



VILNIUS UNIVERSITY  
LIFE SCIENCES CENTER

**LINAS AMBRAZIEJUS**

**The Effect of Chemoenzymatic 5hmC Derivatization on DNA  
Polymerase Readouts**

**Master Thesis**

Molecular Biotechnology Study Programme

*Thesis supervisor*

Dr. Miglė Tomkuvienė

*Thesis completed at*

Department of Biological DNA Modification  
of the LSC Institute of Biotechnology

Vilnius, 2025

# TABLE OF CONTENTS

LIST OF ABBREVIATIONS .....	3
INTRODUCTION.....	4
AIMS AND OBJECTIVES .....	6
1. LITERATURE REVIEW .....	7
1.1. Epigenetic mechanisms .....	7
1.2. Chemical and biological background of hydroxymethylcytosine .....	8
1.3. Detection of hydroxymethylcytosine.....	10
1.3.1. Chemical approaches .....	10
1.3.2. Enzymatic approaches .....	11
1.3.3. Affinity enrichment-based methods .....	13
1.4. Nanopore sequencing .....	13
1.5. Functions of 5hmC .....	15
1.6. The role of 5hmC in human diseases.....	16
1.7. Variety of DNA polymerases .....	17
2. MATERIALS AND METHODS.....	19
2.1. Instruments .....	19
2.2. Materials .....	19
2.3. Model DNA fragment amplification for modifications experiments .....	20
2.4. Hydroxymethylation of the substrate DNA.....	20
2.5. Guanine-cytosine adduct formation.....	21
2.6. Sample preparation for mass spectrometry .....	21
2.7. PCR of the DNA containing G <sup>A</sup> C adduct.....	21
2.8. Nanopore sequencing .....	22
2.8.1. Library preparation .....	22
2.8.2. Sequencing.....	23
2.9. G <sup>A</sup> C enrichment.....	23
3. RESULTS.....	25
3.1. Preparation of substrate DNA by PCR .....	25
3.2. 5hmC-modification of substrate DNA .....	26
3.3. Cytosine-guanine adduct formation.....	27
3.4. PCR processivity on derivatised DNA .....	30
3.5. Enrichment of G <sup>A</sup> C modified DNA for a more prominent evaluation of PCR effects.....	32
3.6. Nanopore sequencing .....	34
4. DISCUSSION .....	37
CONCLUSIONS .....	40
ABSTRACT .....	41
SANTRAUKA .....	42
AUTHOR'S PERSONAL CONTRIBUTION .....	43
ACKNOWLEDGMENTS.....	44
REFERENCES.....	45

## LIST OF ABBREVIATIONS

5caC – 5-carboxylcytosine

5fC – 5-formylcytosine

5hmC – 5-hydroxymethylcytosine

5mC – 5-methylcytosine

ACE-seq – APOBEC-coupled epigenetic sequencing

AID – activation-induced cytidine deaminase

APOBEC – apolipoprotein B mRNA editing enzyme

DNMT – DNA methyltransferase

EM-seq – enzymatic methyl sequencing

hMeDIP – hydroxymethylated DNA immunoprecipitation

HPLC-MS – high-performance liquid chromatography–mass spectrometry

MBD – methyl-CpG-binding domain

PCR – polymerase chain reaction

SAM – S-adenosylmethionine

T4-βGT – T4 beta-glucosyltransferase

TAB-Seq – TET-assisted bisulfite sequencing

TAPS – TET-assisted pyridine borane sequencing

TET – ten-eleven translocation

TLS – translesion synthesis

## INTRODUCTION

Deoxyribonucleic acid (DNA) in all living organisms is responsible for storing genetic information that controls the development, functioning, and reproduction of the cells. This polymer is made of monomeric units called nucleotides comprising 5-carbon sugar deoxyribose, a nitrogenous base and a phosphate group (reviewed in Minchin & Lodge, 2019). DNA is naturally and constantly modified: this phenomenon can impact chromatin composition, structure, as well as gene expression (reviewed in Ludwig et al., 2016). Among these modifications, 5-hydroxymethylcytosine (5hmC) is a stable epigenetic mark with distinctive genomic distribution and has its own unique functions (López et al., 2017).

5hmC is involved in critical biological processes such as embryonic development, cell differentiation, maintenance of cell identity, and ageing. Therefore, 5hmC is suggested as a new epigenetic biomarker for various cancers, including prostate cancer (Patel et al., 2025). There were close to 20 million new cases of cancer in the year 2022 alongside 9.7 million deaths (Bray et al., 2024). And these numbers tend to rise every year (Bizuayehu et al., 2024). For this reason, it is important to know the particular role of 5hmC in the development of various diseases.

In mammalian cells, 5hmC was first detected by thin-layer chromatography and then confirmed quantitatively by the high-performance liquid chromatography and mass spectrometry (HPLC-MS) methods. However, the latter methods are technically complex, require the use of expensive equipment and experienced scientists for accurate results. Also, it does not provide patterns of the distribution of 5hmC across the DNA sequence. Therefore, HPLC-MS was not suitable for clinical diagnostic applications (Kisil et al., 2024). Now, more than 10 various methods are used to differentiate unmodified cytosine from its epigenetic variants. These tests include whole-genome bisulfite sequencing, enzymatic methyl-sequencing (EM-seq) and TET-assisted pyridine borane sequencing (TAPS) (He et al., 2024). However, the gold-standard technologies to study DNA methylation do not always distinguish between 5hmC and a chemically similar DNA modification 5-methylcytosine (5mC) (Skvortsova et al., 2017).

Meanwhile, DNA polymerases catalyse the formation of phosphodiester linkages between nucleotides therefore synthesising DNA strands during replication (Lujan et al., 2016). These enzymes differ in several main properties, including fidelity (the accuracy of replication), processivity (the ability to synthesize long DNA strands), and thermal stability. It is known that unnatural modifications to the DNA, such as chemical adducts or other epigenetic changes can significantly affect the performance of DNA polymerases. A study has shown that adduct-induced changes in the dNTP base binding pocket geometry can lead to decreased fidelity and potential mutagenesis (S. F. Yan et al., 2004).

Previous works at the VU LSC IBT Department of Biological Modification have shown that 5hmC can be enzymatically derivatised with selenols (Liutkevičiūtė et al., 2011). The resulting compounds, when oxidised using sodium periodate, bind to the adjacent guanine base in GC sequences and finally form an intramolecular guanine-cytosine (G<sup>^</sup>C) adduct. We expect that this unnatural adduct of two subsequent bases may induce errors at G5hmC sites upon polymerase chain reaction (PCR), further allowing precise identification of 5hmC in sequencing reads. It is believed that this project can further show chemical plasticity of DNA and pave the way for a new chemoenzymatic 5hmC detection method.

## AIMS AND OBJECTIVES

Project aim:

To analyse the effect of chemoenzymatic 5hmC derivatization to G<sup>+</sup>C adduct on DNA polymerase readouts.

Project objectives:

1. To synthesize model 5hmC-modified DNA in vitro and monitor the modification yield by restriction protection assay;
2. To selectively derivatize 5hmC with a selenol group, followed by oxidation with sodium periodate to generate the G<sup>+</sup>C adduct, and to verify its formation using mass spectrometry;
3. To monitor DNA polymerase activity and accuracy on the generated model DNA substrate containing G<sup>+</sup>C adducts, by PCR and nanopore sequencing of the resulting products.

# 1. LITERATURE REVIEW

## 1.1. Epigenetic mechanisms

Epigenetics is the study of heritable and stable changes in gene expression caused by modifications to chromatin rather than changes in the DNA sequence itself. Although epigenetic mechanisms do not directly change the DNA sequence, they can influence gene expression by chemically modifying DNA bases and the chromosomal structure that organizes the DNA (Al Aboud et al., 2024). There are four main epigenetic mechanisms, including DNA methylation, histone modification, chromatin remodelling, and noncoding RNA (Y.-L. Wu et al., 2023).

DNA methylation is the addition of methyl groups (-CH<sub>3</sub>) to the 5' position of cytosine bases in DNA, usually at CpG dinucleotide sites (Joseph et al., 2018). There are approximately 0.6 billion cytosines in the human genome, and when both DNA strands are considered, 56 million of those are followed by guanines (CpGs) (Vaisvila et al., 2021). Around 80% of CpG dinucleotides in the mammalian genome are methylated and they often reside in clusters called CpG islands, which are generally unmethylated and associated with gene promoters (Zhao & Han, 2009). When these promoters become methylated, the associated gene is typically silenced or repressed (Chatterjee & Vinson, 2012). It has been established that adenine can also be methylated (Varma et al., 2022). The addition of methyl groups is catalysed by a family of enzymes called DNA methyltransferases (DNMTs). Structurally and functionally five different DNMT enzymes have been identified in mammals, including DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L, the latter being only a regulatory partner (Uysal et al., 2016). The methyl-CpG-binding domain proteins (MBD) recognize methylated cytosines and initiate signalling cascades that result in the formation of heterochromatin or euchromatin, thereby controlling gene expression (Coelho et al., 2022). DNA methylation is involved in many biological processes and is essential for controlling chromatin structure and gene expression (Hamidi et al., 2015). However, abnormal increases or decreases in DNA methylation contribute to ageing, the pathogenesis of brain disorders, cancer formation, tumour progression and other diseases (Ehrlich, 2019; Xie et al., 2023).

The dynamic structure of chromatin is another key determinant of gene expression, regulated extensively by histone modifications. To fit in the nucleus, DNA is coiled twice around octameric proteins called histones (H1-4), which facilitate a high degree of DNA organization and control accessibility to the genome (Pease et al., 2013). Highly condensed chromatin (heterochromatin) is inaccessible to transcription factors and transcription machinery, while open chromatin (euchromatin) is associated with transcriptional activation (Shariq & Lines, 2019). The tails of histones can undergo various chemical modifications, including acetylation, methylation, phosphorylation, and

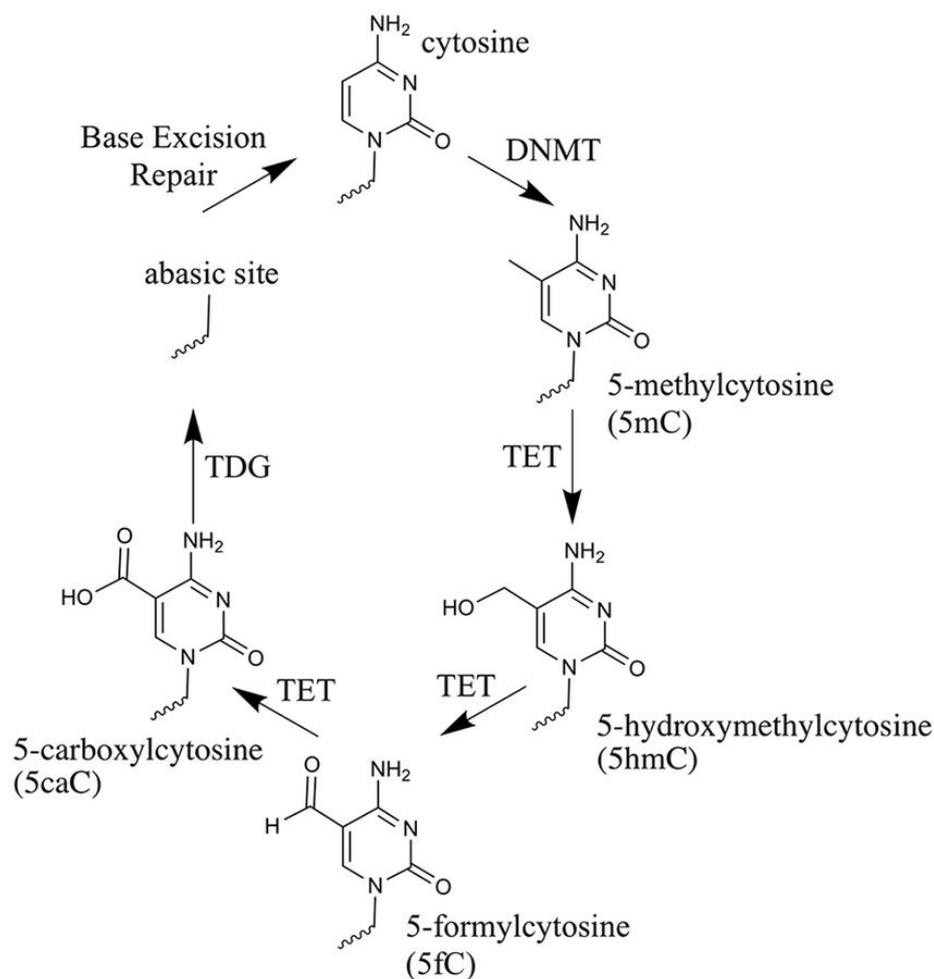
ubiquitination (R. Liu et al., 2023). These modifications are carried out by specific enzymes, such as histone acetyltransferases or histone methyltransferases, and are removed by histone deacetylases or demethylases (Alaskhar Alhamwe et al., 2018). The positively charged lysine residue in the histone tail is neutralized by acetylation, which reduces the strength of the bond between the histone tails and DNA. This process opens up the DNA/histone complex, making it more accessible to transcription factors (Hamilton, 2011). Histone methylation usually occurs on lysine and arginine and does not alter the charge of histones. Instead, it serves as docking sites for specific binding proteins called histone readers (Alam et al., 2015).

Chromatin remodelers regulate transcription and DNA repair by controlling access to genomic DNA (Reyes et al., 2021). There are four distinct families of chromatin remodelling complexes, based on ATPase subunit composition: switch/sucrose non-fermentable, imitation switch, inositol requiring 80, and chromodomain helicase DNA-binding complexes (T. Yang et al., 2022). Moreover, chromatin remodelling factors can control the expression of long noncoding RNAs, which are defined as non-protein coding transcripts longer than 200 nucleotides (P. Han & Chang, 2015). Long noncoding RNAs can interact with DNMT or TET family members to recruit these enzymes to or sequester from specific genome loci, resulting in promotion or repression, respectively, of the DNA methylation and demethylation pathways. They can also regulate transcription and translation of these proteins, thus affecting the DNA modification pathways indirectly too (Z. Yang et al., 2022).

## 1.2. Chemical and biological background of hydroxymethylcytosine

Although DNA methylation has long been considered a highly stable marker, the search for a definitive DNA demethylase has persisted. In 1953, a novel DNA modification form, 5-hydroxymethylcytosine (5hmC), was identified in the T-even bacteriophage (reviewed in Sun et al., 2014). Later, 5hmC was discovered in mammalian genomes. However, the mechanisms and proteins responsible for producing this DNA modification was discovered after the existence of 5hmC in mouse brain cells and mouse embryonic stem cells was re-confirmed (Liutkeviciute et al., 2009).

Now it is acknowledged that the DNA base 5hmC is produced by enzymatic oxidation of 5-methylcytosine (5mC) by 5mC oxidases (Hahn et al., 2014). It is a modified derivative of cytosine, distinguished by the presence of a hydroxymethyl group (-CH<sub>2</sub>OH) attached to the 5th carbon of the cytosine pyrimidine ring (K. Zheng et al., 2024). While 5mC is established and maintained by DNMTs, its further conversion to 5hmC is catalysed by the ten-eleven translocation (TET) family of enzymes (Madrid et al., 2018). TET enzymes are Fe<sup>2+</sup>- and  $\alpha$ -ketoglutarate-dependent oxidases, which perform efficient conversion both *in vitro* and *in vivo* (Vetř et al., 2018). Subsequent oxidation reactions by TET enzymes can further convert 5hmC to 5-formylcytosine (5fC) and then to 5-carboxylcytosine (5caC). This reaction sequence is shown in Figure 1.



**Figure 1. The cytosine methylation-demethylation cycle.** The chemical structures of cytosine, 5-methylcytosine and each oxi-mC within DNA (Jessop et al., 2018)

Furthermore, it was found that thymine DNA glycosylase can remove 5fC and 5caC, creating an abasic site that is further refilled with an unmodified cytosine, following a base excision repair pathway. This discovery enabled the identification of the TET-mediated active demethylation pathway (Jessop et al., 2018). The base excision repair pathway is important for maintaining both the genetic stability and the methylation status of CpG sites (Bellacosa & Drohat, 2015). The latter pathway might also provide a mechanism for reactivation of epigenetically silenced genes (Müller et al., 2021). An alternative pathway for reversal of 5mC could be achieved through the decarboxylation of 5caC to restore cytosine, as it occurs in various mammalian cell lines. Findings suggest that the TET-mediated oxidation of 5mC followed by direct decarboxylation of 5caC represents a novel pathway for active DNA demethylation in mammalian genomes (Feng et al., 2021). Additionally, enzymes activation-induced cytidine deaminase (AID) and apolipoprotein B mRNA editing enzyme (APOBEC) have been proposed to catalyse the conversion of the amino group of 5hmC to a carbonyl group. It results in the formation of 5-hydroxymethyluracil, which can then be converted to unmodified cytosine through the thymine DNA glycosylase-base excision repair pathway (K. Zheng et al., 2024).

Besides methods mentioned above, 5hmC can also be generated *in vitro* using methyltransferase-based methods. One of the methyltransferases used in this process could be M.HhaI, produced by the bacterium *Haemophilus haemolyticus*. DNA cytosine-5-methyltransferase M.HhaI along with HhaI work together to produce protective immunity against viral infections (Horton et al., 2020). M.HhaI recognizes the GCGC sequence and flips the inner cytosine out of DNA helix and into the catalytic site for methylation (Vilkaitis et al., 2000). This methyltransferase can also catalyse the substitution of sulfhydryl groups to hydroxyl groups in 5hmC (An et al., 2023). To create 5hmC, the reaction is carried out in with formaldehyde instead of the natural methyl donor, S-adenosylmethionine (SAM). Here, formaldehyde acts as a donor, enabling M.HhaI to transfer a hydroxymethyl group rather than a methyl group to the cytosine base (Liutkeviciute et al., 2009). Besides M.HhaI, another bacterial C5-MTase, M.SssI, also shows detectable catalytic activity at its target sites (Liutkevičiūtė et al., 2011).

### **1.3. Detection of hydroxymethylcytosine**

The detection of 5hmC at the genomic level remains challenging due to its structural similarity to 5mC, differing only by the presence of a hydroxyl group (-OH) to the methyl group of 5mC, which complicates their differentiation (K. Zheng et al., 2024). Current techniques for detecting DNA methylation can be categorised into four types: bisulfite sequencing, enzymatic approaches, affinity enrichment, and single-molecule techniques. Each type of strategy has its own distinct advantages and disadvantages (Zeng et al., 2015).

#### **1.3.1. Chemical approaches**

For decades, bisulfite sequencing has been the gold standard for mapping DNA modifications (Y. Liu et al., 2019). This method relies on treating DNA with sodium bisulfite, which converts unmethylated cytosine residues to uracil while leaving 5mC unaffected. The methylation status can then be analysed through direct PCR sequencing or cloning sequencing (Y. Li & Tollefsbol, 2011). However, the process involves a harsh chemical reaction that degrades most of the double-stranded DNA, leading to significant information loss (Cao et al., 2023). Also, this method cannot distinguish between 5mC and 5hmC because both resist bisulfite mediated deamination (C. Nestor et al., 2010). For this reason, TET-assisted bisulfite sequencing (TAB-Seq) was developed, incorporating two enzymatic steps prior to bisulfite conversion (Yu et al., 2018). First, 5hmC is selectively glucosylated, making it resistant to further oxidation by TET proteins. Next, 5mC is oxidized to 5caC, is easily deaminated during standard bisulfite treatment. Lastly, 5hmC is sequenced as C, whereas both C and 5mC are sequenced as T, therefore distinguishing 5hmC sites from C and 5mC sites (Yu et al., 2012).

Another similar chemical approach is the oxidative bisulfite sequencing. It was one of the first techniques that enabled absolute quantification of 5mC and 5hmC at single-base resolution (De Borre & Branco, 2021). Here DNA is treated with an oxidizing agent (potassium perruthenate) to convert 5hmC to 5fC. The 5hmC signal is eliminated, thereby, exclusively detecting 5mC. Then, the sequencing results are compared to those from standard bisulfite sequencing to quantify 5hmC (Kirschner et al., 2018; Lee, 2024). In addition, an ultrafast bisulfite sequencing was presented, which uses highly concentrated bisulfite reagents and high reaction temperatures to accelerate the bisulfite reaction by 13-fold. This method minimizes DNA damage and reduces background noise (Dai et al., 2024).

A novel TET-assisted pyridine borane sequencing (TAPS) method involves TET proteins to oxidize mC and 5hmC into 5caC, which is subsequently reduced to dihydrouracil using pyridine borane and sequenced as thymine (T). The key advancements of pyridine-borane chemistry include its ability to directly read modified bases while preserving unmodified cytosine intact and its less destructive nature, which enhances sequencing quality, mapping accuracy, and coverage (reviewed in Kriukienė et al., 2024). Correspondingly, TAPS $\beta$  includes an additional beta-elimination step to further improve the differentiation of 5hmC from 5mC.  $\beta$ -glucosyltransferase ( $\beta$ GT) selectively blocks 5hmC by glucose labelling, followed by TET oxidation and pyridine borane reduction to target 5mC. This step enhances the ability to map and quantify these modifications with greater precision with minimal false positives (Y. Liu, Hu, et al., 2021).

### 1.3.2. Enzymatic approaches

Various methods have been developed progressively to overcome the limitations of the bisulfite conversion method. Recent research indicates that DNA deaminases from the AID/APOBEC family can differentiate between various cytosine modification states, opening new possibilities for sequencing (Lee, 2024). Using the AID/APOBEC family DNA deaminase enzyme as a deamination reagent, a bisulfite-free method detects 5hmC in single-base resolution. T4 beta-glucosyltransferase (T4- $\beta$ GT) catalyses the glucose attachment that protects 5hmC (X. Li et al., 2021). This method, called enzymatic methyl-seq (EM-seq) uses two sets of enzymatic reactions. In the first reaction, TET2 and T4- $\beta$ GT convert 5mC and 5hmC into products that cannot be deaminated by APOBEC3A. In the second reaction, APOBEC3A deaminates unmodified cytosines by converting them to uracils. Therefore, these three enzymes enable the identification of 5mC and 5hmC (Vaisvila et al., 2021). APOBEC-coupled epigenetic sequencing (ACE-seq) can also detect 5hmC as well as distinguish it from 5mC and unmethylated C but does it more precisely than EM-seq. The APOBEC3A enzyme selectively deaminates unmodified cytosine, and converts it to uracil, while leaving 5hmC untouched (Schutsky et al., 2018). EM-seq follows the same workflow as ACE-seq but uses TET2 to oxidize

5mC and 5hmC prior to glucosylation, whereas ACE-seq directly detects 5hmC using APOBEC3A alone (J. J. N. Li et al., 2024).

Similarly, a third enzymatic method called direct enzymatic sequencing has been developed to distinguish 5mC from unmethylated C and 5hmC. In this approach, an engineered methyltransferase uses a SAM analog to create a modified cytosine base resistant to A3A deamination. Methyltransferases with carboxymethyltransferase activity have been developed, enabling efficient use of carboxy-SAM to produce A3A-resistant 5-carboxymethylcytosine at unmethylated cytosines. T4- $\beta$ GT then protects 5hmC, while A3A selectively deaminates 5mC, converting it to thymine during sequencing (Lee, 2024; T. Wang et al., 2023). The main chemical and enzymatic approaches are compared in Figure 2. It concludes that BS-Seq converts cytosine to uracil (sequenced as T), but it does not differentiate between 5mC and 5hmC (both sequenced as C); TAB-Seq enzymatically protects 5hmC, converting it to T while keeping 5mC as C; oxidative bisulfite sequencing chemically modifies 5hmC, leaving only 5mC as C; TAPS and TAPS- $\beta$  enzymatic methods directly sequence 5hmC as T and differentiate it from 5mC, while TAPS- $\beta$  focuses on 5mC-specific detection; ACE-Seq and EM-Seq enzymatic methods offer sensitive and precise 5hmC detection; and direct enzymatic sequencing detects 5mC directly without interference from 5hmC.

	CpG	5mCpG	5hmCpG	<u>protection/</u> <u>modification step</u>	<u>deamination</u> <u>step</u>	<u>bases confounded</u> <u>with 5mC</u>	<u>bases directly</u> <u>detected by C-&gt;T</u>
BS-Seq	T	C	C	N/A	chemical	5hmC	C
TAB-Seq	T	T	C	enzymatic	chemical	C	C, 5mC
oxBS-Seq	T	C	T	chemical	chemical	none	C, 5hmC
TAPS	C	T	T	enzymatic	chemical	5hmC	5mC, 5hmC
TAPS- $\beta$	C	T	C	enzymatic	chemical	none	5mC only
ACE-Seq	T	T	C	enzymatic	enzymatic	C	C, 5mC
EM-Seq	T	C	C	enzymatic	enzymatic	5hmC	C
DM-Seq	C	T	C	enzymatic	enzymatic	none	5mC only

Sequence as:

**Figure 2. Comparison of methods for detecting C, 5mC, and 5hmC based on protection, modification, and deamination strategies.** BS-Seq – bisulfite sequencing, TAB-Seq – TET-assisted bisulfite sequencing, oxBS-Seq – oxidative bisulfite sequencing, TAPS – TET-assisted pyridine borane sequencing, ACE-seq – APOBEC-coupled epigenetic sequencing, EM-seq – enzymatic methyl sequencing, DM-seq – direct enzymatic sequencing (T. Wang et al., 2023)

### 1.3.3. Affinity enrichment-based methods

The three most commonly used affinity-based enrichment techniques are antibody, chemical capture and protein affinity enrichment. They have an ability to accurately and reproducibly report 5hmC (Thomson et al., 2013). Hydroxymethylated DNA immunoprecipitation (hMeDIP) uses specific antibodies that recognize 5hmC to immunoprecipitate DNA fragments containing 5hmC (C. E. Nestor & Meehan, 2014). The comparative hMeDIP-seq method improves conventional hMeDIP-seq by enabling direct comparison of DNA hydroxymethylomes across samples. DNA from different samples is sonicated, barcoded with unique adaptors, and pooled for a single hMeDIP reaction. Following immunoprecipitation and sequencing, reads are organized by barcode and aligned to the genome. This method reduces the experimental variation among samples and allows direct comparison of the DNA hydroxymethylation data across samples (Tan et al., 2013).

To enrich DNA fragments containing 5hmC, scientists use 5hmC chemical labelling (hmC-Seal) strategy, a previously developed method for efficient, unbiased, genome-wide labelling and covalent capture of 5hmC. First, the T4- $\beta$ GT adds an engineered glucose molecule with an azide group to the 5hmC in DNA. Then, a biotin tag is attached to the azide using Huisgen Cycloaddition chemistry. The biotin-tagged 5hmC DNA fragments are captured with avidin beads. Finally, the captured DNA is amplified by PCR and sequenced (D. Han et al., 2016). hmC-Seal has higher sensitivity than hMeDIP-seq, enabling the capture of regions with very low 5hmC content from as few as 1,000 cells (Lee, 2024).

Methyl-CpG-binding domain sequencing (MBD-seq) offers potential advantages compared to antibody-based enrichment (Aberg et al., 2018). Here, the genomic DNA is fragmented and incubated with methyl-binding domain 2 (MBD2) protein or MBD2-based capture beads. MBD2 binds to methylated CpG sites, enriching for methylated regions of the genome. The MBD domain of MBD3 displays preferential binding to 5hmC by electrophoretic mobility shift assays (K. Liu et al., 2018).

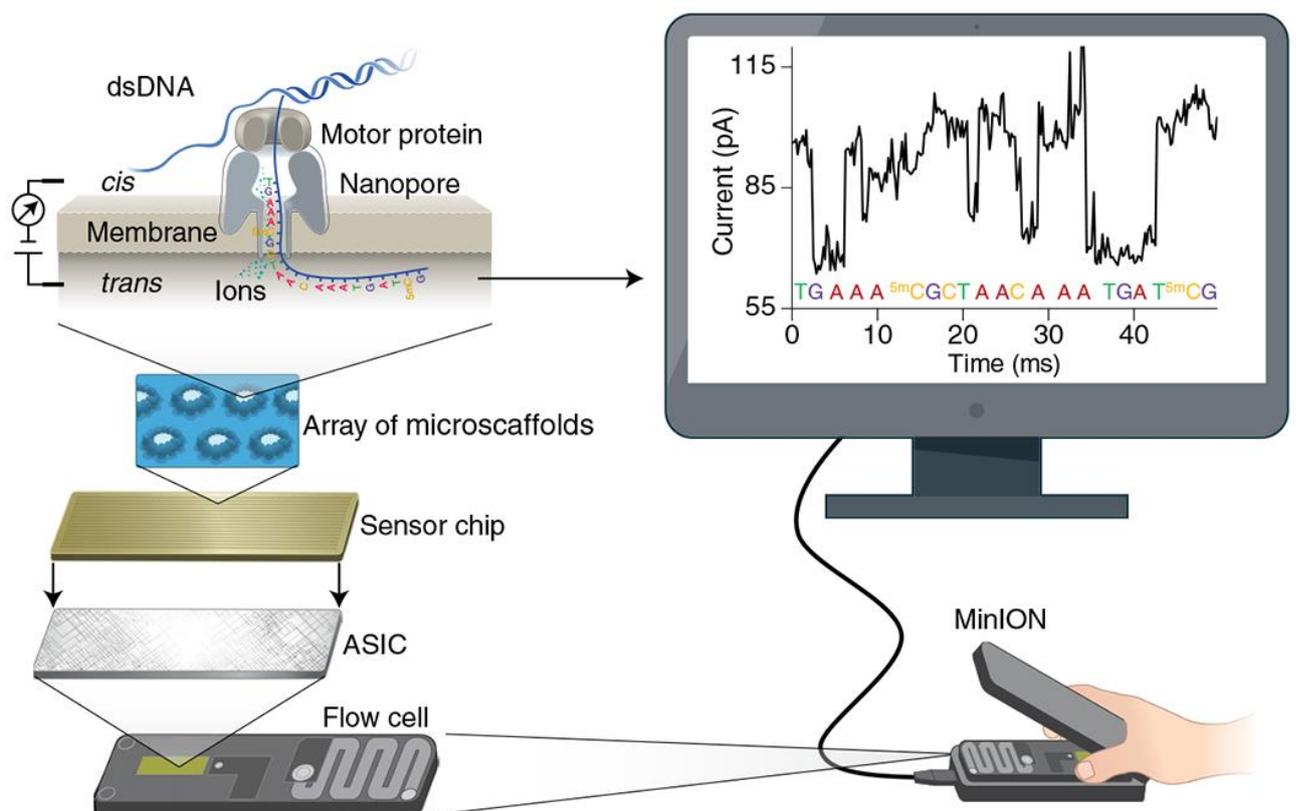
However, affinity-based methods have been criticized for their lack of single-base resolution and bias toward heavily modified regions (Aberg et al., 2020).

## 1.4. Nanopore sequencing

Nanopore sensing is a powerful single-molecule technique based on detecting a modulation in the ionic current due to the partial blockade of nanopores caused by analytes (Zeng et al., 2015). Sometimes it is described as a third-generation sequencing method (Lee, 2024). In nanopore sequencing, nanopores act as biosensors, fixed on a resistive film. Electrodes at both ends form a stable electric field, moving nucleic acids (DNA/RNA) through the nanopores. Motor proteins control the speed of passage. As nucleic acids pass through, changes in nanopore charge alter electron flow, generating unique signals based on base composition or modifications. These signals are recorded

and analysed by the sequencer to determine the base sequence (P. Zheng et al., 2023). In Figure 3 it is schematically shown how nanopore sequencing works.

The main advantages of nanopore sequencing include the ability to directly distinguish signals of nucleic acid modifications. 5mC and 5hmC influence the electronic currents in the pore differently as DNA passes through: 5mC consistently increases the current compared to C, while 5hmC generally decreases the current relative to C. This difference is used to discriminate 5hmC, 5mC, and 5C based on electric signal deviations. (Y. Liu, Rosikiewicz, et al., 2021). DNA methylation calling from Oxford Nanopore sequencing now matches bisulfite sequencing in single base-pair accuracy while offering a portable, low-cost, and rapid workflow for real-time analysis. DNA methylation is a promising circulating cell-free DNA biomarker and is in widespread testing as a cancer screening tool. The turnover of damaged cells in time-sensitive conditions like myocardial infarction, sepsis, and COVID-19 can also be detected using DNA methylation (Katsman et al., 2022). The limitation of the direct DNA modification detection by nanopore sequencing is that the DNA cannot be amplified, potentially restricting input amounts to the microgram scale. However, these methods are still undergoing active development (Lee, 2024).



**Figure 3. Principle of nanopore sequencing and direct DNA modification detection (Y. Wang et al., 2021)**

## 1.5. Functions of 5hmC

Firstly, 5hmC exerts a regulatory role in gene promoters. Its enrichment in these regions highlights its capacity to promote demethylation and activation processes, significantly influencing transcriptional activity and the transcriptional state of genes (K. Zheng et al., 2024). In a 2023 study it was discovered that 5hmC alone can regulate gene expression as a valid epigenetic mark in proliferative somatic cells (Wei et al., 2023). Secondly, 5hmC is significantly enriched at genomic regions that have histone modifications associated with enhancers, such as H3K4me1 and H3K27ac. Moreover, there is a possibility that 5hmC may alter chromatin structure (H. Wu et al., 2011). It is enriched in euchromatin, which is associated with active genes, while 5mC is found in heterochromatin and is associated with gene silencing. The distribution of these methylation marks aligns with specific histone modifications: 5mC with repressive marks (H3K9me3 and H3K27me3) and 5hmC with active marks (H3K4me2) (Y. Chen et al., 2014). 5hmC is also enriched in other protein-DNA interaction sites, such as OCT4 and NANOG binding sites (Stroud et al., 2011). Recent studies have shown that 5hmC enrichment is linked to active gene expression, correlating with marks such as H3K4me3 and RNA polymerase II, and can be used in advanced computational models to predict gene expression (Gonzalez-Avalos et al., 2024). It is known that 5hmC interacts with specific binding proteins, primarily transcriptional regulators. Unlike 5mC, which binds to a different set of proteins, 5hmC is recognized by repair and regulatory proteins, including ribosomal protein L26, pre-mRNA processing factor 8, and the DNA mismatch repair protein malignant hyperthermia susceptibility 6 (Pang et al., 2016). Overall, 5hmC is essential for transcriptional regulation and impacts cellular function and development.

In mammals, 5hmC is involved in methylation reprogramming during early embryonic development (R. Yan et al., 2023). The TET proteins (TET1, TET2, and TET3) are essential for cell reprogramming. During this time 5hmC modification is increased, and knockout of TET proteins prevents the process. TETs are thought to reactivate the OCT4 gene by demethylating its promoter and enhancer regions, with TET1 capable of replacing OCT4 in the OSKM reprogramming cocktail (Basu & Tiwari, 2021). Additionally, TET-mediated oxidation maintains the demethylated state of regulatory regions, particularly enhancers and promoters, which are key targets in reprogramming. This process is important for regulating gene expression and ensuring proper cell fate decisions during development (reviewed in Kriukienė et al., 2024).

Without the things mentioned above, 5hmC is believed to be involved in epigenetic changes during cancer progression, embryonic growth and cellular differentiation (Tong et al., 2024). Furthermore, various data have shown that 5hmC may have functions in the chromatin structure, splicing, self-renewal, transcription, cell adhesion, cell death, development, differentiation, maturation, cytoskeleton, ion transport, and myelopoiesis (J. Wang et al., 2014).

## 1.6. The role of 5hmC in human diseases

The highest known levels of 5hmC are found in brain and in embryonic stem cells (Booth et al., 2013). TAB-Seq showed that 5hmC is 10-fold more abundant in central nervous system and embryonic stem cells than in peripheral tissues (Shi et al., 2017). 5hmC content in brain cells increases with age, suggesting that it is linked to neurodevelopment (Yao et al., 2024). Interestingly, available data indicate that a number of environmental factors, such as stress, exercise, diet modifications, and exposure to exogenous chemicals, are linked to variations in 5hmC levels in different brain regions (Kochmanski & Bernstein, 2020). Several studies strongly indicate the dysregulation of 5hmC could be involved in multiple diseases such as mentioned in Table 1.

**Table 1.** The differential alteration of 5hmC in neurological disorders (Sun et al., 2014)

<b>Diseases</b>	<b>The alteration of 5hmC</b>
Autism (autism spectrum disorders)	Enrichment on autism related genes
Rett syndrome	Global decrease in the genome
Angelman syndrome	Global increase in the genome
Fragile X syndrome	Enriching in disease related genes
Alzheimer's disease	Decrease or increase in the genome
Huntington's disease	Global decrease

Alzheimer's disease is a progressive and most common type of incurable dementia afflicting more than 40 million people globally (Bomasang-Layno & Bronsther, 2021). Several laboratories have demonstrated that human post-mortem brain samples and Alzheimer's disease mouse models exhibit changed 5hmC amounts, and that the genes associated with these changes may regulate Tau-mediated neuronal toxicity (Armstrong et al., 2023). 5hmC markers derived from plasma cell-free DNA can serve as effective, minimally invasive biomarkers for clinical auxiliary diagnosis of late-onset Alzheimer's disease (L. Chen et al., 2022).

Like 5mC, global averages of 5hmC are reduced in cancer tissues. However, unlike 5mC, the enrichment of 5hmC at gene promoters and coding sequences is linked to increased gene expression (Mitrea et al., 2018). Dot blot hybridisation and immunohistochemistry analyses have shown significantly lower 5hmC levels in melanoma, breast, prostate, colon, liver, lung and pancreatic cancer compared with the adjacent normal tissues. This reduction is likely due to impaired TET enzyme activity or decreased expression of TET proteins (Skvortsova et al., 2019). Recent studies have shown that 5hmC patterns in cell-free DNA reflect a critical role in gene expression regulation, as well as in the carcinogenesis of multiple solid tumours (H. Chen et al., 2024). However, in one study nanopore sequencing analysis revealed no detectable difference in global 5hmC content

between healthy and tumour tissue. It suggests that 5hmC changes may not be associated with early-stage breast cancer and instead are a downstream consequence of the disease (Zahid et al., 2024). To conclude, 5hmC plays a crucial role in the development of various diseases, including cancer and neurological disorders. However, further research is needed to determine its full significance and potential as a biomarker for cancer diagnosis and prognosis.

### **1.7. Variety of DNA polymerases**

The primary function of DNA polymerases is to accurately and efficiently replicate the genome to successfully pass on the genetic information through generations (reviewed in Garcia-Diaz & Bebenek, 2007). DNA polymerases are classified into several families (A, B, C, D, X, Y, RT) based on primary amino acid sequence similarities. Their performance varies remarkably across biological contexts in terms of fidelity and processivity. The best-known and one of the earliest DNA polymerase-based biotechnology applications is PCR (reviewed in Gardner & Kelman, 2014). The ability of a DNA adducts to induce mutations is influenced by a few factors including the DNA sequence context, the fidelity of DNA polymerases in translesional synthesis, and the adduct structure. It is known that some bulky groups can block the DNA polymerase and cause frameshift mutations (Hwa Yun et al., 2020).

DNA polymerase fidelity is vital for accurate genome replication, but sometimes, replication errors can contribute to evolution and heritable diseases. Polymerases use mechanisms like nucleotide selectivity and proofreading to maintain high accuracy during this process (de Paz et al., 2018). High-fidelity polymerases replicate complementary DNA strands during S-phase with a low error rate and a fast rate of synthesis (Kaszubowski & Trakselis, 2021). Some of them, such as  $\delta$  and  $\epsilon$  in eukaryotes have 3'→5' exonuclease proofreading activity that corrects misincorporated nucleotides immediately after insertion (Zhou et al., 2021). On the other hand, they are unable to proceed with DNA synthesis when base damage within a template strand occurs (Kaszubowski & Trakselis, 2021). This often leads to replication fork uncoupling, activation of the replication checkpoint and replication stress (Sellés-Baiget et al., 2025). When DNA is damaged, specialized translesion synthesis (TLS) polymerases help in replication substituting the standard high-fidelity polymerase. TLS polymerases have low fidelity and poor processivity. The absence of proofreading activity and the existence of a flexible catalytic site that can accept damaged DNA bases are the reasons for this (de Paz et al., 2018; Sellés-Baiget et al., 2025). Unfortunately, the ability of TLS DNA polymerases to tolerate DNA damage may allow cancer cells to continue proliferating and remaining viable. Now scientists try to use TLS pathways for therapeutic purposes. Interestingly, TLS mechanisms can bypass bulky DNA adducts (Anand et al., 2023).

Organisms distribute various tasks among different DNA polymerases to balance genome stability with the ability to tolerate and bypass DNA damage. Five DNA polymerases are present in *Escherichia coli* (*E. coli*), and the primary replicase of the cell, polymerase III holoenzyme is the cell's main replicase responsible for chromosome duplication by simultaneous coordinated leading and lagging strand synthesis (Vaisman et al., 2021). It is considered a highly accurate enzyme. In contrast, *E. coli* polymerase IV and polymerase V, both belonging to the Y-family, are considered low-fidelity DNA polymerases because of their lack of proofreading activity and their high error rates (Kuban et al., 2004). In eukaryotes, polymerase  $\eta$  plays a critical role in bypassing thymine dimers - lesions induced by ultraviolet radiation. The loss of this polymerase in humans is responsible for the variant of xeroderma pigmentosum (Powers et al., 2018). In this context, at dimer-lesion sites, DNA synthesis is stopped. Then, more prone TLS polymerases carry out bypasses, and mutations build up, which can later cause skin cancer (Bedaiwi et al., 2024). Similarly, DNA polymerase  $\beta$  is involved in the repair of DNA damage during the base excision repair pathway. Located in the centre of this pathway, Pol  $\beta$  finds its substrates quickly to prevent further damage. In searching mode, Pol  $\beta$  uses positively charged lysine residues in the lyase domain to hop along DNA. Upon finding a gap, the lyase domain binds exposed nucleobases, increasing DNA affinity and enabling 31-kDa domain engagement, which initiates damage recognition (Howard et al., 2017). So, both error-prone and high-fidelity polymerase activities are essential to maintain genetic stability for all living organisms – from bacteria to mammals.

Taq DNA polymerase is isolated from the thermophilic bacterium *Thermus aquaticus* (Hernández-Rollán et al., 2024). This polymerase is inactive at low temperatures below 30 °C and its activity continues to increase up to 72 °C (Xue et al., 2021). It also has a half-life of 40 minutes at 95°C and only 5–6 minutes at 97.5°C (Sundarrajan et al., 2018). According to experiments, the error rate for Taq polymerase is  $4.3 \times 10^{-5} \pm 1.8$  mutations/bp per template duplication (McInerney et al., 2014). In this project, a commercial DreamTaq polymerase is used, which according to the manufacturer's protocol has a higher sensitivity and higher yields compared to conventional Taq DNA polymerase. While it has 5'→3' exonuclease activity, it lacks 3'→5' proofreading activity, resulting in relatively low fidelity (Tamás et al., 2022). Another polymerase that is used in this project and does not have 3'→5' proofreading exonuclease activity is DeepVent (exo-) (Lapa et al., 2022). Meanwhile, DeepVent (exo-) polymerase is derived from the archaeon *Pyrococcus* species (Jannasch et al., 1992). According to the manufacturer's protocol, it has a half-life of 23 hours at 95 °C and it's ideal for GC-rich sequences. For this reason, the two polymerases mentioned above were chosen for this project – which lack 5'→3' proofreading activity and could introduce mistakes in the process of DNA replication at chemoenzymatic derivatives of 5hmC.

## 2. MATERIALS AND METHODS

### 2.1. Instruments

Centrifuge “5417R” (Eppendorf)

Centrifuge Microspin FV-2400

Computer “GP66 Leopard 11UH-454NL” (MSI)

Electrophoresis power supply Consort E802

Fluorometer “Qubit 4” (Invitrogen)

Gel imaging system “ChemiDoc” (Bio Rad)

Magnetic stand “DynaMag-2” (Invitrogen)

Microwave oven “BI7036” (LG Intellrowave)

Multi-function mixer “Multi Bio RS-24” (Biosan)

Nucleic acid sequencing system: flongle flow cell, flongle adapter and MinION sequencing system (Oxford Nanopore Technologies)

Scales “GR-202-EC” (A&D Instruments)

Spectrophotometer “NanoDrop 2000” (Thermo Fisher Scientific)

Thermocycler “Labcycler” (SensoQuest)

Thermostat “Fisherbrand 75L Incubator” (Thermo Scientific)

Thermostat FB15101 (Thermo Scientific)

UV lamp “RS 1” (Haiser)

Water filtration system Ultrapure SQPAK™ Mili-Q SQ 2 Series

### 2.2. Materials

*Honeywell*: 96% ethanol, NaOH.

*Roth*: APS, TEMED, acrylamide-bisacrylamide solution Rotiphorese gel 40% (19:1), ethidium bromide, SDS, NaAc, DTT, 90mM boric acid, 2mM EDTA.

*Sigma-Aldrich*: formaldehyde, selenocystamine dihydrochloride.

*Thermo Fisher Scientific*: TopVision agarose tablets, 2mM dNTP, 6x TriTrack loading dye, GeneRuler DNA Ladder Mix, T4 DNA ligase, proteinase K, DreamTaq polymerase, Phusion™ Plus DNA Polymerase (high fidelity) reaction kit, FastAP phosphatase, restriction endonucleases: R.Hin6I, R.EheI; GeneJET PCR purification kit, DNA length standard marker O'GeneRuler DNA Ladder Mix, DNA length standard marker MassRuler Low Range, pUC19 plasmid, 90mM Tris-base, Fast Digest Green buffer.

*Oxford Nanopore Technologies*: DNA Ligaton Sequencing Kit NBD112.24 sequencing library preparation kit.

*New England Biolabs*: ThermoPol buffer, DeepVent (exo-) polymerase, P1 nuclease.

*Zymo Research*: Genomic DNA Clean & Concentrator, DNA Clean & Concentrator-5 kits.

### 2.3. Model DNA fragment amplification for modifications experiments

A 1252 bp fragment was amplified from the pUC19 plasmid using a polymerase chain reaction (PCR). The selected primers F: GATACCGCTCGCCGCAG and R: CACTATTCTCAGAATGACTTGGTTGAG were designed to cover a region with high GCGC content – 10 target sites in total. Amplification was performed using the Phusion™ Plus DNA Polymerase reaction kit following the manufacturer’s protocol. The PCR reactions were carried out using standard cycling conditions appropriate for Phusion™ Plus polymerase (Table 2). To determine PCR condition, a  $T_m$  calculator was used <https://www.thermofisher.com/lt/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html>. The amplified products were subsequently analysed using 1% agarose gel electrophoresis in 1X TBE buffer (90mM Tris-base, 90mM boric acid, 2mM EDTA, pH=8 regulated with NaOH) to confirm the presence and size of the expected 1252 bp fragment. The gel electrophoresis was set to run for a standard duration of 35-45 minutes at 150V with 350A. Samples were dyed with Tritrack loading dye prior to gel loading. Gel was dyed with ethidium bromide for 10 minutes and analysed with Biorad ChemiDoc imaging system. Band sizes were evaluated with Generuler mix ladder. After confirming the presence of the fragment, DNA was purified with GeneJET PCR Purification Kit columns. Later, purified DNA was measured with NanoDrop 2000 spectrophotometer (162 ng/μl) and used as a template for further experiments.

**Table 2.** Phusion™ Plus DNA PCR conditions:

Temperature, °C	Duration	Number of cycles
98	30 s	1
98	10 s	35
60	10 s	
72	60 s	
72	300 s	1

### 2.4. Hydroxymethylation of the substrate DNA

2000 ng of substrate DNA was mixed in a 30 μL reaction with 10x TEN buffer (final concentration 1x), 1% formaldehyde solution (final concentration 0.1%) and 140 mM methyltransferase HhaI (final concentration 2.8 mM; courtesy of Giedrė Urbanavičiūtė). The reaction

was left to proceed overnight at room temperature to ensure complete hydroxymethylation of the DNA. Following overnight incubation, the reaction was further processed with incubation at 65°C for 20 minutes, addition of 10% SDS (final concentration 0.133%) and 0.4 µL Proteinase K. Later, the sample was incubated at 55 °C for 1 hour. Then, substrate DNA was purified using the same DNA purification kit and its concentration measured with Nanodrop.

DNA digestion using restriction enzyme R.Hin6I and Fast Digest Green buffer was performed according to manufacturer recommendations. The amount of purified PCR product used for restriction was ~50 ng. The sample was incubated at 37 °C for 30 minutes before performing electrophoresis with the same conditions as described for PCR products above.

## **2.5. Guanine-cytosine adduct formation**

1M DTT (final concentration 12.5 mM), 125 mM selenol (selenocystamine dihydrochloride; final concentration 25 mM), 10x sodium citrate (pH 5.5) and 10x 0.2 g/ml BSA were incubated together for 20 minutes at room temperature. The selenol attachment to substrate DNA occurred when 5000 ng of substrate (modified DNA) and 140 mM M.HhaI (final concentration 7 mM) were added to the mixture and incubated for 1.5 hours at room temperature. After that, DNA was purified using Zymo purification kit (elution with 50 µL nuclease-free water). Then, 2 µL of sodium periodate and 2 µL of sodium phosphate were added before incubating 40 µL of mixture for 1.5 hours at room temperature in the dark. After this step, the DNA was purified again using the same DNA purification kit.

## **2.6. Sample preparation for mass spectrometry**

At least 8 pmol of target sites were used for analysis. Online calculator [https://www.geneinfinity.org/cc/cc\\_dnaconverter.html](https://www.geneinfinity.org/cc/cc_dnaconverter.html) was used to calculate the amount of DNA needed for the assay. The samples were mixed with P1 buffer (100 mM NaAc, pH 5.5, 10 mM ZnAc; final concentration 1x) and 0.5 U/µL P1 and incubated for 2 hours at 55 °C. Then, FastAP (final concentration 1x) was added and the samples incubated further overnight at 37 °C. The next day the samples were heated for 15 minutes at 75 °C to deactivate FastAP and centrifugated for 30 minutes at 4 °C (13000 rpm). The supernatant was used for HPLC-MS performed by Audronė Rukšėnaitė.

## **2.7. PCR of the DNA containing G<sup>A</sup>C adduct**

After the existence of G<sup>A</sup>C adduct was confirmed with mass spectrometry, the PCR with DeepVent (exo-) polymerase was performed. Also, the same PCR conditions were used with DreamTaq polymerase and the combination of both (Table 3). The amplified DNA fragments were

analysed by agarose gel electrophoresis to verify their size and integrity. After that, the samples were sequenced (2.8).

**Table 3.** ThermoPol PCR conditions:

Temperature, °C	Duration	Number of cycles
95	120 s	1
95	30 s	40
60	30 s	
72	90 s	
75	300 s	1

## 2.8. Nanopore sequencing

### 2.8.1. Library preparation

The library was prepared using the “Native Barcoding Kit 24 V14 (SQK-NBD114.24)” DNA sequencing kit (Oxford Nanopore) and additional enzymes (New England Biolabs) by manufacturer’s protocol: <https://nanoporetech.com/document/ligation-sequencing-amplicons-native-barcoding-v14-sqk-nbd114-24>. Around 120 ng of DNA were used of each sample and all reagents were kept on ice. Each end-prep reaction contained 12.5 µL of DNA sample (adjusted by adding nuclease-free water), 1.75 µL of Ultra II End-Prep Reaction Buffer and 0.75 µL of Ultra II End-Prep Enzyme Mix, bringing the total volume to 15 µL. This step prepares the ends of DNA fragments to be compatible for adapter ligation. The mixture was thoroughly mixed by pipetting, spun down, and incubated at 20 °C for 5 minutes, followed by 65 °C for 5 minutes. The samples were then transferred to clean 1.5 mL Eppendorf DNA LoBind tubes.

For the DNA clean-up, the AMPure XP Beads (AXP) were resuspended by vortexing, and 15 µL was added to each reaction, mixed by flicking, and incubated on a rotating mixer for 5 minutes at room temperature. The magnetic particles were precipitated using a magnetic stand, the supernatant was removed, and the particles were washed twice using 500 µL of 70% ethanol, ensuring the pellet remained undisturbed. After that, the particles were dried for 1 minute, ensuring they did not crack. Later, the beads were resuspended in 10 µL of nuclease-free water, incubated for 2 minutes at room temperature, and pelleted on a magnetic stand. Finally, 10 µL of the eluate was transferred into a clean 1.5 mL Eppendorf DNA LoBind tube, and the pelleted beads were discarded.

In further steps, barcodes are ligated to the DNA. 7.5 µl of the DNA obtained in the previous step is transferred to a new “Low Bind” tube. 2.5 µL of Native Barcode, and 10 µL of Blunt/TA Ligase Master Mix are added. The reaction was gently mixed, briefly spun down, and incubated for 20 minutes at room temperature. The reaction was stopped by adding 2 µl of EDTA, followed by thorough mixing and brief centrifugation. All barcoded samples were pooled into a 1.5 mL Eppendorf

DNA LoBind tube and, depending on the number of samples, 8  $\mu$ l of AXP magnetic particles were added. DNA was purified and eluted analogously to the previous step.

For adapter ligation, 30  $\mu$ L of pooled barcoded sample, 5  $\mu$ L of Native Adapter, 10  $\mu$ L of NEBNext Quick Ligation Reaction Buffer (5X), and 5  $\mu$ L of Quick T4 DNA Ligase were added to a fresh tube. The reaction was gently mixed, briefly spun down, and incubated for 20 minutes at room temperature. The DNA mixture was purified by adding 20  $\mu$ l of AXP magnetic particles and incubated on rotating mixer for 10 minutes at room temperature. The tube was placed in a magnetic stand and the particles were pelleted. After removing the supernatant, the particles were washed using 125  $\mu$ l of Long Fragment Buffer (LFB) and spun down. Any residual buffer was removed after an additional spin.

The beads were resuspended in 15  $\mu$ L of Elution Buffer (EB), spun down, and incubated at 37°C for 10 minutes, with gentle flicking every 2 minutes to encourage DNA elution. The beads were pelleted on a magnet for at least 1 minute until the eluate was clear and colourless. Then, 15  $\mu$ L of eluate containing the DNA library was transferred to a clean 1.5 mL Eppendorf DNA LoBind tube, and the beads were discarded.

The final concentration of the library is measured with a Qubit fluorometer using a broad-spectrum dgDNA dye. The library can be used further for sequencing or frozen at -80°C

### **2.8.2. Sequencing**

DNA library sequencing was performed using the ONT MinION instrument and Flongle flow cells. Before sequencing, the flow cell and the number of intact pores were checked. 4 ng of freshly prepared or frozen library is diluted to 12  $\mu$ L, 37.5  $\mu$ L of sequencing buffer (SBII) and 25.5  $\mu$ L of library loading particles (LBII) are added. The library is gently mixed by pipetting and immediately loaded onto the flow cell. This action was done slowly to avoid air bubbles. The obtained sequencing data were analysed on the IVG 2.19.1 platform.

### **2.9. G<sup>A</sup>C enrichment**

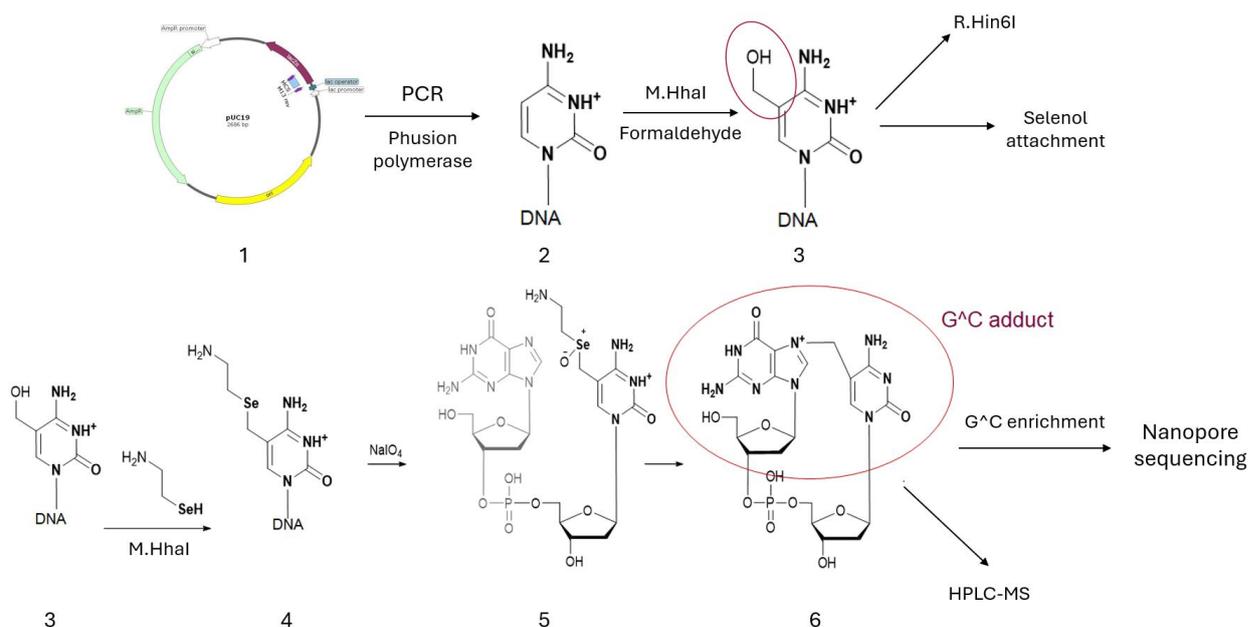
G<sup>A</sup>C enrichment was performed using R.EheI enzymatic digestion and PCR techniques. Initially, samples were digested in a 40  $\mu$ l reaction containing 0.4  $\mu$ l 10x R.EheI enzyme, the required amount of Fast Digest Green buffer and nuclease-free water. The amounts of DNA varied: control (~100 ng), 5hmC (~100 ng), Se (~10 ng), GC (~99 ng). Digestion was performed for 30 min. at 37 °C, after which the samples were purified with Zymo DNA purification kit and eluted into 20  $\mu$ l water.

A 25  $\mu$ l volume PCR reaction with two combined polymerases (Deep Vent (exo-) and DreamTaq) was performed. Amplification conditions were the same as in previous steps with ThermoPol buffer. The resulting PCR products were digested with R.EheI again, to leave intact only the ones that had altered sequences and therefore resistance to R.EheI treatment. The resulting intact 1252 bp fragment was excised from the 5% PAA gel and diffused into 1 ml of TE buffer overnight and purified with Zymo DNA purification kit.

## 3. RESULTS

### 3.1. Preparation of substrate DNA by PCR

For this study, we selected a model DNA substrate from pUC19 plasmid featuring 10 GCGC target sites for M.HhaI, with a length suitable for detection and analysis via PCR, gel electrophoresis, and nanopore sequencing. The main steps of this project are shown in Figure 4.

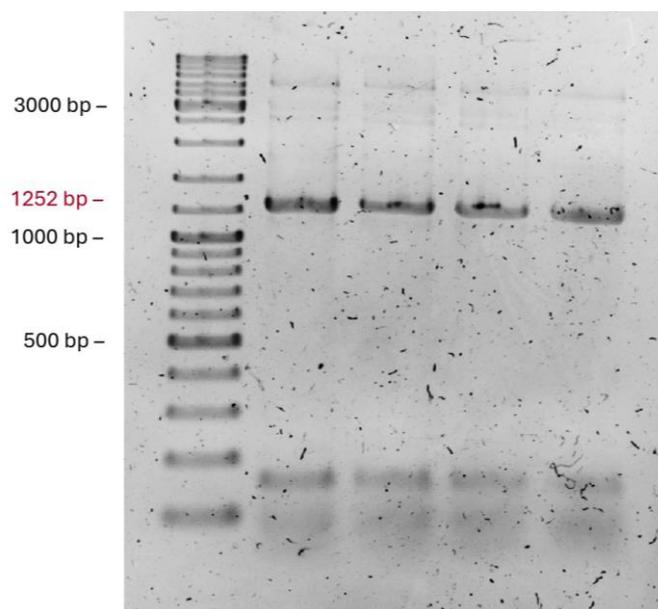


**Figure 4. Overview of the project workflow.** 1 – 2686 bp pUC19 plasmid, 2 – cytosine in 1252 bp DNA sequence, 3 – 5-hydroxymethylcytosine, 4-5 – intermediate selenol compounds, 6 – G<sup>^</sup>C adduct. Each step is followed by DNA purification

PCR amplification with Phusion polymerase was performed to obtain a model DNA fragment. This high-fidelity polymerase was selected to avoid any errors, enabling selective detection of mutations caused by the adduct in subsequent experiments. The success of amplification was assessed via agarose gel electrophoresis. As shown in Figure 5, a distinct band was observed at approximately 1250 bp, corresponding to the expected 1252 bp size of the amplified product. Four replicates of PCR reactions were made to produce more DNA substrate. Although primer dimers were detected, but the prominent intensity of the target bands suggested that this PCR product could be purified using Thermofisher DNA purification kit and used for further experiments.

Following PCR, the DNA was purified using a silica column-based method to remove excess primers, nucleotides, and polymerase. The purified DNA was quantified using a NanoDrop spectrophotometer, which indicated a concentration of 127.4 ng/ $\mu$ L and an A<sub>260</sub>/A<sub>280</sub> ratio of 1.85. These values confirm high purity and minimal protein contamination. However, an A<sub>260</sub>/A<sub>230</sub> ratio

was lower than expected – 1.48. It indicates that there were still some salts left from the PCR. Total amount of substrate DNA produced was around 8000 ng, enough for further experiments.



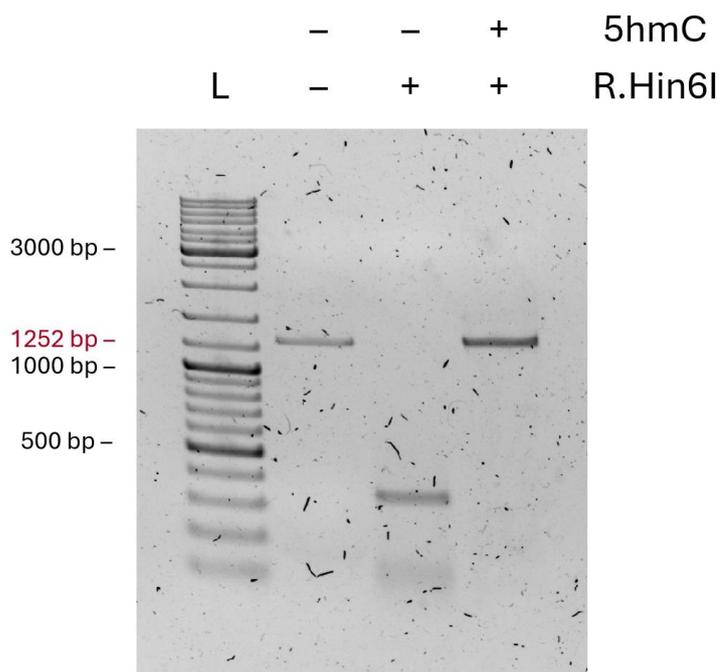
**Figure 5. 1252 bp amplification product from pUC19 plasmid.** Production of substrate DNA by phusion polymerase; four identical reactions. Ladder – GeneRuler DNA Ladder Mix

### 3.2. 5hmC-modification of substrate DNA

To introduce 5hmC into the DNA, the previously produced PCR product was treated with methyltransferase HhaI and formaldehyde for the unconventional enzymatic reaction to directly convert cytosine to 5hmC at M.HhaI target sites GCGC (the underlined C is modified) (Liutkevičiūtė et al., 2011).

5hmC modification protects DNA hydrolysis by restriction enzyme Hin6I, recognising the same target sequence GCGC. This was leveraged to estimate the yield of modification. Therefore, aliquots of the modified DNA were subjected to R.Hin6I digestion. In the control unmodified DNA sample, complete digestion was observed, as evidenced by the appearance of the expected fragments (~300 bp and less) on the gel (Figure 6). In contrast, the 5hmC-modified DNA was undigested, meaning that nearly 100% of GCGC sites were protected from digestion.

5hmC-modified DNA was purified using Thermofisher DNA purification kit and used for the following reaction steps.

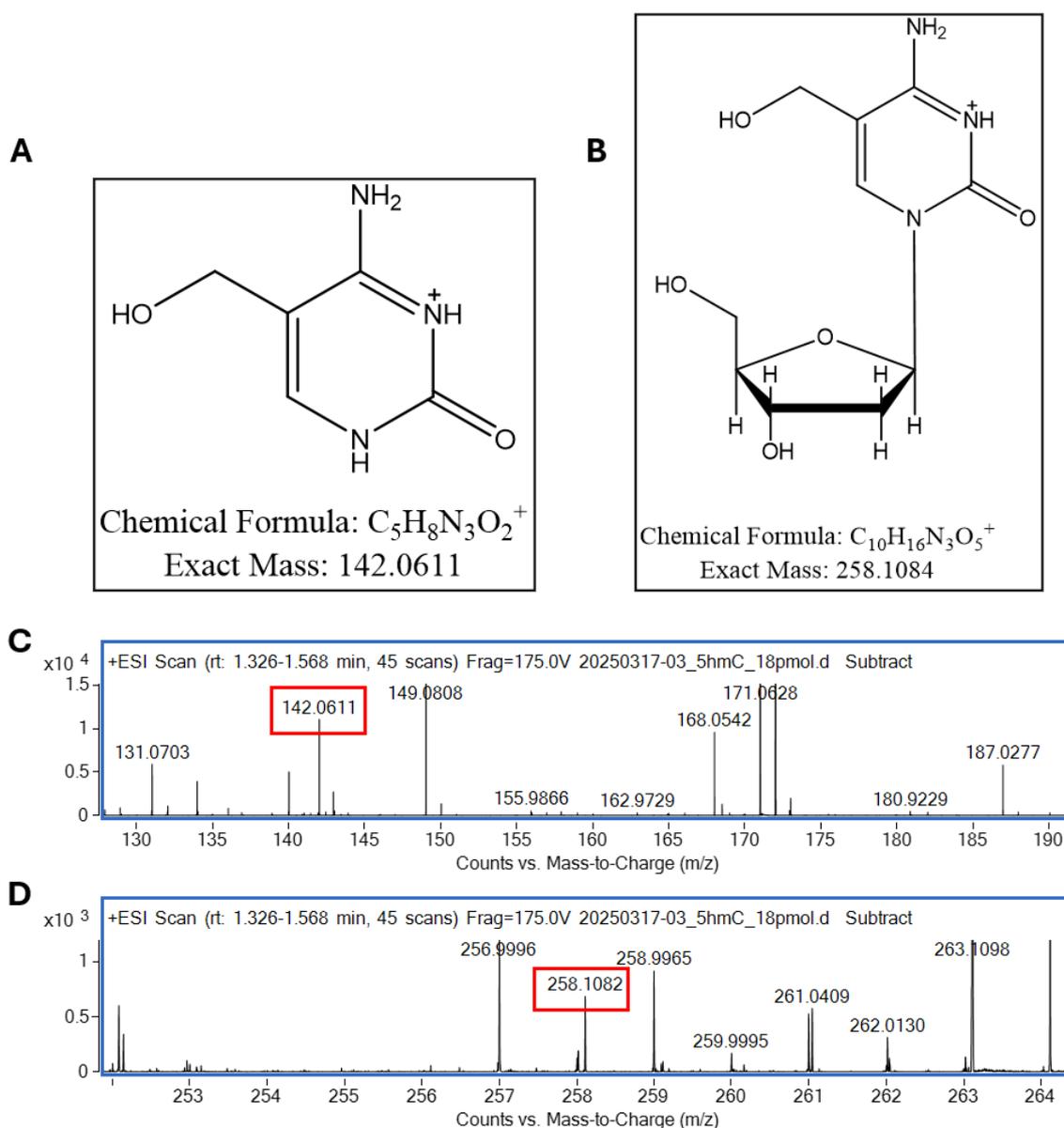


**Figure 6. Comparison of R.Hin6I digestion products in modified and unmodified DNA samples.** 5hmC-modified substrate DNA is protected from digestion by a cognate restriction enzyme. Ladder – GeneRuler DNA Ladder Mix

### 3.3. Cytosine-guanine adduct formation

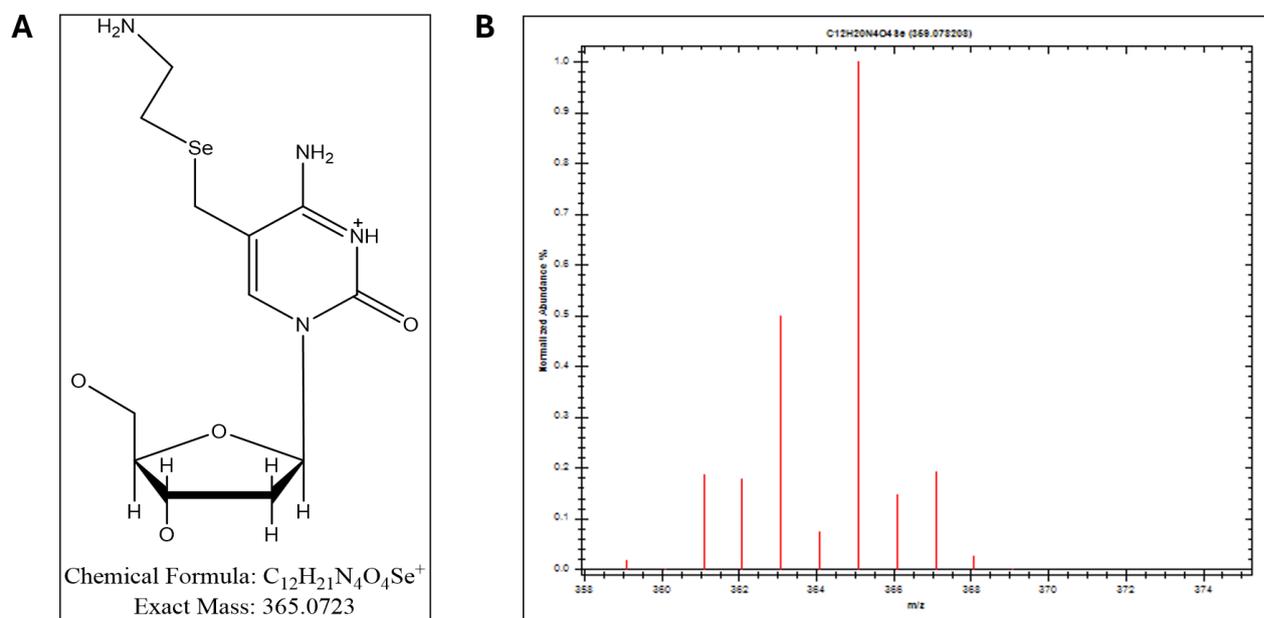
After optimizing the ratio of DTT and selenol (selenocystamine dihydrochloride), adduct formation began by incubating these two reagents along with sodium citrate buffer, and BSA at room temperature for 20 minutes. Substrate DNA and M.HhaI were then added to enable selenol attachment to the modified DNA, followed by a 1.5 hours incubation. DNA was purified, then oxidized with sodium periodate and sodium phosphate in the dark for another 1.5 hours. A second purification completed the adduct formation process.

Following the chemical modifications of 5hmC, four samples were analysed using HPLC-MS to assess the resulting composition and modifications: control unmodified, 5hmC, selenol intermediate product 5-(2-aminoethylseleno)methyl-dC and G<sup>+</sup>C adduct. The presence of 5hmC modification was already confirmed using restriction enzyme digestion. To further validate chemical modification, 18 pmol of the sample was subjected to molecular weight analysis. The detected [M+H]<sup>+</sup> mass for 5hmdC was 258.1082, which closely aligned with the theoretical value of 258.1084, and the detected [M+H]<sup>+</sup> mass for 5hmC base was 142.0611, which matched with its theoretical value (142.0611), confirming successful modification (Figure 7). The RT for 5hmdC was approximately 1.5 minutes.



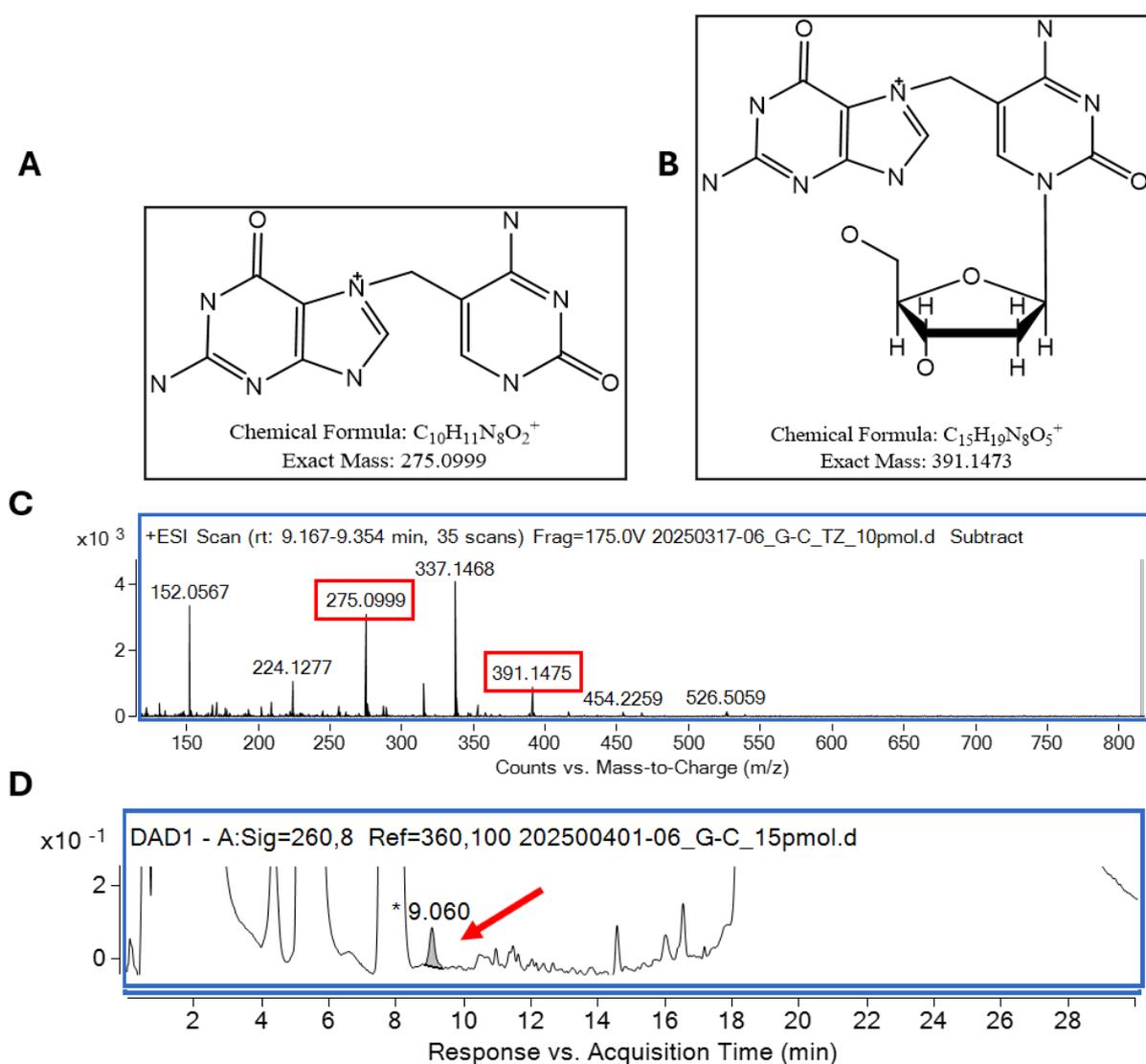
**Figure 7. HPLS-Mass Spectrometry profiles of 5hmC.** A – chemical formula and exact mass of protonated fragment of 5hmC molecule, B – chemical formula and exact mass of a full protonated 5hmC compound. Panels C and D show molecular identity using mass spectrometry, with specific ions (m/z 258.1082 and 142.0611) representing 5hmC. ESI Scan – electrospray ionization, rt – retention time, frag – fragmentor voltage

The same testing was made with 5hmC sample that underwent selenol addition reaction (10 pmol). Unfortunately, no intermediate selenol compounds with a theoretical  $[M+H]^+$  mass of 365,0723 were seen in the HPLC-MS analysis (Figure 8). The intermediate compound was not detected as a distinct monoisotopic peak in the HPLC-MS data, likely due to the presence of multiple naturally occurring isotopes leading to signal dispersion across several m/z values. However, its existence was confirmed because later the G<sup>^</sup>C adduct was formed.



**Figure 8. 5-(2-aminoethylseleno)methyl-dC.** A – chemical formula and exact mass of a full protonated intermediate compound, B – isotopic distribution

10 pmol of oxidized DNA was used in the analysis. The detected  $[M+H]^+$  mass for dG<sup>^</sup>C was 391.1475, which closely aligned with the theoretical value of 391.1473, and the detected  $[M+H]^+$  mass for G<sup>^</sup>C base was 275.0999, which matched with its theoretical value (275.0999), confirming successful adduct formation (Figure 9A, 9B and 9C). The retention time for G<sup>^</sup>C was approximately 9.2 minutes. Diode array detector (DAD) chromatogram uses UV absorbance to monitor the elution of compounds from HPLC. Figure 9D shows chromatographic separation and UV-detectable presence of the G<sup>^</sup>C adduct, with a clear peak at ~9 min., and further confirms its detection.

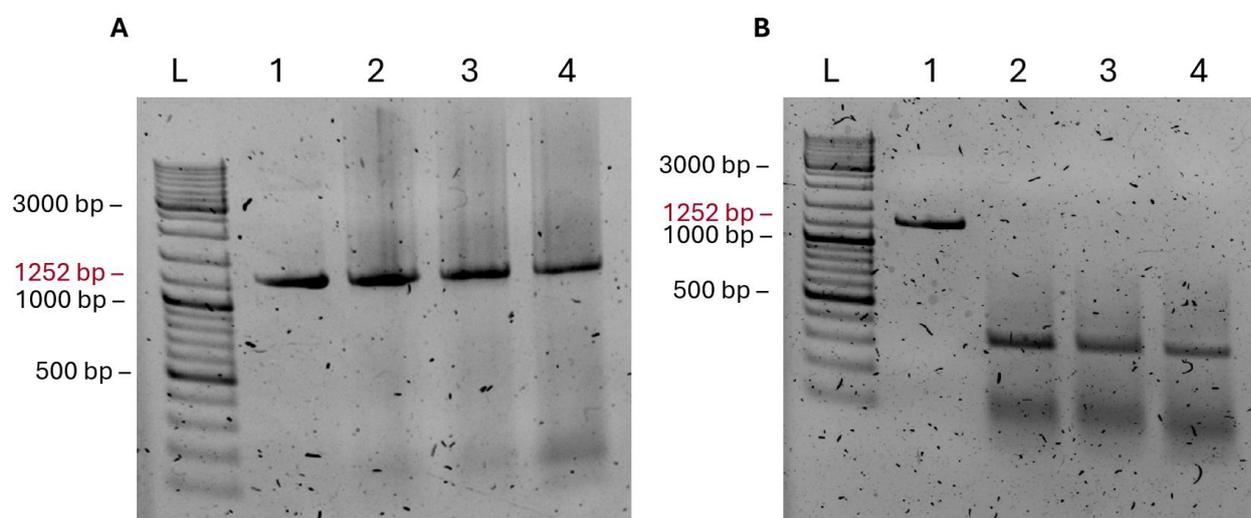


**Figure 9. HPLS-Mass Spectrometry profiles of G<sup>C</sup> adduct substrate DNA.** A – chemical formula and exact mass of protonated fragment of guanine-cytosine adduct, B – chemical formula and exact mass of a full protonated adduct compound, C – molecular identity using mass spectrometry, with specific ions (m/z 391.1475 and 275.0999) representing G<sup>C</sup> adduct, D – chromatographic separation and UV-detectable presence of the product. ESI Scan – electrospray ionization, rt – retention time, frag – fragmentor voltage, DAD – diode array detector

### 3.4. PCR processivity on derivatised DNA

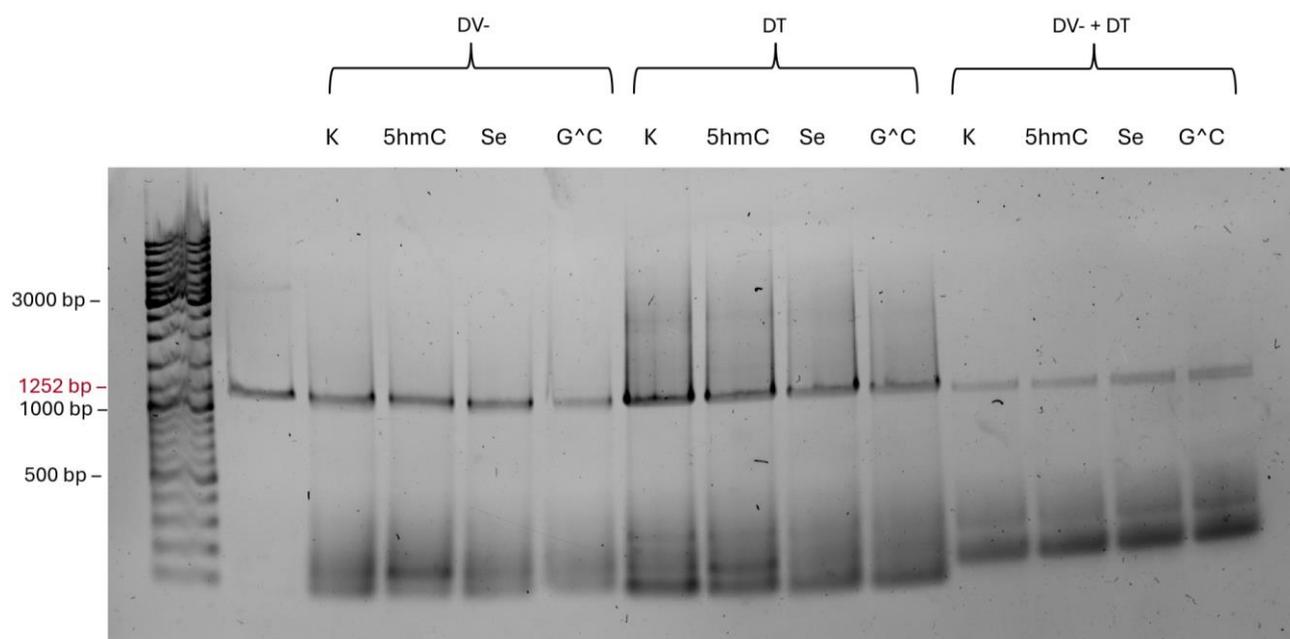
Initially, different quantities of reagents were used to attach selenol to 5hmC compared to those described in the Methods section (65 mM DTT instead of 12.5 mM and 12.5 mM selenol instead of 25 mM; final concentrations). PCR was performed using Deep Vent (exo-) polymerase to evaluate the amplification efficiency of modified DNA samples with chemical steps versus unmodified samples (Figure 10A). Analysis of the PCR products using ImageLab software revealed a significant reduction in product yield from the modified sample, suggesting decreased polymerase efficiency on derivatized DNA. Based on band intensity measurements, the control sample was equated to 100%, the unmodified sample yielded 94.94%, and the modified sample was only 69.00%. These results

indicate that the chemical steps involved in chemoenzymatic derivatization may reduce PCR efficiency. However, no difference in digestion patterns was observed post-restriction enzyme treatment, implying that no significant errors (insertions or deletions) were introduced during amplification (Figure 10B). This may suggest that the majority of amplified molecules originated from unmodified templates.



**Figure 10. Agarose gel analysis.** A – PCR with DeepVent (exo-) polymerase of modified and unmodified samples B – Enzymatic digestion by R.Hin6I of the PCR products shown in panel A. 1 – control sample without PCR, 2 – control unmodified after PCR, 3 – chemically oxidized control sample after PCR, 4 – 5hmC after chemical oxidation and PCR. Ladder – GeneRuler DNA Ladder Mix

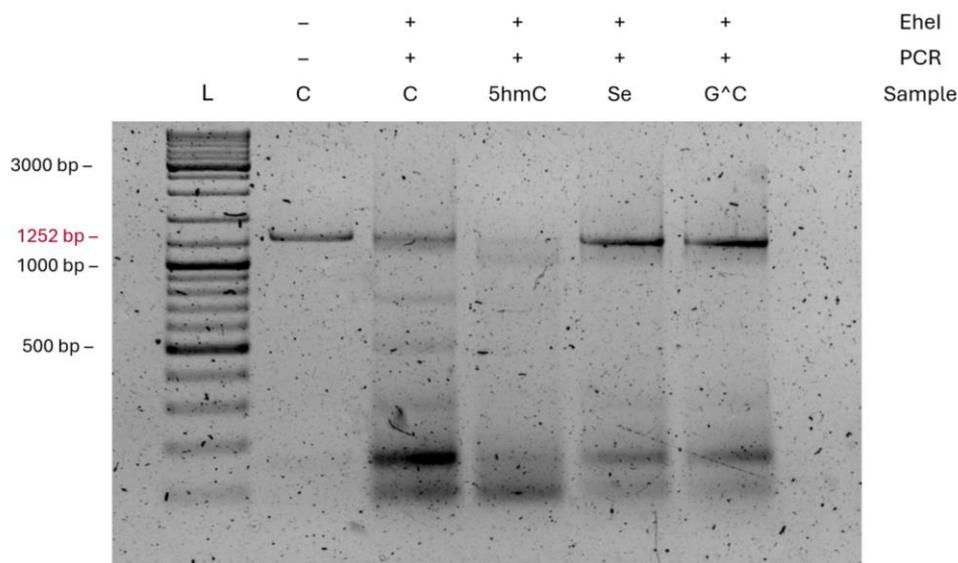
Further, 5hmC derivatization conditions were optimised and after the G<sup>^</sup>C adduct was confirmed by HPLC-MS, the PCR productivity and fidelity on the derivatised DNA was assessed again including additional polymerase DreamTaq, alone and in combination with DeepVent (exo-). Figure 11 gel electrophoresis image shows comparison of the performance and fidelity of two DNA polymerases: DeepVent (exo-) (DV-) and DreamTaq (DT), as well as their combined application (DV+ DT), across various DNA modifications: unmodified control (K), 5hmC-modified, selenol-attached DNA (Se), and a G<sup>^</sup>C context. The band at approximately 1252 bp represents the expected product size. Under DV- alone, all templates show strong and specific amplification with minimal smearing, indicating high fidelity and efficiency. G<sup>^</sup>C adduct substrate sample yielded noticeably less PCR product compared to the others. DT alone also shows strong amplification, but with slightly less product in the Se and G<sup>^</sup>C samples. The combination of DV- and DT results in less intense and more diffuse bands, indicating reduced amplification efficiency when both enzymes are used together.



**Figure 11. Evaluation of polymerase performance on derivatized model DNA.** DV- – DeepVent (exo-) polymerase, DT – DreamTaq polymerase, K – unmodified control, 5hmC – 5-hydroxymethylcytosine-modified, Se – selenol-attached DNA, G<sup>^</sup>C – cytosine-guanine adduct containing substrate. Ladder – GeneRuler DNA Ladder Mix

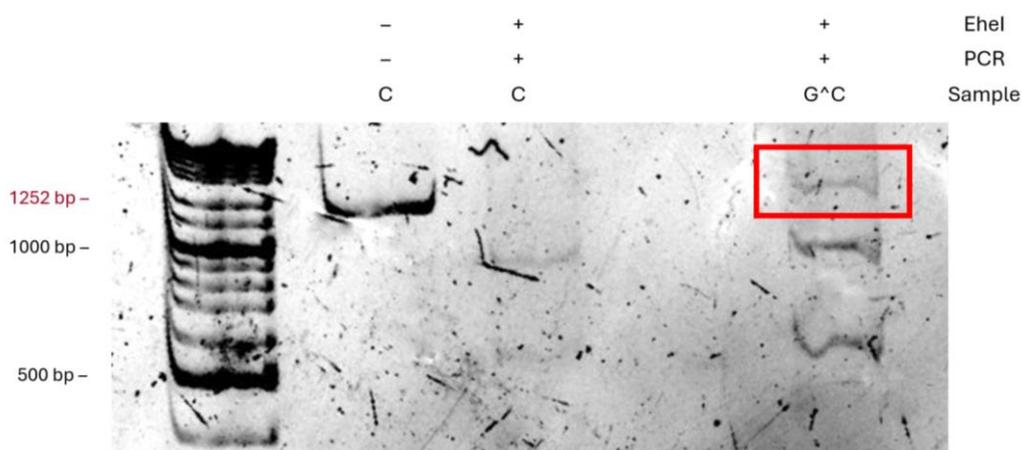
### 3.5. Enrichment of G<sup>^</sup>C modified DNA for a more prominent evaluation of PCR effects

The estimated G<sup>^</sup>C yield, based on HPLC-MS results was rather low, approximately 1% of all GCGC targets. To enrich the occurrence of the G<sup>^</sup>C adduct at a higher frequency, ideally at least one in ten target sites, the restriction enzyme R.EheI was selected. It recognizes the sequence GGC<sup>^</sup>GCC, which overlaps with the M.HhaI recognition site at position 509 bp of the substrate DNA. R.EheI enzymatic digestion of modified DNA removes full length substrates that are not derivatised at R.EheI target, therefore only modified ones can be further amplified by PCR. After enzymatic treatment and purification, sample DNA (G<sup>^</sup>C, together with unmodified (C), 5hmC and Se samples) was amplified using the combined two-enzyme (Deep Vent (exo-) and DreamTaq) system. All samples were successfully amplified, except from 5hmC sample which suffered from an unidentified problem (Figure 12). Unmodified control was also unexpectedly amplified, most probably due to incomplete digestion.



**Figure 12. PCR Amplification of substrate DNA after restriction-protection-based enrichment of modifications at R.EheI target sites.** Polymerases: combined DeepVent (exo-) and DreamTaq. Ladder – GeneRuler DNA Ladder Mix

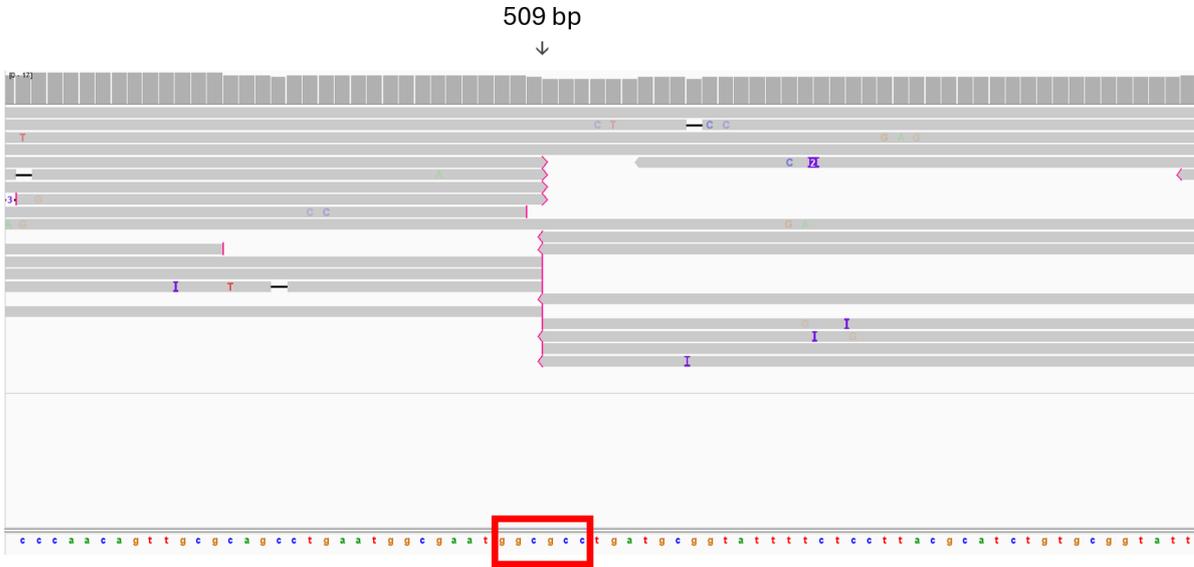
The resulting PCR products were subjected to R.EheI digestion once again, this time to evaluate if the modified target sites resulted in sequence alterations after PCR, which should also protect from R.EheI digestion if the sequence alteration disrupts the target sequence. In the subsequent digestion reactions, a 1252 bp fragment was observed in the PAA gel only in the G<sup>^</sup>C sample. This fragment remained uncut even after prolonged exposure to the enzyme, indicating that the G<sup>^</sup>C-modified DNA bore a PCR product resistant to R.EheI. Fragments from other samples were completely or almost completely digested. The 1252 bp fragment from the G<sup>^</sup>C sample was excised from the gel (Figure 13), eluted, and column-concentrated for further analyses (final concentration – 2.6 ng/μL).



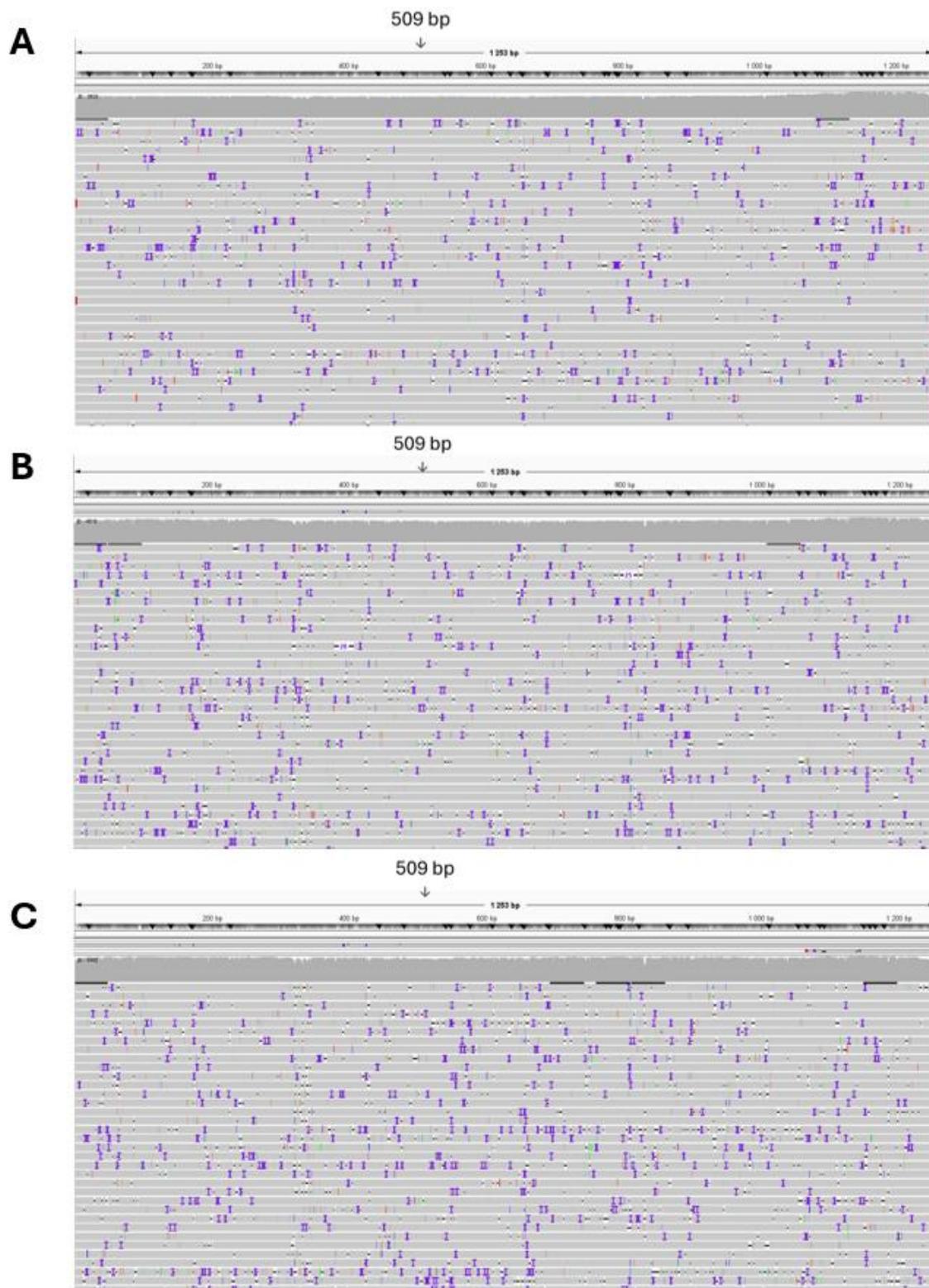
**Figure 13. Selective escape of PCR products of G<sup>^</sup>C-modified DNA from R.EheI digestion in 5% PAA gel.** Uncut full-length fragment is visible only in the G<sup>^</sup>C sample, which is circled in red. This fragment was cut out of the gel for further analysis. Polymerases: DeepVent (exo-) and DreamTaq. Ladder – GeneRuler DNA Ladder Mix

### 3.6. Nanopore sequencing

The derived PCR product was further subjected to nanopore sequencing to detect the exact sequence alterations at the R.EheI target site that were induced by G<sup>^</sup>C adduct. The sequencing image (Figure 14) shows a pattern of soft clipping at the site where the restriction enzyme R.EheI recognizes the sequence GGC<sup>^</sup>GCC. In this modified G<sup>^</sup>C adduct sample, the majority of aligned reads exhibit soft clipping precisely at this recognition site. Soft clipping typically occurs when the sequenced reads cannot be aligned to the reference genome and recognised by IGV program. The disruption observed here suggests that the recognition site may be heavily mutated. Importantly, this pattern is absent in the control samples (Figure 15), indicating that the effect is specific to the G<sup>^</sup>C sample only.



**Figure 14. G<sup>^</sup>C sample nanopore sequencing result analysis in IGV.** Detected error-induced soft clipping disruption of R.EheI (GGC<sup>^</sup>GCC) site at 509 bp



**Figure 15. Nanopore sequencing of the control samples, analysis in IGV.** No specific systemic errors are seen at or around the 509 bp site (A – control unmodified, B – 5hmC, C – selenol-attached sample). The control samples were prepared directly from PCR products after modification enrichment (no second enrichment of the alters targets, as only the G<sup>A</sup>C sample escaped full digestion at that step)



## 4. DISCUSSION

The aim of this project was to evaluate the impact of chemoenzymatic derivatization of 5hmC on DNA polymerase activity and nanopore sequencing readouts. With the experiments that included PCR, enzymatic and chemical derivatization, mass spectrometry, restriction digestion, and nanopore sequencing, we demonstrated successful formation of G<sup>^</sup>C adduct and its impact on fidelity and processivity of DeepVent (exo-) and DreamTaq polymerases.

In the beginning, PCR amplification yielded a high-quality 1252 bp DNA fragment from pUC19 plasmid, suitable for enzymatic and chemical modification as well as further analysis assays. 5hmC was introduced enzymatically using M.HhaI and formaldehyde, and its functional incorporation was assessed by restriction enzyme digestion. The 5hmC-modified DNA displayed complete resistance to Hin6I digestion, in contrast to unmodified DNA, which was fully digested into smaller fragments. This effect occurs when cytosine is replaced by 5hmC in the GCGC site, so Hin6I fails to recognize the modified site properly and does not cut. This result confirms efficient M.HhaI-mediated hydroxymethylation.

Following successful 5hmC creation, chemical derivatization was performed to generate a 5-(2-aminoethylseleno)methyl-dC intermediate compound, with the ultimate goal of forming a G<sup>^</sup>C adduct upon oxidation with sodium periodate. Mass spectrometry confirmed the molecular identity of both 5hmdC and its base component, showing close alignment between the theoretical and observed values. These data support the successful production and structural integrity of 5hmC. Although the intermediate selenol compound was not observed as a distinct one isotopic peak (likely due to isotopic dispersion and detection limits), the subsequent successful formation of the G<sup>^</sup>C adduct supports the fact that at least a small quantity of intermediate compound had been formed. At first, neither the selenol intermediate compound nor the adduct was confirmed by mass spectrometry, because different proportions of reaction compounds were used: 65 mM DTT and 12.5 mM selenol (final concentrations). Later, it was assumed that DTT potentially binds together with other components and prevents adduct formation, therefore the reaction was done without it, but the results of mass spectrometry were negative again. After that, it was decided to use only freshly prepared reagents and effective reaction conditions described in the Methods section. Finally, the presence of the adduct was confirmed with HPLC-MS, with observed masses for both the dG<sup>^</sup>C nucleoside and base component matching theoretical values. The clear chromatographic separation at approximately 9.2 minutes proves the chemical identity, purity and existence of the adduct. These findings demonstrate the feasibility of one more method of chemoenzymatic 5hmC derivatization into a unique complex compound.

This study also aimed to assess how G<sup>^</sup>C derivatization affects polymerase activity, to facilitate specific further 5hmC detection in nanopore sequencing readouts. Using Deep Vent (exo-) polymerase, we observed a 31% reduction in amplification efficiency from modified templates, suggesting that G<sup>^</sup>C adduct interferes with DNA polymerization. Despite this, restriction digestion with Hin6I showed no evidence of insertions or deletions, indicating that most likely unmodified templates dominated amplification or that the polymerase bypassed modified sites without major errors. It was expected that DeepVent (exo-) would slow down at the modified sites, so we tried to perform the PCR more efficiently by adding DreamTaq polymerase along with DeepVent (exo-) because it could help polymerize the unmodified parts of the substrate more efficiently. When mixed, DeepVent (exo-) should have introduced errors and helped skip the modified sites, while DreamTaq should have improved overall yield.

The overall G<sup>^</sup>C yield was rather low. Therefore, we needed to emphasise the G<sup>^</sup>C adduct effects on DNA readouts. This was done by further enrichment of the modified sequences without amplifying the unmodified ones. At this stage, we focused on one target instead of ten, as we could enrich the modified full-length substrates by digesting the ones that lacked modification at the single R.EheI site. One site monitoring rendered a more simple mode of result interpretation in this complex pipeline of multiple reaction steps.

A key finding of this study is the ability of G<sup>^</sup>C-modified DNA to escape R.EheI digestion by altering the target sequence in resulting PCR products generated by DeepVent (exo-) and DreamTaq polymerases combined in one reaction. Nanopore sequencing further confirmed this effect. IGV analysis showed soft clipping at the GGC<sup>^</sup>GCC site only in the G<sup>^</sup>C-modified sample, indicating disruption of polymerization fidelity. Such site-specific disruption provided a functional readout of derivatized 5hmC modification. However, the errors introduced in the sequences are rather complex and long-ranged. The reads downstream of the modification site are automatically soft-clipped as too divergent from the reference. Manual inspection of the soft-clipped sequences shows some common patterns. However, they are not completely consistent and can not be reliably aligned at any other site of the substrate.

In conclusion, these results demonstrate that chemoenzymatic derivatization of 5hmC into a G<sup>^</sup>C adduct is not only possible but also can influence behaviour of DNA polymerases. The presence of the G<sup>^</sup>C adduct creates soft-clipping sites in nanopore sequencing reads of derived PCR products. Importantly, such polymerase-based research can be the first step for developing novel chemoenzymatic detection strategies for 5hmC. Now, it was observed that the sequencing data of derived PCR products was characterized by an overly increased error rate, which made it impossible to align the modification downstream sequences to the reference. This indicates a limitation of the method. To enable reliable, modification-sensitive applications in science and diagnostics and to

apply such methods for a novel 5hmC detection, future work should focus on testing other polymerases and their combinations.

## CONCLUSIONS

1. As it was shown before, 5hmC in DNA can be derivatized into a G<sup>+</sup>C base adduct using a combined enzymatic and chemical approach, with strong analytical evidence in this work confirming the specificity and reproducibility of the modification.
2. G<sup>+</sup>C adduct disrupts the activity of DNA polymerases DeepVent (exo-) and DreamTaq, as shown by reduced amplification efficiency and induction of extensive DNA sequence alterations downstream of the modification site. While the approach holds promise for applications in epigenetic research, low 5hmC derivatisation yield and overly erroneous PCR product sequencing readouts induced by the G<sup>+</sup>C adduct highlight the need for further optimisation and analysis.

# ABSTRACT

VILNIUS UNIVERSITY

DEPARTMENT OF BIOLOGICAL DNA MODIFICATION AT LIFE SCIENCES CENTER

**Linas Ambraziejus**

Master Thesis

## **The Effect of Chemoenzymatic 5hmC Derivatization on DNA Polymerase Readouts**

DNA modifications, particularly the epigenetic marks 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), play critical roles in gene regulation, embryonic and disease development, including cancer. Despite its biological importance, the detection and mapping of 5hmC remain technically challenging due to its chemical similarity to 5-methylcytosine and limitations of current sequencing methods. This study aimed to explore a novel chemoenzymatic strategy for 5hmC detection by evaluating the effect of G<sup>^</sup>C adduct formation, which is a product of 5hmC derivatization, on DNA polymerase fidelity and nanopore sequencing readouts.

In this project we aimed to synthesize model DNA containing 5hmC with M.HhaI and formaldehyde; to derivatize 5hmC into a G<sup>^</sup>C adduct using a selenol-based chemoenzymatic modification; and to evaluate the impact of this modification on DNA polymerase DeepVent (exo-) and DreamTaq performance and sequencing accuracy. PCR analysis with polymerases revealed a 31% decrease in amplification efficiency from G<sup>^</sup>C-modified templates, indicating interference with polymerase activity. Nanopore sequencing further confirmed the functional impact of the G<sup>^</sup>C adduct. Soft-clipping was observed at the modification site, and downstream sequences showed complex, long-range deviations from the reference, supporting the hypothesis that the G<sup>^</sup>C adduct alters polymerase fidelity. These effects were not observed in the control samples.

This study demonstrates that 5hmC can be converted into a unique intramolecular G<sup>^</sup>C adduct that significantly affects DNA polymerase behaviour and sequencing output. While the derivatization strategy shows potential as a novel method for 5hmC detection, the low modification yield and high error rate in sequencing readouts underscore the need for further optimization. Nonetheless, the approach provides a foundation for developing sensitive, modification-specific detection strategies in epigenetic research and diagnostics.

# SANTRAUKA

VILNIAUS UNIVERSITETAS  
GYVYBĖS MOKSLŲ CENTRO DNR MODIFIKACIJŲ TYRIMŲ SKYRIUS

**Linas Ambraziejus**

Magistro baigiamasis darbas

## **Chemofermentinės 5hmC derivatizacijos poveikis DNR polimerazių tikslumui**

DNR modifikacijos, ypač epigenetiniai žymenys 5-metilcitozinas (5mC) ir 5-hidroksimetilcitozinas (5hmC), atlieka svarbų vaidmenį genų reguliacijoje, embriono ir ligų, įskaitant vėžį, vystymesi. Nepaisant biologinės svarbos, 5hmC aptikimas ir tikslios vietos nustatymas išlieka techniškai sudėtingas dėl jo cheminio panašumo į 5-metilcitoziną bei dabartinių sekvenavimo metodų apribojimų. Šio tyrimo tikslas buvo ištirti naują chemofermentinę 5hmC aptikimo strategiją, įvertinant G<sup>C</sup> adukto, kuris yra 5hmC derivatizacijos produktas, susidarymo poveikį DNR polimerazės tikslumui ir nanoporų sekvenavimo rodmenims.

Šiame projekte siekėme susintetinti modelinę DNR su 5hmC, naudojant M.HhaI ir formaldehidą; derivatizuoti 5hmC į G<sup>C</sup> aduktą naudojant netipinį metiltransferazių aktyvumą su selenoliais ir cheminę oksidaciją; ir įvertinti šios modifikacijos poveikį DNR polimerazių DeepVent (exo-) ir DreamTaq veikimui bei DNR nuskaitymo tikslumui. PGR analizė su polimerazėmis parodė 31 % sumažėjusį amplifikacijos efektyvumą nuo G<sup>C</sup> modifikuotų mėginių, o tai parodo polimerazės aktyvumo sumažėjimą. Nanoporų sekvenavimas dar kartą patvirtino G<sup>C</sup> adukto funkcinį poveikį. Modifikacijos vietoje buvo pastebėtas „bioinformatinis iškirpimas“ (angl. *soft-clipping*), o tolesnėse sekose buvo matomi sudėtingi, dideli nukrypimai nuo referentinės sekos, patvirtinantys hipotezę, kad G<sup>C</sup> aduktas keičia polimerazės tikslumą. Šis poveikis nebuvo pastebėtas kontroliniuose mėginiuose.

Šis tyrimas parodo, kad 5hmC gali būti paverstas unikaliu intramolekuliniu G<sup>C</sup> aduktu, kuris reikšmingai veikia DNR polimerazių elgseną ir DNR nuskaitymo rezultatus. Nors ši derivatizacijos metodika rodo naujo 5hmC aptikimo metodo potencialą, maža modifikavimo išeiga ir didelis klaidų dažnis DNR nuskaitymo rodmenyse pabrėžia tolesnio optimizavimo poreikį. Nepaisant to, šis metodas suteikia pagrindą kurti jautrias, 5hmC modifikacijai specifines aptikimo metodikas epigenetiniuose tyrimuose ir diagnostikoje.

## **AUTHOR'S PERSONAL CONTRIBUTION**

PCR, gel electrophoresis and analysis, chemical and enzymatic reactions, library preparation for sequencing as well as Master thesis were done and written by me, reviewed by dr. Miglė Tomkuvienė. Nanopore sequencing was assisted by Joris Balčiūnas. HPLC-MS was performed by Audronė Rukšėnaitė.

## ACKNOWLEDGMENTS

I would like to express sincere gratitude to my supervisor dr. Miglė Tomkuvienė for accepting me into her team at the last minute when I needed it the most and providing an interesting topic. Not to mention continuous guidance, valuable insights, and support throughout the course of this project. In addition, I want to compliment my supervisor's endless kindness and patience in answering all the questions asked, even the silliest ones.

I also thank the members of the laboratory for keeping a friendly atmosphere and for technical assistance, especially Joris Balčiūnas, who helped with nanopore sequencing and data analysis. I could not have done my project without Audronė Rukšėnaitė, who performed mass spectrometry and consulted on related questions.

Lastly, I want to thank my whole family for believing and support while I was studying at this university.

## REFERENCES

- Aberg, K. A., Chan, R. F., & van den Oord, E. J. C. G. (2020). MBD-seq - realities of a misunderstood method for high-quality methylome-wide association studies. *Epigenetics*, *15*(4), 431–438. <https://doi.org/10.1080/15592294.2019.1695339>
- Aberg, K. A., Chan, R. F., Xie, L., Shabalin, A. A., & van den Oord, E. J. C. G. (2018). *Methyl-CpG-Binding Domain Sequencing: MBD-seq* (pp. 171–189). [https://doi.org/10.1007/978-1-4939-7481-8\\_10](https://doi.org/10.1007/978-1-4939-7481-8_10)
- Al Aboud, N. M., Tupper, C., & Jialal, I. (2024). *Genetics, Epigenetic Mechanism* (2023 Aug 14). Treasure Island (FL): StatPearls Publishing.
- Alam, H., Gu, B., & Lee, M. G. (2015). Histone methylation modifiers in cellular signaling pathways. *Cellular and Molecular Life Sciences : CMLS*, *72*(23), 4577–4592. <https://doi.org/10.1007/s00018-015-2023-y>
- Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., von Bülow, V., Harb, H., Alhamdan, F., Hii, C. S., Prescott, S. L., Ferrante, A., Renz, H., Garn, H., & Potaczek, D. P. (2018). Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, Asthma, and Clinical Immunology : Official Journal of the Canadian Society of Allergy and Clinical Immunology*, *14*, 39. <https://doi.org/10.1186/s13223-018-0259-4>
- Anand, J., Chiou, L., Sciandra, C., Zhang, X., Hong, J., Wu, D., Zhou, P., & Vaziri, C. (2023). Roles of trans-lesion synthesis (TLS) DNA polymerases in tumorigenesis and cancer therapy. *NAR Cancer*, *5*(1). <https://doi.org/10.1093/narcan/zcad005>
- Armstrong, M. J., Jin, Y., Vattathil, S. M., Huang, Y., Schroeder, J. P., Bennet, D. A., Qin, Z. S., Wingo, T. S., & Jin, P. (2023). Role of TET1-mediated epigenetic modulation in Alzheimer's disease. *Neurobiology of Disease*, *185*, 106257. <https://doi.org/10.1016/j.nbd.2023.106257>
- Basu, A., & Tiwari, V. K. (2021). Epigenetic reprogramming of cell identity: lessons from development for regenerative medicine. *Clinical Epigenetics*, *13*(1), 144. <https://doi.org/10.1186/s13148-021-01131-4>
- Bedaiwi, S., Usmani, A., & Carty, M. P. (2024). Canonical and Non-Canonical Roles of Human DNA Polymerase  $\eta$ . *Genes*, *15*(10). <https://doi.org/10.3390/genes15101271>
- Bellacosa, A., & Drohat, A. C. (2015). Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA Repair*, *32*, 33–42. <https://doi.org/10.1016/j.dnarep.2015.04.011>
- Bizuayehu, H. M., Ahmed, K. Y., Kibret, G. D., Dadi, A. F., Belachew, S. A., Bagade, T., Tegegne, T. K., Venchiarutti, R. L., Kibret, K. T., Hailegebireal, A. H., Assefa, Y., Khan, M. N., Abajobir, A., Alene, K. A., Mengesha, Z., Erku, D., Enquobahrie, D. A., Minas, T. Z., Misgan, E., & Ross, A. G. (2024). Global Disparities of Cancer and Its Projected Burden in 2050. *JAMA Network Open*, *7*(11), e2443198. <https://doi.org/10.1001/jamanetworkopen.2024.43198>
- Bomasang-Layno, E., & Bronsther, R. (2021). Diagnosis and Treatment of Alzheimer's Disease: *Delaware Journal of Public Health*, *7*(4), 74–85. <https://doi.org/10.32481/djph.2021.09.009>
- Booth, M. J., Ost, T. W. B., Beraldi, D., Bell, N. M., Branco, M. R., Reik, W., & Balasubramanian, S. (2013). Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature Protocols*, *8*(10), 1841–1851. <https://doi.org/10.1038/nprot.2013.115>
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. <https://doi.org/10.3322/caac.21834>

- Cao, Y., Bai, Y., Yuan, T., Song, L., Fan, Y., Ren, L., Song, W., Peng, J., An, R., Gu, Q., Zheng, Y., & Xie, X. S. (2023). Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. *Proceedings of the National Academy of Sciences*, *120*(49). <https://doi.org/10.1073/pnas.2310367120>
- Chatterjee, R., & Vinson, C. (2012). CpG methylation recruits sequence specific transcription factors essential for tissue specific gene expression. *Biochimica et Biophysica Acta*, *1819*(7), 763–770. <https://doi.org/10.1016/j.bbagr.2012.02.014>
- Chen, H., Duolikun, M., & Zhu, H.-C. (2024). 5-Hydroxymethylcytosine profilings in circulating cell-free DNA as diagnostic biomarkers for DLBCL. *Frontiers in Cell and Developmental Biology*, *12*. <https://doi.org/10.3389/fcell.2024.1387959>
- Chen, L., Shen, Q., Xu, S., Yu, H., Pei, S., Zhang, Y., He, X., Wang, Q., & Li, D. (2022). 5-Hydroxymethylcytosine Signatures in Circulating Cell-Free DNA as Diagnostic Biomarkers for Late-Onset Alzheimer's Disease. *Journal of Alzheimer's Disease*, *85*(2), 573–585. <https://doi.org/10.3233/JAD-215217>
- Chen, Y., Damayanti, N. P., Irudayaraj, J., Dunn, K., & Zhou, F. C. (2014). Diversity of two forms of DNA methylation in the brain. *Frontiers in Genetics*, *5*. <https://doi.org/10.3389/fgene.2014.00046>
- Coelho, F. S., Sangi, S., Moraes, J. L., Santos, W. da S., Gamosa, E. A., Fernandes, K. V. S., & Grativol, C. (2022). Methyl-CpG binding proteins (MBD) family evolution and conservation in plants. *Gene*, *824*, 146404. <https://doi.org/10.1016/j.gene.2022.146404>
- Dai, Q., Ye, C., Irkliyenko, I., Wang, Y., Sun, H.-L., Gao, Y., Liu, Y., Beadell, A., Perea, J., Goel, A., & He, C. (2024). Ultrafast bisulfite sequencing detection of 5-methylcytosine in DNA and RNA. *Nature Biotechnology*, *42*(10), 1559–1570. <https://doi.org/10.1038/s41587-023-02034-w>
- De Borre, M., & Branco, M. R. (2021). *Oxidative Bisulfite Sequencing: An Experimental and Computational Protocol* (pp. 333–348). [https://doi.org/10.1007/978-1-0716-0876-0\\_26](https://doi.org/10.1007/978-1-0716-0876-0_26)
- de Paz, A. M., Cybulski, T. R., Marblestone, A. H., Zamft, B. M., Church, G. M., Boyden, E. S., Kording, K. P., & Tyo, K. E. J. (2018). High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic Acids Research*, *46*(13), e78. <https://doi.org/10.1093/nar/gky296>
- Ehrlich, M. (2019). DNA hypermethylation in disease: mechanisms and clinical relevance. *Epigenetics*, *14*(12), 1141–1163. <https://doi.org/10.1080/15592294.2019.1638701>
- Feng, Y., Chen, J.-J., Xie, N.-B., Ding, J.-H., You, X.-J., Tao, W.-B., Zhang, X., Yi, C., Zhou, X., Yuan, B.-F., & Feng, Y.-Q. (2021). Direct decarboxylation of ten-eleven translocation-produced 5-carboxylcytosine in mammalian genomes forms a new mechanism for active DNA demethylation. *Chemical Science*, *12*(34), 11322–11329. <https://doi.org/10.1039/D1SC02161C>
- Garcia-Diaz, M., & Bebenek, K. (2007). Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences*, *26*(2), 105–122. <https://doi.org/10.1080/07352680701252817>
- Gardner, A. F., & Kelman, Z. (2014). DNA polymerases in biotechnology. *Frontiers in Microbiology*, *5*, 659. <https://doi.org/10.3389/fmicb.2014.00659>
- Gonzalez-Avalos, E., Onodera, A., Samaniego-Castruita, D., Rao, A., & Ay, F. (2024). Predicting gene expression state and prioritizing putative enhancers using 5hmC signal. *Genome Biology*, *25*(1), 142. <https://doi.org/10.1186/s13059-024-03273-z>
- Hahn, M. A., Szabó, P. E., & Pfeifer, G. P. (2014). 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics*, *104*(5), 314–323. <https://doi.org/10.1016/j.ygeno.2014.08.015>
- Hamidi, T., Singh, A. K., & Chen, T. (2015). Genetic Alterations of DNA Methylation Machinery in Human Diseases. *Epigenomics*, *7*(2), 247–265. <https://doi.org/10.2217/epi.14.80>

- Hamilton, J. P. (2011). Epigenetics: principles and practice. *Digestive Diseases (Basel, Switzerland)*, 29(2), 130–135. <https://doi.org/10.1159/000323874>
- Han, D., Lu, X., Shih, A. H., Nie, J., You, Q., Xu, M. M., Melnick, A. M., Levine, R. L., & He, C. (2016). A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Molecular Cell*, 63(4), 711–719. <https://doi.org/10.1016/j.molcel.2016.06.028>
- Han, P., & Chang, C.-P. (2015). Long non-coding RNA and chromatin remodeling. *RNA Biology*, 12(10), 1094–1098. <https://doi.org/10.1080/15476286.2015.1063770>
- He, B., Yao, H., & Yi, C. (2024). Advances in the joint profiling technologies of 5mC and 5hmC. *RSC Chemical Biology*, 5(6), 500–507. <https://doi.org/10.1039/d4cb00034j>
- Hernández-Rollán, C., Ehrmann, A. K., Vlassis, A., Kandasamy, V., & Nørholm, M. H. H. (2024). Neq2X7: a multi-purpose and open-source fusion DNA polymerase for advanced DNA engineering and diagnostics PCR. *BMC Biotechnology*, 24(1), 17. <https://doi.org/10.1186/s12896-024-00844-7>
- Horton, J. R., Yang, J., Zhang, X., Petronzio, T., Fomenkov, A., Wilson, G. G., Roberts, R. J., & Cheng, X. (2020). Structure of HhaI endonuclease with cognate DNA at an atomic resolution of 1.0 Å. *Nucleic Acids Research*, 48(3), 1466–1478. <https://doi.org/10.1093/nar/gkz1195>
- Howard, M. J., Rodriguez, Y., & Wilson, S. H. (2017). DNA polymerase  $\beta$  uses its lyase domain in a processive search for DNA damage. *Nucleic Acids Research*, 45(7), 3822–3832. <https://doi.org/10.1093/nar/gkx047>
- Hwa Yun, B., Guo, J., Bellamri, M., & Turesky, R. J. (2020). DNA adducts: Formation, biological effects, and new biospecimens for mass spectrometric measurements in humans. *Mass Spectrometry Reviews*, 39(1–2), 55–82. <https://doi.org/10.1002/mas.21570>
- Jannasch, H. W., Wirsén, C. O., Molyneaux, S. J., & Langworthy, T. A. (1992). Comparative Physiological Studies on Hyperthermophilic Archaea Isolated from Deep-Sea Hot Vents with Emphasis on Pyrococcus Strain GB-D. *Applied and Environmental Microbiology*, 58(11), 3472–3481. <https://doi.org/10.1128/aem.58.11.3472-3481.1992>
- Jessop, P., Ruzov, A., & Gering, M. (2018). Developmental Functions of the Dynamic DNA Methylome and Hydroxymethylome in the Mouse and Zebrafish: Similarities and Differences. *Frontiers in Cell and Developmental Biology*, 6. <https://doi.org/10.3389/fcell.2018.00027>
- Joseph, D. B., Strand, D. W., & Vezina, C. M. (2018). DNA methylation in development and disease: an overview for prostate researchers. *American Journal of Clinical and Experimental Urology*, 6(6), 197–218.
- Kaszubowski, J. D., & Trakselis, M. A. (2021). Beyond the Lesion: Back to High Fidelity DNA Synthesis. *Frontiers in Molecular Biosciences*, 8, 811540. <https://doi.org/10.3389/fmolb.2021.811540>
- Katsman, E., Orlanski, S., Martignano, F., Fox-Fisher, I., Shemer, R., Dor, Y., Zick, A., Eden, A., Petrini, I., Conticello, S. G., & Berman, B. P. (2022). Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing. *Genome Biology*, 23(1), 158. <https://doi.org/10.1186/s13059-022-02710-1>
- Kirschner, K., Krueger, F., Green, A. R., & Chandra, T. (2018). *Multiplexing for Oxidative Bisulfite Sequencing (oxBS-seq)* (pp. 665–678). [https://doi.org/10.1007/978-1-4939-7481-8\\_34](https://doi.org/10.1007/978-1-4939-7481-8_34)
- Kisil, O., Sergeev, A., Bacheva, A., & Zvereva, M. (2024). Methods for Detection and Mapping of Methylated and Hydroxymethylated Cytosine in DNA. *Biomolecules*, 14(11). <https://doi.org/10.3390/biom14111346>
- Kochmanski, J., & Bernstein, A. I. (2020). The Impact of Environmental Factors on 5-Hydroxymethylcytosine in the Brain. *Current Environmental Health Reports*, 7(2), 109–120. <https://doi.org/10.1007/s40572-020-00268-3>
- Kriukienė, E., Tomkuvienė, M., & Klimašauskas, S. (2024). 5-Hydroxymethylcytosine: the many faces of the sixth base of mammalian DNA. *Chemical Society Reviews*, 53(5), 2264–2283. <https://doi.org/10.1039/D3CS00858D>

- Kuban, W., Jonczyk, P., Gawel, D., Malanowska, K., Schaaper, R. M., & Fijalkowska, I. J. (2004). Role of Escherichia coli DNA polymerase IV in in vivo replication fidelity. *Journal of Bacteriology*, *186*(14), 4802–4807. <https://doi.org/10.1128/JB.186.14.4802-4807.2004>
- Lapa, S. A., Volkova, O. S., Kuznetsova, V. E., Zasedatelev, A. S., & Chudinov, A. V. (2022). Study of Multiple Enzymatic Incorporation of Modified Nucleotides of Purine and Pyrimidine Nature in the Growing DNA Chain. *Molecular Biology*, *56*(1), 115–123. <https://doi.org/10.1134/S0026893322010046>
- Lee, S.-M. (2024). Detecting DNA hydroxymethylation: exploring its role in genome regulation. *BMB Reports*, *57*(3), 135–142. <https://doi.org/10.5483/BMBRep.2023-0250>
- Li, J. J. N., Liu, G., & Lok, B. H. (2024). Cell-Free DNA Hydroxymethylation in Cancer: Current and Emerging Detection Methods and Clinical Applications. *Genes*, *15*(9), 1160. <https://doi.org/10.3390/genes15091160>
- Li, X., Shi, X., Gong, Y., Guo, W., Liu, Y., Peng, C., & Xu, Y. (2021). Selective Chemical Labeling and Sequencing of 5-Hydroxymethylcytosine in DNA at Single-Base Resolution. *Frontiers in Genetics*, *12*, 749211. <https://doi.org/10.3389/fgene.2021.749211>
- Li, Y., & Tollefsbol, T. O. (2011). DNA methylation detection: bisulfite genomic sequencing analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *791*, 11–21. [https://doi.org/10.1007/978-1-61779-316-5\\_2](https://doi.org/10.1007/978-1-61779-316-5_2)
- Liu, K., Xu, C., Lei, M., Yang, A., Loppnau, P., Hughes, T. R., & Min, J. (2018). Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. *Journal of Biological Chemistry*, *293*(19), 7344–7354. <https://doi.org/10.1074/jbc.RA118.001785>
- Liu, R., Wu, J., Guo, H., Yao, W., Li, S., Lu, Y., Jia, Y., Liang, X., Tang, J., & Zhang, H. (2023). Post-translational modifications of histones: Mechanisms, biological functions, and therapeutic targets. *MedComm*, *4*(3). <https://doi.org/10.1002/mco2.292>
- Liu, Y., Hu, Z., Cheng, J., Siejka-Zielińska, P., Chen, J., Inoue, M., Ahmed, A. A., & Song, C.-X. (2021). Subtraction-free and bisulfite-free specific sequencing of 5-methylcytosine and its oxidized derivatives at base resolution. *Nature Communications*, *12*(1), 618. <https://doi.org/10.1038/s41467-021-20920-2>
- Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foux, 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>
- Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., & Song, C.-X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology*, *37*(4), 424–429. <https://doi.org/10.1038/s41587-019-0041-2>
- Liutkevičiūtė, Z., Kriukienė, E., Grigaitytė, I., Masevičius, V., & Klimašauskas, S. (2011). Methyltransferase-directed derivatization of 5-hydroxymethylcytosine in DNA. *Angewandte Chemie (International Ed. in English)*, *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>
- López, V., Fernández, A. F., & Fraga, M. F. (2017). The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. *Ageing Research Reviews*, *37*, 28–38. <https://doi.org/10.1016/j.arr.2017.05.002>
- Ludwig, A. K., Zhang, P., & Cardoso, M. C. (2016). Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease. *Frontiers in Genetics*, *7*, 115. <https://doi.org/10.3389/fgene.2016.00115>

- Lujan, S. A., Williams, J. S., & Kunkel, T. A. (2016). DNA Polymerases Divide the Labor of Genome Replication. *Trends in Cell Biology*, 26(9), 640–654. <https://doi.org/10.1016/j.tcb.2016.04.012>
- Madrid, A., Chopra, P., & Alisch, R. S. (2018). Species-Specific 5 mC and 5 hmC Genomic Landscapes Indicate Epigenetic Contribution to Human Brain Evolution. *Frontiers in Molecular Neuroscience*, 11, 39. <https://doi.org/10.3389/fnmol.2018.00039>
- McInerney, P., Adams, P., & Hadi, M. Z. (2014). Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International*, 2014, 287430. <https://doi.org/10.1155/2014/287430>
- Minchin, S., & Lodge, J. (2019). Understanding biochemistry: structure and function of nucleic acids. *Essays in Biochemistry*, 63(4), 433–456. <https://doi.org/10.1042/EBC20180038>
- Mitrea, C., Wijesinghe, P., Dyson, G., Kruger, A., Ruden, D. M., Drăghici, S., & Bollig-Fischer, A. (2018). Integrating 5hmC and gene expression data to infer regulatory mechanisms. *Bioinformatics*, 34(9), 1441–1447. <https://doi.org/10.1093/bioinformatics/btx777>
- Müller, N., Ponkkonen, E., Carell, T., & Khobta, A. (2021). Direct and Base Excision Repair-Mediated Regulation of a GC-Rich cis-Element in Response to 5-Formylcytosine and 5-Carboxycytosine. *International Journal of Molecular Sciences*, 22(20). <https://doi.org/10.3390/ijms222011025>
- Nestor, C. E., & Meehan, R. R. (2014). *Hydroxymethylated DNA Immunoprecipitation (hmeDIP)* (pp. 259–267). [https://doi.org/10.1007/978-1-62703-706-8\\_20](https://doi.org/10.1007/978-1-62703-706-8_20)
- Nestor, C., Ruzov, A., Meehan, R. R., & Dunican, D. S. (2010). Enzymatic Approaches and Bisulfite Sequencing Cannot Distinguish Between 5-Methylcytosine and 5-Hydroxymethylcytosine in DNA. *BioTechniques*, 48(4), 317–319. <https://doi.org/10.2144/000113403>
- Pang, A. P. S., Sugai, C., & Maunakea, A. K. (2016). High-throughput sequencing offers new insights into 5-hydroxymethylcytosine. *Biomolecular Concepts*, 7(3), 169–178. <https://doi.org/10.1515/bmc-2016-0011>
- Patel, M. S., Almubarak, M., Matta, J., Ortiz-Sanchez, C., Encarnacion, J., Ruiz-Deya, G., Dutil, J., Dhillon, J., Yamoah, K., Berglund, A., Park, H., Kilari, D., Balagurunathan, Y., Wang, L., & Park, J. Y. (2025). 5hmC-profiles in Puerto Rican Hispanic/Latino men with aggressive prostate cancer. *Frontiers in Oncology*, 15, 1541878. <https://doi.org/10.3389/fonc.2025.1541878>
- Pease, M., Ling, C., Mack, W. J., Wang, K., & Zada, G. (2013). The Role of Epigenetic Modification in Tumorigenesis and Progression of Pituitary Adenomas: A Systematic Review of the Literature. *PLoS ONE*, 8(12), e82619. <https://doi.org/10.1371/journal.pone.0082619>
- Powers, K. T., Elcock, A. H., & Washington, M. T. (2018). The C-terminal region of translesion synthesis DNA polymerase  $\eta$  is partially unstructured and has high conformational flexibility. *Nucleic Acids Research*, 46(4), 2107–2120. <https://doi.org/10.1093/nar/gky031>
- Reyes, A. A., Marcum, R. D., & He, Y. (2021). Structure and Function of Chromatin Remodelers. *Journal of Molecular Biology*, 433(14), 166929. <https://doi.org/10.1016/j.jmb.2021.166929>
- Schutsky, E. K., DeNizio, J. E., Hu, P., Liu, M. Y., Nabel, C. S., Fabyanic, E. B., Hwang, Y., Bushman, F. D., Wu, H., & Kohli, R. M. (2018). Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. *Nature Biotechnology*, 36(11), 1083–1090. <https://doi.org/10.1038/nbt.4204>
- Sellés-Baiget, S., Ambjørn, S. M., Carli, A., Hendriks, I. A., Gallina, I., Davey, N. E., Benedict, B., Zarantonello, A., Gadi, S. A., Meeusen, B., Hertz, E. P. T., Slappendel, L., Semlow, D., Sturla, S., Nielsen, M. L., Nilsson, J., Miller, T. C. R., & Duxin, J. P. (2025). Catalytic and noncatalytic functions of DNA polymerase  $\kappa$  in translesion DNA synthesis. *Nature Structural & Molecular Biology*, 32(2), 300–314. <https://doi.org/10.1038/s41594-024-01395-3>

- Shariq, O. A., & Lines, K. E. (2019). Epigenetic Dysregulation in Pituitary Tumors. *International Journal of Endocrine Oncology*, 6(3). <https://doi.org/10.2217/ije-2019-0006>
- Shi, D.-Q., Ali, I., Tang, J., & Yang, W.-C. (2017). New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Frontiers in Genetics*, 8. <https://doi.org/10.3389/fgene.2017.00100>
- Skvortsova, K., Stirzaker, C., & Taberlay, P. (2019). The DNA methylation landscape in cancer. *Essays in Biochemistry*, 63(6), 797–811. <https://doi.org/10.1042/EBC20190037>
- Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., & Jacobsen, S. E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biology*, 12(6), R54. <https://doi.org/10.1186/gb-2011-12-6-r54>
- Sun, W., Zang, L., Shu, Q., & Li, X. (2014). From development to diseases: The role of 5hmC in brain. *Genomics*, 104(5), 347–351. <https://doi.org/10.1016/j.ygeno.2014.08.021>
- Sundarrajan, S., Parambath, S., Suresh, S., Rao, S., & Padmanabhan, S. (2018). Novel properties of recombinant Sso7d-Taq DNA polymerase purified using aqueous two-phase extraction: Utilities of the enzyme in viral diagnosis. *Biotechnology Reports (Amsterdam, Netherlands)*, 19, e00270. <https://doi.org/10.1016/j.btre.2018.e00270>
- Tamás, V., Mészáros, I., Olasz, F., Kiss, I., Homonnay, Z. G., Mortensen, P., & Zádori, Z. (2022). Allele-Specific Dual PCRs to Identify Members of the 27a Cluster of PPV. *Viruses*, 14(7). <https://doi.org/10.3390/v14071500>
- Tan, L., Xiong, L., Xu, W., Wu, F., Huang, N., Xu, Y., Kong, L., Zheng, L., Schwartz, L., Shi, Y., & Shi, Y. G. (2013). Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Research*, 41(7), e84–e84. <https://doi.org/10.1093/nar/gkt091>
- Thomson, J. P., Hunter, J. M., Nestor, C. E., Dunican, D. S., Terranova, R., Moggs, J. G., & Meehan, R. R. (2013). Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Research*, 41(22), e206–e206. <https://doi.org/10.1093/nar/gkt1080>
- Uysal, F., Akkoyunlu, G., & Ozturk, S. (2016). DNA methyltransferases exhibit dynamic expression during spermatogenesis. *Reproductive BioMedicine Online*, 33(6), 690–702. <https://doi.org/10.1016/j.rbmo.2016.08.022>
- Vaisman, A., Łazowski, K., Reijns, M. A. M., Walsh, E., McDonald, J. P., Moreno, K. C., Quiros, D. R., Schmidt, M., Kranz, H., Yang, W., Makiela-Dzbenka, K., & Woodgate, R. (2021). Novel Escherichia coli active site dnaE alleles with altered base and sugar selectivity. *Molecular Microbiology*, 116(3), 909–925. <https://doi.org/10.1111/mmi.14779>
- Vaisvila, R., Ponnaluri, V. K. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., Samaranyake, M., Samuelson, J. C., Church, H. E., Tamanaha, E., Corrêa, I. R., Pradhan, S., Dimalanta, E. T., Evans, T. C., Williams, L., & Davis, T. B. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*, 31(7), 1280–1289. <https://doi.org/10.1101/gr.266551.120>
- Varma, S. J., Calvani, E., Grüning, N.-M., Messner, C. B., Grayson, N., Capuano, F., Mülleder, M., & Ralser, M. (2022). Global analysis of cytosine and adenine DNA modifications across the tree of life. *ELife*, 11. <https://doi.org/10.7554/eLife.81002>
- Vető, B., Szabó, P., Bacquet, C., Apró, A., Hathy, E., Kiss, J., Réthelyi, J. M., Szeri, F., Szüts, D., & Arányi, T. (2018). Inhibition of DNA methyltransferase leads to increased genomic 5-hydroxymethylcytosine levels in hematopoietic cells. *FEBS Open Bio*, 8(4), 584–592. <https://doi.org/10.1002/2211-5463.12392>
- Wang, T., Fowler, J. M., Liu, L., Loo, C. E., Luo, M., Schutsky, E. K., Berríos, K. N., DeNizio, J. E., Dvorak, A., Downey, N., Montermoso, S., Pingul, B. Y., Nasrallah, M., Gosal, W. S., Wu, H., & Kohli, R. M. (2023). Direct enzymatic

- sequencing of 5-methylcytosine at single-base resolution. *Nature Chemical Biology*, 19(8), 1004–1012. <https://doi.org/10.1038/s41589-023-01318-1>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Wei, A., Zhang, H., Qiu, Q., Fabyanic, E. B., Hu, P., & Wu, H. (2023). 5-hydroxymethylcytosines regulate gene expression as a passive DNA demethylation resisting epigenetic mark in proliferative somatic cells. <https://doi.org/10.1101/2023.09.26.559662>
- Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., & Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & Development*, 25(7), 679–684. <https://doi.org/10.1101/gad.2036011>
- Wu, Y.-L., Lin, Z.-J., Li, C.-C., Lin, X., Shan, S.-K., Guo, B., Zheng, M.-H., Li, F., Yuan, L.-Q., & Li, Z. (2023). Epigenetic regulation in metabolic diseases: mechanisms and advances in clinical study. *Signal Transduction and Targeted Therapy*, 8(1), 98. <https://doi.org/10.1038/s41392-023-01333-7>
- Xie, J., Xie, L., Wei, H., Li, X.-J., & Lin, L. (2023). Dynamic Regulation of DNA Methylation and Brain Functions. *Biology*, 12(2). <https://doi.org/10.3390/biology12020152>
- Xue, Y., Braslavsky, I., & Quake, S. R. (2021). Temperature effect on polymerase fidelity. *The Journal of Biological Chemistry*, 297(5), 101270. <https://doi.org/10.1016/j.jbc.2021.101270>
- Yan, R., Cheng, X., Gu, C., Xu, Y., Long, X., Zhai, J., Sun, F., Qian, J., Du, Y., Wang, H., & Guo, F. (2023). Dynamics of DNA hydroxymethylation and methylation during mouse embryonic and germline development. *Nature Genetics*, 55(1), 130–143. <https://doi.org/10.1038/s41588-022-01258-x>
- Yan, S. F., Wu, M., Geacintov, N. E., & Broyde, S. (2004). Altering DNA polymerase incorporation fidelity by distorting the dNTP binding pocket with a bulky carcinogen-damaged template. *Biochemistry*, 43(24), 7750–7765. <https://doi.org/10.1021/bi0499516>
- Yang, T., Wang, D., Tian, G., Sun, L., Yang, M., Yin, X., Xiao, J., Sheng, Y., Zhu, D., He, H., & Zhou, Y. (2022). Chromatin remodeling complexes regulate genome architecture in Arabidopsis. *The Plant Cell*, 34(7), 2638–2651. <https://doi.org/10.1093/plcell/koac117>
- Yang, Z., Xu, F., Teschendorff, A. E., Zhao, Y., Yao, L., Li, J., & He, Y. (2022). Insights into the role of long non-coding RNAs in DNA methylation mediated transcriptional regulation. *Frontiers in Molecular Biosciences*, 9. <https://doi.org/10.3389/fmolb.2022.1067406>
- Yao, B., Hsu, C., Goldner, G., Michaeli, Y., Ebenstein, Y., & Listgarten, J. (2024). Effective training of nanopore callers for epigenetic marks with limited labelled data. *Open Biology*, 14(6). <https://doi.org/10.1098/rsob.230449>
- Yu, M., Han, D., Hon, G. C., & He, C. (2018). Tet-Assisted Bisulfite Sequencing (TAB-seq). *Methods in Molecular Biology (Clifton, N.J.)*, 1708, 645–663. [https://doi.org/10.1007/978-1-4939-7481-8\\_33](https://doi.org/10.1007/978-1-4939-7481-8_33)
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C.-X., Jin, P., Ren, B., & He, C. (2012). Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nature Protocols*, 7(12), 2159–2170. <https://doi.org/10.1038/nprot.2012.137>
- Zahid, O. K., Rivas, F., Wang, F., Sethi, K., Reiss, K., Bearden, S., & Hall, A. R. (2024). Corrigendum to “Solid-state nanopore analysis of human genomic DNA shows unaltered global 5-hydroxymethylcytosine content associated with early-stage breast cancer” [Nanomed. Nanotechnol. Biol. Med. 35(2021)102407]. *Nanomedicine: Nanotechnology, Biology and Medicine*, 58, 102746. <https://doi.org/10.1016/j.nano.2024.102746>
- Zeng, T., Liu, L., Li, T., Li, Y., Gao, J., Zhao, Y., & Wu, H.-C. (2015). Detection of 5-methylcytosine and 5-hydroxymethylcytosine in DNA via host–guest interactions inside  $\alpha$ -hemolysin nanopores. *Chemical Science*, 6(10), 5628–5634. <https://doi.org/10.1039/C5SC01436K>

- Zhao, Z., & Han, L. (2009). CpG islands: algorithms and applications in methylation studies. *Biochemical and Biophysical Research Communications*, 382(4), 643–645. <https://doi.org/10.1016/j.bbrc.2009.03.076>
- Zheng, K., Lyu, Z., Chen, J., & Chen, G. (2024). 5-Hydroxymethylcytosine: Far Beyond the Intermediate of DNA Demethylation. *International Journal of Molecular Sciences*, 25(21). <https://doi.org/10.3390/ijms252111780>
- Zheng, P., Zhou, C., Ding, Y., Liu, B., Lu, L., Zhu, F., & Duan, S. (2023). Nanopore sequencing technology and its applications. *MedComm*, 4(4). <https://doi.org/10.1002/mco2.316>
- Zhou, Z.-X., Lujan, S. A., Burkholder, A. B., St Charles, J., Dahl, J., Farrell, C. E., Williams, J. S., & Kunkel, T. A. (2021). How asymmetric DNA replication achieves symmetrical fidelity. *Nature Structural & Molecular Biology*, 28(12), 1020–1028. <https://doi.org/10.1038/s41594-021-00691-6>