

End-Prep

NEBNext Ultra II End Repair / dA-tailing Module reagents and DNA Control Sample (DCS) was prepared according to manufacturer's protocol. At least 30 ng of each sample was aliquoted into 0.2 ml thin-walled PCR tubes. Each sample was made up to 11.5 µl using nuclease-free water. Samples were incorporated with 1 µl of diluted DCS, 1.75 µl of Ultra II End-prep Reaction Buffer, and 0.75 µl of ultra II End-prep Enzyme Mix, with a final reaction volume of 15 µl. Mixtures were incubated at 20°C for 5 minutes and then 65°C for 5 minutes. Each sample was transferred into 1.5 ml Eppendorf DNA LoBind tube. 15 µl of resuspended AMPure XP beads (AXP) were added to each reaction and incubated in a rotator mixer for 5 minutes. Beads were then pelleted in each sample on a magnet until elute is clear and colorless and then washed with freshly prepared 80% ethanol.

Native Barcode ligation

NEB Blunt/TA Ligase Master Mix was prepared according to manufacturer's protocol. Then, a unique barcode was assigned to each of the samples.

In a 0.2 ml PCR-tube, each mixture was composed of 7.5 µl of End-prepped DNA, 2.5 µl of Native Barcode, 10 µl of Blunt/TA Ligase Master Mix, with a total reaction volume of 20 µl. Following incubation for 20 min at RT, 2 µl clear cap EDTA and 4 µl blue cap EDTA were added to each reaction mixture. Barcoded samples are then pooled into a 1.5 ml Eppendorf DNA LoBind tube. 0.4X AMPure XP Beads (AXP) to the pooled reaction and incubated at RT for 10 min in a rotator. After pelleting on a magnet, and sample are resuspended in freshly prepared 80% ethanol, washed, and repeated. Finally, 35 µl of eluate are retained into a 1.5 ml Eppendorf DNA LoBind tube.

Adapter Ligation and Clean up

NEBNext Quick Ligation Reaction Module was prepared according to the manufacturer's protocol. The reaction mixture was composed of 30 µl Pooled barcoded sample, Native Adapter 5 µl, NEBNext Quick Ligation Reaction Buffer (5X) 10 µl, Quick T4 DNA Ligase 5 µl, with a total reaction volume of 50 µl.

Sample sequencing

Sample sequencing was performed on MinION device using a R10.4.1 flow cell. The reads were base called with Dorado software using super high accuracy model. While the mC/hmC detection used model, which is provided by Oxford Nanopore, synthetic modification basecalling was performed using a custom Remora model trained for identification of 5odyC (trained by Joris Balčiūnas). Sequences and modifications were observed on Integrative Genomic Viewer (IGV) program.

# 3. RESULTS

## 3.1. PCR Products

The initial step of the project was to produce targeted PCR products ensuring the amplifications of the regions of interest. Genes of interest were comprised of *FAT1*; *Sfi1*; *H1fnt*; *TMEM267*; *Gm26917*. Regions of interest included the promoter regions of the genes (producing shorter fragments) with the first reverse primers designating the genes *FAT1*_1; *Sfi1*_1; *H1fnt*_1; *TMEM267*_1. While, whole gene regions (producing longer fragments) were to be produced with the second reverse primer, designating the genes *FAT1*_2; *Sfi1*_2; *H1fnt*_2; *TMEM267*_2; *Gm26917* (assigned only one primer for whole gene region). The polymerase, buffer, and cyclic conditions were adjusted according to a qualitative assessment of the gel electrophoresis.

The initial round of amplification (Figure 3.1) was conducted using Phusion Plus polymerase throughout the whole gene set, under the cyclic conditions of table 2.1. However, the buffer was the differentiating factor of the procedure, with an attempt at using Phusion Plus Buffer and Phusion GC Buffer to identify which fragments with clearer band intensity would be selected for further validation of gene products.

The Phusion plus buffer was adopted for the following genes in upcoming procedures of amplification: *FAT1*_1, *FAT1*_2, *Sfi1*_1, and *Sfi1*_2. While Phusion GC Buffer was adopted for *Gm26917*. Moreover, first reverse primer (producing the shorter products) amplified gene products were selected for further upscaling and amplification given they showed either none or very little non-specific amplification.

In this attempt, *H1Fnt*_1 and *H1Fnt*_2 failed to be amplified in the Phusion Plus buffer set. However, while *H1fnt*_2 failed to show any amplification in both buffers, H1fnt_1 showed a band light enough to exhibit a preferential condition of Phusion GC Buffer; thus, requiring an amendment to the reagents and/or cyclic conditions. On the other hand, *TMEM267*_1 and *TMEM267*_2 elicited non-specific fragment amplification, indicating the binding of the primers to the wrong targets and amplifying artifact products.

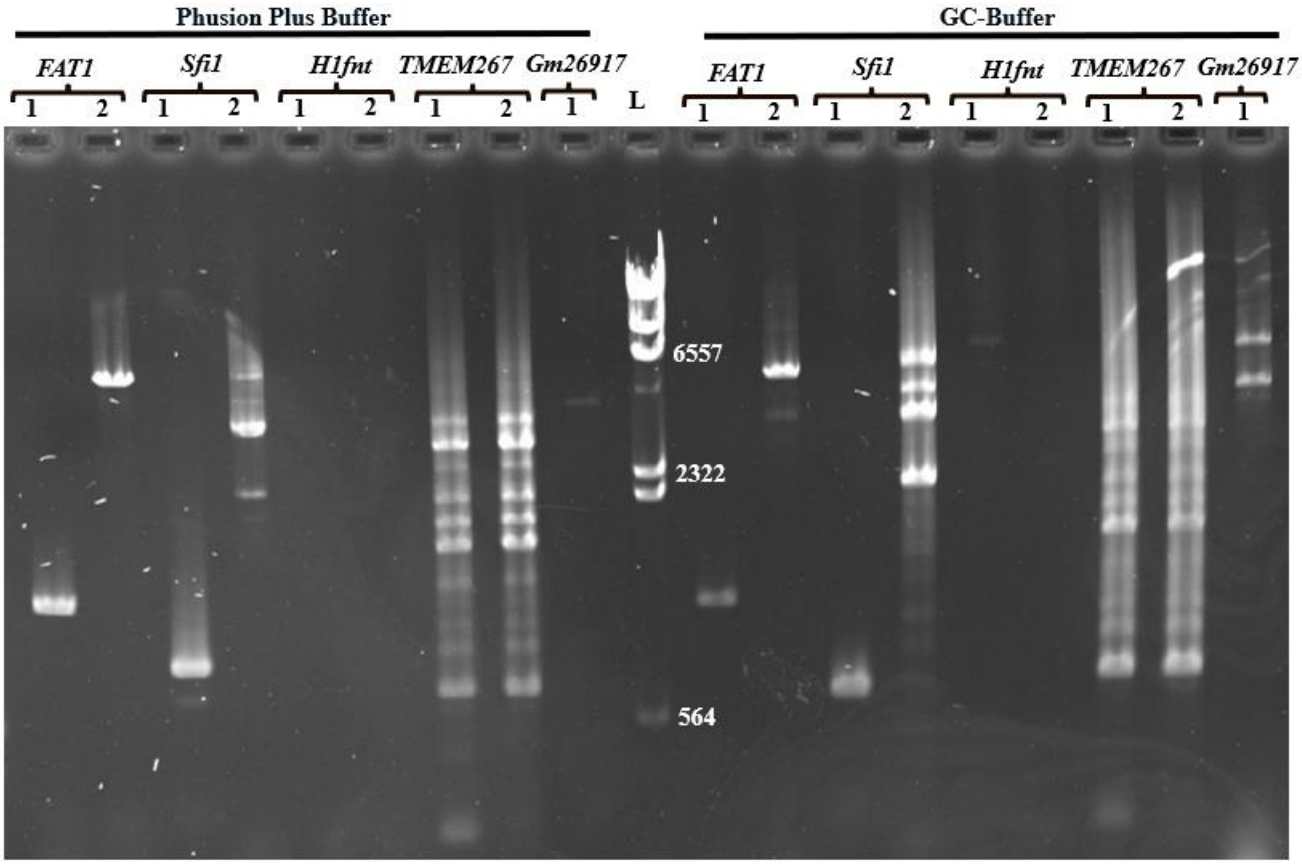


**Figure 3.1. PCR Amplification of the Selected Murine Gene Set.** The 1% agarose gel shows entire gene sets PCR amplified using either Phusion Plus Buffer or GC-Buffer. 1 indicates first reverse primer and 2 indicates the second reverse primer. L indicates the Ladder. The agarose gel was stained with Ethidium Bromide

After several attempts at altering the cyclic conditions of *H1fnt* and *TMEM267*, an attempt was made using a different reagent set. Therefore, DreamTaq polymerase, along with its corresponding buffer DreamTaq, was adopted (Figure 3.2). *TMEM267*_1 and *TMEM267*_2 still underscored non-specific fragment amplification. However, with the chosen reagents and cyclic conditions, *H1fnt*_1 exhibited a banding pattern compatible with the expected product size, and therefore, DreamTaq polymerase and its corresponding buffer were adopted for any future amplification plans.

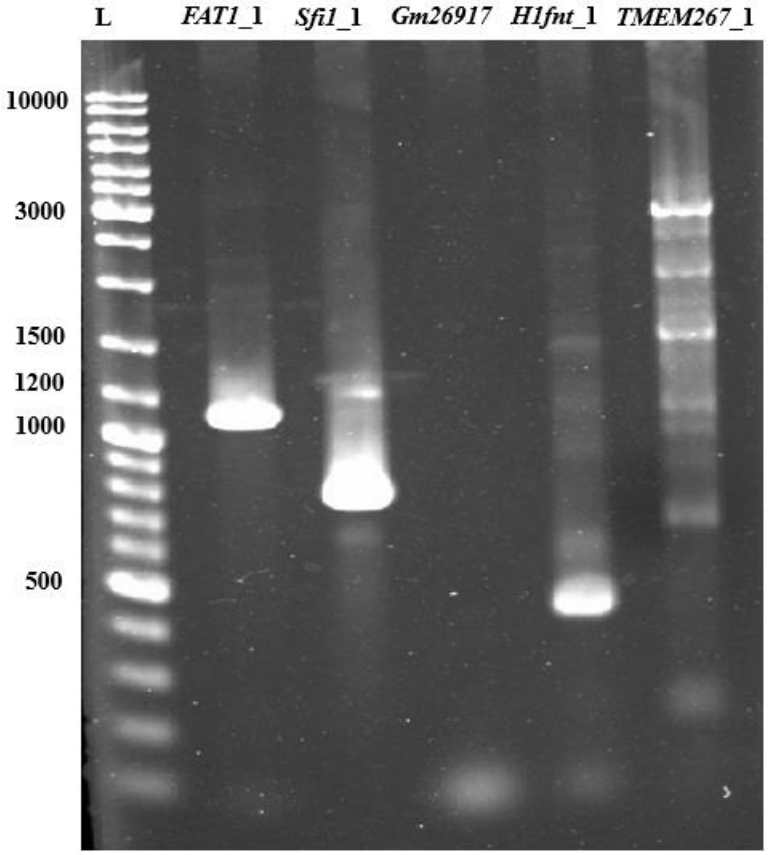


**Figure 3.2. DreamTaq PCR Amplification of the Selected Murine Gene Set.** The 1% agarose gel shows the PCR amplification of each gene using the first reverse primer. L indicates the Ladder. The agarose gel was stained with Ethidium Bromide

PCR Product Confirmation

While the PCR amplification yielded products of the expected size, it's important to validate that they are also of the intended sequence. For the initial verification we chose restriction analysis, as a robust method for DNA sequence pattern evaluation. R.LguI was the restriction enzyme selected for *FAT1*_2 and *Gm26917* for three distinct reasons: 1) given that its recognition sequence was present in both gene fragments, 2) is capable of yielding fragment sizes large enough to confidently validate within the gel qualitatively, 3) as well as the availability of the enzyme at hand. Results confirmed that *FAT1*_2 and *Gm26917* restriction products were of the expected size and quantity (Figure 3.3).

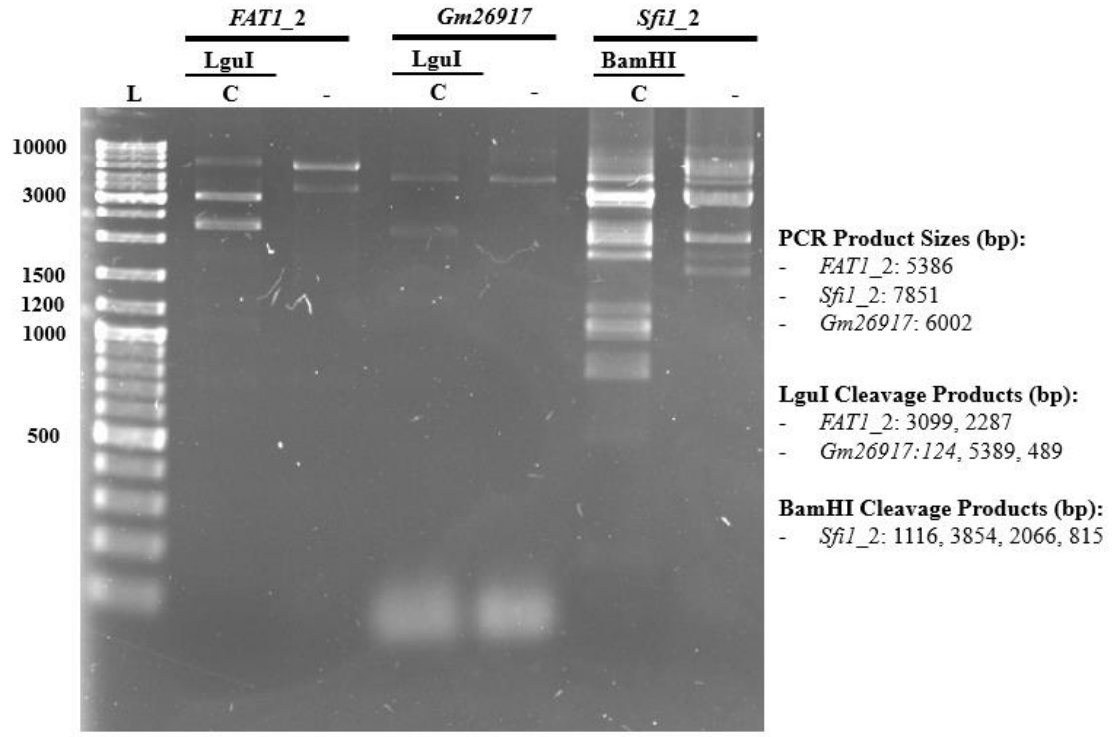


**Figure 3.3. Restriction Validation of PCR Products.** The 1% agarose gel validates the PCR products of *FAT1*_2 and *Gm26917* with R.LguI treatment; the PCR product of *Sfi*_2 with R.BamHI treatment."-" indicates untreated control samples. L indicates the Ladder. The agarose gel was stained with Ethidium Bromide.

On the other hand, R.BamHI was adopted for *Sfi1*_2 for the same aforementioned reasons. However, the initial restriction analysis of *Sfi*_2 did not yield the conclusive evidence required for ensuring that the intended fragments were amplified. It rather exhibited a restriction banding pattern that doesn't align with the expected outcome. Therefore, a second analysis was required. Analysis by XmaJI (Figure 3.4) confirmed the intended fragment of *Sfi*_1, however, *Sfi*_2 reiterated an ambiguous banding pattern; hence, *Sfi*_1 was adopted for any future amplification and modifications plans.

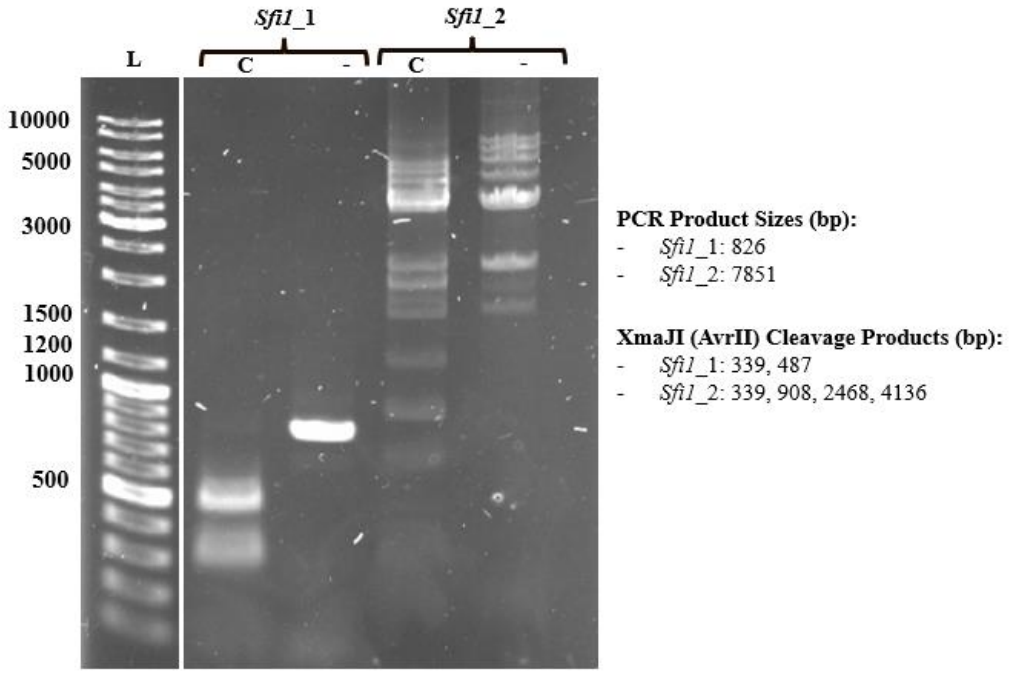


**Figure 3.4. Repeated Restriction Validation of *Sfi1* Products.** The 1% agarose gel validates the PCR products of *Sfi*_1 and *Sfi*_2 with R.XmaJI treatment. "-" indicates untreated control samples. L indicates the Ladder. The agarose gel was stained with Ethidium Bromide

Therefore, *FAT1*_2, *Gm26917,* and *Sfi1*_1 were selected to proceed with into the following modification reactions.

3.2. Modifications

Prior to initiating the rounds of modifications, the total volume of the amplification reactions was upscaled in order to obtain the selected PCR products in at least 1 µg amounts, given that this obtained amount would be distributed and utilized for further downstream procedures. The amounts would be distributed in the following consecutive order:

(1) Restriction analysis for validation of PCR product;

(2) Synthetic modification of said product;

(3) Restriction-protection analysis for evaluation of yielded modifications;

(4) hmC and mC modification of said product;

(5) Restriction-protection analysis of the mC/hmC for evaluation of yielded modifications;

(6) Library preparation of the modified samples for nanopore sequencing.

Steps (1, 2, 3) comprise the beginning of the modification reactions using the engineered MTases, along with any of the synthetic cofactor (By, Oct, HxN3) individually in order to deposit the synthetic groups of either CG or GCGC sites. Synthetic modifications were done prior to mC/hmC to optimize the reactions and attempt to yield (through qualitative assessment of the gel restriction analysis) the highest level of synthetic modification obtainable. Remaining steps (4, 5) were to further furnish the synthetically modified fragments with, first, hmC, and second, mC. hmC modifications were conducted prior to mC modifications to ensure that 5hmC modifications will be conferred, since in a co-factor independent modification reaction, MTases could yield an activated cytosine intermediate, capable of reacting with formaldehyde, and producing hmC. In the same reaction mixture, SAM, a natural cofactor, would eventually be introduced and overpowers the exogenous formaldehyde, thus yielding 5mC modifications unto the rest of CG sites. Such series of modifications was essential for step 6, in producing a fragment composite of synthetic, hmC, mC modifications, and possibly even lack thereof, at CG sites. This range of modifications itself would be utilized for differentiating the signal output of each modification. This will be a composite substrate for testing the ability of the modification detection program to discern between the natural and synthetic modifications.

Synthetic Modifications

The initial round of synthetic modifications was performed on validated PCR fragments, *FAT*_2, *Sfi1*_1, and *Gm26917* (Figure 3.5). Firstly, the modification utilized the engineered methyltransferase eM.HhaIQN targeting GCGC regions of the fragments; whereby, the deposition of the synthetic group targets the first cytosine nucleotide of the MTase's recognition sequence. Secondly, LuxS and MTAN were incorporated into the reaction since they serve a crucial role in degrading byproduct of the methyl cycle, the S-adenosylhomocysteine, also known as an inhibitor of methyltransferases. Thirdly, we aimed at exploring three different synthetic modifications (By, Oct, HxN3) and therefore incorporated any one of them in separate reactions. On the other hand, the control samples with -$CH_3$ modifications on the same target site (G<u>C</u>GC), were obtained using the wtM.HhaI MTase, and the cognate SAM cofactor.

Figure 3.5 depicts a restriction analysis using R.Hin6I, with the same corresponding restriction target sequence as that of the modification, G<u>C</u>GC. Any C5 modification at the inner cytosine blocks R.Hin6I activity. Therefore, we use it as a validator of successful modification reactions. Partial modification results in partial protection, which can be used as an estimation of modification yields.

*FAT1*_2 and *Sfi1*_1 had both been successfully modified by AdoOctadiyn, AdoHx$N_3$, and SAM; however, modification by AdoButyn was incomplete, rendering both protected and unprotected fragments in either of the samples. However, with regards to *Gm26917*, SAM was the only cofactor that had conferred a successful modification; all the remaining R.Hin6I treated samples of the said gene depicted a banding pattern resembling fully digested product controls (Figure 3.5).

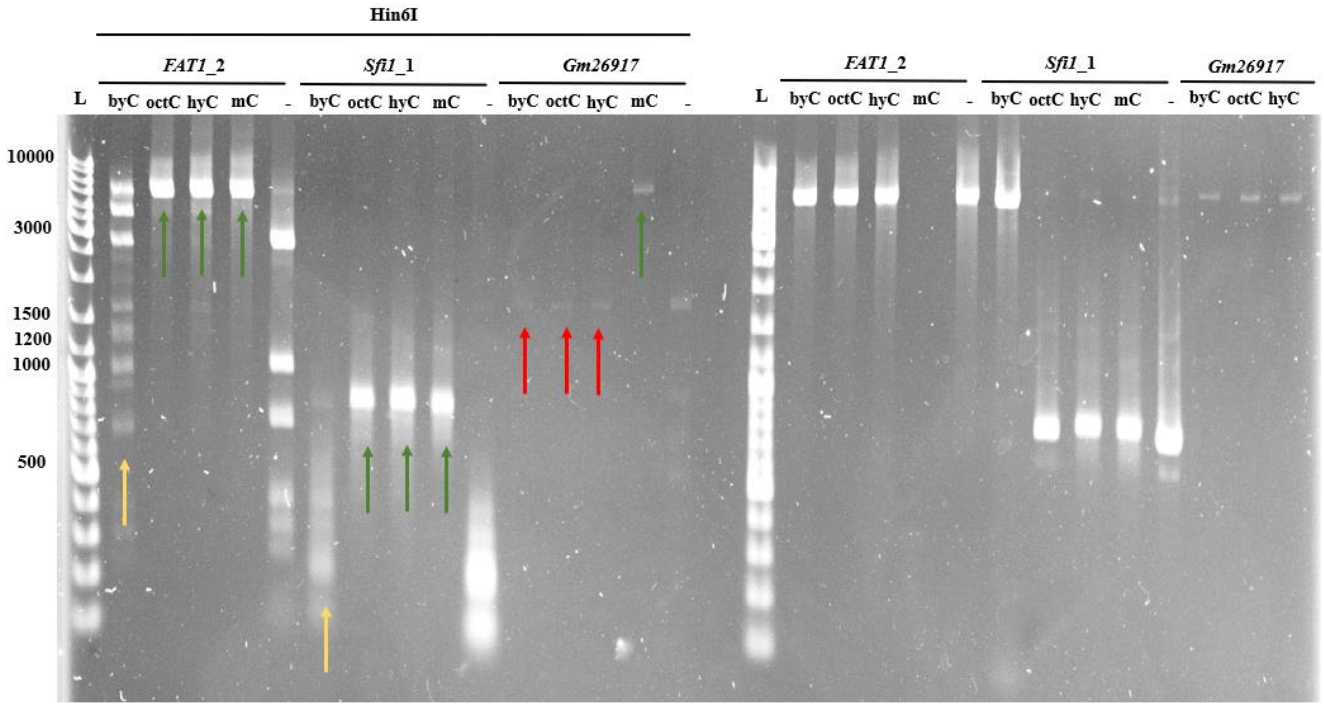


**Figure 3.5. Restriction Analysis Validation of Successful Modifications.** L indicates the ladder. Modifications of gene fragments *FAT1*_1; *Sfi1*_1; *Gm26917* with the following byC(Butyn), octC(Octadiyn), hyC(Hexynazide), mC(methyl); "-" indicates non-modified fragment. Hin6I indicates treatment with respective restriction enzyme. Green arrow indicates full protection; yellow arrow indicates partial protection; red arrow indicates no protection

A second round of modifications was performed (Figure 3.6) utilizing eM.SssIQN MTase targeting CG sites of the amplified fragments. Targeting CG sites with synthetic modifications would allow us to unify all targeted modification regions (i.e. synthetic/hmC/mC) as a means of comparing signal differentiation in downstream sequencing analysis. By this time point, Butyn modifications were

adopted to continue attempting its optimization since it was comprehensively unsuccessful in *FAT1*_1, *Sfi1*_1, and *H1fnt*_1.

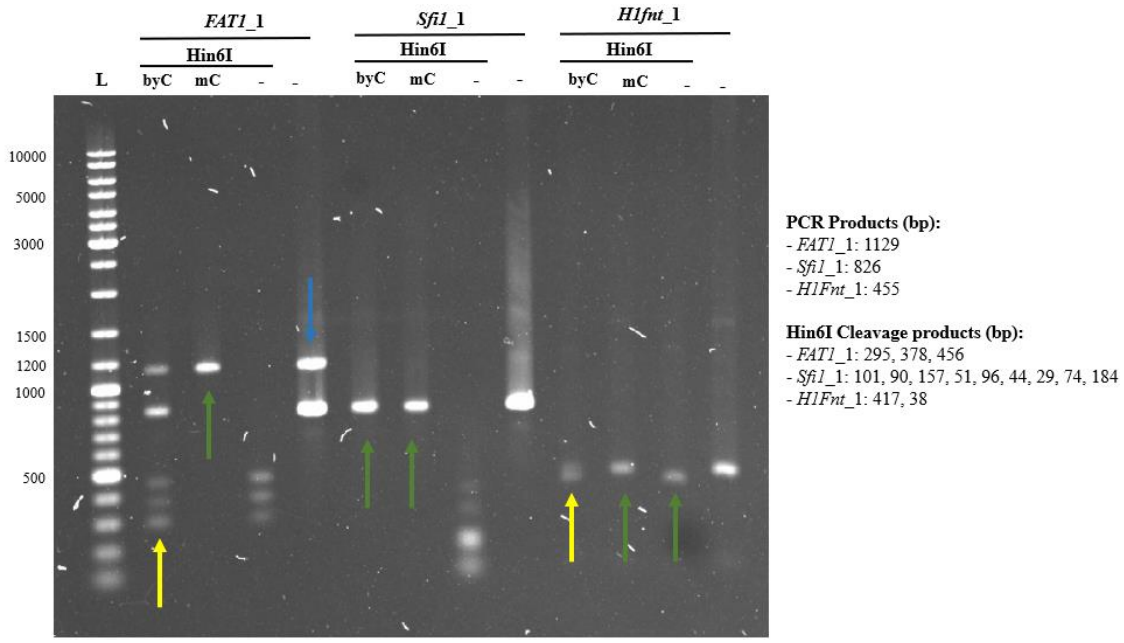


**Figure 3.6 Restriction Analysis Validation of Butyn Modifications.** L indicates the ladder. Modifications of gene fragments *FAT1*_1; *Sfi1*_1; *Gm26917*. with the following byC(Butyn), octC(Octadiyn), hyC(Hexynazide), mC(methyl); "-" indicates non-modified fragment. Hin6I indicates treatment with respective restriction enzyme. Green arrow indicates full protection; yellow arrow indicates partial protection; blue arrow indicates pipetting error of unmodified *FAT1*_1 and *Sfi1*_1 loaded within the same well

While all gene fragments exhibited full modification by SAM in the R.Hin6I treatment reaction, *Sfi1*_1 was the one product with the full protection by synthetic Butyn modification. Both *H1fnt*_1 and *FAT*_1 had been partially protected by Butyn. This partial protection/modification encouraged the need to validate the extent of modification by utilizing a different restriction enzyme; hence, targeting sites of modification. Additionally, hmC modifications were initiated as a comparison.

hmC and mC modifications

At this point, the results as aforementioned were promising in order to initiate hmC modifications. We used M.SssI with the target recognition CG for introducing natural DNA modifications. However, there is no known restriction enzyme which targets CG singularly. Therefore, to assess the modification efficiency, we used two different restriction enzymes with 4-bp targets (R.TaiI targeting ACGT; R.Hin6I targeting GCGC) including CG in different contexts, as a reduced representation of the aforementioned dinucleotides. However, due to MTase targeted sequence context preferences, it's important to assess the modifications at different endonuclease target regions of the gene set in question. By comparing the restriction analysis of the modified gene sets, there was an apparent difference in protection levels depending on the recognition sequence (Figure 3.7). Meaning that, the CG sites displayed differential degrees of modification when looking at only one recognition sequence of a given endonuclease – in this case, R.Hin6I. Another note with regards to the hmC modification, it would appear that *FAT1*_1 and *H1fnt*_1 had some degree of modification, while *Sfi1*_1 rendered almost non-apparent modification. The results, however, encouraged moving onwards with *FAT1*_1 and *H1fnt1*_1 to further decorate the synthetically modified samples with hmC followed by mC.

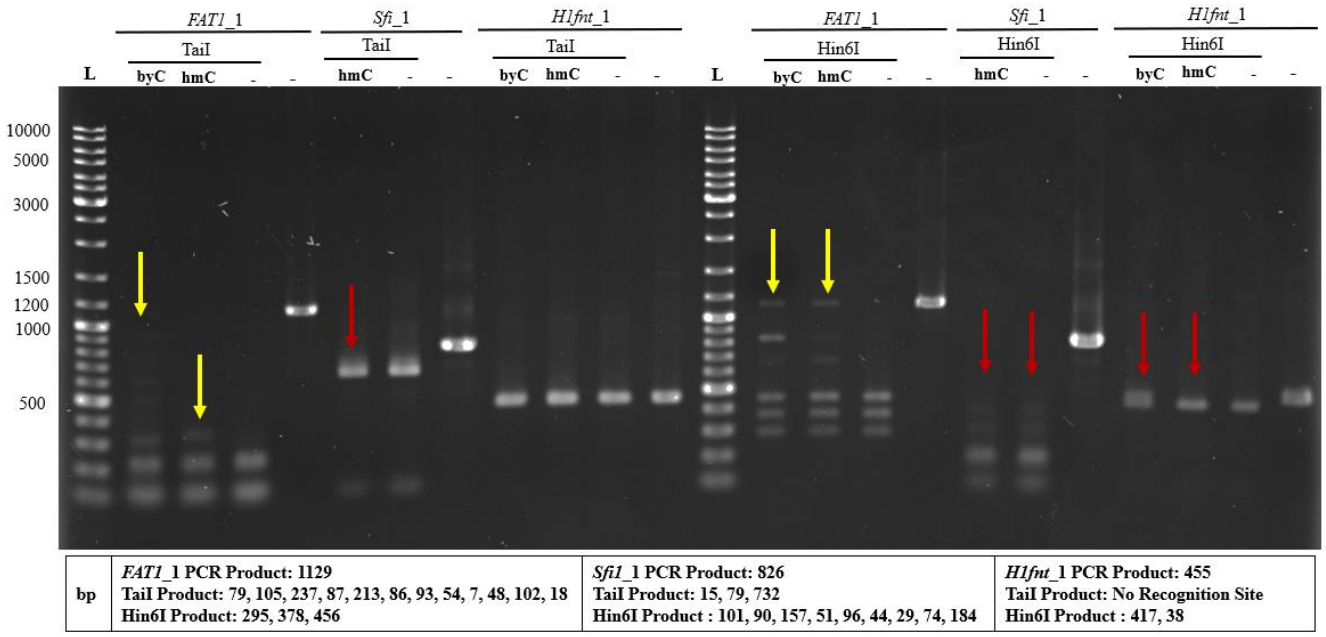


**Figure 3.7. Restriction Analysis Validation Targeting Different Modified Regions.** L indicates the ladder. Modifications of gene fragments *FAT1*_1; *Sfi1*_1; *H1fnt*_1 with byC (Butyn) and hmC(5-hydroxymethyl-). "-" indicates non-modified fragment. TaiI indicates treatment with respective

restriction enzyme. The same set of genes and modifications are subjected to Hin6I treatment. Yellow arrows indicate partial protection. Red Arrows indicate no protection

In order to further furnish the synthetically modified fragments with hmC at CG dinucleotides, we used the cofactor-independent aldehyde transfer activity of the wtM.SssI (Liutkeviciute et al., 2009; Rietjens et al., 2022).

Figure 3.8. clearly depicts the successful modification of *FAT1*_1 and *H1fnt1*_1 with byC, hmC, and mC. In addition to a higher degree of hmC protection after a second round of modification reaction. Essentially, qualitatively evaluating each degree of modification is based on visual comparison of band intensity between the modified fragment and non-modified fragment (both enzymatically cleaved). Moreover, fragment lengths found in the modified fragment lanes larger than those found in the non-modified fragment lanes are indicative of some degree of protection in the former.

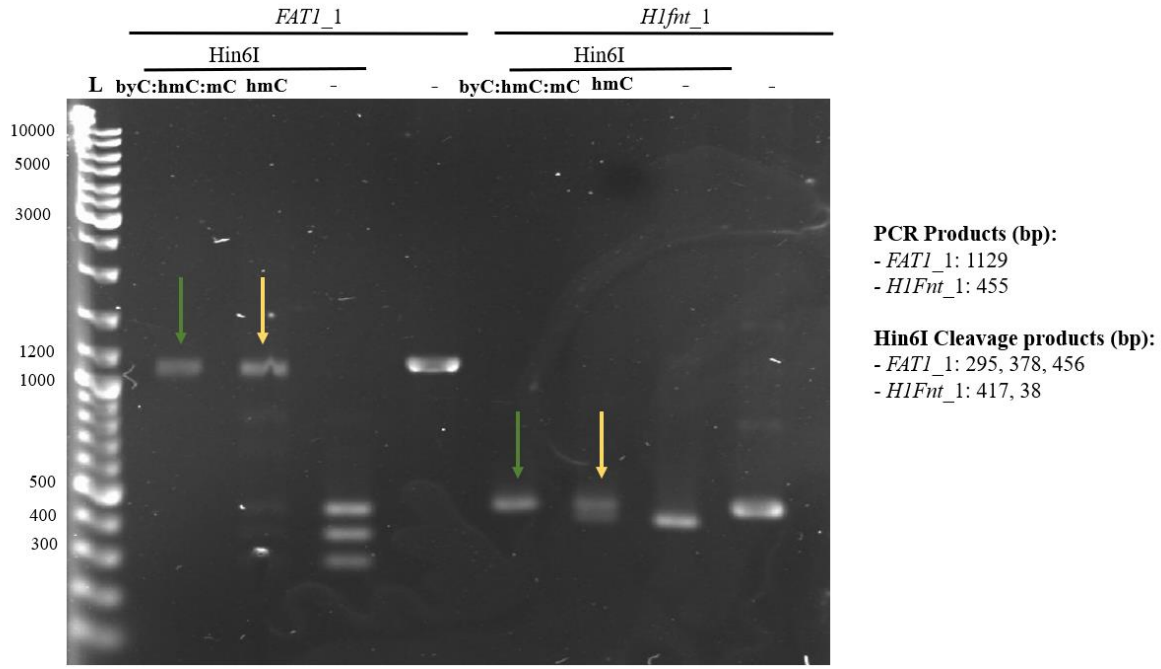


**Figure 3.8. Modification Restriction Analysis Validation of Sample Selection for Nanopore Sequencing.** L indicates the ladder. Modifications of gene fragments *FAT1*_1 (successful protection – selected for sequencing), with compound modification *FAT1*:byC:hmC:mC and *FAT1*:hmC. *H1fnt* (unsuccessful – not selected), with compound modifications *H1fnt*:byC:hmC:mC, and *H1fnt*:hmC. "-" indicates non-modified fragment. Hin6I indicates treatment with respective restriction enzyme

3.3.Nanopore sequencing of modified DNA

Upon attempting to upscale the reaction mixtures of PCR amplification, *FAT1*_1 was most consistent with the yield in purity, quality. and modification degree. Therefore, the aforementioned gene was singled out for sequencing. However, we had to change the synthetic modification from the optimized Butyn to a new one Oct, as the other parallel research in our lab showed it as the most promising one for precise direct detection by nanopore sequencing.

Eventually, a minimal starting amount of 30 ng of *FAT1*_1 PCR product was utilized as starting material for nanopore library preparation. Whereby a minimum of 1 ng was sufficient enough, according to manufacturer's protocol, to sequence. The final *FAT1*_1 sample Oct-modified by eM.HhaI at GCGC regions (at approx. 50% saturation, Figure 3.9A) was adopted and utilized for the nanopore sequencing. Additionally, *FAT1*_1:octC was further modified with hmC and mC at CG sites with wtM.SssI, yielding *FAT1*:octC:hmC:mC; while ensuring that some regions were left unmodified in order to emulate a more realistic representation of a natural DNA molecule extracted from cells (figure 3.9B).

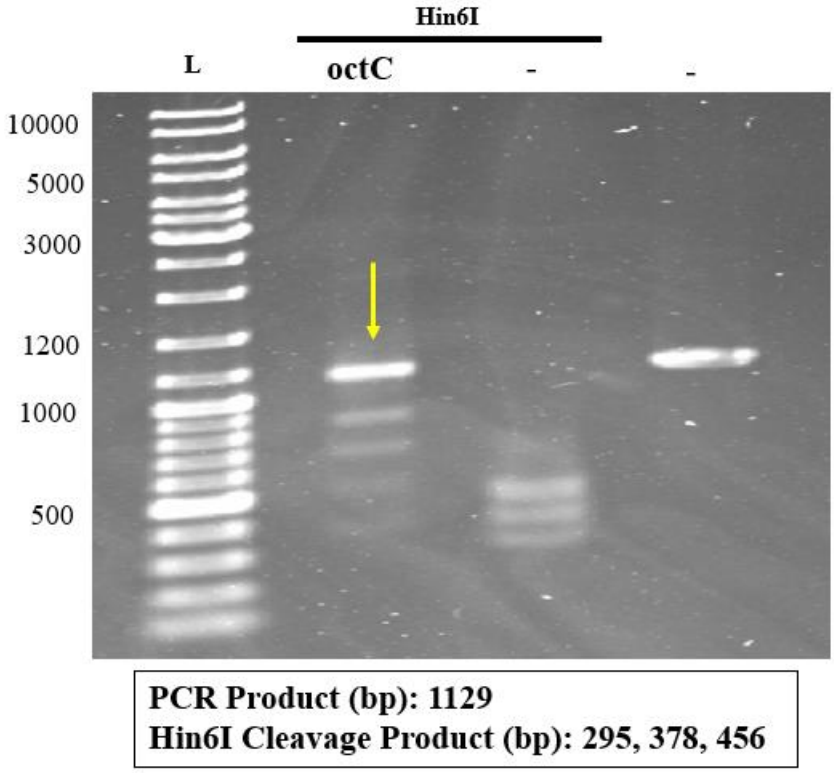

**Figure 3.9 (A) Restriction Analysis Validation of Successful *FAT1*_1 Compound Modification Selected for Nanopore Sequencing.** L indicates the ladder. Modifications of gene fragments *FAT1*_1 with octC,; "-" indicates non-modified fragment. R.Hin6I indicates treatment with respective restriction enzyme. Yellow arrow indicates partial protection

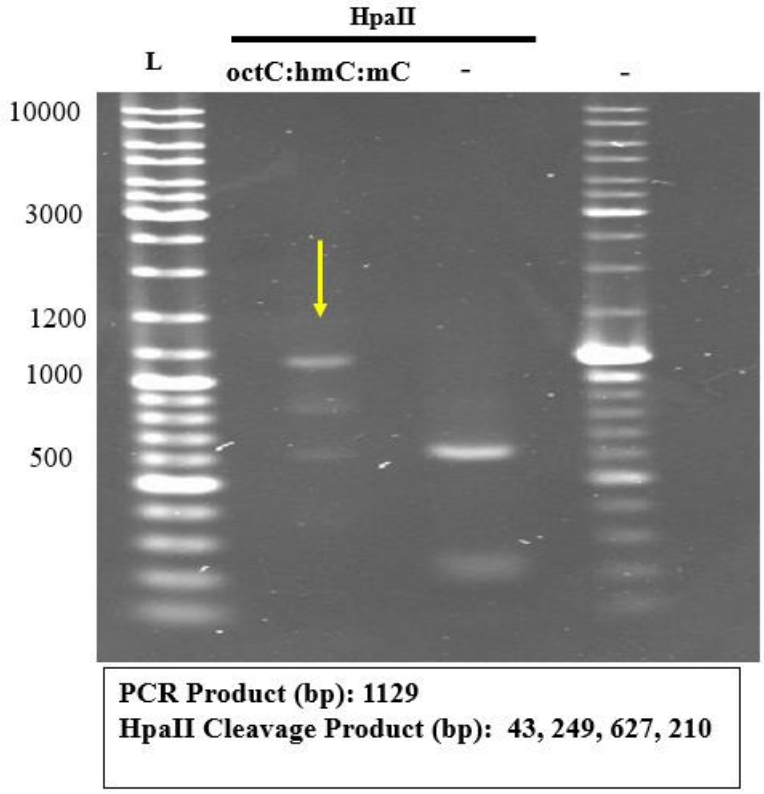


**Figure 3.9 (B). Restriction Analysis Validation of Successful *FAT1*_1 Compound Modification Selected for Nanopore Sequencning.** L indicates the ladder. Modifications of gene fragments *FAT1*_1 collectively with octC, hmC, and mC; "-" indicates non-modified fragment. HpaII indicates treatment with respective restriction enzyme. Yellow arrow indicates partial protection. Final lane "-" includes a technical error in adding ladder with non-modified, non-treated control fragment

Upon preparing the library, the samples were barcoded prior to loading for sequencing, in order to sequence the modified and unmodified ones in parallel. As the DNA goes through the pore, the electrical current signal is recorded and wrote out to a file and the signal changes can then be translated into DNA bases. This process is called basecalling. It's noted that ONT (Oxford Nanopore Technologies) basecalling models are used to translate electrical signal changes into canonical DNA

bases. However, per the manufacturer's commercialization, any DNA modifications can also be detected through the training of a supplementary modified basecalling algorithm. Currently, ONT provides commercially available models for some of the natural DNA modifications such as mC and hmC. However, to identify other DNA modifications such as octC a separate model has been trained before in the Department of Biological DNA Modification, Vilnius University. Previously, this model has only been tested on significantly modified DNA consisting of only octC modifications. Therefore, to test this model's efficiency in more naturalistic scenarios, several versions of variably modified *FAT*_1 PCR products were sequenced. The samples were then basecalled for the modified octC base (performed by Joris Balčiūnas). The *FAT1*:octC PCR product showed expected results, where the semi-modified GCGC position was partially identified as modified. However, the *FAT1*:octC:mC:hmC sample included quite a few false-positive signals, meaning the octC modification was called in regions other than G<u>C</u>GC, for example. This is likely due to the natural mC, hmC modifications in the sample causing the octC model to be less accurate.

Moreover, when looking at the percentage reads coverage, *FAT1*:octC:hmC:mC showed to have a higher percentage than that of *FAT1*:octC. Per example of figure 3.10, coverage at base 296 of *FAT1*:octC showed a modified base coverage of only 40% at G<u>C</u>GC. This coverage should be expected to be the same with *FAT1*:octC:hmC:mC since it's the same synthetically modified sample as *FAT1*:octC, only further modified with hmC:mC.

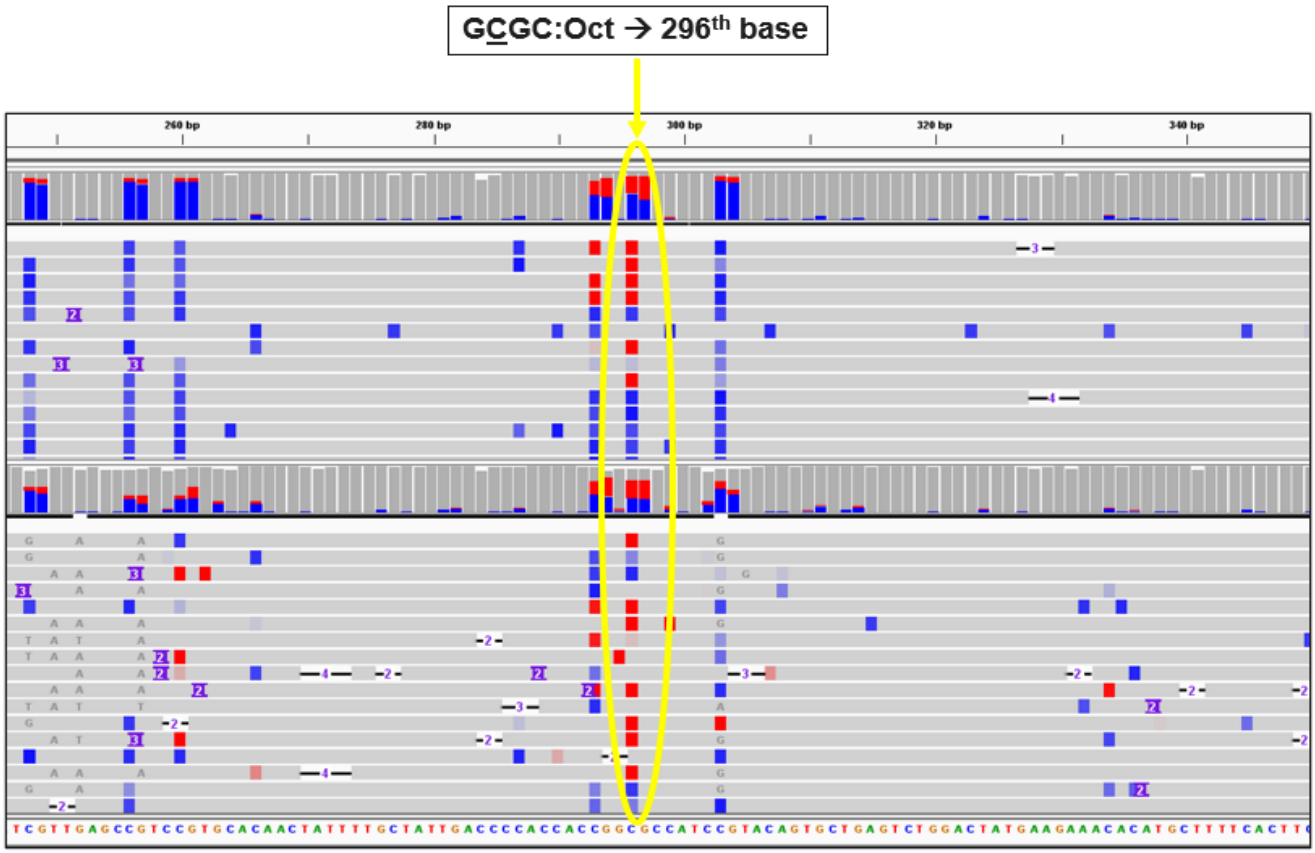


**Figure 3.10. Integrative Genomic Viewer (IGV) excerpt of *FAT1*_1.** Blue squares indicate <50% degree of confidence; Red squares indicate >50% degree of confidence. Each grey line represents one read strand in 5' – 3' direction (left to right). Upper lane indicates reads of *FAT1*:octC; Lower lane indicates reads of *FAT1*:octC:hmC:mC. Circled part represents expected synthetic modification (octC) at 296$^{th}$ base of *FAT1*_1 in GCGC site

However, *FAT1*:octC:hmC:mC displayed a much higher coverage of 60%; implicating that this 20% difference was the false-positive basecalling of hmC:mC modifications as octC. Such error is even more exaggerated when looking at the 678th base in GCGC (Figure 3.11.) Where *FAT1*:octC coverage was at 29%, while *FAT1*:octC:hmC:mC displayed a coverage of 66% more than double the former.

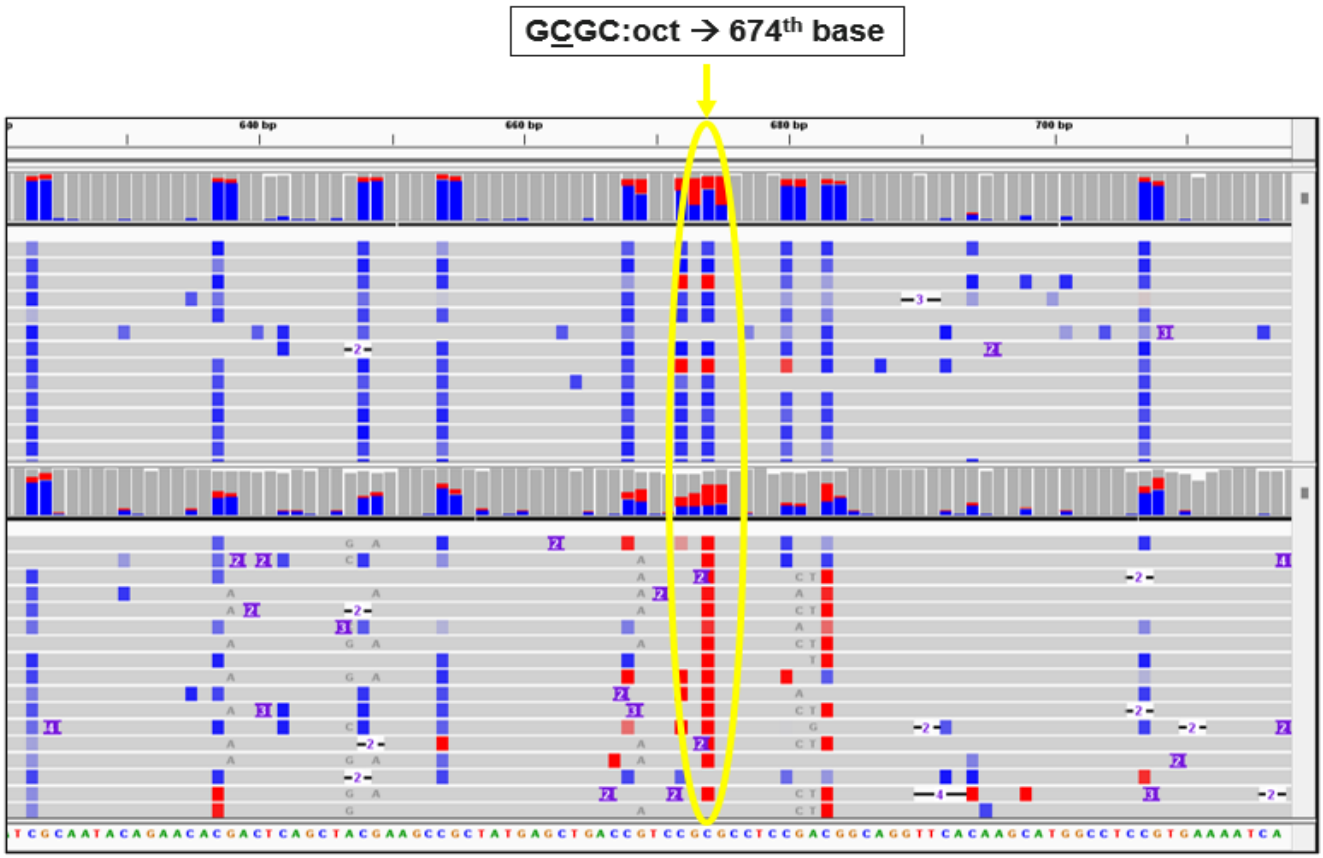


**Figure 3.11. Integrative Genomic Viewer (IGV) excerpt of *FAT1*_1.** Blue squares indicate <50% degree of confidence; Red squares indicate >50% degree of confidence. Each grey line represents one read strand in 5' – 3' direction (left to right). Upper lane indicates reads of *FAT1*:octC; Lower lane indicates reads of *FAT1*:octC:hmC:mC. Circled part represents expected synthetic modification (octC) at 296$^{th}$ base of *FAT1*_1 in GCGC site.

This shows that variably modified substrates cause issues with the model. It is not capable of correctly differentiating octC from mC or hmC, since we know for a fact, that the amount of octC in *FAT1*:octC:hmC:mC sample cannot be above the number of mod base coverage *FAT1*:octC.

# 4. DISCUSSION OF RESULTS

The first objective within the project was to optimize the PCR conditions for the set of selected murine genes. As results have shown, the PCR was mostly successful in each gene when amplified with its first reverse primer, resulting in shorter products. Several PCR targets, including *FAT1*_1, *Sfi1*_1, *H1fnt*_1, and *Gm26917*, showed good PCR yields and purity and were selected for further preparative scale-up and modification reactions.

With regards to the synthetic modifications, the primary goal was obtaining fragments of varying types of modifications. Nevertheless, *FAT*_1 substrate was selected to be focused on with the modifications and sequencing due to the consistency of its yield in PCR and modification reactions. Therefore, several types of modifications were to be induced in an attempt to optimize them; however, out of the variable modification, byC was singled out in needing further improvement of modification conditions. At the same time, parallel work was being done in our group in which Joris Balčiūnas had shown that octC modification is the most promising for detection by nanopore sequencing. Therefore, byC optimization was proceeded in parallel to octC. Additionally, it's also important for us to be able to distinguish among the natural and synthetic modifications, therefore modified PCR substrates were done separately, as well as in combinations. Finally, a modified substrate with combined modifications was prepared and nanopore sequencing showed that we could directly detect the octC modification in DNA sequence. However, currently it cannot be readily discerned from the natural mC and hmC modifications.

While nanopore sequencing is very much capable at identifying base modifications, training a model for a specific base might be challenging. This study describes the challenges of modified base identification in variably modified DNA substrates. To overcome these issues, further refinement of the octC model is needed. This could be achieved through the usage of a larger training dataset and/or the usage of multiple training datasets consisting of uniquely modified bases. Using multiple datasets would allow the training of a model capable of identifying multiple modified bases such as octC and mC at once, further increasing the model accuracy and diminishing the false-positive call rates.

# CONCLUSIONS

1- PCR conditions were optimized for the production of *FAT1*_1; *Sfi1*_1; and *H1fnt*_1 gene promoter regions as well as whole *Gm26917* gene region, with high specificity and yield.
2- Optimized modification reaction conditions allowed production of DNA substrates, modified both by each 5mC, 5hmC and synthetic modifications separately, and in combination.
3- Nanopore sequencing results showed that Octadiyn-C modification can be directly identified when it's the only modification present in the DNA substrate. However, it renders higher error rates when placed together with mC and hmC modifications in the same DNA substrate.

DESCRIPTION OF PERSONAL INPUT

The experiments that I had executed involve the PCR amplifications, restriction analysis of PCR products, gel electrophoresis, modification reactions, restriction analysis of modification reactions, library preparation and barcoding of modified samples, as well as final visualization on IGV.

ACKNOWLEDGEMENTS

I would like to thank my lab mates, doctors, and anyone kind enough to have had a part in what I had learned along this project. An important part I had included within my work was thanks to Joris Balčiūnas. He had a major part in the direction my project took.

I extend my deepest appreciation and gratitude to my supervisor, Dr Miglė Tomkuvienė. She guided me through every step of the way and was more than patient with me and my many mistakes. She was beyond enlightening, inspirational, and kind. I had the greatest opportunity and luck of being her student.

And while they're very far away from the Life Sciences Center, this acknowledgement wouldn't be complete without the immeasurable appreciation I have for my father, sister, and dear family friend. For if it weren't for them, their encouragement, and support, I never would've had the privilege of possibly the most priceless experience of my life.

LIST OF REFERENCES

Aberg, K. A., McClay, J. L., Nerella, S., Xie, L. Y., Clark, S. L., Hudson, A. D., Bukszár, J., Adkins, D., Swedish Schizophrenia Consortium, Hultman, C. M., Sullivan, P. F., Magnusson, P. K., & Van Den Oord, E. J. (2012). MBD-Seq as a Cost-Effective Approach for Methylome-Wide Association Studies: Demonstration in 1500 Case–Control Samples. *Epigenomics*, *4*(6), 605–621. https://doi.org/10.2217/epi.12.59

Ahmed, A. F. (2015). *Role of FAT1 cadherin in neuronal differentiation* [University of Newcastle]. https://ogma.newcastle.edu.au/vital/access/manager/Repository/uon:21081;jsessionid=7718AA965B3E3C100530F287DF4E43E9?exact=sm_creator%3A%22Ahmed%2C+Abdulrzag+Faraj%22&sort=sort_ss_title%2F

Andrews, S., Krueger, C., Mellado-Lopez, M., Hemberger, M., Dean, W., Perez-Garcia, V., & Hanna, C. W. (2023). Mechanisms and function of de novo DNA methylation in placental development reveals an essential role for DNMT3B. *Nature Communications*, *14*(1), 371. https://doi.org/10.1038/s41467-023-36019-9

Aran, D., Toperoff, G., Rosenberg, M., & Hellman, A. (2011). Replication timing-related and gene body-specific methylation of active human genes. *Human Molecular Genetics*, *20*(4), 670–680. https://doi.org/10.1093/hmg/ddq513

Arand, J., Chiang, H., Martin, D., Snyder, M., Sage, J., Reijo Pera, R., & Wossidlo, M. (2021). Tet enzymes are essential for early embryogenesis and completion of embryonic genome activation. *EMBO Reports*, *23*. https://doi.org/10.15252/embr.202153968

Arnaud, P. (2010). Genomic imprinting in germ cells: Imprints are under control. *Reproduction*, *140*(3), 411–423. https://doi.org/10.1530/REP-10-0173

Bajrami, E., & Spiroski, M. (2016). Genomic Imprinting. *Open Access Macedonian Journal of Medical Sciences*, *4*(1), 181–184. https://doi.org/10.3889/oamjms.2016.028

Bali, P., Im, H.-I., & Kenny, P. J. (2011). Methylation, memory and addiction. *Epigenetics*, *6*(6), 671–674. https://doi.org/10.4161/epi.6.6.15905

Barros-Silva, D., Marques, C. J., Henrique, R., & Jerónimo, C. (2018). Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. *Genes*, *9*(9), 429. https://doi.org/10.3390/genes9090429

Bommarito, P. A., & Fry, R. C. (2019). The Role of DNA Methylation in Gene Regulation. In *Toxicoepigenetics* (pp. 127–151). Elsevier. https://doi.org/10.1016/B978-0-12-812433-8.00005-8

Borgel, J., Guibert, S., & Weber, M. (2012). Methylated DNA Immunoprecipitation (MeDIP) from Low Amounts of Cells. In N. Engel (Ed.), *Genomic Imprinting* (Vol. 925, pp. 149–158). Humana Press. https://doi.org/10.1007/978-1-62703-011-3_9

Boyes, J., & Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell*, *64*(6), 1123–1134. https://doi.org/10.1016/0092-8674(91)90267-3

Chen, Z., Bu, N., Qiao, X., Zuo, Z., Shu, Y., Liu, Z., Qian, Z., Chen, J., & Hou, Y. (2018). Forkhead Box M1 Transcriptionally Regulates the Expression of Long Noncoding RNAs Snhg8 and Gm26917 to Promote Proliferation and Survival of Muscle Satellite Cells. *Stem Cells*, *36*(7), 1097–1108. https://doi.org/10.1002/stem.2824

Clark, T. A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S. W., He, C., & Korlach, J. (2013). Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biology*, *11*(1), 4. https://doi.org/10.1186/1741-7007-11-4

Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V., & Chuzhanova, N. A. (2010). Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Human Genomics*, *4*(6), 406–410. https://doi.org/10.1186/1479-7364-4-6-406

Cross, S. H., Meehan, R. R., Nan, X., & Bird, A. (1997). A component of the transcriptional represser MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature Genetics*, *16*(3), 256–259. https://doi.org/10.1038/ng0797-256

Daks, A., Vasileva, E., Fedorova, O., Shuvalov, O., & Barlev, N. A. (2022). The Role of Lysine Methyltransferase SET7/9 in Proliferation and Cell Stress Response. *Life*, *12*(3), 362. https://doi.org/10.3390/life12030362

Delaney, C., Garg, S. K., & Yung, R. (2015). Analysis of DNA Methylation by Pyrosequencing. *Methods in Molecular Biology (Clifton, N.J.)*, *1343*, 249–264. https://doi.org/10.1007/978-1-4939-2963-4_19

Drohat, A. C., & Coey, C. T. (2016). Role of Base Excision "Repair" Enzymes in Erasing Epigenetic Marks from DNA. *Chemical Reviews*, *116*(20), 12711–12729. https://doi.org/10.1021/acs.chemrev.6b00191

Fan, G., Martinowich, K., Chin, M. H., He, F., Fouse, S. D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., ten Hoeve, J., Shuai, K., & Sun, Y. E. (2005). DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. *Development (Cambridge, England)*, *132*(15), 3345–3356. https://doi.org/10.1242/dev.01912

Fan, Y., Nikitina, T., Morin-Kensicki, E. M., Zhao, J., Magnuson, T. R., Woodcock, C. L., & Skoultchi, A. I. (2003). H1 Linker Histones Are Essential for Mouse Development and Affect Nucleosome Spacing In Vivo. *Molecular and Cellular Biology*, *23*(13), 4559–4572. https://doi.org/10.1128/MCB.23.13.4559-4572.2003

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I., & Leonhardt, H. (2011). Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. *PLOS ONE*, *6*(6), e21306. https://doi.org/10.1371/journal.pone.0021306

Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity*, *105*(1), 4–13. https://doi.org/10.1038/hdy.2010.54

Goto, K., Numata, M., Komura, J. I., Ono, T., Bestor, T. H., & Kondo, H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice.

*Differentiation; Research in Biological Diversity*, *56*(1–2), 39–44. https://doi.org/10.1046/j.1432-0436.1994.56120039.x

Guo, J. U., Ma, D. K., Mo, H., Ball, M. P., Jang, M.-H., Bonaguidi, M. A., Balazer, J. A., Eaves, H. L., Xie, B., Ford, E., Zhang, K., Ming, G., Gao, Y., & Song, H. (2011). Neuronal activity modifies DNA methylation landscape in the adult brain. *Nature Neuroscience*, *14*(10), 1345–1351. https://doi.org/10.1038/nn.2900

Hellman, A., & Chess, A. (2007). Gene Body-Specific Methylation on the Active X Chromosome. *Science*, *315*(5815), 1141–1143. https://doi.org/10.1126/science.1136352

Hendrich, B., & Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Molecular and Cellular Biology*, *18*(11), 6538–6547.

Hervouet, E., Vallette, F. M., & Cartron, P.-F. (2010). Dnmt1/Transcription Factor Interactions. *Genes & Cancer*, *1*(5), 434–443. https://doi.org/10.1177/1947601910373794

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), 3156. https://doi.org/10.1186/gb-2013-14-10-r115

Huang, Z., Bassil, C. F., & Murphy, S. K. (2013). Methylation-specific PCR. *Methods in Molecular Biology (Clifton, N.J.)*, *1049*, 75–82. https://doi.org/10.1007/978-1-62703-547-7_7

Jacinto, F. V., Ballestar, E., & Esteller, M. (2008). Methyl-DNA immunoprecipitation (MeDIP): Hunting down the DNA methylome. *BioTechniques*, *44*(1), 35–43. https://doi.org/10.2144/000112708

Jjingo, D., Conley, A. B., Yi, S. V., Lunyak, V. V., & Jordan, I. K. (2012). On the presence and role of human gene-body DNA methylation. *Oncotarget*, *3*(4), 462–474.

Kim, Y.-I. (2005). Nutritional Epigenetics: Impact of Folate Deficiency on DNA Methylation and Colon Cancer Susceptibility12. *The Journal of Nutrition*, *135*(11), 2703–2709. https://doi.org/10.1093/jn/135.11.2703

Kishi, Y., Fujii, Y., Hirabayashi, Y., & Gotoh, Y. (2012). HMGA regulates the global chromatin state and neurogenic potential in neocortical precursor cells. *Nature Neuroscience*, *15*(8), 1127–1133. https://doi.org/10.1038/nn.3165

Kriukienė, E., Labrie, V., Khare, T., Urbanavičiūtė, G., Lapinaitė, A., Koncevičius, K., Li, D., Wang, T., Pai, S., Ptak, C., Gordevičius, J., Wang, S.-C., Petronis, A., & Klimašauskas, S. (2013). DNA unmethylome profiling by covalent capture of CpG sites. *Nature Communications*, *4*(1), 2190. https://doi.org/10.1038/ncomms3190

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915–926. https://doi.org/10.1016/0092-8674(92)90611-f

Liao, J., Karnik, R., Gu, H., Ziller, M. J., Clement, K., Tsankov, A. M., Akopian, V., Gifford, C. A., Donaghey, J., Galonska, C., Pop, R., Reyon, D., Tsai, S. Q., Mallard, W., Joung, J. K., Rinn, J. L., Gnirke, A., & Meissner, A. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nature Genetics*, *47*(5), 469–478. https://doi.org/10.1038/ng.3258

Lin, Q., Sirotkin, A., & Skoultchi, A. I. (2000). Normal Spermatogenesis in Mice Lacking the Testis-Specific Linker Histone H1t. *Molecular and Cellular Biology*, *20*(6), 2122–2128.

Liu, Q., Georgieva, D. C., Egli, D., & Wang, K. (2019). NanoMod: A computational tool to detect DNA modifications using Nanopore long-read sequencing data. *BMC Genomics*, *20*(1), 78. https://doi.org/10.1186/s12864-018-5372-8

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., & Li, S. (2021). DNA methylation-calling tools for Oxford Nanopore sequencing: A survey and human epigenome-wide evaluation. *Genome Biology*, *22*(1), 295. https://doi.org/10.1186/s13059-021-02510-z

Liutkevičiūtė, Z., Kriukienė, E., Grigaitytė, I., Masevičius, V., & Klimašauskas, S. (2011). Methyltransferase-Directed Derivatization of 5-Hydroxymethylcytosine in DNA. *Angewandte Chemie International Edition*, *50*(9), 2090–2093. https://doi.org/10.1002/anie.201007169

Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., & Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chemical Biology*, *5*(6), 400–402. https://doi.org/10.1038/nchembio.172

Lorincz, M. C., Dickerson, D. R., Schmitt, M., & Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural & Molecular Biology*, *11*(11), 1068–1075. https://doi.org/10.1038/nsmb840

Lukinavičius, G., Lapienė, V., Staševskij, Z., Dalhoff, C., Weinhold, E., & Klimašauskas, S. (2007). Targeted Labeling of DNA by Methyltransferase-Directed Transfer of Activated Groups (mTAG). *Journal of the American Chemical Society*, *129*(10), 2758–2759. https://doi.org/10.1021/ja0691876

Mayer, W., Niveleau, A., Walter, J., Fundele, R., & Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature*, *403*(6769), 501–502. https://doi.org/10.1038/35000656

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., & Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell*, *151*(7), 1417–1430. https://doi.org/10.1016/j.cell.2012.11.022

Menezo, Y., Clement, P., Clement, A., & Elder, K. (2020). Methylation: An Ineluctable Biochemical and Physiological Process Essential to the Transmission of Life. *International Journal of Molecular Sciences*, *21*(23), 9311. https://doi.org/10.3390/ijms21239311

Mizuno, S., Hirota, J., Ishii, C., Iwasaki, H., Sano, Y., & Furuichi, T. (2020). Comprehensive Profiling of Gene Expression in the Cerebral Cortex and Striatum of BTBRTF/ArtRbrc Mice Compared to C57BL/6J Mice. *Frontiers in Cellular Neuroscience*, *14*, 595607. https://doi.org/10.3389/fncel.2020.595607

Mohn, F., Weber, M., Schübeler, D., & Roloff, T.-C. (2009). Methylated DNA Immunoprecipitation (MeDIP). In J. Tost (Ed.), *DNA Methylation: Methods and Protocols* (pp. 55–64). Humana Press. https://doi.org/10.1007/978-1-59745-522-0_5

Montero, L. M., Filipski, J., Gil, P., Capel, J., Martínez-Zapater, J. M., & Salinas, J. (1992). The distribution of 5-methylcytosine in the nuclear genome of plants. *Nucleic Acids Research*, *20*(12), 3207–3210. https://doi.org/10.1093/nar/20.12.3207

Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Mostafa, H. H. (2024). An evolution of Nanopore next-generation sequencing technology: Implications for medical microbiology and public health. *Journal of Clinical Microbiology*, *62*(5), e00246-24. https://doi.org/10.1128/jcm.00246-24

Nan, X., Ng, H.-H., Johnson, C., Laherty, C., Turner, B., Eisenman, R., & Bird, A. (1998). Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird ATranscriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386-389. *Nature*, *393*, 386–389. https://doi.org/10.1038/30764

Newkirk, S. J., & An, W. (2020). UHRF1: A jack of all trades, and a master epigenetic regulator during spermatogenesis. *Biology of Reproduction*, *102*(6), 1147–1152. https://doi.org/10.1093/biolre/ioaa026

Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., & Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Developmental Dynamics*, *236*(6), 1663–1676. https://doi.org/10.1002/dvdy.21176

Nicholson, T. B., Veland, N., & Chen, T. (2015). Writers, Readers, and Erasers of Epigenetic Marks. In *Epigenetic Cancer Therapy* (pp. 31–66). Elsevier. https://doi.org/10.1016/B978-0-12-800206-3.00003-3

Nohesara, S., Abdolmaleky, H. M., & Thiagalingam, S. (2023). Epigenetic Aberrations in Major Psychiatric Diseases Related to Diet and Gut Microbiome Alterations. *Genes*, *14*(7), 1506. https://doi.org/10.3390/genes14071506

Oda, M., & Greally, J. M. (2009). The Help Assay. In J. Tost (Ed.), *DNA Methylation: Methods and Protocols* (pp. 77–87). Humana Press. https://doi.org/10.1007/978-1-59745-522-0_7

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*, *99*(3), 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Pfeifer, G. P., & Rauch, T. A. (2009). DNA methylation patterns in lung carcinomas. *Seminars in Cancer Biology*, *19*(3), 181–187. https://doi.org/10.1016/j.semcancer.2009.02.008

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*(1), 90. https://doi.org/10.1186/s13059-018-1462-9

Rasmussen, K. D., & Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development*, *30*(7), 733–750. https://doi.org/10.1101/gad.276568.115

Ravichandran, M., Jurkowska, R., & Jurkowski, T. (2017). Target specificity of mammalian DNA methylation and demethylation machinery. *Organic & Biomolecular Chemistry*, *16*. https://doi.org/10.1039/C7OB02574B

Rietjens, I., Michael, A., Bolt, H., Bourdoux, S., Andrea, H., Hinrichsen, N., Kalisch, C., Angela, M., Gloria, P., Ribera, D., Natalie, T., & Eisenbrand, G. (2022). The role of endogenous versus exogenous sources in the exposome of putative genotoxins and consequences for risk assessment. *Archives of Toxicology*, *96*. https://doi.org/10.1007/s00204-022-03242-0

Saenz-de-Juano, M. D., Ivanova, E., Billooye, K., Herta, A.-C., Smitz, J., Kelsey, G., & Anckaert, E. (2019). Genome-wide assessment of DNA methylation in mouse oocytes reveals effects associated with in vitro growth, superovulation, and sexual maturity. *Clinical Epigenetics*, *11*, 197. https://doi.org/10.1186/s13148-019-0794-y

Saxonov, S., Berg, P., & Brutlag, D. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 1412–1417. https://doi.org/10.1073/pnas.0510310103

Segurado, R., Conroy, J., Meally, E., Fitzgerald, M., Gill, M., & Gallagher, L. (2005). Confirmation of Association Between Autism and the Mitochondrial Aspartate/Glutamate Carrier SLC25A12 Gene on Chromosome 2q31. *American Journal of Psychiatry*, *162*(11), 2182–2184. https://doi.org/10.1176/appi.ajp.162.11.2182

Skinner, M. K. (2011). Role of epigenetics in developmental biology and transgenerational inheritance. *Birth Defects Research Part C: Embryo Today: Reviews*, *93*(1), 51–55. https://doi.org/10.1002/bdrc.20199

Stankevičius, V., Gibas, P., Masiulionytė, B., Gasiulė, L., Masevičius, V., Klimašauskas, S., & Vilkaitis, G. (2022). Selective chemical tracking of Dnmt1 catalytic activity in live cells. *Molecular Cell*, *82*(5), 1053-1065.e8. https://doi.org/10.1016/j.molcel.2022.02.008

Staševskij, Z., Gibas, P., Gordevičius, J., Kriukienė, E., & Klimašauskas, S. (2017). Tethered Oligonucleotide-Primed Sequencing, TOP-Seq: A High-Resolution Economical Approach for DNA Epigenome Profiling. *Molecular Cell*, *65*(3), 554-564.e6. https://doi.org/10.1016/j.molcel.2016.12.012

Tate, P. H., & Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics & Development*, *3*(2), 226–231. https://doi.org/10.1016/0959-437X(93)90027-M

*Tm Calculator—LT*. (n.d.). Retrieved May 23, 2024, from https://www.thermofisher.com/tr/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html

Tomkuvienė, M., Kriukienė, E., & Klimašauskas, S. (2022). DNA Labeling Using DNA Methyltransferases. In A. Jeltsch & R. Z. Jurkowska (Eds.), *DNA Methyltransferases—Role and Function* (Vol. 1389, pp. 535–562). Springer International Publishing. https://doi.org/10.1007/978-3-031-11454-0_19

Tomkuvienė, M., Mickutė, M., Vilkaitis, G., & Klimašauskas, S. (2019). Repurposing enzymatic transferase reactions for targeted labeling and analysis of DNA and RNA. *Current Opinion in Biotechnology*, *55*, 114–123. https://doi.org/10.1016/j.copbio.2018.09.008

Trent, S., Fry, J. P., Ojarikre, O. A., & Davies, W. (2014). Altered brain gene expression but not steroid biochemistry in a genetic mouse model of neurodevelopmental disorder. *Molecular Autism*, *5*(1), 21. https://doi.org/10.1186/2040-2392-5-21

Tse, O. Y. O., Jiang, P., Cheng, S. H., Peng, W., Shang, H., Wong, J., Chan, S. L., Poon, L. C. Y., Leung, T. Y., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proceedings of the National Academy of Sciences*, *118*(5), e2019768118. https://doi.org/10.1073/pnas.2019768118

Veland, N., Lu, Y., Hardikar, S., Gaddis, S., Zeng, Y., Liu, B., Estecio, M. R., Takata, Y., Lin, K., Tomida, M. W., Shen, J., Saha, D., Gowher, H., Zhao, H., & Chen, T. (2019). DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Research*, *47*(1), 152–167. https://doi.org/10.1093/nar/gky947

Wu, H., & Zhang, Y. (2014). Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell*, *156*(1–2), 45–68. https://doi.org/10.1016/j.cell.2013.12.019

Yang, A. Y., Kim, H., Li, W., & Kong, A.-N. T. (2016). Natural compound-derived epigenetic regulators targeting epigenetic readers, writers and erasers. *Current Topics in Medicinal Chemistry*, *16*(7), 697–713.

Yang, S.-M., Kim, B. J., Norwood Toro, L., & Skoultchi, A. I. (2013). H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(5), 1708–1713. https://doi.org/10.1073/pnas.1213266110

Yang, Z., Shi, Y., Liao, W., Yin, H., & Ai, S. (2016). A novel signal-on photoelectrochemical biosensor for detection of 5-hydroxymethylcytosine based on in situ electron donor producing strategy and all wavelengths of light irradiation. *Sensors and Actuators B: Chemical*, *223*, 621–625. https://doi.org/10.1016/j.snb.2015.09.159

Yebra, M. J., & Bhagwat, A. S. (1995). A Cytosine Methyltransferase Converts 5-Methylcytosine in DNA to Thymine. *Biochemistry*, *34*(45), 14752–14757. https://doi.org/10.1021/bi00045a016

Yildirim, O., Li, R., Hung, J.-H., Chen, P. B., Dong, X., Ee, L.-S., Weng, Z., Rando, O. J., & Fazzio, T. G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell*, *147*(7), 1498–1510. https://doi.org/10.1016/j.cell.2011.11.054

Zhang, H., Liu, L., & Li, M. (2024). Mini-review of DNA Methylation Detection Techniques and Their Potential Applications in Disease Diagnosis, Prognosis, and Treatment. *ACS Sensors*, *9*(3), 1089–1103. https://doi.org/10.1021/acssensors.3c02328

Zwier, M. V., Verhulst, E. C., Zwahlen, R. D., Beukeboom, L. W., & van de Zande, L. (2012). DNA methylation plays a crucial role during early Nasonia development. *Insect Molecular Biology*, *21*(1), 129–138. https://doi.org/10.1111/j.1365-2583.2011.01121.x