THE UNIVERSITY *of* EDINBURGH

Using synthetic biology to understand DNA methylation maintenance in cancer

Andreanna Wright

Doctor of Philosophy

The University of Edinburgh

2024

Declaration

I declare that the work presented in this thesis is my own work, unless otherwise stated. This thesis has not been submitted for any other degree, diploma, or professional qualification.

Andreanna Wright

31st August 2024

Acknowledgements

Firstly, I would like to thank my supervisor Duncan Sproul for his continual support and guidance throughout this project. Your door has always been open for help and advice. I would especially like to thank you for your patience with teaching me how to use Eddie and R, even when I forget the most basic commands! I've learned so much over the last 4 years and it has been a pleasure to be part of the Sproul lab.

Thank you to every member of the Sproul lab, past and present. Francesca, I am incredibly grateful for your wisdom, and you will always inspire me to be more organised. Ioannis, thank you for teaching me cell culture. Hazel, thank you for your support with the project and for keeping the lab running. Roza, you have always been so kind and helpful, thank you so much for your advice! Lyndsay, thank you for your support and patience with my naïve coding questions. Willow, it's always a great day when you bless us with a visit from King's! Moira, thanks for your enthusiasm and patience with my first supervising experience. And Christine, who has kept me sane! Thanks for always being up for a rant or a chat, I couldn't have made it through without you.

I would like to give a big thanks to the Meehan lab, for welcoming me as an unofficial member. Richard, thank you for going above and beyond, your support and positive energy has been invaluable! Thank you to the members of the Meehan group: Katie, Sisi, Mihail, Olga Stepanova, Donncha and David, for the interesting scientific discussions and the laughs.

Thank you to all the people who have contributed to this project. To the members of my thesis committee, Greg Kudla and Hannah Long, thank you for your advice and guidance throughout the project. Thanks to all the technical services staff, especially to Jeff and Stephen in sequencing! Thank you to the FACS facility, and a special thanks to Lizzie - it's always a joy to have a chat with you when I come to count my cells. A big thank you to Gogo, from the bioinformatics team, for your help analysing my RNA-sequencing data, and thank you to Ian for your help with mapping the transposable elements.

I'm grateful to many people around the IGC who have filled my time here with fun and laughter. A special shout out to Shelagh, thank you for teaching me FISH and for blessing us

all with your amazing cakes. And thanks to members of the chatty office corner, past and present: Alba, Beth and Matilda. You all make the IGC a great place to be!

I was extremely lucky to join the IGC as part of a cohort of amazing students: Ailish, Eimear, Grace, Jan, Jon, Katie, Kelsey, Lizzy, Olivia, Yipeng, Yolanda. Thank you all for making lonch the highlight of my day. A special thanks to Grace, who brought us together for Wednesday zoom calls back when we weren't allowed to meet in person, for always making me laugh and being the best desk neighbour! A shout-out to my fellow yapper, Katie (and to E2, for putting up with us). Lizzy, thanks for being my late-night office companion and for your excellent book recommendations. Yipeng, thank you for your dedication to booking us in to the Black Cat pub quiz every week. Thanks to Kelsey, Ailish and Grace for somehow getting us all to the less-windy side of Rhodes and back in one piece. Big love to you all! <3

Thank you to my mum, who has always been there for me, encouraged me and given me the confidence that I can achieve whatever I set my mind to. I'm so lucky to have you. Thanks for always picking up the phone, even if it's 1am and I just want to rant! Thank you to Anne, for your love, support and cute dog pics.

Finally, thank you to Pablo. Your love and enthusiasm have kept me positive throughout this experience and your support means the world to me. I can't wait for our next adventure in London!!

Abstract

Loss of DNA methylation is common to the vast majority of human cancers and is considered a hallmark of cancer epigenomes. Current models propose that improper maintenance of DNA methylation could underpin this hypomethylation. However, these models are based on evidence derived from static snapshots of cancer methylomes. In addition, DNA methylation represents an important therapeutic target in cancer treatment. DNA hypomethylating agents (HMAs) are evidenced to promote antitumour immunity, however, existing HMAs have limited applications due to poor pharmacometrics and cytotoxicity. Thus, advancing our understanding of DNA hypomethylation in cancer, alongside identifying more effective drugs to target DNA methylation, is crucial for enhancing therapeutic strategies in cancer treatment.

To gain insight into the dynamics of DNA methylation maintenance in colorectal cancer cells, I developed a methylation-sensitive reporter system. This piggyBac transposon-based system allows stable integration of a methylated eGFP transgene (eGFPme) at distinct locations across the genome. I show that methylation of this reporter construct represses its expression in colorectal cancer cells. Reporter integrations were identified across the whole genome, enabling tracking of genome-wide DNA methylation maintenance. To test whether HCT116 cells are able to faithfully propagate DNA methylation, cells containing premethylated reporters were maintained in culture for up to 8 weeks, equivalent to ~75 cell divisions. The majority of methylated reporters remained repressed and maintained their methylated status over this time course. These findings demonstrate that methylation patterns remain consistently preserved at synthetic reporters in HCT116 colorectal cancer cells. My results highlight the functionality of the DNA methylation maintenance machinery in colorectal cancer cells, despite the presence of extensive hypomethylation.

Having demonstrated that DNA methylation is maintained in colorectal cancer cells, I next investigated the contribution of different components of the DNA methylation maintenance machinery to methylation maintenance. To do this, I assessed reporter activity in knock-out (KO) and degron cell lines. DNMT1 is the primary maintenance methyltransferase, responsible for maintaining DNA methylation following replication. As expected, reporter repression and methylation were not maintained in HCT116 cells lacking DNMT1 function. It has been previously reported that cooperation between DNMT1 and the *de novo* methyltransferase

DNMT3B is required to maintain DNA methylation in mouse embryonic stem cells. However, this has not previously been tested in human cancer cells. To investigate the contribution of DNMT3B to DNA methylation maintenance, I assessed reporter activity in DNMT3B^{KO} cells. No significant differences in reporter expression or methylation were detected in DNMT3B^{KO} cells compared to HCT116. My results indicate that DNMT3B does not have a detectable contribution, in this experimental system, to methylation maintenance in human HCT116 cells. Next, I investigated the contribution of UHRF1 to DNA methylation maintenance. UHRF1 is an emerging therapeutic target of interest due to its essential role in DNA methylation maintenance, and its overexpression in some cancers. I integrated the eGFPme reporter into DNMT1 and UHRF1 degron cell lines. This degron system enables rapid depletion of the degron-tagged proteins, and I monitored reporter activity in response to protein depletion. Following DNMT1 or UHRF1 depletion, rapid upregulation of the reporter gene was detected. These results support that targeting UHRF1 is an efficient method to deplete DNA methylation colorectal cancer cells. Furthermore, these findings demonstrate that the reporter is activated in response to acute loss of DNA methylation.

Finally, I sought to test whether the reporter system developed in this study could be utilised as a drug screening platform, to investigate compounds that modify DNA methylation. My previous results provided evidence for therapeutic targeting of DNMT1 and UHRF1. Therefore, I screened several novel inhibitors of DNMT1 and UHRF1. Removal of DNA methylation maintenance using the DNMT1 inhibitor GSK-3484862 resulted in a ~60% reduction of genome-wide methylation and rapid reporter activation. Importantly, I showed that GSK-3484862 treatment led to the activation of HERV-1 elements and IFN-1 signalling, which is thought to be important for the immunostimulatory applications of hypomethylating agents in cancer treatment. Together, my findings provide proof-of-concept for the use of this reporter system in large-scale compound library screen, to identify novel hypomethylating agents.

The methylation-sensitive reporter system developed in this study was utilised to provide novel insights into the mechanisms underlying DNA methylation maintenance in cancer. This reporter system has future applications as a screening platform, to identify genetic factors involved in DNA methylation, and novel hypomethylating compounds with therapeutic potential.

Lay abstract

Cancer is one of the leading causes of death worldwide, yet we still do not understand many of the changes that happen within a cancer cell. Cancer occurs when DNA is damaged, leading to changes in gene expression and abnormal activity of proteins that cause uncontrolled growth of cells. Each cancer is unique, but there are some common features that are seen in many different types of cancer. One alteration that is seen in cancer cells, but not healthy cells, is loss of an important chemical signal known as DNA methylation. DNA methylation is a chemical mark that is attached to the DNA, and it helps to provide instructions to the cell, regulating which genes get expressed and which ones do not. When DNA methylation is present at the regulatory region of a gene, this usually turns off expression of that gene. Correct patterns of DNA methylation are essential for healthy cells to function. However, in cancer, large sections of DNA lose the methylation signal. This loss of DNA methylation always occurs in the same locations, and these changes are observed in most human cancers. We do not currently understand why the DNA methylation signal is altered. However, we do know that some cancer cells depend on the remaining DNA methylation to keep growing. Drugs that can remove DNA methylation from the whole genome activate immune signalling from cancer cells which can help the body recognise and eliminate the cancer. These drugs have been successfully used to treat some cancers but have limited applications due to toxicity. Therefore, advancing our understanding of the processes driving changes in DNA methylation, alongside developing more effective drugs to target DNA methylation, is crucial for enhancing therapeutic strategies in cancer treatment.

In my PhD research, I aimed to develop a new tool to help us understand the changes to DNA methylation seen in cancer cells. To do this, I designed a methylation-sensitive reporter system using the Green Fluorescent Protein (GFP) reporter gene. A reporter gene is a gene that produces a protein, usually a fluorescent protein, that can be used to measure and visualise the activity of a gene. Reporter genes are powerful tools that can reveal the effects of different cellular process on expression of the reporter gene. To develop a methylation-sensitive reporter system, I first demonstrated that it was possible to add DNA methylation to the reporter gene, and that the methyl mark prevented expression of GFP in human cells.

Next, I introduced the GFP reporter gene into cancer cells and showed that it was possible to integrate this methylated reporter gene into the genome while maintaining its repression.

Having established a methylation-sensitive reporter system, I next used this system to investigate the changes to DNA methylation in cancer. We hypothesised that loss of DNA methylation in cancer was caused by a failure of mechanisms within the cell that maintain the methylation signature when the cell divides. To test this, I grew cancer cells that contained the methylated reporter gene for 8 weeks, which is equivalent to 75 cell divisions, and measured expression of GFP over this time. The reporter gene remained turned off and the methylation was faithfully maintained. These results provide evidence, for the first time, that the proteins that maintain DNA methylation are fully functional in cancer cells.

I was next interested in the potential application of this reporter system as a screening platform that could be used to identify new drug targets and therapeutic compounds. Identifying drug targets is a critical step in the development of effective treatments. Therefore, I investigated which proteins are involved in maintaining DNA methylation. I introduced the methylated reporter gene into cells that were genetically engineered to remove key proteins involved in DNA methylation and showed that two proteins, DNMT1 and UHRF1, are primarily responsible for maintaining methylation in human cells. Next, I utilised the reporter system to screen new drugs that inhibit activity of DNMT1 and UHRF1. Inhibitors are a vital type of drug used in cancer therapy, that offer a targeted approach to disrupt pathways that cancer cells rely on for growth. I showed that a novel inhibitor of DNMT1 activated reporter expression, was well-tolerated by cells and activated immune signalling, making it a promising candidate for future cancer treatment strategies.

In summary, the methylation-sensitive reporter system developed in this thesis was utilised to provide novel insights into the mechanisms underlying DNA methylation maintenance in cancer. This reporter system has future applications in screening, to identify new genes involved in DNA methylation and new drugs that can modify methylation for cancer therapy.

Abbreviations

5-aza	5-aza-cytidine
5-aza-dC	5-aza-2'-deoxycytidine
5mC	5-Methylcytosine
AID	Auxin-inducible degron
BS	Bisulfite
C	Cytosine
CAG	Synthetic promoter containing CMV enhancer and chicken β-actin promoter
CGI	CpG Island
CMV	Cytomegalovirus
CpG	Cytosine and guanine separated by one phosphate group
CTA	Cancer testis antigen
DE	Differentially expressed
DKO	Double knock-out of DNMT1 and DNMT3B in a HCT116 cell line
DNA	Deoxyribonucleic Acid
DNMT1	DNA Methyltransferase 1
DNMT1^{KO}	DNMT1 knock-out HCT116 cell line
DNMT3B	DNA Methyltransferase 3B
DNMT3B^{KO}	DNMT3B knock-out HCT116 cell line
dsRNA	Double-stranded RNA
eGFP	Enhanced Green Fluorescent Protein
eGFPme	Methylated eGFP reporter gene
EM	Enzymatic-Methyl
ESC	Embryonic Stem Cell
G	Guanine
gDNA	Genomic DNA
GFP	Green fluorescent protein
GSEA	Gene set enrichment analysis
GSK-3484862	Small molecule inhibitor of DNMT1
HCT116	Human colorectal cancer cell line

HERV	Human endogenous retrovirus
HMA	Hypomethylating agent
IFN	Interferon
ITR	Inverted Terminal Repeat
LINE1	Long interspaced nuclear element
LTR	Long terminal repeat
mCpG	Methylated CpG site
MFI	Median Fluorescence Intensity
PCR	Polymerase Chain Reaction
PMD	Partially Methylated Domain
RKO	Human colon carcinoma cell line
RMCE	Recombination-Mediated Cassette Exchange
RNA	Ribonucleic acid
SAM	S-adenosylmethionine
TSS	Transcriptional Start Site
UHRF1	Ubiquitin-like with PHD and RING finger domains 1, E3-ubiquitin ligase
WT	Wild-type

Table of contents

1 Introduction

1.1 Epigenetics

The field of genetics encompasses not only the study of the DNA sequence and individual genes, but also the entire genome – an organisms complete set of genetic material – and the regulatory factors that govern genome organisation and gene expression. Proper genome regulation is essential for ensuring the correct spatial and temporal expression of genes throughout development and for maintaining the identity of specialized cell types. This regulation is dynamic, allowing cells to respond to internal and external stimuli. Epigenetic mechanisms play a crucial role in this regulation (Jaenisch and Bird, 2003).

The term ‘epigenetics’ was first introduced by C.H. Waddington in 1942, to describe how a single genotype can give rise to multiple phenotypes through complex developmental processes (Waddington, 1942, 2012). It was not until the 1970s that DNA methylation was identified as a chemical modification that could influence gene expression, providing the first evidence in support of Waddington's ideas (Holliday and Pugh, 1975; Riggs, 1975). Since then, numerous mechanisms of gene regulation have been identified (Gibney and Nolan, 2010). The term ‘epigenetics’ is now used to refer to mechanisms that influence gene expression without altering the underlying DNA sequence (Russo, Martienssen and Riggs, 1996; Waterland, 2006). Bird (2007) described these epigenetic mechanisms as the “structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states” (Bird, 2007).

In eukaryotes, the genome is organised into chromosomes within the nucleus of a cell (Fig. 1). DNA is wrapped around nucleosomes, which consist of an octamer of histone proteins, providing structural support (Luger, 1997). The resulting DNA-histone complex is known as chromatin (Kornberg, 1974; Olins and Olins, 1978). Chromatin is compacted into higher-order structures, typically classified as either euchromatin or heterochromatin. Euchromatin is less condensed, which is thought to make the DNA more accessible for transcription and facilitate gene expression. In contrast, heterochromatin is tightly packed, potentially restricting access to the DNA and repressing gene expression (Smith and Peterson, 2004; Murakami, 2013). During cell division, chromatin further condenses into chromatids, which then form

chromosomes (Antonin and Neumann, 2016). Epigenetic factors play a crucial role in modifying the chromatin state and influencing the interaction of proteins with the DNA, thereby regulating gene expression. These processes include DNA modifications, such as DNA methylation, and histone tail modifications, among others (Gibney and Nolan, 2010). This thesis will focus primarily on the epigenetic signal DNA methylation.

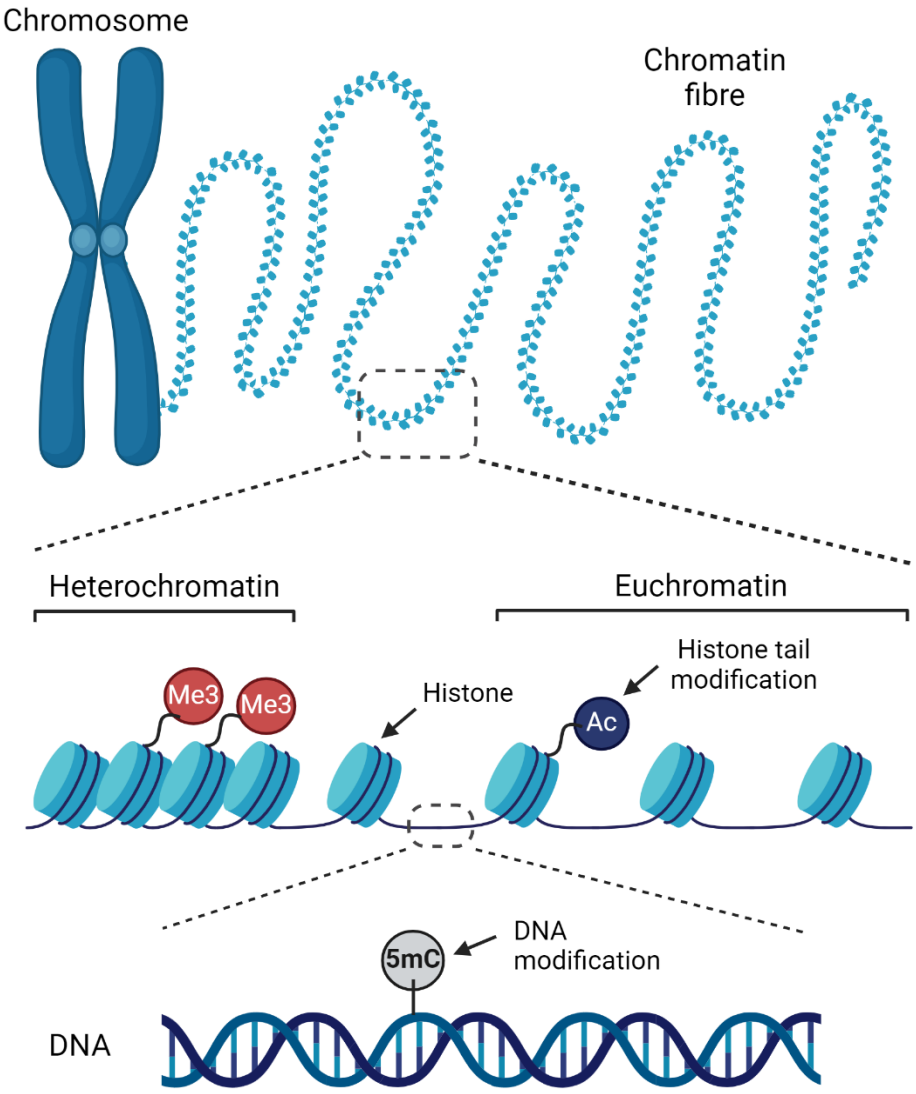


Figure 1-1: Chromatin structure and organisation of DNA in the nucleus. *Graphical representation of DNA organisation into chromosomes. The DNA double-helix is composed of four nucleotides, A, T, G and C which form base-pairs. DNA can be chemically modified, for example by the addition of a methyl group to Cytosine, which acts as an 'epigenetic' signal. DNA is wrapped around proteins known as histones and this structure is known as the nucleosome. Histone tails can also be chemically modified, an important type of epigenetic mark that can influence chromatin structure. Nucleosomes form the chromatin fibre, which can be condensed (heterochromatin) or open (euchromatin). This is thought to influence the accessibility of DNA, thereby regulating gene expression (Jaenisch and Bird, 2003). Figure created using BioRender.com.*

1.2 DNA Methylation

1.2.1 DNA methylation and its function

DNA methylation is a heritable epigenetic modification with an important role in mammalian development and genome regulation (Moore, Le and Fan, 2013). The DNA molecule is a polynucleotide, composed of four different nucleotide subunits: adenine, thymine, cytosine and guanine (Alberts *et al.*, 2002). In mammals, methylation of DNA involves the covalent transfer of a methyl group from the methyl donor S-adenosyl-L-methionine (SAM) to carbon-5 (C5) of cytosine residues, generating a chemically modified nucleotide known as 5-methylcytosine (5mC, Fig. 1.2) (Moore, Le and Fan, 2013). DNA methylation primarily occurs at cytosine residues in mammals (Cooper, 1983). Cytosine methylation does not alter the Watson-Crick base pairing of DNA. However, the methyl group is located within the major groove of the cytosine base, which can influence how proteins interact with the DNA (Jeltsch, 2002a; Dantas Machado *et al.*, 2015). Thus, DNA methylation acts as an epigenetic signal, which is recognised by proteins and influences cellular function.

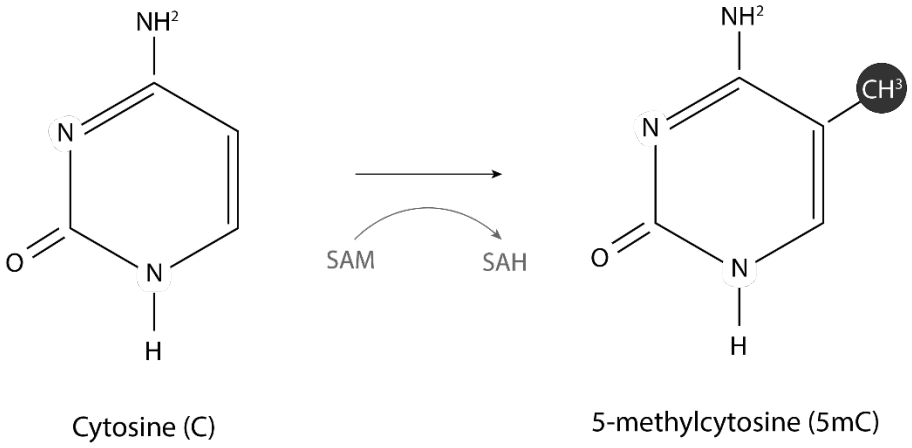


Figure 1-2: Chemical structures of cytosine and 5-methylcytosine. *Cytosine methylation involves the covalent transfer of a methyl group to the 5th carbon of Cytosine, forming the modified nucleotide 5-methylcytosine (5mC). The methyl group is provided by S-adenosyl-L-methionine (SAM) which is converted to S-adenosylhomocysteine (SAH) during this process (Moore, Le and Fan, 2013).*

Cytosine methylation is generally found in the context of symmetrical CpG dinucleotides, that is a cytosine followed by a guanine in the 5'-3' direction of the DNA strand (Schübeler, 2015). Overall, the human genome is largely depleted of CpG sites. 5mC is mutagenic by nature and

can spontaneously deaminate to thymine. It has been postulated that as a result of this, the genome has become depleted of CpG sites over the course of evolution (Coulondre *et al.*, 1978; Bird, 1980). However, CpG dinucleotides are enriched at gene regulatory regions in clusters known as CpG islands (CGIs) (Bird *et al.*, 1985; Cross *et al.*, 1994). CGIs are short regions of DNA around 1 kilobase (kb) in size and constitute approximately 1% of the human genome. These regions have a high GC base composition and around 10 times higher frequency of CpG dinucleotides than the remaining genome (Bird *et al.*, 1985; Bird, 1986; Cross *et al.*, 1994; Illingworth *et al.*, 2010).

CGIs are believed to act as regulatory elements that serve as transcription initiation sites and promote gene expression (Cain, Montibus and Oakey, 2022). Approximately 59% of human CGIs coincide the transcription start site of protein-coding genes, which accounts for ~72% of known transcription start sites (Saxonov, Berg and Brutlag, 2006; Illingworth *et al.*, 2010). Around 20% of CGIs display promoter activity but are located distally from annotated genes (Illingworth *et al.*, 2010). CGIs lack typical promoter elements, such as TATA boxes (Reynolds *et al.*, 1984; Carninci *et al.*, 2006), but are less likely to assemble into stable nucleosomes (Ramirez-Carrozzi *et al.*, 2009). This may promote the formation of an accessible, transcriptionally permissive environment at the gene promoter region. Most CGIs associated with gene promoters remain unmethylated in normal tissues (Fig. 1.3) (Bird *et al.*, 1987; Shen *et al.*, 2007). Unmethylated CGIs facilitate the binding of chromatin modifiers and transcription factors, thereby promoting transcription of downstream genes (Long, Blackledge and Klose, 2013; Blackledge *et al.*, 2014). It has been proposed that the binding of these factors protects CGIs from gaining methylation, thereby maintaining their unmethylated status (Lienert *et al.*, 2011; Krebs *et al.*, 2014; Saadeh and Schulz, 2014).

Methylation of CGIs located within the regulatory region of a gene is associated with repression of transcription and gene silencing (Bird, 1986; Mohn *et al.*, 2008). This association has led to the conventional description of DNA methylation as a repressive epigenetic signal. In support of this, DNA methylation has been evidenced to directly repress gene expression by inhibiting the binding of transcription factors to their recognition motif (Watt and Molloy, 1988; Bednarik *et al.*, 1991; Campanero, Armstrong and Flemington, 2000; Kaluscha *et al.*, 2022). Furthermore, recognition of DNA methylation by methyl-binding proteins facilitates the formation of repressive heterochromatin environments (Bogdanović and Veenstra, 2009;

Rose and Klose, 2014). However, the role of DNA methylation in regulating gene expression is debatable. Only ~11% of all CGIs in somatic tissue are stably methylated (Illingworth *et al.*, 2010). Shen *et al*. (2007) showed that ~4% of CGIs associated with gene promoters were highly methylated in normal blood. This represents a small, but nonetheless notable, selection of genes that are stably repressed by methylation in somatic cells. Germline genes, such as the cancer testis antigen (CTA) gene family, genes involved in pluripotency and differentiation, and imprinted genes are among those that are repressed by methylation (De Smet *et al.*, 1999; Farthing *et al.*, 2008; Meissner *et al.*, 2008; Mohn *et al.*, 2008; Butz *et al.*, 2022). It has been suggested that promoter CGI methylation is associated with lineage-specific gene silencing and serves to maintain tissue-specific gene expression profiles (Shen *et al.*, 2007; Mohn *et al.*, 2008). Whether DNA methylation establishes gene repression, or is a consequence of it, remains unclear. Alternative mechanisms of gene silencing, for example repressive histone marks, have been evidenced to precede DNA methylation, suggesting that methylation may serve to stabilise the repressive state (Mutskov and Felsenfeld, 2004; Payer and Lee, 2008). However, DNA methylation is required for maintaining gene inactivation, and removal of DNA methylation leads to reactivation of silenced genes (De Smet *et al.*, 1999; Sado *et al.*, 2000).

Outside of CGIs, CpG sites are highly methylated across the human genome (Fig. 1.3). Approximately 80% of all CpG sites are methylated in somatic tissues (Ehrlich *et al.*, 1982; Gama-Sosa *et al.*, 1983; Suzuki and Bird, 2008; Li and Zhang, 2014). A key role for DNA methylation outside of CGI gene regulation is repression of repeat sequences and transposable elements (Fig. 1.3). Retrotransposons and repeats account for ~45% of the human genome (Schulz, Steinhoff and Florl, 2006; Cordaux and Batzer, 2009). These are mobile genetic elements which can contribute to genome instability by generating insertion mutations (Cordaux and Batzer, 2009; Crichton *et al.*, 2014). Some of the most abundant retrotransposons in the human genome are long terminal repeat (LTR) elements, also known as human endogenous retroviruses (HERVs), and the non-LTR small interspaced nuclear elements (SINEs) and long interspaced nuclear elements (LINEs). Transposable elements are transcribed into RNA intermediates then reverse-transcribed into DNA, allowing them to transpose through a 'copy and paste' mechanism (Schulz, Steinhoff and Florl, 2006). In order to protect against the activity of retrotransposons, these elements are silenced by DNA

methylation. Retrotransposable elements are highly enriched for methylated CpG sites, which may function to directly block transcription factor binding or to recruit other repressive mechanisms (Lee *et al.*, 2010; Rowe and Trono, 2011; Crichton *et al.*, 2014). Furthermore, over time, C to T deamination mutations at methylated CpG sites contribute to inactivation of the transposon sequence (Yoder, Walsh and Bestor, 1997). Loss of DNA methylation leads to activation of retroviral transcription, particularly of ERV elements (Li, Bestor and Jaenisch, 1992; Walsh, Chaillet and Bestor, 1998; Okano *et al.*, 1999; Maksakova, Mager and Reiss, 2008).

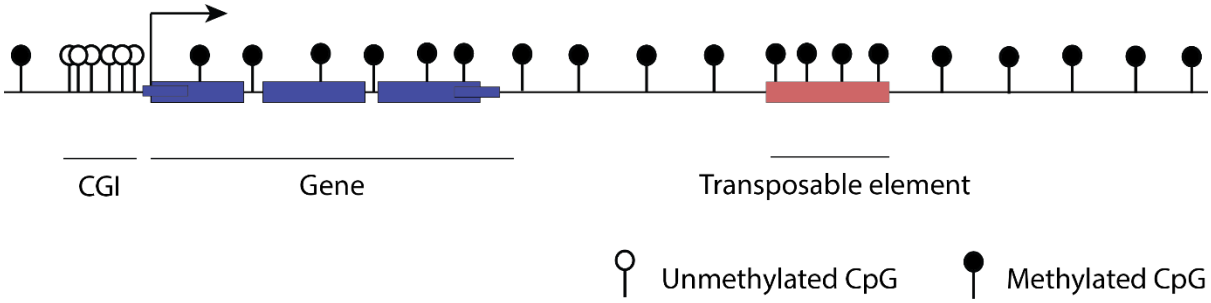


Figure 1-3: DNA methylation patterns in normal cells. *In most healthy tissues, the vast majority of CpG sites (>80%) across the genome are fully methylated, except for CGIs. CGIs are enriched at gene regulatory regions and remain unmethylated, facilitating gene expression. Transposable elements are highly methylated in cells, which is thought to maintain their repression (Li, Bestor and Jaenisch, 1992; Illingworth* et al.*, 2010).*

1.2.2 The DNA methyltransferases (DNMTs)

DNA methylation is catalysed by a family of enzymes known as the DNA methyltransferases (DNMTs). The DNMTs catalyse the transfer of a methyl group from the methyl-donor S-adenosylmethionine (SAM) to cytosine residues in the DNA sequence (Lyko, 2018). Briefly, this process involves binding of the DNMT to the DNA and rotation of the target nucleotide out of the double-helix so that it is accessible, a process known as 'base flipping' (Jeltsch, 2002a). The guanine that was base-paired with the cytosine is held in place by hydrogen bonding with the DNMT (Cheng and Blumenthal, 1996). The DNMT then forms a covalent intermediate through nucleophilic attack on cytosine C6, which activates C5. Subsequently, the methyl group is transferred from SAM to the activated C5. The reaction intermediate is

resolved and the DNMT releases, resulting in the formation of 5mC (Cheng and Blumenthal, 1996; Jeltsch, 2002a; Walbott *et al.*, 2007).

In humans, the DNMTs are divided into two families: the DNMT3 family, which is primarily responsible for establishing methylation on previously unmethylated DNA during development, and are therefore known as the *de novo* methyltransferases, and DNMT1, which propagates methylation patterns onto newly synthesized DNA during cell division (Fig. 1.4) (Okano *et al.*, 1999; Robert *et al.*, 2003; Hermann, Goyal and Jeltsch, 2004). There are three catalytically active DNMTs in humans: DNMT3A, DNMT3B and DNMT1. The catalytic methyltransferase domain of these enzymes is highly conserved between mammals, plants and bacteria (Bestor *et al.*, 1988). There is a third, catalytically inactive member of the DNMT3 family, DNMT3L (Aapola et al., 2000). Although DNMT3L does not have methyltransferase activity, it is expressed during development and has a role in facilitating *de novo* methylation by stimulating activity of DNMT3A and DNMT3B (Chédin, Lieber and Hsieh, 2002; Hervouet et al., 2018; Veland et al., 2019).

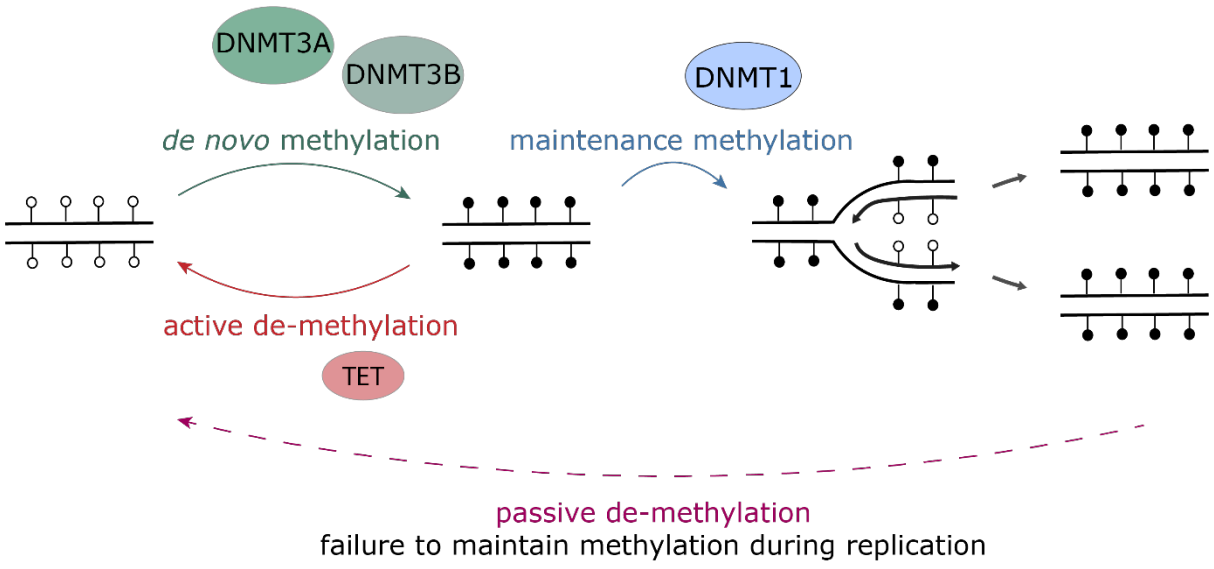


Figure 1-4: The DNA methyltransferases (DNMTs) catalyse methylation of DNA. *Initial methylation patterns are established on previously unmethylated DNA by DNMT3A and DNMT3B. Following DNA replication, DNMT1 maintains methylation on the newly synthesised DNA. Failure to correctly maintain DNA methylation can lead to passive demethylation over time. DNA methylation can also be actively removed by the TET enzymes. Adapted from (Turpin and Salbert, 2022).*

The catalytic domain is located at the C-terminus of the protein, and this domain contains conserved sequence motifs that are required for recognition and binding of DNA, cytosine base-flipping and methyl group transfer (Fig. 1.5) (Cheng and Blumenthal, 1996; Jeltsch, 2002b). The N-terminal region is more variable between the different DNMTs. The N-terminal domain directs enzyme-specific nuclear localisation and chromatin binding (Hervouet *et al.*, 2018) . The N-terminal domains of DNMT3A and DNMT3B contain a Proline-Tryptophan-Tryptophan-Proline (PWWP) domain which is involved in DNA binding and chromatin interaction, and a plant homeodomain (PHD)-like ATRX-DNMT3-DNMT3L (ADD) domain which is involved in protein-protein interactions (Jeltsch and Jurkowska, 2016). The N-terminal of DNMT1 contains various domains that facilitate the protein's function including the DNMT1-Associated Protein 1 domain (DMAPD), the Proliferating Cell Nuclear Antigen (PCNA)-binding domain (PBD), the Replication Foci-Targeting domain (RFTD), the CXXC zinc finger domain, and the tandem Bromo Adjacent Homology (BAH) 1 and 2 domains (Leonhardt *et al.*, 1992; Egger *et al.*, 2006; Song *et al.*, 2011; Hervouet *et al.*, 2018).

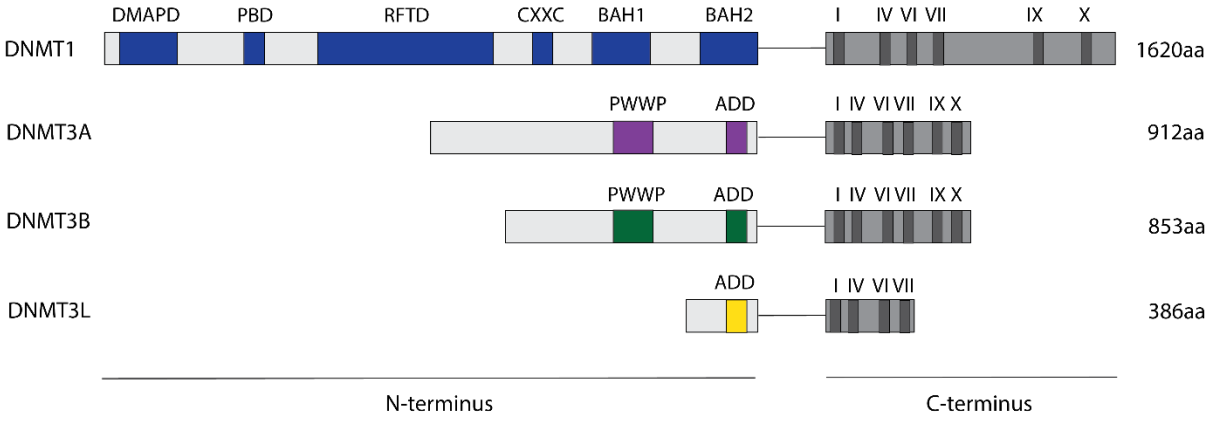


Figure 1-5: DNA methyltransferase (DNMT) domain structure. *Graphical representation of the DNMT enzymes. The enzyme is divided into the C-terminus, which is the catalytic methyltransferase domain, and the N-terminus, which is involved in enzyme localisation, chromatin interaction and protein-protein interactions. The domains are linked through a lysine-glycine repeat (GK)n linker. DNMT3L is catalytically inactive. Adapted from (Jeltsch and Jurkowska, 2016)*

1.2.3 *De novo* methylation

During development, DNA methylation patterns must be removed to enable expression of pluripotency factors, and then re-established (Reik, Dean and Walter, 2001). A first wave of

genome-wide demethylation occurs in the primordial germ cells, the precursors of mature male and female germ cells. Re-methylation occurs in the germ line several days later, and the mature sperm and oocytes display high methylation levels. After fertilisation, a second wave of demethylation occurs in the early embryo. At the blastocyst stage, around the time of implantation, methylation patterns are re-established (Reik, Dean and Walter, 2001; Hemberger, Dean and Reik, 2009). It is worth noting that some repeat sequences, and methylated imprinted genes, do not become demethylated (Reik, Dean and Walter, 2001). Methylation patterns are established during development by the DNMT3 family of *de novo* methyltransferases. This is an essential process, which has a role in regulating genomic imprinting, X-chromosome inactivation, retrotransposon silencing and reinforcing cell-lineage specific gene expression patterns (Okano, Xie and Li, 1998; Bourc'his and Bestor, 2004; Auclair *et al.*, 2014). DNA methylation is primarily targeted to the X-chromosome, and to germline and lineage-specific genes (Borgel *et al.*, 2010; Auclair *et al.*, 2014). During differentiation, lineage-specific genes are demethylated to allow their expression (Borgel *et al.*, 2010).

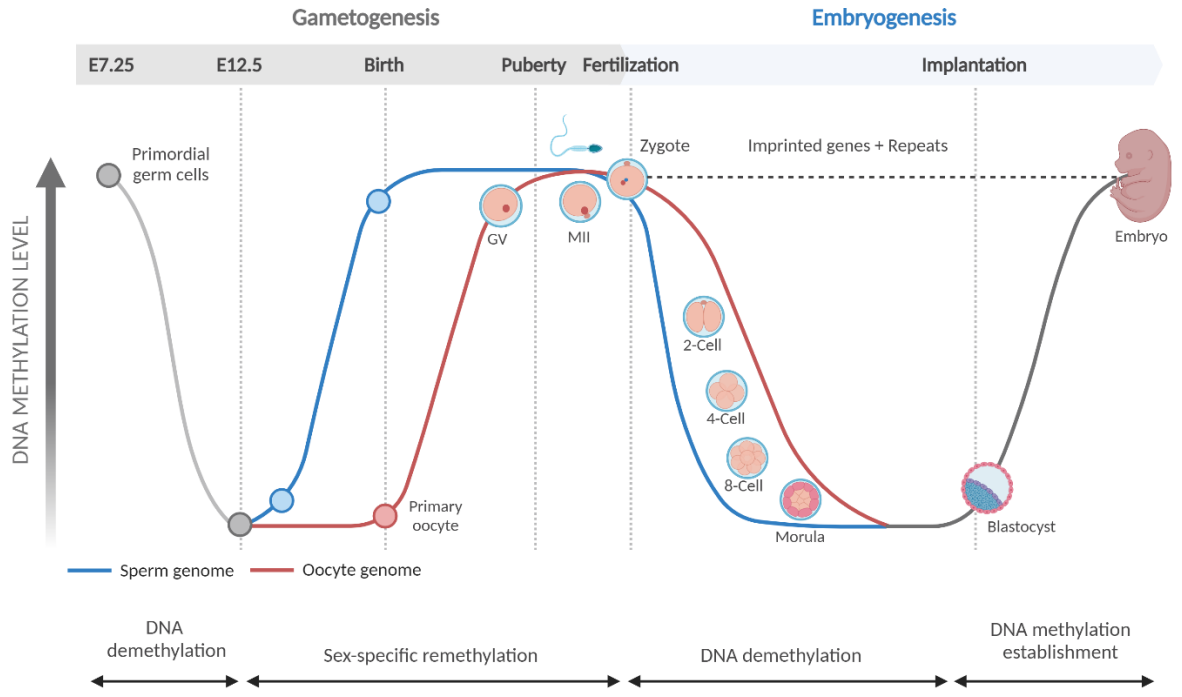


Figure 1-6: Process of developmental DNA methylation. *DNA methylation patterns are reset during gametogenesis, and again during embryogenesis. This process is necessary to enable activation of pluripotency genes. Importantly, some imprinted genes and repeat sequences are not demethylated during embryogenesis (Reik, Dean and Walter, 2001). Figure adapted from "DNA Methylation Levels During Mammalian Development" BioRender Template (Smallwood and Kelsey, 2012).*

Several findings have indicated that the DNMT3 family are the primary enzymes involved in *de novo* methylation. The catalytically active DNMT3A and DNMT3B have been shown to methylate DNA *in vitro* with no preference towards unmethylated or hemi-methylated DNA (Okano, Xie and Li, 1998). Okano *et al*. (1999) demonstrated that inactivation of DNMT3A and DNM3B prevents *de novo* methylation in mouse embryonic stem cells (ESCs), with no effect on pre-existing methylation patterns (Okano *et al.*, 1999). Furthermore, DNMT3A and DNMT3B are highly expressed during embryonic development and their expression correlates with the global establishment of methylation (Sperger *et al.*, 2003; Enver *et al.*, 2005; Auclair *et al.*, 2014). Their expression is reduced in somatic cells, emphasising their role in developmental methylation (Okano, Xie and Li, 1998; Li and Zhang, 2014).

The *de novo* methyltransferases are essential for normal development. Mice lacking activity of DNMT3A develop to term but die shortly after birth. Loss of DNMT3B results in embryonic lethality. Mice lacking activity of DNMT3B develop normally until embryonic age E9.5 but die by E11.5-15.5. Combined loss of DNMT3A and DNMT3B causes growth and morphological defects, and death by E11.5 (Li, Bestor and Jaenisch, 1992; Okano *et al.*, 1999). In humans, mutations in DNMT3B are associated with Immunodeficiency, Centromere instability and Facial anomalies (ICF) Syndrome (Smeets *et al.*, 1994). This condition is characterised by reduced DNMT3B activity causing hypomethylation at heterochromatic regions, repeat sequences and transposons (Heyn *et al.*, 2012). Patients present with immunodeficiency, facial development anomalies and learning difficulties (Ehrlich, Jackson and Weemaes, 2006). In addition, DNMT3B is overexpressed in several human cancers (Peralta-Arrieta *et al.*, 2017). Germline variants in DNMT3A are associated with Tatton-Brown-Rahman syndrome (TBRS) and cause overgrowth, obesity, and learning difficulties (Ostrowski and Tatton-Brown, 1993; Smith *et al.*, 2021). Gain-of-function mutations in DNMT3A have been associated with microcephalic dwarfism. These mutations result in mistargeting of DNMT3A, leading to aberrant *de novo* methylation at developmental genes and Polycomb-marked regions (Heyn *et al.*, 2012). Furthermore, DNMT3A is frequently mutated in acute myeloid leukaemia (AML) and myelodysplastic syndromes (MDS). Mutation of DNMT3A can initiate carcinogenesis (Ley *et al.*, 2010; Walter *et al.*, 2011). The two DNMT3 isoforms display ~40% sequence similarity, but these phenotypes indicate that the two proteins have non-redundant functions (Okano *et al.*, 1999).

During embryogenesis, DNMT3A and DNMT3B display different spatial and temporal expression patterns (Watanabe *et al.*, 2002; Watanabe, Uchiyama and Hanaoka, 2006). In developing mouse embryos, DNMT3B is expressed earlier than DNMT3A. Watanabe *et al*. (2002) detected expression of DNMT3B between E4.5-8.5, during which time genome-wide methylation occurs, and its expression decreases after E10.5. Expression of DNMT3A was not detected until E10.5, and it was expressed ubiquitously thereafter. The *de novo* DNMTs display tissue-specific expression patterns and are active in different cell types (Watanabe *et al.*, 2002; Watanabe, Uchiyama and Hanaoka, 2006). Together, these findings indicate that DNMT3A and DNMT3B display distinct functions during development.

The *de novo* DNMTs display unique localisation patterns, which may explain their distinct functions (Bachman, Rountree and Baylin, 2001). Both DNMT3A and DNMT3B have an important role in methylating repeat sequences, but are targeted to different regions. DNMT3A displays preference towards methylating major satellite repeats, located in the pericentromeric region, and endogenous retroviral DNA. In contrast, DNMT3B preferentially methylates centromeric minor satellite repeats (Okano *et al.*, 1999; Chen *et al.*, 2003). DNMT3A and DNMT3B also show preference towards different genic regions (Chen *et al.*, 2003).

Localisation of DNMT3A and DNMT3B is thought to be driven by interaction with chromatin through the N-terminal domain. The N-terminal domain contains two functional domains: the ADD domain and the PWWP domain (Fig. 1.5). The ADD domain facilitates interaction with chromatin, partly through recognition of methylated histones. The ADD domain of DNMT3A recognises unmethylated lysine 4 in histone 3 (H3K4), which recruits DNMT3A to chromatin and switches the enzyme into an active conformation. This may facilitate DNMT3A activity in unmethylated H3K4 marked regions or serve to exclude it from actively transcribed H3K4me regions (Otani *et al.*, 2009). Conversely, although DNMT3B also interacts with H3K4, its activity appears to be independent of this interaction (Boyko *et al.*, 2022). The ADD domain additionally has a role in recruiting DNMT3A and DNMT3B to H3K9 marked heterochromatic regions (Fuks *et al.*, 2003; Lehnertz *et al.*, 2003).The PWWP domain is also involved in the functional specialisation of DNMT3A and DNMT3B. The PWWP domain of DNMT3A displays low DNA binding activity, whereas the DNMT3B PWWP domain binds to DNA non-specifically (Chen, Tsujimoto and Li, 2004). The PWWP domain facilitates interaction with H3K36me.

DNMT3B appears to be preferentially targeted towards H3K36me3 and is suggested to have a role in methylating the bodies of transcribed genes (Baubec *et al.*, 2015; Rondelet *et al.*, 2016). DNMT3A recognises H3K36me3 but enriched at H3K36me2 regions, where it may have a role in methylating intergenic regions (Dhayalan *et al.*, 2010; Weinberg *et al.*, 2019). The remaining N-terminal is the least conserved region of the DNMT3 proteins (Manzo *et al.*, 2017). It is unstructured and less well-understood, but appears to contribute to region-specific targeting. DNMT3A is directed towards H3K27me3 marked regions via its N-terminal domain (Manzo *et al.*, 2017; Weinberg *et al.*, 2021). Furthermore, the N-terminal domain of DNMT3A is evidence to interact with H2AK119ub1 (Chen *et al.*, 2024). This interaction targets DNMT3A to bivalent promoters in mouse brain cortical cells (Parry and Reik, 2022), and is essential for mouse postnatal development (Gu *et al.*, 2022). Instead, the N-terminal region of DNMT3B facilitates its recruitment to H3K9me3 marked heterochromatin through interaction with HP1- α (Taglini *et al.*, 2024).

1.2.4 Maintenance of DNA methylation

Once DNA methylation is established, it is essential that methylation patterns are stably maintained. When cells divide and DNA is replicated, the newly synthesised daughter strand is unmethylated. In this state the DNA is referred to as being hemi-methylated, meaning that the CpG site is methylated only on one strand of the DNA (Greenberg and Bourc'his, 2019). The process of copying methylation onto the newly synthesised DNA is known as DNA methylation maintenance and is predominately accomplished through the activity of DNMT1. The role of DNMT1 as a maintenance methyltransferase is highlighted by its strong preference towards targeting hemi-methylated CpG sites *in vitro* (Bestor and Ingram, 1983; Hermann, Goyal and Jeltsch, 2004). DNMT1 is recruited to replication foci during S-phase (Leonhardt *et al.*, 1992). In non-S-phase cells, DNMT1 displays more diffuse distribution in the nucleus but remains associated with highly methylated heterochromatic regions throughout G2 and M (Leonhardt *et al.*, 1992; Easwaran *et al.*, 2004). DNMT1 is predominately expressed in dividing tissues and its expression is reduced in terminally differentiated, non-dividing tissue (Robertson *et al.*, 1999; Sen *et al.*, 2010).

Expression of DNMT1 is essential. Loss of its activity leads to global hypomethylation which is detrimental during development and in somatic cells (Chen *et al.*, 2007; Liao *et al.*, 2015). Loss of DNMT1 activity is lethal in all somatic dividing tissues, emphasising the importance of methylation maintenance (Li, Bestor and Jaenisch, 1992; Jackson-Grusby *et al.*, 2001; Trowbridge *et al.*, 2009; Sen *et al.*, 2010). Loss of DNMT1 activity in developing mouse embryos impedes development, with embryos showing 30% reduced methylation levels compared to wild-type, and ultimately results in embryonic lethality (Li, Bestor and Jaenisch, 1992). In mouse ESCs, cells tolerate loss of DNMT1 and remain viable but fail to differentiate upon induction of differentiation (Jackson *et al.*, 2004). These findings indicate that maintenance of DNA methylation is essential for differentiation during embryonic development. In human ESCs, DNMT1 is essential for viability. Conditional depletion of DNMT1 in human ESCs results in increased DNA damage, cell cycle arrest at G1, and rapid cell death (Liao *et al.*, 2015). Cell viability relies specifically on the catalytic domain of DNMT1. Liao *et al*. (2015) demonstrated that the phenotype of DNMT1$^{-/-}$ human ESCs could not be rescued by catalytically inactive DNMT1. Furthermore, somatic cells tolerate partial deletion of DNMT1 by expressing a catalytically active, truncated protein (Egger *et al.*, 2006).

The specificity and targeting of DNMT1 is regulated through various intermolecular and intramolecular interactions, which serve to ensure the fidelity of methylation maintenance and stable epigenetic inheritance (Jeltsch and Jurkowska, 2016). Multiple autoinhibitory mechanisms prevent aberrant methylation by DNMT1. In the absence of DNA, the RFTD domain is inserted into the catalytic pocket of DNMT1, preventing its activity (Syeda *et al.*, 2011; Takeshita *et al.*, 2011). In the presence of unmethylated DNA, the CXXC domain of DNMT1 selectively binds unmethylated CpG dinucleotides (Frauer *et al.*, 2011). This positions the CXXC-BAH1 linker within the active site of DNMT1, preventing aberrant *de novo* methylation (Song *et al.*, 2011). DNMT1 is targeted to replicating DNA during early S-phase through interaction of its PBD domain with PCNA at the replication fork (Chuang *et al.*, 1997; Egger *et al.*, 2006). The PBD domain is required for PCNA-mediated recruitment to replication foci (Easwaran *et al.*, 2004), but is not strictly required for methylation maintenance (Spada *et al.*, 2007). The RFTD domain also has a role in targeting DNMT1 to replication sites (Easwaran *et al.*, 2004; Schneider *et al.*, 2013). The recruitment of DNMT1 to hemi-

methylated DNA is dependent upon interaction of the RFTD domain with Ubiquitin-like, containing PHD and RING Finger domains 1 (UHRF1) (Achour *et al.*, 2008; Qin *et al.*, 2015)

UHRF1 has a crucial role in DNA methylation maintenance. UHRF1 is a multi-functional chromatin binding protein with 5 domains: a Ubiquitin-like domain (UBL), Tandem Tudor Domain (TTD), plant homeodomain (PHD), SET and RING-associated (SRA) domain and Really Interesting New Gene (RING) domain (Fig. 1.6) (Mancini, Magnani, Macchi and I. M. Bonapace, 2021). The SRA domain of UHRF1 recognises and preferentially binds to hemi-methylated CpG sites (Bostick *et al.*, 2007; Arita *et al.*, 2008; Avvakumov *et al.*, 2008). The SRA domain of UHRF1 interacts with the RFTD domain of DNMT1, tethering DNMT1 to chromatin (Bostick *et al.*, 2007). UHRF1 also ubiquitinates histone H3 lysine residues 14, 18, 23 and 27 through combined activity of its UBL and RING finger domains (Qin *et al.*, 2015; Harrison *et al.*, 2016). H3K18Ub and H3K23Ub bind to the Ubiquitin-Interacting Motif (UIM) within the RFTD domain of DNMT1, resulting in spatial rearrangement that displaces the RFTD domain and opens the active site of DNMT1 (Bashtrykov *et al.*, 2014; Ishiyama *et al.*, 2017; Nishiyama *et al.*, 2020). Furthermore, UHRF1 has a role in directing DNMT1 to replication sites through its interaction with DNA Ligase 1 (LIG1), a component of the DNA replication machinery (Ferry *et al.*, 2017). UHRF1 also recognises H3K9me3 through its TTD and has a role in recruiting DNMT1 to heterochromatic regions marked by H3K9me3 (Nady *et al.*, 2011; Rothbart *et al.*, 2012; Liu *et al.*, 2013). Thus, UHRF1 acts as an upstream activator of DNMT1, increasing DNMT1 specificity and stimulating methylation maintenance and (Bashtrykov *et al.*, 2014; Berkyurek *et al.*, 2014). UHRF1 appears to be essential for stable propagation of DNA methylation. Loss of UHRF1 has been evidenced to impair global methylation maintenance to a similar extent as loss of DNMT1 activity, in mouse embryos and ESCs (Bostick *et al.*, 2007; Sharif *et al.*, 2007). In HCT116 cells depletion of UHRF1 induces greater methylation loss than depletion of DNMT1, suggesting that UHRF1 may have roles outside of DNMT1 recruitment (Yamaguchi *et al.*, 2023). UHRF1 has been evidenced to interact with DNMT3A and DNMT3B, and antagonise TET demethylation activity, which may explain the more severe hypomethylation phenotype (Yamaguchi *et al.*, 2023).

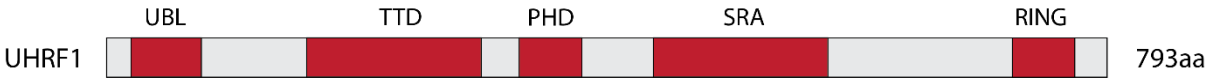


Figure 1-7: UHRF1 domain structure. *Graphical representation of UHRF1 protein structure. UHRF1 is composed of a Ubiquitin-like domain (UBL), Tandem Tudor Domain (TTD), plant homeodomain (PHD), SET and RING-associated (SRA) domain and Really Interesting New Gene (RING) domain. Adapted from (Mancini, Magnani, Macchi and I. Bonapace, 2021).*

Methylation of the daughter strand is a complex process, which must be co-ordinated with processive DNA synthesis and chromatin assembly. Once recruited to replication foci, DNMT1 acts processivity to methylate newly synthesised DNA (Hermann, Goyal and Jeltsch, 2004; Vilkaitis *et al.*, 2005). Obtaining complete transmission of methylation patterns can take hours, and a global lag between replication and complete maintenance of methylation has been reported (Charlton *et al.*, 2018; Ming *et al.*, 2020; Stewart-Morgan *et al.*, 2023).

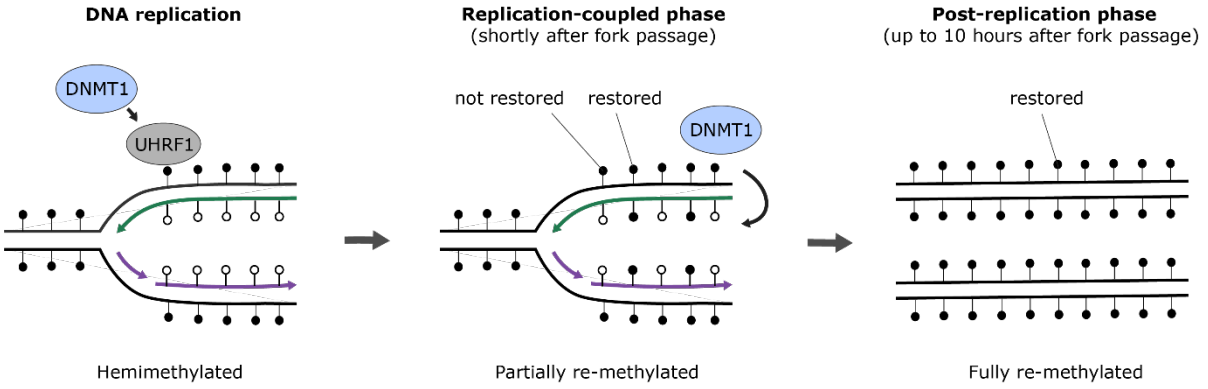


Figure 1-8: DNA methylation maintenance. *DNMT1 is recruited to newly synthesised, hemi-methylated DNA by UHRF1. DNMT1 then copies methylation patterns onto the unmethylated strand. There is an initial wave of re-methylation shortly after passage of the replication fork, but complete re-methylation can take up to 10 hours (Charlton* et al.*, 2018; Ming* et al.*, 2020).*

1.2.5 Cooperation between the DNA methyltransferases

Despite the canonical role for DNMT1 as the maintenance methyltransferase, and DNMT3A and DNMT3B as the *de novo* methyltransferases, this is an over-simplified view. There is evidence for cooperation between the enzymes during establishment and maintenance of DNA methylation. Although DNMT1 displays strong preferential binding towards hemi-

methylated CpG sites, it is capable of methylating previously unmethylated CpGs (Fatemi *et al.*, 2001; Wang *et al.*, 2020). Reintroduction of DNMT1 into mouse ESCs lacking activity of all three DNMTs results in *de novo* gain of methylation at transposable elements, suggesting a role for DNMT1 is establishing repression of these sequences (Haggerty *et al.*, 2021). The *de novo* activity of DNMT1 at transposable elements is dependent on UHRF1, potentially through UHRF1-mediated recruitment to H3K9me3-marked regions (Haggerty *et al.*, 2021).

The DNMT3 enzymes have been suggested to have a role in maintaining DNA methylation. In human colorectal cancer cells lacking ~90% of wild-type DNMT1 activity, global methylation levels are only reduced by ~20% and gene silencing is maintained (Rhee *et al.*, 2000). This finding indicates that there are alternative mechanisms of maintaining DNA methylation. Combined deletion of DNMT1 and DNMT3B was shown to reduce methylation levels by up to 95%, while deletion of DNMT3B alone had minimal impact on global methylation (Rhee *et al.*, 2002). In mouse ESCs, DNMT3A and DNMT3B are evidenced to cooperate with DNMT1 to maintain methylation at certain genomic regions, including endogenous repeat sequences (Liang *et al.*, 2002). DNMT3B has been evidenced to compensate for loss of DNMT1 in mouse intestinal epithelium by maintaining DNA methylation (Elliott, Sheaffer and Kaestner, 2016). Furthermore, combined knock-out of DNMT3A and DNMT3B in human ESCs results in progressive loss of DNA methylation over cell passages. The maintenance fidelity of DNMT1 in the absence of DNMT3A and DNMT3B was estimated to be ~99.7% (Liao *et al.*, 2015). The imprecise activity of DNMT1 has been suggested to facilitate epigenetic plasticity (Wang *et al.*, 2020). Together, these findings suggest that the DNMT3 enzymes may cooperate with DNMT1 to propagate DNA methylation patterns.

1.2.6 DNA demethylation

DNA demethylation is the process by which the DNA methylation mark is removed from the DNA, generating an unmethylated molecule. DNA demethylation can be an active or passive process (Moore, Le and Fan, 2013). Passive demethylation occurs when DNA methylation patterns are not faithfully maintained. During DNA replication, if DNMT1 fails to propagate methylation onto the newly synthesised stand of DNA, this results in the formation of a hemi-methylated molecule. Over consecutive rounds of cell division, the methyl signature is diluted

out of the cell population and the DNA becomes unmethylated (Kagiwada *et al.*, 2013). Passive DNA demethylation can be achieved pharmacologically by inhibiting DNMT1 activity, for example by 5-azacytidine treatment (Neidhart, 2016).

Active demethylation involves the enzymatic modification and removal of the methyl group from 5mC. Demethylation is catalysed by the Ten-Eleven Translocation (TET) family of enzymes, which include TET1, TET2 and TET3 (Wu and Zhang, 2017). The TET proteins mediate demethylation of DNA through a sequential oxidation reaction. In the presence of oxygen, TET proteins catalyse oxidative decarbonation of α-ketoglutarate to generate Fe(IV)-oxo, an iron-oxygen intermediate that converts 5mC to 5-hydroxy-methylcytosine (5hmC) (Kohli and Zhang, 2013). 5hmC is a stable epigenetic mark that is found at actively transcribed genes (Pfeifer, 2023). Subsequent oxidation products are 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which are less chemically stable (Kohli and Zhang, 2013; Wu and Zhang, 2017). 5fC and 5caC are recognised by the base excision repair DNA repair pathway and converted into unmethylated cytosines (He *et al.*, 2011; Maiti and Drohat, 2011).

1.3 Aberrant DNA methylation in cancer

Cancer is a complex and diverse group of diseases characterized by the uncontrolled growth and spread of abnormal cells. In healthy individuals, cell division and growth are carefully regulated by multiple cellular pathways, to ensure that normal tissue architecture and function is maintained. A cancer develops when these regulatory mechanisms malfunction, allowing cells to proliferate uncontrollably (Hanahan and Weinberg, 2000). Uncontrolled cell proliferation can result in the formation of tumours, which may be benign (non-cancerous) or malignant (cancerous). Malignant tumours can acquire the ability to invade nearby tissues through a process called metastasis. In order to continue growing, tumours develop a growth-promoting microenvironment and acquire the ability to evade immune detection (Weinberg and Weinberg, 2013).

Cancer is a genetic disease, and it arises through the progressive accumulation of mutations that deregulate gene expression and promote abnormal activity of proteins. These mutations enable the cell to sustain proliferative signalling, evade growth suppressors and resist cell

death signalling (Hanahan and Weinberg, 2000, 2011). There are two main categories of genes that are involved in cancer: oncogenes ad tumour-suppressor genes. Proto-oncogenes are genes that function to promote normal cell growth and division (Dakal *et al.*, 2024). Mutations that cause the over-expression or aberrant activity of proto-oncogenes can result in deregulated cellular activity. The mutated proto-oncogene is referred to as an oncogene, meaning that they drive cancer development. Tumour-suppressor genes normally function to regulate the cell cycle, repair damaged DNA or promote apoptosis. Mutations that inactivate tumour suppressor genes enable uncontrolled cell proliferation and promote genetic instability (Dakal *et al.*, 2024). Epigenetic alterations are frequently observed in cancer cells, which could lead to deregulated gene expression and promote carcinogenesis (Esteller, 2006; Jones and Baylin, 2007). DNA methylation patterns are altered the vast majority of human cancer cells, such that aberrant DNA methylation is considered a hallmark of cancer. The primary alteration to DNA methylation patterns that is observed in human cancer is loss of DNA methylation across large regions of the genome (Jones and Baylin, 2002, 2007; Feinberg and Tycko, 2004; Esteller, 2006).

1.3.1 Hypomethylation at Partially Methylated Domains (PMDs)

Loss of DNA methylation is a hallmark of cancer epigenomes. Early studies of DNA methylation levels in cancer identified genome-wide reduced methylation across multiple cancer types, compared to analogous healthy tissue (Feinberg and Vogelstein, 1983; Gama-Sosa *et al.*, 1983). With advancements in whole-genome bisulphite sequencing (WGBS) technology, researchers were able to obtain a more detailed view of methylation patterns in cancer cells. This work revealed that hypomethylation occurs across large, defined regions of the genome, while other regions maintain high methylation levels (Lister *et al.*, 2009; Hansen *et al.*, 2011; Berman *et al.*, 2012). An example of the DNA methylation landscape in colorectal cancer cells is shown in Figure 1.7. These hypomethylated regions are known as Partially Methylated Domains (PMDs). The size of a PMD can range from ~100kb-20Mb (mean size ~150kb) and they cover approximately 30-50% of the genome (Salhab *et al.*, 2018). PMDs have since been identified as a universal defining feature of most human solid tumours (Timp *et al.*, 2014). Importantly, although these regions display broadly decreased methylation levels, the

decrease is not a consistent reduction at all CpGs. Instead, methylation levels at individual CpGs within PMDs are highly variable and range from 0-100% (Gaidatzis *et al.*, 2014)

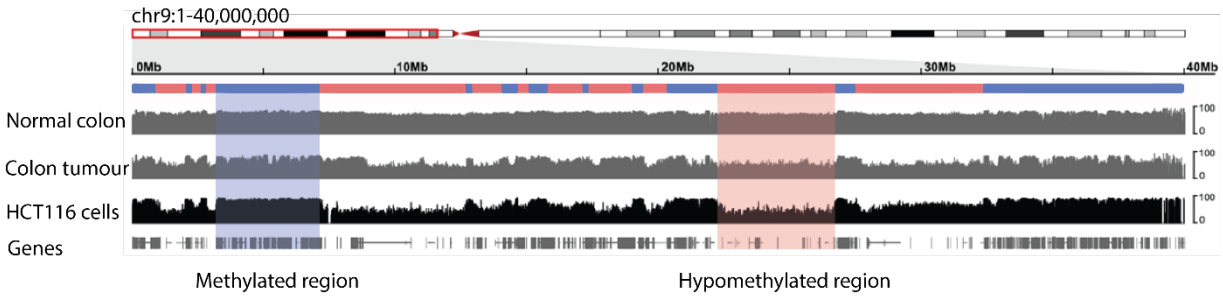


Figure 1-9: Hypomethylation in cancer occurs at Partially Methylated Domains (PMDs). *Genome browser tracks (UCSC) showing whole-genome bisulfite sequencing data (Sproul, unpublished) from normal colon tissue, colon tumour, and HCT116 cells, a colorectal cancer derived cell lines. The percentage of methylated CpGs are indicated on the right. Regions that maintain methylation (blue) and those that lose methylation (red) are shown. Example data shown for the short arm of chromosome 9.*

The positions of PMDs are highly conserved across different cell types. Hypomethylation in cancer generally coincides with H3K9me3 and H3K27me3 marked heterochromatic regions. Conversely, the regions that retain high methylation levels are enriched for H3K36me3, which is associated with transcribed gene bodies (Hon *et al.*, 2012; Zhou *et al.*, 2018; Zheng *et al.*, 2023). PMDs have also been shown to overlap with lamina associated domains (LADs), which are repressive chromatin domains (Berman *et al.*, 2012). Furthermore, PMDs coincide with late-replicating genomic regions (Aran *et al.*, 2011; Berman *et al.*, 2012; Salhab *et al.*, 2018; Zhou *et al.*, 2018; Decato *et al.*, 2020). PMDs display low gene density and have lower CpG density, compared to regions that retain methylation in cancer (Gaidatzis *et al.*, 2014; Zhou *et al.*, 2018; Decato *et al.*, 2020). Interestingly, PMDs have also been identified in normal dividing tissues, although a far greater degree of hypomethylation is observed in cancer tissue and immortalised cell lines (Zhou *et al.*, 2018)

The molecular mechanism underlying PMD formation remains unclear. Analogous normal tissues display some degree of hypomethylation at PMDs, but retain higher methylation levels than cancer tissues. Immortalised cell lines display more pronounced hypomethylation at PMDs than cancer tissue, thus it has been suggested that hypomethylation may be linked to

proliferation (Salhab *et al.*, 2018; Zhou *et al.*, 2018; Decato *et al.*, 2020). Zhou *et al*. (2018) showed that cell samples from younger individuals displayed higher methylation levels than samples from older individuals, linking hypomethylation to ageing. By comparing hypomethylation to somatic mutation densities, Zhou *et al*. (2018) also demonstrated a link between PMD methylation loss and cumulative cell divisions. These findings suggest that DNA hypomethylation may be a progressive process and that PMDs may form as a result of improper methylation maintenance during cell division. Interestingly, the most severely hypomethylated regions in cancer display the latest replication timings (Zhou *et al.*, 2018; Decato *et al.*, 2020). A genome-wide lag between DNA replication and re-methylation of newly synthesised DNA has been reported (Charlton *et al.*, 2018). The increased proliferation rates in cancer could prevent maintenance of methylation at late replicating regions, leading to progressive loss of DNA methylation at these regions over time (Endicott *et al.*, 2022). However, the ability of cancer cells to maintain methylation at PMDs has not been directly tested.

The lower density of CpG sites within PMDs could contribute to hypomethylation (Gaidatzis *et al.*, 2014; Zhou *et al.*, 2018). Local CpG density has been shown to influence levels of methylation at neighbouring CpGs sites, and there is evidence for collaboration between local CpG sites which may influence methylation maintenance (Gaidatzis *et al.*, 2014; Wang *et al.*, 2020). The maintenance methyltransferase, DNMT1, acts processivity on the DNA and has been reported to be more efficient in the presence of multiple CpG sites (Hermann, Goyal and Jeltsch, 2004). Therefore, the low CpG content within PMDs could contribute to a breakdown in collaboratively, leading to loss of methylation over time, but again this has not been directly tested. Furthermore, the DNA sequences flanking CpGs have been shown to influence DNMT1 activity. Computational modelling and experimental evidence show that solo-CpG sites, those flanked by an A or a T, are more prone to methylation loss than those flanked by a C or a G (Gaidatzis *et al.*, 2014; Zhou *et al.*, 2018). DNMT1 displays preference for CpGs flanked by a cytosine at the -1 position (Adam *et al.*, 2020). Solo-CpGs are enriched in PMDs, which could contribute to incomplete maintenance of methylation at PMDs (Zhou *et al.*, 2018).

These models for passive loss of methylation at PMDs through insufficient maintenance activity suggest that there may be malfunction of DNMT1, UHRF1 or other components of the methylation maintenance machinery in cancer cells. However, misexpression of DNMTs and

UHRF1 has not been liked to hypomethylation (Zhou *et al.*, 2018). Active demethylation may also play a role. Dysfunction of the TET enzymes could contribute to loss of methylation at PMDs, although deregulation of the TET family proteins has not been associated with hypomethylation (Zhou *et al.*, 2018). Finally, the *de novo* methyltransferases may contribute to maintaining methylation at the regions which are protected from hypomethylation in cancer. DNMT3B is directed towards H3K36me3 marked regions, and may explain the higher methylation at H3K36me3 regions observed in cancer (Baubec *et al.*, 2015; Rondelet *et al.*, 2016; Zhou *et al.*, 2018; Masalmeh *et al.*, 2021). The *de novo* activity of DNMT3A has also been suggested to play a role in maintaining DNA methylation at PMDs in the absence of DNMT1 activity (Kafetzopoulos, 2022).

Hypomethylation at PMDs is an early event in tumour development that precedes malignancy and has been associated with future risk of malignant transformation (Goelz *et al.*, 1985; Teschendorff *et al.*, 2012; Timp *et al.*, 2014). DNA hypomethylation has been evidenced to promote genome instability, which may contribute to cancer progression (Sheaffer, Elliott and Kaestner, 2016; Besselink *et al.*, 2023). Human PMDs display a high level of instability at the genetic, epigenetic and transcriptome levels, which may contribute to tumour cell heterogeneity (Brinkman *et al.*, 2019). Mouse models with a hypomorphic mutation in DNMT1 that reduced DNMT1 expression by ~90% displayed genome-wide hypomethylation in all tissues, high levels of chromosomal instability and developed T-cell lymphomas, linking hypomethylation with carcinogenesis (Gaudet *et al.*, 2003). Thus, PMDs have been suggested to act as a potential driving force underlying cancer progression. However, the precise functional consequences of this hypomethylation are yet to be determined.

1.3.2 Hypermethylation of CpG Islands

Alongside hypomethylation at PMDs, hypermethylation of certain CGIs associated with gene promoters has been identified in wide variety of cancers (Baylin and Herman, 2000; Ehrlich, 2002; Berman *et al.*, 2011). Aberrant hypermethylation tends to occur in a more tumour-type specific manner than hypomethylation at PMDs, although some patterns of hypermethylation are shared between tumour types (Costello *et al.*, 2000). Hypermethylation of CGIs associated

with tumour-suppressor genes has been suggested to result in aberrant silencing of these genes, which may promote carcinogenesis (Geissler *et al.*, 2024).

Hypermethylation was first identified at the retinoblastoma tumour-suppressor gene (Greger *et al.*, 1989), and was later observed at a subset of CGIs in colon, breast, glioma and acute myeloid leukaemia (AML), among others (Costello *et al.*, 2000). A subset of colorectal tumours display hypermethylation at certain CGIs, and these are referred to as CpG Island Methylator Phenotype (CIMP) positive tumours. Seven CGIs that were only methylated in cancer cells were identified, and CIMP+ tumours were shown to have hypermethylation at the promoter of MLH1, a DNA mismatch repair gene (Toyota *et al.*, 1999). Hypermethylation of the MLH1 CGI was associated with loss of MLH1 protein expression, which was restored upon treatment with 5-aza-2'-deoxycytidine, a hypomethylating drug (Herman *et al.*, 1998). Furthermore, re-expression of MLH1 restored mismatch repair function in colorectal cancer cell lines, directly linking CGI hypermethylation to silencing of the MLH1 tumour-suppressor (Herman *et al.*, 1998). Another example of a tumour-suppressor gene with a CGI promoter that is hypermethylated in cancer is CDKN2A, which has a role in cell cycle regulation and senescence. Hypermethylation of the CDKN2A promoter is frequently observed in a variety of human cancers, including AML, MDS (Drexler, 1998), lung (Belinsky *et al.*, 1998), breast (Sinha *et al.*, 2008) and colon cancers (Shima *et al.*, 2011). Hypermethylation of CDKN2A is an early event in tumour development, and is linked to CDKN2A silencing which is thought to help enable the cancers escape from growth regulation and apoptotic signalling (Belinsky *et al.*, 1998; Zhao *et al.*, 2016). The related tumour-suppressor, CDKN2B, is also frequently hypermethylated and down-regulated in cancer, particularly AML and MDS (Drexler, 1998; Arya *et al.*, 2017; Jang *et al.*, 2019). As a final example, BRCA1, a breast-cancer susceptibility gene, is a well-studied example of a tumour-suppressor gene that frequently acquires hypermethylation in cancer. Hypermethylation and inactivation of BRCA1 is associated with cancer progression (Esteller *et al.*, 2000; Glodzik *et al.*, 2020).

However, while some CGIs are prone to hypermethylation in cancer, the role that this hypermethylation plays in aberrant gene repression is disputed. The majority of promoters that acquire aberrant hypermethylation in cancer are associated with genes that are already silenced in the corresponding normal tissue (Sproul *et al.*, 2011, 2012). These genes are silenced in normal tissue through repressive chromatin marks and the Polycomb Group (PcG)

repressive complex. The genes with promoter hypermethylation in cancer often exist in a bivalent state in normal tissue (Mikkelsen *et al.*, 2007; Easwaran *et al.*, 2012). Bivalent chromatin is simultaneously marked with repressive histone marks, PcG marks and active transcription histone marks. This state has been proposed to maintain genes in a transcriptionally 'off' state, but poised for transcription (Mikkelsen *et al.*, 2007; Macrae, Fothergill-Robinson and Ramalho-Santos, 2023). Thus, it has been suggested that aberrant hypermethylation in cancer may reinforce the silent state, preventing reactivation of the gene. Notably, aberrant hypermethylation occurs in a cell lineage-specific pattern (Sproul *et al.*, 2011). DNA hypermethylation may, therefore, not contribute directly to cancer development through aberrant gene silencing but instead act as a marker of cell lineage.

1.4 DNA methylation as a therapeutic target

1.4.1 Hypomethylating agents in cancer treatment

In recent years, there has been a growing interest surrounding targeting epigenetic mechanisms in cancer treatment, particularly in combination therapies. As discussed in the previous section, aberrant DNA methylation is common to the vast majority of human cancers and has been linked to poor prognosis in patients (Jones and Baylin, 2002). The dynamic and reversible nature of DNA methylation makes it a promising target for therapeutic intervention. Hypomethylating agents (HMAs) are compounds that induce global hypomethylation by inhibiting DNA methylation and represent a promising class of anti-cancer drugs that have been approved for clinical use (Stomper *et al.*, 2021).

The first generation of HMAs, 5-azacytidine (5-aza) and 5-aza-2'-deoxycytidine (5-aza-dC), are cytidine analogues (Futscher, 2011). These compounds were synthesised in the 1960s, and when tested against cancer cells were shown to have a range of anti-metabolic activities (Pískala and Šorm, 1964). It was later shown that these cytidine analogues are incorporated into DNA (Čihák, Veselý and Šorm, 1965). Once incorporated into the DNA, they bind to DNMTs during methyl group transfer and form a covalent complex. Normally, following methyl transfer, the DNMT is released from this complex by a beta(β)-elimination reaction

with the C5 proton. However, 5-aza and 5-aza-dC lack a proton at position C5, preventing β-elimination and thus trapping the methyltransferase on the DNA. This subsequently depletes the free enzyme pool, impeding methylation maintenance and resulting in passive hypomethylation (Santi, Garrett and Barr, 1983; Jüttermann, Li and Jaenisch, 1994; Schermelleh *et al.*, 2005). The trapped enzyme-DNA complex inhibits DNA synthesis, forming adducts that are mutagenic if not repaired and causing considerable cytotoxicity (Jüttermann, Li and Jaenisch, 1994; Christman, 2002). Cytidine analogues have been widely used as HMAs, both in research and in the clinic. They have been shown to be effective in the treatment of haematologic malignancies, and are currently approved for the treatment of acute myeloid leukaemia (AML) and myelodysplastic syndromes (MDS) (Stomper *et al.*, 2021). However, these compounds are irreversibility incorporated into DNA, resulting in dose-limiting toxicities. Furthermore, 5-aza and 5-aza-dC exhibit poor chemical stability and pharmacokinetics, making drug delivery to solid tumours challenging (Stresemann and Lyko, 2008). These properties have limited the use of cytidine analogues as chemotherapeutic agents for solid tumours, despite promising pre-clinical evidence, and highlight the need for improved HMAs.

More recently, a new class of non-nucleoside methylation inhibitors have been identified. These compounds, GSK-3484862 and GSK-3685032, are small-molecule inhibitors that selectively inhibit DNMT1 and induce rapid hypomethylation (Pappalardi *et al.*, 2021). Co-crystallisation studies showed that these inhibitors compete with the active site loop of DNMT1 for penetration into hemi-methylated CpG dinucleotides. Pappalardi *et al*. (2021) reported that these compounds were highly selective towards DNMT1 and, unlike cytidine analogues, did not bind non-specifically to the *de novo* DNMTs. GSK-3685032 interacts with DNMT1 primarily through hydrogen bonding, and with DNA through pi-stacking and van der Waals interactions. The lack of covalent interactions meant that selective DNMT1 inhibition was less cytotoxic than cytidine analogues, and reversible. In addition, GSK-3685032 displayed improved efficacy *in vivo* compared to 5-aza-dC and elicited similar cellular responses, indicating that these compounds may be preferable for clinical use, over cytosine analogues.

1.4.2 Hypomethylating agents induce immune activation pathways

The precise mechanism through which HMAs exert their clinical benefit remains unclear. Historically, abnormal DNA methylation in cancer has been described as leading to aberrant silencing of tumour-suppressor genes through promoter hypermethylation (Jones and Baylin, 2002). As such, HMAs were initially thought to help re-express aberrantly silenced genes. However, as demonstrated by Sproul *et al*. (Sproul *et al.*, 2011, 2012), very few genes are aberrantly silenced by DNA methylation in cancer. In fact, most genes that gain aberrant hypermethylation are already silenced in the corresponding pre-neoplastic tissue. HMAs have been reported to reverse aberrant DNA hypermethylation (Jones and Baylin, 2002), but whether re-expression of silenced tumour suppressor genes is the clinical mode of action for HMAs is ambiguous. Hypermethylation resulting in inactivation of cyclin-dependant kinase inhibitor 2B (CDKN2B), a proposed tumour suppressor gene, is frequently reported in AML and MDS patients (Drexler, 1998; Jang *et al.*, 2019). In AML cell lines treated with 5-aza-dC, re-expression of CDKN2B has been reported (Bender, Pao and Jones, 1998; Lakshmikuttyamma *et al.*, 2010). Pre-treatment hypermethylation of the CDKN2B promoter has been associated with poor clinical outcomes in AML and MDS patients (Raj *et al.*, 2007; Kamali Dolatabadi *et al.*, 2017). However, demethylation and re-expression of CDKN2B does not appear to correlate with patient responses to 5-aza treatment (Bender, Pao and Jones, 1998; Raj *et al.*, 2007; Fandy *et al.*, 2009; Kamali Dolatabadi *et al.*, 2017). This implies that CDKN2B re-expression is not necessary for positive clinical responses to HMAs. However, reduced global methylation has been significantly correlated with positive clinical responses to HMAs, suggesting that demethylation may promote alternative anti-tumour effects (Shen *et al.*, 2010).

Upregulation of the germline Cancer Testis Antigen (CTA) family genes has been widely reported in response to 5-aza/5-aza-dC (Güre *et al.*, 2002; James, Link and Karpf, 2006; Weon and Potts, 2015; Weber, Jung and Doerfler, 2016). The CTA genes are known to be regulated by promoter methylation in somatic cells, and are derepressed following hypomethylation by HMAs (De Smet *et al.*, 1999; Weber, Jung and Doerfler, 2016). CTA genes are a family of tumour-associated antigens (TTAs) which may be able to induce tumour immunogenicity via T-cell activation (Loo Yau *et al.*, 2021). Increased CTA-associated antigen expression was

reported to induce a tumour-specific cytotoxic $CD8^+$ T-cell response in patients treated with 5-aza and a histone deacetylase (HDAC) inhibitor (Goodyear *et al.*, 2010). This TTA-specific T-cell response resulted in major clinical response in 8/11 of patients, with 4 patients achieving complete remission (Goodyear *et al.*, 2010). As such, there has been growing interest surrounding immune activation in response to HMAs.

Activation of genes implicated in interferon (IFN) signalling in response to 5-aza-dC was first reported in HT29 colon adenocarcinoma cells by Karpf et al. (1999). Activation of the signal transductors and activators of transcription (STAT) family of transcription factors, which promote transcription of genes involved in immune signalling, along with a range of interferon-stimulated genes (ISGs), was observed (Karpf *et al.*, 1999). HT29 cells treated with 5-aza-dC were sensitised to the growth-inhibitory effects of IFN-alpha, which is used as an immune-stimulating drug in cancer treatment (Karpf *et al.*, 1999). This finding suggested that immune activation by HMAs could be exploited for therapeutic benefit. Further characterisation of the pathways upregulated by 5-aza revealed that interferon signalling, cytokine signalling, antigen presentation and inflammation were enriched following 5-aza treatment in multiple cancer cell lines (Wrangle *et al.*, 2013; Li *et al.*, 2014). Specifically, a Type-I interferon (IFN-I) response is triggered by 5-aza (Chiappinelli *et al.*, 2015). The IFN-I signalling pathway is an adaptive immune response which is activated upon detection of viral nucleic acid in the cytosol. IFN-I signalling leads to the activation of ISGs and the production of pro-inflammatory chemokines and cytokines, which promote recruitment of immune cells. IFN-I signalling also induces increased antigen presentation and apoptotic signalling (Platanias, 2005; Lukhele, Boukhaled and Brooks, 2019).

The immune response genes induced by 5-aza treatment are not repressed by promoter DNA methylation prior to treatment, suggesting that promoter hypomethylation is not the mechanism underlying immune activation (Li *et al.*, 2014; Roulois *et al.*, 2015). However, immune activation by 5-aza is dependent on the activity of cytosolic sensors of viral RNA, RIG1 and MDA5 (also known as IFIH1) (Chiappinelli *et al.*, 2015; Roulois *et al.*, 2015). Upon detection of viral RNA, RIG1 and MDA5 form homo-oligomers and interact with MAVS, which is required for down-stream IFN-I signalling (Wu and Hur, 2015). Both Chiappinelli *et al*. (2015) and Roulois *et al*. (2015) demonstrated that knock-down of MAVS decreased 5-aza/5-aza-dC-induced immune signalling, implicating detection of dsRNA in this process. Therefore,

increased expression of endogenous retroviruses in response to hypomethylation was investigated as a potential mechanism of immune activation. Upregulation of Human Endogenous Retrovirus 1 (HERV1) family elements has been reported in response to HMAs in multiple cancer cell lines (Chiappinelli *et al.*, 2015; Y. Kong *et al.*, 2019; Ohtani *et al.*, 2020; Pappalardi *et al.*, 2021). HERV elements represent ~8% of the human genome and are normally highly methylated in cells to repress their activity (Weber, Jung and Doerfler, 2016; Alcazer, Bonaventura and Depil, 2020). When transcribed, HERV elements are hypothesised to form dsRNA structures which could be detected by viral sensing mechanisms (Alcazer, Bonaventura and Depil, 2020; Sadeq *et al.*, 2021). Furthermore, HERV elements are evidenced to form peptides that are presented by MHC molecules and trigger T- and B-cell activation (Smith *et al.*, 2018; Y. Kong *et al.*, 2019). Over-expression of HERV elements leads to activation of the same IFN genes induced by 5-aza, further implicating HERV expression in this process (Chiappinelli *et al.*, 2015; Ohtani *et al.*, 2020). Thus, it has been proposed that hypomethylation of HERV elements leads to their transcription and activates immune signalling, inducing a 'viral mimicry' state in cancer cells which may help improve immune infiltration into the tumour microenvironment (Roulois *et al.*, 2015).

This evidence collectively suggests that while hypomethylation can, in some cases, lead to the re-expression of aberrantly silenced tumour-suppressor genes, this may not be the primary mechanism through which HMAs confer clinical benefits. Rather than 'reprogramming' tumour cells back to a normal state, HMAs remove the remaining DNA methylation which cancer cells may rely on to continue proliferating (Kulaeva *et al.*, 2003). The resulting hypomethylation leads to activation of immune signalling and increases immunogenicity. Activation of immune signalling by HMAs is closely associated with positive clinical responses to 5-aza (Roulois *et al.*, 2015; Ohtani *et al.*, 2020). Therefore, exploiting immune activation by HMAs may be a beneficial therapeutic avenue for cancer treatment.

1.4.3 Towards hypomethylating agent combination strategies

Hypomethylating agents are increasingly being explored in combination treatment strategies as a method to sensitise patients to immunotherapeutic agents. Initial clinical studies have

demonstrated that HMAs, in combination with programmed cell death ligand 1 (PD-L1) inhibitors, an immune checkpoint blockade approach, can potentiate immune responses in AML patients (Apostolova *et al.*, 2023). A number of trials are ongoing in both MDS and AML patients, and have shown positive results (Daver *et al.*, 2018). Epigenetic priming with 5-aza-dC has also been shown to augment the patient responses to PD-I inhibition in ovarian cancer (Chen *et al.*, 2022; Chiappinelli and Baylin, 2022). Pre-clinical studies have indicated that HMAs have been effective in enhancing responses to immune checkpoint blockade in multiple cancer types, including colorectal cancer (Villanueva, Álvarez-Errico and Esteller, 2020). Combined inhibition of DNMT1 with GSK3484862 with EZH2 inhibitors has been evidence to enhance the viral mimicry response to hypomethylation in colorectal cancer cells (Chomiak *et al.*, 2024). This suggests that the HMAs may have therapeutic applications outside of haematologic malignancies, and they are currently being trialled for the treatment of solid tumours (Linnekamp *et al.*, 2017; 2021).

1.4.4 Future perspectives for hypomethylating agents

A remaining challenge in delivery of HMAs is obtaining sufficient hypomethylation *in vivo*, partly due to the dose-limiting toxicities of 5-aza/5-aza-dC, and emerging resistance (Zeidan, Kharfan-Dabaja and Komrokji, 2014). The new generation of non-nucleoside DNMT1 inhibitors may help overcome this challenge, although this awaits to be seen. Alternatively, combination therapies targeting multiple epigenetic modulators may be a more effective method to induce hypomethylation in patients. Depletion of UHRF1 in HCT116 degron cells has been shown to induce greater loss of DNA methylation than targeting DNMT1. Combined depletion of DNMT1 and UHRF1 further reduces methylation levels (Yamaguchi *et al.*, 2023). Importantly, inhibition of UHRF1 has been shown to induce expression of HERV elements and immune activation *in vitro* (Irwin *et al.*, 2023). UHRF1 is overexpressed in a variety of cancers, including colorectal, breast and lung, and has been associated with poor prognosis, making it an attractive therapeutic target (Jin *et al.*, 2010; Ashraf *et al.*, 2017; X. Kong *et al.*, 2019). However, there are currently no clinically approved inhibitors of UHRF1, and inhibiting UHRF1 activity has proved challenging (Liu *et al.*, 2022).

Alternative compounds that are able to induce similar immunomodulatory effects as 5-aza and 5-aza-dC are currently being pursued. An example is RRx-001, a first-in-class compound recently identified in a phenotypic drug screen (Oronsky *et al.*, 2016, 2021). Phenotypic drug screening is a drug discovery approach that involves evaluating the effect of compounds on a disease phenotype or biomarker, without prior knowledge of the specific molecular targets. This method allows for the identification of drug candidates based on their overall biological effects, potentially uncovering new therapeutic mechanisms or targets (Vincent *et al.*, 2022). RRx-001 has been shown to decrease DNA methylation levels and upregulate HERV expression and IFN signalling in HCT116 cells (Zhao *et al.*, 2017). This novel compound is currently in phase II clinical trials for multiple diseases, including colorectal cancer, and has displayed vastly improved tolerability compared to 5-aza and 5-aza-dC (Morgensztern *et al.*, 2019; Reid *et al.*, 2023).

The immune-stimulating effects of DNA hypomethylation make it attractive as a therapeutic avenue. However, emerging resistance has limited patient responses. Some cells retain high levels of methylation following HMA treatment, which has been suggested to facilitate survival of cancer cells (Bond *et al.*, 2024). This justifies the ongoing efforts to develop new HMAs with enhanced therapeutic potential.

1.5 HCT116 colorectal cancer cell model

Colorectal cancer (CRC) is a form of adenocarcinoma that originates in the bowel and is the second most fatal cancer worldwide (Hossain *et al.*, 2022). HCT116 cells are a CRC derived cell line which have been used extensively as an *in vitro* cancer model. These cells display a relatively stable, near-diploid karyotype, and high transfection efficiency, making them convenient for experimental use (Lengauer, Kinzler and Vogelstein, 1997). Furthermore, HCT116 cells are a well-established model system for studying DNA methylation due to the established contribution of DNA methylation to the disease. The CpG Island Methylator Phenotype (CIMP) was first identified in CRC, which prompted widespread study of aberrant DNA methylation in HCT116 cells (Toyota *et al.*, 1999, p. 200; Weisenberger *et al.*, 2006).

In this study, I utilise HCT116 cells and their derivatives as an *in vitro* cancer model. The DNMT1 knock-out (henceforth, DNMT1^{KO}) cells used in this study were generated by Rhee *et al*. (2000). Homologous recombination was used to replace Dnmt1 exons 3, 4 and 5 with a promoter-less Hygromycin resistance gene. This disruption resulted in a 20% reduction in global 5mC content. The high retention of DNA methylation following knock-out of DNMT1 implied that the activity of other proteins may contribute to methylation maintenance. Therefore, in a follow-up study, Rhee *et al*. (2002) generated a HCT116 DNMT3B knock-out (DNMT3B^{KO}) cell line which is also utilised in this study. Homologous recombination was used to replace exons 2 to 21, resulting in a deletion that removed the catalytic active site. DNMT3B^{KO} cells retained >97% of wild-type 5mC content, suggesting that loss of DNMT3B does not have a substantial impact on DNA methylation maintenance. A double knock-out (DKO) of DNMT1 and DNMT3B was also generated (Rhee *et al.*, 2002). DKO cells display 57.5-95% reduction in m^5C content, indicating that DNMT1 and DNMT3B may cooperate to maintain DNA methylation levels in human cells. I also utilise HCT116 degron cell lines in this study. Recently, Yamaguchi *et al*. (2024) developed degron-tagged alleles of DNMT1 and UHRF1 in HCT116 cells using Auxin-Inducible Degron (AID) technology. Degron alleles allow rapid depletion of target proteins and are useful tools to interrogate protein function.

Our group and others have performed whole-genome bisulfite sequencing in HCT116 cells, and their DNMT^{KO} derivatives (Masalmeh et al., 2021; Taglini et al., 2024). As a result, the DNA methylation landscape, including the locations and degree of hypomethylation in PMDs, has

been mapped in these cells. Furthermore, our group has previously mapped histone modifications and replication timing across the genome in HCT116 cells and the DNMT^{KO} derivatives (Masalmeh et al., 2021; Kafetzopoulos, 2022; Taglini et al., 2024). Thus, HCT116 cells represent ideal model system for this study due to the well-characterized epigenetic landscape.

1.6 Thesis objectives

Widespread loss of DNA methylation is a hallmark of human cancer cells. Hypomethylation in cancer may develop due to incomplete maintenance of DNA methylation. Yet, this model for hypomethylation in cancer is based on evidence derived from static snapshots of cancer methylomes. Less is known about the dynamics of DNA methylation maintenance in cancer cells. Nonetheless, it is clear that DNA methylation represents an important therapeutic target in cancer treatment. Drugs that are able to induce further hypomethylation can trigger immune activation in cancer cells, potentially increasing tumour immunogenicity and boosting the efficacy of immunotherapy in patients. However, existing hypomethylating compounds have limited clinical applications due to cytotoxicity and poor pharmacokinetics.

In order to gain insight into the dynamics of DNA methylation maintenance in colorectal cancer cells, I aimed to develop an experimental tool to measure DNA methylation maintenance *in vitro*. Therefore, in the first stage of this study I establish a methylation-sensitive eGFP (eGFPme) reporter system. I demonstrate that this methylated reporter can be efficiently integrated in host cells and that expression of the reporter gene corelates to underlying DNA methylation levels. I utilise this reporter system to track DNA methylation maintenance fidelity in human cancer cells. Next, I aimed to determine which proteins are the primary contributors to DNA methylation maintenance in cancer cells by assessing eGFPme reporter activity in cancer cell lines lacking activity of the DNMTs and UHRF1. Finally, I tested reporter activity in response to pharmacological inhibition of DNMT1, to determine whether the methylation-sensitive reporter system could be utilised as a phenotypic drug screening tool to screen novel hypomethylating compounds with therapeutic potential.

Key objectives:

- Develop a methylation-sensitive reporter system to measure the fidelity of DNA methylation maintenance *in vitro*.
- Assess the ability of human cancer cells to maintain DNA methylation patterns.
- Interrogate which proteins are required to maintain DNA methylation in cancer cells.
- Utilise the methylation-sensitive reporter as a phenotypic drug screening tool, to screen novel hypomethylating compounds *in vitro*.

2 Methods

2.1 Mammalian cell culture

2.1.1 Cell lines

HCT116 wild-type, DNMT1 knock-out ($DNMT1^{KO}$), DNMT3B knock-out ($DNMT3B^{KO}$) and double knock-out (DKO) were provided by Ioannis Kafetzopoulos, Sproul lab, IGC. These cells were originally obtained from Richard Meehan's research group, IGC. The knock-out cell lines were generated by Rhee *et al.* (2000;2002) and gifted by Bert Vogelstein (Rhee *et al.*, 2000, 2002). RKO cells were obtained from Francesca Taglini, Sproul lab, IGC.

HCT116 AID-tagged DNMT1, UHRF1 and DNMT1-UHRF1 cell lines were provided by Agnieszka Gambus, University of Birmingham. These cell were generated by Yamaguchi *et al.* (2024) and gifted by Pierre-Antoine Defossez (Yamaguchi *et al.*, 2024).

2.1.2 Culture conditions

All cells were maintained at 37°C under humidified air containing 5% CO_2. When passaging to new culture flasks, cells were washed with Phosphate Buffered Saline (PBS, IGC technical services) and detached using TrypLE Express Enzyme (Thermo Fisher, 12605010). Cells were passaged at a 1:3-1:5 ratio. Cell lines were tested routinely for mycoplasma contamination.

HCT116 WT, $DNMT1^{KO}$, $DNMT3B^{KO}$ and DKO cells were cultured in McCoy's 5A modified medium (Thermo Fisher, 16600082) supplemented with 10% Foetal Calf Serum (FCS, IGC technical services) and 1% Penicillin-Streptomycin (Pen/Strep, 140 µg/ml and 400 µg/ml, respectively, IGC technical services).

RKO cells were cultured in Dulbecco's Modified Eagles Medium (Sigma-Aldrich, D5796) supplemented with 10% FCS and 1% Pen/Strep.

HCT116-AID cells were cultured in McCoy's 5A media (Thermo Fisher, 16600082) supplemented with 10% FCS, 1% Pen/Strep, and 2nM L-glutamine (Thermo Fisher, 25030081).

2.1.3 Mammalian cell transfection

All cells were transfected using a forward transfection method. In 6-well plates 3-5x10^5 cells were seeded 24 hours prior to transfection, to be ~60% confluent at the time of transfection. To mediate piggyBac transposition, cells were co-transfected with 0.5µg transposase donor plasmid and 1.5µg piggyBac CAG_eGFP_IRES_Puro transposon donor plasmid (1:3 ratio). The piggyBac CAG_eGFP_IRES_Puro plasmid was originally cloned by Francesca Taglini, Sproul lab, IGC (Fig. 2.1). Transfection was performed using FuGene HD (Promega, E2311) and Opti-MEM (Thermo Fisher Scientific, 31985062). A DNA:FuGene ratio of 1:3 was used. Transfection mixture was added to wells in a drop-wise manner. Cells were cultured with the transfection mixture overnight at 37°C and medium was refreshed the following day. After 24 hours, transfection efficiency was assessed under the fluorescent microscope. After 48 hours, transfection efficacy could be quantified by flow cytometry.

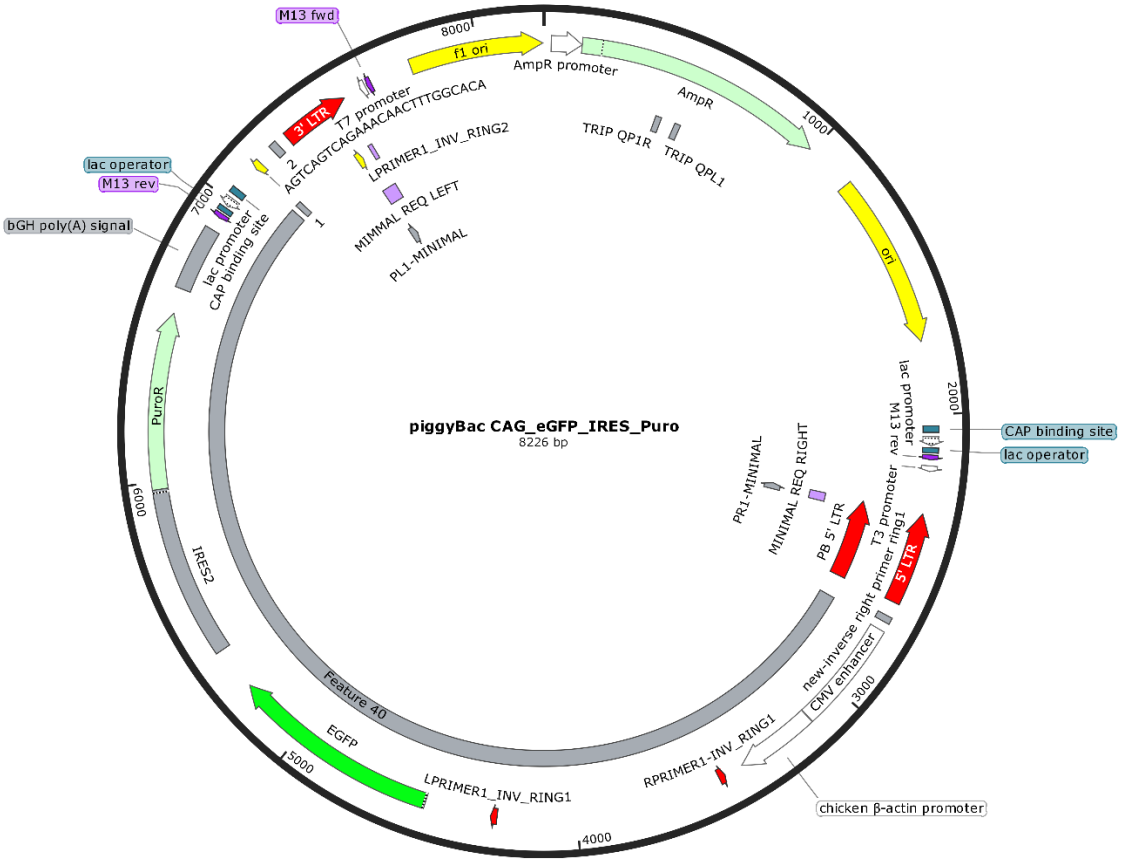


Figure 2-1: piggyBac CAG_eGFP_IRES_Puro plasmid map. *PiggyBac inverted terminal repeats (ITRs) flank the CAG synthetic promoter, enhanced green fluorescent protein (eGFP) reporter gene, internal ribosome entry site (IRES) and Puromycin resistance gene. Plasmid backbone encodes Ampicillin resistance. Plasmid map generated on SnapGene v5.1.7.*

2.1.4 Colony formation assay

HCT116 cells ($5x10^5$) were transfected with the PB_CAG_eGFP_IRES_Puro plasmid as described in 1.1.3. After 48 hours, cells were selected using 1µg/ml Puromycin antibiotic (Thermo Fisher, A1113803) and maintained under selection for 2 weeks. Medium was changed every 3 days. Surviving colonies were fixed using 4% paraformaldehyde and stained using 0.2% Methylene Blue at room temperature. Colonies were manually counted using a cell counter clicker. The percentage of colonies was calculated from the initial number of cells plated, as a proxy for vector integration efficiency.

2.1.5 Induction of AID-fused protein degradation

To induce expression of OsTIR1, HCT116-AID cells were seeded and incubated with medium containing 2.0µg/ml doxycycline (Dox, Apexbio, A4052, gifted by Christine Roger) for 24 hours. Subsequently, to activate protein degradation, medium was replaced with fresh medium containing 2.0µg/ml Dox and 500µM indole-3-acetic acid (IAA), a natural auxin. IAA was gifted by Emma Ramsay (Wood lab, IGC).

2.1.6 GSK-3484862 treatment

The DNMT1 inhibitor GSK-3484862 was gifted by Katie Pickup (Meehan lab, IGC). Cells ($5x10^5$) were seeded in 6-well plates and incubated with medium containing a final concentration of 0.5µM (unless otherwise stated) GSK-3484862 diluted in dimethylsulfoxide (DMSO, Sigma-Aldrich, D2650). Cell treated with only 0.5µM DMSO were maintained as vehicle controls. Media and inhibitor were replenished every 24 hours during the treatment time course.

2.1.7 NSC232005, NSC34716 and NSC43513 treatment

These compounds were gifted by Vassilios Myrianthopoulos, University of Athens. Cells ($5x10^5$) were seeded in 6-well plates and treated with a final concentration of 25µM drug

diluted in DMSO. Cell treated only with DMSO were maintained as vehicle controls. Media and inhibitor were replenished every 48 hours during the treatment time course.

2.2 Bacterial culture and manipulation

2.2.1 Bacterial transformation and plating

Subcloning Efficiency DH5α Competent Cells (Invitrogen, 18265017) were used for bacterial transformation. Cells were stored at -70^{o}C when not in use. Prior to use, cells were thawed on ice. For transformation, 50µl of cells were incubated with ~10ng of DNA for 30 minutes on ice. Cells were heat shocked at 42^{o}C for 30 seconds then allowed to rest on ice for 5 minutes. Subsequently, cells were incubated in 250µl SOC medium (Super Optimal broth with Catabolite repression, IGC technical services) for 1 hour at 37^{o}C in a shaking incubator (250RPM). Cells (250µl) were spread on Luria-Bertani (LB) plates supplemented with 1.5% agar and appropriate antibiotic. For ampicillin selection, LB plates containing 50mg/ml Ampicillin were used. For blue/white colony selection (pGEM-T Easy vector), LB plates with 50mg/ml Ampicillin and 20mg/ml X-Gal were used. Plates were incubated at 37^{o}C overnight

2.2.2 Bacterial liquid culture

Single colonies were selected using sterilised plastic tips and grown in 5ml LB broth containing 50mg/ml Ampicillin. Cultures were incubated overnight at 37^{o}C with 250RPM shaking. For larger cultures, the initial 5ml culture was incubated for ~6 hours then transferred into 200µl LB broth supplemented with Ampicillin (50mg/ml) and grown overnight at 37^{o}C with shaking.

2.2.3 Glycerol stocks

For long-term storage of plasmids, overnight liquid culture of DH5α transformed with plasmid were mixed in a 1:1 ratio with glycerol. Glycerol stocks were stored at -70^{o}C. Glycerol stocks were used for plasmid DNA isolation (described in 1.4.1). To grow bacteria from glycerol stock,

the stock was kept on ice without allowing it thaw and a small amount was streaked on an Ampicillin agar plate and grown up overnight. Single colonies were then selected and grown in liquid culture, as in 1.3.2.

2.3 DNA methods

2.3.1 Plasmid DNA isolation

Plasmid DNA was isolated from overnight liquid culture using either QIAprep Spin Miniprep Kit (QIAGEN, 27104) or EndoFree Plasmid Maxi Kit (QIAGEN, 12362) kit, according to the manufacture's protocol. When preparing plasmid DNA for mammalian cell transfection, the EndoFree Plasmid Maxi Kit was used. Briefly, liquid culture was harvested by centrifugation 6000 g for 15 minutes at 4°C and resuspended in buffer. Cells were lysed for 5 minutes, then the lysis reaction was neutralised. Supernatant was loaded onto a column matrix and DNA was allowed to bind the column by centrifugation. The column was washed twice, and DNA eluted in nuclease free water.

2.3.2 Phenol Chloroform isolation of genomic DNA

For Phenol Chloroform genomic DNA (gNDA) extraction, cells were detached and pelleted by centrifugation (1000 RMP, 5mins). Cell pellets were resuspended in 250µl PBS and combined with an equal volume 2XGLB to lyse cells. The solution was pipetted 10-20 times using a 21G needle syringe to facilitate cell lysis and fragment gDNA. Solution was then incubated with 20µl Proteinase K (20mg/ml, Roche) overnight at 55^{o}C with shaking (250 RMP). To isolate DNA, an equal volume (500µl, adjusted based on sample volume at each stage) buffered-phenol chloroform (BPC, pH8.0, Sigma-Aldrich, P0269) was mixed with the cell sample by shaking and centrifuged for 5 minutes. The upper layer of the phase-separated mixture was transferred to a new Eppendorf and the BPC treatment was repeated. A total of two BPC extractions were performed, followed by one extraction using pure chloroform. RNA was removed by incubation with RNAse A/T1 Cocktail (Ambion, AM2286) for 1 hour at 37^{o}C with 250 RPM shaking. Extraction was then repeated once with BPC and once with pure

chloroform. DNA was precipitated using 1 volume isopropanol and 1/10th volume NaAc. Samples were incubated for 5 minutes than centrifuged at 13,000 RPM at room temperature for 15 minutes. The DNA pellet was washed with 70% EtOH and air dried for 5 minutes. The pellet was resuspended in nuclease free water (20-50µl, depending on expected yield).

2.3.3 Kit isolation of genomic DNA

Genomic DNA was isolated using the Monarch Genomic DNA Purification Kit (NEB, T3010) according to the manufacturer's protocol. Briefly, frozen cell pellets (~1x10^6 cells) were thawed on ice and resuspended in 100µl chilled PBS. Proteinase K and RNAse A were added to the resuspended pellet along with lysis buffer and samples were incubated at 56^{o}C with shaking (1400 RPM) for 5 minutes. The samples were then combined with binding buffer and applied to the purification column. Samples were centrifuged at 3 minutes at 1000g to bind gDNA and then for 1 minute at maximum speed (>12,000g) to clear the membrane. The column was transferred to a fresh column and washed twice with gDNA wash buffer. To clear the column a final centrifugation step at maximum speed was performed. DNA was eluted in water that was pre-warmed to 65^{o}C per the kit recommendations.

2.3.4 Nucleic acid quantification

The Nanodrop spectrophotometer (Nanodrop ND-1000, Thermo Fisher) was used to assess quality and quantity of DNA. Absorption at 260nm was used to determine DNA concentration. The ratio of absorbance at 260nm and 280nm was used to assess the purity of DNA. The absorbance ratio at 260nm and 230nm was used as an indication of contamination. For more sensitive DNA quantification, the Qubit 3 Fluorometer (Invitrogen, 15387293) was utilised. Quantification was performed according to the manufacturer's protocol. The appropriate Qubit reagents were used, depending on the sample type. For DNA, typically the dsDNA Broad Range buffer was used for detection.

2.3.5 Restriction digest

Restriction digests were performed at 37°C (unless otherwise stated) with 250 RPM shaking. Appropriate buffer was used for each individual reaction according to the NEB double-digest tool. Restriction enzymes did not exceed 10% of total reaction volume. Restriction products were resolved on an agarose gel by gel electrophoresis.

2.3.6 Agarose gel electrophoresis

Agarose gel was prepared at the required concentration of 0.8-2.0%, depending on the size of the DNA fragment to be visualised. UltraPure Agarose (Invitrogen, 16500500) was mixed with 1x TBE and microwaved for 2-3 minutes with intermittent shaking to mix. Ethidium bromide (EtBR, 1µl) was added to the gel to stain DNA fragments for visualisation under UV light. Purple loading dye (NEB, B7024) was added to samples when loading. Size selection markers for 1kb (NEB, N3232) or 100bp (NEB, N3231) were included when loading the gel.

2.3.7 Methylation of plasmid DNA

Plasmid DNA was methylated *in vitro* using the CpG methyltransferase *M.SssI* (NEB, M0226), according to the manufacturer's protocol. For transfections, larger amounts of DNA were required, therefore, the protocol was adjusted for methylating up to 8µg DNA by increasing the concentration of methyl-donor S-Adenosyl methionine (SAM) to 3720µM. Briefly, plasmid DNA (8µg) was incubated with M.SssI (1µl per µg DNA), 10µl SAM (3720µM), 10µl NEB 10X buffer and water (total volume 100µl) at 37°C overnight. Reaction was deactivated by heating to 65°C for 20 minutes. To assess plasmid methylation, DNA was digested using MspI (NEB, R0106) and HpaII (NEB, R0171) restriction enzymes for up to 1 hour at 37°C. 1 hour of digestion is required to obtain identical digestion patterns between unmethylated and methylated plasmids, but 15 minutes is sufficient to observe digestion. Restriction products were resolved on a 2% agarose gel. Methylated DNA was purified using a QIAGEN Mini-prep spin column (QIAGEN, 27104). To increase purification yield, DNA was incubated for 5 minutes with 1 volume 100% EtOH prior to purification. The DNA/EtOH mixture was applied to the

column and centrifuged at 13,000 RPM**.** The column was washed with 80% EtOH and DNA was eluted in 30µl nuclease free water.

2.3.8 qPCR

Quantitative PCR (qPCR) was conducted using LightCycler 480 SYBR Green I Master Mix (Roche, 04707516001), according to the manufacturer's protocol. Briefly, ~25ng gDNA was combined with 5µM primers and SYBR Green Master Mix. The qPCR reaction was conducted on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad) according to the conditions in table 1. Triplicate technical replicates were included for each sample. Data was analysed using the ΔΔCT method. BRCA2 was used as an internal genomic control. The integration rate of methylated piggyBac was calculated by normalising it against unmethylated piggyBac integration. For primers, see Appendix.

Table 1: qPCR conditions

	Temperature (°C)	Time	
Initial Denature	95	5 min	x50
Denature	95	6 sec	
Anneal	60	6 sec	
Extend	72	6 sec	
Image capture			

2.3.9 Mass Spectrometry 5mC quantification

DNA samples were isolated by phenol-chloroform extraction, as described in 2.3.2. Liquid chromatography–mass spectrometry (LC–MS) quantification of 5mC levels in HCT116 cells treated with GSK-3484862 was conducted by Nandini Mozumdar, Kriaucionis lab, University of Oxford. Briefly, nucleic acid hydrolysis was performed at 37°C for 2 hours using Degradase Reaction buffer (Zymo Research (Zymo Research, E2020) and Degradase Plus enzyme (Zymo research, E2020). Samples were added to a Supor filtration plate (Agilent, 203940-100) mounted on a (V-shaped) 96-well (Agilent, 5043-9313) MS plate and centrifuged at 3220g for

35mins and 4°C. The volume in each well was made up to 40µl with water. The 96-well plate was sealed with sealing mat (Agilent, 5043-9317). Nucleic acid quantification was conducted using the Agilent 6495 triple quadrupole LC/MS system. Column: Zorbax C18 (1.8 µm, 2.1 mm 150 mm; Agilent, 859700-902). Data was acquired in dMRM mode.

2.4 Flow cytometry

2.4.1 Flow cytometry sample preparation and data acquisition

To prepare samples, cells were detached from culture flasks as described in 1.1.2. Cell suspension was transferred to 15ml Falcon 1.5ml Eppendorf tubes, and centrifuged (1000RPM for 5 minutes). Cell pellets were washed with 1ml PBS. Cells were resuspended in 1-2ml PBS, volume adjusted depending on cell density. Cells were kept on ice prior to analysis. DAPI (10µg/ml) was used to stain dead cells. Data was acquired on a LSR Fortessa II with DIVA software.

2.4.2 Flow cytometry data analysis with FlowJo

Flow cytometry data was analysed using FlowJo software version 10.10.0 (TreeStar). The following gating strategy was used: first, samples were assessed on side scatter (SSC-A) and forward-side scatter (FSC-A) and debris was removed by gating for cells based on density, remove any overly large or small events. Next, single cell selection was performed using forward scatter heigh (FSC-H) and forward side scatter area (FSC-A). DAPI-negative cells were than selected, removing DAPI-positive dead cells. A sample without DAPI staining was used to identify DAPI-negative cells. GFP positive cells were identified using a GFP-negative control cell population, for example HCT116 cells without eGFP. The distribution of GFP fluorescence intensity from cells was assessed on a histogram.

The gates for high and low GFP expression were defined using a GFP-positive control population. To do this, GFP-positive cells containing the unmethylated eGFP reporter were used. A GFP-positive control was used for each individual cell line and experimental replicate

conducted, as an internal control. Differences in fluorescence intensity were noted between samples and cell lines, therefore, inclusion of an unmethylated control is essential for each experiment. The GFP-positive population was used to set the high and low GFP gating for each corresponding methylated reporter sample. In stably transfected cells, the high/low GFP gate was set such that it included 95% of the unmethylated GFP-positive control population. This threshold was determined based on the fluorescence intensity of GFP positive cells, assed on a histogram. The flow cytometry gate was set such that 95% of the events in the unmethylated GFP-positive population fell within the gate. This gate was then applied to the corresponding methylated reporter population, ensuring unbiased delineation of high GFP-expressing cells within the methylated reporter population. This approach was used to analyse all stably transfected cell lines to ensure consistent gating of high GFP-expressing cells between samples and cell lines. In transiently transfected cells, a higher level of background fluorescence was observed in cells. This resulted in a tail of distribution when assessing GFP intensity on a histogram. Therefore, in transiently transfected cells, the 'high GFP' gate was set by eye at a consistent GFP intensity (10^4), to more accurately capture the peak of GFP expression in these cells.

2.4.3 Cell cycle profile analysis

Cells were pelleted by centrifugation (1000RPM for 5 minutes) then washed with PBS and re-pelleted. Cells were then resuspended in an equal parts mixture of PBS and FCS. Live cells were fixed by adding ice-cold freshly made 70% EtOH in a drop-wise manner while vortexing. Cells were fixed overnight at 4^{o}C then washed with PBS. Cells were stained with DAPI (1µg/ml) and incubated in the dark for 1 hour. Cells were stored on ice prior to analysis by flow cytometry. Flow cytometry data was acquired on the LSR Fortessa II with DIVA software and analysed using FlowJo software version 10.10.0 (TreeStar), as above. Data was fit to the Watson Pragmatic algorithm, a mathematical model to estimate the proportion of cells within each phase of the cell cycle (Watson, Chambers and Smith, 1987).

2.5 Fluorescence microscopy

2.5.1 Two-dimensional epifluorescence imaging

For two-dimensional live cell imagining of HCT116 cells, $5x10^5$ cells were seeded in 6 well glass bottom plates with black frames (Cellvis, P06-1.5H-N) 24 hours prior to imaging. Environmental control of the cultured cells was maintained during imaging with a Okolab bold line stage top incubation chamber incorporating temperature and humidified CO_2 control (OkolabS.R.L, Italy). All images were captured on a Nikon Eclipse Ti-E inverted microscope (Nikon Instruments UK Ltd, Kingston Upon Thames, UK) using the 10x or 20x objective lens. Widefield illumination was provided by a CoolLED PE-4000 LED light source (CoolLED Ltd, Andover, UK) combined with LED-CFP/YFP/mCherry-3X-A (Pinkel) filter sets (IDEX Health & Science, LLC, Center of Excellence, New York, United States). Images were acquired with a Photometrics Prime BSI sCMOS camera (Teledyne Photometrics, Tucson, AZ 85706). Image capture was performed manually using Micromanager software (Version 1.4). During image capture the objectives and magnifications were set on the Micromanager software and scale bars were applied to the images. The exposure was set at a consistent value, ensuring consistency between images.

2.6 Mapping piggyBac integration sites

2.6.1 Asymmetric-nested PCR

PiggyBac insert sites were mapped using a asymmetric PCR-based approach (Franke, 2017). The process involved a two round of PCR amplification. In the first PCR, primers specific to the piggyBac reporter construct (PL1 or PR1) targeting the edge of the LTRs and pointing outwards from the insert sequence, and a set of random primers, were used. The random primers contain a 5' adaptor sequence (KmonP), a random nucleotide sequence, and a 'seed' sequence. The seed sequence is designed to bind sequences within the genome every ~200-300bp. A multi-step PCR reaction was used. The first ten cycles of the PCR are designed to favour linear amplification from the piggyBac specific primers by using high annealing temperatures. In the remaining cycles, PCR asymmetry is maintained by interchanging high

and low annealing temperatures. The PCR was performed using Platinum Green Hot Start PCR Master Mix (Thermo Fisher, 13001012) according to the manufacturers protocol. Reaction: 50ng of input DNA, 20µM PL1/PR1 primer and 100µM KmonP-N7-CTAG, PCR conditions shown in Table 2. Note: it is essential to use a polymerase that leaves 3'-A overhangs for downstream cloning and sequencing.

Table 2: Round 1, asymmetric-nested PCR conditions.

	Temperature (°C)	Time	
Initial denaturation	95	3 min	
Denature	94	20 sec	x9
Anneal	64	45 sec	
Extend	72	3 min	
Denature	94	20 sec	x12
Anneal	64	45 sec	
Extend	72	3 min	
Denature	94	20 sec	
Anneal	64	45 sec	
Extend	72	3 min	
Denature	94	20 sec	
Anneal	44	1 min	
Extend	72	3 min	
Final Extension	72	7min	

In the second round, a nested PCR is used to further amplify the products from the first round, while minimising off-target amplification from the random primers. In this round a 1:100 dilution of the first PCR was used as a template. Primers that target the primer sequences from the first PCR round were used. The PL2/PR2 primers target within PL1/PR1. The KmonP primer anneals to the adaptor sequence contained in the KmonP-N7-CTAG random primer. The final amplification product should contain the sequence flanking the piggyBac integration site. Again, Platinum Green Hot Start PCR Master Mix (Thermo Fisher, 13001012) was used as this Taq leaves 3'-A overhangs, which are required for the cloning steps. Primers: PL2/PR2 20µM and KmonP 20µM, PCR conditions shown in Table 3. PCR products were resolved on a 1% agarose gel. Successful amplification resulted in a smear with different banding patterns. For primers, see Appendix.

Table 3: Round 2, asymmetric-nested PCR conditions.

	Temperature (°C)	Time	
Initial denaturation	95	4 min	
Denature	94	20 sec	x9
Anneal	59	45 sec	
Extend	72	3 min	
Denature	94	20 sec	
Anneal	59	45 sec	
Extend	72	3 min	
Denature	94	20 sec	
Anneal	53	45 sec	
Extend	72	3 min	
Denature	94	20 sec	
Anneal	59	45 sec	x8
Extend	72	3 min	
Denature	94	20 sec	
Anneal	53	45 sec	
Extend	72	3 min	
Final Extension	72	7 min	

2.6.2 Cloning, sequencing and mapping

The asymmetric-nested PCR products from the second round of PCR amplification in 1.6.1 were purified using AMPure XP magnetic beads (Beckman Coulter, A63881) at a 1.8X ratio in order to retain all fragment sizes. The purified PCR product was then ligated into the pGEM-T Easy blue-white screening vector (Promega, A1360) at 4°C overnight. Next, 3µl of the vector ligation mixture was transformed with Subcloning Efficiency DH5α Competent Cells (Thermo Fisher, 18265017) and grown overnight at 37°C on X-Gal Ampicillin plates, as described in 2.3.1. White colonies were selected and inoculated in 1ml LB media. Mini-preparation and Sanger sequencing was performed by the IGC Sequencing Facility. The SP6 sequencing primer, provided by the IGC Sequencing Facility, was used for sequencing. This primer binds to the pGEM-T Easy Vector backbone at a site flanking the multiple cloning site, allowing sequencing of the ligated PCR fragment. The resulting sequence was used for a BLAT search (UCSC genome browser) against the hg38 human genome (Kent, 2002). The insert locations were converted to BED files for downstream analysis.

2.7 Methylation sequencing

2.7.1 Methyl primer design

MethPrimer software was used to design primers for Bisulfite (BS) and Enzymatic-Methyl (EM) PCR (Li and Dahiya, 2002). The software coverts the target DNA sequence into BS converted sequence, by performing C to T conversion. The software was initially developed for BS-converted DNA, but is suitable for use with EM converted DNA because the sequence of the final DNA product is the same after the two conversion methods. Primers were designed to be 20-30bp long and the size of the amplicon varied between 250-350bp. The NEB Tm calculator was used to calculate primer melting temperature and appropriate PCR conditions.

2.7.2 Bisulfite PCR

Bisulfite conversion of genomic DNA was performed using Zymo Research EZ DNA methylation Gold Kit (Zymo, D5005). Modifications were made to the manufacturer's protocol. As standard, 500ng of input DNA was used. Input DNA was sonicated using the Biorupter Plus sonication device (Biagenode, B01020001). 100µl of gDNA at a concentration of 50ng/µl was sonicated for 3 cycles of 30 seconds on/30 seconds off in 4^{o}C water. This step shears the gDNA and was found to increase conversion efficiency. Converted DNA was eluted in 10µl M-elution buffer with 5 minutes incubation at room temperature. A second elution was performed with a further 10µl to increase yield, for a total of 20µl eluate. The remainder of the protocol was performed according to the manufacturer's recommendations. Target regions of the genome or the piggyBac CAG_eGFP_IRES_Puro reporter construct were amplified by PCR using primers designed to target BS converted DNA (see 2.7.1). The TaKaRa EpiTaq Hot Start polymerase (TaKaRa Bio, R110A) using ~50ng input DNA and primers at 5µM. The PCR conditions are listed in Table 4 and for primers, see Appendix.

Table 4: Bisulfite PCR conditions

	Temperature (°C)	Time	
Initial denaturation	95	30 sec	
Denature	95	20 sec	x30-40
Anneal	55	30 sec	
Extend	72	30 sec	
Final Extension	72	10 sec	

2.7.3 Enzymatic-Methyl PCR

Enzymatic-Methyl (EM) conversion was conducted using the NEBNext Enzymatic Methyl-seq Conversion Module (NEB, E7125). The manufacturer's protocol was followed closely, with one adjustment: the input gDNA was sonicated prior to starting the conversion. 100µl of gDNA at a concentration of 50ng/µl was sonicated for 2 cycles of 30 seconds on/30 seconds off in 4°C water. Then, 4µl of sonicated DNA (for a final total of 200ng input DNA) was diluted in a final volume of 28µl of water, as recommended by the protocol. AMPure XP magnetic beads (Beckman Coulter, A63881) were used for the purification steps. Beads were washed with freshly made 80% EtOH. Target regions were amplified from converted DNA by PCR using primers designed to target EM converted DNA (see 2.7.1). The PCR was performed using TaKaRa EpiTaq Hot Start polymerase (TaKaRa Bio, R110A) with the recommended reaction mixture and ~50ng of input DNA. The primers are listed in the Appendix. For PCR conditions, see Table 3.

2.7.4 Sanger sequencing (BS/EM-seq) and analysis

For both BS- and EM-PCR, PCR products were purified using QIAGEN MinElute PCR Purification Kit (QIAGEN, 28004). Purified PCR products (3µl) were ligated into the PGEM® T-Easy Vector (Promega, A1360) at 4°C overnight. Ligated vectors were transformed with Subcloning Efficiency DH5α Competent Cells (Thermo Fisher, 18265017), as described in 2.3.1. For blue/white screening, cells were plated on Ampicillin X-Gal agar plates and grown overnight

at 37°C. White colonies were selected and cultured overnight in LB medium supplemented with Ampicillin. Plasmid DNA was either extracted manually using the QIAprep Spin Miniprep kit (QIAGEN, 27104), or minipreparation was performed by the IGC Sequencing Facility. Samples were Sanger-sequenced by the IGC Sequencing Facility. Sequencing data was analysed using the Quantification tool for Methylation Analysis (QUMA) software, and the graphical representation of methylation sequencing data was generated using this software (Kumaki, Oda and Okano, 2008).

2.7.5 Illumina iSeq library preparation (EM-seq)

Targeted deep amplicon EM-sequencing was performed by Illumina next-generation sequencing. To prepare sequencing libraries, EM-PCR was performed as described in 2.7.3 using locus-specific primers containing a 4bp unique molecular identifier and Nextera adaptor sequence. The PCR products were resolved on a 1.5% agarose gel then purified using AMPure XP magnetic beads (Beckman Coulter, A63881) at 1X to remove primer dimer. DNA was quantified on the Qbit and amplicons from the same genotype were pooled. Sequencing libraries were amplified using unique dual index (UDP) Illumina adapters and NEBNext Ultra II Q5 Master Mix (NEB, M0544). The PCR conditions are listed in Table 5. Libraries were purified using AMPure XP magnetic beads (Beckman Coulter, A63881) at 0.7X. Libraries were quantified on the Bioanalyser (Agilent, 2100) using the DNA 1000 chip by the IGC Sequencing Facility. Libraries were brought to 10nM and samples with unique UDPs were pooled for sequencing. Sequencing (2x150 paired-end) was performed on the Illumina iSeq 100 platform using the iSeq 100 i1 Reagent v2 (300 cycle) Kit. Due to the low sequence diversity of the library, a PhiX Control v3 (Illumina) was spiked into the sequencing run at a concentration of 50% to help improve cluster resolution and avoid run failure. Sequencing was performed by the Edinburgh Clinical Research Facility Genetics Core facility. For primers, see Appendix.

Table 5: Illumina iSeq library amplification PCR conditions.

	Temperature (°C)	Time	
Initial denaturation	98	30 sec	
Denature	98	10 sec	x6
Anneal/Extension	65	75 sec	
Final Extension	65	5 min	

2.8 Southern blotting

2.8.1 Human Satellite II probe labelling

The Southern blot Human Satellite II (HSatII) probe was created using a Digoxigenin (DIG)-11-uridine triphosphate (UTP) (Roche, 12352200) labelled PCR. The HSatII fragment was amplified by PCR from the template plasmid p375M2.4 (Jackson, Mole and Ponder, 1992) using Taq DNA Polymerase recombinant (Invitrogen, 10342020). A 25µl PCR reaction was assembled according to the manufacturer's protocol using DIG-11-dUTP in a ratio of 1:19 to dTTP and 10ng of template DNA (see Table 6 for PCR conditions). Successful probe labelling was assessed by running the labelled probe on an agarose gel and comparing the size to an unlabelled PCR. The labelled probe is larger than the unlabelled probe, product size: 1-1.5kb.

Table 6: DIG-UTP probe labelling PCR conditions.

	Temperature (°C)	Time	
Initial denaturation	95	3 min	
Denature	95	45 sec	x25
Anneal	53	30 sec	
Extend	72	2 min	
Final Extension	72	3 min	

2.8.2 Human Satellite II Southern blot

To detect methylation levels at Human Satellite II sequences, gDNA (1.5-3µg) was digested with 20-50U BstBI (NEB, R0519) restriction enzyme at 65°C overnight. Digested DNA was run on a 0.8% agarose gel at 4°C. To obtain clear separation of bands, 40 volts were applied for 15 minutes and then increased to 80V and run for ~4 hours, or until the loading dye had migrated through ¾ of the gel. To visualise the DNA, the gel was post-stained with EtBR and assessed under UV light.

To denature the DNA, the gel was submerged in denaturation solution (0.5M NaOH, 1.5M NaCl) for 2x15 minutes at room temperature with rocking, then washed with ddH_2O. The gel was then submerged in neutralisation solution (0.5 Tris Ph7.5. 1.5M NaCl) for 2x15 minutes at room temperature with rocking. The gel was equilibrated for 10 minutes in 20X SSC. DNA was then transferred to Hybond N+ hybridization membrane (Merck, 41105339) by overnight

blot. The following day, DNA was crosslinked to membrane by UV irradiation at 0.15J. The blot was incubated with DIG Easy Hyb (Merck, 41105326) pre-hybridisation buffer at 42^{o}C for 30 minutes with rocking. The blot was then probed for HSatII sequence using DIG-labelled HSatII probe for 4 hours at 42^{o}C with rocking. The blot was washed with low stringency buffer (2XSSC, 0.1% SDS) for 2x5 minutes at room temperature with rocking, then high stringency buffer (0.5XSSC, 0.1% SDS) 2x15 minutes at 65^{o}C. Probe detection was performed using the DIG Luminescent Detection Kit (Roche, 11363514910) according to the manufacturer's protocol. Briefly, the blot was washed in MABT wash buffer (1X Maleic acid buffer pH7.5, 0.1M Maleic Acid 0.15M NaCl; MABT: 1X Maleic Acid Buffer, 0.3% Tween-20). Next, the blot was submerged in 1X bocking solution (Blocking reagent, 11096176001) for 30 minutes at room temperature. The blot was then incubated with 20µl Anti-Digoxigenin-AP (Merck, 11093274910) antibody solution for 30 minutes at room temperature with agitation, then washed with MABT wash buffer. The blot was equilibrated in detection buffer (0.1M Tris-HCl, 0.1M NaCl, pH 9.5) for 5 minutes at room temperature. Finally, the blot was sealed in a plastic bag with CSPD solution (Roche, 11755633001) and incubated for 5 minutes prior to image acquisition.

Images were acquired using the LightCycler and ImageQuant LAS 4000 (GE Health Care Life Sciences) imaging system using the chemiluminescent setting. Gel band intensity was quantified using the gel analysis tool in Fiji ImageJ software (Schindelin *et al.*, 2012). The relative level of hypomethylation was calculated by taking the band intensity of methylated over unmethylated regions for each sample.

2.8.3 LINE1 Southern blot

The LINE1 DNA probe, which targeted a specific fragment of the LINE1 sequence, was prepared and gifted by Mihail Peychev. The probe was labelled with DIG-dUTP using the Roche DIG DNA Labelling Kit (Roche, 11175033910) following the manufacturer's guidelines. The probe was stored at -20^{o}C. Before its application in the Southern blot, the probe was denatured in 0.4M NaOH and 10 mM EDTA, followed by neutralization with 0.5M Tris-HCl (pH 7.5). To detect methylation levels at LINE1 sequences, gDNA (1.5-3µg) was digested with MspI

(NEB, R0106) or HpaII (NEB, R0171) restriction enzymes at 37°C overnight. The remaining Southern blot procedure was performed as described in 2.5.2, apart from use of the LINE1 probe where appropriate.

Analysis of blot of the Southern blot signal was performed using the gel analysis tool in Fiji ImageJ software (Schindelin *et al.*, 2012). The degree of digestion by MspI and HpaII was assessed by generating a scan profile plot for each lane of the gel. This was done by analysing signal intensity in each lane, by highlighting the entire lane and generating a plot of the signal intensity down the lane. The resulting line representations of band intensity were exported to Adobe Illustrator and used to create a line plot.

2.9 RNA methods

2.9.1 RNA isolation and handling

RNA was extracted from cell pellets using the RNeasy Mini Kit (QIAGEN, 74104) according to the manufacturer's protocol. Briefly, cell pellets (~$1x10^6$ cells) were thawed on ice and mixed with Buffer RLT lysis buffer to lyse cells. Cell lysates were homogenized by passing through a 21G syringe and needle 5-10 times. Samples were transferred to the RNeasy spin column and centrifuged at ≥8000g to facilitate sample binding to the column. On-column DNase digestion was performed. The column was washed twice using Buffer RPE, then transferred to a fresh collection tube and centrifuged for 1 minute to remove residual wash buffer. RNA was eluted in RNAse-free water. The provided RNAse-free tubes were used throughout, and work area was cleaned with EtOH prior to RNA extraction to avoid contamination. RNA samples were aliquoted in 10-20µl aliquots and stored at -70°C.

2.9.2 RNA sequencing

RNA quality and quantity was assessed on the Bioanalyser (Agilent, 2100) using the RNA 6000 Nano chip by the IGC Sequencing Facility. Total RNA sequencing libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760) with the with

the NEBNext rRNA Depletion Kit v2 (NEB, E7400) for ribonucleotide depletion. Libraries were sequenced on the Illumina NextSeq 2000 RNA-Sequencing platform using the P3 Kit (200 cycles, 2x150 paired-end, 1.2B reads). Library preparation and RNA-sequencing was performed by the Edinburgh Clinical Research Facility Genetics Core facility.

2.10 Bioinformatics

2.10.1 Data analysis and visualisation

Data visualisation and analysis was primarily conducted using GraphPad Prism v10.2.3. Bar plots, box plots, violin plots and histograms were generated in GraphPad. The GraphPad built-in statistical analysis tool was used to analysed data, the specific statistical analyses are indicated where relevant. RNA-Sequencing data was analysed using R Studio v4.4.0. Volcano plots were generated using the EnhancedVolcano package.

2.10.2 piggyBac insert site analysis

Bioinformatic analysis of piggyBac integration sites was performed using command line tools. The BEDTools v2.31.0 intersect feature was used to assess overlaps between piggyBac insert locations and genomic features such as PMDs. The *-wa -u -a* options were used to ensure that each overlap was only counted once. BED file annotations of PMD locations in HCT116 cells generated by Ioannis Kafetzopoulos (Sproul lab, IGC) were used for intersect comparisons.

2.10.3 iSeq data processing

RNA-Sequencing reads were demultiplexed and the Illumina indexes were removed from the FASTQ headers for compatibility with UMI tools. The extract function of UMI tools (v1.0.0 with --extract-method=string --bc-pattern = NNNN --bc-pattern2 = NNNN settings) was used to remove the UMIs from the FASTQ files and place them in the header (Smith, Heger and Sudbery, 2017). For methylation sequencing analysis, a custom ‘bisulfite-treated’ (C to T

conversion) reference genome was created in the command line. The reference genome included the hg38 human genome and the piggyBac CAG_eGFP_IRES_PURO vector sequence. A custom command line script written by Lyndsay Kerr (former Sproul lab) was used for analysis. The reads were trimmed (TrimGalore v0.6.6 with the options: --paired), aligned to the reference genome (Bismark v0.22.3 with options: --multicore 3 -N 0 -L 20), indexed using Samtools (v1.13) and deduplicated using the dedup option of UMI tools (v1.0.0 with options: --paired --method=unique). The mean percentage of methylated CpG sites in each sequencing read was calculated using the command line. The bismark_methylation_extractor function of Bismark (v0.22.3 with options: -p --no_header --no_overlap) was utilised to determine the methylation state of CpGs on individual reads. The command line was then used to identify the number CpGs per read, the number of methylated CpGs per read, and the overall mean level of methylation for each read.

2.10.4 RNA sequencing data processing

The RNA-Sequencing FASTQ data was processed, and gene differential expression analysis was performed by Dr Philippe Gautier, IGC. Briefly, data was processed using the nf-core RNA-Seq v3.10.1 package of the nf-core collection workflow (Ewels *et al.*, 2020). The pipeline was executed with Nextflow v22.10.7 (Di Tommaso *et al.*, 2017). Breifly, reads were aligned to the human genome (GRCh38) using STAR (v2.7.10a). Reads aligned to transcripts were counted using Salmon (v1.9.0). Multiple quality control of the sequence and alignment quality were made and gathered in a report (multiqc_report.html). Differential expression analysis was performed using the Deseq2 R package (v1.34.0, running on R v4.1.2). The package vignette was followed, using the raw count file produced by nf-core. A basic pre-filter was applied beforehand, removing all genes with no expression in all samples. The apeglm method was used to output both unshrunken or shrunken log2fold values. The normalised counts data, which is produced during sample processing, was used for downstream GSEA analysis (see 2.10.5). GO term enrichment analysis was performed using the enrichGO package from ClusterProfiler v3.0.4 on R. Genes with adjusted p-value <0.05 and an absolute Log2 fold change >2 were analysed.

2.10.5 Transposable element analysis

To assess transposable element expression, RNA-sequencing data was processed and transposable elements were mapped by Prof Ian Adams, IGC, using a custom script. Briefly, the RNA-sequencing reads were trimmed to remove the adaptor sequence using TrimGalore! v.0.4.1 and the paired-end Illumina stringency 3 setting. Reads were than aligned to the human genome (hg38) using TopHat v2.1. Stringent thresholds were set, to ensure reads were mapped accurately. Thresholds included: inner distance set to 54 ± 79, with no coverage search, and the maximum multi-hits set to 1. The output from TopHat was intersected with RepeatMasker (UCSC genome browser) track coordinates for repeat sequence locations using the bedtools intersect feature of BEDTools v2.31.0 The number of reads identified within each repeat element family was summed. The reads that were mapped to repeat elements were then analysed using the Deseq2 R package (v1.34.0). I then performed further analysis in R Studio v4.4.0 to plot differentially expressed transposable elements.

2.10.6 Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using RNA-seq normalised count data in GSEA v4.4.3 (Subramanian *et al.*, 2005). The default parameters were maintained. The Hallmark gene sets were derived from the Human Molecular Signatures Database (MSigDB) (Liberzon *et al.*, 2015). Results with a false discover rate of <0.25 and a normalised enrichment score of >1.5 were considered statistically significant.

3 A methylation-sensitive reporter system to measure DNA methylation maintenance

3.1 Introduction

I hypothesised that the widespread hypomethylation observed in cancer cells, resulting in the formation of partially methylated domains (PMDs), is caused by a failure to maintain DNA methylation patterns in these cells. In order to test this hypothesis, I aimed to develop a reporter system that would allow us to track the degree to which human cancer cells are able to maintain and propagate methylation states across the genome.

Existing work that has investigated DNA hypomethylation in cancer has been based on whole-genome bisulfite sequencing (WGBS) data, a method that enables sequencing and detection of methylated CpG sites throughout the genome (Salhab *et al.*, 2018; Brinkman *et al.*, 2019). This data provides a static snapshot of methylation levels across the genome at a single point in time from a bulk population of cells. Although informative, WGBS data provides a limited insight into the dynamic process of DNA methylation maintenance. For the purpose of my study, I was interested in tracing the fidelity DNA methylation maintenance over time in cancer cells. Some previous work has interrogated DNA methylation and maintenance of the methyl mark by integrating methylated sequences and assessing stability of the signal over time. A previous study used recombination-mediated cassette exchange (RMCE), a site-specific recombinase approach, to introduce an *in vitro* methylated green fluorescent protein (GFP) reporter construct (Schübeler *et al.*, 2000). This study demonstrated that the GFP reporter was repressed by methylation, and the methyl mark was faithfully propagated at the promoter and gene body, but not the enhancer, over time (Schübeler *et al.*, 2000). A second study demonstrated that methylation density influences the propagation of methylation states, by introducing fully and partially-methylated GFP reporter constructs into cells and monitoring the expression of the reporter and its methylation state over time (Lorincz *et al.*, 2002). However, both studies utilised the same site-directed RMCE integration approach. While this approach has some benefits, such as controlling for the site of integration and minimising variable effects of different integration sites on transgene expression, it also limits the number of sites that can be studied to a small number of candidate regions. It has been suggested that genomic features such as replication timing or chromatin structure can

influence methylation maintenance (Zhou *et al.*, 2018). Both studies integrated the methylated transgene into the RL5 and RL6 sites in Murine erythroleukemia (MEL) cells, both shown to be early replicating sites and in non-repressive chromatin environments (Feng *et al.*, 1999, 2001; Schübeler *et al.*, 2000). Therefore, these studies may not be able to capture potential positional effects on methylation maintenance. In addition, these studies were both performed in MEL cells and, therefore, may not be representative of processes occurring in human cells.

Building on this work, I sought to develop a tool to track DNA methylation maintenance in human cells at different genomic loci. To do this, I decided to develop a methylation-sensitive reporter system. Reporter systems are useful tools to track and visualise gene expression, using reporter genes. Reporter genes are typically genes that encode a fluorescence protein, which enables visualisation of gene expression from cells by fluorescence microscopy or flow cytometry.

I was interested in sampling many different genomic locations, to gain insight into how methylation is maintained across the genome. If I was to use a site-directed recombination approach to introduce to reporter gene, I would first need to establish a number of cell lines with recombination cassettes within desired locations. Alternatively, the reporter construct could be randomly integrated across the genome using a transposon. Transposable elements are mobile genetic elements, which move through a DNA intermediate that is excised from one site and integrated into another (Uren *et al.*, 2005). Transposon-mediated integration can be used to obtain thousands of different integrations within a cell population in a single transfection (Akhtar *et al.*, 2014). The piggyBac transposon is reported as the most efficient transposon, in terms of transposition activity (Wu *et al.*, 2006; Liang *et al.*, 2009). The piggyBac transposon has been previously used to stably integrate reporter genes across the genome to study positional effects on gene expression (Akhtar *et al.*, 2013, 2014). Furthermore, the piggyBac transposon has been utilised to integrate methylated and unmethylated CpG islands (CGIs) in mouse ESCs and human cancer cells, and was shown to recapitulate the expected methylation state following integration (Stelzer *et al.*, 2015; Masalmeh, 2019; Masalmeh *et al.*, 2021). This indicates that the piggyBac transposon system is not hindered by DNA methylation and does not alter the methylation state of the integrated DNA sequence, making it an ideal candidate for use in this reporter system.

3.2 Experimental design

In this chapter, I sought to develop a methylation-sensitive reporter system to measure maintenance of DNA methylation in cells. I chose to make use of the enhanced GFP (eGFP) reporter gene, a fluorescent reporter that quantitively reports gene expression levels (Soboleski, Oaks and Halford, 2005). This reporter gene was placed under control of the CAG promoter, a synthetic CGI promoter that efficiently drives gene expression in a variety of cell types (Alexopoulou, Couchman and Whiteford, 2008). I hypothesised that the reporter gene could be repressed by methylation at the CGI promoter, based on evidence from previous studies, as discussed above (Schübeler *et al.*, 2000; Lorincz *et al.*, 2002). Therefore, the ability of cells to maintain methylation at the promoter could be inferred by assessing expression of the reporter gene. To perform random transgene integration, I chose to utilise the piggyBac transposon system. The integrated reporter construct contained the eGFP reporter gene under control of the synthetic CAG promoter construct, which contains the cytomegalovirus (CMV) enhancer element fused to the chicken beta-actin promoter, and splice acceptor of the rabbit beta-globin gene (Hitoshi, Ken-ichi and Jun-ichi, 1991). The construct also contained an internal ribosome entry site (IRES) and encodes Puromycin resistance, for antibiotic selection of cells containing integrations. These elements are flanked by inverted terminal repeat (ITR) sequences which are integrated into genomic DNA (gDNA) alongside the reporter construct (Fig. 3.1A schematic). This vector was used as a control plasmid in Taglini *et al.* (2024) (Taglini *et al.*, 2024).

In order to study maintenance of DNA methylation, it was necessary to add methylation to the reporter construct. The aim was to begin from a known baseline state of a sequence with all CpG sites methylated and track maintenance over time. I methylated the delivery vector *in vitro* using the *M.SssI* methyltransferase. This methyltransferase has been used in previous studies (Schübeler *et al.*, 2000; Lorincz *et al.*, 2002), and efficiently methylates all cytosines within the CpG dinucleotide context (Renbaum *et al.*, 1990). The entire vector was methylated, including the CAG promoter, eGFP reporter gene, the ITRs and plasmid backbone (see Fig. 3.1A for schematic).

To deliver the PB_CAG_eGFP_IRES_Puro reporter construct to cells, I performed a co-transfection with the piggyBac vector containing the eGFP reporter gene and a separate

vector expressing the piggyBac transposase enzyme (Fig. 3.1B). The transposase recognises the sequence of the ITRs, excises the DNA contained within the ITRs, and mediates insertion of this sequence into genomic DNA (gDNA) through a 'cut and paste' mechanism (Fraser *et al.*, 1996; Bauser, Elick and Fraser, 1999; Wilson, Coates and George, 2007).

Following integration of the PB_CAG_eGFP_IRES_Puro construct, I investigated maintenance of the DNA methylation mark over time (Fig. 3.1C schematic). Unless otherwise stated, cells were maintained in a mixed population and no selection was carried out. I chose not to select cells using the Puromycin resistance encoded in the reporter construct, in order to avoid removing any cells containing reporters that were silenced by the methylation. Expression of the fluorescent reporter gene was assessed by fluorescence microscopy and flow cytometry. To assess methylation of the reporter construct, I extracted gDNA and carried out either bisulfite (BS) treatment or enzymatic-methyl (EM) conversion. I performed PCR on the converted DNA using primers specific to the reporter construct and conducted sequencing to determine the methylation state.

In this chapter, I test whether the reporter gene can be repressed by methylation and stably integrated into the genome through piggyBac transposition. I also test whether expression of the reporter gene corresponds to its underlying methylation state following integration into human cancer cells.

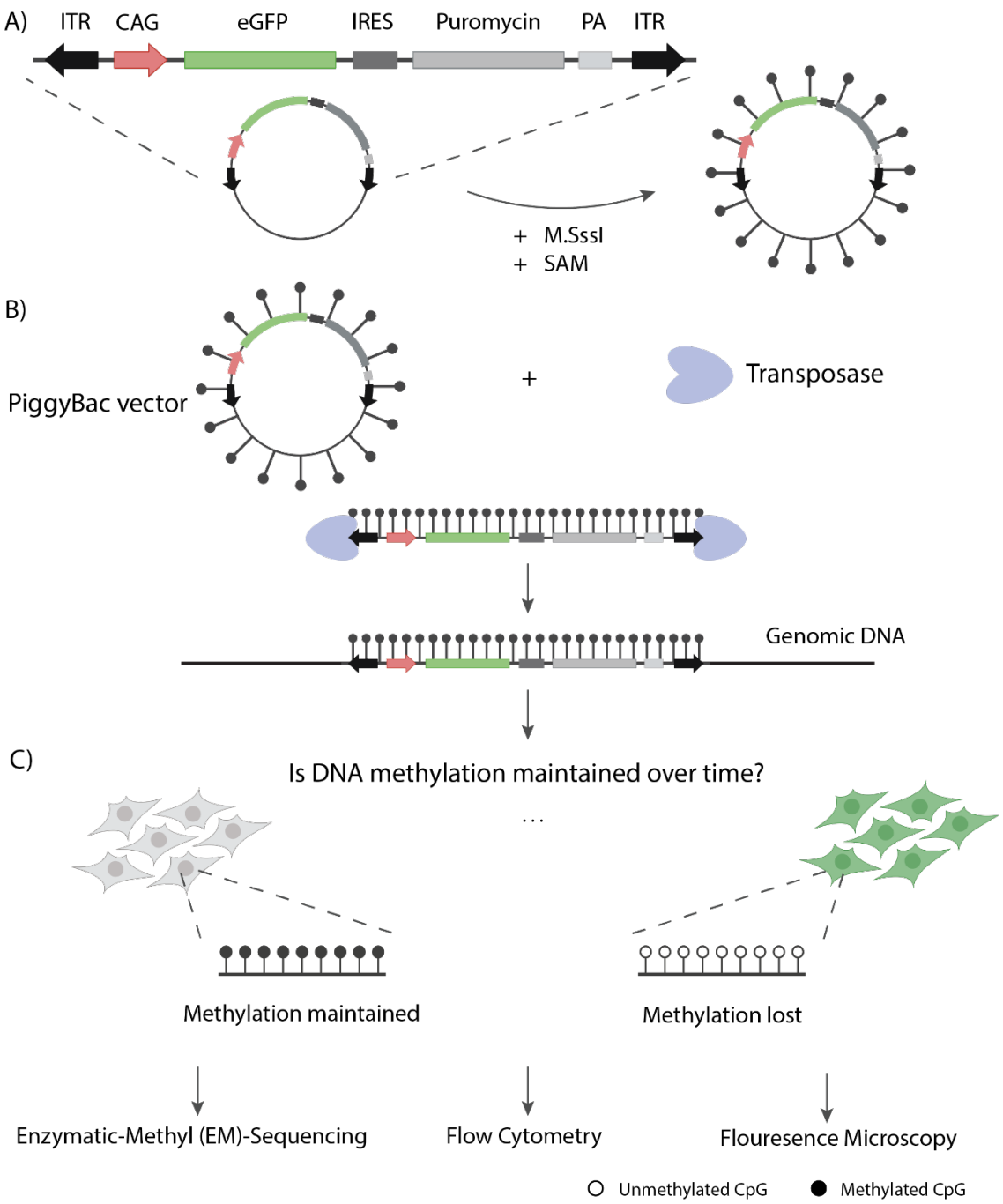


Figure 3-1: Schematic representation of methylation-sensitive reporter system experimental design. A) *The reporter construct contains the CAG synthetic promoter, eGFP reporter gene, an IRES, Puromycin resistance gene and a Poly-A tail, all contained within the piggyBac ITRs. The piggyBac vector was methylated using the M.SssI methyltransferase and methyl group donor (SAM).* ***B)*** *To deliver the reporter transgene to cells, the piggyBac vector was co-transfected alongside a separate vector expressing the transposase enzyme. The transposase mediates integration of the reporter into genomic DNA.* ***C)*** *Following integration, reporter expression was monitored by fluorescence microscopy and flow cytometry. Maintenance of DNA methylation was assessed by extracting gDNA from cells and performing targeted EM-sequencing of the PB_CAG_eGFP_IRES_Puro reporter construct.*

3.3 Evaluating piggyBac transposon integration efficiency

Having chosen the piggyBac transposon as the method of delivery for the reporter transgene, I first sought to characterise the integration efficiency of piggyBac transposition in HCT116 cells. I aimed to establish the integration efficiency of an unmodified piggyBac vector, in order to determine whether this rate would be suitable for delivery of the reporter construct. I tested piggyBac integration efficiency by colony formation assay, which is a method to assess the ability of a single cell to grow into a colony (Franken *et al.*, 2006). I transfected cells with the PB_CAG_eGFP_IRES_Puro construct and selected cells using Puromycin, therefore, only those that took up the vector should survive. It has been reported that, in some instances, plasmid backbone material can integrate spontaneously into the host cell genome (Troyanovsky *et al.*, 2016). This is a random process, not mediated by the transposase. Therefore, to investigate the rate of spontaneous integration, a separate transfection was carried out without the transposase-expressing vector. The mean integration rate of the piggyBac transposon when co-transfected with the transposase was estimated at 0.88% in HCT116 cells (Fig. 3.2A-B). In the absence of transposase, the rate of colony formation was reduced to 0.03%, indicating a significant ($p=1.75x10^{-2}$) reduction in integration efficiency without the transposase. The rate of piggyBac integration has been reported previously to average 10% but can vary by cell line from 0.7-14.5% (Wu *et al.*, 2006; Wilson, Coates and George, 2007). Previous tests in our lab have shown an ~1.6% piggyBac integration efficiency (Masalmeh, 2019), therefore, this rate is within expected range.

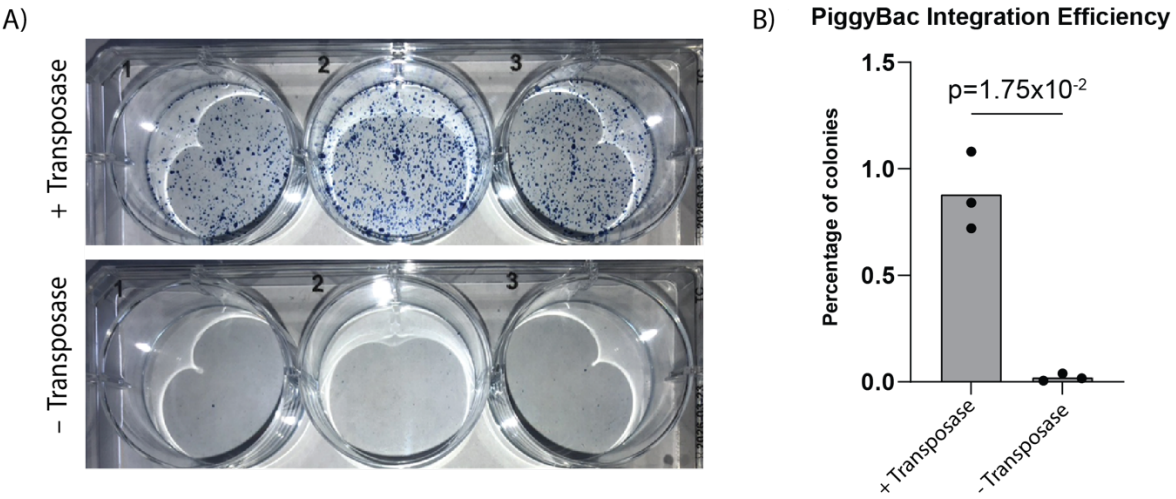


Figure 3-2: Evaluation of piggyBac transposon integration efficiency by colony formation assay. *A) Representative photographs of colony formation assay for HCT116 cells transfected*

with piggyBac transposon vector, with (top) and without (bottom) transposase. One biological replicate with three (n=3) technical replicates shown. B) Bar plot depicting the percentage of colonies as a function of transfected cells. Bars represent the mean of three biological replicates, points represent the mean of three technical replicates. P-value was determined by a paired t-test.

3.4 Unintegrated plasmid is lost within 3 weeks

Next, I sought to determine the time point from which I would be able assess stably integrated reporters. Following transfection, unintegrated plasmid persists in the cell and is gradually diluted out of the population by cell division (Akhtar *et al.*, 2014). For the purpose of this study, I was interested in only assessing reporter expression and DNA methylation from reporters stably integrated into the genome of cells. Therefore, it was necessary to establish the time required to dilute unintegrated plasmid out of the cell population. To do this, I performed separate transfections with and without transposase and assessed expression of eGFP on a weekly basis in HCT116 (Fig 3.3). Having established a low level of spontaneous vector integration without the presence of the transposase (Fig. 3.2), I anticipated that a low level of eGFP expression would persist permanently from these cells. The vast majority of eGFP positive cells were lost in cells transfected without transposase after 21 days (3 weeks). The mean proportion of eGFP positive cells in this population plateaued at 0.08% after 3 weeks (Fig 3.3B). Whereas, GFP expression persisted in cells transfected with the transposase to mediate stable vector integration. These data suggest that the unintegrated plasmid was been lost from cells, and the remaining eGFP expression in cells that were not transfected with transposase likely represented spontaneously integrated vector. Therefore, I determined that a minimum of three weeks passage was required to dilute unintegrated plasmid out of the population, in line with previous reports (Akhtar *et al.*, 2014; Masalmeh, 2019). Thus, subsequent measurements were taken from the 4-week timepoint onwards.

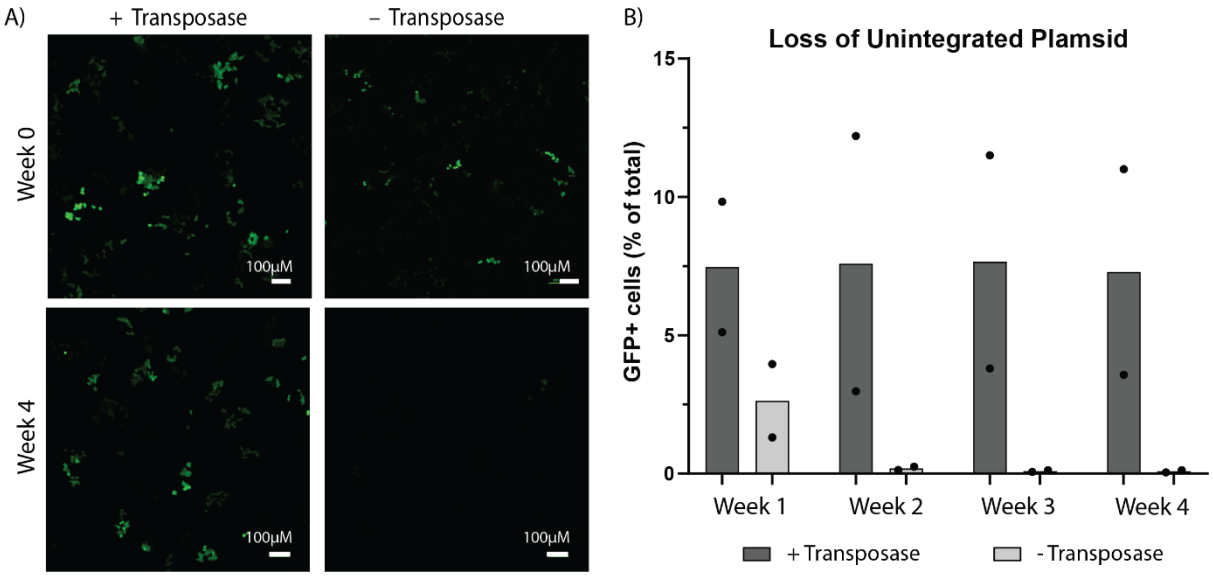


Figure 3-3: *Unintegrated plasmid is lost within 3 weeks.* ***A)*** *Representative fluorescence microscopy images of cells transfected with or without transposase, the week of transfection and after 4-weeks passage.* ***B)*** *The mean proportion of GFP-positive cells quantified by flow cytometry in cells transfected with or without transposase. Bars represent the mean of two biological replicates, with individual replicates shown as data points.*

3.5 All CpG sites in the piggyBac vector containing the eGFP reporter gene are methylated prior to integration

To generate a methylated PB_CAG_eGFP_IRES_Puro reporter construct, I chose to methylate the vector *in vitro* prior to integration into the genome. Purified plasmid DNA was methylated using the *M.SssI* methyltransferase, which methylates all cytosines in a CpG dinucleotide. I verified the efficacy of the methylation reaction through two separate methods. Firstly, I performed a diagnostic restriction digest on methylated and unmodified plasmid DNA. This reaction was carried out using the methylation-sensitive restriction endonuclease HpaII, which cleaves the internal C residue of its recognition sequence 5'-CCGG-3' but is blocked by the presence of a 5-methyl group. I used HpaII in combination with its isoschizomer MspI, which cleaves DNA irrespective of the presence of a methyl group (Waalwijk and Flavell, 1978). As shown in Figure 3.4B, no digestion by HpaII was observed in the vector that had been methylated, indicating that all CpG sites in a 5'-CCGG-3' context within the vector were methylated.

To further confirm methylation at all CpG sites within the reporter sequence, I assessed the methylation states of the unmodified and methylated vectors by EM-sequencing. Plasmid DNA was EM-converted, PCR-amplified using primers specific to the eGFP gene-body and sequenced. This technique allows detection at single-base resolution of any methylated cytosine in DNA. The mean percentage of methylated CpG sites was 0% in unmodified vector. The mean percentage of methylated CpG sites in the methylated vector was 100% (Fig. 3.4C-D). Therefore, I concluded that the vector sequence had been fully methylated and that the unmodified vector could be used in subsequent experiments as an unmethylated control.

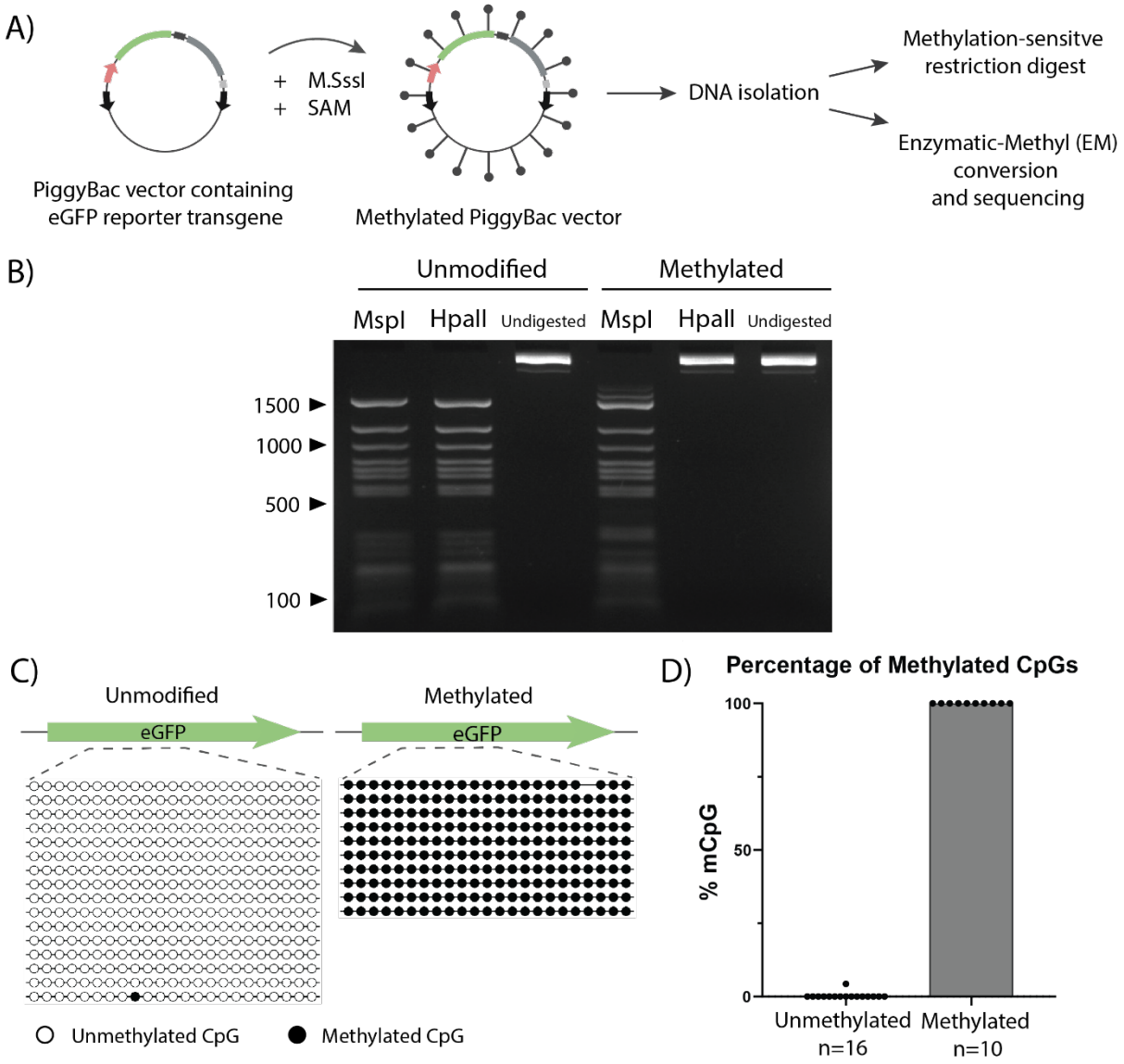


Figure 3-4: All CpG sites in the piggyBac vector containing the eGFP reporter gene are methylated prior to integration. A) *Schematic of experimental strategy for in vitro plasmid methylation and methylation state analysis.* ***B)*** *Representative image of EtBR stained agarose gel showing methylation-sensitive restriction endonuclease digestion of unmodified and*

methylated piggyBac plasmid. MspI cleaves 5'-CCGG-3' sites regardless of methylation state and its isoschizomer HpaII is blocked by methylation of the internal C reside. ***C)*** *Schematic representation of EM-sequencing results from the eGFP gene body region of the unmodified and methylated plasmids. Empty circle represents an unmethylated CpG site, and a filled circle represents a methylated CpG site. Each row represents sequencing from a single clone, corresponding to an individual plasmid. Analysis performed using QUMA (Kumaki, Oda and Okano, 2008).* ***D)*** *Violin plot showing the percentage of methylated CpG sites per individual clone from C) EM-sequencing.*

3.6 DNA methylation represses the CAG promoter and eGFP expression in HCT116 cells

In order to determine whether the chosen eGFP reporter construct would be suitable for use as a methylation-sensitive reporter, I next aimed to determine the impact of methylation on expression of the reporter gene. I aimed to assess expression of GFP from the vector itself prior to integration into the genome, therefore, for the purpose of this experiment the transposase-expressing vector was not co-transfected alongside the expression vector. I transiently transfected HCT116 cells with either the unmethylated or methylated pB_CAG_eGFP_IRES_Puro vector and assessed GFP expression by flow cytometry after 48-hours. The mean proportion of cells that were transfected with an unmethylated vector and GFP positive was 29.1% after 48 hours. In cells transfected with a methylated vector, the mean proportion of GFP positive cells was significantly lower (3.90%, $p=2.34x10^{-2}$, paired t-test) (Fig. 3.5).

I next assessed the fluorescence intensity distribution from cells containing a methylated or unmethylated vector (Fig 3.5B). In cells containing an unmethylated vector, the median fluorescence intensity (MFI) of GFP positive cells $1.32x10^{4}$. In cells containing a methylated vector, the MFI intensity was significantly lower ($1.73x10^{3}$, $p=5.00x10^{-2}$, paired t-test). These results indicate that cells containing the methylated pB_CAG_eGFP_IRES_Puro vector expressed a lower level of GFP fluorescence. I next sought to categorise high-GFP and low-GFP expression, as a method to distinguish between GFP expression from the methylated and unmethylated vector. To do this, I gated 'high-GFP' expressing cells in the flow cytometry data. The 'high-GFP' threshold was determined based on the peak of GFP fluorescence intensity in cells transiently transfected with the unmethylated vector. This gate was then applied to the population of cells containing the methylated pB_CAG_eGFP_IRES_Puro

plasmid. The proportion of high-GFP cells was significantly reduced ($p=3.67x10^{-2}$, paired t-test) when the vector was methylated. Approximately 15.02% of cells expressed a high level of GFP fluorescence when the vector was methylated. These data indicate that methylation of the reporter construct represses the CAG promoter, thereby repressing expression of the eGFP reporter gene. Methylation of a GFP reporter construct through *in-vitro* methylation using *M.SssI* has previously been reported to repress, but not completely silence, expression of GFP (Schübeler *et al.*, 2000; Lorincz *et al.*, 2002). Therefore, these results were in line with what was expected, based on published work. I concluded that this reporter construct could be repressed by DNA methylation and would be suitable for use as a methylation-sensitive reporter.

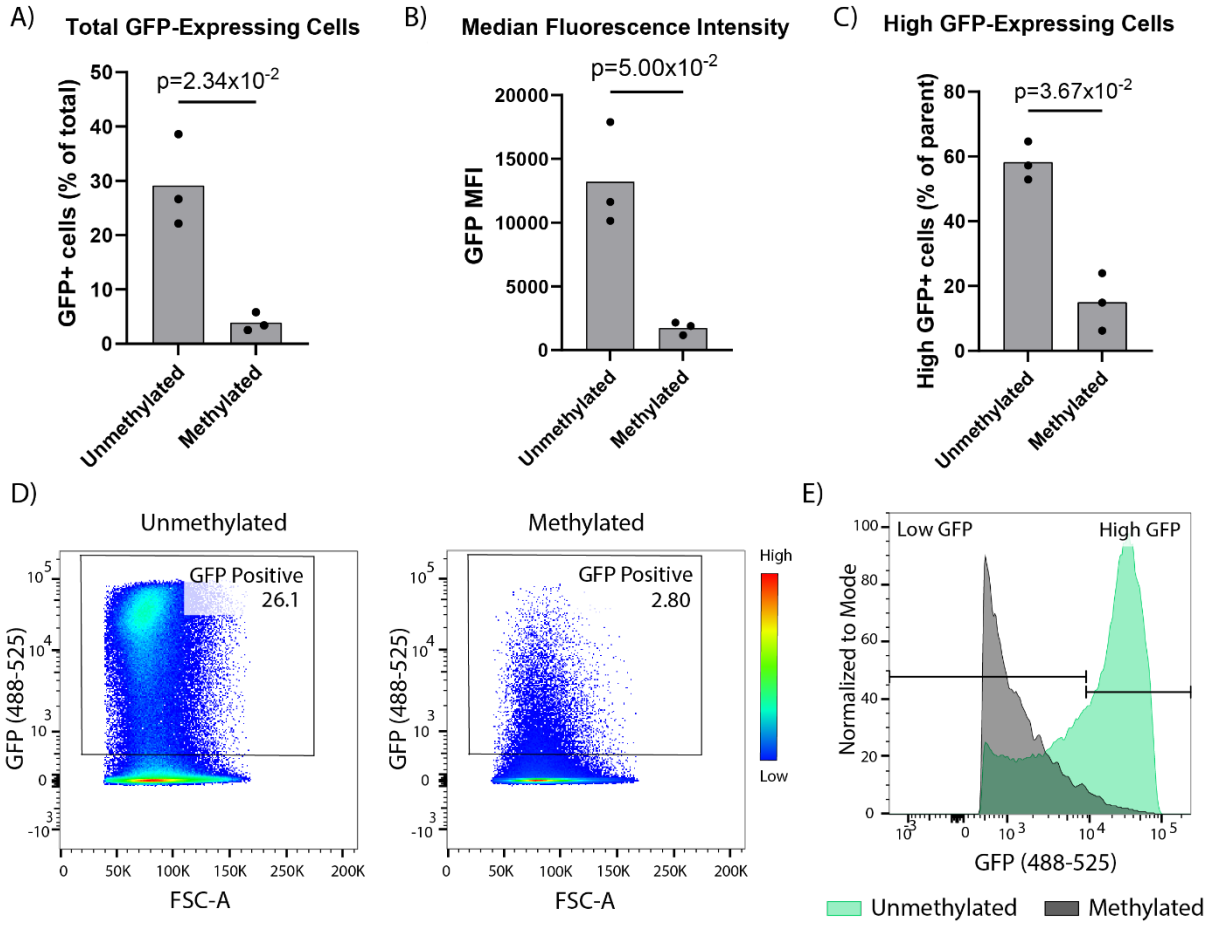


Figure 3-5: DNA methylation represses eGFP reporter gene expression. A) *Total GFP-positive cells quantified by flow cytometry.* ***B)*** *Median fluorescence intensity (MFI) of all GFP positive cells.* ***C)*** *Proportion of GFP positive cells (parent) that express a high level of GFP. Paired t-test. For all graphs, bars represent the mean of n=3 biological replicates. Each point represents the mean of two technical replicates. P-values determined by paired t-test.* ***D)*** *Representative flow cytometry density scatter plot from a single experimental replicate. GFP expression from*

HCT116 cells transfected with unmethylated or methylated vectors is shown. Total cells measured = 100,000. Each point represents a single cell, with the colour gradient key showing high (red) to low (blue) density. ***E)*** *Histogram showing fluorescence intensity distributions from flow cytometry analysis of cells transfected with unmethylated (green) and methylated (grey) vectors. GFP expression intensity is shown on the X-axis, and the count of cells is shown on the Y-axis, normalised to the mode.*

3.7 The piggyBac transposon system is suitable for use with methylated DNA

Next, I assessed whether the piggyBac transposition system was functional when the ITR sequence was methylated. It has been previously reported that methylation of the piggyBac transposon by *M.SssI* can reduce the efficiency of transposition 12-fold (Wang *et al.*, 2008). Additionally, a lower proportion of GFP positive cells were observed following transient transfection of HCT116 cells with a methylated eGFP piggyBac vector. This could correspond to repression of reporter expression or could indicate a reduced transfection efficiency that might coincide with a lower integration efficiency.

Therefore, I sought to confirm that DNA could be integrated into the genome though piggyBac transposition when the piggyBac vector is methylated. I investigated the integration efficiency of the piggyBac transposon when methylated, compared to the unmethylated plasmid. To determine the integration efficiency of the piggyBac transposase, I stably integrated unmethylated and methylated reporter constructs and expanded cells to 4 weeks before extracting gDNA. I then performed quantitative PCR (qPCR) targeting the eGFP gene body and used the BRCA2 region as an internal genomic control. Quantification of piggyBac integration by qPCR showed that piggyBac transposition was not inhibited by DNA methylation, in contrast to the findings of Wang *et al*. (Wang *et al.*, 2008). In fact, methylation of the piggyBac transposon appeared to enhance integration efficiency. The rate of methylated piggyBac integration was 4.89-fold higher than unmethylated integration in the first biological replicate of the experiment, and 7.71-fold higher in the second replicate, suggesting that this is a consistent effect. This result indicates that the piggyBac transposon system is suitable for use with methylated DNA.

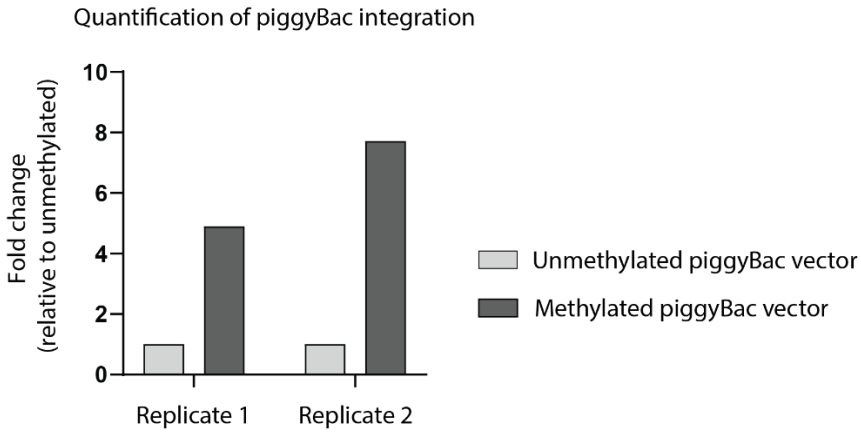


Figure 3-6: piggyBac integration efficiency is enhanced by DNA methylation. *Integration rate of the piggyBac transposon when the vector is methylated, compared to unmethylated. Copy number of piggyBac transposon insert sequence quantified by qPCR targeting the eGFP gene body, as a measure of integration efficiency. One (n=1) biological replicate, with n=3 technical replicates. Fold change calculated using the ΔΔCt method.*

3.7.1 PiggyBac integrations are obtained across PMD and non-PMD genomic regions

Having established that the piggyBac integration rate was not negatively impacted by methylation of the construct, I next investigated the insertion locations of the piggyBac transposon in HCT116 cells. I was interested in determining where the transposon integrated and assessing the types of genomic regions that the piggyBac transposon integrated into. For the purpose of this study, I was interested in obtaining reporter integrations across different genomic regions, particularly within PMD and non-PMD regions. It was also of interest to determine whether there were any notable differences in the rate of integration in different genomic regions when the piggyBac vector was methylated, compared to the unmethylated vector.

To obtain a comprehensive view of the distribution of piggyBac integrations, insertion sites were mapped using asymmetric nested PCR. This method involved amplifying the genomic region flanking the different piggyBac integration sites by PCR and then sequencing the amplicons. The genomic locations of insertion were determined by mapping the sequences to the Ensembl human genome database (Harrison *et al.*, 2024). A total of 78 unique integration sites were mapped for the unmethylated vector, and 66 unique sites were

determined for the methylated piggyBac integrations. Analysis of these integration sites revealed that integrations were obtained across all chromosomes for both the methylated and unmethylated vectors, except for chromosome 21 and the sex chromosomes (Fig. 3.6.2A). No integrations were mapped to chromosome 21 or the sex chromosome for either the methylated or unmethylated vector. Of the unique integration sites mapped, 12.82% of unmethylated piggyBac integrations were found within PMDs in HCT116 cells (Fig. 3.6.2B). With the methylated vector, 16.66% of integrations were mapped to PMD regions in HCT116 cells (Fig. 3.6.2C). A similar proportion of insert sites were mapped to PMDs, indicating that vector methylation did not have a substantial impact on the likelihood of the piggyBac transposon to integrate in PMDs. Thus, I was able to obtain reporter integrations across the genome and in both PMD and non-PMD regions.

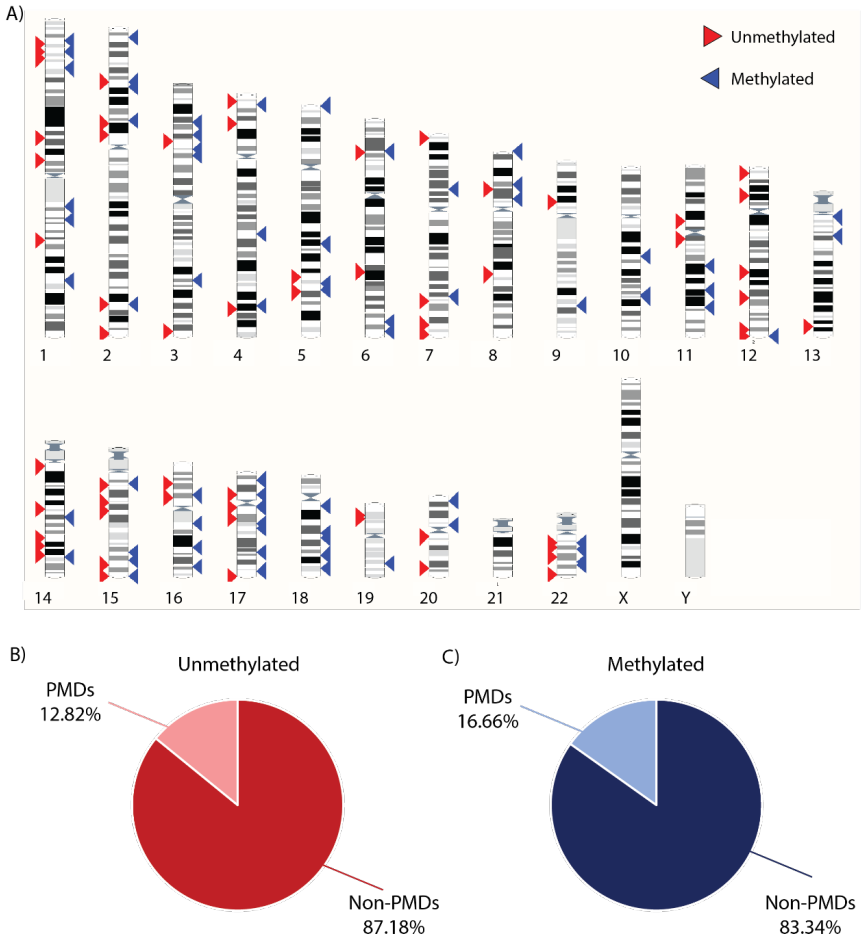


Figure 3-7-2: Genomic distribution of piggyBac integrations. A) *Chromosome ideogram showing the distribution of piggyBac insert sites from methylated and unmethylated vectors. Red arrows represent integration sites from the unmethylated vector, and blue arrows represent integration sites of the methylated vector. G/R bands: Dark bands (G-bands) represent heterochromatic regions and light bands (R-bands) represent euchromatic regions.*

Insertion sites mapped to human genome assembly GRCh38.p14. Figure generated using Ensembl v112 (Harrison et al.*, 2024).* ***B)*** *Pie chart showing the proportion of unmethylated piggyBac inserts in PMDs and non-PMDs.* ***C)*** *Pie chart showing the proportion of methylated piggyBac inserts in PMDs and non-PMDs.*

3.8 Reporter repression is maintained at stably integrated reporters

Having established that the reporter construct can be methylated, repressing reporter expression, and stably integrated into the genome, I next assessed maintenance of reporter repression following stable integration. I co-transfected HCT116 cells with the piggyBac expression vector and transposase-expressing vector to mediate stable integration into the genome. Cells were expanded for 4 weeks to dilute out unintegrated plasmid, and expression of GFP was assessed by fluorescence microscopy and flow cytometry.

First, I assessed expression of GFP from cells under the fluorescent microscope. HCT116 cells containing unmethylated reporters (HCT116-eGFP cells) expressed eGFP consistently across the 4-week time course and appeared bright green under the microscope. Cells containing methylated reporters (HCT116-eGFPme cells) were largely non-fluorescent at week 1, and this was unaltered at week 4 (Fig. 3.7A). I quantified expression of eGFP by flow cytometry. The distribution of fluorescence intensities from HCT116-eGFP and HCT116-eGFPme cells were distinct, with a clear separation of distributions depending on methylation state, and were comparable between week 1 and week 4 (Fig. 3.7B). At week 4, the MFI of HCT116-eGFP cells was $5.97x10^3$, whereas the MFI of HCT116-eGFPme cells reduced substantially to $4.98x10^2$ (Fig. 3.7C). Next, I assessed the proportion of HCT116-eGFPme that expressed a high level of GFP (Fig. 3.7D). In cells stably transfected with the reporter cassette, the 'high GFP' gate was set so that 95% of the events (single cells) in the unmethylated GFP-positive population fell within the gate. This gate was then applied to the HCT116-eGFPme population of cells. The mean proportion of HCT116-eGFPme cells that expressed a high level of eGFP at week 4 was 13.80%. This proportion of high GFP-positive cells was comparable to that of the transiently transfected vector, as show in Figure 3.5. Together, these data indicate that the reporter gene remained repressed following stable integration into the genome of HCT116 cells.

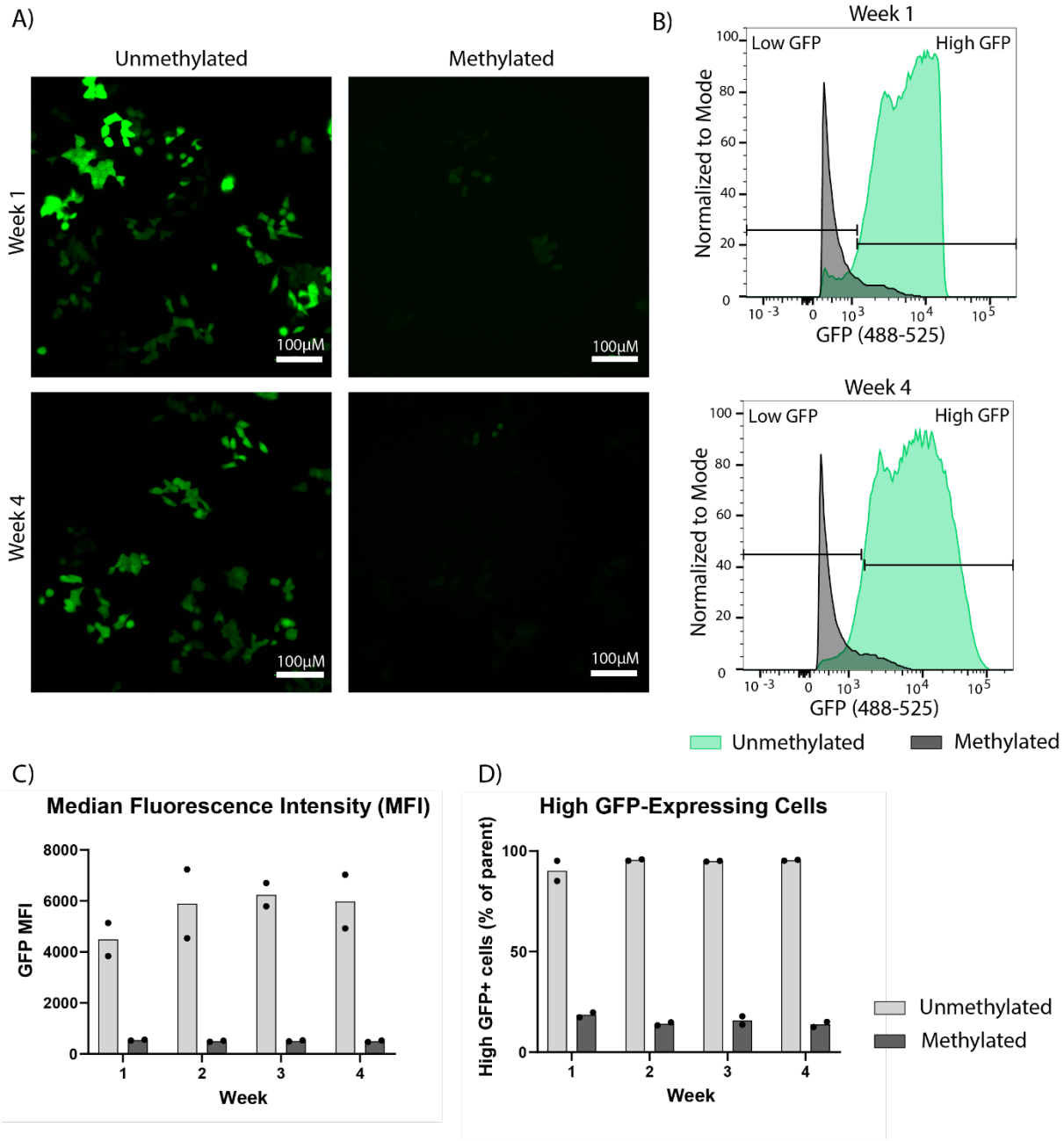


Figure 3-8: Reporter repression is maintained at stably integrated reporters in HCT116 cells. ***A)*** *Representative fluorescence microscopy images of HCTT16 cells containing either a methylated or unmethylated eGFP reporter at 1-week and 4-weeks post-transfection.* ***B)*** *Representative flow cytometry histograms showing fluorescence intensity distributions from methylated (grey) and unmethylated (green) vectors at week 1 and week 4. The gates for 'high' and 'low' GFP expression re indicated by horizontal lines.* ***C)*** *Median fluorescence intensity from HCT116 cells containing unmethylated (light grey) or methylated (dark grey) reporter constructs.* ***D)*** *Proportion of cells expressing a high level of GFP, from unmethylated (light grey) or methylated (dark grey) reporter constructs. Bars represent the mean of n=2 biological replicates. Data points show individual replicates. Flow cytometry data represents measurement from 100,000 cells. Example data shown from experimental replicate 2.*

3.9 DNA methylation is largely maintained at stably integrated reporters

In order to determine whether the expression of the reporter gene corresponded to the underlying methylation state of the construct, I next assessed whether the methylation states of the integrated reporters were maintained following stable integration into the genome of HCT116 cells.

3.9.1 Reporter methylation is largely maintained following stable integration

In the first instance, I conducted an initial analysis of the methylation state of stably integrated reporters at the week 4 time point. To do this, I isolated genomic DNA from cells and performed BS-conversion followed by PCR targeting the eGFP gene body, and subsequently sequenced these regions by Sanger sequencing. Following stable integration, the unmethylated vector remained largely unmethylated, with a median 0% methylated CpG sites (Fig. 3.8.1). A very low level of *de novo* methylation was observed at a subset of reporter sequences. The median percentage of methylated CpG sites in methylated reporters was 100%, indicating that the methylated state was largely maintained following stable integration in HCT116 cells (Fig. 3.8.1). One clone was found to have become partially demethylated. Together, these data indicate that the methylation states of synthetic reporters are recapitulated following stable integration into the genome, mediated by the piggyBac transposase. However, due to the low number of reporters able to be sequenced through this method, these data may provide a limited insight into maintenance of methylation across different reporter integrations.

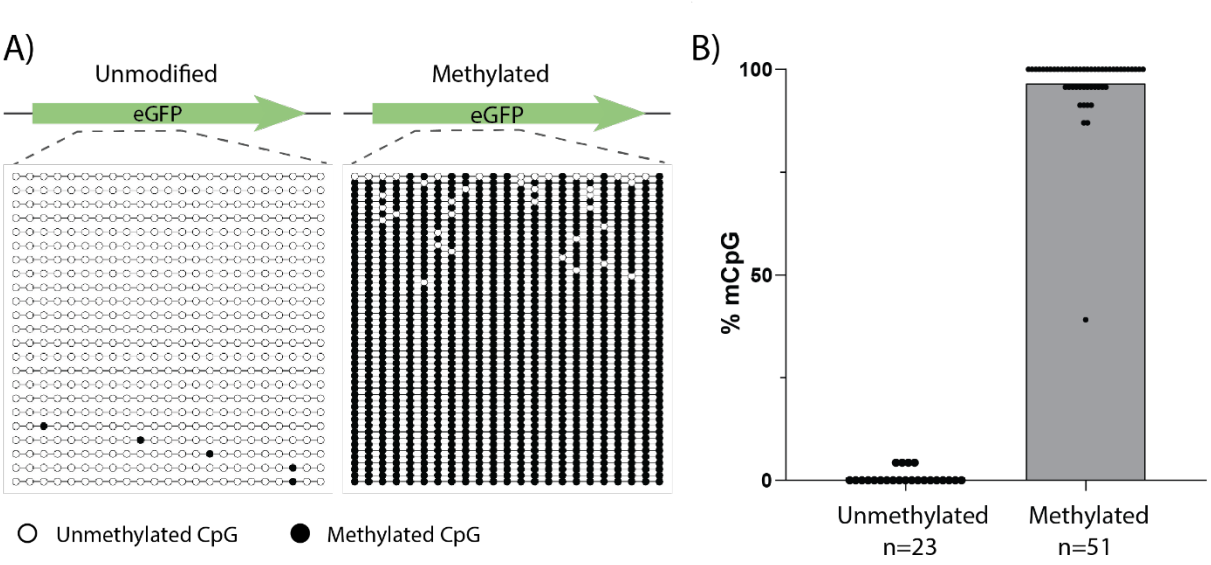


Figure 3-9-1: Reporter methylation is largely maintained following stable integration. A) *Schematic representation of DNA methylation at individual CpG sites within the BS-sequenced region of the eGFP gene body. Filled circles represent a methylated CpG and an unfilled circle represents an unmethylated CpG. Each row represents an individual clone sequences, corresponding to the sequence of a single integrated reporter. Figure generated using QUMA (Kumaki, Oda and Okano, 2008)* ***B)*** *Bar plot showing the percentage of methylated CpGs (%mCpG) within eGFP gene body region, as shown in A), for the methylated and unmethylated vectors. Example data shown from experimental replicate 2.*

3.9.2 The CMV enhancer is more likely to lose methylation than the eGFP gene body

In order to gain a more comprehensive insight into the methylation levels across different integrated the reporters, and to assess methylation levels at different regions within the construct, I next conducted deep amplicon sequencing targeting the CMV enhancer and eGFP gene body. I designed primers compatible with the Illumina sequencing platform with unique molecular identifiers (UMIs), targeting the CMV enhancer portion of the CAG promoter and the eGFP gene body. Genomic DNA was EM-converted, and the target regions of interest were amplified by PCR. I pooled CMV and eGFP amplicons from the same genotype and amplified Illumina sequencing libraries by PCR (see Methods 2.7.4). This method was adapted from Taglini *et al*. (Taglini *et al.*, 2024). High-throughput sequencing was performed on the Illumina iSeq platform. During analysis, I removed PCR duplicates using the UMI sequence, to ensure that each sequencing read came from a unique reporter sequence. I then calculated the

number of CpG sites within the sequenced region, the number of methylated CpG sites and the percentage of methylated CpG sites (%mCpG) for each individual sequencing read.

I assessed the median percentage of methylated CpG sites at the CMV and eGFP regions of the reporter, in HCT116-eGFP and HCT116-eGFPme cells. In HCT116-eGFP cells, the median methylation level at the CMV enhancer and eGFP gene body was 0% (Fig. 3.8.2A). The frequency distribution of %mCpG values for the CMV and eGFP regions are shown in Figure 3.8.2B, and show that the majority of reporters were <10% methylated. These results are in line with the previous sequencing data (Fig. 3.8.1) and demonstrate that the unmethylated vector remained largely unmethylated following stable integration in HCT116 cells. In HCT116-eGFPme cells, the median methylation level at the CMV enhancer was 80%, and at the eGFP gene body it was 100% (Fig. 3.8.2A). The frequency distribution of %mCpG values for the CMV and eGFP regions in HCT116-eGFPme cells are shown in Figure 3.8.2D. The enhancer was more likely to lose methylation than the gene body, in line with results previously reported by Schübeler *et al.* (2000) (Schübeler *et al.*, 2000). The proportion of reporters that lost methylation at the enhancer corresponds approximately to the proportion of cells expressing a high level of GFP at week 4 (see Fig. 3.7D). These results indicate that expression of the eGFP reporter gene corresponds to the underlying level of DNA methylation at the CMV.

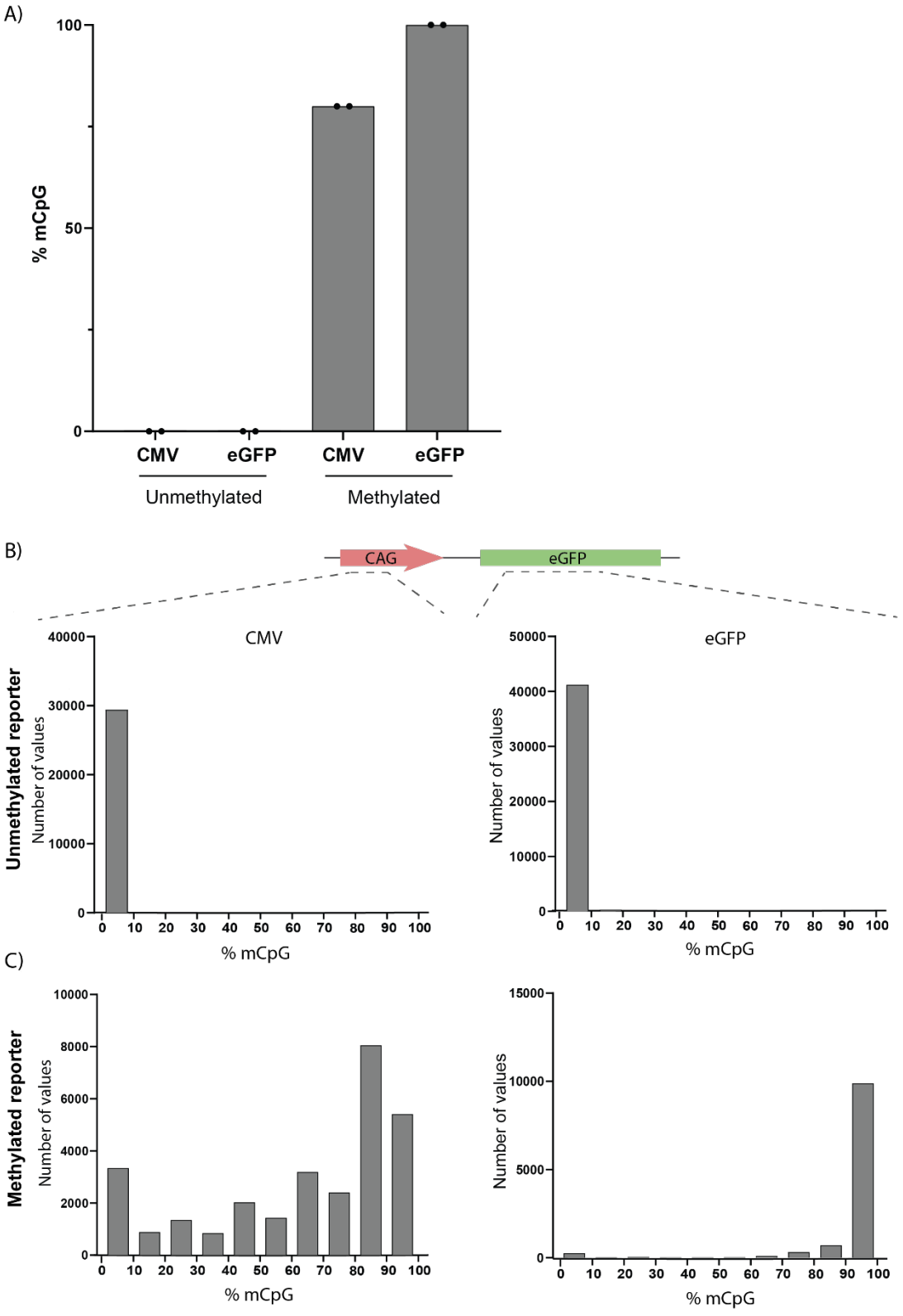


Figure 3-8-2: The CMV enhancer is more likely to lose methylation than the eGFP gene body in HCT116 cells. A) *Bar plot showing the median percentage of methylated CpG sites (%mCpG) at the CMV and eGFP gene body regions of the reporter. Methylation levels assessed in HCT116 cells containing methylated or unmethylated reporters. Bars show the median level of methylation. Data points represent the median %mCpG from n=2 biological replicates shown. Total number of sequencing reads = Unmethylated replicate 1: CMV, n=33,500, eGFP,*

n=42,063; Unmethylated replicate 2: CMV, n=29,565, eGFP n=41,628; Methylated replicate 1: CMV, n=39,106, eGFP, n= 42,062; Methylated replicate 2: CMV, n=36,174, eGFP, n=46,062 ***B)*** *Representative histogram showing the frequency distribution of %mCpG values for the CMV and eGFP regions of reporters in HCT116-eGFP cells.* ***C)*** *Representative histogram showing the frequency distribution of %mCpG values for the CMV and eGFP regions of reporters in HCT116-eGFPme cells. Example data shown from experimental replicate 2.*

3.10 Reporter repression, but not methylation, is maintained in RKO cells.

3.10.1 Reporter repression is maintained in RKO cells

Next, I asked whether a methylated eGFP reporter could be integrated into an alternative cancer cell line, and whether the reporter expression and methylation state would be maintained, consistent with the results in HCT116 cells. To do this, I stably integrated the pB_CAG_eGFP_IRES_Puro reporter construct into an alternative colorectal cancer cell line, RKO cells. Cells were expanded to 4 weeks to remove unintegrated plasmid. At week 4, eGFP expression was analysed by flow cytometry and gDNA was extracted from cells to assess methylation levels at the reporter construct by EM-sequencing. At week 4, the proportion of cells expressing a high level of GFP was significantly reduced ($p=6.01x10^{-4}$, paired t-test) when the reporter construct was methylated (Fig. 3.9). These results are in line with the previous data from HCT116 cells, indicating that the methylated reporter construct was repressed following stable integration into RKO cells.

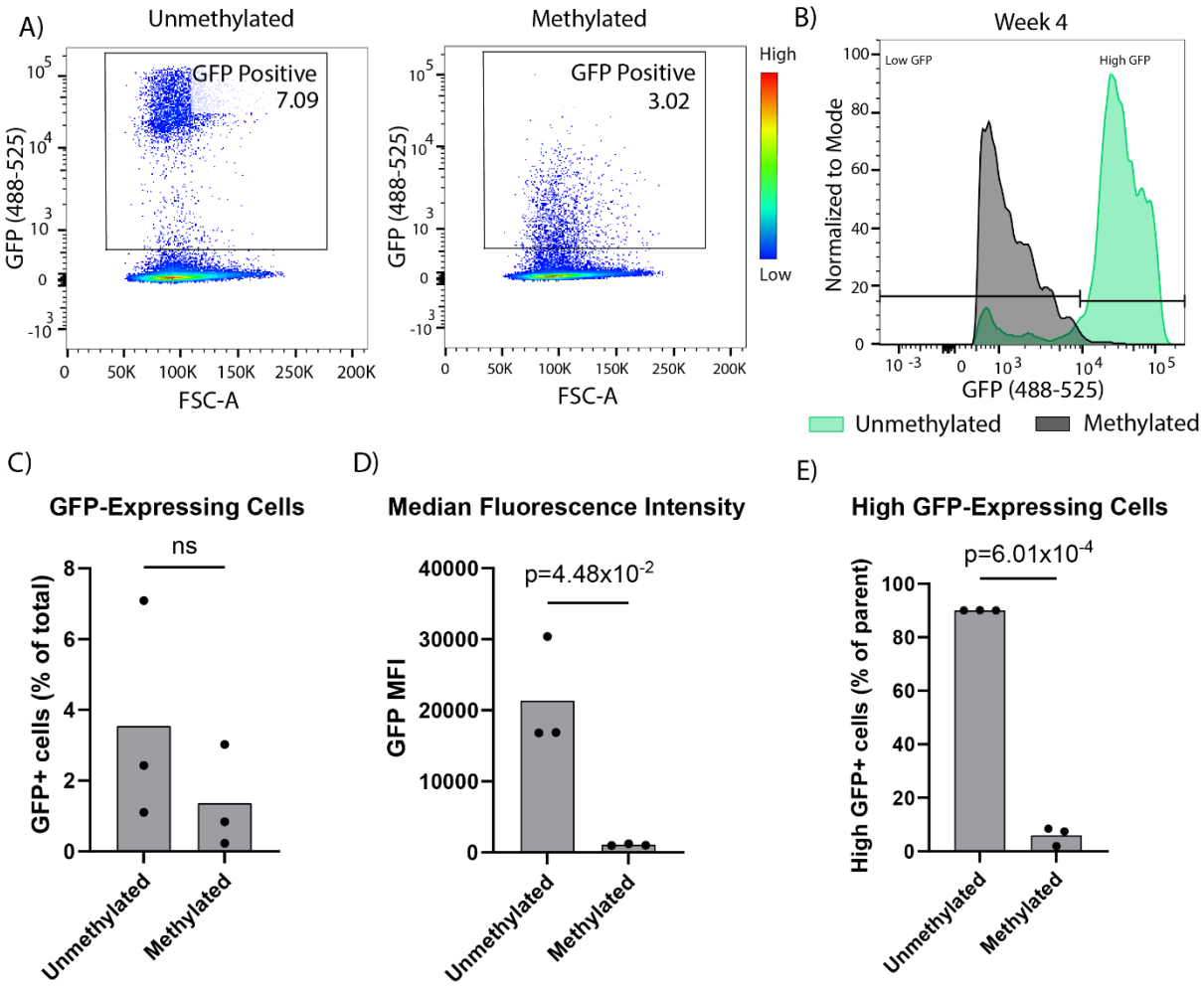


Figure 3-10: Reporter repression is maintained following stable integration into RKO cells. ***A)*** *Representative flow cytometry density scatter plot from a single experimental replicate. GFP expression from RKO cells transfected with unmethylated or methylated vectors are shown. Total cells measured = 100,000. Each point represents a single cell, with the colour gradient key showing high (red) to low (blue) density.* ***B)*** *Representative histogram showing fluorescence intensity distributions from RKO cells containing either unmethylated (green) or methylated (grey) reporter constructs. GFP expression intensity is shown on the X-axis, and the count of cells is shown on the Y-axis, normalised to the mode.* ***C)*** *Total proportion of GFP positive cells.* ***D)*** *Median fluorescence intensity of cells.* ***E)*** *Proportion of GFP positive cells expressing a high level of GFP. For all graphs, bars represent the mean and each data point represents the mean of three biological replicates (n=3). P-values determined by paired t-test.*

3.10.2 Reporter methylation is not maintained in RKO cells

I assessed methylation levels at the reporter construct in RKO cells containing either methylated or unmethylated reporters. I conducted deep amplicon EM-sequencing targeting the CMV enhancer and eGFP gene body, as described previously. In RKO cells containing unmethylated reporters, the median methylation level at the CMV enhancer and eGFP gene body was 0% (Fig. 3.10.2A). The frequency distribution of %mCpG values for the CMV and eGFP regions are shown in Figure 3.10.2B, and show that the majority of reporters were <10% methylated, but a notable proportion of reads displayed a higher level of methylation. These results indicate that the reporter construct was more likely to be targeted for *de novo* methylation in RKO cells than in HCT116 cells. However, overall, the unmethylated vector remained largely unmethylated following stable integration in RKO cells. In RKO cells containing methylated reporters, the median methylation level at the CMV enhancer and eGFP gene body it was 0% (Fig. 3.10.2A). The frequency distribution of %mCpG values for the CMV and eGFP regions in RKO-eGFPme cells show that the majority of reads were fully demethylated (Fig. 3.10.2D). Some sequencing reads from the CM enhancer retained a higher methylation level, whereas the vast majority of reads from the eGFP gene body were <10% methylated. Together, these results indicate that methylation was not faithfully maintained at the pB_CAG_eGFP_IRES_Puro reporter construct in RKO cells. These findings are in contrast to the reporter expression data (see Fig. 3.9), which showed that the reporter remained repressed in RKO cells.

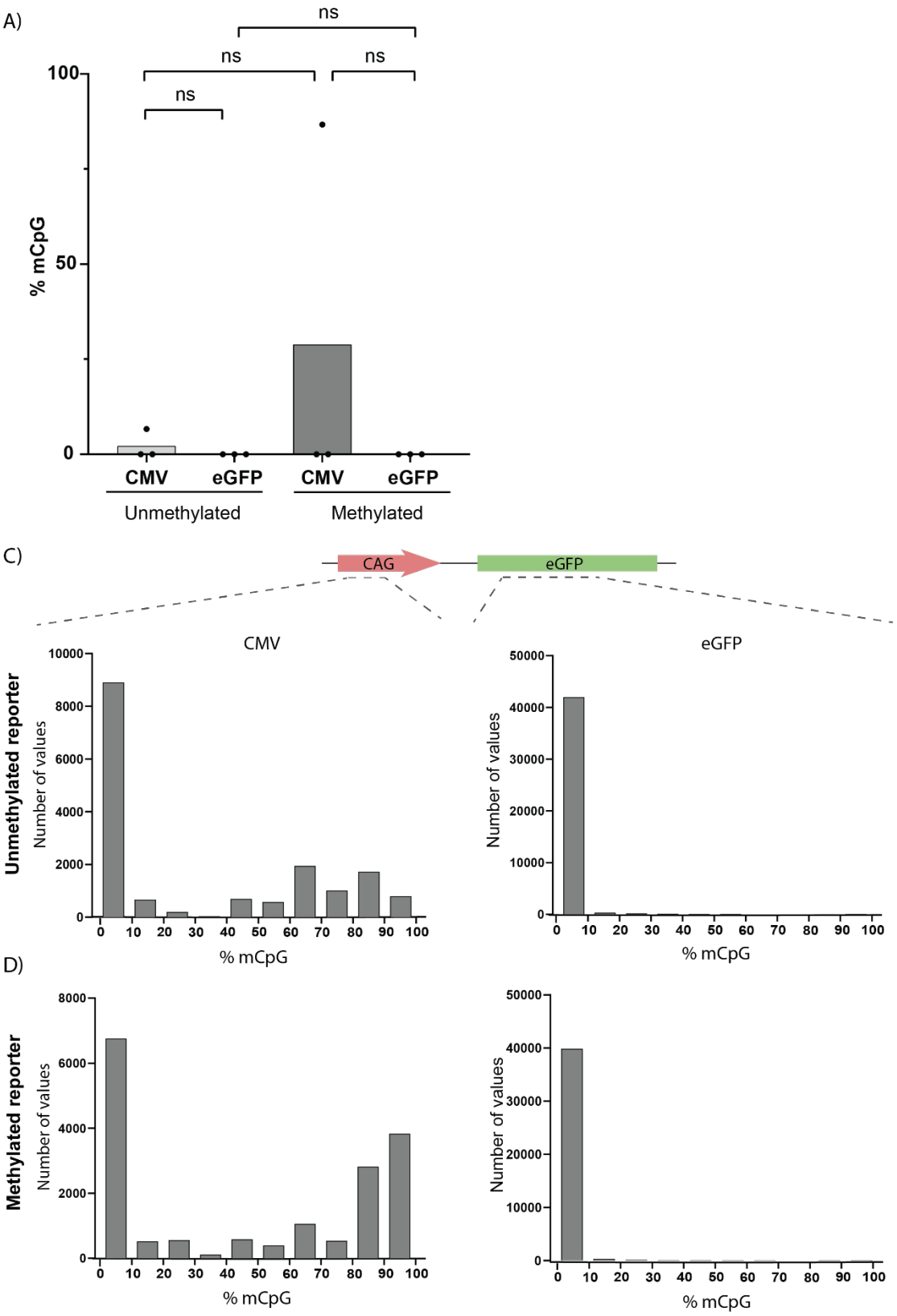


Figure 3-11: Reporter methylation is not maintained in RKO cells. A) *Bar plot showing the median percentage of methylated CpG sites (%mCpG) at the CMV and eGFP gene body regions of the reporter. Methylation levels were measured by EM-sequencing in RKO cells containing methylated or unmethylated reporters. Bars show the median level of methylation. Data points represent the median %mCpG from n=2 biological replicates shown. Total number of*

sequencing reads = Unmethylated replicate 1: CMV, n=17,093, eGFP, n=43,016; Unmethylated replicate 2: CMV, n=21,779, eGFP n=50,183; Unmethylated replicate 3: CMV, n=36,031, eGFP, n=43,996; Methylated replicate 1: CMV, n=23,743, eGFP, n= 40,717; Methylated replicate 2: CMV, n=33,125, eGFP, n=38,673; Methylated replicate 3: CMV, n=34,925 , eGFP, n=46,479 ***B)*** *Representative histogram showing the frequency distribution of %mCpG values for the CMV and eGFP regions of reporters in RKO-eGFP cells.* ***C)*** *Representative histogram showing the frequency distribution of %mCpG values for the CMV and eGFP regions of reporters in RKO-eGFPme cells. Example data shown from experimental replicate 2.*

3.11 Reporter repression is maintained over prolonged periods in HCT116 cells.

I next sought to utilise the reporter system to investigate the fidelity of DNA methylation maintenance in HCT116 cells over prolonged culture periods. The loss of DNA methylation in cancer has been proposed to be a result of incomplete maintenance of methylation, leading to progressive loss over time. Therefore, I hypothesised that I would observe gradual activation of the reporter gene if the underlying methylation of the reporter cassette was not maintained in HCT116 cells during prolonged culture. I utilised the HCT116-eGFP and HCT116-eGFPme cells previously generated (section 3.8) and maintained these cells in culture for a further 4 weeks, for a total of 8 weeks in culture. HCT116 cells have a doubling time of ~18 hours, therefore this time frame corresponds to approximately 75 cell divisions. As shown in Figure 3.11, the reporter construct remained largely repressed at 8 weeks post-integration. The MFI intensity of HCT116-eGFPme cells, and the proportion of cells expressing a high level of GFP, remained largely consistent across the time course.

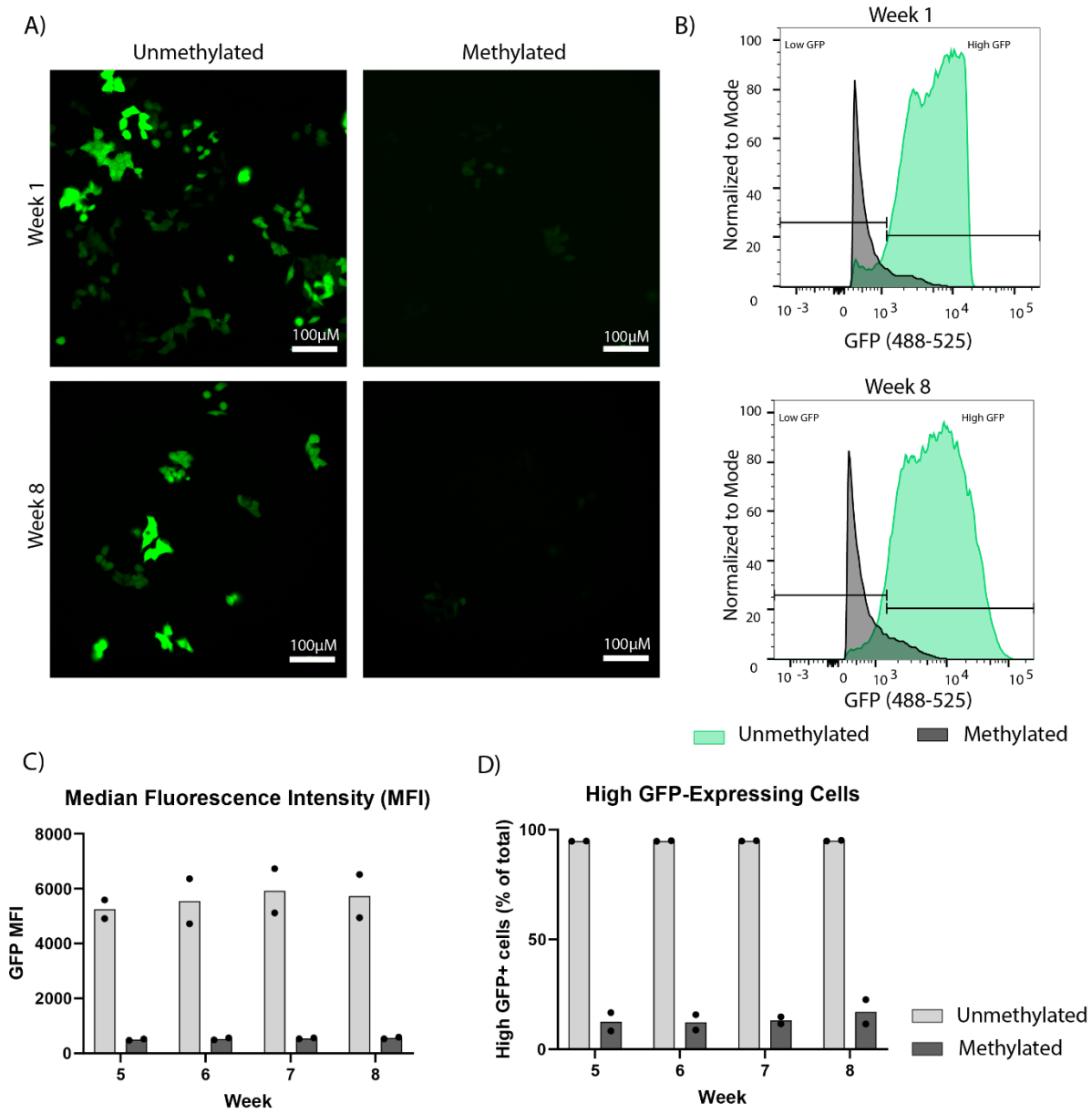


Figure 3-12: Reporter repression is maintained over prolonged periods in HCT116 cells. A) *Representative fluorescence microscopy images of HCTT16 cells containing either a methylated or unmethylated eGFP reporter at week 1 and week 8.* ***B)*** *Representative flow cytometry histograms showing fluorescence intensity distributions from methylated (grey) and unmethylated (green) vectors at week 1 and week 8. The gates for 'high' and 'low' GFP expression re indicated by horizontal lines. Week 1 data is repeated from Figure 3.7.* ***C)*** *Median fluorescence intensity from HCT116 cells containing unmethylated (light grey) or methylated (dark grey) reporter constructs.* ***D)*** *Proportion of cells expressing a high level of GFP, from unmethylated (light grey) or methylated (dark grey) reporter constructs. Bars represent the mean of n=2 biological replicates. Data points show individual replicates. Flow cytometry data represents measurement from 100,000 cells. Example data shown from experimental replicate 2.*

3.12 Reporter methylation is largely maintained over prolonged periods in HCT116 cells.

To assess methylation levels at integrated reporters at week 8, I isolated genomic DNA from HCT116-eGFP and HCT116-eGFPme cells and analysed reporter methylation through two methods. First, I conducted BS-PCR and Sanger sequencing at the eGFP gene body region of the reporter construct. As shown in Figure 3.12, the methylated reporter remained largely methylated at week 8, with a median %mCpG of 100%. The unmethylated reporter was largely unmethylated, with a median %mCpG of 0%.

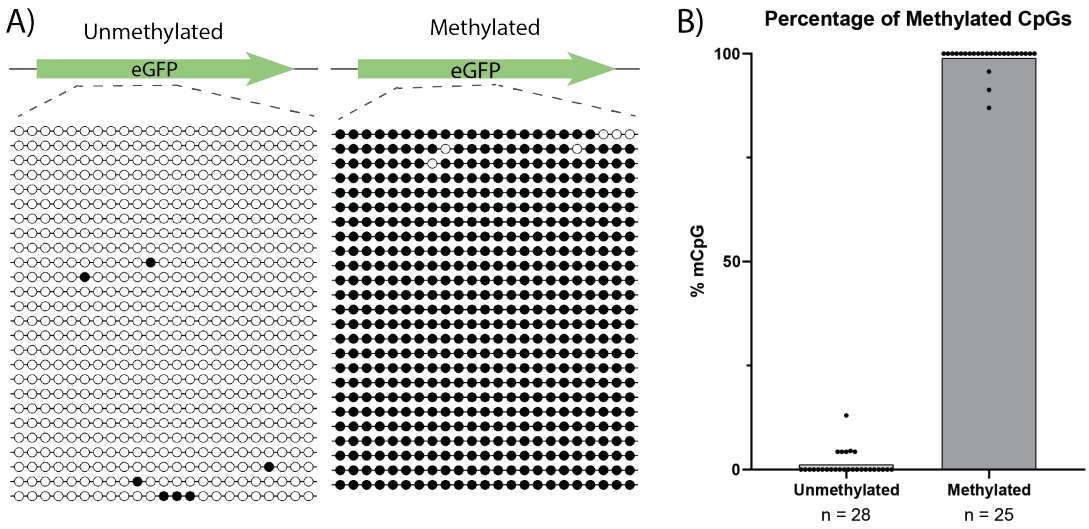


Figure 3-13: Reporter methylation is maintained at week 8. A) *Schematic representation of DNA methylation at individual CpG sites within the BS-sequenced region of the eGFP gene body in HCT116-eGFPme cells at week 8. Filled circles represent a methylated CpG and an unfilled circle represents an unmethylated CpG. Each row represents an individual clone sequences, corresponding to the sequence of a single integrated reporter. Figure generated using QUMA (Kumaki, Oda and Okano, 2008)* ***B)*** *Bar plot showing the percentage of methylated CpGs (%mCpG) within eGFP gene body region, as shown in A), for the methylated and unmethylated vectors. Example data shown from experimental replicate 2.*

To obtain a more detailed insight into the methylation levels across different reporter integrations, I conducted deep amplicon EM-sequencing targeting the CMV enhancer and eGFP gene body. The median %mCpG in the unmethylated control cells was 0% at week 8, at both regions of the reporter that were sequenced. In HCT116-eGFPme cells, the median %mCpG at the CMV enhancer was 70%, and at the eGFP gene body it was 100% (Fig. 3.13). A reduction of ~10% mCpG was observed at the CMV enhancer, while the methylation at the

eGFP gene body was unchanged, between week 4 and week 8. The frequency distribution of %mCpG values for the CMV and eGFP regions in HCT116-eGFP and HCT116-eGFPme are shown. Together, these data indicate that methylation patterns are faithfully maintained at synthetic reporters in HCT116 over consecutive cell divisions.

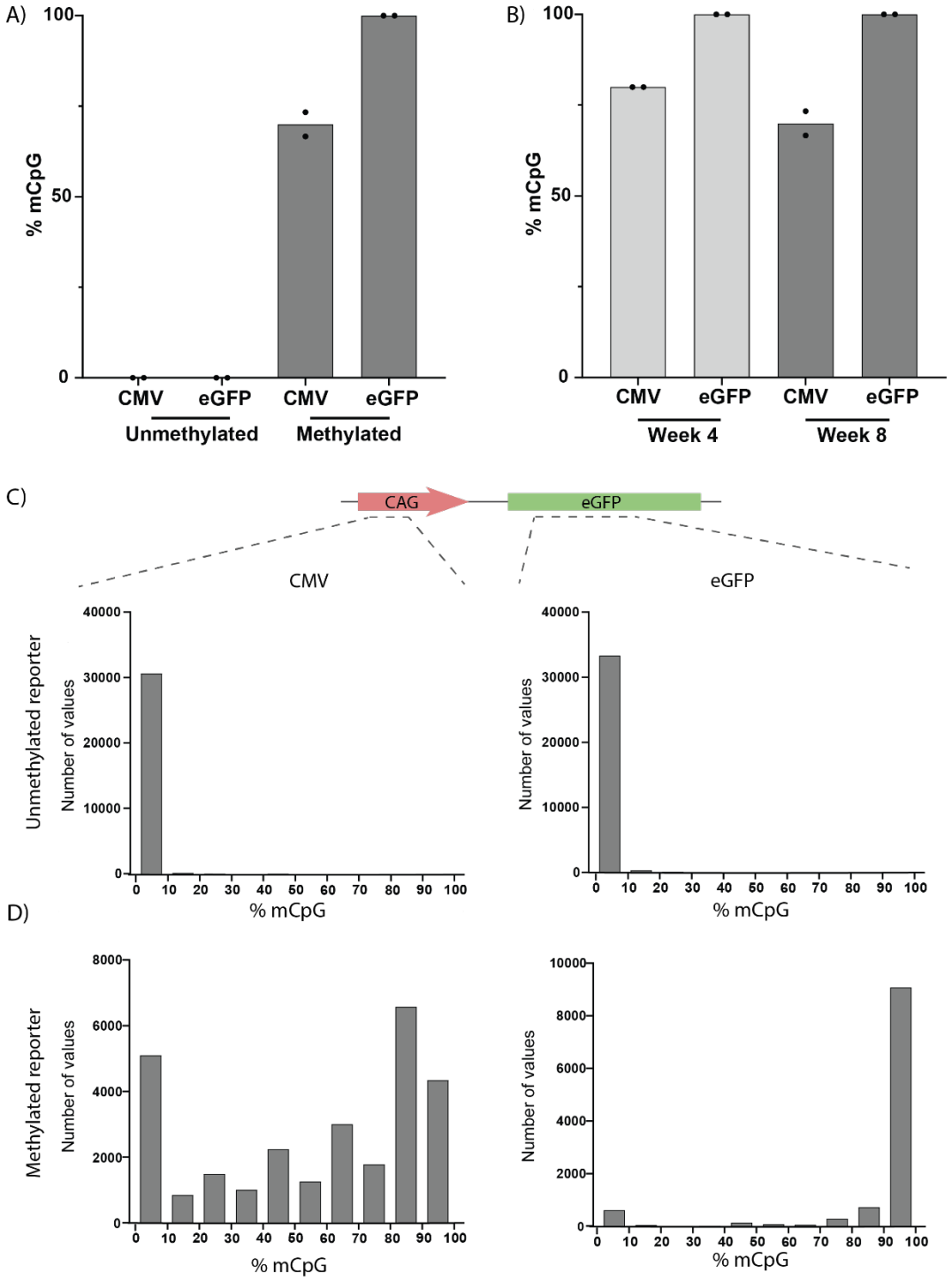


Figure 3-14: Reporter methylation is largely maintained over prolonged periods in HCT116 cells. A) *Bar plot showing the median percentage of methylated CpG sites (%mCpG) at the CMV and eGFP gene body regions of the reporter at week 8. Methylation levels assessed in*

HCT116 cells containing methylated or unmethylated reporters. Bars show the median level of methylation. Data points represent the median %mCpG from n=2 biological replicates shown. Total number of sequencing reads: Unmethylated replicate 1: CMV, n=29,844, eGFP, n=28,702; Unmethylated replicate 2: CMV, n=30,789, eGFP n=34,081; Methylated replicate 1: CMV, n=25,266, eGFP, n= 39,804; Methylated replicate 2: CMV, n=32,981, eGFP, n=42,811. ***B)*** *Representative histogram showing the frequency distribution of %mCpG values for the CMV and eGFP regions of reporters in HCT116-eGFP cells.* ***C)*** *Representative histogram showing the frequency distribution of %mCpG values for the CMV and eGFP regions of reporters in HCT116-eGFPme cells. Example data shown from experimental replicate 2.*

3.13 Discussion

In this chapter, I established a methylation-sensitive reporter system in the human colorectal cancer cell line, HCT116. I demonstrated that methylation of the CAG promoter represses its activity in HCT116 cells, resulting in repression of eGFP expression. Repression of the reporter, along with its underlying methylation state, were faithfully recapitulated following stable integration into the human genome. Expression of the eGFP reporter gene was shown to accurately correspond to its underlying methylation state in HCT116 cells. Thus, this reporter system can be utilised to track the fidelity of DNA methylation maintenance *in vitro*.

Reporters were stably integrated into HCT116 and RKO cells using the piggyBac transposition system. A minimum of three weeks passage was required to dilute out unintegrated plasmid from the cell population, in line with previous reports (Akhtar *et al.*, 2014; Masalmeh, 2019). This was an essential step, to ensure that downstream analysis of the reporter corresponded only to those that were stably integrated into the genome. Following stable integration, the initial methylation state of the reporter was recapitulated in HCT116 cells. This agrees with previous work, which demonstrated that the piggyBac transposon could be used to integrate methylated CGIs that maintained the expected methylation state following integration (Stelzer *et al.*, 2015). Furthermore, the rate of piggyBac integration was enhanced when the vector was methylated, compared to the unmethylated vector. This is in contrast to previous work that indicated that methylation can inhibit piggyBac transposition (Wang *et al.*, 2008). However, DNA methylation has been previously reported to enhance the transposition of the *Sleeping Beauty* transposon (Yusa, Takeda and Horie, 2004; Park *et al.*, 2005). Together, my

data indicates that the piggyBac transposon system is not impeded by DNA methylation and can be used to integrate methylated DNA into the human genome.

I mapped the insertion sites of the piggyBac transposon, when the vector was unmethylated and methylated, and identified integrations distributed throughout the genome under both conditions. The distribution of integration sites appeared to be largely random. Integrations were identified on all chromosomes except chromosomes 21, X and Y. HCT116 cells are male, and thus contain a single copy of the X and Y chromosomes. The lack of integrations found on these chromosomes may be due to the relatively small number of integration sites that were mapped. Previous work has shown that the piggyBac transposon can integrate on any chromosome (Wang *et al.*, 2008; Galvan *et al.*, 2009; Liang *et al.*, 2009; Meir *et al.*, 2011). Furthermore, the piggyBac transposon has been shown to mediate thousands of integrations in a single pool of transfected cells (Akhtar *et al.*, 2013, 2014). Therefore, these mapped integrations likely represent only a small fraction of the total integrations obtained. A higher throughput sequencing method would allow us to obtain a more complete picture of the piggyBac integration landscape.

Methylation of the CAG promoter construct was shown to repress, but not completely silence, its activity in HCT116 and RKO cells. This finding is in line with previous work in MEL cells, which showed that eGFP expression was reduced, but still above background fluorescence levels, when the reporter cassette was methylated *in vitro* prior to integration into cells (Schübeler *et al.*, 2000; Lorincz *et al.*, 2002). The expression profiles of the methylated and unmethylated vectors were clearly distinct, allowing us to easily differentiate between them. Expression of the reporter gene was shown to correspond to its underlying methylation state in HCT116 cells, demonstrating that this system accurately reports DNA methylation levels when stably integrated into the human genome in this cell line. However, this was not recapitulated in RKO cells. The reporter was shown to become demethylated in RKO cells, but, interestingly, repression of the reporter was maintained. It may be that alternative repressive mechanisms, such as repressive histone marks and the Polycomb repressive complex, are recruited to the CAG promoter CGI and maintain its repression in the absence of DNA methylation (Reddington *et al.*, 2013; Barrasa *et al.*, 2023). This would be interesting to explore further, perhaps by investigating the type of histone modifications present at integrated reporters in RKO cells.

DNA methylation was shown to be largely maintained following stable integration into the genome of HCT116 cells. The CMV enhancer region was more likely to lose methylation than the eGFP gene body. These results are in line with previously reported findings from Schübeler *et al.* (2000) (Schübeler *et al.*, 2000). The 2000 study by Schübeler *et al.* demonstrated that the HS2 fragment of the human β-globin locus control region, the enhancer of the report construct used in this study, was more likely to become demethylated following integration into MEL cells than the human β-globin promoter or GFP gene-body. It was not possible to design primers to amplify the human β-globin promoter construct used in this study. The sequence downstream of the CAG promoter in the pB_CAG_eGFP_Puro construct differed from that of the construct used in Schübeler *et al.* (2000), meaning it was not possible to utilise the same PCR primers (Schübeler *et al.*, 2000). The downstream region of the construct used in this study has a high density of CpG sites, making unfavourable for targeting with BS/EM-PCR primers. However, it would be of interest to obtain methylation sequencing from the promoter region of the reporter, in order to determine whether the methylation state of the promoter is in line with results from Schübeler *et al.* (2000).

The results demonstrating that methylation was largely maintained at synthetic reporters in HCT116 cells have interesting implications on the dynamics of DNA methylation maintenance in human cancer cells. Previous work has suggested that DNA hypomethylation in cancer is a progressive process, arising through incomplete maintenance of methylation patterns over consecutive cell divisions (Salhab *et al.*, 2018; Zhou *et al.*, 2018; Decato *et al.*, 2020). However, my data shows that HCT116 cells are capable of faithfully propagating DNA methylation states over 75 cell divisions. This finding suggests that the process of hypomethylation may not be ongoing and highlights the functionality of the DNA methylation maintenance machinery in HCT116 cells. DNA hypomethylation arises as an early event in cancer development (Goelz *et al.*, 1985; Timp *et al.*, 2014). Therefore, the mechanism underlying PMD hypomethylation may be a more active process occurring during malignant transformation, as opposed to a gradual, passive process.

There are some limitations to this reporter system that should be considered. The effect of *in vitro* methylation of the CAG promoter on expression of the eGFP reporter gene has been robustly characterised within this experimental system. However, this reporter cassette consists of a synthetic promoter construct and an exogenous reporter gene. Therefore, this

artificial system may not accurately represent the behaviour of an endogenous human sequence. The 2015 study by Stelzer *at al*. characterised a methylation sensitive reporter system based on a minimal Snrpn promoter, which was shown to accurately report *in vivo* demethylation (Stelzer *et al.*, 2015). The Snrpn promoter could be utilised within this experimental system to more reliably estimate the behaviour of a human endogenous sequence. However, the Snrpn promoter has shown weak expression in HC116 cells when tested by our research group.

Another caveat of this study is that reporter construct used here has a highly CpG-dense sequence. Local CpG density has been proposed to influence methylation maintenance, with high CpG dense regions potentially reinforcing maintenance of methylation (Higham *et al.*, 2022). Therefore, the high CpG density of the reporter construct may promote maintenance of methylation. It would be of interest to repeat these experiments using a reporter construct with a lower CpG density, to gain insight into how CpG density might influence methylation maintenance.

In summary, in this chapter I present data detailing the development of a novel transposon-based fluorescent reporter system that accurately reports the underlying methylation state of the reporter construct. Due to the flexibility of the piggyBac transposition system, this reporter cassette could be integrated into any transfectable cell line to study DNA methylation maintenance. Additionally, any element of the reporter construct could be replaced. Alternative promoters could be tested, or the reporter gene could be exchanged with a different reporter, a human endogenous gene, or any genetic element of interest. This flexibility widens the potential applications of the methylation-sensitive reporter system developed here.

4 Methylation is maintained at integrated reporters primarily through activity of DNMT1 and UHRF1

4.1 Introduction

The results presented in Chapter 1 demonstrate that methylation was largely maintained at synthetic reporters in HCT116 cells over a prolonged period of time, as was repression of the reporter gene. These data suggest that HCT116 cells are capable of faithfully propagating DNA methylation patterns. Next, I was interested in the potential applications of the methylation-sensitive reporter system as a screening tool, to identify factors that are responsible for methylation maintenance in cancer cells and to screen for therapeutic targets. To investigate whether the system would be suitable for this purpose, I sought to test whether the reporter would be activated in response to changes in global methylation.

To do this, I utilised HCT116 cell lines with disruption of key proteins previously reported to have a role in DNA methylation maintenance, and assessed how removal of these proteins impacted reporter expression and maintenance of DNA methylation. DNA methyltransferase 1 (DNMT1) is the primary methyltransferase responsible for maintaining DNA methylation following DNA replication (Moore, Le and Fan, 2013). DNA methylation loss in cancer has been linked to cell division; it has been proposed that DNMT1 is unable to faithfully propagate DNA methylation in rapidly proliferating cells, leading to progressive loss of DNA methylation cancer (Charlton *et al.*, 2018; Zhou *et al.*, 2018). The results presented in Chapter 1 demonstrate that DNA methylation was largely maintained at synthetic reporters in HCT116 cells. These data suggest that DNMT1 is capable of faithfully propagating DNA methylation in HCT116 cells. To confirm that loss of DNMT1 activity would lead to reporter activation, I assessed reporter expression and methylation in HCT116 DNMT1 knock-out (henceforth, $DNMT1^{KO}$) cells. The $DNMT1^{KO}$ cells used in this study were generated by Rhee *et al*. (2000). Overall, $DNMT1^{KO}$ cells displayed a 20% reduction in global m^5C content (Rhee *et al.*, 2000).

The high retention of DNA methylation following knock-out of DNMT1 implies that the activity of other proteins may contribute to methylation maintenance. DNA methyltransferase 3B (DNMT3B) is a *de novo* methyltransferase, with a role in establishing initial methylation patterns during embryonic development (Moore, Le and Fan, 2013). In a follow-up study,

Rhee et al. (2002) generated a HCT116 DNMT3B knock-out (henceforth, DNMT3B^{KO}) cell line (Rhee *et al.*, 2002). DNMT3B^{KO} retained >97% of wild-type m^{5}C content, suggesting that loss of DNMT3B does not have a substantial impact on DNA methylation maintenance. Interestingly, double knock-out (DKO) of DNMT1 and DNMT3B resulted in an overall 57.5-95% reduction in m^{5}C content in different clones tested. This finding indicated that DNMT1 and DNMT3B may cooperate to maintain DNA methylation levels in human cells. To test whether I could recapitulate these findings using the reporter system developed in this study, I assessed reporter expression in DNMT3B^{KO} and DKO cell lines.

Another protein that has been reported to play an important role in DNA methylation maintenance is UHRF1 (Mancini, Magnani, Macchi and I. M. Bonapace, 2021). UHRF1 is recruited to replicating DNA and binds with a preference for hemi-methylated CpG sites (Arita *et al.*, 2008). UHRF1 ubiquitinates H3K18Ub and H3K23Ub, which bind to the RFTD domain of DNMT1, resulting in spatial rearrangement and opening of the active site of DNMT1 (Qin *et al.*, 2015; Ishiyama *et al.*, 2017; Nishiyama *et al.*, 2020). Thus, UHRF1 acts as an upstream activator of DNMT1 and is essential for DNA methylation maintenance. Recently, Yamaguchi *et al*. (2024) utilised Auxin-Inducible Degron (AID) technology to develop degron-tagged alleles for UHRF1 and DNMT1 in HCT116 cells (Yamaguchi *et al.*, 2023). The AID system is a method that allows rapid and conditional depletion of proteins in human cells (Nishimura *et al.*, 2009; Natsume *et al.*, 2016). This is done by fusing a destabilising domain, known as a degron, to either the N or C terminal of a protein (Chen, Gaynor and Chen, 2019). The degron tag is an amino acid sequence that is recognised by proteases, leading to degradation of the tagged protein (Worley *et al.*, 2000; Ramos *et al.*, 2001). Within the AID system, induction of protein degradation is controlled by two components. The auxin responsive F-box transport inhibitor response 1 (TIR1) is the first component. Yamaguchi et al. (2024) used TIR1 derived from *Oryza sativa* (OsTIR1), under doxycycline-inducible control (Natsume *et al.*, 2016). The presence of auxin is then required to activate degradation. Auxins, such as indole-3-acetic acid (IAA), are a family of plant hormones with a role in mediating gene expression (Teale, Paponov and Palme, 2006). Auxins bind to OsTIR1 and promote ubiquitination of OsTIR1 target proteins, a process that marks these target proteins for degradation by the proteosome (Ruegger *et al.*, 1998; Teale, Paponov and Palme, 2006).

Degradation of AID-tagged UHRF1 and DNMT1 in HCT116 cells was shown to result in rapid depletion of the tagged proteins, accompanied by global reduction in m^5C (Yamaguchi *et al.*, 2023). Interestingly, depletion of UHRF1 was reported to result in a greater reduction of m^5C than DNMT1 depletion. In cells depleted of both DNMT1 and UHRF1, slightly more loss of methylation was observed. These data suggest that UHRF1 may have additional roles in DNA methylation maintenance outside of its interaction with DNMT1. In fact, Yamaguchi *et al*. demonstrated that UHRF1 interacts with DNMT3A and DNMT3B, further implicating the *de novo* methyltransferases in DNA methylation maintenance. Thus, the degron system developed by Yamaguchi *et al*. (2023) is a useful tool to interrogate the functionality of DNMT1 and UHRF1.

In this chapter, I utilise HCT116 knock-out and HCT116-AID tagged cell lines to investigate DNA methylation maintenance at synthetic reporters in cells lacking activity of DNMT1, DNMT3B and UHRF1. I show that the reporter is activated in response to DNMT1 knock-out and acute depletion of DNMT1 and/or UHRF1. In $DNMT1^{KO}$ cells, reporter activation was accompanied with demethylation of the reporter construct. Reporter repression and methylation were maintained in $DMT3B^{KO}$ cells, indicating that loss of DNMT3B alone does not greatly impact methylation maintenance fidelity. Together, these results indicate that DNA methylation is maintained primarily through the activity of DNMT1 and UHRF1 in HCT116 cells. Furthermore, these data demonstrate that the methylation-sensitive reporter system developed in this study is responsive to global methylation changes. This finding suggests that the reporter could be used as a screening tool, to identify novel factors involved in DNA methylation maintenance.

4.2 Experimental design

In order to assess how loss of DNMT1 and/or DNMT3B activity impacted reporter gene expression and maintenance of DNA methylation, I stably integrated methylated reporters into HCT116 WT, $DNMT1^{KO}$, $DNMT3B^{KO}$ and DKO cells. Cells were expanded for 28 days (4 weeks), to ensure unintegrated plasmid was diluted out of the population prior to analysing reporter expression and methylation (Fig. 4.1). Transfection and passaging of cells were

conducted in collaboration with Hazel Davidson-Smith, Sproul lab Research Assistant (IGC). Failure to maintain DNA methylation following cell division can result in passive demethylation within a single DNA replication cycle. Successive cycles of cell division in the absence of functional methylation maintenance progressively dilute global levels of m^5C (Moore, Le and Fan, 2013; Sandoval and Serra, 2016). HCT116 cells have a doubling time of approximately 18 hours (Gongora *et al.*, 2008; Zhao *et al.*, 2021). Therefore, I hypothesised that 28 days in culture, equivalent to approximately 37 cell divisions, would be sufficient to observe changes to expression of the reporter gene if DNA methylation was not faithfully maintained during that time period. Reporter gene expression was quantified by flow cytometry 28 days after introduction into the cell genome. To determine DNA methylation levels at integrated reporters, genomic DNA was extracted from cells at the 28-day time point and EM-converted. To obtain a broad picture of the methylation landscape across distinct integrated reporters, I performed deep amplicon EM-sequencing targeting the CMV enhancer and eGFP gene body regions of the reporter cassette.

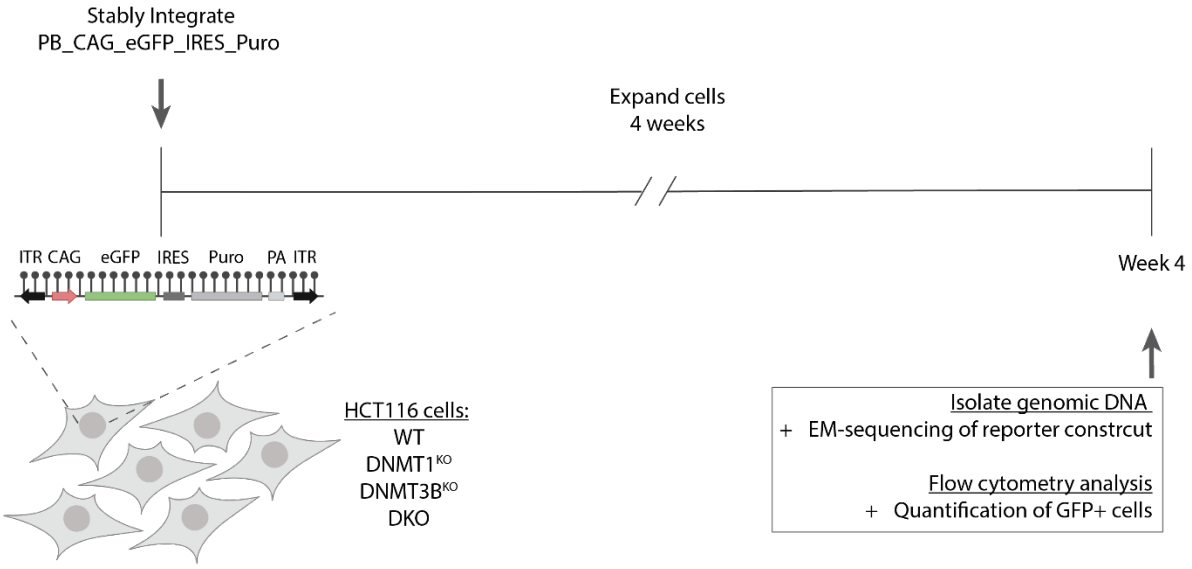


Figure 4-1: Experimental schematic for reporter analysis in HCT116 knock-out cell lines. *The PB_CAG_eGFP_IRES_Puro reporter construct was stably integrated into HCT116 WT, DNMT1^{KO}, DNMT3B^{KO}, and DKO cells. Cells were expanded for 28 days (4 weeks) to dilute out unintegrated plasmid. I hypothesised that 4 weeks culture, corresponding to ~37 cell divisions, would be sufficient to observe reporter activation if DNA methylation was not maintained at the reporter construct during this time period. At week 4, reporter expression was quantified by flow cytometry and genomic DNA was isolated from cells. Reporter DNA methylation levels were assessed by EM-sequencing.*

4.3 Reporter repression is lost in DNMT1^{KO} cells

First, I asked whether reporter repression would be maintained in DNMT1^{KO} cells. I stably integrated reporters into DNMT1^{KO} cells and expanded cells for 28-days before quantifying eGFP expression by flow cytometry (Fig. 4.2). The MFI of GFP positive DNMT1^{KO} cells containing unmethylated reporters (DNMT1^{KO}-eGFP) was 2.03×10^{4}, while the MFI of GFP positive cells containing premethylated reporters (DNMT1^{KO}-eGFPme) was 5.19×10^{3}. The MFI of cells was significantly reduced ($p=2.41x10^{-2}$, paired t-test) when the reporter was methylated, compared to the unmethylated reporter (Fig. 4.2C). Next, I compared MFI between HCT116-eGFPme and DNMT1^{KO}-eGFPme cells. The MFI of GFP positive DNMT1^{KO}-eGFPme cells was significantly ($p=3.69x10^{-3}$, paired t-test) higher than HCT116-eGFPme cells (Fig. 4.2C). The mean proportion of DNMT1^{KO}-eGFPme cells expressing a high level of GFP was 52.47% (Fig. 4.2D). This proportion of high-GFP DNMT1^{KO}-eGFPme cells was significantly ($p=7.45x10^{-3}$, paired t-test) lower than cells with unmethylated reporters, but a significant 21.16-fold increase (p=5.21x10-3, paired-t-test) compared to HCT116-eGFPme cells (Fig. 4.2D). Together, these data indicate that reporter repression was lost in DNMT1^{KO} cells. This suggests that DNA methylation was not faithfully maintained at integrated reporters in DNMT1^{KO} cells, however, it was necessary to confirm methylation of integrated reporters by EM-sequencing.

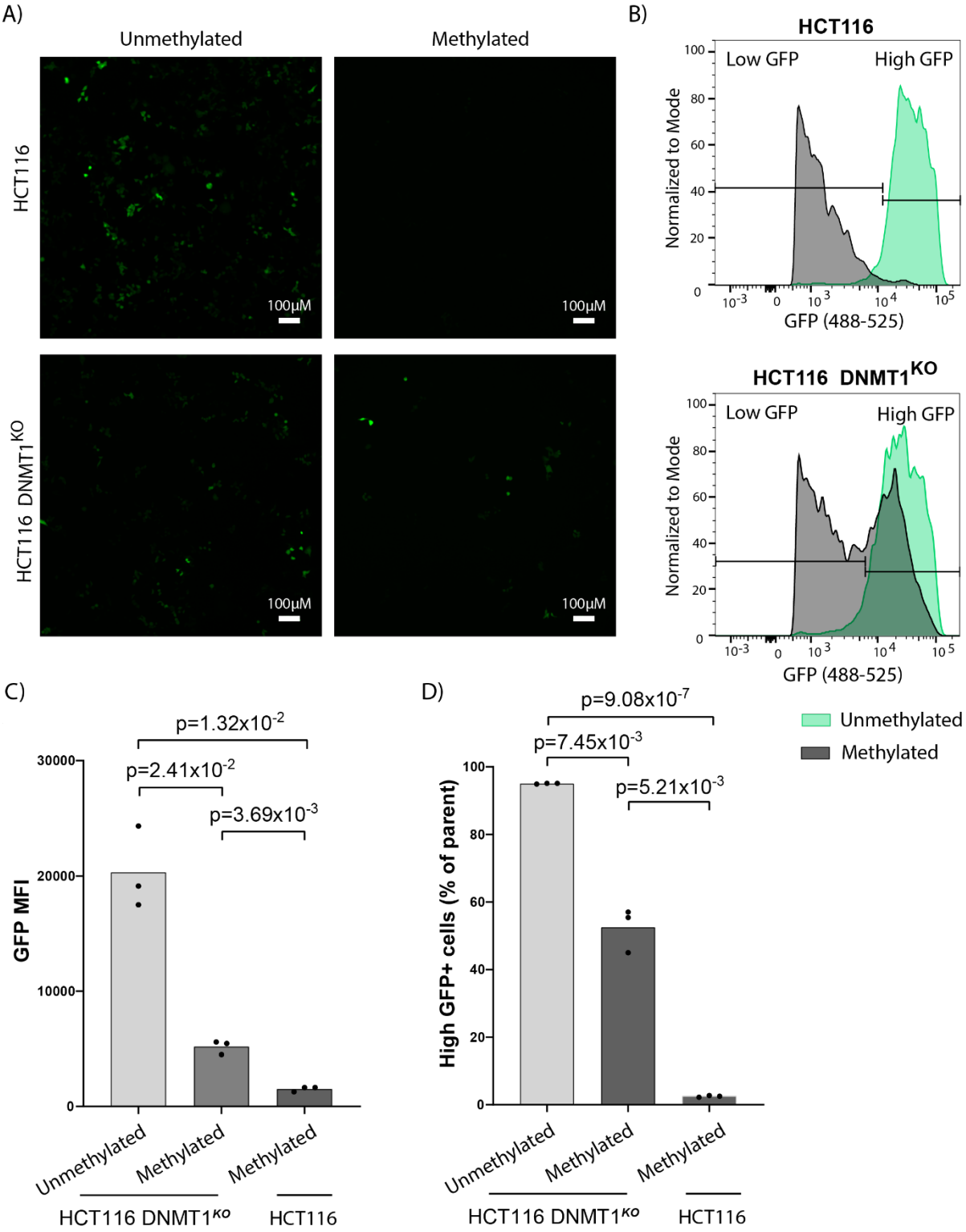


Figure 4-2: Reporter repression is lost in DNMT1^{KO} cells. A) *Representative fluorescence microscopy images of HCT116 WT (top) and DNMT1^{KO} cells (bottom) containing unmethylated or premethylated reporters. Scale bar = 100uM.* ***B)*** *Representative histograms showing fluorescence intensity distributions from GFP+ cells containing either unmethylated (green) or methylated (grey) reporters. HCT116 cells (top) are shown in comparison to DNMT1^{KO} cells (bottom). Y axis = total count of cells, normalized to the mode. X-axis = GFP fluorescence intensity. Gates for high and low GFP expression are indicated by horizontal bars and were set*

so that 95% of the events in the unmethylated GFP-positive population fell within the high-GFP gate. ***C)*** *Median fluorescence intensity (MFI) of all GFP positive cells.* ***D)*** *Proportion of GFP positive cells (parent) that express a high level of GFP. For bar graphs, bars represent the mean of n=3 biological replicates. Data points represent the mean value from 100,00 cells measured per biological replicate. P-values determined by paired t-tests.*

4.4 Methylation is reduced at stably integrated reporters in DNMT1^{KO} cells

Next, I assessed the underlying methylation of reporters integrated in DNMT1^{KO} cells by deep amplicon EM-sequencing targeting the CMV enhancer region and eGFP gene body regions of the reporter construct (Fig. 4.3). In DNMT1^{KO}-eGFP cells, the median %mCpG at the CMV enhancer was 0%, and the median %mCpG observed at the eGFP gene body region was 0%. Some minor gain of methylation was observed at both the enhancer and gene body, but there was no significant (p= 0.42, paired t-test) difference in mean methylation levels between the two regions (Fig. 4.3A). In DNMT1^{KO}-eGFPme cells, the median %mCpG at the CMV enhancer region was 40.00%. The median %mCpG at the eGFP gene body in these cells was 85.71%. Levels of DNA methylation at reporters integrated into DNMT1^{KO} cells were compared to HCT116 WT cells. In HCT116-eGFPme cells, the median %mCpG at the enhancer was 73.33% and at the gene body 100%. There were no significant differences between the %mCpG at the CMV enhancer (p=0.13, paired t-test) or the eGFP gene body (p= 0.35, paired t-test) when comparing DNMT1^{KO} cells to HCT116 WT. However, the median %mCpG was substantially lower across both regions of the reporter construct in DNMT1^{KO} cells, compared to HCT116 WT. The median %mCpG at the CMV enhancer was ~1.85-fold reduced in DNMT1^{KO}-eGFPme cells compared to HCT116-eGFPme. The mean percentage of methylation at the gene body was also ~1.16-fold reduced in DNMT1^{KO} cells. Unlike HCT116-eGFPme cells, there was no significant difference (p=0.18, paired t-test) between the methylation level of the enhancer region compared to the gene body region in DNMT1^{KO}-eGFPme cells. Together, these results indicate that DNA methylation was not as faithfully maintained at integrated reporters in DNMT1^{KO} cells, in line with what was suggested by the reporter expression data.

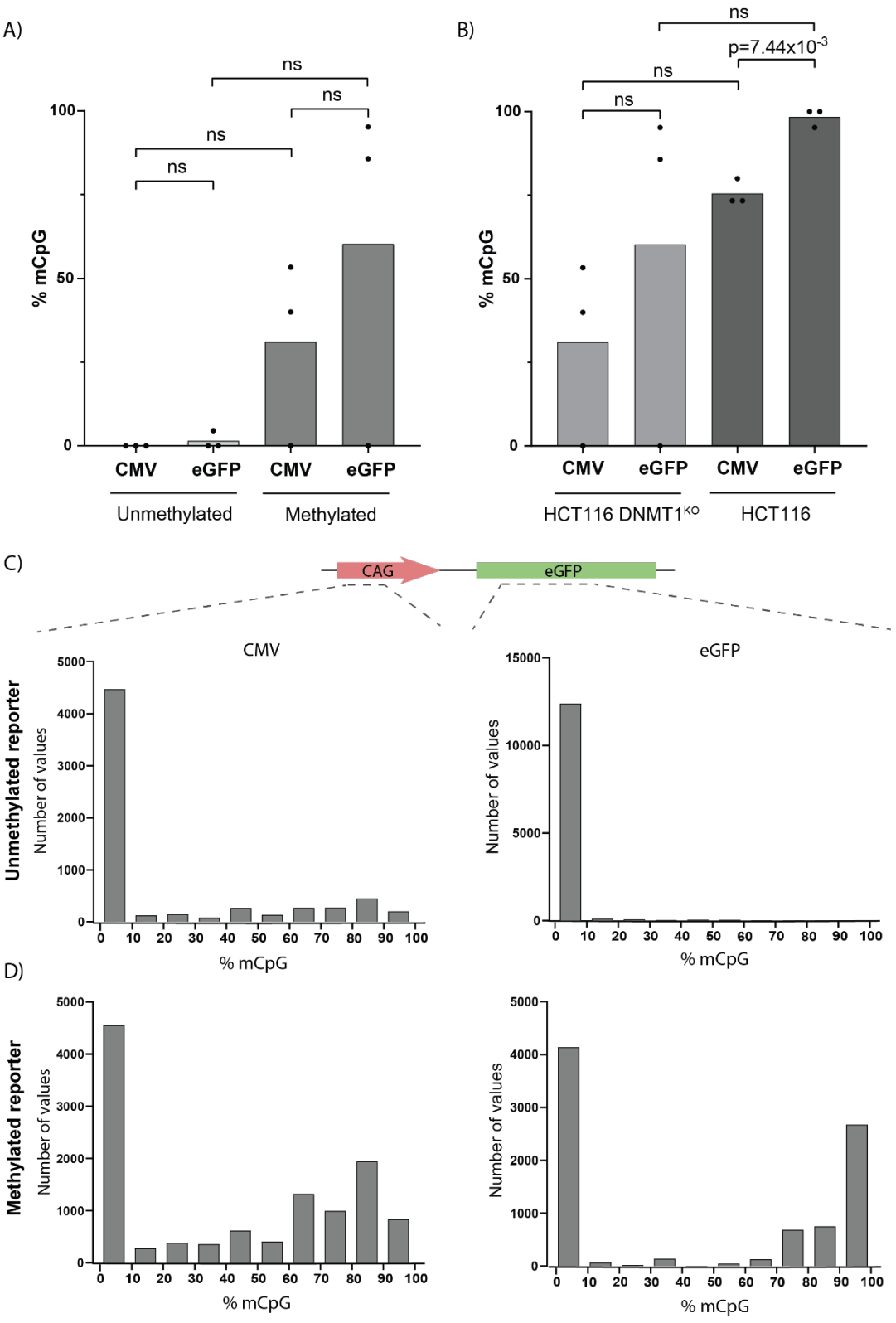


Figure 4-3: Methylation is reduced at stably integrated reporters in DNMT1^{KO} cells. A) *The median percentage of methylated CpG sites (%mCpG) within the CMV enhancer and eGFP gene body in DNMT1^{KO} cells containing either unmethylated or methylated reporter constructs.* ***B)*** *The median %mCpG reporters in DNMT1^{KO}-eGFPme compared to HCT116-eGFPme cells. Bars represent the mean of n=3 biological replicates, with data points representing the median of each individual replicate. P-values determined by paired t-test.*

Total number of sequencing reads: Unmethylated replicate 1: CMV, n=8855, eGFP, n=15,850; Unmethylated replicate 2: CMV, n=6697, eGFP n=12,760; Unmethylated replicate 2: CMV, n=25,246, eGFP n=28568; Methylated replicate 1: CMV, n=10,844, eGFP, n= 9621; Methylated replicate 2: CMV, n=11,979, eGFP, n=11,534; Methylated replicate 3: CMV, n=27,026, eGFP, n=24,625. ***C)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 6697) and eGFP gene body (n=12,760) of unmethylated reporter constructs in DNMT1^{KO} cells.* ***D)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 11,979) and eGFP gene body (n= 11,534) of methylated reporter constructs in DNMT1^{KO} cells. Example histograms from experimental replicate 2. Bin width = 10.*

4.5 Reporter repression is maintained in DNMT3B^{KO} cells

Next, I asked whether reporter repression would be maintained in DNMT3B^{KO} cells. I stably integrated methylated reporters into DNMT3B^{KO} cells and expanded cells for 28-days before quantifying eGFP expression by flow cytometry (Fig. 4.4). The MFI of GFP positive cells in DNMT3B^{KO} cells containing unmethylated reporters (DNMT3B^{KO}-eGFP) was 2.81×10^{4}. The MFI of GFP positive DNMT3B^{KO} cells containing methylated reporters (DNMT3B^{KO}-eGFPme) was 1.11×10^{3}. This represents a significant ($p=6.48\times10^{-3}$, paired t-test) reduction in GFP MFI when the reporter was methylated. When compared to the MFI of GFP positive HCT116-eGFPme cells, there were no significant (p= 0.13, paired t-test) differences in GFP MFI between HCT116 and DNMT3B^{KO} cells. The mean proportion of DNMT3B^{KO}-eGFPme that expressed a high level of GFP was 2.99%. There was no significant (p=0.67, paired t-test) difference between in the proportion of cells expressing a high level of GFP in DNMT3B^{KO}-eGFPme cells compared with HCT116-eGFP. Together, these data indicate that reporter repression was largely maintained in DNMT3B^{KO} cells. This suggests that there were no detectable differences in maintenance of DNA methylation at reporter constructs in DNMT3B^{KO} cells.

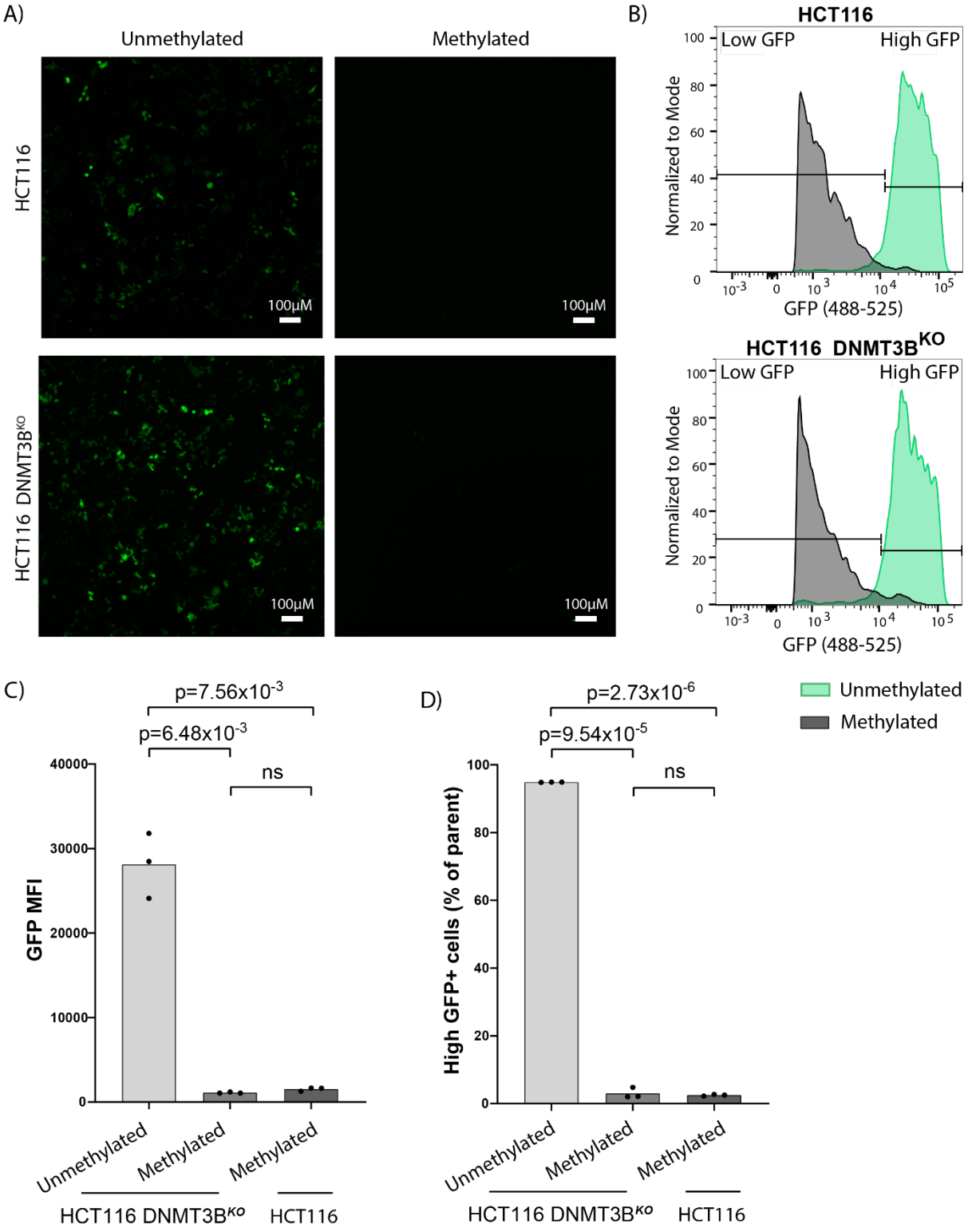


Figure 4-4: Reporter repression is maintained in DNMT3B^{KO} cells. ***A)*** *Representative fluorescence microscopy images of HCT116 WT (top) and DNMT3B^{KO} cells (bottom) containing unmethylated or premethylated reporters. Scale bar = 100uM.* ***B)*** *Representative histograms showing fluorescence intensity distributions from GFP+ cells containing either unmethylated (green) or methylated (grey) reporters. HCT116 cells (top) are shown in comparison to DNMT3B^{KO} cells (bottom). Y axis = total count of cells, normalized to the mode. X-axis = GFP fluorescence intensity. Gates for high and low GFP expression are indicated by horizontal bars,*

and were set so that 95% of the events in the unmethylated GFP-positive population fell within the high-GFP gate. ***C)*** *Median fluorescence intensity (MFI) of all GFP positive cells.* ***D)*** *Proportion of GFP positive cells (parent) that express a high level of GFP. Bars represent the mean of n=3 biological replicates. Data points represent the mean value from 100,000 cells measured per biological replicate. P-values determined by paired t-tests. Data from HCT116 WT cells is repeated from previous figure.*

4.6 Reporter methylation is maintained in DNMT3B^{KO} cells

In order to assess methylation at stably integrated reporters in DNMT3B^{KO} cells, I performed deep amplicon EM-sequencing targeting the CMV enhancer region and eGFP gene body regions of the reporter construct (Fig. 4.5). In DNMT3B^{KO}-eGFP cells median %mCpG at the CMV enhancer was 0%, and at the eGFP gene body region in these cells the median %mCpG was also 0%. Low levels of methylation gain were observed, with no significant (p=0.42, paired t-test) differences in methylation level between the enhancer and gene body regions. In DNMT3B^{KO}-eGFPme cells, the median %mCpG at the CMV enhancer was 73.33%. The median %mCpG observed at the eGFP gene body region in DNMT3B^{KO}-eGFPme cells was 100% (Fig. 4.5A). Similar to previous results in HCT116 WT cells, in DNMT3B^{KO}-eGFPme cells the enhancer was significantly ($p=5.85x10^{-3}$, paired t-test) more demethylated than the gene body region, which largely maintained methylation (Fig. 4.5A). There were no significant differences between the median methylation levels at the enhancer (p=0.18, paired t-test) or gene body (p=0.42, paired t-test) regions in DNMT3B^{KO}-eGFPme cells compared to HCT116-eGFPme (Fig. 4.5B). Together, these data indicate that there were no significant differences in DNA methylation maintenance fidelity in DNMT3B^{KO} cells compared to HCT116 WT cells. These results are in line with what was expected, based on reporter gene expression.

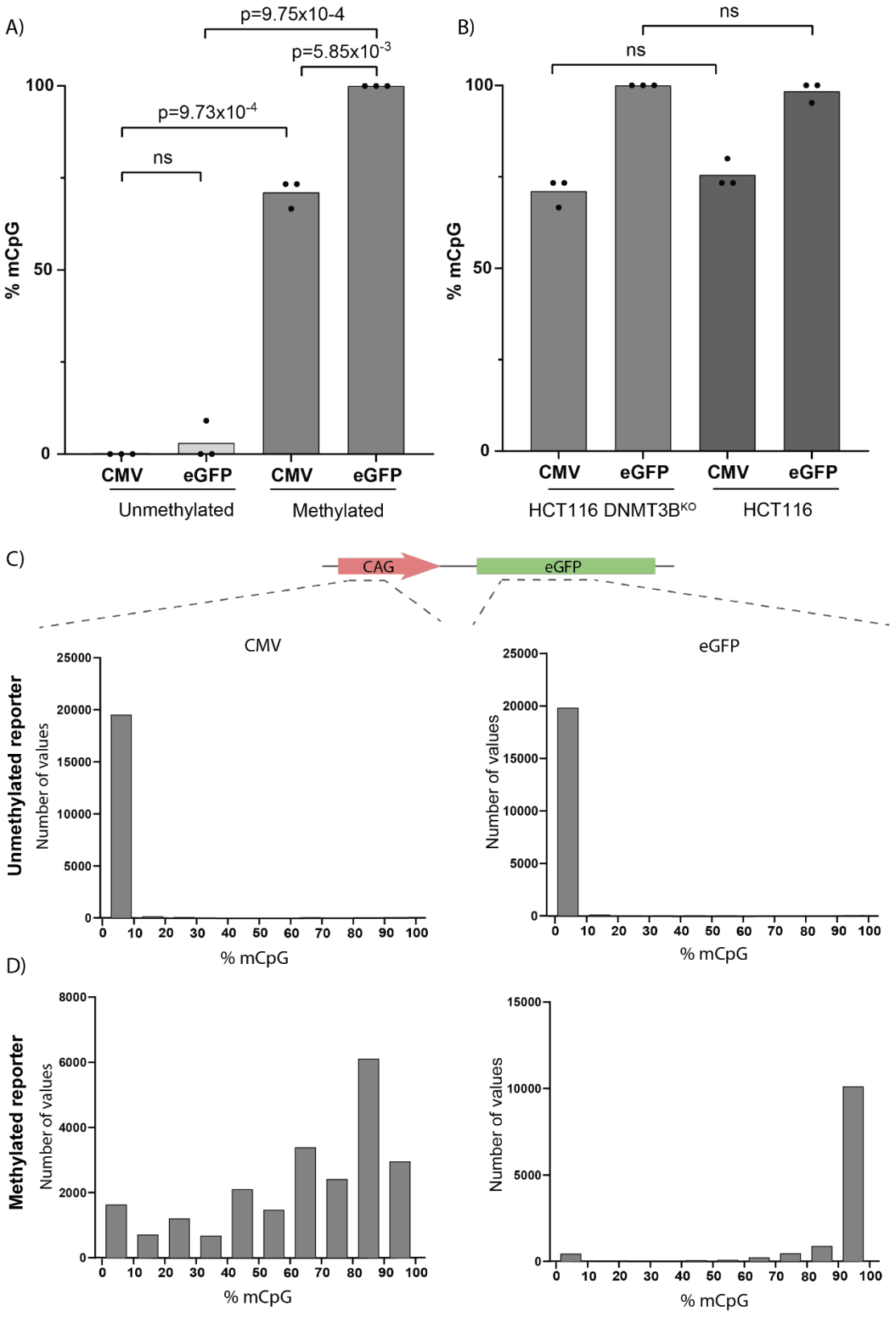


Figure 4-5: Methylation is largely maintained at stably integrated reporters in DNMT3B^{KO} cells. A) *The median percentage of methylated CpG sites (%mCpG) within the CMV enhancer and eGFP gene body in DNMT3B^{KO} cells containing either unmethylated or methylated reporter constructs.* ***B)*** *The median %mCpG in DNMT3B^{KO}-eGFPme cells compared to HCT116-eGFPme cells. For bars plots, bars represent the mean of n=3 biological replicates, with data*

points representing the median of each individual replicate. P-values determined by paired t-test. WT data is repeated from previous figure. Total number of sequencing reads: Unmethylated replicate 1: CMV, n=6121, eGFP, n=13,956; Unmethylated replicate 2: CMV, n=6319, eGFP n=10,361; Unmethylated replicate 2: CMV, n=19,959, eGFP n=20,352; Methylated replicate 1: CMV, n=24399, eGFP, n= 30,695; Methylated replicate 2: CMV, n=10,050, eGFP, n=7475; Methylated replicate 3: CMV, n=5591, eGFP, n=24,625. ***C)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 19,959) and eGFP gene body (n= 20,352) of unmethylated reporter constructs.* ***D)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 10,050) and eGFP gene body (n= 7475) of methylated reporter construct in one experimental replicate. Example histograms from experimental replicate 2. Bin width = 10.*

4.7 Reporter repression is lost in a subset of DKO cells

I next asked whether reporter repression would be maintained in DKO cells. I stably integrated methylated reporters into DKO cells and expanded cells for 28 days before analysing reporter expression by flow cytometry (Fig. 4.6). The MFI of GFP positive DKO cells containing unmethylated reporters (DKO-eGFP) was 3.37×10^{4}. The MFI of GFP positive DKO cells containing methylated reporters (DKO-eGFPme) was 1.19×10^{3}. The MFI of GFP positive cells was significantly ($p=1.19x10^{-3}$, paired t-test) lower when the reporter construct was premethylated in DKO cells (Fig. 4.6C). There was no significant (p=0.08, paired t-test) difference between the MFI of GFP positive DKO-eGFPme cells compared to HCT116-eGFPme cells. The mean proportion of DKO-eGFPme cells that expressed a high level of GFP was 21.33% (Fig. 4.6D). This was a significant ($p=2.19x10^{-2}$, paired t-test) 8.60-fold increase in high GFP positive cells compared to HCT116-eGFPme cells. These results indicate that reporter repression was lost in a subset of DKO cells, suggesting that methylation was not as faithfully maintained in DKO-eGFPme cells.

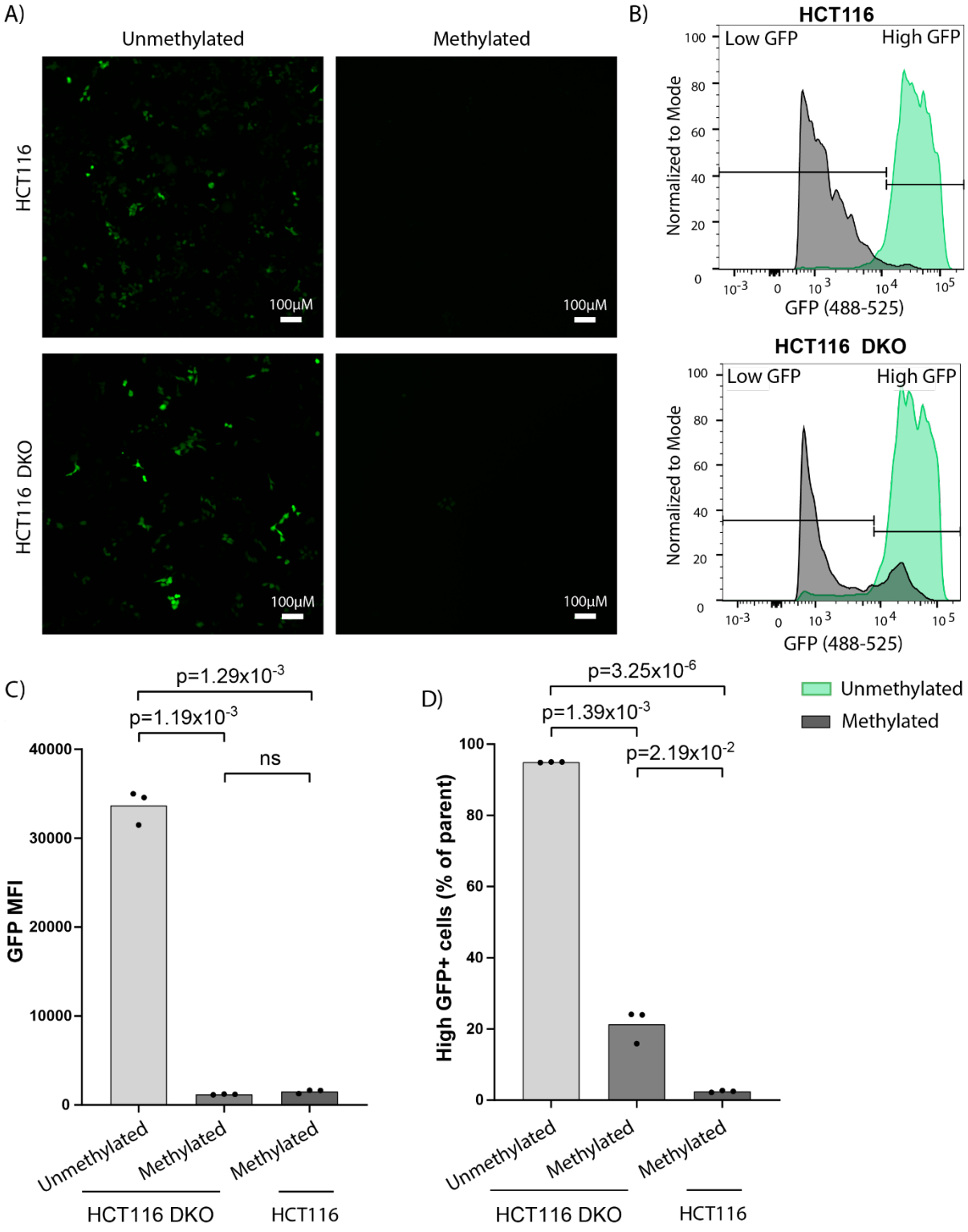


Figure 4-6: Reporter repression is lost in a subset of DKO cells. ***A)*** *Representative fluorescence microscopy images of HCT116 WT (top) and DKO cells (bottom) containing unmethylated or premethylated reporters. Scale bar = 100uM.* ***B)*** *Representative histograms showing fluorescence intensity distributions from GFP+ cells containing either unmethylated (green) or methylated (grey) reporters. HCT116 cells (top) are shown in comparison to DKO cells (bottom). Y axis = total count of cells, normalized to the mode. X-axis = GFP fluorescence intensity. Gates for high and low GFP expression are indicated, and were set so that 95% of*

the events in the unmethylated GFP-positive population fell within the high-GFP gate. ***C)*** *Median fluorescence intensity (MFI) of all GFP positive cells.* ***D)*** *Proportion of GFP positive cells (parent) that express a high level of GFP. For bar graphs, bars represent the mean of n=3 biological replicates. Data points represent the mean value from 100,000 cells measured per biological replicate. P-values determined by paired t-tests. Data from HCT116 WT cells is repeated from previous figure.*

4.8 Reporter methylation largely maintained in DKO cells

Next, I assessed the underlying methylation state of reporters integrated in DKO cells by deep amplicon EM-sequencing targeting the CMV enhancer region and eGFP gene body regions of the reporter construct (Fig. 4.7). In DKO-eGFP cells, the median %mCpG the CMV enhancer was 0%, and the median %mCpG at the eGFP gene body region in these cells was also 0% (Fig. 4.7A). As observed in the previous cell lines, some minor gain methylation was observed enhancer and gene body, with no significant (p=0.42, paired t-test) difference between gain of methylation at these two regions. In DKO-eGFPme cells, median %mCpG at the CMV enhancer was 93.33%. The median %mCpG observed at the eGFP gene body region 95.24% (Fig. 4.7A). There was no significant (p=0.16, paired t-test) difference in the methylation level between the enhancer and gene body regions in DKO-eGFPme cells. This differs from results in HCT116 WT cells, in which the enhancer was more likely to become demethylated. Methylation levels in DKO-eGFPme cells were compared to those in HCT116-eGFPme cells. DKO-eGFPme cells had a 1.30-fold higher level of methylation at the enhancer region, a significant ($p=1.52x10^{-2}$, paired t-test) increase compared to HCT116-eGFPme cells. This suggests that methylation may have been maintained more faithfully at the enhancer in DKO-eGFPme cells. There was no significant (p=0.42, paired t-test) difference between the %mCpG at the gene body region in DKO-eGFPme cells compared with HCT116-eGFPme cells. Together, these data indicate that DNA methylation was maintained in DKO cells, despite the lack of DNMT1 and DNMT3B activity in these cells. These results do not align with what was expected, based the reporter expression data, and may suggest that the reporter escaped repression through a mechanism unrelated to methylation at the reporter.

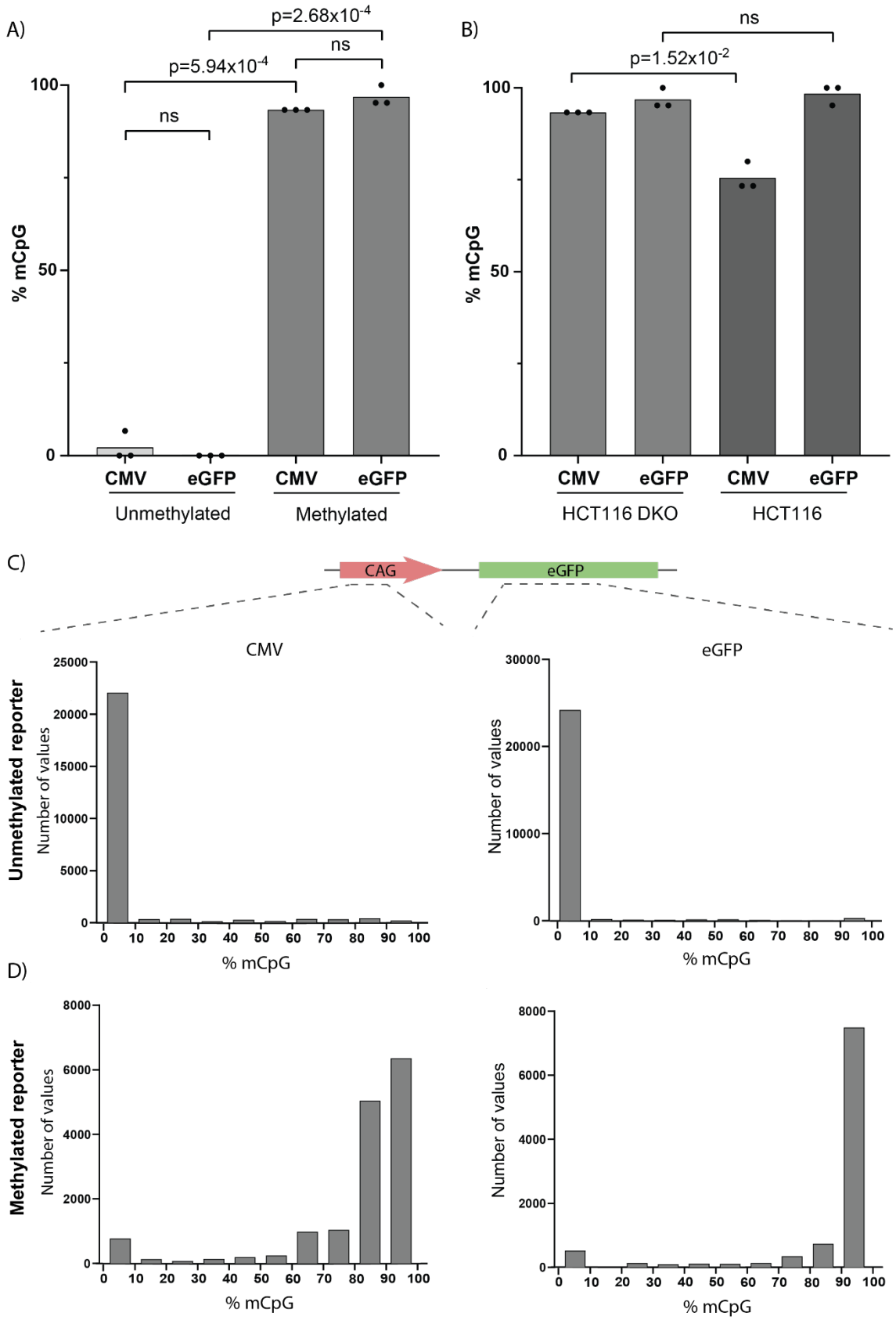


Figure 4-7: Methylation is maintained at stably integrated reporters in DKO cells. A) *The median percentage of methylated CpG sites (%mCpG) within the CMV enhancer and eGFP gene body in in DKO cells containing either unmethylated or methylated reporter constructs.*

B) *The median %mCpGs in DKO-eGFPme cells compared to HCT116-eGFPme cells. For bars plots, bars represent the mean of n=3 biological replicates, with data points representing the median of each individual replicate. P-values determined by paired t-test. WT data is repeated from previous figure. Total number of sequencing reads: Unmethylated replicate 1: CMV, n=5891, eGFP, n=10,380; Unmethylated replicate 2: CMV, n=7062, eGFP n=10,228; Unmethylated replicate 2: CMV, n=24,899, eGFP n=25,889; Methylated replicate 1: CMV, n=12,184, eGFP, n= 11,533; Methylated replicate 2: CMV, n=9994, eGFP, n=9182; Methylated replicate 3: CMV, n=26,646, eGFP, n=26,073.* ***C)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n=24,899) and eGFP gene body (n=25,889) of unmethylated reporter constructs. Bin width = 10.* ***D)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 9994) and eGFP gene body (n=9182) of methylated reporter construct in one experimental replicate. Example histograms from experimental replicate 2. Bin width = 10.*

4.9 Reporter repression is lost following depletion of UHRF1

Having established that reporter repression was lost in DNMT1^{KO}-eGFPme and DKO-eGFPme cell lines, I was next interested in testing the response of the reporter to loss of UHRF1. To do this, I utilised HCT116-AID degron cell lines, which enable rapid depletion of AID-tagged proteins (Yamaguchi *et al.*, 2024). Using this system, I was able to induce depletion of DNMT1 and/or UHRF1, allowing me to investigate the impact of UHRF1 depletion on reporter expression and maintenance of methylation. The DNMT1-AID cells were included as a positive control, as my previous results indicated that reporter activation should be observed in response to loss of DNMT1. Furthermore, utilising the DNMT1-AID cells allowed me to replicate the findings from DNMT1^{KO} cells in an alternative experimental system. I was also interested in testing whether reporter activation would be observed in response to acute depletion of DNMT1 and/or UHRF1.

To test whether depletion of DNMT1 and/or UHRF1 would lead to activation of the methylated reporter construct, I stably integrated methylated reporters into HCT116 WT, HCT116 DNMT1-AID, HCT116 UHRF1-AID and HCT116 DNMT1-AID-UHRF1-AID degron cell lines (Yamaguchi *et al.*, 2024). Cells were again expanded for 28 days, to ensure that unintegrated plasmid was diluted out of the population prior to analysing reporter expression (Fig. 4.8). Following expansion, cells were treated with Doxycycline to induce expression of OsTIR1 and IAA to activate degradation of tagged proteins. Yamaguchi *et al*. (2024) reported

that degron-tagged proteins were largely depleted 2 hours after induction, and hypomethylation of cells was observed by 48 hours. HCT116 cells are reported to grow with a doubling time of approximately 18 hours (Gongora *et al.*, 2008; Zhao *et al.*, 2021). Therefore, I hypothesised that two rounds of cell division would be sufficient to obtain hypomethylation and observe reporter activation. Thus, I grew cells in the presence of doxycycline and IAA for 48 hours before assessing reporter expression by flow cytometry.

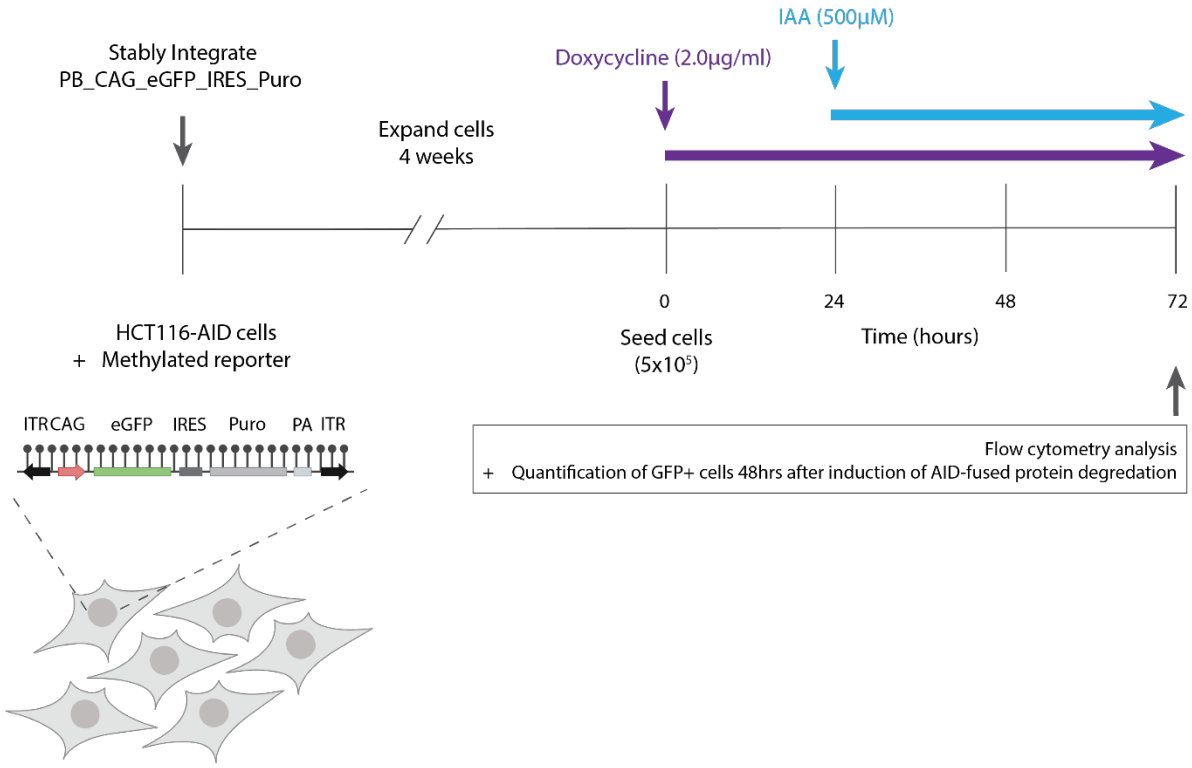


Figure 4-8: Experimental schematic for eGFPme reporter analysis in HCT116-AID cell lines. *The PB_CAG_eGFP_IRES_Puro reporter construct was stably integrated into HCT116-AID cells lines. Cells were expanded for 4 weeks (28 days) to dilute out unintegrated plasmid. Following expansion, 5x105 cells were seeded in 6-well plates and treated with doxycycline (Dox) for 24 hours to activate OdTIR1 expression. After 24 hours, media was replaced with fresh media containing Dox and IAA to induce protein degradation. Cells were grown with Dox and IAA for 48 hours, then reporter expression was analysed by flow cytometry.*

In HCT116-AID cells containing methylated reporter constructs, the mean proportion of high GFP+ in cells not treated with IAA 3.43%. and in cells treated with IAA was 2.93%. There was no significant (p=0.37, paired t-test) difference in the proportion of GFP+ cells expressing a high level of GFP following treatment with IAA in WT HCT116-AID cells (Fig. 4.9B). In DNMT1-

AID cells, the mean proportion of high GFP+ cells without IAA treatment was 10.99%, and following IAA treatment 29.67% of DNMT1-AID cells were high GFP+. There was a significant ($p=1.76x10^{-2}$, paired t-test) increase in the proportion of GFP+ cells expressing a high level of GFP after induction of DNMT1 degradation by IAA. This represents a 2.70-fold increase in high GFP+ cells. These results are in line with the reporter expression data obtained from $DNMT1^{KO}$ cells, and support that loss of DNMT1 activity leads to hypomethylation and increased reporter expression. In UHRF1-AID cells, the mean proportion of high GFP+ without IAA treatment was 15.20% and following IAA treatment was 25.03%. There was a significant 1.65-fold increase ($p=2.68x10^{-2}$, paired t-test) in the proportion of high GFP+ cells following induction of UHRF1 depletion by IAA. This data suggests that depletion of UHRF1 also leads to hypomethylation and activation of integrated reporters in HCT116 cells. Finally, the mean proportion of high GFP+ cells in DNMT1-AID-UHRF1-AID cells without IAA treatment was 17.07% and following IAA treatment was 33.97%. There was a significant 1.99-fold increase ($p=2.00x10^{-3}$) in the proportion of high GFP+ DNMT1-AID-UHRF1-AID cells observed following treatment with IAA.

There were no significant differences between the proportion of high GFP+ cells in DNMT1-AID and UHRF1-AID (p=0.38, paired t-test), DNMT1-AID and DNMT1-AID-UHRF1-AID (p=0.57, paired t-test), and UHRF1-AID and DNMT1-AID-UHRF1-AID cells (p=0.06, paired t-test). However, the mean proportion of high GFP+ cells was largest in DNMT1-AID-UHRF1-AID cells, with the lowest p-value when compared to WT cells. This data may suggest more robust evidence for loss of methylation and activation of reporter expression in DNMT1-AID-UHRF1-AID cells than in DNMT1-AID or UHRF1-AID cells. Notably, higher proportions of high GFP+ cells were observed in all degron cell lines without IAA treatment compared to HCT116 WT cells, suggesting that there may be some leaky degradation leading to incomplete maintenance of methylation in this system.

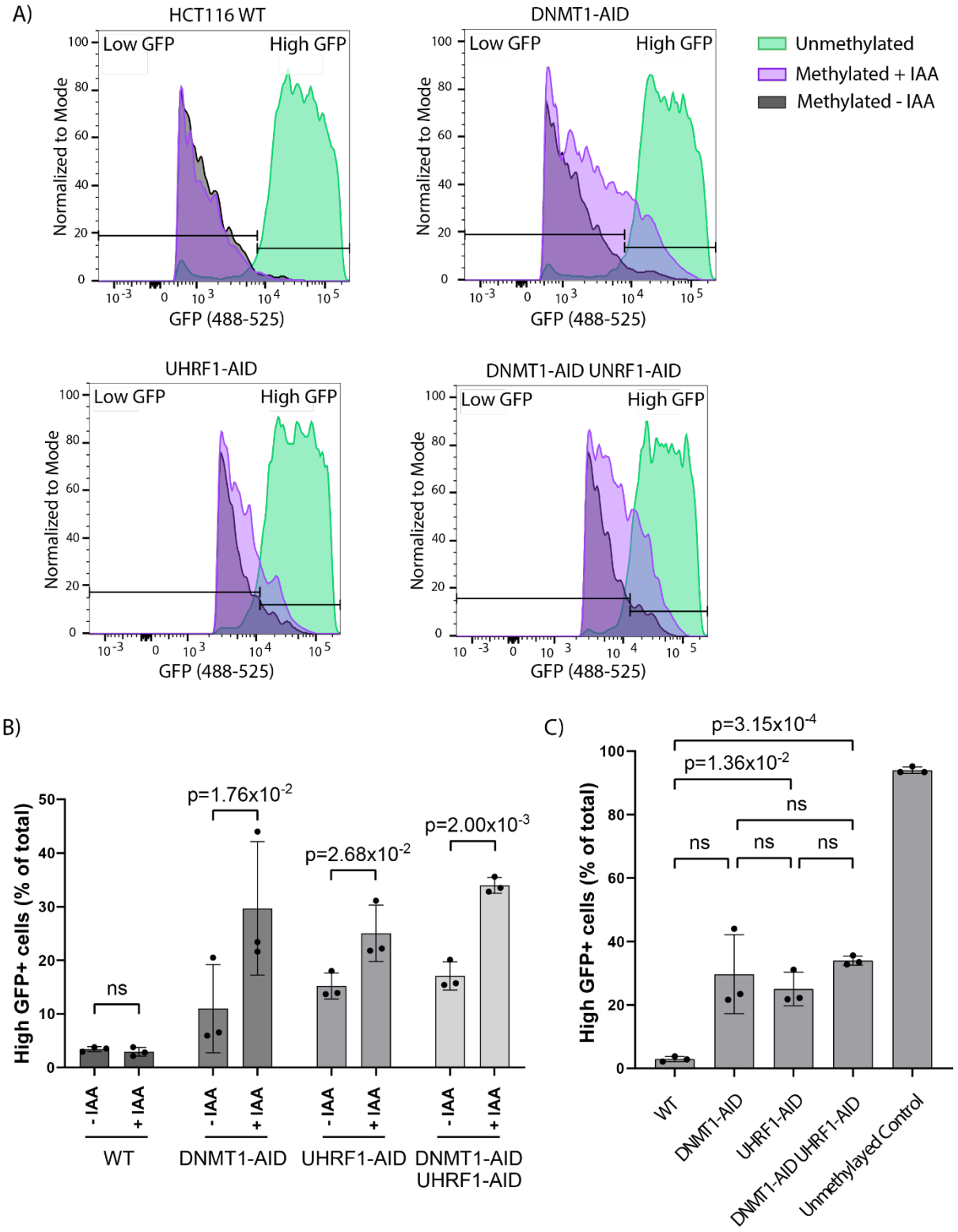


Figure 4-9: Reporter repression is lost following depletion of UHRF1. **A)** *Representative histograms showing fluorescence intensity distributions from GFP+ cells. HCT116 WT cells (top left), DNMT1-AID (top right), UHRF1-AID (bottom left) and DNMT1-AID-UHRF1-AID (bottom right) are shown. Cells contain either methylated reporters treated with IAA (purple), methylated reporters without IAA (grey), or unmethylated reporters (green). Y axis = total count of cells, normalized to the mode. X-axis = GFP fluorescence intensity. Gates for high and low GFP expression are indicated with horizontal bars, and were set so that 95% of the events*

in the unmethylated GFP-positive population fell within the high-GFP gate, based on an internal unmethylated reporter control for each individual cell line. ***B)*** *Proportion of GFP positive cells (parent) that express a high level of GFP, in HCT116 WT, DNMT1-AID, UHRF1-AID and DNMT1-AID-UHRF1-AID cells with or without IAA.* ***C)*** *Proportion of GFP positive cells (parent) that express a high level of GFP, only in cells treated with IAA. Bars represent the mean of n=3 biological replicates. Data points represent the mean value from 100,000 cells analysed for each biological replicate. P-values determined by paired t-tests. Error bars = SD.*

4.10 Discussion

In this chapter, I explored the activity of the methylated reporter system developed in this study in response to changes in global methylation. I investigated expression of the reporter in HCT116 cell lines with disruption of key proteins suggested to be involved in maintaining DNA methylation: the DNA methyltransferases DNMT1 and DNMT3B, and the chromatin-binding protein UHRF1. I aimed to establish whether the methylated reporter construct would become activated in HCT116 cells in response to loss of DNMT1, DNMT3B or UHRF1 activity, and whether changes in reporter expression would correspond to methylation changes at the reporter cassette. I demonstrated that reporter repression was lost following depletion of DNMT1 in both DNMT1^{KO} and DNMT1-AID cell lines. Reporter repression was also lost following depletion of UHRF1 in UHRF1-AID cells. These findings indicate that methylation is maintained at synthetic reporters in HCT116 cells primarily through the activity of DNMT1 and UHRF1. Furthermore, these results demonstrate that eGFPme reports acute changes to global methylation levels, indicating that this system could be utilised as a screening tool, to identify factors responsible for methylation maintenance in cancer cells and to screen for therapeutic targets.

The results I presented in this chapter show that reporter repression was lost in 52.47% of DNMT1^{KO}-eGFPme cells. This represents a 21.16-fold increase in high GFP+ cells compared to HCT116-eGFPme cells. This increase in high GFP+ cells corresponded to reduced methylation levels at the reporter cassette. These results indicate that incomplete maintenance of methylation at the reporter in DNMT1^{KO} cells led to activation of the reporter gene. The finding that loss of DNMT1 activity results in reporter activation was replicated in DNMT1-AID cells following acute depletion of DNMT1. However, in DNMT^{KO} cells, reporter repression was

maintained at the remaining 47.53% and the cassette remained partially methylated. It is interesting that reporter repression and methylation was maintained in nearly half of the cells analysed, despite these cells undergoing ~37 cell divisions. The DNMT1^{KO} cells used in this study were generated through homologous recombination that replaced Dnmt1 exons 3, 4 and 5 with a promoter-less Hygromycin resistance gene (Rhee *et al.*, 2000). Initially, it was thought that this generated a complete knock-out. However, it was later shown that the disruption resulted in the expression of a truncated protein, generated through alternative splicing of the mutated DNMT1 transcript (Egger *et al.*, 2006). The truncated protein retained its catalytic C-terminal domain but lacked part of the DMAP1 interaction domain and proliferating cell nuclear antigen (PCNA)-binding domain (Egger *et al.*, 2006; Spada *et al.*, 2007). The truncated protein displayed lower methyltransferase activity and reduced expression compared to wild-type DNMT1, but retained its catalytic activity (Rhee *et al.*, 2000; Egger *et al.*, 2006; Spada *et al.*, 2007). The DNMT1^{KO} cells express the truncated DNMT1 protein at approximately ~20% of DNMT1 WT levels and were reported to retain ~80% global m^{5}C content (Rhee *et al.*, 2000; Egger *et al.*, 2006). Therefore, this may explain the maintenance of methylation observed in DNMT1^{KO} cells.

There were no detectable differences in reporter expression or maintenance of methylation at reporters stably integrated into DNMT3B^{KO} cells. Disruption of DNMT3B in DNMT3B^{KO} cells was achieved through homologous recombination was used to replace exons 2 to 21, resulting in a deletion that removed the catalytic active site (Rhee *et al.*, 2002; Egger *et al.*, 2006). Loss of DNMT3B was reported to result in only minor changes to global m^{5}C content, with DNMT3B^{KO} cells retaining ~97% methylation levels of HCT116 WT cells (Rhee *et al.*, 2002). Therefore, my results are in line with these findings and support that disruption of DNMT3B alone does not result in substantial differences in methylation maintenance.

However, this is in contrast to work that has suggested a role for DNMT3B in methylation maintenance. The maintenance activity of DNMT1 can be imprecise (Liao *et al.*, 2015; Ming *et al.*, 2020), and DNMT3B (along with DNMT3A) has been evidenced to collaborate with DNMT1 and compensate for inefficient maintenance by DNMT1 (Rhee *et al.*, 2002; Liao *et al.*, 2015; Elliott, Sheaffer and Kaestner, 2016). DNMT3B is upregulated in DNMT1^{KO} mice and compensates for lack of methylation maintenance by DNMT1, which was required for animal survival (Elliott, Sheaffer and Kaestner, 2016). Recently, a CRISPR-based genetic knock-out

screen identified DNMT3B (along with DNMT1, DNMT3A, DNMT3L and UHRF1) as being required for both the establishment and maintenance of epigenetic states in mouse ESCs (Tatarakis, Saini and Moazed, 2023). However, DNMT3B maintenance activity has been evidenced to be directed towards specific sequences, such as LINE-1 elements in mouse ESCs (Liang *et al.*, 2002). Therefore, it is possible that I did not obtain sufficient reporters within regions targeted by DNMT3B to observe an effect. Furthermore, most studies that have suggested a role for DNMT3B in methylation maintenance have only observed this effect in combination with loss of DNMT1 (Rhee *et al.*, 2002; Elliott, Sheaffer and Kaestner, 2016). Therefore, DNMT1 is likely able to compensate for loss of DNMT3B, which could explain my results.

I expected to observe a higher degree of reporter activation and hypomethylation in DKO cells compared to DNMT1^{KO} cells, due to the additive effect of loss of DNMT1 and DNMT3B activity reported by Rhee *et al*. (Rhee *et al.*, 2002). In DKO cells, an 8.60-fold increase in the proportion of high GFP+ were observed when compared to HCT116 WT, and reporter repression was lost in 21.33% of cells. Unexpectedly, the proportion of high GFP+ cells was lower in DKO cells than in DNMT1^{KO} cells. Methylation of the reporter cassette was well-maintained in DKO cells, with a 1.30-fold higher %mCpG observed in the enhancer region when compared to WT cells. These results do not align with what was expected, based the reporter expression data, and may suggest that the reporter escaped repression through a mechanism unrelated to methylation at the enhancer or gene body region. For example, methylation at the promoter region may be the mechanism underlying methylation-induced repression of the transgene, as suggested by Schübeler *et al.* (Schübeler *et al.*, 2000).

The DKO cells used in this study were generated by knock-out of DNMT3B, followed by introduction of the same disruption of DNMT1 as in DNMT1^{KO} cells, as described above (Rhee *et al.*, 2000, 2002). Rhee *et al*. (2002) generated three DKO cell lines and the DKO cells used in my study are DKO8. The DKO8 cells retained ~42.5% of WT m^5C content (Rhee *et al.*, 2002). DKO8 cells were shown to express the truncated DNMT1 transcript at ~40% the level of WT DNMT1, 2-fold higher than the DNMT^{KO} cell line (Egger *et al.*, 2006). Spada *et al.* (2007) suggested that the residual activity of the truncated DNMT1 transcript was responsible for the majority of methyltransferase function in DNMT1^{KO} and DKO cells. Thus, the increased

expression of the truncated DNMT1 protein may account for the improved methylation maintenance observed in DKO cells.

Additionally, DKO cells may have adapted other means to maintain DNA methylation, in the absence of fully functional DNMT1 and DNMT3B. The canonical *de novo* methyltransferase DNMT3A has been suggested to have methylation maintenance activity. Like DNMT3B, DNMT3A has been evidenced to have a role in compensating for incomplete DNA methylation maintenance by DNMT1 (Liao *et al.*, 2015), particularly at repetitive elements (Liang *et al.*, 2002). DNMT3A and DNMT3B are largely redundant in the regions that they target, and these target regions only lose methylation when both proteins are inactive (Liao *et al.*, 2015). This indicates that DNMT3A is able to compensation for loss of DNMT3B. Therefore, the activity of DNMT3A may compensate for reduced DNMT1 and DNMT3B activity in DKO cells.

UHRF1 plays a crucial role in the maintenance of DNA methylation through interaction with DNMT1. Furthermore, UHRF1 has increasingly been recognised for its non-canonical roles, including interaction with DNMT3s, and its involvement in diseases such as cancer (Newkirk and An, 2020; Kostyrko *et al.*, 2023; Yamaguchi *et al.*, 2024). I assessed reporter expression following depletion of UHRF1. Depletion of UHRF1 in UHRF1-AID cells resulted in a statistically significant 1.65-fold increase in high GFP+ cells, indicating that depletion of UHRF1 led to loss of reporter repression. These data support that UHRF1 also has an important role in maintaining DNA methylation in HCT116 cells, as previously reported (Bostick *et al.*, 2007; Nishiyama *et al.*, 2020; Yamaguchi *et al.*, 2023). One caveat of this experiment is that reporter expression was assessed after only 48 hours of protein depletion. Yamaguchi et al. (2024) showed that DNA methylation levels decrease continuously in cells depleted of UHRF1 and/or DNMT1 up to 12 days after initiating protein degradation (Yamaguchi *et al.*, 2024). Therefore, it would be of interest to track reporter expression over a longer time period, to learn more about the dynamics of reporter activation in response to UHRF1 depletion.

Together, the results presented in this chapter demonstrate that the eGFPme reporter is activated in response to acute changes in DNA methylation levels. Furthermore, my findings suggest that there may be other enzymes capable of maintaining DNA methylation in cancer cells. A potential future application of the methylation-sensitive reporter system developed in this study could be its use as a cell-based screening assay. CRISPR knock-out screens employ CRISPR-Cas9-directed genome editing, with a library of guide RNAs targeting different genes,

to introduce perturbations into a pool of cells. The effects of these perturbations are then evaluated through different methods, in order to identify genes that are involved in the biological pathway of interest (Bock *et al.*, 2022; Tatarakis, Saini and Moazed, 2023). This methylation-sensitive reporter could be utilised in a CRISPR-based genetic screen, to identify genes with a role in maintaining DNA methylation and gene repression. Furthermore, identification of novel factors involved in DNA methylation maintenance could lead to the discovery to important drug targets for cancer therapy.

5 Methylation-sensitive reporter system as drug screening platform

5.1 Introduction

In Chapter 4, I demonstrated that the methylated eGFP reporter was activated in response to hypomethylation in DNMT1^{KO} cells, and following depletion of DNMT1 and UHRF1 in HCT116-AID cell lines. This indicates that the eGFPme reporter is responsive to changes in methylation and that targeting DNMT1 and UHRF1 is an effective method to induce hypomethylation in HCT116 cells. Given the widespread changes to DNA methylation that occur in cancer, targeting DNA methylation is an attractive strategy for cancer treatment. Hypomethylating agents (HMAs) have shown clinical benefits, but have limited applicability due to their toxicity (Issa and Kantarjian, 2009; Connor *et al.*, 2023). Therefore, identifying new HMAs with reduced cytotoxicity is important to improve therapeutic avenues in cancer treatment. In this chapter, I aimed to explore whether the methylation-sensitive reporter system developed during my PhD could be utilised as a phenotypic drug screening tool, to identify novel HMAs.

Phenotypic screening is an approach used in research and drug discovery, to evaluate compounds based on their effect on observable phenotypes, without predefined molecular targets. This approach is particularly beneficial for identifying drugs that modulate complex biological processes and disease phenotypes, as it captures the response of biological systems to treatments (Moffat *et al.*, 2017; Vincent *et al.*, 2022). Phenotypic screening offers advantages in identifying novel mechanisms of action and potential therapeutic targets that might be otherwise overlooked in target-based approaches (Swinney and Anthony, 2011).

The hypomethylating compounds 5-azacytidine (5-aza) and 5-aza-2'-deoxycytidine (5-aza-dC) are cytidine analogues. These compounds were first synthesised in the 1960s, and when tested against cancer cells were shown to have a range of anti-metabolic activities (Píškala and Šorm, 1964). It was later shown that these cytidine analogues are incorporated into DNA, where they covalently bind and trap DNMT1, DNMT3A and DNMT3B, thus preventing DNA methylation and leading to passive hypomethylation (Jüttermann, Li and Jaenisch, 1994). As such, 5-aza and 5-aza-dC have been widely used as HMAs, both in research and in the clinic. In the clinic, 5-aza and 5-aza-dC have been particularly effective in enhancing responses to

immunotherapies (Banerjee *et al.*, 2023). This is thought to be due to hypomethylation triggering an antiviral interferon response (Chiappinelli *et al.*, 2015; Ohtani *et al.*, 2020). However, these compounds are irreversibility incorporated into DNA, resulting in dose-limiting toxicities in patients. Furthermore, 5-aza and 5-aza-dC exhibit poor chemical stability and pharmacokinetics, making delivery to solid tumours inefficient (Stresemann and Lyko, 2008). These properties have limited the use of cytidine analogues as chemotherapeutic agents and highlight the need for improved HMAs.

Recently, GlaxoSmithKline (GSK) reported the discovery of two highly-selective small molecule inhibitors of DNMT1, GSK-3484862 and GSK-3685032 (Pappalardi *et al.*, 2021). Importantly, these compounds were chemically stable, did not covalently bind to DNMT1 or non-specifically bind to DNA, and induced rapid and reversible hypomethylation. Therefore, these DNMT1-specific small molecule inhibitors may be preferable for clinical use, over cytosine analogous.

DNA hypomethylation can alternatively be achieved by targeting other components of the DNA methylation machinery, such as UHRF1. In Chapter 4, I showed that depletion of UHRF1 resulted in activation of the methylated eGFP reporter. Depletion of UHRF1, in AID-tagged cells and through RNA interference, has been shown to result in global hypomethylation and an interferon response comparable to 5-aza and 5-aza-dC (Irwin *et al.*, 2023; Yamaguchi *et al.*, 2023). In recent years, UHRF1 has increasingly been recognised to have an important role in epigenetic regulation and disease pathology. Importantly, UHRF1 can act as an oncogene (Ashraf *et al.*, 2017). Upregulation of UHRF1 has been observed in a variety of human cancers, including CRC, lung and breast cancer. Overexpression of UHRF1 has been linked to poor prognosis in AML and CRC patients (X. Kong *et al.*, 2019; Xue *et al.*, 2019; Reardon *et al.*, 2021; Hu *et al.*, 2022; Wang *et al.*, 2023). Therefore, UHRF1 is a therapeutic target of interest and inhibiting UHRF1 has been proposed as a viable treatment avenue (Reardon *et al.*, 2021). There are currently no clinically approved UHRF1 inhibitors. Targeting UHRF1 has proved challenging due to its primary functions occurring via protein–protein or protein-DNA interactions, which are difficult to target therapeutically (Chang *et al.*, 2021). Thus, it is of interest to identify cell-active, potent inhibitors of UHRF1.

In this chapter, I investigated whether the methylation-responsive reporter system developed during my PhD could be used as a drug screening tool to identify HMAs. To do this, I tested

the system using proposed inhibitors of DNMT1 and UHRF1. The methylated eGFP reporter was activated in HCT116 cells following pharmacological inhibition of DNMT1 with GSK-3484862, providing proof-of-concept for this idea. Three proposed inhibitors of UHRF1, NSC232005, NSC34716 and NSC43513 (Myrianthopoulos *et al.*, 2016), were screened using this approach but did not result in reporter activation. I further characterised the response of HCT116 cells to GSK-3484862. Reporter activation coincided with global hypomethylation and increased expression of human endogenous retrovirus 1 (HERV1) elements and antiviral interferon signalling. Together, these findings support that the eGFPme reporter is reliably activated in response to hypomethylation, and that global hypomethylation is consistently accompanied by an interferon response in HCT116 cells. I therefore propose that this reporter system could be utilised to efficiently identify novel cell-active hypomethylating compounds with therapeutic potential.

5.2 Characterising GSK-3484862 hypomethylating activity in HCT116 cells

In order to provide proof-of-concept for use of the eGFPme reporter system as a drug screening platform to identify HMAs, I sought to test whether the reporter would be activated in response to pharmacological inhibition of DNMT1. Having established in Chapter 4 that the reporter was activated in DNMT1^{KO} cells and following acute depletion of DNMT1 in DNMT1-AID cells, I hypothesised that I would observe reporter activation after treating cells with a DNMT1 inhibitor. I chose to test the system using the small molecule DNMT1 inhibitor GSK-3484862, as this compound is reported to be a less cytotoxic and more specific to DNMT1 compared to nucleoside analogues such as 5-Aza and 5-Aza-dC (Pappalardi *et al.*, 2021). At the time of this study, GSK-3484862 was the only commercially available compound of the two non-nucleoside inhibitors developed by Pappalardi *et al.* (Pappalardi *et al.*, 2021). The GSK-3484862 inhibitor was kindly gifted by Katie Pickup (Meehan Lab). The activity of GSK-3484862 has yet to be extensively characterised in HCT116 cells. Pappalardi et al. (2021) reported that biochemical engagement of DNMT1 was dose-dependent, and maximum inhibition of DNMT1 was achieved at ~10μM, as assessed by orthogonal fluorescence-coupled breaklight assay. In HCT116 cells, hypomethylation by GSK-3484862 was also dose-dependent, with maximal loss of DNA methylation at the promoter of vimentin observed at a concentration of ~0.5μM. Pappalardi *et al*. (2021) did not report any further characterisation of GSK-3484862 activity in HCT116 cells. Therefore, I first aimed to determine the optimal dose and hypomethylating activity of GSK-3484862 in HCT116 cells.

5.2.1 GSK-3484862 slows growth, but does not alter the cell cycle profile, of HCT116 cells

I assessed the growth of HCT116 cells treated with 0.5μM, 2μM or 10μM GSK-3484862. I chose these concentrations based on the dose-dependent activity of GSK-3484862 in biochemical assays and in cells, as described by Pappalardi *et al*. (2021). HCT116 cells were seeded and treated with GSK-3484862 or DMSO (vehicle control). I replenished media and compound every 24 hours and counted cells after 48 and 96 hours of treatment. After 48

hours of GSK-3484862 treatment, no substantial differences in cell count were observed at any of the tested drug concentrations. After 96 hours, HCT116 cell count was ~2.43-fold lower in cells treated with 10µM GSK-3484862, compared to DMSO control cells (Fig. 5.1A). Cell count was ~1.40-fold lower in cells treated with 2µM GSK-3484862, compared to DMSO control cells. Cells treated with 0.5µM GSK-3484862 did not display a substantially reduced cell count. These data indicate that >2µM GSK-3484862 may slow the growth of HCT116 cells.

I next asked whether GSK-3484862 caused cell cycle dysregulation, which could explain the observed growth effects. I assessed the cell cycle profile of HCT116 cells treated with GSK-3484862 by flow cytometry. The distribution of cells at different stages of the cell cycle can be detected by flow cytometry using DNA-specific fluorescent dyes, such as DAPI. The intensity of fluorescence corresponds to DNA content and the resulting profile can be used to accurately identify the major cell cycle components; G0/G1, S-phase, G2/M (Kim and Sederstrom, 2015). I treated HCT116 cells with 0.5µM, 2µM or 10µM GSK-3484862, or DMSO, for 96 hours then fixed and stained cells with DAPI before quantifying DAPI intensity by flow cytometry (Fig. 5.1B-C). The proportion of cells within G0/G1, S-phase and G2/M were estimated using the Watson Pragmatic model (Watson, Chambers and Smith, 1987). No significant ($p= 9.88\times10^{-1}$) differences in the proportion of cells within each phase of the cell cycle were detected in HCT116 cells treated with GSK-3484862, compared with DMSO control cells. These results indicate that GSK-3484862 does not alter cell cycle progression in HCT116 cells.

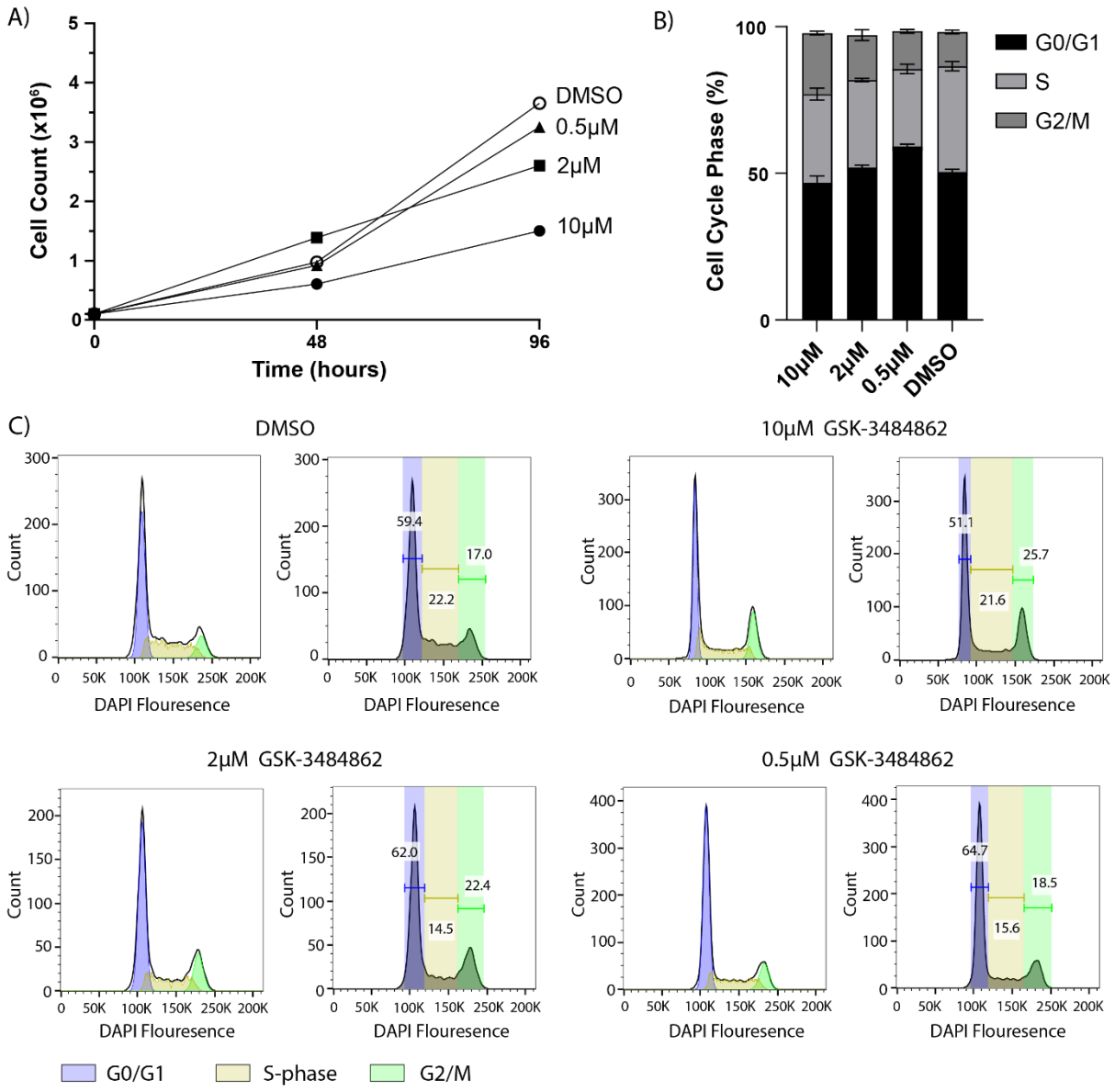


Figure 5-1: GSK-3484862 slows HCT116 cell growth but does not alter cell cycle profile. A) *Growth curve line plot showing the cell count for HCT116 cells treated with GSK-3484862 or DMSO. Cells counted at seeding and after 48 and 96 hours of treatment. Points represent the mean of n=2 biological replicates.* ***B)*** *Stacked bar plot showing the proportion of cells within G0/G1, S-phase and G2/M, in HCT116 cells treated with GSK-3484862, or DMSO. Cell cycle profile estimated using the Watson model. Bars represent the mean of n=3 biological replicates. 2-Way ANOVA, p=0.99.* ***C)*** *Representative flow cytometry data showing the cell cycle profile for HCT116 cells treated with GSK-3484862 or DMSO. Bars represent gates for G0/G1 (blue), S-phase (yellow) and G2/M (green), as estimated by the Watson model. The percentage of the cell population within each cell cycle phase is shown. Representation flow cytometry data from a single biological replicate.*

5.2.2 GSK-3484862 induces hypomethylation of Human Satellite II sequences

Next, I examined DNA hypomethylation induced by GSK-3484862 at Human Satellite II (SatII) repeat sequences. Repetitive elements and retrotransposons are normally methylated in human cells, which is important to repress their activity and maintain genome stability (Rowe and Trono, 2011; Pappalardo and Barra, 2021). I measured hypomethylation at SatII sequences using a methylation-sensitive Southern blot (Taglini *et al.*, 2024). To do this, I isolated genomic DNA from HCT116 cells that were treated with GSK-3484862 (0.5µM, 2µM or 10µM) or DMSO for 120 hours. I digested DNA with the methylation-sensitive restriction enzyme, BstBI. BstBI recognises the sequence 5'-TT/CGAA-3', which occurs within the SatII repeat motif. BstBI is unable to cleave methylated CpG sites, therefore the degree of DNA digestion is proportional to the degree of DNA methylation. I detected SatII fragments using a fluorescently-labelled SatII probe. SatII sequences were digested at a similar level in HCT116 cells treated with 0.5µM, 2µM or 10µM GSK-3484862 (Fig. 5.2). In cells treated with DMSO, DNA was more protected from digestion. These results suggest that GSK-3484862 induces hypomethylation of SatII repeat sequences in HCT116 cells.

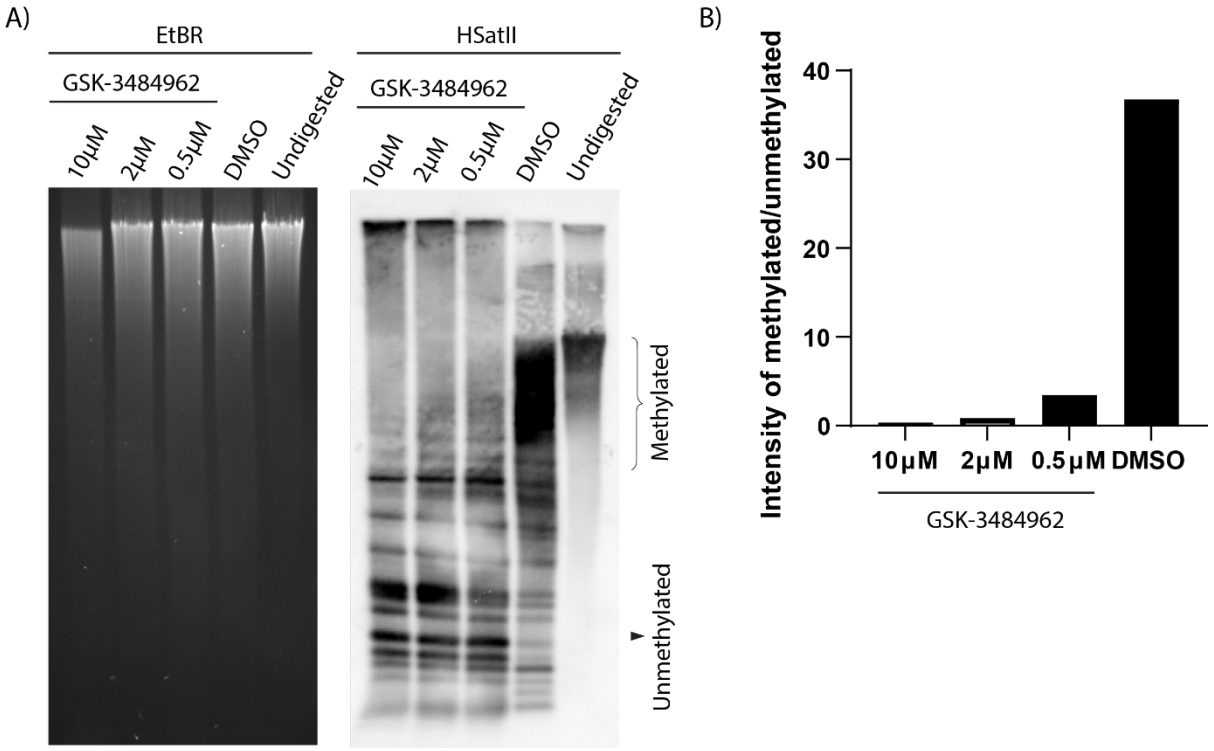


Figure 5-2 GSK-3484862 induces hypomethylation of SatII repeat sequences in HCT116 cells. ***A)*** *Methylation-sensitive southern blot showing digestion of human satellite II sequences by*

BstBI in HCT116 cells treated with GSK-3484862 or DMSO. Ethidium bromide (EtBr)-stained gel shown as loading control (left). Southern blot hybridisation probed with SatII DIG-labelled probe (right). Undigested, methylated regions and digested, hypomethylated regions are indicated. ***B)*** *Bar plot showing signal quantification of the Southern blot using the ratio of methylated over unmethylated regions. Signal intensity quantified using the Gel Analyser tool in ImageJ.*

5.2.3 GSK-3484862 induces hypomethylation within two cell divisions

I next asked whether hypomethylation at SatII sequences was observable at an earlier time point. In Chapter 4, I showed that acute depletion of DNMT1 and UHRF1 resulted in activation of the methylated reporter gene within 48 hours. HCT116 cells have a double time of ~18 hours, therefore, reporter activation was observed after approximately two cell divisions. I therefore hypothesised that inhibition of DNMT1 would cause DNA hypomethylation within 48 hours. To test this, I treated HCT116 cells with GSK-3484862 (0.5µM, 2µM or 10µM) or DMSO for either 48 or 120 hours, and isolated genomic DNA from cells. Again, I carried out a methylation-sensitive Southern blot by digesting DNA with BstBI and probing for SatII fragments. I compared digestion of treated HCT116 samples with digestion of DNMT1^{KO}, DNMT3B^{KO} and DKO cell lines, to determine whether the degree of hypomethylation obtained with GSK-3484862 was comparable to the genetic knock-out.

As shown in Figure 5.3, DMSO-treated cells were more protected from digestion by BstBI than cells treated with GSK-3484862. The extent of hypomethylation, indicated by the decrease in the intensity of large DNA fragments and an increase in smaller, digested fragments, was comparable between the 48- and 120-hour time points. This suggests that hypomethylation is induced within 2 cell divisions, and that the degree of hypomethylation obtained does not increase over time. The degree of hypomethylation obtained was also similar between all GSK-3484862 concentrations tested. Therefore, for future experiments I chose to work with 0.5µM GSK-3484862. Cells treated with GSK-3484862 appear to be more highly methylated than DNMT3B^{KO} cells, with a similar level of hypomethylation as DNMT1^{KO} and DKO cells (Fig. 5.3). This suggests that acute depletion of DNMT1 activity through pharmacological inhibition is an efficient method to induce hypomethylation in HCT116 cells.

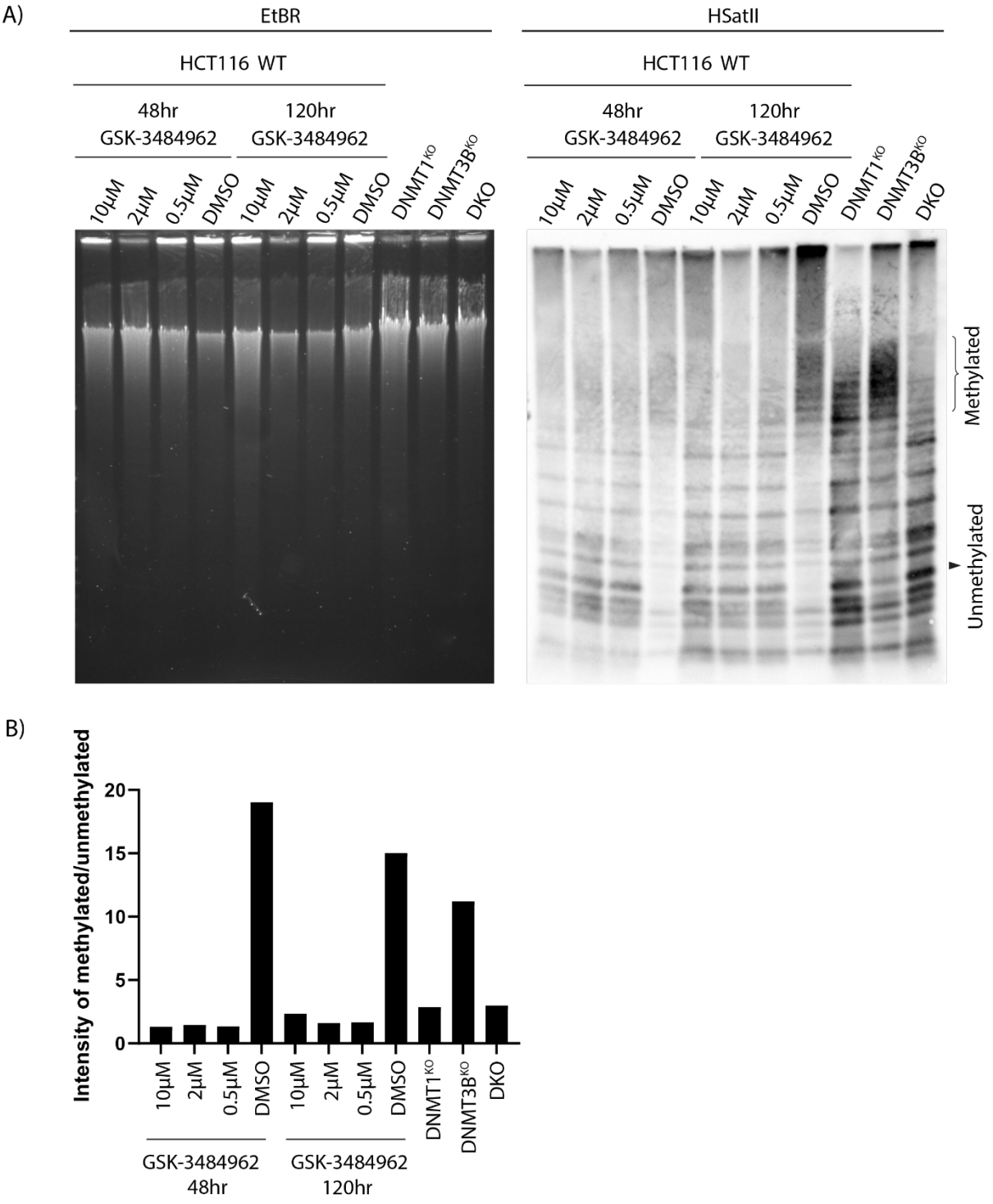


Figure 5-3: GSK-3484862 induces hypomethylation within two cell divisions. A) *Methylation-sensitive southern blot showing digestion of human satellite II sequences by BstBI in HCT116 cells treated with GSK-3484862 (0.5μM, 2μM or 10μM) or DMSO after 48-hours or 120-hours exposure. DNMT1^{KO}, DNMT3B^{KO} and DKO cells also shown. Ethidium bromide (EtBr)-stained gel shown as loading control (left). Southern blot hybridisation probed with SatII DIG-labelled probe (right). Undigested, methylated regions and digested, hypomethylated regions are indicated.* ***B)*** *Bar plot showing signal quantification of the Southern blot using the ratio of methylated over unmethylated regions. Signal intensity quantified using the Gel Analyser tool in ImageJ.*

5.2.4 GSK-3484862 induces hypomethylation of LINE1 sequences

To further examine the degree of hypomethylation induced by GSK-3484862 in HCT116 cells, I assessed methylation of Long Interspaced Nuclear Element 1 (LINE1) sequences by methylation-sensitive Southern blot. Similar to SatII sequences, LINE1 sequences are generally methylated in human cells (Rowe and Trono, 2011). Measuring hypomethylation at LINE1 sequences has previous been used as a proxy method to infer genome-wide methylation changes (Yang *et al.*, 2004). Having established in that similar degrees of hypomethylation were obtained with 0.5µM, 2µM or 10µM GSK-3484862, I chose to use 0.5µM for subsequent experiments, in order to minimise impact on HCT116 cell growth. HCT116 cells were treated with 0.5µM GSK-3484862 or DMSO for either 48- or 120-hours and genomic DNA was isolated from cells. Here, the restriction enzyme MspI and its isoschizomer HpaII were used to digest DNA. The MspI and HpaII enzymes both cleave the 5'-CCGG-3' recognition sequence, but HpaII digestion is blocked by methylation of the internal CpG (Waalwijk and Flavell, 1978). Digestion of LINE1 sequences using MspI and HpaII is an established technique to assess methylation of these elements (Florl *et al.*, 1999). LINE1 sequences were detected using fluorescently labelled LINE1 probe, prepared and gifted by Mihail Peychev (Meehan Lab). I compared digestion of treated HCT116 cells to DNMT1^{KO}, DNMT3B^{KO} and DKO cell lines.

As shown in Figure 5.4, GSK-3484862-treated HCT116 cells were less protected from HpaII digestion at LINE1 sequences than DMSO-treated cells. Quantification of signal intensity for digested bands demonstrated that slightly more digestion was found at 120 hours than 48 hours. This suggests that LINE1 sequences may be more hypomethylated after 120 hours exposure to GSK-3484862. However, overall, digestion patterns are similar between the 48- and 120-hour time points. Again, the degree of digestion in HCT116 GSK-3484862-treated cells was comparable to DNMT1^{KO} and DKO cell lines. These data indicate that 0.5µM GSK-34848623 induces hypomethylation of LINE1 sequences within 48 hours, or two cell divisions.

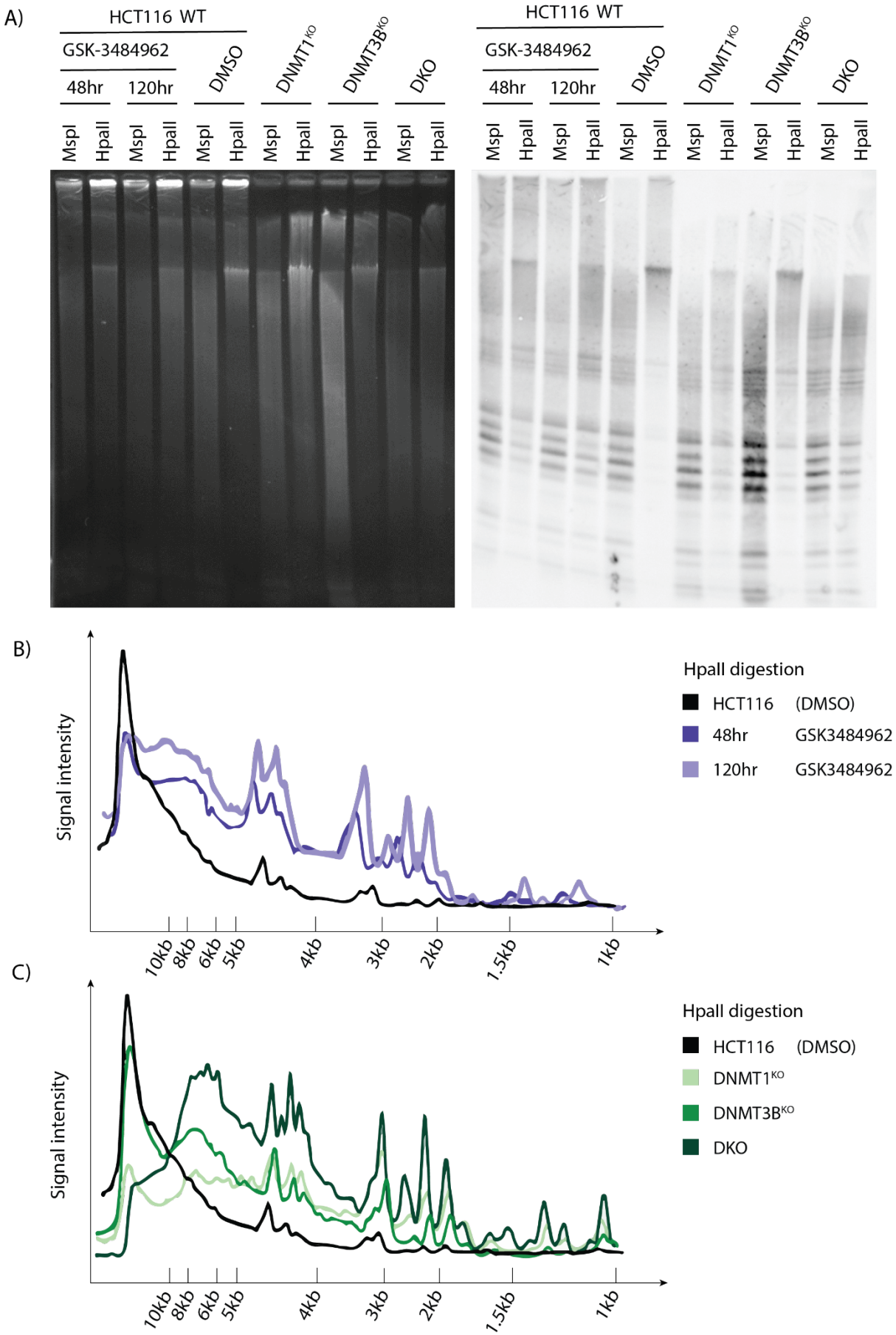

A)
HCT116 WT
GSK-3484962
48hr
120hr
DMSO
DNMT1KO
DNMT3BKO
DKO
MspI
HpaII
B)
Signal intensity
Hpall digestion
HCT116 (DMSO)
48hr GSK3484962
120hr GSK3484962
10kb
8kb
6kb
5kb
4kb
3kb
2kb
1.5kb
1kb
C)
Signal intensity
Hpall digestion
HCT116 (DMSO)
DNMT1KO
DNMT3BKO
DKO

Figure 5-4: GSK-3484862 induces hypomethylation of LINE1 sequences. A) *Methylation-sensitive southern blot showing digestion of LINE1 sequences by MspI and HpaII in HCT116 cells treated with 0.5µM GSK-3484862 or DMSO after 48-hours or 120-hours exposure. DNMT1^{KO}, DNMT3B^{KO} and DKO cells also shown. Ethidium bromide (EtBr)-stained gel shown as loading control (left). Southern blot hybridised with probe against LINE1 sequences (right).* ***B)*** *Scan profile plot showing signal quantification across lanes for HpaII-digested DNA. Signal intensity was quantified in ImageJ from the top (left) to bottom (right) of the lane, using the Gel Analyser tool.*

5.3 Reporter repression is lost following treatment with GSK-3484862

Having established that GSK-3484862 induces hypomethylation in HCT116 cells, I next sought to test the response of my methylation-sensitive reporter to GSK-3484862 treatment. I utilised HCT116 cells containing stably integrated, methylated reporter constructs (HCT116-eGFPme cells) that were generated in Chapter 3. To obtain a comprehensive picture of reporter activity, I conducted a 120-hour time course treatment with GSK-3484862 (0.5µM) or DMSO and assessed GFP+ cells at 24-hour intervals by flow cytometry (Fig. 5.5). I hypothesised that I would observe reporter activation within 48 hours, based on the reporter activation results in HCT116-AID cells from Chapter 4 and the evidence for hypomethylation at SatII and LINE1 sequences at 48 hours, presented earlier in this Chapter.

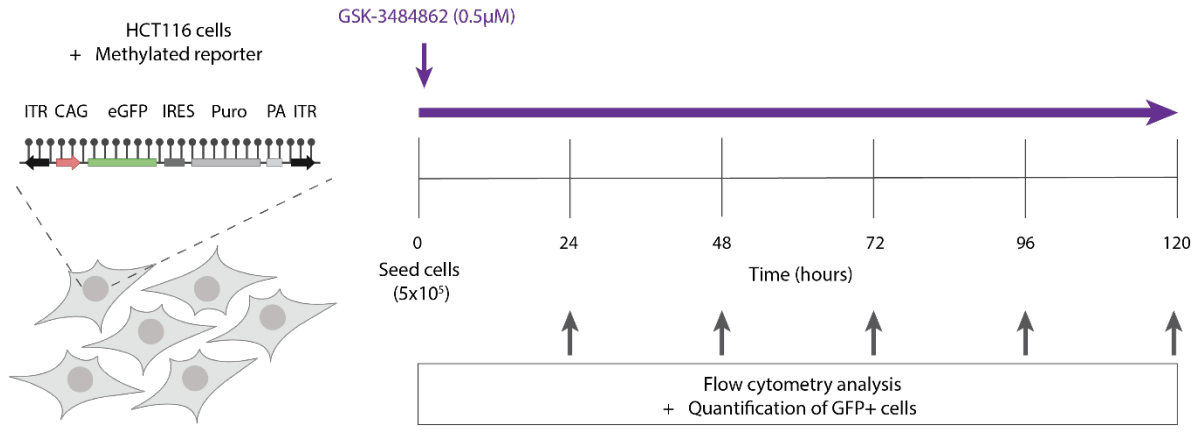


Figure 5-5: Experimental schematic of GSK-3484862 120-hour treatment time course in HCT116-eGFPme cells. *HCT116 cells containing the stably integrated methylated eGFP reporter construct were seeded (5x10^5 cells) and treated with 0.5µM GSK-3484862. Media and*

inhibitor were replenished every 24 hours. Flow cytometry analysis of GFP+ cells was conducted at 24-hour intervals to assess reporter expression.

In HCT116-eGFPme cells treated with GSK-3484862, the mean proportion of high GFP+ cells after 24-hours increased slightly from 4.35% in DMSO-treated cells to 10.82% in GSK-3484862 (Fig. 5.6). However, no significant difference (paired t-test, $p=7.19\times10^{-2}$) in the proportion of high GFP+ was detected after 24 hours of treatment, compared to DMSO cells. At 48 hours GSK-3484862 exposure the mean proportion of high GFP+ cells was 47.57%, a significant (paired t-test, $p=8.29\times10^{-3}$) increase compared to DMSO-treated cells. The proportion of high GFP+ continued to increase in GSK-3484862-treated cells, to 72.77% at 72 hours, 81.13% at 96 hours, and began to plateau at 82.90% at 120 hours. Increased GFP expression was visible under the fluorescent microscope at 48 hours and was not visually increased at 120 hours. Analysis of the distribution of fluorescence intensities by flow cytometry cells showed that GSK-3484862-treated cells had intermediate fluorescence intensity at 48-hours, compared to the methylated (DMSO) and unmethylated controls. At 120 hours, the distribution of fluorescence intensities in GSK-3484862-treated cells overlapped with that of the unmethylated HCT116-eGFP control cells. Taken together, these data demonstrate that repression of the methylated reporter construct was lost in response to GSK-3484862 treatment.

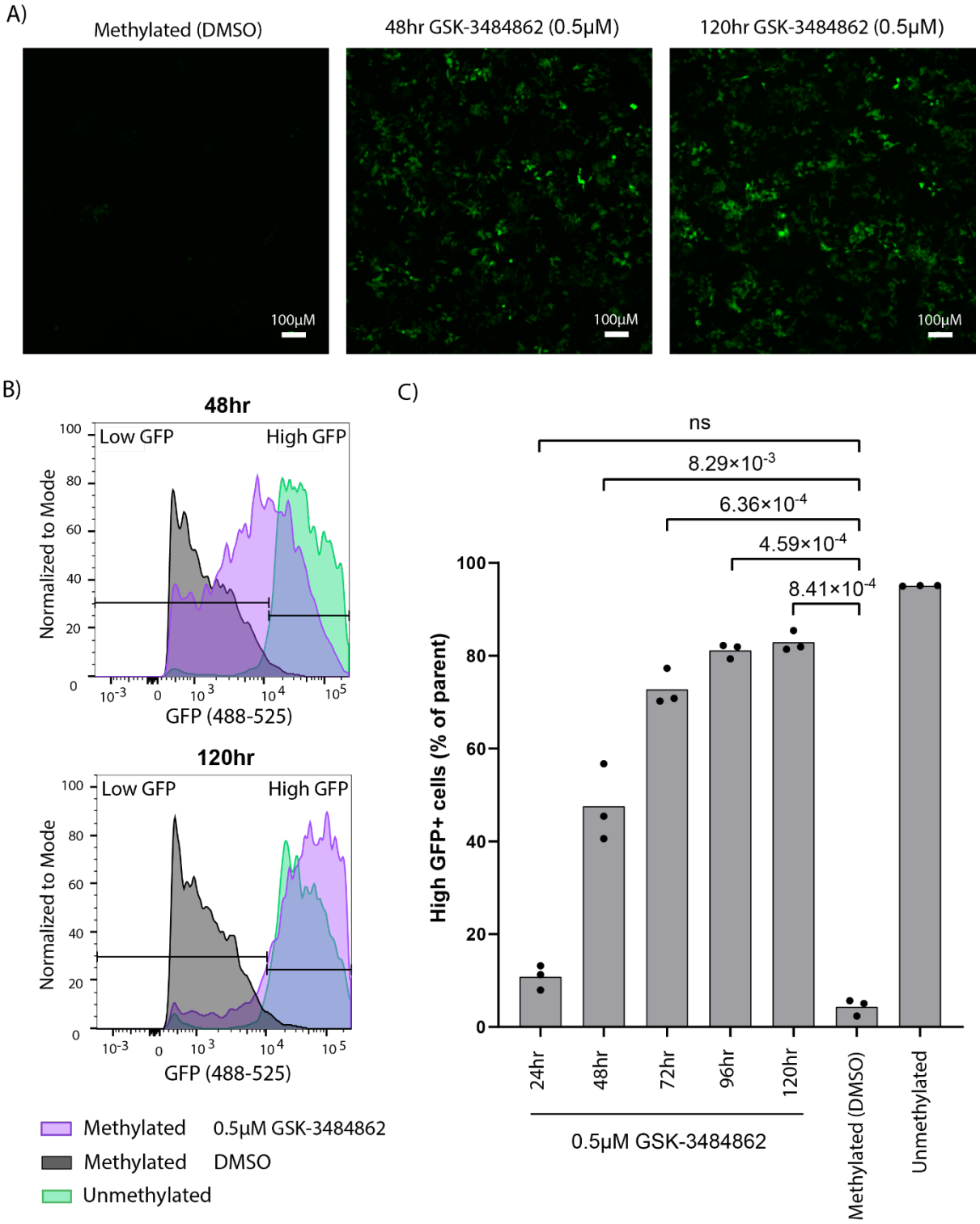


Figure 5-6: Reporter repression is lost following treatment with GSK-3484862. A) *Representative fluorescence microscopy images of HCT116-eGFPme cells treated with DMSO or 0.5µM GSK-3484862 for 48- or 120-hours. Scale bar = 100µM* ***B)*** *Representative fluorescence intensity distribution histogram, showing intensity of eGFP expression from HCT116-eGFPme cells treated with DMSO (grey) or GSK-3484862N (purple) compared to unmethylated HCT116-eGFP cells (green). Horizontal bars represent the high and low GFP gates. Gates for high and low GFP expression are indicated and were set so that 95% of the events in the unmethylated GFP-positive population fell within the high-GFP gate. Y axis =*

proportion of cells, normalized to the mode. X-axis = GFP fluorescence intensity. ***C)*** *Quantification of high GFP+ cells shown as a percentage of all GFP+ cells. 120-hour GSK-3484862 treatment time course, in HCT116-eGFPme cells treated with 0.5µM GSK-3484862 or DMSO, compared to unmethylated HCT116-eGFP control cells. Bars represent the mean of n=3 biological replicates. Data points represent the mean value from 1x10^5 cells measured for each individual replicate Paired t-tests, p-values shown.*

5.4 Reporter methylation is reduced following treated with GSK-3484862

I next asked whether the reporter cassette lost methylation following treatment with GSK-3484862. To do this, I isolated genomic DNA samples from cells during the 120-hour GSK-3484862 time-course experiment. I conducted high-throughput EM-sequencing targeting the CMV enhancer and eGFP gene body regions of the reporter cassette in cells treated for 48- and 120-hours. As shown in Figure 5.7, the median percentage of methylated CpG sites (%mCpG) at the CMV enhancer in cells treated with GSK-3484862 for 48 hours was 13.33%, and at the eGFP gene body 54.76%. This was substantially lower than the median %mCpG in DMSO-treated cells at this time point, which was 73.34% at the enhancer and 100.00% at the gene body. After 120-hours GSK-3484862 treatment, the median %mCpG at the enhancer was 0.00% and at the gene body 42.86%. In DMSO-treated cells, some unexpected demethylation was observed. The median %mCpG in DMSO-treated cells was 36.67% at the enhancer and 50.00% and the gene body at the 120-hour time point. Hypomethylation in DMSO treated cells was only observed in one experimental replicate that was sequenced, and the other replicate retained high methylation levels at the reporter cassette. No differences in reporter expression were observed in any DMSO-treated cells, shown in Figure 5.6. It would be of interest to repeat this experiment and sequencing of these samples, to determine whether this result represents genuine hypomethylation in DMSO-treated cells.

Overall, these results indicate that GSK-3484862 induced hypomethylation at stably integrated eGFP reporters in HCT116 cells. The degree of hypomethylation increased between 48- and 120-hours of treatment. The partial hypomethylation of the reporter construct observed at 48 hours was sufficient for the reporter gene to escape repression (see Fig. 5.6). The CMV enhancer became fully demethylated at the 120-hour time point whereas the eGFP gene body retained 42.86% mCpG, perhaps because the enhancer region was already partially hypomethylated prior to GSK-3484862 exposure, as evidenced in Chapter 3

and in DMSO-treated cells. Taken together with the reporter expression data (Fig. 5.6), this data provides proof-of-concept for using HCT116-eGFPme cells as a phenotypic screening tool, to screen for compounds that induce hypomethylation in cancer cells.

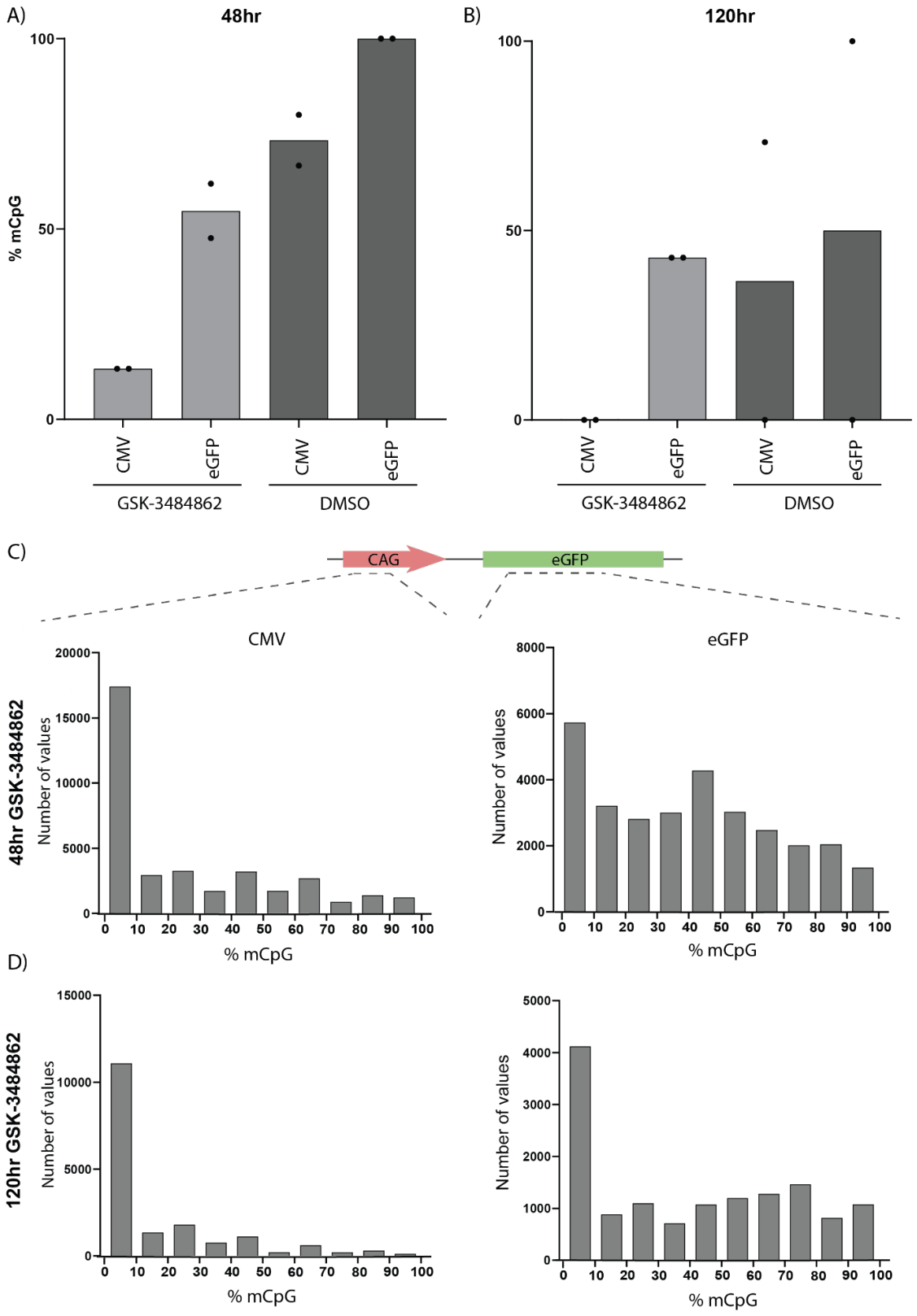


Figure 5-7: Reporter methylation is reduced following treated with GSK-3484862. A) *Median percentage of methylated CpG sites (%mCpG) at the CMV enhancer and eGFP gene body regions of the eGFP reporter cassette in HCT116 cells treated with GSK-3484862 or DMSO for 48 hours.* ***B)*** *Median percentage of methylated CpG sites (%mCpG) at the CMV enhancer and eGFP gene body regions of the eGFP reporter cassette in HCT116 cells treated with GSK-3484862 or DMSO for 120 hours. Bars represent the median of n=3 biological replicates, with data points representing the median of each individual replicate.* ***C)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 37,539) and eGFP gene body (n=34,419) from cells treated with GSK-3484862 for 48 hours. Bin width = 10.* ***D)*** *Representative histograms showing the frequency distribution of %mCpG values for the CMV enhancer (n= 17,626) and eGFP gene body (n= 14,314) from cells treated with GSK-3484862 for 120 hours. Bin width = 10.*

5.5 Proposed UHRF1 inhibitors NSC23005, NSC34716 and NSC43513 do not induce reporter activation

Having demonstrated that the methylation-sensitive reporter was activated in response to hypomethylation induced by pharmacological inhibition of DNMT1, I next sought to utilise this system to screen proposed inhibitors of UHRF1. UHRF1 has a key role in methylation maintenance and is overexpressed in several cancers, making it an attractive therapeutic target. However, difficulties targeting UHRF1 for inhibition mean that there are currently no potent or clinically improved inhibitors of UHRF1. In Chapter 4 I showed that the methylated eGFP reporter was activated in response to depletion of UHRF1 in UHRF1-AID cells. Therefore, I anticipated that a similar response would be observed with pharmacological inhibition of UHRF1. I tested three proposed inhibitors of UHRF1 that were identified in an *in silico* screen by our collaborators Myrianthopoulos *et al*. (Myrianthopoulos *et al.*, 2016). The compound NSC232003 was identified as a primary hit following screening and is proposed to bind to the SRA domain of UHRF1 by *in silico* modelling. This compound was reported to reduce DNMT1-UHRF1 interaction *in vitro*, tested by P-LISA assay (Myrianthopoulos *et al.*, 2016). Myrianthopoulos *et al*. (2016) refined this compound to generate NSC232005, which is yet to be tested as a UHRF1 inhibitor in human cells. The compounds NSC34716 and NSC43513 were also identified in by *in silico* screening and were evidenced to reduce mass spectrometry measurement of global 5mC content in MCF7 cells (Myrianthopoulos *et al.*, 2016). I tested NSC232005, NSC34716 and NSC43513 in HCT116-eGFPme cells (Fig. 5.8). I conducted a 168-

hour time course treatment and expression of the eGFP reporter was quantified after 168 hours by flow cytometry.

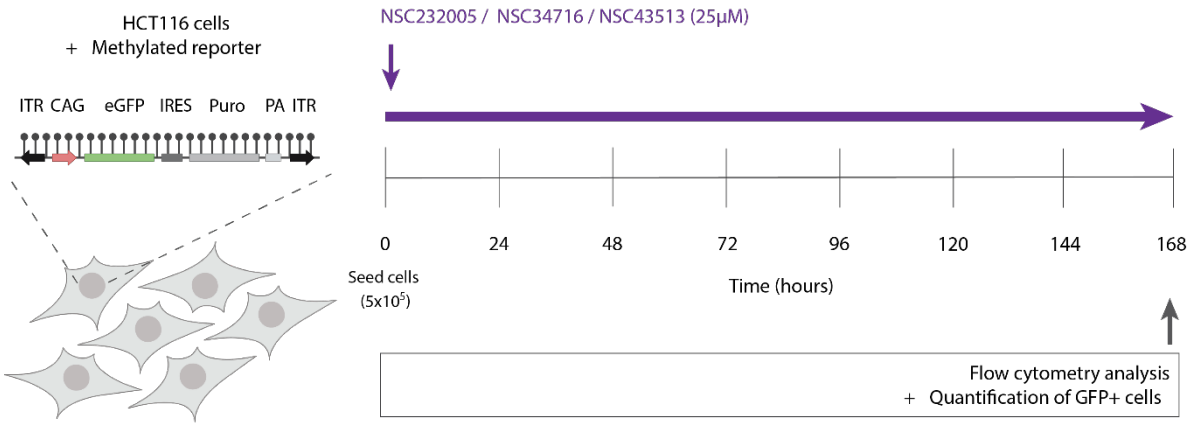


Figure 5-8: Experimental schematic of NSC232005, NSC34716 and NSC43513 168-hour treatment time course in HCT116-eGFPme cells. *HCT116 cells containing the stably integrated methylated eGFP reporter construct were seeded (5x10^5 cells) and treated with 25µM NSC232005, NSC34716 or NSC43513. Media and inhibitor were replenished every 48 hours. Flow cytometry analysis of GFP+ cells was conducted after 168-hour to assess reporter expression.*

I examined expression of eGFP in HCT116-eGFPme cells after 168-hours exposure to proposed inhibitors of UHRF1: NSC232005, NSC34716 or NSC43513 (Fig. 5.9). In cells treated with NSC232005, there was no significant change in the median fluorescence intensity (MFI) of GFP+ cells (paired t-test, $p=5.56x10^{-1}$), or the proportion of high GFP+ cells compared to DMSO-treated control cells (paired t-test, $p=2.02\times10^{-1}$). In cells treated with NSC34716, there were also no significant changes in the MFI of GFP+ cells (paired t-test, $p=7.37x10^{-1}$) or the proportion of high GFP+ cells (paired t-test, $p=8.48\times10^{-1}$). Due to the limited quantity of NSC43513 I had access to, I was only able to conduct one biological replicate with this compound. However, no substantial differences in the MFI of GFP+ cells, or the proportion of high GFP+ cells, were detected compared to DMSO-treated cells. Overall, the reporter remained repressed in cells treated with NSC232005, NSC34716 and NSC43513. These results suggest that neither NSC232005, NSC34716 nor NSC43513 are able to efficiently induce DNA hypomethylation through inhibition of UHRF1 in HCT116 cells. However, this would need to

be confirmed by EM-sequencing of the reporter construct or by assessing global methylation levels in treated cells.

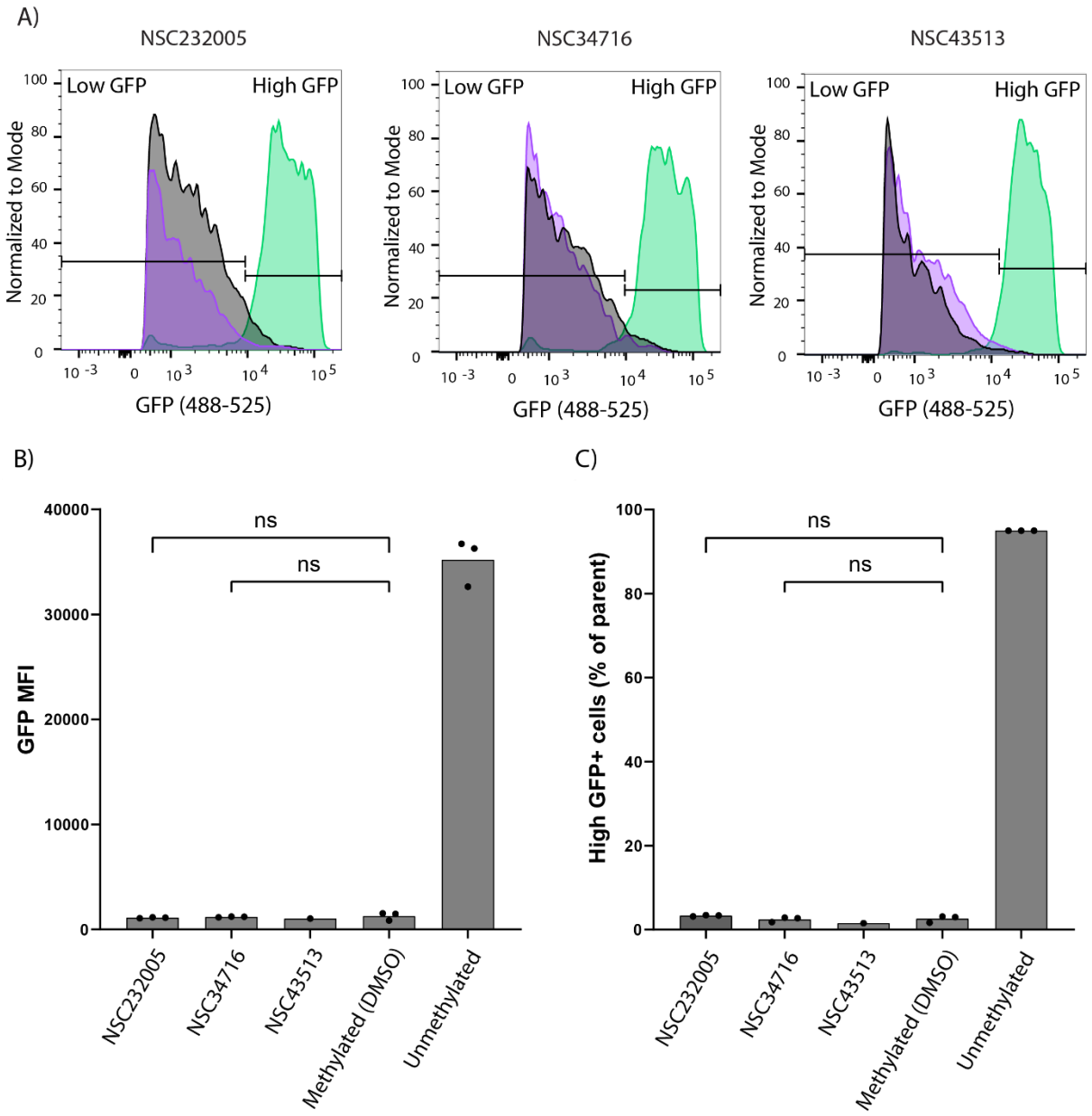


Figure 5-9: Proposed UHRF1 inhibitors NSC23005, NSC34716 and NSC43513 do not induce reporter activation. A) *Representative fluorescence intensity distribution histogram, showing intensity of eGFP expression from HCT116-eGFPme cells treated with DMSO (grey) or UHRF1 inhibitors (purple) compared to unmethylated HCT116-eGFP cells (green). Horizontal bars represent the high and low GFP gates. Gates for high and low GFP expression are indicated and were set so that 95% of the events in the unmethylated GFP-positive population fell within the high-GFP gate. Y axis = proportion of cells, normalized to the mode. X-axis = GFP fluorescence intensity.* ***B)*** *Median fluorescence intensity of all GFP+ cells at 168-hours. HCT116-eGFPme cells treated with 25µM compound or DMSO, compared to unmethylated HCT116-eGFP control cells. Paired t-tests, p-values shown.* ***C)*** *High GFP+ cells shown as a*

percentage of all GFP+ cells. HCT116-eGFPme cells treated with 25µM compound or DMSO for 168 hours, compared to unmethylated HCT116-eGFP control cells. Paired t-tests, p-values shown. NSC23005, n=3 biological replicates. NSC34716, n=3 biological replicates. NSC43513, n=1 biological replicate. Bars represent the mean of biological replicates. Data points represent the mean value from 1x10^5 cells measured for each individual replicate

5.6 Characterising the transcriptional response of HCT116 cells to GSK-3484862

In order to understand more about the transcriptional changes that occur in HCT116 cells in response to GSK-3484862 treatment, I carried out whole-transcriptome RNA sequencing. I treated HCT116 cells with 0.5µM GSK-3484862 and isolated samples at the time of seeding, prior to beginning treatment, and after 48- and 120-hours of exposure, to investigate the temporal dynamics of the transcriptional response (Fig 5.10). In tandem, HCT116 cells were treated with DMSO, as a vehicle control. I isolated both RNA and DNA from each time point. To verify hypomethylation, genomic DNA samples were analysed by liquid chromatography–mass spectrometry (LC-MS) to quantify global 5-methyl-deoxy-cytidine (5mdC) content. LC-MS was carried out by Nandini Mozumdar (Kriaucionis Lab, the University of Oxford). Library preparation and RNA sequencing was performed by the Wellcome Trust Clinical Research Facility, Edinburgh. Illumina paired-end (PE) sequencing libraries were prepared, with ribodepletion, for each sample and sequenced on the NextSeq 2000 platform.

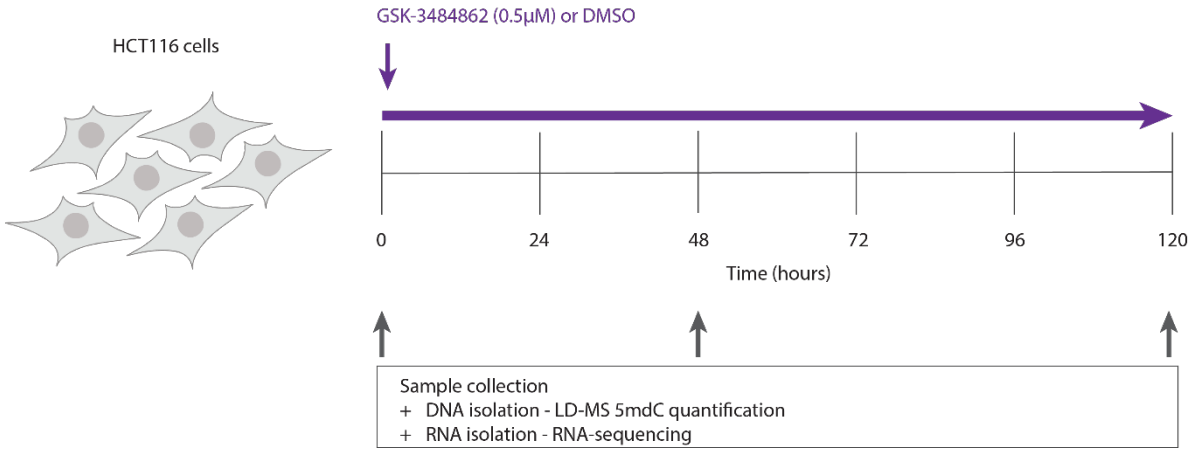


Figure 5-10: Experimental schematic for RNS sequencing of GSK-3484862-treated HCT116 cells. *HCT116 cells were treated with 0.5µM GSK-3484862 for a total of 120-hours. Media and inhibitor were refreshed every 24 hours. Samples were collected at the time of seeding, prior*

to beginning treatment, and after 48- and 120-hours of GSK-3484862 exposure. The 5mdC content of cells was quantified by LD-MS. Transcriptional changes were investigated by RNA sequencing. HCT116 cells treated with DMSO were included as a vehicle control and analysed at the 120-hour time point by LD-MS and RNA sequencing. The experiment was conducted in triplicate (n=3 biological replicates).

5.6.1 Global methylation content is ~60% reduced following GSK-3484862 treatment

To verify hypomethylation in HCT116 cells treated with 0.5µM GSK-3484862, samples were analysed by LD-MS to quantify global methylation levels (Fig. 5.11). HCT116 cells exposed to GSK-3484862 for 48 hours were found to have ~60% reduced global methylation levels compared to untreated cells (paired t-test, $p=2.97x10^{-2}$). After 120-hours exposure, global methylation levels remained reduced by ~60% compared to untreated cells, with no significant difference in relative methyl-cytosine content between 48- and 120-hours (paired t-test, p=0.95). There was no significant difference in global methylation content in DMSO cells and untreated cells. These findings indicate that maximum global hypomethylation is obtained within 48 hours, or after approximately two cell divisions. This result is in support of the Southern blot data (see Figures 5.3 and 5.4), which suggested that the degree of hypomethylation was similar after 48- and 120-hours GSK-3484862 treatment. However, this is not in line the reporter expression (Fig. 5.6) or reporter EM-sequencing (Fig 5.7) data, which indicated that hypomethylation was time-dependant.

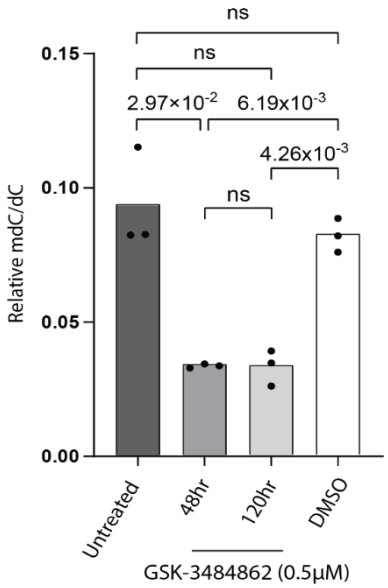


Figure 5-11: Global methylation content is ~60% reduced following GSK-3484862 treatment. *Bar plot showing the relative abundance of methyl-deoxy-cytosine (mdC) over deoxy-cytosine (dC) in HCT116 cells. Quantified by liquid chromatography–mass spectrometry (LC-MS) by Nandini Mozumdar, Kriaucionis Lab. Bars represent the mean of n=3 biological replicates. Data points represent the mean of two technical replicates for each biological replicate. Paired t-tests, p-values shown.*

5.6.2 Quality control and clustering of RNA sequencing samples

The raw RNA sequencing data were processed by Dr Philippe Gautier, Institute of Genetics and Cancer Bioinformatics Analysis Core. Details of the processing and analysis pipelines are provided in the Materials and Methods section. Due to a problem with one of the sequencing lanes during the first sequencing run, all samples were re-sequenced. Therefore, in addition to the full sequencing results from run 2, we obtained sequencing from one lane of run 1. In order to utilise these additional reads, we first examined whether we could merge them with the second sequencing experiment. The analysis pipeline was first carried out with all sequences from runs 1 and 2 labelled separately. The results are summarised in a Principal Component Analysis (PCA) plot in Figure 5.12. PCA is method to reduce the dimensionality of data, enabling the identification of patterns and similarities between samples by projecting the data into a lower-dimensional space that captures the most significant variance (Greenacre *et al.*, 2022). This enables visualisation and clustering of samples based on their gene expression profiles. Figure 5.12 shows that the two runs cluster closely. Based on this, we chose to merge the two runs and repeat the analysis, treating them as technical replicates.

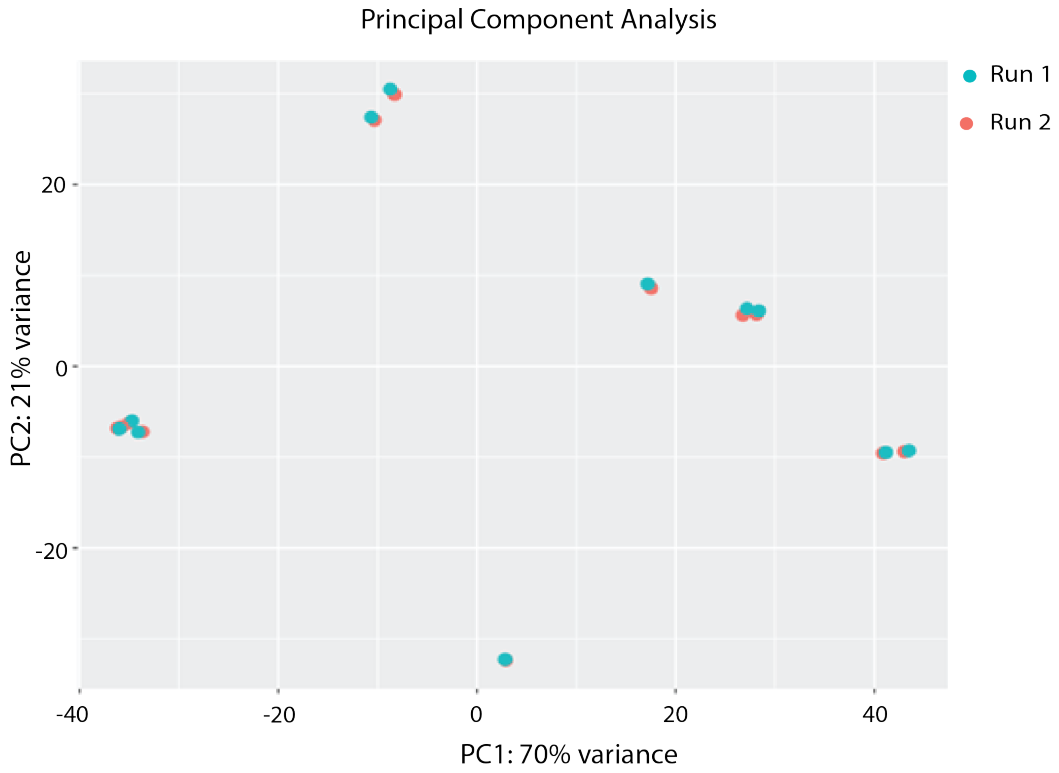


Figure 5-12: Principal component analysis (PCA) showing clustering of gene expression profiles from RNA-sequencing run 1 and run 2. *The two runs display a near-identical profile, clustering closely in both dimensions of variance.*

As a result of merging the two sequencing runs, we obtained a very high number of reads per sample. The multiQC report showed high quality for all sequencing and mapping measurements. The PCA was repeated for the merged analysis, and the results are shown in Figure 5.13. Two outliers were identified: replicate 1 of the DMSO-treated cells and replicate 3 of the GSK-3484862-treated 120-hour timepoint. The outlier DMSO-treated sample clustered close to the untreated samples. The remaining DMSO samples are distinct from untreated samples along PCA 2, which accounts for 18% of the total transcriptome variance. The outlier 120-hour GSK-3484862 sample clustered away from other replicates from this treatment time point along PCA 1, which accounts for 73% of the transcriptome variance. This outlier is clustered towards the 48-hour GSK-3484862 treatment along PCA 1 but is segregated from the 48-hour timepoint samples along PCA 2. No differences in the outlier samples were observed during the experiment and all were treated identically, therefore we chose not to exclude these outliers from further analysis. The remaining replicate samples cluster closely and show consistency between the replicate data. The GSK-3484862-treated samples cluster away from untreated and DMSO-treated samples along PCA 1 and PCA 2.

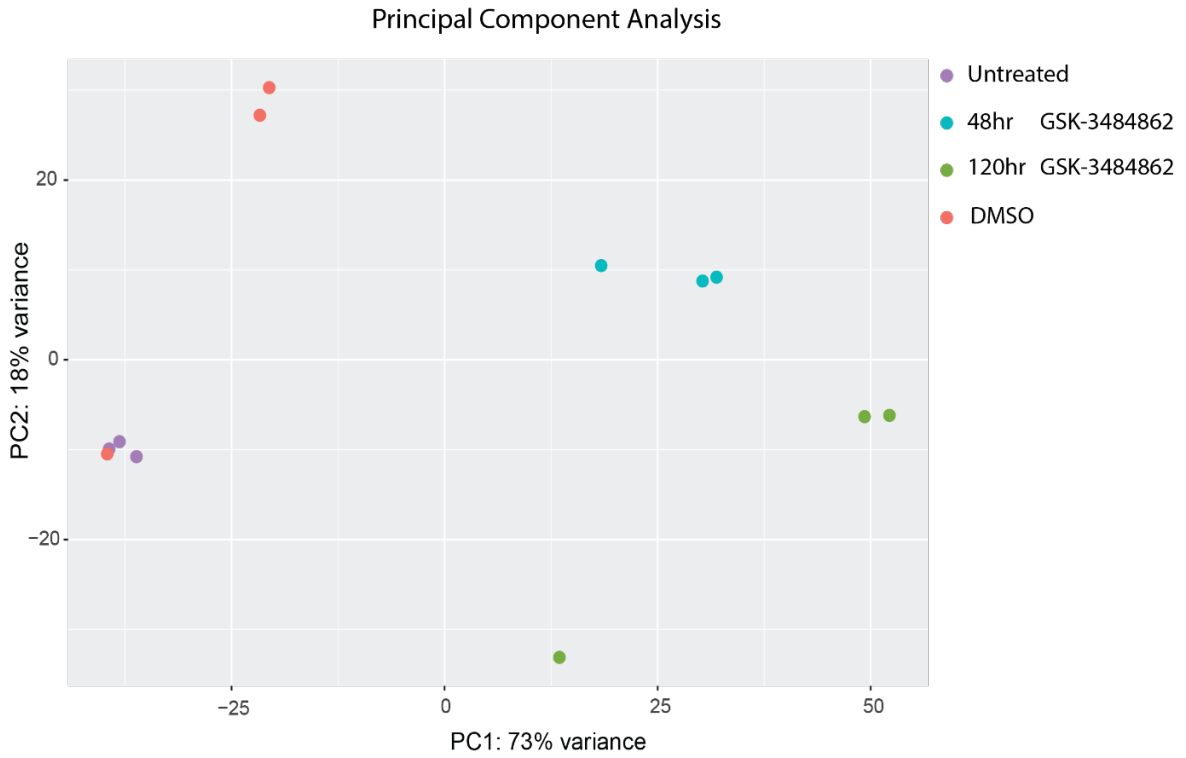


Figure 5-13: Principal component analysis (PCA) showing gene expression profiles of HCT116 cells treated with GSK-3484862, compared to untreated cells and DMSO controls. *Untreated sample replicates cluster closely, as do 48-hour GSK-3484862 treated replicates, demonstrating consistency between the triplicate samples. Outliers were identified in DMSO control and 120-hour GSK-3484862 cells but not excluded from further analysis. The remaining DMSO and 120-hour GSK3484862 samples cluster closely in PCA 1 and PCA 2. GSK-3484862 samples cluster separately from untreated and DMSO control cells along PCA 1 and PCA 2.*

5.7 Hypomethylation induced by GSK-3484862 results in upregulation of genes repressed by methylation

To determine the impact of hypomethylation induced by GSK-3484862 on HCT116 gene expression, Dr Gautier carried out differential gene expression analysis using the DESeq2 package (Love, Huber and Anders, 2014). The differential expression changes between untreated HCT116 cells and those treated with GSK-3484862 for 48 hours are shown in Figure 5.14. Thresholding the results for genes with a Log_2 fold change >2, we observe 1669 significantly up-regulated genes (padj <0.05). 311 genes were significantly down-regulated (Log_2 fold change <2, padj <0.05). The volcano plot shows a strong bias towards genes becoming upregulated. This is in line with what we expected to observe after inhibiting DNA

methylation, which is typically considered to have a role in repressing gene expression. Notably, re-expression of *Dazl* was observed. DAZL is an RNA-binding transcriptional regulator that is active during meiosis and essential for gamete development (Yang *et al.*, 2020). This gene is known to be directly repressed by DNA methylation in somatic cells. Activation of *Dazl* has been previously reported in breast cancer cells lines after treatment with 5-aza-dC (Sproul *et al.*, 2011; Hackett *et al.*, 2012). Activation of the cancer testis antigen (CTA) family genes was also observed, including MAGE, GAGE and SSX family CTAs (Fig. 5.17). Upregulation of CTA genes in response to demethylation is in line with published work (Weber *et al.*, 1994; Güre *et al.*, 2002; James, Link and Karpf, 2006; Weon and Potts, 2015).

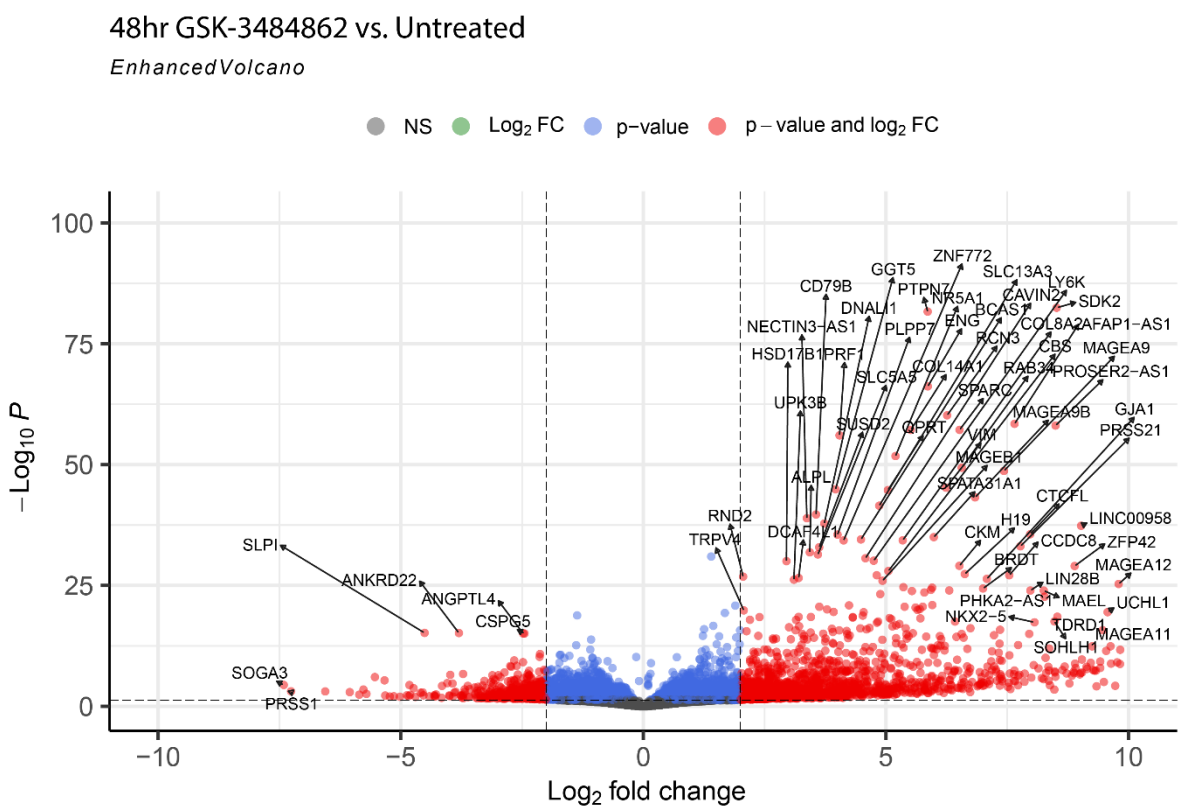


Figure 5-14: Hypomethylation induced by GSK-3484862 results in up-regulation of genes at 48 hours. *Enhanced volcano plot showing differentially expressed genes in HCT116 cells treated with GSK-3484862 for 48 hours compared to untreated cells. In total, 6997 genes were differentially expressed (padj<0.05). Of these, 1669 were significantly up-regulated (Log_2 fold change >2, red dots) and 311 were significantly down-regulated (Log_2 fold change <-2, red dots). Blue dots represent genes that were differentially expressed but did not meet the Log_2 fold change threshold. Black dots represent genes that were not significantly differentially expressed between conditions. Total genes shown = 40,397.*

Differential expression changes between untreated HCT116 cells and those treated with GSK-3484862 for 120 hours are shown in Figure 5.15. Again, a strong bias towards up-regulation of genes was observed, supporting the theory that DNA methylation acts as a repressive epigenetic mark. Thresholding the results for genes with a Log_2 fold change >2, 1980 genes were up-regulated (padj<0.05). 107 were down-regulated (Log_2 fold change <2, padj<0.05). More genes were upregulated and fewer downregulated at 120-hours than 48-hours, suggesting a trend towards increased gene expression. The methylation-responsive *Dazl* and CTA gene family were again shown to be activated in response to hypomethylation (Fig 5.17). ANKRD22 is significantly down-regulated at 48- and 120-hours GSK-3484862 exposure. ANKRD22 promotes cell proliferation, and its knock-down is reported to inhibit tumour growth (Yin *et al.*, 2017; Utsumi *et al.*, 2021)

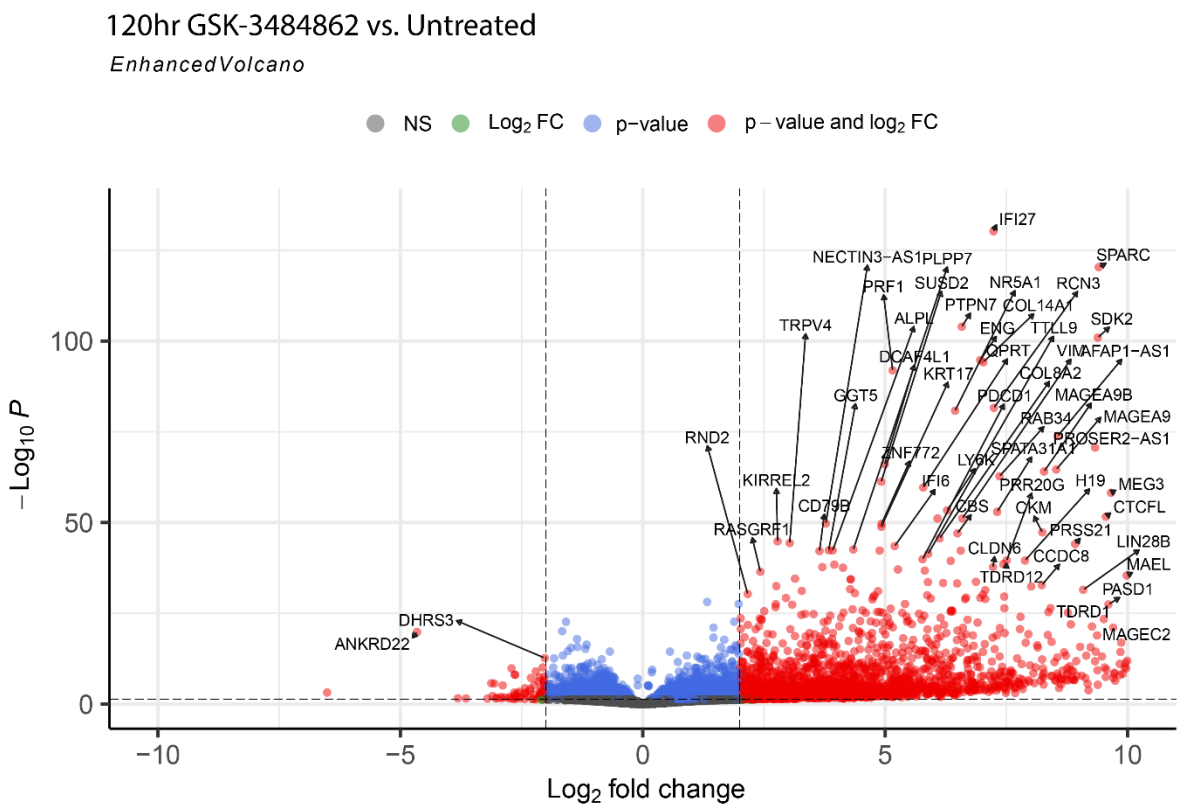


Figure 5-15: Hypomethylation induced by GSK-3484862 results in up-regulation of genes at 120 hours. *Enhanced volcano plot showing differentially expressed genes in HCT116 cells treated with GSK-3484862 for 120 hours compared to untreated cells. 6709 genes were significantly differentially expressed (padj<0.05). Of these, 1980 were up-regulated (Log_2 fold change >2, red dots) and 107 were down-regulated (Log_2 fold change <-2, red dots). Blue dots represent genes that were differentially expressed but did not meet the Log_2 fold change threshold. Black dots represent genes that were not significantly differentially expressed between conditions. Total genes shown = 40,397.*

Next, we assessed differential expression changes between HCT116 cells treated with GSK-3484862 for 48- and 120-hours (Fig. 5.16). Overall, 1659 genes were differentially expressed between the 48- and 120-hour time points. Thresholding the results for genes with a Log_2 fold change >2, 362 genes were up-regulated (padj<0.05). Only 8 were down-regulated (Log_2 fold change <2, padj<0.05). Interestingly, the genes most up-regulated between 48- and 120-hours GSK-3484862 treatment were genes involved in antiviral immune response and induction of type-I interferon signalling. These include genes involved in sensing viral RNA, such as the OAS family genes, and various interferons, including IFI27, IFI6, IFIH1 and IFI44, among others. Activation of interferon signalling has previously been reported in cancer cells in response to 5-aza exposure (Chiappinelli *et al.*, 2015). I will discuss activation of interferon signalling by GSK-3484862 in more detail later in this chapter. Of the down-regulated genes, H2AB3 is a variant of histone 2A (Martire and Banaszynski, 2020). The RAS oncogene family members RAB26 and RAB7B were also down-regulated.

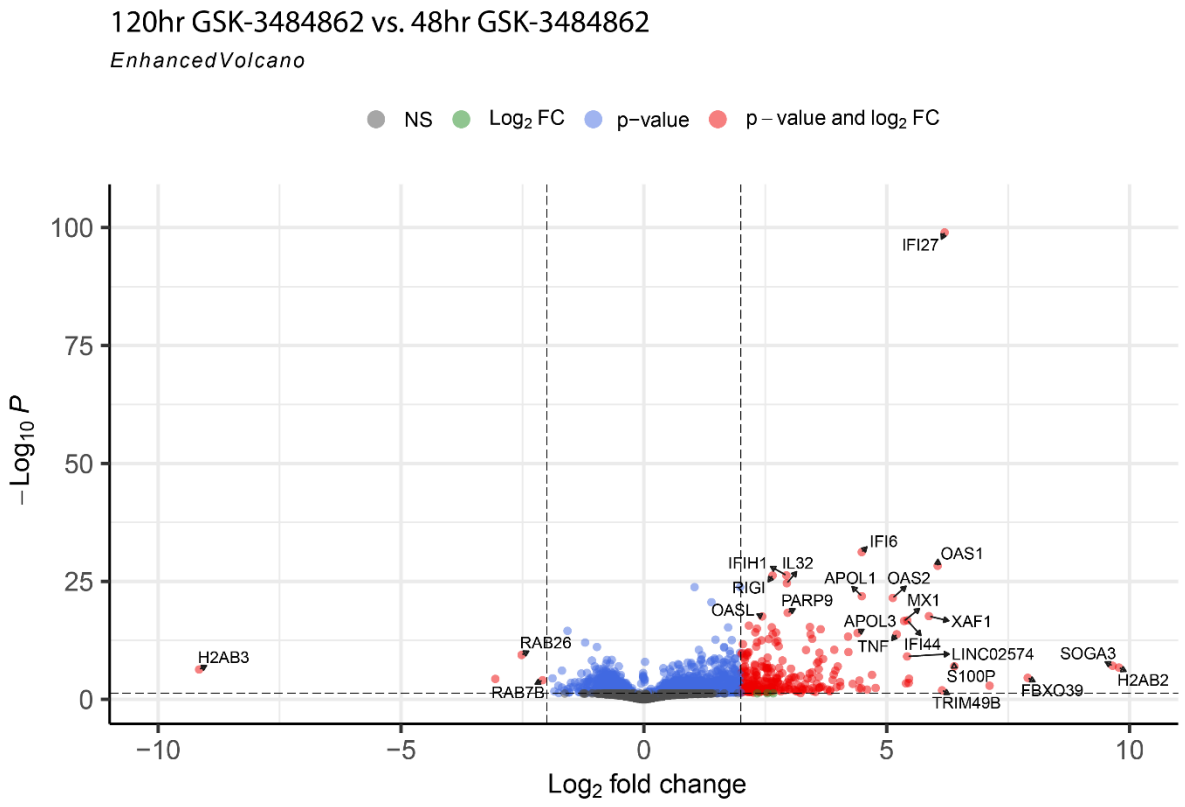


Figure 5-16: Differential gene expression changes between HCT116 cells treated with GSK-3484862 for 48- and 120-hours. *Enhanced volcano plot showing differentially expressed genes in HCT116 cells treated with GSK-3484862 for 120 hours compared to 48 hours. 1659 genes were significantly differentially expressed (padj<0.05). Of these, 362 were up-regulated (Log_2 fold change >2, red dots) and 8 were down-regulated (Log_2 fold change <-2, red dots). Blue*

dots represent genes that were differentially expressed but did not meet the Log_2 fold change threshold. Black dots represent genes that were not significantly differentially expressed between conditions. Total genes shown = 40,397.

I plotted the Log_2 fold change values for a selection of genes typically repressed by methylation, to better visualise their upregulation (Fig 5.17). Those plotted include Dazl, and a selection of CTA genes from the MAGE, GAGE and SSX families. Expression of each of these marker genes increased between the 48- and 120-hour time points, indicating a time-dependant increase in expression in response to hypomethylation. Activation of these methylation-responsive marker genes supports that GSK-3484862 induces hypomethylation through inhibition of DNMT1, leading to de-repression of genes regulated by methylation.

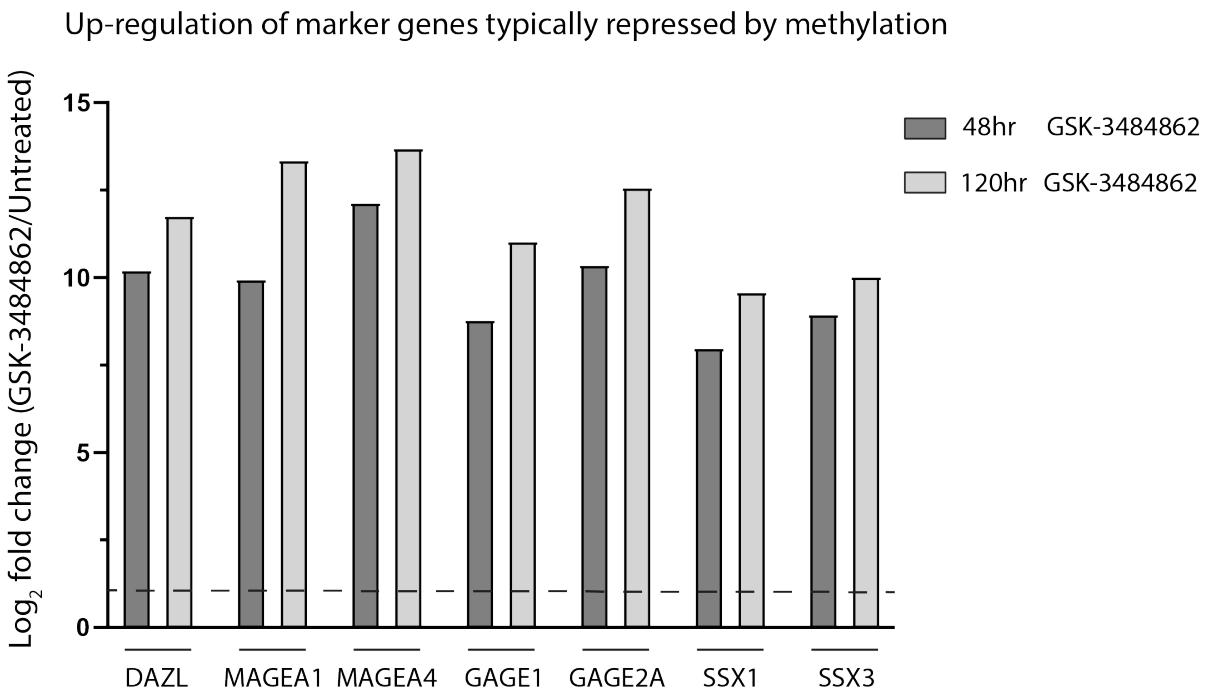


***Figure 5-17**: **Hypomethylation induced by GSK-3484862 results in upregulation of marker genes typically repressed by methylation.** Bar plot showing Log_2 fold change values for a selection of genes known to be regulated by DNA methylation and typically repressed in somatic cells. Log_2 fold change values for 48-hours GSK-3484862 compared to untreated cells (dark grey) and 120-hours GSK-3484862 compared to untreated cells (light grey). Y-axis = Log_2 fold change (padj<0.05). Dashed line at Log_2 fold change = 1.*

5.8 GSK-3484862-treated cells are enriched for genes associated with 5-aza-dC and 5-aza responses

Next, I was interested in determining how similar the transcriptional response of HCT116 cells to GSK-3484862 was to responses to the traditional HMAs, 5-aza and 5-aza-dC. To do this, I utilised a publicly available RNA sequencing dataset for HCT116 cells treated with 1µM 5-aza-dC for a total of 5 days (Cristofari *et al*., 2023; Lanciano *et al.*, 2023). Dr Philippe Gautier processed the raw RNA sequencing data, performed differential gene expression analysis using the DESeq2 pipeline and assessed overlaps between GSK-3484862 and 5-aza-dC differentially expressed (DE) genes. DE genes in 120-hour GSK-3484862 treated HCT116 cells were compared to DE genes from the 5-day 5-aza-dC treated HCT116 dataset (Fig. 5.18A-B). In GSK-3484862 treated cells, a total of 6709 significantly differentially expressed genes were identified (padj = <0.05). Of these, 1980 had a Log_2 fold change >2. In 5-aza-dC treated cells, a total of 11,144 significantly differentially expressed genes were found (padj = <0.05), and 5327 of these had a Log_2 fold change >2. There were 1419 DE genes that were shared between the GSK-3484862 and 5-aza-dC datasets (Log_2 fold change >2). Overall, 72% of the GSK-3484862 DE genes overlap with 5-aza-dC DE genes, at a threshold of Log_2 fold change >2, indicating that the GSK-3484862 DE are highly similar to 5-aza-dC DE genes. However, overlapping genes only account for 27% of the 5-aza-dC DE genes. This may be due to GSK-3484862 being a more specific inhibitor of DNMT1. Whereas, 5-aza-dC is a cytidine analogue that is incorporated into the DNA and induces a range of cellular changes apart from inhibiting DNMT1 activity.

To gain further insight into the GSK-3484862 gene expression changes associated with cytosine analogues, I performed a gene set enrichment analysis (GSEA). GSEA is a computational method used to determine whether a predefined set of genes, associated with a particular phenotype, show over- or under- representation within a ranked list of DE genes from a large sequencing dataset (Subramanian *et al.*, 2005). Instead of focusing on individual genes, GSEA evaluates groups of genes that are involved in common biological pathways. The method involves ranking all genes in a sequencing dataset by their expression differences and then assessing whether the members of a gene set are randomly distributed throughout the ranking, or are primarily found at the top or bottom, indicating enrichment. GSEA produces a

graphical overview of the distribution of genes from the predefined gene set along the ranked list.

I examined gene set enrichment in 120-hour GSK-3484862 treated cells compared with untreated cells. Having shown that GSK-484862 treatment elicits a similar transcriptional response to 5-aza-dC in HCT116 cells, I was next interested in comparing gene expression in GSK-3484862 treated cells to genes reported to be upregulated in response to 5-aza. Using RNA-sequencing normalised count data, generated by Dr Gautier during DESeq2 processing, I examined the enrichment of genes reported to be frequently upregulated in response to 5-aza and TSA treatment (Zhong *et al.*, 2007). This gene set contained 185 genes that were upregulated in response to 5-aza and TSA treatment in at least 3 out of 4 non-small cell lung cancer (NSCLC) cell lines tested. This was the only available gene list for 5-aza, however, a notable caveat of this gene set is the inclusion of TSA treatment, a histone deacetylase inhibitor. The results of the gene set enrichment analysis are shown in Figure 5.18C. The GSEA algorithm generates an enrichment profile, by 'walking' down the ranked list of differentially expressed genes and adding a positive increment to a running-sum statistic when a gene is in the gene set, and a negative increment when a gene is not in the gene set. An enrichment score (ES) is calculated based on the maximum deviation from zero encountered during the walk, representing the point at which the gene set is most enriched in the ranked list. The enrichment profile is illustrated by the green line in Figure 5.18C. The black bars beneath the enrichment profile show the positions of genes within the predefined list ('hits') within the ranked differential expression list. The relative change in gene expression is visualised with a colour scale, where 'red' indicates increased expression and 'blue' indicates decreased expression. At the bottom of the plot, the ranked genes and their relative expression are shown as a 'ranking metric score' in 'grey'. The Normalized Enrichment Score (NES) is derived by normalizing the ES to account for differences in gene set size, making it comparable across different gene sets. The NES reflects the degree to which a gene set is overrepresented at the top or bottom of a ranked list, with higher NES values indicating stronger enrichment relative to what would be expected by chance. A NES score of >1.5 indicates strong positive enrichment, while a NES <-1.5 indicates strong negative enrichment. Finally, the algorithm corrects for multiple testing and calculates a false discovery rate (FDR) q-value for the NES

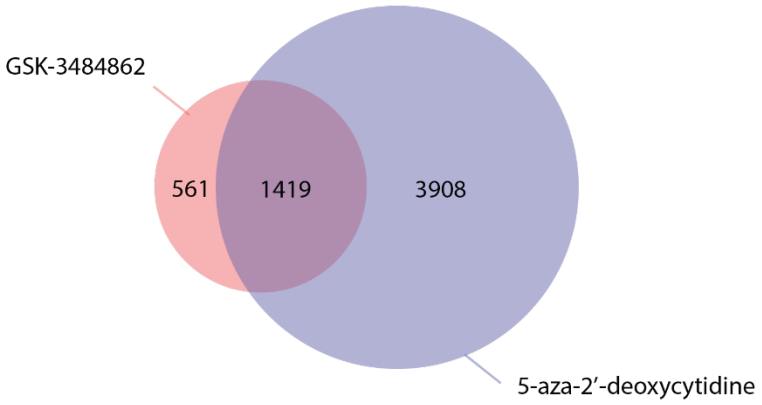
(Subramanian *et al.*, 2005). The NES for the 5-aza and TSA gene set was 2.09 (q=0.00), indicating positive significant enrichment of this gene set in 120-hour GSK-3484862 cells.

To validate the GSEA findings, I plotted the DESeq2 Log_2 fold changes values for all genes differentially expressed at 120-hour GSK-3484862 treatment against only those from the Aza and TSA gene set (Fig. 5.18D). Genes activated in response to 5-aza and TSA were significantly enriched (Student's t-test, $p=8.61x10^{-9}$) compared to all genes. In line with the GSEA, this finding indicates that GSK-3484862 treated cells display a transcriptional bias toward expression of genes associated with 5-aza and TSA treatment. Together, these data indicate that GSK-3484862 elicits a similar transcriptional response in cancer cells as 5-aza and 5-aza-dC.

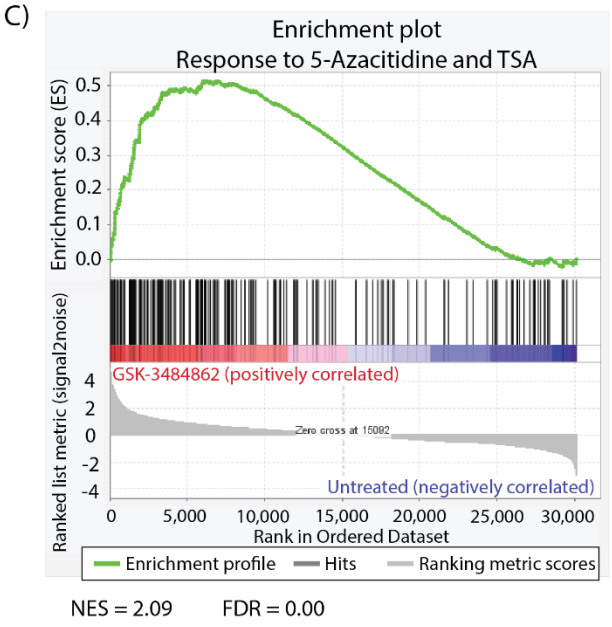
A)

	Count			Percentage overlap	
Log2 Fold Change	GSK-3484862	5-aza-2'-deoxycytidine	Overlaps	GSK-3484862	5-aza-2'-deoxycytidine
no Log filters	6709	11144	3720	55%	33%
log>2	1980	5327	1419	72%	27%
log>5	657	2771	444	68%	16%

B)



C)



D)

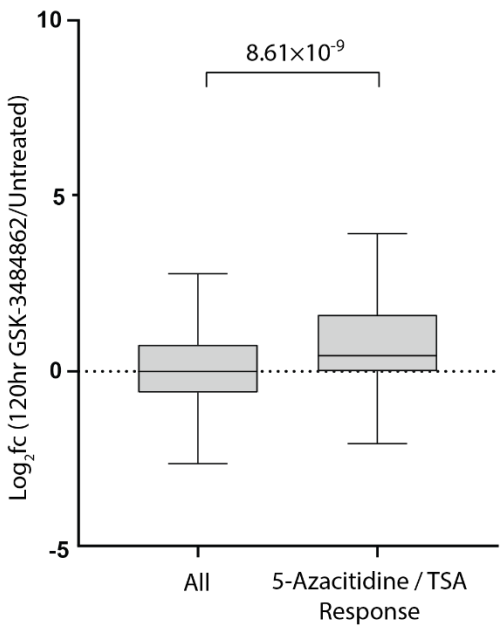


Figure 5-18: GSK-3484862-treated cells are enriched for genes associated with 5-aza-dC and 5-aza responses. A) *Table showing significantly differentially expressed (padj <0.05) genes in GSK-3484862 and 5-aza-dC treated cells at 120 hours. Absolute counts and percentage overlap for shared genes are shown.* ***B)*** *Venn diagram showing overlap between differentially expressed genes with a Log_2 fold change >2 from GSK-3484863 (n=1980) and 5-aza-dC (n=5327) treated cells.* ***C)*** *GSEA enrichment plot showing enrichment of genes associated with responses to 5-aza and TSA. Enrichment profile shown in green. Black bars show the position of genes within the gene set ('hits') relative to the ranked list of differentially expressed genes from the GSK-3484862 120-hour cells. Ranked genes and their relative expression are shown as a 'ranking metric score' in grey. Normalised enrichment score (NES) = 2.09. False-discovery rate (FDR) = 0.00.* ***D)*** *Box plot showing Log_2 fold change values for GSK-3484862 120-hour cells compared to untreated cells. All genes (n=29,518) and 5-aza/TSA-responsive gene set (n=185).*

Box, 25th–75th percentile; whiskers, 1.5× interquartile range; centre line, median. Student's t-test, p=8.61x10^{9}.

5.9 Interferons and interferon-responsive genes are upregulated in response to GSK-3484862

Having identified upregulation of genes interferons and gene associated with immune signalling between 48- and 120-hours GSK-3484862 treatment (see Fig. 5.16), I was interested in exploring this response in more detail. The activation of viral defence responses has previously been reported in human cancer cells following hypomethylation. This response has been observed in cancer cells treated with 5-aza, and in response to shRNA knockdown of UHRF1 (Chiappinelli *et al.*, 2015; Irwin *et al.*, 2023). It has been proposed that this innate immune response can help improve patient responses to immune checkpoint therapy in cancer treatment. Therefore, I was interested in investigating whether GSK-3488462 treated cells were enriched for genes associated with immune signalling. Gene ontology (GO) term enrichment for differentially expressed genes between 48- and 120hours GSK-3484862 treatment using enrichGO, an R package for gene ontology analysis. GO terms analysis was conducted by Phillipe Gautier. This technique classifies genes that are differentially expressed in an RNA-sequencing data according to their biological process and then ranks the processes that are most enriched in the sequencing dataset. The GO terms that were most enriched in the 120hr vs. 48hr GK-3484862 dataset all related to innate immune responses, response to biotic stimuli and viral defence pathways (Fig. 5.19). This data indicates that genes related to immune signalling and viral defence pathways are highly enriched between 48- and 120-hours GSK-3484862 treatment.

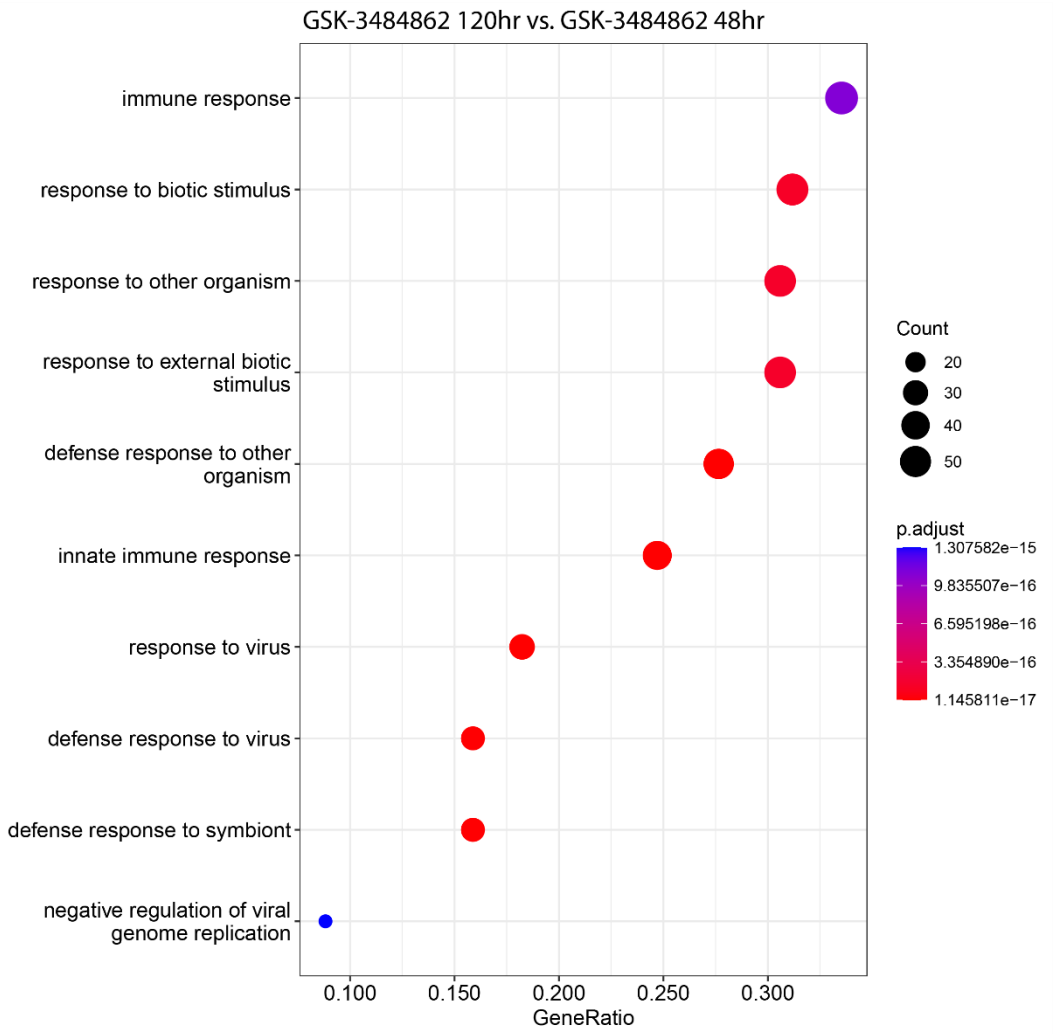


Figure 5-19: GSK-3484862 treated cells are enriched for GO terms associated with immune response and viral defence. *Biological process GO term enrichment analysis for HCT116 cells treated with GSK-3484862 for 120 hours compared to cells treated for 48 hours. Statistical significance is indicated by padj value, colour coded according to the key. The size of the note indicates the count of genes within the GO term. The GeneRatio is the percentage of the number of genes present in this GO term in the sequencing dataset over the total number of genes in the GO term category. Plot generated by Phillipe Gautier using enrichGO.*

To look in more detail at specific gene lists that were enriched, I performed a hallmark GSEA using RNA-sequencing normalised count data, generated by Dr Gautier during DESeq2 processing. The hallmark gene set collection is a curated collection of gene sets that represent well-defined biological states or processes, such as pathways involved in cancer, immune response, or metabolism. The hallmark gene sets were designed to reduce redundancy and improve the interpretability of GSEA results by grouping genes that have well-defined biological functions (Liberzon *et al.*, 2015). The hallmark analysis provides an unbiased

overview of the biological processes that are altered in a large sequencing dataset. I assessed gene set enrichment in 120-hour GSK-3484862 treated cells compared with untreated cells. Of the hallmark gene sets, interferon alpha response was most enriched, with a NES score of 2.52 (q=0.00, Fig. 5.19A). Interferon gamma response genes were also a top hit in the GSEA, with a NES score of 2.13 (q=0.00, Fig 5.19B).

I tested an additional set of 66 interferon-responsive genes (Browne *et al.*, 2001). This gene set was curated by Browne *et al*. (2001) and contains genes that were found to be up-regulated in human primary fibroblast culture after treatment with interferon alpha. GSEA for the interferon-responsive gene set showed significant positive correlation with the GSK-3484862 120-hour dataset (NES=2.59, q=0.00, Fig. 5.19C).

To validate the results of the GSEA, I plotted the DESeq2 Log_2 fold change values for all genes differentially expressed at 120-hour GSK-3484862 treatment against only those from the tested gene sets (Fig. 5.19D). Interferon alpha response genes were significantly enriched (Student's t-test, p=4.42x10-^{13}), as were interferon gamma response genes (Student's t-test, p=3.44x10^{-4}) and interferon responsive genes (Student's t-test, p=1.38x10^{-7}). Finally, I assessed the overlap between these three gene sets using BioVenn, a tool that produces Venn diagrams from lists containing biological identifiers, in this case gene names (Hulsen, de Vlieg and Alkema, 2008). I identified some overlap between the three gene sets, but each contained unique genes. This suggests that separate interferon pathways are upregulated in the 120-hour GSK-3484862 dataset. Together, these data support that interferon signalling is activated in GSK-3484862 treated cancer cells.

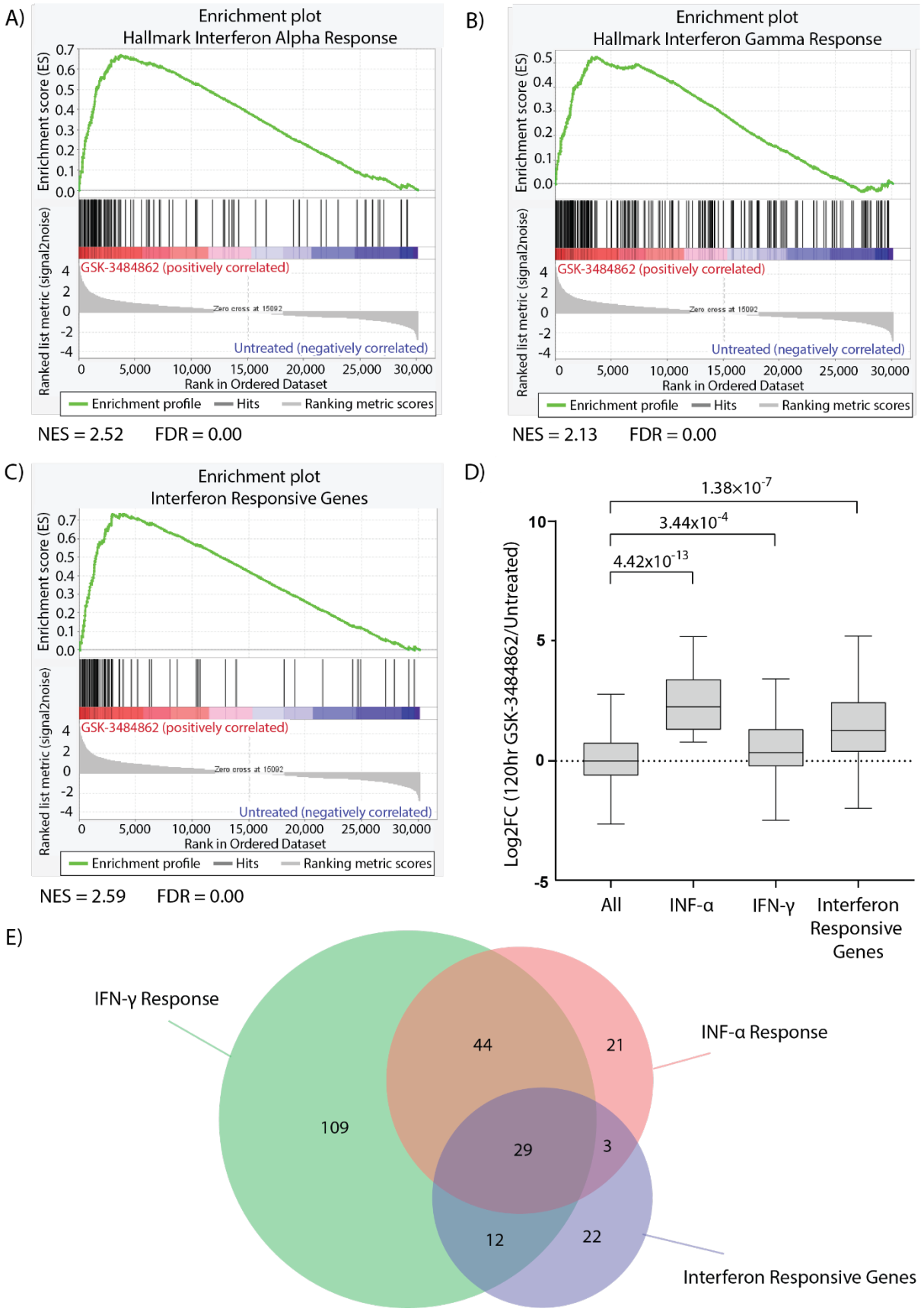


Figure 5-20: Interferon-signalling and interferon-responsive genes are upregulated by GSK-3484862. A) *GSEA enrichment plot showing enrichment of genes associated interferon alpha signalling. Normalised enrichment score (NES) = 2.52. False-discovery rate (FDR) = 0.00.* ***B)***

GSEA enrichment plot showing enrichment of genes associated interferon gamma signalling. NES=2.13. FDR=0.00. Hallmark gene sets curated by Liberzon et al*. (2015).* ***C)*** *GSEA enrichment plot showing enrichment of interferon responsive genes (Browne* et al*., 2001). NES=2.59. FDR=0.00.* ***GSEA plots:*** *Enrichment profile shown in green. Black bars show the position of genes within the gene set ('hits') relative to the ranked list of differentially expressed genes from the GSK-3484862 120-hour cells. Ranked genes and their relative expression are shown as a 'ranking metric score' in grey.* ***D)*** *Box plot showing Log_2 fold change values for GSK-3484862 120-hour cells compared to untreated cells. All genes (n=29,518), IFN-α (n=97), IFN-γ (n=194) and interferon-responsive (n=66). Box, 25th–75th percentile; whiskers, 1.5× interquartile range; centre line, median. Student's t-test, p-values shown.* ***E)*** *Venn diagram showing the number of overlapping and unique genes from the three gene sets tested. Generated using BioVenn (Hulsen, de Vlieg and Alkema, 2008).*

5.10 Interferon-induced antiviral response genes are upregulated in response to GSK-3484862

Interferons are involved in innate immune signalling and antiviral defence. Therefore, I next investigated enrichment of gene sets specifically associated with the interferon antiviral response. I performed the GSEA using RNA-sequencing normalised count data and compared 120-hour GSK-3484862 treated cells to untreated cells. I utilised two gene sets related to interferon-induced antiviral signalling. The first gene set contained 78 genes that were characterised as being involved in interferon-induced antiviral signalling pathways in asthmatic individuals (Bosco *et al.*, 2010). This gene set showed significant positive correlation with the GSK-3484862 dataset, indicating enrichment of these genes in inhibitor-treated cells (NES=2.29, q=0.00, Fig. 5.20A). The second gene set contained 152 genes related to antiviral response by interferon-stimulated genes, characterised by the Reactome project (Milacic *et al.*, 2023). This gene set was significantly positively correlated with the GSK-3484862 dataset, suggesting enrichment of genes associated with antiviral signalling through interferon-stimulated genes in GSK-3484862 treated cells (NES=1.74, q=0.00, Fig. 5.20B). The Log_2 fold change values for genes from these two gene sets are plotted against those for all genes in the GSK-3484862 120-hour dataset in Figure 5.20C. Genes from the Bosco *et al.* (2010) gene set were significantly enriched (Student's t-test, $p=6.84x10^{-7}$), as were genes from the Reactome gene set (Student's t-test, $p=6.26x10^{-6}$). These gene sets displayed very little overlap, suggesting that separate antiviral response pathways may be activated in response

to GSK-3484862 treatment (Fig. 5.20D). Together, these data indicate that GSK-3484862 induces interferon signalling and activation of interferon-stimulated genes that go on to activate innate and adaptive immune responses in HCT116 cells.

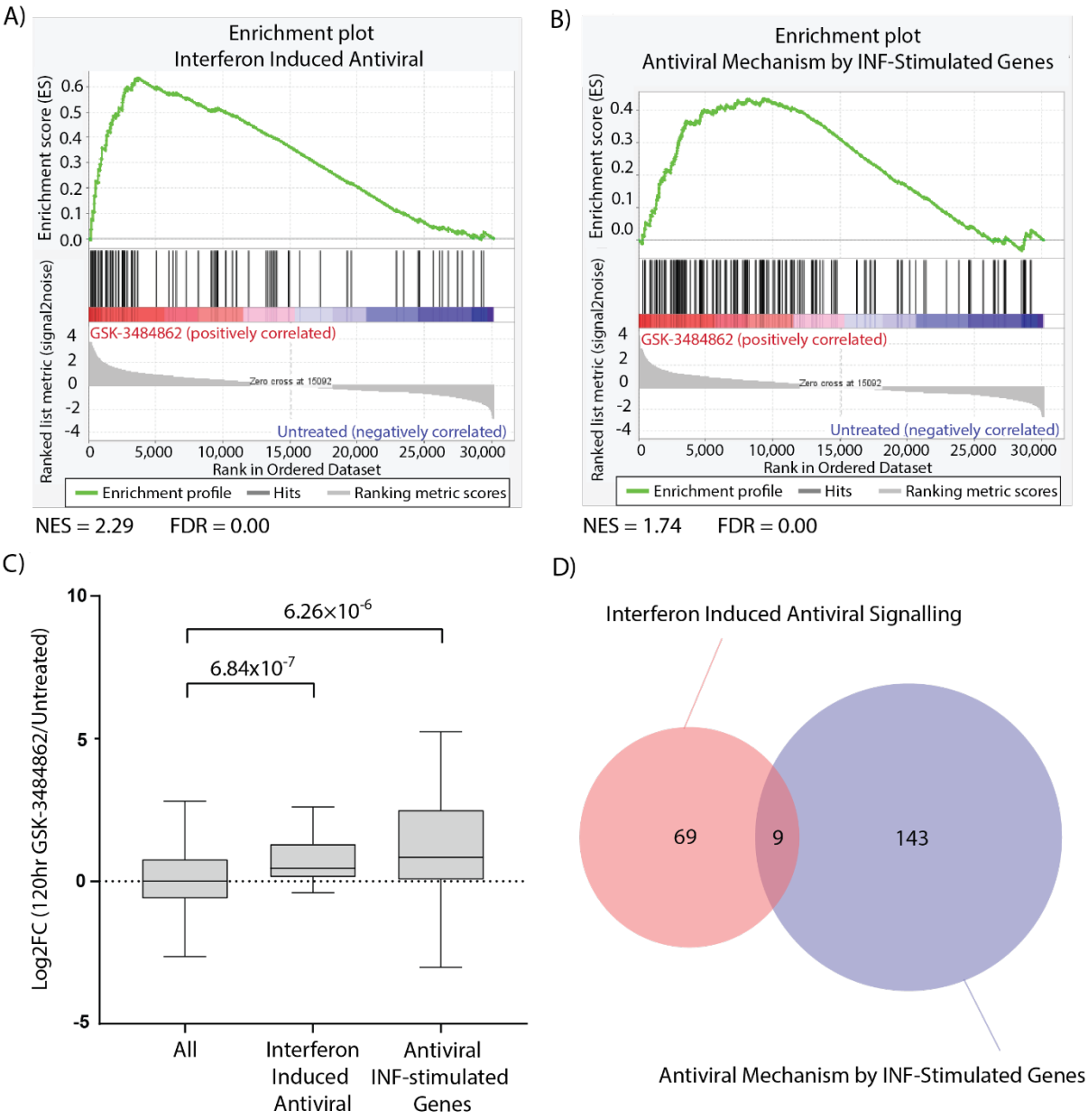


Figure 5-21: Interferon-induced antiviral response genes are upregulated by GSK-3484862. ***A)*** *GSEA enrichment plot showing enrichment of interferon induced antiviral genes (Bosco* et al.*, 2010). Normalised enrichment score (NES)=2.29. False discovery rate (FDR)=0.00.* ***B)*** *Enrichment of genes associated with antiviral mechanisms through interferon-stimulated genes (Milacic* et al.*, 2023). NES=1.74. FDR=0.00.* ***GSEA plots:*** *Enrichment profile shown in green. Black bars show the position of genes within the gene set ('hits') relative to the ranked list of differentially expressed genes from the GSK-3484862 120-hour cells. Ranked genes and their relative expression are shown as a 'ranking metric score' in grey.* ***C)*** *Box plot showing*

Log_2 fold change values for GSK-3484862 120-hour cells compared to untreated cells. All genes (n=29,518), gene set A) (n=78) and gene set B) (n=152). Box, 25th–75th percentile; whiskers, 1.5× interquartile range; centre line, median. Student's t-test, p-values shown. D) Venn diagram showing overlap between the two gene sets. Generated using BioVenn (Hulsen, de Vlieg and Alkema, 2008).

5.11 GSK-3484862 triggers expression of HERV1 viral elements.

Having found that antiviral signalling pathways were enriched in GSK-3484862 treated cells, I was next interested in investigating why these pathways became activated. Antiviral response genes, such as those most upregulated between 48- and 120-hours GSK-3484862 treatment (Fig. 5.16), are not reported to be silenced by DNA methylation at their promoter regions (McGarvey *et al.*, 2008; Li *et al.*, 2014; Chiappinelli *et al.*, 2015). Therefore, it is possible that activation of immune signalling is a downstream effect of GSK-3484862 treatment. The 2'-5'-oligoadenylate synthetase (OAS) family of enzymes, OAS1, OAS2, OAS3 and OASL, are some of the most highly upregulated genes between 48- and 120-hours GSK-3484862 treatment (Fig. 5.16). OAS1-3 play a critical role in limiting viral propagation by detecting cytosolic viral double-stranded (ds) RNA and activating the enzyme RNase L, which degrades host and viral RNA (Hornung *et al.*, 2014). OASL mediates induction of interferon signalling through interaction of RIGI (Zhu *et al.*, 2014). RIGI also has a role in detecting cytosolic viral RNA, along with MDA5 (also known as IFIH1), and further enhancing immune activation. Both RIGI and MDA5 are upregulated after 120-hours GSK-3484862 exposure. Therefore, I was interested in exploring whether expression of dsRNA could be the mechanism through which interferon signalling is activated by GSK-3484862. As discussed earlier in this chapter, repetitive sequences and transposable elements are normally methylated in human cells to repress their activity (Rowe and Trono, 2011). Expression of transposable elements as dsRNA has been previously reported in response to 5-aza/5-aza-dC and this has been linked to activation of immune signalling (Chiappinelli *et al.*, 2015; Roulois *et al.*, 2015; Y. Kong *et al.*, 2019; Ohtani *et al.*, 2020).

I assessed the differential expression of transposable element transcripts in GSK-3484862 treated HCT116 cells. The raw RNA sequencing data was processed, and repetitive and transposable element transcripts were mapped, by Professor Ian Adams (see Methods

2.10.5). I then plotted differential expression changes between untreated HCT116 cells and those treated with GSK-3484862. Differential expression of transposable elements at 48 hours are shown in Figure 5.21. Thresholding the results for elements with a Log_2 fold change >2, we observe 1 significantly upregulated, Human Endogenous Retrovirus (HERV)-Fc2-int (padj <0.05). There were no significantly down-regulated elements detected that met a Log_2 fold change threshold of <-2. HERV-Fc2-int is a member of the ERV1 superfamily of endogenous retroviruses which are normally silenced by methylation (Kojima, 2018; Durnaoglu, Lee and Ahnn, 2021).

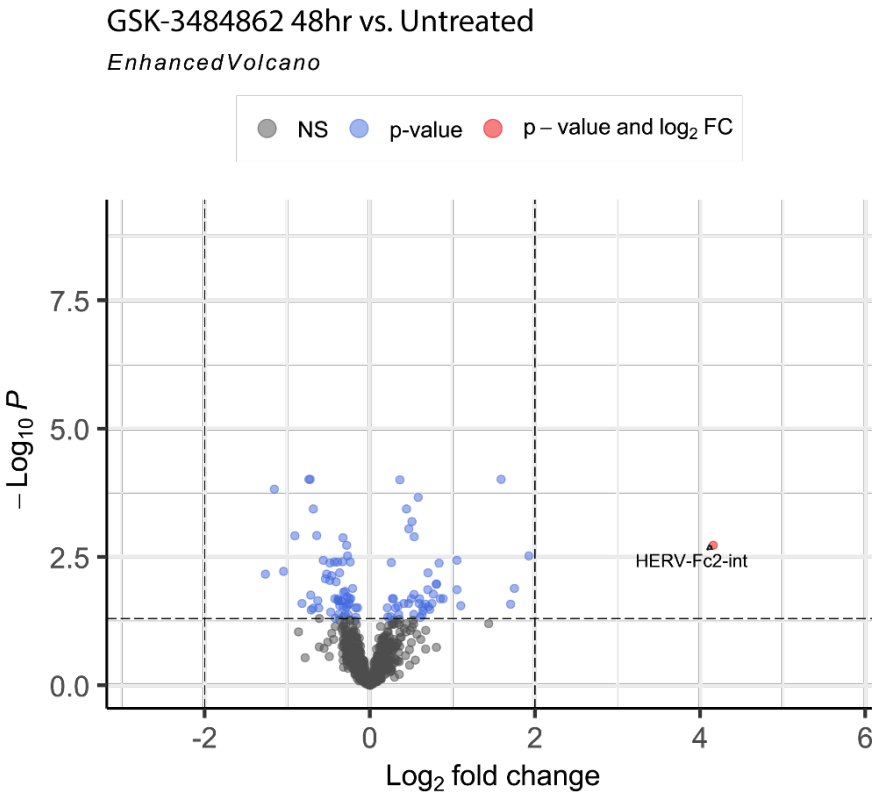


Figure 5-22: Differential expression of transposable elements induced by 48-hours GSK-3484862 treatment. *Enhanced volcano plot showing differentially expressed transposable elements in HCT116 cells treated with GSK-3484862 for 48 hours compared to untreated cells. HERV-Fc2-int was significantly up-regulated (padj<0.05, Log_2 fold change >2, red dot). Blue dots represent genes that were differentially expressed but did not meet the Log_2 fold change threshold. Black dots represent genes that were not significantly differentially expressed between conditions. Total elements shown = 843.*

Next, I assessed differential expression of transposable elements after 120-hours GSK-3484862 exposure (Fig. 5.22). Thresholding the results for genes with a Log_2 fold change >2, a total of 4 elements were found to be significantly up-regulated (padj <0.05). The most highly

differentially expressed element was HERV2-Fc2-int. The other up-regulated elements were also members of the ERV1 superfamily, HERV-Fc1-LTR1, HERV-Fc1-int and LTR12D. Again, there were no significantly down-regulated transposable elements that met the Log_2 fold change <-2 threshold. Together, these data indicate that hypomethylation induced by GSK-3484862 triggers expression of a subset of HERV1 family retrotransposons.

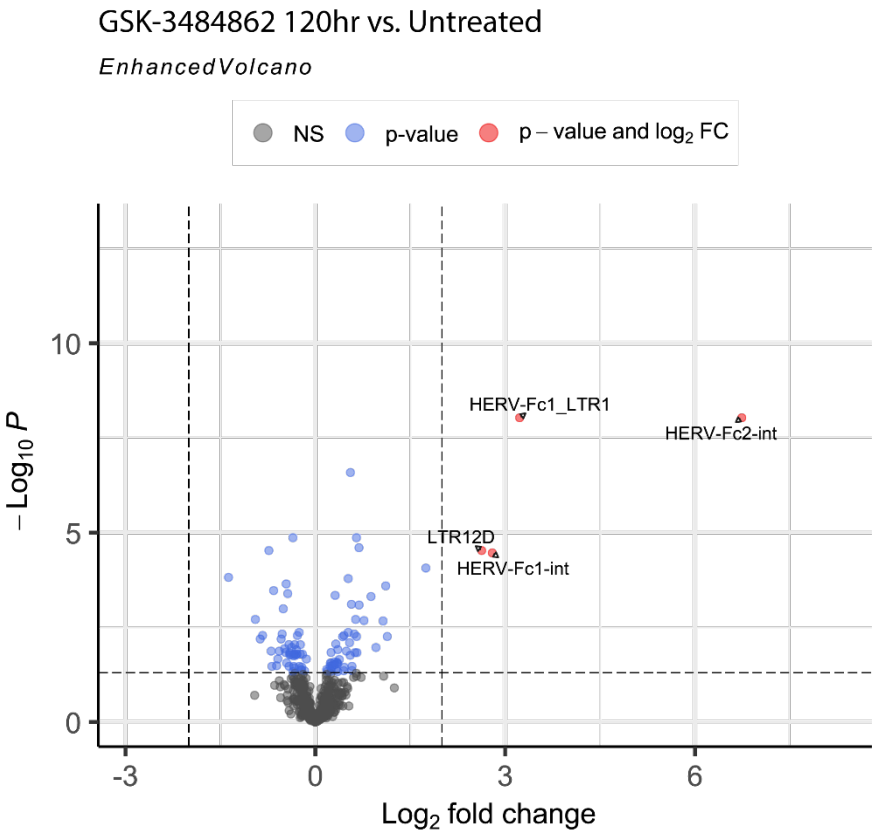


Figure 5-23: Differential expression of transposable elements induced by 120-hours GSK-3484862 treatment. *Enhanced volcano plot showing differentially expressed transposable elements in HCT116 cells treated with GSK-3484862 for 120 hours compared to untreated cells. Overall, 4 transposable elements were significantly up-regulated (padj<0.05, Log_2 fold change >2, red dots). Blue dots represent genes that were differentially expressed but did not meet the Log_2 fold change threshold. Black dots represent genes that were not significantly differentially expressed between conditions. Total elements shown = 843.*

5.12 Hypomethylation induced by GSK-3484862 triggers expression of HERV1 elements and a type-I interferon response

Several members of the ERV1 family transposable elements were shown to be transcriptionally upregulated, most notably HERV-Fc2-int (Fig. 5.21-5.22). In Figure 5.23A, I show Log_2 fold change values for transposable elements from 120-hour GSK-3484862 treated cells compared with untreated cells. Here, I have plotted all significantly differentially expressed (padj <0.05) transposable elements that met the Log_2 fold change threshold of >1, to look more broadly at the changes induced by GSK-3484862. At this threshold we observe upregulation of *Gypsy* transposable elements. *Gypsy* elements are inactive in mammalian genomes, although sequences derived from *Gypsy* are present and many have been domesticated to have roles in cellular functions, such as meiosis (Volff, 2009). The transcriptional regulator DAZL, which was shown be activated in response to GSK-3484862 treatment (Fig. 5.14-5.17), has been reported to regulate *Gypsy*-derived elements (Laureau *et al.*, 2021). Activation of DAZL may explain the increased expression of *Gypsy* elements in response to GSK-3484862. No upregulation of SatII or LINE1 sequences were observed, despite evidence that these sequences become hypomethylated following GSK-3484862 treatment (Fig. 5.2-5.4). ERV1 family members are the primary class of transposable element upregulated. Activation of ERV1 family elements has been previously reported in response to 5-aza/5-aza-dC treatment, further supporting that GSK-3484862 induces a similar cellular response as traditional HMAs (Chiappinelli *et al.*, 2015).

In Figure 5.23B, I have plotted the Log_2 fold change values (Log2 fold change >1, padj <0.05) for a selection of interferon response genes that have previously been reported to become upregulated in response to hypomethylation (Chiappinelli *et al.*, 2015; Irwin *et al.*, 2023). The activation of genes involved in sensing viral dsRNA and subsequent Type I interferon response by 5-aza is closely replicated by GSK-3484862. Taken together, the RNA-sequencing results presented in this chapter demonstrate that hypomethylation induced by GSK-3484862 elicits a similar transcriptional response in HCT116 cells as previously reported in 5-aza treated cancer cells.

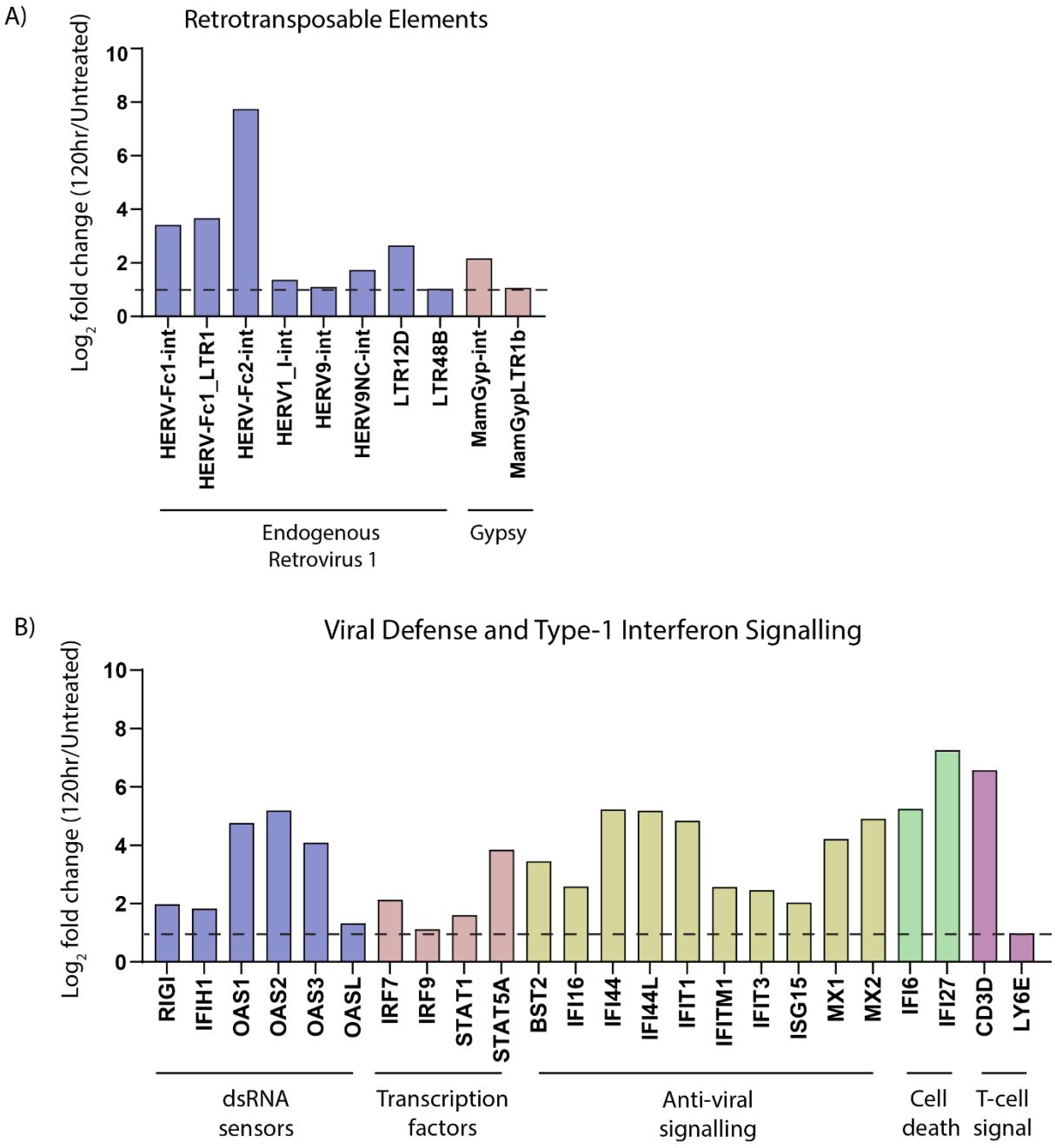


Figure 5-24: Hypomethylation induced by GSK-3484862 triggers expression of HERV1 elements and a type-I interferon response. A) *Bar plot showing Log_2 fold change (padj<0.05) values for all differentially expressed transposable elements with a Log_2 fold change >1 at 120-hours GSK-3484862 exposure.* ***B)*** *Bar plot showing Log_2 fold change (padj<0.05) values for a selection of antiviral and interferon type-I response genes at 120-hours GSK-3484862 exposure. Gene list selected based on genes previously reported to become activated in response to hypomethylation. Y-axis = Log_2 fold change (padj<0.05) for 120-hours GSK-3484862 compared to untreated cells. Dashed line at Log_2 fold change = 1.*

5.13 Discussion

In this chapter, I explored the potential application of the methylation-sensitive reporter system developed during this thesis as a drug screening tool, to identify HMAs. In cancer treatment HMAs have been shown to enhance patient responses to immunotherapy, but existing HMAs have limited use due to their toxicity and poor pharmacokinetics (Stresemann and Lyko, 2008). Therefore, it is of interest to identify new compounds that can induce hypomethylation. I provided proof-of-concept for this idea by demonstrating that the eGFPme reporter was activated in response to pharmacological inhibition of DNMT1 with GSK-3484862 (Fig. 5.6). Activation of the reporter gene coincided with a ~60% reduction in global methylation levels (Fig. 5.11). Importantly, hypomethylation by GSK-3484862 induced activation of immune signalling, which has been reported to be important for the therapeutic applications of HMAs (Li *et al.*, 2014; Chiappinelli *et al.*, 2015; Roulois *et al.*, 2015). These results show that hypomethylation leading to reporter activation also results in activation of immune signalling, indicating that this system could be utilised to screen for compounds that induce similar effects *in vitro*.

This work provides the first in-depth characterisation of the transcriptional response of HCT116 cells to inhibition of DNMT1 by GSK-3484862. I observed a bias towards transcriptional up-regulation, which is in line with inhibiting DNA methylation, which is typically considered to act as a repressive epigenetic signal. In Figures 5.14-5.17, I show that *Dazl* and CTA family genes, marker genes known to be repressed by promoter DNA methylation, were significantly upregulated by GSK-3484862 (Sproul *et al.*, 2011; Weon and Potts, 2015). In GSK-3484862-treated cells, I observed upregulation of genes involved in Type-I interferon (IFN-I) signalling, a pathway that is activated in response of detection of viral dsRNA in the cytosol (Fig. 5.16, 5.19-20 and 5.23). Activation of immune signalling is important for the anti-cancer activity of HMAs. Therefore, GSK-3484862 may be a promising candidate for therapeutic use. It has been suggested that immune signalling is induced by sensing of dsRNA formed by transposable element transcripts. To directly test whether this is the mechanism through which GSK-3484862 activates immune signalling, it would be interesting to knock-down sensors of dsRNA, such as RIGI, IFIH1 or the OAS genes, to determine whether this limits immune induction.

I showed that GSK-3484862 induces activation of ERV1 family transposable elements, which are normally repressed by methylation in human cells (Fig. 5.21-5.23). However, I did not directly test if HERV elements became demethylated following GSK-3484862 treatment. It would be of interest to assess this by EM-seq in future work. HERV activity has previously been associated with responses to immunotherapy (Smith *et al.*, 2018). Members of the ERV1 family HERV1-Fc2-int, HERV-Fc1-LTR, LTR12D and HERV-Fc2-int were the most upregulated, with HERV-Fc2-int displaying the highest Log_2 fold change in 120-hour GSK-3484862 treated cells. This is in line with published work that has reported activation of ERV1 family transposons in response to hypomethylation (Chiappinelli *et al.*, 2015; Y. Kong *et al.*, 2019). Chiappinelli *et al*. (2019) reported that HERV-Fc2 transcripts were highly upregulated in response to 5-aza. Furthermore, a novel immunomodulatory drug currently in Phase III clinical trials was reported to activate HERV-Fc2 expression through demethylation, and increase interferon signalling (Zhao *et al.*, 2017; Oronsky *et al.*, 2019). This further implicates the expression of HERV-Fc2 elements in the anti-cancer activity of HMAs. Expression of LTR12D in GSK-3484862 treated cells is also of interest. It has been suggested that LTR12-derived peptides can be presented as antigens on the cell surface, which may help improve immune recognition of cancer cells (Y. Kong *et al.*, 2019; Goyal *et al.*, 2023). Together, my findings support that demethylating agents activate ERV1 elements and provoke an immune response, which may represent a viable therapeutic avenue for cancer treatment. However, although HERV elements have been widely reported to become upregulated in response to hypomethylation, whether HERV elements are directly responsible for activation of immune signalling remains unclear. Chiappinelli *et al.* (2015) reported that over-expression of vectors containing ERV cDNA genes increased the same IFN genes as 5-Aza treatment. Conversely, one study implicated Alu elements as the primary immunogenic dsRNA induced by hypomethylation, as opposed to HERV (Mehdipour *et al.*, 2020). Immune stimulation by 5-aza/5-azadC is dependent on the pattern recognition receptor MDA5/IFIH1 (Goelz *et al.*, 1985; Chiappinelli *et al.*, 2015; Roulois *et al.*, 2015). It would be interesting to test whether HERV elements alone can stimulate immune signalling in HCT116 cells, by overexpressing HERV RNA or cDNA *in vitro*.

I investigated the temporal dynamics of the transcriptional response to GSK-3484862. My results indicate that immune signalling was most upregulated between 48- to 120-hours GSK-

3484862 exposure (Fig. 5.14-5.16). This further supports that immune signalling is a downstream response to DNA hypomethylation. An important challenge of developing new therapeutics is identifying a treatment window. My results indicate that immune activation in response to GSK-3484862 is time-dependant, which may help inform the optimum window for immunotherapy responsiveness.

Inhibiting UHRF1 has been reported to induce expression of HERV elements and immune signalling, and may therefore represent a promising drug target (Irwin *et al.*, 2023). I utilised the eGFPme reporter system to screen three proposed inhibitors of UHRF1, NSC232005, NSC34716 or NSC43513. These compounds were identified in an *in-silico* screen by Myrianthopoulos *et al*. (2016). None of these compounds activated reporter expression, suggesting that they were not able to efficiently induce hypomethylation through inhibition of UHRF1 (Fig. 5.9). It is possible that the reporter system is not sensitive enough to detect subtle changes to DNA methylation levels. However, reporter activation was detected within 48-hours in response to GSK-3484862, which supports that that the system is rapidly responsive to hypomethylation. Furthermore, reporter activation was observed in response to depletion of UHRF1 in UHRF1-AID cells, demonstrating that the eGFPme reporter detects loss of UHRF1 activity. Although target-based drug discovery approaches are valuable, this result provides support for the idea that unbiased phenotypic screening using the eGFPme reporter could be a beneficial method to identify cell-active hypomethylating compounds. Furthermore, the eGFPme reporter could also be employed to test other UHRF1 inhibitors.

There are some caveats to results presented in this Chapter which should be acknowledged. Firstly, I explored the transcriptional response of HCT116 cells to GSK-3484862, a novel small molecule inhibitor identified by Pappalardi *et al*. (2021). However, Pappalardi *et al*. (2021) developed a second DNMT1 inhibitor, GSK-3685032. This compound was optimised to improve potency, solubility and lipophilicity, and may therefore more suitable for pharmacological use. However, at the time of this study GSK-3685032 was not commercially available. In MV4-11 cells, GSK-3484862 and GSK-3685032 induced similar gene expression changes (Pappalardi *et al.*, 2021). GSK-3685032 also induced HERV expression, and displayed improved *in vivo* efficacy in mouse models compared to 5-aza-dC(Pappalardi *et al.*, 2021). This supports that small-molecule inhibitors of DNMT1 may be more effective in cancer treatment than cytidine analogues.

I conducted my experiments in HCT116 colorectal cancer cells, as this is the cell line that I initially established the eGFPme reporter system. However, HMAs have been primarily used to treat the hematologic malignancies, such as acute myeloid leukaemia (AML) and myelodysplastic syndromes (MDS) (Stomper *et al.*, 2021). Therefore, if the eGFPme reporter system were to be utilised as a HMA drug screening tool, I suggest that this is additionally done in an AML cell line, such as MV4-11, and an MDS cell line, such as SKM-1. This would enable compound screening to be carried out in more disease-relevant cell lines. Nonetheless, valuable insights were gained from initial screening in HCT116 cells. There is ongoing interest surrounding the use of HMAs for the treatment of solid tumours, particularly in CRC. Clinical data has indicated that the lack of success in CRC patients is due to insufficient hypomethylation, likely due to challenges with delivery and pharmacokinetics (Linnekamp *et al.*, 2021). Pre-clinical data has demonstrated that colon, breast and lung cancer cells, among others, respond to HMAs with an interferon response (Li et al., 2014; Wrangle *et al*., 2013). Therefore, it is essential to identify hypomethylating compounds that could be more efficient for the treatment of solid tumours, and this highlights the clinical relevance of the data presented in this thesis.

6 Conclusions and future perspectives

In this thesis, I developed and characterised a methylation-sensitive reporter system, as a tool to study the dynamics of DNA methylation maintenance *in vitro*. I utilised this system to examine the fidelity of DNA methylation maintenance in HCT116 cells over time. My results indicate that despite the widespread hypomethylation observed in these cells, a hallmark of human cancers, DNA methylation is largely maintained at synthetic reporter sequences. These findings highlight the functionality of the DNA methylation maintenance machinery in HCT116 cells, suggesting that improper maintenance may not be the driving force underlying DNA hypomethylation in cancer.

To test whether the eGFPme reporter could accurately report DNA methylation changes, I assessed reporter activity in cell lines lacking activity of the DNMTs or UHRF1. Reporter activation was observed in cell lines lacking DNMT1 or UHRF1 activity, but not in DNMT3B^{KO} cells. These results suggest that loss of DNMT3B activity alone has minimal impact on genome-wide DNA methylation maintenance, in line with previous findings. Expression of the reporter was shown to coincide with incomplete methylation maintenance, demonstrating that the eGFPme reporter accurately responds to changes in methylation levels. This indicates that the reporter system can be used to visualise genomic methylation levels and obtain a readout of methylation changes in real-time. Together, these results support that the eGFPme reporter could be utilised as a screening platform to identify novel factors that contribute to DNA methylation maintenance.

Following on from this, I was interested in exploring the potential application of the eGFPme reporter system as a phenotypic drug screening platform. My previous results indicated that targeting DNMT1 or UHRF1 for inhibition was an efficient method to induce hypomethylation and reporter activation. Therefore, to investigate whether the methylation-sensitive reporter system developed in this study could be utilised as a drug screening tool, I tested the response of the eGFPme reporter to pharmacological inhibition of DNMT1. Rapid reporter activation was observed in response to DNMT1 inhibition, providing proof of concept for this idea. Furthermore, reporter activation was accompanied by HERV expression and immune activation in inhibitor-treated cells. These factors are crucial for the clinical application of HMAs. Taken together, these findings indicate that the methylation-sensitive fluorescent

reporter system developed in this study could provide an effective screening platform for the identification of hypomethylating and immuno-stimulatory compounds.

6.1 Methylation-sensitive reporter system as a tool to study DNA methylation maintenance

The methylation-sensitive reporter system developed in this study allows tracking of genome-wide methylation maintenance in real-time. The design of the system takes advantage of the eGFP fluorescent reporter, which enables convenient, quantitative readout of reporter gene expression under the florescence microscope or by flow cytometry (Soboleski, Oaks and Halford, 2005). The reporter is under control of the synthetic CAG promoter, a CGI promoter, which has been shown to efficiently drive gene expression in a range of cell types (Jun-ichi *et al.*, 1989; Alexopoulou, Couchman and Whiteford, 2008; Kang *et al.*, 2022). Thus, this reporter construct could be utilised in a range of cell models. I demonstrated that expression of the reporter is repressed by methylation at the promoter and gene body, and that loss of this methylation leads to activation of eGFP expression (Chapter 3, Figure 3.5; Chapter 4, Figure 4.3). The reporter cassette is not targeted for *de novo* methylation, as indicated by the lack of methylation gain at the unmethylated reporter sequence (Chapter 3, Figure 3.8.2). This is in contrast to some CGI sequences with tissue-specific expression patterns, such as the Dazl-associated CGI promoter, which can become *de novo* methylated following ectopic integration in ESCs (Sabag *et al.*, 2014; Stelzer *et al.*, 2015). Therefore, repression of the reporter is dependent on maintenance of DNA methylation, meaning that this system can be used to track methylation maintenance fidelity *in vitro*.

6.2 Understanding the dynamics of DNA methylation maintenance in cancer

Existing work exploring hypomethylation in cancer has been based on static views of the DNA methylation landscape in tumours or cancer cell lines, obtained through bulk whole-genome bisulfite sequencing. Some studies have suggested that loss of methylation in cancer is a progressive process, arising through incomplete maintenance of methylation over successive

cell divisions (Zhou *et al.*, 2018). However, DNA methylation is a dynamic process, and these studies have not directly tested the ability of cancer cells to propagate methylation patterns across the genome. In this study, I attempted to address this question. My results demonstrate that DNA methylation was largely maintained at synthetic reporters over an 8-week period, equivalent to approximately 75 cell divisions (Chapter 3). This finding indicates that HCT116 cells are capable of maintaining methylation states over prolonged periods. Therefore, my data does not support the model for progressive methylation loss in cancer cells (Fig 6.1A). However, it should be noted that this result is within the context of a synthetic reporter construct, and maintenance of methylation at endogenous sequences could differ.

If PMD formation is not a progressive process, then the formation of PMDs could occur at a specific early stage in the cancer transformation process. Analysis of abnormal non-neoplastic tissue and early-stage tumours has shown that DNA hypomethylation is present at a very early stage in cancer development (Goelz *et al.*, 1985; Feinberg *et al.*, 1988; Jackson *et al.*, 2004; Teschendorff *et al.*, 2012). This evidence supports a model of hypomethylation in cancer that is not based on progressive loss of methylation coupled to cell proliferation. In this model, a reduced methylation level could be established, for example through active de-methylation by the TET enzymes (Fig. 6.1B). Lower methylation levels could then become fixed through maintenance of the new, reduced methylation state. Alternatively, a continual de-methylation by TETs combined with *de novo* and maintenance DNMT activity may maintain partial methylation in PMDs. This could explain the highly disordered methylation seen in PMDs (Gaidatzis *et al.*, 2014).

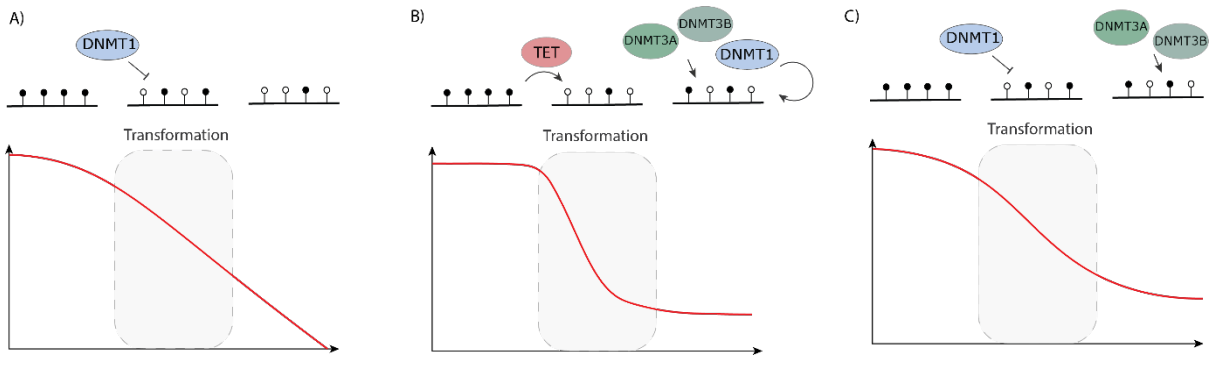


Figure 6-1: Models for hypomethylation in cancer. A) *Hypomethylation as a progressive process wherein DNMT1 fails to maintain methylation at some genomic regions, leading to passive loss of methylation over time.* ***B)*** *Hypomethylation as an active process occurring during the cancer transformation process. For example, TET enzymes could actively remove*

DNA methylation. However, continual demethylation may not be beneficial, and methylation could be re-established by the de novo *methyltransferases and maintained by DNMT1.* ***C)*** *Combined model wherein DNA hypomethylation is a progressive process resulting from incomplete maintenance by DNMT1, but the DNMT3 enzymes compensate and re-establish or maintain methylation.*

I had initially anticipated that whatever process results in methylation loss would likely be ongoing in the cells, considering that the cancer clonally expands and should retain the genetic alterations of its predecessors (Yates and Campbell, 2012; Abbosh *et al.*, 2017). However, tumours are highly heterogenous and display plasticity (Boumahdi and de Sauvage, 2020). Furthermore, hypomethylation in cancer has not been linked to a specific genetic alteration, and epigenetic mechanisms are plastic in nature (Flavahan, Gaskell and Bernstein, 2017). Studies have shown that complete removal of DNMT1 leads to extensive hypomethylation and is lethal to cancer cells (Kulaeva *et al.*, 2003; Chen *et al.*, 2007). Therefore, it could be that, at an early stage of cancer development, DNA hypomethylation is progressive and promotes genetic instability and cancer development (Sheaffer, Elliott and Kaestner, 2016; Brinkman *et al.*, 2019), but continual demethylation is detrimental to cells. This may mean that the cancer must reach an equilibrium or steady state that maintains some level of DNA methylation, possibly through compensation by the DNMT3 enzymes (Fig. 6.1C). This may be required to maintain repression of repeats and transposable elements, which would otherwise promote anti-tumour immunity, as evidenced by my work and others (Chiappinelli *et al.*, 2015; Ohtani *et al.*, 2020). To elucidate this process, it would be interesting to assess the expression and localisation of epigenetic factors, for example the DNMTs and TET enzymes, at different stages throughout tumour progression. This would likely be unfeasible in human patients because tumour biopsy is an invasive process, but blood samples from leukaemia patients or an *in vivo* mouse model could be utilised as an alternative.

Another interesting way to test this model would be to track methylation maintenance across the cancer transformation process *in vitro*. This could be done by introducing the eGFPme reporter into a primary cell line and subsequently inducing immortalization. Induction of carcinogenesis *in vitro* has been used as a method to study mutational landscapes in human cancers (Olivier *et al.*, 2014). Thus, taking a similar approach to explore alterations to the

epigenetic landscape in cancer would be very interesting. Expression of the reporter construct could be monitored, to determine whether the immortalization process alters the cells' ability to maintain DNA methylation. Live-cell imaging could be utilised to monitor reporter expression in real-time, as a proxy for methylation levels, in order to learn more about the dynamics of this process.

The DNA sequence context and CpG density has additionally been evidenced to influence methylation maintenance fidelity. DNMT1 has been shown to be more efficient in the presence of multiple CpG sites (Hermann, Goyal and Jeltsch, 2004), and displays a preference towards CpG sites that are flanked by a cytosine at the -1 position (Adam *et al.*, 2020). PMDs have a low CpG density and are enriched for CpG sites flanked by A or T, which has been suggested to influence hypomethylation at these regions. It would be interesting to explore the influence of CpG density on methylation maintenance within the methylation-sensitive reporter system. This could be done by utilising an alternative low-CpG reporter construct. The Kudla research group (IGC) have generated a library of synonymous eGFP variants that differ in CpG content (Mordstein *et al.*, 2021). These eGFP variants could be utilised within the methylation-sensitive reporter system, to investigate the impact of CpG density of DNA methylation maintenance.

6.3 Immunostimulatory effects of hypomethylating agents in cancer and therapeutic implications

Inhibiting DNA methylation has shown clinical benefits in cancer treatment, particularly in enhancing patient responses to immunomodulatory drugs (Daver *et al.*, 2018; Apostolova *et al.*, 2023). Demethylation appears to activate expression of HERV elements, which are normally repressed by methylation. Activation of HERV elements is proposed to induce immune signalling (Chiappinelli *et al.*, 2015; Kong *et al.*, 2019; Ohtani *et al.*, 2020), potentially leading to the presentation of HERV peptides on the cell surface, further promoting immune recognition (Smith *et al.*, 2018; Kong *et al.*, 2019). My data supports that DNA hypomethylation activates HERV expression and IFN-I signalling in cancer cells (Fig. 6.2). Furthermore, my data indicates that that the new generation of non-nucleoside DNMT1

inhibitors could be effective in cancer treatment by sensitising patients to immunotherapy (Chapter 5).

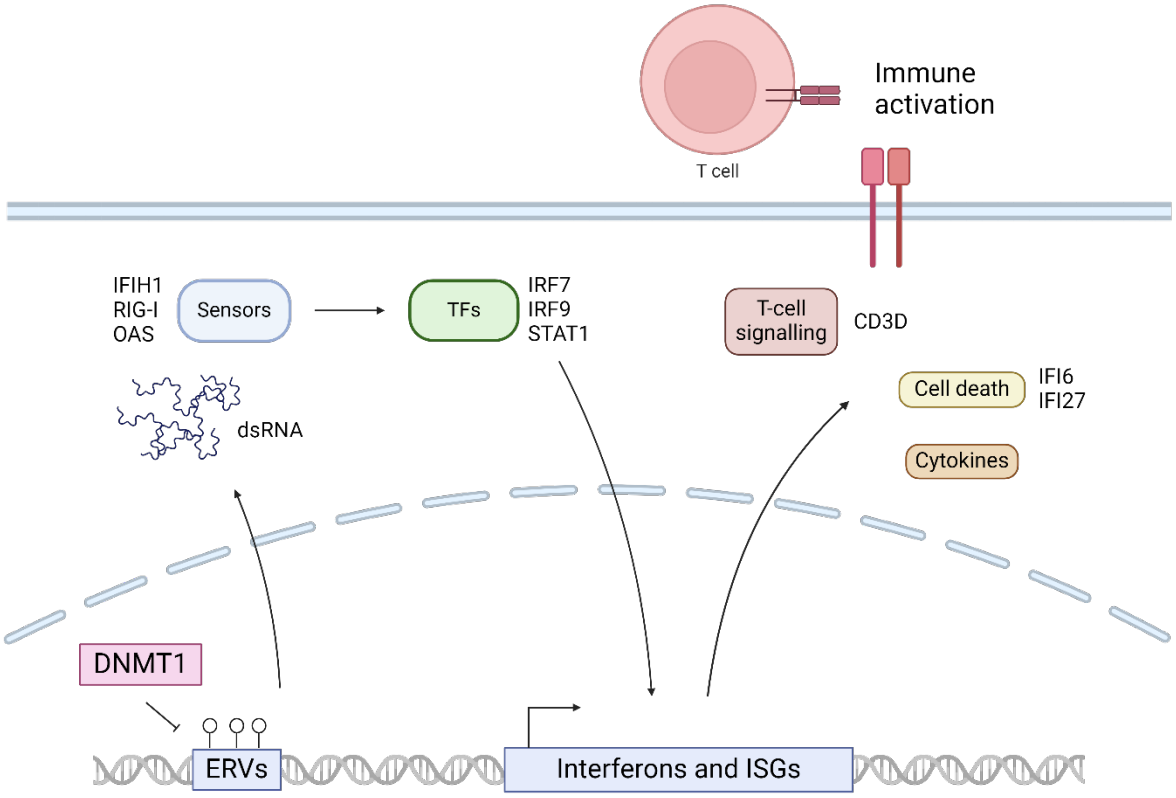


Figure 6-2: Model for DNMT1 inhibition and subsequent activation of HERV elements and immune signalling. *Based on my RNA-sequencing data, following inhibition of DNMT1 with GSK-3484862, I observe activation of HERV elements and anti-viral immune signalling. I propose that HERV elements could become hypomethylated, leading to translation of the sequence as RNA. HERV elements may then go on to form dsRNA molecules, which could be detected in the cytoplasm by pattern recognition receptor IFIH1/MDA5 and other cytosolic sensors of foreign RNA, such as RIGI and the OAS genes. This would trigger immune signalling, leading to activation of interferons and interferon-stimulated genes (ISGs). Subsequent expression of apoptotic factors, cytokines and T-cell signalling could promote immune recognition and infiltration into the tumour microenvironment. Genes highlighted in this figure are upregulated in response GSK-3484862 in HCT116 cells (Chapter 5).*

It has not yet been robustly demonstrated that expression of HERV RNA transcripts is directly responsible for immune activation. Overexpression of vectors containing *ERV-3, EnvW2* and *Syncytin-1* have been reported to activate interferon genes (Chiappinelli *et al*., 2015). Conversely, others have implicated Alu elements as the primary drivers of hypomethylation-induced immune response (Mehdipour *et al.*, 2020). Immune activation in response to hypomethylation is dependent on the activity of MDA5/IFIH1, an RNA sensor and pattern

recognition receptor (Chiappinelli *et al.*, 2015; Roulois *et al.*, 2015). It would be interesting to test whether HERV elements can stimulate immune activation in HCT116 cells, by overexpressing HERV RNA elements *in vitro*. If expression HERV elements is sufficient to activate an immune response, it would be of interest to determine whether this is reliant on MDA5/IFIH1. This could be done by testing if knock-down of MDA5/IFIH1 impedes HERV-mediated immune activation.

Finally, immune activation and inflammation can be a double-edged sword. While IFN signalling can potentiate responses to immunotherapy and promote immune infiltration into the tumour microenvironment, prolonged inflammation can lead to immune cell exhaustion (Chow *et al.*, 2022; Zhang *et al.*, 2021). DNA hypomethylation has been evidenced to stimulate cytotoxic T-cell activity and may be able to promote reinvigoration of exhausted T-cells (Loo Yau *et al.*, 2021). However, demethylation at the promoter of PD-1, an immunoinhibitory receptor, has also been reported in response to HMA treatment in patients. PD-1 promoter demethylation was linked poor outcomes, T-cell exhaustion and may represent a mode of resistance to HMAs (Ørskov *et al.*, 2015). Combined treatment with HMAs and PD-1 inhibitors has shown positive clinical responses (Apostolova *et al.*, 2023), as have alternative combination treatments (Stomper *et al.*, 2021). This should be considered when designing patient treatment plans.

Furthermore, therapeutic treatment with DNA hypomethylating drugs could induce an IFN response in normal tissues. While immune activation can be beneficial, excessive activity in normal tissues may lead to uncontrolled inflammation or autoimmunity, potentially causing damage to healthy cells and tissues (Chen *et al*., 2017). As evidenced in Chapter 5, treatment of HCT116 with GSK-3484862 resulted in upregulation of pro-inflammatory Type-I IFN signalling and cytokine production. Excessive inflammatory cytokine production has been associated with tissue damage, metabolic dysfunction, organ failure and cancer promotion (Chen *et al*., 2017; Greten and Grivennikov, 2019). Healthy cells have been reported to be less sensitive to the effects of hypomethylating drugs, potentially due to the higher levels of methylation present in these cells compared to cancer cells (Píškala and Šorm, 1964). However, the effects of DNA hypomethylating compounds on healthy tissues requires further investigation. The potential negative outcomes of IFN signalling represent a safety concern that should be closely monitored in patients.

6.4 Future perspectives

Despite the promising immunostimulatory effects of DNA hypomethylation, cytotoxicity, poor pharmacokinetics and emerging resistance are major limitations of existing demethylating drugs, highlighting the need for improved alternatives. The methylation-sensitive reporter system developed in this study accurately reports genome-wide hypomethylation and represents a useful phenotypic screening tool to identify hypomethylating compounds. Phenotypic drug discovery has regained popularity in recent years, as it has proven effective in identifying first-in-class drugs with novel mechanisms of action (Vincent *et al.*, 2022). More first-in-class drugs were identified through phenotypic screens than target-based approaches between 1999 to 2008 (Swinney and Anthony, 2011). This approach involves screening compounds based on their effect on disease-relevant phenotypes, rather than targeting specific genes or molecular pathways. While target-based approaches offer the benefit of understanding the drugs mechanism of action, many compounds identified *in silico* do not ultimately function *in vivo*. This is evidenced by my results in Chapter 5, where I show that the proposed inhibitors of UHRF1 identified *in silico* did not appear to be active in cells, providing justification for a phenotype-based screening approach. Notably, there are currently only two known drug targets for inducing global hypomethylation, DNMT1 and UHRF1. Leveraging the eGFPme reporter as a screening tool, whether in a CRISPR knock-out screen or as a drug screening platform, could allow identification of novel drug targets. Some remaining challenges of a phenotype-based screening approach would include hit validation, and confirmation of hypomethylation and immune activation. Nonetheless, utilising the eGFPme reporter could provide an efficient and unbiased platform to identify new cell-active compounds that can induce hypomethylation.

An essential requirement for a successful screening assay is rigorous characterisation and optimisation, to improve robustness, reproducibility and signal interpretation. Efforts should be made to optimise the transfection process, to minimise variability between experimental replicates and cell lines. Transfection by electroporation using the MaxCyte electroporation system was tested during this study and found to be an efficient method to transfect many cells in a single experiment. The MaxCyte is reportedly less toxic, designed for clinical use, and may represent a promising alternative to chemical transfection (Fratantoni *et al*., 2003).

Appropriate controls for screening should also be considered. A positive control for reporter activation that produces a well-characterised response within the system, for example GSK-3484862, should be utilised. Similarly, a negative control that does not induce hypomethylation should be used. DMSO as a negative control should be carefully tested, to determine whether the demethylation detected in DMSO-treated cells (Chapter 5) represents true demethylation caused by DMSO treatment. Furthermore, internal unmethylated and methylated controls should be characterised for each cell line that the reporter is introduced to.

One limitation of the current system is that only a fraction of the total reporter insertion sites have been mapped (Chapter 3). The piggyBac transposon has been shown to mediate thousands of integrations in a single pool of transfected cells (Akhtar *et al.*, 2013, 2014). In future work, I suggest that a higher throughput sequencing method, such as the iSeq method employed in this study, is utilised to map the integrated reporter loci. This would enable more detailed characterisation of the piggyBac integration landscape and better aid the interpretation of results. In addition, the chromatin state of integrated reporters should be investigated. The PB_CAG_eGFP_IRES_Puro reporter construct is 4.38kb and could influence its own chromatin environment, or the chromatin environment of the region surrounding the insertion site. Alternatively, the reporter cassette may adopt the chromatin environment of the region it lands within. To better understand this, Chromatin Immunoprecipitation (ChIP)-qPCR could be utilised to assess the chromatin marks present at integrated reporters. It would also be of interest to investigate the chromatin state of unmethylated reporters compared to methylated reporters. Additionally, the chromatin state of methylated reporters when methylated and following removal of methylation, for example after treatment with GSK-3484862, should be tested to gain insight into the impact of hypomethylation on the chromatin marks present at integrated reporters.

An additional caveat of the current approach is that the mixed population of cells containing random reporter integrations may complicate signal interpretation. A site-directed integration approach, such as Recombination-Mediated Cassette Exchange (RMCE), could be used to introduce a single copy of the reporter construct to a designated genomic locus. Alternatively, single clones from a pool of piggyBac integrations could be isolated and expanded. The reporter integrations contained within these lines could then be characterised

in detail, by mapping the insert sites and assessing chromatin marks present at the reporter. These clonal lines could then be maintained separately or pooled together as needed. This would allow the formation of a library of known reporter integrations with the option to customise screening pools. Screening pools could be varied to contain reporter integrations within different genomic locations, depending on experimental requirements. Together, these refinements will enable the eGFPme reporter system to be developed as a viable screening platform.

Appendix

6.5 PCR primers

	Name	Sequence (5'-3')
BS/EM-PCR	bs1-TNFRSF1A-F	GTTTTGTTGGTGAGGATATTTAAAA
	bs1-TNFRSF1A-R	CAAAAACAAACAAAAACTAACAACC
	bs1-BRCA2-F	GGGGAATAGGTTTTGAGAGAATATT
	bs1-BRCA2-R	CCAAACCACCCTACTTAAAAAAAC
	bs2-BRCA2-F	TATGTTTGAGAGAAAGGTTTTTGTT
	bs2-BRCA2-R	TATTTCCAAAAAATTTCAAATAAC
	emPCR_eGFP_F	GTAAGTTGATTTTGAAGTTTATTTGTA
	emPCR_eGFP_R	AACTTATACCCCAAAATATTACC
	emPCR_CMV_F	GGGGTTATTAGTTTATAGTTTATATATGGA
	emPCR_CMV_R	ATACTACCAAATAAAAAAATCCCATAAAAT
qPCR	hBRCA2-2-F	AGTGGTGGTGGTAGTGGGTTGG
	hBRCA2-2-R	TTGGCAGAGACAAAAGGGCAAG
qPCR (Akhtar et al. 2014)	qPCR_eGFP_F_TRIP	CGACAACCACTACCTGAGCA
	qPCR_eGFP_R_TRIP	GAACTCCAGCAGGACCATGT
	qPCR_Amp_F_TRIP	CGATCGTTGTCAGAAGTAAGTTGG
	qPCR_Amp_R_TRIP	CACAGAAAAGCATCTTACGGATGG
Mapping piggyBac insert (Martin, 2017; Masalmeh, 2018)	PL1-MIN	CATGCGTCAATTTTACGCATGATTATCTTTAACG
	PR1-MIN	GCATGCGTCAATTTTACGCAGACTATCTTTCTA
	PL2-MIN	CATGCGTCAATTTTACGCATGATTATCTTTAACG
	PR2-MIN	GCATGCGTCAATTTTACGCAG
	KmonP-N7-CTCAG	GTACGAGAATCGCTGTCCTNNNNNNNCTCAG
	KmonP-N7-TCCTG	GTACGAGAATCGCTGTCCTNNNNNNNTCCTG
	KmonP	GTACGAGAATCGCTGTCCT

6.6 iSeq loci-specific primers

	Name	Sequence (5'-3') with adaptors and UMI
iSeq locus specific primers	eGFP_emPCR_iSeq_F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGTAAGTTGATTTTGAAGTTTATTTGTA
	eGFP_emPCR_iSeq_R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNAACTTATACCCCAAAATATTACC
	CMVe_emPCR_iSeq_F1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNGGGGTTATTAGTTTATAGTTTATATATGGA
	CMVe_emPCR_iSeq_R1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNATACTACCAAATAAAAAAATCCCATAAAAT

6.7 Illumina UDP library amplification primers

UDP0193_i1	CAAGCAGAAGACGGCATACGAGATTATCATGAGAGTCTCGTGGGCTCGG
UDP0194_i1	CAAGCAGAAGACGGCATACGAGATCTTGGCCTCGGTCTCGTGGGCTCGG
UDP0195_i1	CAAGCAGAAGACGGCATACGAGATGTCTCGTGAAGTCTCGTGGGCTCGG
UDP0196_i1	CAAGCAGAAGACGGCATACGAGATCCATCCACGCGTCTCGTGGGCTCGG
UDP0197_i1	CAAGCAGAAGACGGCATACGAGATACAACCAGGAGTCTCGTGGGCTCGG
UDP0198_i1	CAAGCAGAAGACGGCATACGAGATAGCAGAATTAGTCTCGTGGGCTCGG
UDP0199_i1	CAAGCAGAAGACGGCATACGAGATCAGTCGTGCGGTCTCGTGGGCTCGG
UDP0200_i1	CAAGCAGAAGACGGCATACGAGATGTCTAACCTCGTCTCGTGGGCTCGG
UDP0201_i1	CAAGCAGAAGACGGCATACGAGATGAACTCGGTTGTCTCGTGGGCTCGG
UDP0202_i1	CAAGCAGAAGACGGCATACGAGATAGTTATCACAGTCTCGTGGGCTCGG
UDP0203_i1	CAAGCAGAAGACGGCATACGAGATGTAGCATACTGTCTCGTGGGCTCGG
UDP0204_i1	CAAGCAGAAGACGGCATACGAGATCTTCAGTTACGTCTCGTGGGCTCGG
UDP0205_i1	CAAGCAGAAGACGGCATACGAGATAGTCCGAGGAGTCTCGTGGGCTCGG
UDP0206_i1	CAAGCAGAAGACGGCATACGAGATACAGTTCCAGGTCTCGTGGGCTCGG
UDP0207_i1	CAAGCAGAAGACGGCATACGAGATCCGCATATTCGTCTCGTGGGCTCGG
UDP0208_i1	CAAGCAGAAGACGGCATACGAGATTTATCCGATCGTCTCGTGGGCTCGG
UDP0209_i1	CAAGCAGAAGACGGCATACGAGATATAGTCTAGCGTCTCGTGGGCTCGG
UDP0210_i1	CAAGCAGAAGACGGCATACGAGATTATAGTAGCTGTCTCGTGGGCTCGG
UDP0193_i2	AATGATACGGCGACCACCGAGATCTACACAACACGTGGATCGTCGGCAGCGTC
UDP0194_i2	AATGATACGGCGACCACCGAGATCTACACGTGTTACCGGTCGTCGGCAGCGTC
UDP0195_i2	AATGATACGGCGACCACCGAGATCTACACAGATTGTTACTCGTCGGCAGCGTC
UDP0196_i2	AATGATACGGCGACCACCGAGATCTACACTTGACCAATGTCGTCGGCAGCGTC
UDP0197_i2	AATGATACGGCGACCACCGAGATCTACACCTGACCGGCATCGTCGGCAGCGTC
UDP0198_i2	AATGATACGGCGACCACCGAGATCTACACTCTCATCAATTCGTCGGCAGCGTC
UDP0199_i2	AATGATACGGCGACCACCGAGATCTACACGGACCAACAGTCGTCGGCAGCGTC
UDP0200_i2	AATGATACGGCGACCACCGAGATCTACACAATGTATTGCTCGTCGGCAGCGTC
UDP0201_i2	AATGATACGGCGACCACCGAGATCTACACGATCTCTGGATCGTCGGCAGCGTC
UDP0202_i2	AATGATACGGCGACCACCGAGATCTACACCAGGCGCCATTCGTCGGCAGCGTC
UDP0203_i2	AATGATACGGCGACCACCGAGATCTACACTTAATAGACCTCGTCGGCAGCGTC
UDP0204_i2	AATGATACGGCGACCACCGAGATCTACACGGAGTCGCGATCGTCGGCAGCGTC
UDP0205_i2	AATGATACGGCGACCACCGAGATCTACACAACGCCAGAGTCGTCGGCAGCGTC
UDP0206_i2	AATGATACGGCGACCACCGAGATCTACACCGTAATTAACTCGTCGGCAGCGTC
UDP0207_i2	AATGATACGGCGACCACCGAGATCTACACACGAGACTGATCGTCGGCAGCGTC
UDP0208_i2	AATGATACGGCGACCACCGAGATCTACACGTATCGGCCGTCGTCGGCAGCGTC
UDP0209_i2	AATGATACGGCGACCACCGAGATCTACACAATACGACATTCGTCGGCAGCGTC
UDP0210_i2	AATGATACGGCGACCACCGAGATCTACACGTTATATGGCTCGTCGGCAGCGTC

References

Abbosh, C. *et al.* (2017) 'Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution', *Nature*, 545(7655), pp. 446–451. Available at: https://doi.org/10.1038/nature22364.

Achour, M. *et al.* (2008) 'The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression', *Oncogene*, 27(15), pp. 2187–2197. Available at: https://doi.org/10.1038/sj.onc.1210855.

Adam, S. *et al.* (2020) 'DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation', *Nature Communications*, 11(1), p. 3723. Available at: https://doi.org/10.1038/s41467-020-17531-8.

Akhtar, W. *et al.* (2013) 'Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel', *Cell*, 154(4), pp. 914–927. Available at: https://doi.org/10.1016/j.cell.2013.07.018.

Akhtar, W. *et al.* (2014) 'Using TRIP for genome-wide position effect analysis in cultured cells', *Nature Protocols*, 9(6), pp. 1255–1281. Available at: https://doi.org/10.1038/nprot.2014.072.

Alberts, B. *et al.* (2002) 'The Structure and Function of DNA', in *Molecular Biology of the Cell. 4th edition*. Garland Science. Available at: https://www.ncbi.nlm.nih.gov/books/NBK26821/ (Accessed: 15 August 2024).

Alcazer, V., Bonaventura, P. and Depil, S. (2020) 'Human Endogenous Retroviruses (HERVs): Shaping the Innate Immune Response in Cancers', *Cancers*, 12(3), p. 610. Available at: https://doi.org/10.3390/cancers12030610.

Alexopoulou, A.N., Couchman, J.R. and Whiteford, J.R. (2008) 'The CMV early enhancer/chicken β actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors', *BMC Cell Biology*, 9(1), p. 2. Available at: https://doi.org/10.1186/1471-2121-9-2.

Antonin, W. and Neumann, H. (2016) 'Chromosome condensation and decondensation during mitosis', *Current Opinion in Cell Biology*, 40, pp. 15–22. Available at: https://doi.org/10.1016/j.ceb.2016.01.013.

Apostolova, P. *et al.* (2023) 'Phase II trial of hypomethylating agent combined with nivolumab for acute myeloid leukaemia relapse after allogeneic haematopoietic cell transplantation—Immune signature correlates with response', *British Journal of Haematology*, 203(2), pp. 264–281. Available at: https://doi.org/10.1111/bjh.19007.

Aran, D. *et al.* (2011) 'Replication timing-related and gene body-specific methylation of active human genes', *Human Molecular Genetics*, 20(4), pp. 670–680. Available at: https://doi.org/10.1093/hmg/ddq513.

Arita, K. *et al.* (2008) 'Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism', *Nature*, 455(7214), pp. 818–821. Available at: https://doi.org/10.1038/nature07249.

Arya, A.K. *et al.* (2017) 'Promoter hypermethylation inactivates CDKN2A, CDKN2B and RASSF1A genes in sporadic parathyroid adenomas', *Scientific Reports*, 7, p. 3123. Available at: https://doi.org/10.1038/s41598-017-03143-8.

Ashraf, W. *et al.* (2017) 'The epigenetic integrator UHRF1: on the road to become a universal biomarker for cancer', *Oncotarget*, 8(31), pp. 51946–51962. Available at: https://doi.org/10.18632/oncotarget.17393.

Auclair, G. *et al.* (2014) 'Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse', *Genome Biology*, 15(12), p. 545. Available at: https://doi.org/10.1186/s13059-014-0545-5.

Avvakumov, G.V. *et al.* (2008) 'Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1', *Nature*, 455(7214), pp. 822–825. Available at: https://doi.org/10.1038/nature07273.

Bachman, K.E., Rountree, M.R. and Baylin, S.B. (2001) 'Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin *', *Journal of Biological Chemistry*, 276(34), pp. 32282–32287. Available at: https://doi.org/10.1074/jbc.M104661200.

Banerjee, S.M. *et al.* (2023) 'Combination of verteporfin-photodynamic therapy with 5-aza-2'-deoxycytidine enhances the anti-tumour immune response in triple negative breast cancer', *Frontiers in Immunology*, 14. Available at: https://doi.org/10.3389/fimmu.2023.1188087.

Barrasa, J.I. *et al.* (2023) 'DNA elements tether canonical Polycomb Repressive Complex 1 to human genes', *Nucleic Acids Research*, 51(21), pp. 11613–11633. Available at: https://doi.org/10.1093/nar/gkad889.

Bashtrykov, P. *et al.* (2014) 'The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism', *The Journal of Biological Chemistry*, 289(7), pp. 4106–4115. Available at: https://doi.org/10.1074/jbc.M113.528893.

Baubec, T. *et al.* (2015) 'Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation', *Nature*, 520(7546), pp. 243–247. Available at: https://doi.org/10.1038/nature14176.

Bauser, C.A., Elick, T.A. and Fraser, M.J. (1999) 'Proteins from nuclear extracts of two lepidopteran cell lines recognize the ends of TTAA-specific transposons piggyBac and tagalong', *Insect Molecular Biology*, 8(2), pp. 223–230. Available at: https://doi.org/10.1046/j.1365-2583.1999.820223.x.

Baylin, S.B. and Herman, J.G. (2000) 'DNA hypermethylation in tumorigenesis: epigenetics joins genetics', *Trends in Genetics*, 16(4), pp. 168–174. Available at: https://doi.org/10.1016/S0168-9525(99)01971-X.

Bednarik, D.P. *et al.* (1991) 'DNA CpG methylation inhibits binding of NF-kappa B proteins to the HIV-1 long terminal repeat cognate DNA motifs', *The New biologist*, 3(10), pp. 969–976.

Belinsky, S.A. *et al.* (1998) 'Aberrant methylation of p16INK4a is an early event in lung cancer and a potential biomarker for early diagnosis', *Proceedings of the National Academy of Sciences of the United States of America*, 95(20), pp. 11891–11896.

Bender, C.M., Pao, M.M. and Jones, P.A. (1998) 'Inhibition of DNA methylation by 5-aza-2'-deoxycytidine suppresses the growth of human tumor cell lines', *Cancer Research*, 58(1), pp. 95–101.

Berkyurek, A.C. *et al.* (2014) 'The DNA Methyltransferase Dnmt1 Directly Interacts with the SET and RING Finger-associated (SRA) Domain of the Multifunctional Protein Uhrf1 to Facilitate Accession of the Catalytic Center to Hemi-methylated DNA *', *Journal of Biological Chemistry*, 289(1), pp. 379–386. Available at: https://doi.org/10.1074/jbc.M113.523209.

Berman, B.P. *et al.* (2011) 'Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains', *Nature Genetics*, 44(1), pp. 40–46. Available at: https://doi.org/10.1038/ng.969.

Berman, B.P. *et al.* (2012) 'Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains', *Nature Genetics*, 44(1), pp. 40–46. Available at: https://doi.org/10.1038/ng.969.

Besselink, N. *et al.* (2023) 'The genome-wide mutational consequences of DNA hypomethylation', *Scientific Reports*, 13(1), p. 6874. Available at: https://doi.org/10.1038/s41598-023-33932-3.

Bestor, T. *et al.* (1988) 'Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases', *Journal of Molecular Biology*, 203(4), pp. 971–983. Available at: https://doi.org/10.1016/0022-2836(88)90122-2.

Bestor, T.H. and Ingram, V.M. (1983) 'Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA.', *Proceedings of the National Academy of Sciences*, 80(18), pp. 5559–5563. Available at: https://doi.org/10.1073/pnas.80.18.5559.

Bird, A. *et al.* (1985) 'A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA', *Cell*, 40(1), pp. 91–99. Available at: https://doi.org/10.1016/0092-8674(85)90312-5.

Bird, A. (2007) 'Perceptions of epigenetics', *Nature*, 447(7143), pp. 396–398. Available at: https://doi.org/10.1038/nature05913.

Bird, A.P. (1980) 'DNA methylation and the frequency of CpG in animal DNA', *Nucleic Acids Research*, 8(7), pp. 1499–1504. Available at: https://doi.org/10.1093/nar/8.7.1499.

Bird, A.P. (1986) 'CpG-rich islands and the function of DNA methylation', *Nature*, 321(6067), pp. 209–213. Available at: https://doi.org/10.1038/321209a0.

Bird, A.P. *et al.* (1987) 'Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene.', *The EMBO Journal*, 6(4), pp. 999–1004. Available at: https://doi.org/10.1002/j.1460-2075.1987.tb04851.x.

Blackledge, N.P. *et al.* (2014) 'Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation', *Cell*, 157(6), pp. 1445–1459. Available at: https://doi.org/10.1016/j.cell.2014.05.004.

Bock, C. *et al.* (2022) 'High-content CRISPR screening', *Nature Reviews Methods Primers*, 2(1), pp. 1–23. Available at: https://doi.org/10.1038/s43586-021-00093-4.

Bogdanović, O. and Veenstra, G.J.C. (2009) 'DNA methylation and methyl-CpG binding proteins: developmental requirements and function', *Chromosoma*, 118(5), pp. 549–565. Available at: https://doi.org/10.1007/s00412-009-0221-9.

Bond, D.R. *et al.* (2024) 'Upregulated cholesterol biosynthesis facilitates the survival of methylation-retaining AML cells following decitabine treatment'. bioRxiv, p. 2024.01.30.577864. Available at: https://doi.org/10.1101/2024.01.30.577864.

Borgel, J. *et al.* (2010) 'Targets and dynamics of promoter DNA methylation during early mouse development', *Nature Genetics*, 42(12), pp. 1093–1100. Available at: https://doi.org/10.1038/ng.708.

Bosco, A. *et al.* (2010) 'Decreased activation of inflammatory networks during acute asthma exacerbations is associated with chronic airflow obstruction', *Mucosal Immunology*, 3(4), pp. 399–409. Available at: https://doi.org/10.1038/mi.2010.13.

Bostick, M. *et al.* (2007a) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science (New York, N.Y.)*, 317(5845), pp. 1760–1764. Available at: https://doi.org/10.1126/science.1147939.

Bostick, M. *et al.* (2007b) 'UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells', *Science*, 317(5845), pp. 1760–1764. Available at: https://doi.org/10.1126/science.1147939.

Boumahdi, S. and de Sauvage, F.J. (2020) 'The great escape: tumour cell plasticity in resistance to targeted therapy', *Nature Reviews Drug Discovery*, 19(1), pp. 39–56. Available at: https://doi.org/10.1038/s41573-019-0044-1.

Bourc'his, D. and Bestor, T.H. (2004) 'Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L', *Nature*, 431(7004), pp. 96–99. Available at: https://doi.org/10.1038/nature02886.

Boyko, K. *et al.* (2022) 'Structure of the DNMT3B ADD domain suggests the absence of a DNMT3A-like autoinhibitory mechanism', *Biochemical and Biophysical Research Communications*, 619, pp. 124–129. Available at: https://doi.org/10.1016/j.bbrc.2022.06.036.

Brinkman, A.B. *et al.* (2019) 'Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation', *Nature Communications*, 10(1), p. 1749. Available at: https://doi.org/10.1038/s41467-019-09828-0.

Browne, E.P. *et al.* (2001) 'Altered cellular mRNA levels in human cytomegalovirus-infected fibroblasts: viral block to the accumulation of antiviral mRNAs', *Journal of Virology*, 75(24), pp. 12319–12330. Available at: https://doi.org/10.1128/JVI.75.24.12319-12330.2001.

Butz, S. *et al.* (2022) 'DNA sequence and chromatin modifiers cooperate to confer epigenetic bistability at imprinting control regions', *Nature Genetics*, 54(11), pp. 1702–1710. Available at: https://doi.org/10.1038/s41588-022-01210-z.

Cain, J.A., Montibus, B. and Oakey, R.J. (2022) 'Intragenic CpG Islands and Their Impact on Gene Regulation', *Frontiers in Cell and Developmental Biology*, 10. Available at: https://doi.org/10.3389/fcell.2022.832348.

Campanero, M.R., Armstrong, M.I. and Flemington, E.K. (2000) 'CpG methylation as a mechanism for the regulation of E2F activity', *Proceedings of the National Academy of Sciences*, 97(12), pp. 6481–6486. Available at: https://doi.org/10.1073/pnas.100340697.

Carninci, P. *et al.* (2006) 'Genome-wide analysis of mammalian promoter architecture and evolution', *Nature Genetics*, 38(6), pp. 626–635. Available at: https://doi.org/10.1038/ng1789.

Chang, L. *et al.* (2021) 'Discovery of small molecules targeting the tandem tudor domain of the epigenetic factor UHRF1 using fragment-based ligand discovery', *Scientific Reports*, 11(1), p. 1121. Available at: https://doi.org/10.1038/s41598-020-80588-4.

Charlton, J. *et al.* (2018) 'Global delay in nascent strand DNA methylation', *Nature Structural & Molecular Biology*, 25(4), pp. 327–332. Available at: https://doi.org/10.1038/s41594-018-0046-4.

Chen, L. *et al*. (2017) 'Inflammatory responses and inflammation-associated diseases in organs', *Oncotarget*, 5(3), pp.578-598. Available at: doi: 10.18632/oncotarget.1782.

Chen, R.P., Gaynor, A.S. and Chen, W. (2019) 'Synthetic biology approaches for targeted protein degradation', *Biotechnology Advances*, 37(8), p. 107446. Available at: https://doi.org/10.1016/j.biotechadv.2019.107446.

Chen, S. *et al.* (2022) 'Epigenetic priming enhances antitumor immunity in platinum-resistant ovarian cancer', *The Journal of Clinical Investigation*, 132(14). Available at: https://doi.org/10.1172/JCI158800.

Chen, T. *et al.* (2003) 'Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b', *Molecular and Cellular Biology*, 23(16), pp. 5594–5605. Available at: https://doi.org/10.1128/MCB.23.16.5594-5605.2003.

Chen, T. *et al.* (2007) 'Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells', *Nature Genetics*, 39(3), pp. 391–396. Available at: https://doi.org/10.1038/ng1982.

Chen, T., Tsujimoto, N. and Li, E. (2004) 'The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin', *Molecular and Cellular Biology*, 24(20), pp. 9048–9058. Available at: https://doi.org/10.1128/MCB.24.20.9048-9058.2004.

Chen, X. *et al.* (2024) 'Structural basis for the H2AK119ub1-specific DNMT3A-nucleosome interaction', *Nature Communications*, 15(1), p. 6217. Available at: https://doi.org/10.1038/s41467-024-50526-3.

Cheng, X. and Blumenthal, R.M. (1996) 'Finding a basis for flipping bases', *Structure*, 4(6), pp. 639–645. Available at: https://doi.org/10.1016/S0969-2126(96)00068-8.

Chiappinelli, K.B. *et al.* (2015) 'Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses', *Cell*, 162(5), pp. 974–986. Available at: https://doi.org/10.1016/j.cell.2015.07.011.

Chiappinelli, K.B. and Baylin, S.B. (2022) 'Inhibiting DNA methylation improves antitumor immunity in ovarian cancer', *The Journal of Clinical Investigation*, 132(14). Available at: https://doi.org/10.1172/JCI160186.

Chomiak, A.A. *et al.* (2024) 'Select EZH2 inhibitors enhance viral mimicry effects of DNMT inhibition through a mechanism involving NFAT:AP-1 signaling', *Science Advances*, 10(13), p. eadk4423. Available at: https://doi.org/10.1126/sciadv.adk4423.

Chow, A. *et al.* (2022) 'Clinical implications of T cell exhaustion for cancer immunotherapy', *Nature Reviews Clinical Oncology*, 19(12), pp. 775–790. Available at: https://doi.org/10.1038/s41571-022-00689-z.

Christman, J.K. (2002) '5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy', *Oncogene*, 21(35), pp. 5483–5495. Available at: https://doi.org/10.1038/sj.onc.1205699.

Chuang, L.S.-H. *et al.* (1997) 'Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1', *Science*, 277(5334), pp. 1996–2000. Available at: https://doi.org/10.1126/science.277.5334.1996.

Čihák, A., Veselý, J. and Šorm, F. (1965) 'Incorporation of 5-azacytidine into liver ribonucleic acids of leukemic mice sensitive and resistant to 5-azacytidine', *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 108(3), pp. 516–518. Available at: https://doi.org/10.1016/0005-2787(65)90046-8.

Connor, M.P. *et al.* (2023) 'Hypomethylating Agents Are Associated with High Rates of Hematologic Toxicity in Patients with Secondary MDS/AML That Develops after Acquired Aplastic Anemia', *Blood*, 142(Supplement 1), p. 2720. Available at: https://doi.org/10.1182/blood-2023-179383.

Cooper, D.N. (1983) 'Eukaryotic DNA methylation', *Human Genetics*, 64(4), pp. 315–333. Available at: https://doi.org/10.1007/BF00292363.

Cordaux, R. and Batzer, M.A. (2009) 'The impact of retrotransposons on human genome evolution', *Nature reviews. Genetics*, 10(10), pp. 691–703. Available at: https://doi.org/10.1038/nrg2640.

Costello, J.F. *et al.* (2000) 'Aberrant CpG-island methylation has non-random and tumour-type–specific patterns', *Nature Genetics*, 24(2), pp. 132–138. Available at: https://doi.org/10.1038/72785.

Coulondre, C. *et al.* (1978) 'Molecular basis of base substitution hotspots in Escherichia coli', *Nature*, 274(5673), pp. 775–780. Available at: https://doi.org/10.1038/274775a0.

Crichton, J.H. *et al.* (2014) 'Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline', *Cellular and Molecular Life Sciences*, 71(9), pp. 1581–1605. Available at: https://doi.org/10.1007/s00018-013-1468-0.

Cross, S.H. *et al.* (1994) 'Purification of CpG islands using a methylated DNA binding column', *Nature Genetics*, 6(3), pp. 236–244. Available at: https://doi.org/10.1038/ng0394-236.

Dakal, T.C. *et al.* (2024) 'Oncogenes and tumor suppressor genes: functions and roles in cancers', *MedComm*, 5(6), p. e582. Available at: https://doi.org/10.1002/mco2.582.

Dantas Machado, A.C. *et al.* (2015) 'Evolving insights on how cytosine methylation affects protein–DNA binding', *Briefings in Functional Genomics*, 14(1), pp. 61–73. Available at: https://doi.org/10.1093/bfgp/elu040.

Daver, N. *et al.* (2018) 'Hypomethylating agents in combination with immune checkpoint inhibitors in acute myeloid leukemia and myelodysplastic syndromes', *Leukemia*, 32(5), pp. 1094–1105. Available at: https://doi.org/10.1038/s41375-018-0070-8.

De Smet, C. *et al.* (1999) 'DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter', *Molecular and Cellular Biology*, 19(11), pp. 7327–7335. Available at: https://doi.org/10.1128/MCB.19.11.7327.

Decato, B.E. *et al.* (2020) 'Characterization of universal features of partially methylated domains across tissues and species', *Epigenetics & Chromatin*, 13(1), p. 39. Available at: https://doi.org/10.1186/s13072-020-00363-7.

Dhayalan, A. *et al.* (2010) 'The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation', *The Journal of Biological Chemistry*, 285(34), pp. 26114–26120. Available at: https://doi.org/10.1074/jbc.M109.089433.

Di Tommaso, P. *et al.* (2017) 'Nextflow enables reproducible computational workflows', *Nature Biotechnology*, 35(4), pp. 316–319. Available at: https://doi.org/10.1038/nbt.3820.

Drexler, H.G. (1998) 'Review of alterations of the cyclin-dependent kinase inhibitor INK4 family genes p15, p16, p18 and p19 in human leukemia-lymphoma cells', *Leukemia*, 12(6), pp. 845–859. Available at: https://doi.org/10.1038/sj.leu.2401043.

Durnaoglu, S., Lee, S.-K. and Ahnn, J. (2021) 'Human Endogenous Retroviruses as Gene Expression Regulators: Insights from Animal Models into Human Diseases', *Molecules and Cells*, 44(12), pp. 861–878. Available at: https://doi.org/10.14348/molcells.2021.5016.

Easwaran, H. *et al.* (2012) 'A DNA hypermethylation module for the stem/progenitor cell signature of cancer', *Genome Research*, 22(5), pp. 837–849. Available at: https://doi.org/10.1101/gr.131169.111.

Easwaran, H.P. *et al.* (2004) 'Replication-independent chromatin loading of Dnmt1 during G2 and M phases', *EMBO Reports*, 5(12), pp. 1181–1186. Available at: https://doi.org/10.1038/sj.embor.7400295.

Egger, G. *et al.* (2006) 'Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival', *Proceedings of the National Academy of Sciences of the United States of America*, 103(38), pp. 14080–14085. Available at: https://doi.org/10.1073/pnas.0604602103.

Ehrlich, M. *et al.* (1982) 'Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells.', *Nucleic Acids Research*, 10(8), pp. 2709–2721.

Ehrlich, M. (2002) 'DNA methylation in cancer: too much, but also too little', *Oncogene*, 21(35), pp. 5400–5413. Available at: https://doi.org/10.1038/sj.onc.1205651.

Ehrlich, M., Jackson, K. and Weemaes, C. (2006) 'Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF)', *Orphanet Journal of Rare Diseases*, 1, p. 2. Available at: https://doi.org/10.1186/1750-1172-1-2.

Elliott, E.N., Sheaffer, K.L. and Kaestner, K.H. (2016) *The 'de novo' DNA methyltransferase Dnmt3b compensates the Dnmt1-deficient intestinal epithelium*, *eLife*. eLife Sciences Publications Limited. Available at: https://doi.org/10.7554/eLife.12975.

Endicott, J.L. *et al.* (2022) 'Cell division drives DNA methylation loss in late-replicating domains in primary human cells', *Nature Communications*, 13(1), p. 6659. Available at: https://doi.org/10.1038/s41467-022-34268-8.

Enver, T. *et al.* (2005) 'Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells', *Human Molecular Genetics*, 14(21), pp. 3129–3140. Available at: https://doi.org/10.1093/hmg/ddi345.

Esteller, M. *et al.* (2000) 'Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors', *JNCI: Journal of the National Cancer Institute*, 92(7), pp. 564–569. Available at: https://doi.org/10.1093/jnci/92.7.564.

Esteller, M. (2006) 'Epigenetics provides a new generation of oncogenes and tumour-suppressor genes', *British Journal of Cancer*, 94(2), pp. 179–183. Available at: https://doi.org/10.1038/sj.bjc.6602918.

Ewels, P.A. *et al.* (2020) 'The nf-core framework for community-curated bioinformatics pipelines', *Nature Biotechnology*, 38(3), pp. 276–278. Available at: https://doi.org/10.1038/s41587-020-0439-x.

Fandy, T.E. *et al.* (2009) 'Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in patients with myeloid malignancies', *Blood*, 114(13), pp. 2764–2773. Available at: https://doi.org/10.1182/blood-2009-02-203547.

Farthing, C.R. *et al.* (2008) 'Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes', *PLoS Genetics*, 4(6), p. e1000116. Available at: https://doi.org/10.1371/journal.pgen.1000116.

Fatemi, M. *et al.* (2001) 'The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA', *Journal of Molecular Biology*, 309(5), pp. 1189–1199. Available at: https://doi.org/10.1006/jmbi.2001.4709.

Feinberg, A.P. *et al.* (1988) 'Reduced genomic 5-methylcytosine content in human colonic neoplasia', *Cancer Research*, 48(5), pp. 1159–1161.

Feinberg, A.P. and Tycko, B. (2004) 'The history of cancer epigenetics', *Nature Reviews Cancer*, 4(2), pp. 143–153. Available at: https://doi.org/10.1038/nrc1279.

Feinberg, A.P. and Vogelstein, B. (1983) 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, 301(5895), pp. 89–92. Available at: https://doi.org/10.1038/301089a0.

Feng, Y.-Q. *et al.* (1999) 'Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange1', *Journal of Molecular Biology*, 292(4), pp. 779–785. Available at: https://doi.org/10.1006/jmbi.1999.3113.

Feng, Y.-Q. *et al.* (2001) 'Position Effects Are Influenced by the Orientation of a Transgene with Respect to Flanking Chromatin', *Molecular and Cellular Biology*, 21(1), pp. 298–309. Available at: https://doi.org/10.1128/MCB.21.1.298-309.2001.

Ferry, L. *et al.* (2017) 'Methylation of DNA Ligase 1 by G9a/GLP Recruits UHRF1 to Replicating DNA and Regulates DNA Methylation', *Molecular Cell*, 67(4), pp. 550-565.e5. Available at: https://doi.org/10.1016/j.molcel.2017.07.012.

Flavahan, W.A., Gaskell, E. and Bernstein, B.E. (2017) 'Epigenetic plasticity and the hallmarks of cancer', *Science (New York, N.Y.)*, 357(6348), p. eaal2380. Available at: https://doi.org/10.1126/science.aal2380.

Florl, A.R. *et al.* (1999) 'DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas', *British Journal of Cancer*, 80(9), pp. 1312–1321. Available at: https://doi.org/10.1038/sj.bjc.6690524.

Franke, M. (2017) *The Role of Higher-Order Chromatin Organization at the SOX9 Locus in Gene Regulation and Disease*.

Franken, N.A.P. *et al.* (2006) 'Clonogenic assay of cells in vitro', *Nature Protocols*, 1(5), pp. 2315–2319. Available at: https://doi.org/10.1038/nprot.2006.339.

Fraser, M.J. *et al.* (1996) 'Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera',

Insect Molecular Biology, 5(2), pp. 141–151. Available at: https://doi.org/10.1111/j.1365-2583.1996.tb00048.x.

Fratantoni, J.C. *et al.* (2003) 'A non-viral gene delivery system designed for clinical use', *Cytotherapy,* 5(3), pp.208-10. Available at: doi: 10.1080/14653240310001479.

Frauer, C. *et al.* (2011) 'Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1', *PLoS ONE*, 6(2), p. e16627. Available at: https://doi.org/10.1371/journal.pone.0016627.

Fuks, F. *et al.* (2003) 'The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase', *Nucleic Acids Research*, 31(9), pp. 2305–2312. Available at: https://doi.org/10.1093/nar/gkg332.

Futscher, B.W. (2011) '5-aza-2'Deoxycytidine', in M. Schwab (ed.) *Encyclopedia of Cancer*. Berlin, Heidelberg: Springer, pp. 325–328. Available at: https://doi.org/10.1007/978-3-642-16483-5_503.

Gaidatzis, D. *et al.* (2014) 'DNA Sequence Explains Seemingly Disordered Methylation Levels in Partially Methylated Domains of Mammalian Genomes', *PLOS Genetics*, 10(2), p. e1004143. Available at: https://doi.org/10.1371/journal.pgen.1004143.

Galvan, D.L. *et al.* (2009) 'Genome-Wide Mapping of PiggyBac Transposon Integrations in Primary Human T Cells', *Journal of immunotherapy (Hagerstown, Md. : 1997)*, 32(8), pp. 837–844. Available at: https://doi.org/10.1097/CJI.0b013e3181b2914c.

Gama-Sosa, M.A. *et al.* (1983) 'The 5-methylcytosine content of DNA from human tumors.', *Nucleic Acids Research*, 11(19), pp. 6883–6894.

Gaudet, F. *et al.* (2003) 'Induction of Tumors in Mice by Genomic Hypomethylation', *Science*, 300(5618), pp. 489–492. Available at: https://doi.org/10.1126/science.1083558.

Geissler, F. *et al.* (2024) 'The role of aberrant DNA methylation in cancer initiation and clinical impacts', *Therapeutic Advances in Medical Oncology*, 16, p. 17588359231220512. Available at: https://doi.org/10.1177/17588359231220511.

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

Glodzik, D. *et al.* (2020) 'Comprehensive molecular comparison of BRCA1 hypermethylated and BRCA1 mutated triple negative breast cancers', *Nature Communications*, 11(1), p. 3747. Available at: https://doi.org/10.1038/s41467-020-17537-2.

Goelz, S.E. *et al.* (1985) 'Hypomethylation of DNA from Benign and Malignant Human Colon Neoplasms', *Science*, 228(4696), pp. 187–190. Available at: https://doi.org/10.1126/science.2579435.

Gongora, C. *et al.* (2008) 'Altered expression of cell proliferation-related and interferon-stimulated genes in colon cancer cells resistant to SN38', *Cancer Biology & Therapy*, 7(6), pp. 822–832. Available at: https://doi.org/10.4161/cbt.7.6.5838.

Goodyear, O. *et al.* (2010) 'Induction of a CD8+ T-cell response to the MAGE cancer testis antigen by combined treatment with azacitidine and sodium valproate in patients with acute myeloid leukemia and myelodysplasia', *Blood*, 116(11), pp. 1908–1918. Available at: https://doi.org/10.1182/blood-2009-11-249474.

Goyal, A. *et al.* (2023) 'DNMT and HDAC inhibition induces immunogenic neoantigens from human endogenous retroviral element-derived transcripts', *Nature Communications*, 14(1), p. 6731. Available at: https://doi.org/10.1038/s41467-023-42417-w.

Greenacre, M. *et al.* (2022) 'Principal component analysis', *Nature Reviews Methods Primers*, 2(1), pp. 1–21. Available at: https://doi.org/10.1038/s43586-022-00184-w.

Greenberg, M.V.C. and Bourc'his, D. (2019) 'The diverse roles of DNA methylation in mammalian development and disease', *Nature Reviews Molecular Cell Biology*, 20(10), pp. 590–607. Available at: https://doi.org/10.1038/s41580-019-0159-6.

Greger, V. *et al.* (1989) 'Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma', *Human Genetics*, 83(2), pp. 155–158. Available at: https://doi.org/10.1007/BF00286709.

Greten, F.R. and Grivennikov S.I. (2019) 'Inflammation and Cancer: Triggers, Mechanisms, and Consequences. Immunity', 51(1), pp. 27-41. Available at: doi: 10.1016/j.immuni.2019.06.025.

Gu, T. *et al.* (2022) 'The disordered N-terminal domain of DNMT3A recognizes H2AK119ub and is required for postnatal development', *Nature Genetics*, 54(5), pp. 625–636. Available at: https://doi.org/10.1038/s41588-022-01063-6.

Güre, A.O. *et al.* (2002) 'The SSX gene family: characterization of 9 complete genes', *International Journal of Cancer*, 101(5), pp. 448–453. Available at: https://doi.org/10.1002/ijc.10634.

Hackett, J.A. *et al.* (2012) 'Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline', *Development*, 139(19), pp. 3623–3632. Available at: https://doi.org/10.1242/dev.081661.

Haggerty, C. *et al.* (2021) 'Dnmt1 has de novo activity targeted to transposable elements', *Nature Structural & Molecular Biology*, 28(7), pp. 594–603. Available at: https://doi.org/10.1038/s41594-021-00603-8.

Hanahan, D. and Weinberg, R.A. (2000) 'The Hallmarks of Cancer', *Cell*, 100(1), pp. 57–70. Available at: https://doi.org/10.1016/S0092-8674(00)81683-9.

Hanahan, D. and Weinberg, R.A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646–674. Available at: https://doi.org/10.1016/j.cell.2011.02.013.

Hansen, K.D. *et al.* (2011) 'Increased methylation variation in epigenetic domains across cancer types', *Nature Genetics*, 43(8), pp. 768–775. Available at: https://doi.org/10.1038/ng.865.

Harrison, J.S. *et al.* (2016) 'Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1', *eLife*. Edited by J.L. Workman, 5, p. e17101. Available at: https://doi.org/10.7554/eLife.17101.

Harrison, P.W. *et al.* (2024) 'Ensembl 2024', *Nucleic Acids Research*, 52(D1), pp. D891–D899. Available at: https://doi.org/10.1093/nar/gkad1049.

He, Y.-F. *et al.* (2011) 'Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA', *Science*, 333(6047), pp. 1303–1307. Available at: https://doi.org/10.1126/science.1210944.

Hemberger, M., Dean, W. and Reik, W. (2009) 'Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal', *Nature Reviews Molecular Cell Biology*, 10(8), pp. 526–537. Available at: https://doi.org/10.1038/nrm2727.

Herman, J.G. *et al.* (1998) 'Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma', *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), pp. 6870–6875. Available at: https://doi.org/10.1073/pnas.95.12.6870.

Hermann, A., Goyal, R. and Jeltsch, A. (2004) 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites *', *Journal of Biological Chemistry*, 279(46), pp. 48350–48359. Available at: https://doi.org/10.1074/jbc.M403427200.

Hervouet, E. *et al.* (2018) 'Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma', *Clinical Epigenetics*, 10(1), p. 17. Available at: https://doi.org/10.1186/s13148-018-0450-y.

Heyn, H. *et al.* (2012) 'Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient', *Epigenetics*, 7(6), pp. 542–550. Available at: https://doi.org/10.4161/epi.20523.

Higham, J. *et al.* (2022) 'Local CpG density affects the trajectory and variance of age-associated DNA methylation changes', *Genome Biology*, 23(1), p. 216. Available at: https://doi.org/10.1186/s13059-022-02787-8.

Hitoshi, N., Ken-ichi, Y. and Jun-ichi, M. (1991) 'Efficient selection for high-expression transfectants with a novel eukaryotic vector', *Gene*, 108(2), pp. 193–199. Available at: https://doi.org/10.1016/0378-1119(91)90434-D.

Holliday, R. and Pugh, J.E. (1975) 'DNA modification mechanisms and gene activity during development', *Science (New York, N.Y.)*, 187(4173), pp. 226–232.

Hon, G.C. *et al.* (2012) 'Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer', *Genome Research*, 22(2), pp. 246–258. Available at: https://doi.org/10.1101/gr.125872.111.

Hornung, V. *et al.* (2014) 'OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids', *Nature Reviews. Immunology*, 14(8), pp. 521–528. Available at: https://doi.org/10.1038/nri3719.

Hossain, Md.S. *et al.* (2022) 'Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies', *Cancers*, 14(7), p. 1732. Available at: https://doi.org/10.3390/cancers14071732.

Hu, C.-L. *et al.* (2022) 'Targeting UHRF1-SAP30-MXD4 axis for leukemia initiating cell eradication in myeloid leukemia', *Cell Research*, 32(12), pp. 1105–1123. Available at: https://doi.org/10.1038/s41422-022-00735-6.

Hulsen, T., de Vlieg, J. and Alkema, W. (2008) 'BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams', *BMC Genomics*, 9(1), p. 488. Available at: https://doi.org/10.1186/1471-2164-9-488.

Illingworth, R.S. *et al.* (2010) 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS genetics*, 6(9), p. e1001134. Available at: https://doi.org/10.1371/journal.pgen.1001134.

Irwin, R. *et al.* (2023) 'The UHRF1 protein is a key regulator of retrotransposable elements and innate immune response to viral RNA in human cells', *Epigenetics*, 18(1), p. 2216005. Available at: https://doi.org/10.1080/15592294.2023.2216005.

Ishiyama, S. *et al.* (2017) 'Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance', *Molecular Cell*, 68(2), pp. 350-360.e7. Available at: https://doi.org/10.1016/j.molcel.2017.09.037.

Issa, J.-P.J. and Kantarjian, H.M. (2009) 'Targeting DNA Methylation', *Clinical cancer research : an official journal of the American Association for Cancer Research*, 15(12), pp. 3938–3946. Available at: https://doi.org/10.1158/1078-0432.CCR-08-2783.

Jackson, K. *et al.* (2004) 'DNA hypomethylation is prevalent even in low-grade breast cancers', *Cancer Biology & Therapy*, 3(12), pp. 1225–1231. Available at: https://doi.org/10.4161/cbt.3.12.1222.

Jackson, M. *et al.* (2004) 'Severe Global DNA Hypomethylation Blocks Differentiation and Induces Histone Hyperacetylation in Embryonic Stem Cells', *Molecular and Cellular Biology*, 24(20), pp. 8862–8871. Available at: https://doi.org/10.1128/MCB.24.20.8862-8871.2004.

Jackson, M.S., Mole, S.E. and Ponder, B.A.J. (1992) 'Characterisation of a boundary between satellite III and aiphoid sequences on human chromosome 10', *Nucleic Acids Research*, 20(18), pp. 4781–4787. Available at: https://doi.org/10.1093/nar/20.18.4781.

Jackson-Grusby, L. *et al.* (2001) 'Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation', *Nature Genetics*, 27(1), pp. 31–39. Available at: https://doi.org/10.1038/83730.

Jaenisch, R. and Bird, A. (2003) 'Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals', *Nature Genetics*, 33(3), pp. 245–254. Available at: https://doi.org/10.1038/ng1089.

James, S.R., Link, P.A. and Karpf, A.R. (2006) 'Epigenetic regulation of X-linked cancer/germline antigen genes by DNMT1 and DNMT3b', *Oncogene*, 25(52), pp. 6975–6985. Available at: https://doi.org/10.1038/sj.onc.1209678.

Jang, W. *et al.* (2019) 'CDKN2B downregulation and other genetic characteristics in T-acute lymphoblastic leukemia', *Experimental & Molecular Medicine*, 51(1), pp. 1–15. Available at: https://doi.org/10.1038/s12276-018-0195-x.

Jeltsch, A. (2002a) 'Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases', *ChemBioChem*, 3(4), pp. 274–293. Available at: https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S.

Jeltsch, A. (2002b) 'Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases', *Chembiochem: A European Journal of Chemical Biology*, 3(4), pp. 274–293. Available at: https://doi.org/10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S.

Jeltsch, A. and Jurkowska, R.Z. (2016) 'Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm', *Nucleic Acids Research*, 44(18), pp. 8556–8575. Available at: https://doi.org/10.1093/nar/gkw723.

Jin, W. *et al.* (2010) 'UHRF1 is associated with epigenetic silencing of BRCA1 in sporadic breast cancer', *Breast Cancer Research and Treatment*, 123(2), pp. 359–373. Available at: https://doi.org/10.1007/s10549-009-0652-2.

Jones, P.A. and Baylin, S.B. (2002) 'The fundamental role of epigenetic events in cancer', *Nature Reviews. Genetics*, 3(6), pp. 415–428. Available at: https://doi.org/10.1038/nrg816.

Jones, P.A. and Baylin, S.B. (2007) 'The Epigenomics of Cancer', *Cell*, 128(4), pp. 683–692. Available at: https://doi.org/10.1016/j.cell.2007.01.029.

Jun-ichi, M. *et al.* (1989) 'Expression vector system based on the chicken β-actin promoter directs efficient production of interleukin-5', *Gene*, 79(2), pp. 269–277. Available at: https://doi.org/10.1016/0378-1119(89)90209-6.

Jüttermann, R., Li, E. and Jaenisch, R. (1994) 'Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation.', *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), pp. 11797–11801.

Kafetzopoulos, I. (2022) 'Reiterative de novo methylation maintainsmethylation levels in somatic cells'. Available at: https://doi.org/10.7488/era/2637.

Kagiwada, S. *et al.* (2013) 'Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice', *The EMBO Journal*, 32(3), pp. 340–353. Available at: https://doi.org/10.1038/emboj.2012.331.

Kaluscha, S. *et al.* (2022) 'Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation', *Nature Genetics*, 54(12), pp. 1895–1906. Available at: https://doi.org/10.1038/s41588-022-01241-6.

Kamali Dolatabadi, E. *et al.* (2017) 'CDKN2B Methylation Correlates with Survival in AML Patients', *Iranian Journal of Pharmaceutical Research : IJPR*, 16(4), pp. 1600–1611.

Kang, J. *et al.* (2022) 'Promoter CAG is more efficient than hepatocyte-targeting TBG for transgene expression via rAAV8 in liver tissues', *Molecular Medicine Reports*, 25(1), pp. 1–9. Available at: https://doi.org/10.3892/mmr.2021.12532.

Karpf, A.R. *et al.* (1999) 'Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells', *Proceedings of the National Academy of Sciences*, 96(24), pp. 14007–14012. Available at: https://doi.org/10.1073/pnas.96.24.14007.

Kent, W.J. (2002) 'BLAT—The BLAST-Like Alignment Tool', *Genome Research*, 12(4), pp. 656–664. Available at: https://doi.org/10.1101/gr.229202.

Kim, K.H. and Sederstrom, J.M. (2015) 'Assaying cell cycle status using flow cytometry', *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, 111, p. 28.6.1-28.6.11. Available at: https://doi.org/10.1002/0471142727.mb2806s111.

Kohli, R.M. and Zhang, Y. (2013) 'TET enzymes, TDG and the dynamics of DNA demethylation', *Nature*, 502(7472), pp. 472–479. Available at: https://doi.org/10.1038/nature12750.

Kojima, K.K. (2018) 'Human transposable elements in Repbase: genomic footprints from fish to humans', *Mobile DNA*, 9, p. 2. Available at: https://doi.org/10.1186/s13100-017-0107-y.

Kong, X. *et al.* (2019) 'Defining UHRF1 Domains That Support Maintenance of Human Colon Cancer DNA Methylation and Oncogenic Properties', *Cancer cell*, 35(4), pp. 633-648.e7. Available at: https://doi.org/10.1016/j.ccell.2019.03.003.

Kong, Y. *et al.* (2019) 'Transposable element expression in tumors is associated with immune infiltration and increased antigenicity', *Nature Communications*, 10(1), p. 5228. Available at: https://doi.org/10.1038/s41467-019-13035-2.

Kornberg, R.D. (1974) 'Chromatin Structure: A Repeating Unit of Histones and DNA', *Science*, 184(4139), pp. 868–871. Available at: https://doi.org/10.1126/science.184.4139.868.

Kostyrko, K. *et al.* (2023) 'UHRF1 is a mediator of KRAS driven oncogenesis in lung adenocarcinoma', *Nature Communications*, 14(1), p. 3966. Available at: https://doi.org/10.1038/s41467-023-39591-2.

Krebs, A.R. *et al.* (2014) 'High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions', *eLife*. Edited by A.C. Ferguson-Smith, 3, p. e04094. Available at: https://doi.org/10.7554/eLife.04094.

Kulaeva, O.I. *et al.* (2003) 'Epigenetic silencing of multiple interferon pathway genes after cellular immortalization', *Oncogene*, 22(26), pp. 4118–4127. Available at: https://doi.org/10.1038/sj.onc.1206594.

Kumaki, Y., Oda, M. and Okano, M. (2008) 'QUMA: quantification tool for methylation analysis', *Nucleic Acids Research*, 36(suppl_2), pp. W170–W175. Available at: https://doi.org/10.1093/nar/gkn294.

Lakshmikuttyamma, A. *et al.* (2010) 'Reexpression of epigenetically silenced AML tumor suppressor genes by SUV39H1 inhibition', *Oncogene*, 29(4), pp. 576–588. Available at: https://doi.org/10.1038/onc.2009.361.

Lanciano, S. *et al.* (2023) 'Resolving the heterogeneity of L1 DNA methylation reveals the epigenetic and transcriptional interplay between L1s and their integration sites'. bioRxiv, p. 2023.01.03.522582. Available at: https://doi.org/10.1101/2023.01.03.522582.

Laureau, R. *et al.* (2021) 'Meiotic cells counteract programmed retrotransposon activation via RNA-binding translational repressor assemblies', *Developmental cell*, 56(1), pp. 22-35.e7. Available at: https://doi.org/10.1016/j.devcel.2020.11.008.

Lee, S.-H. *et al.* (2010) 'The Impact of CpG Island on Defining Transcriptional Activation of the Mouse L1 Retrotransposable Elements', *PLoS ONE*, 5(6), p. e11353. Available at: https://doi.org/10.1371/journal.pone.0011353.

Lehnertz, B. *et al.* (2003) 'Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin', *Current Biology*, 13(14), pp. 1192–1200. Available at: https://doi.org/10.1016/S0960-9822(03)00432-9.

Lengauer, C., Kinzler, K.W. and Vogelstein, B. (1997) 'Genetic instability in colorectal cancers', *Nature*, 386(6625), pp. 623–627. Available at: https://doi.org/10.1038/386623a0.

Leonhardt, H. *et al.* (1992) 'A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei', *Cell*, 71(5), pp. 865–873. Available at: https://doi.org/10.1016/0092-8674(92)90561-P.

Ley, T.J. *et al.* (2010) 'DNMT3A Mutations in Acute Myeloid Leukemia', *New England Journal of Medicine*, 363(25), pp. 2424–2433. Available at: https://doi.org/10.1056/NEJMoa1005143.

Li, E., Bestor, T.H. and Jaenisch, R. (1992) 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell*, 69(6), pp. 915–926. Available at: https://doi.org/10.1016/0092-8674(92)90611-F.

Li, E. and Zhang, Y. (2014) 'DNA Methylation in Mammals', *Cold Spring Harbor Perspectives in Biology*, 6(5), p. a019133. Available at: https://doi.org/10.1101/cshperspect.a019133.

Li, H. *et al.* (2014) 'Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers', *Oncotarget*, 5(3), pp. 587–598.

Li, L.-C. and Dahiya, R. (2002) 'MethPrimer: designing primers for methylation PCRs', *Bioinformatics (Oxford, England)*, 18(11), pp. 1427–1431. Available at: https://doi.org/10.1093/bioinformatics/18.11.1427.

Liang, G. *et al.* (2002) 'Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements', *Molecular and Cellular Biology*, 22(2), pp. 480–491. Available at: https://doi.org/10.1128/MCB.22.2.480-491.2002.

Liang, Q. *et al.* (2009) 'Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons', *genesis*, 47(6), pp. 404–408. Available at: https://doi.org/10.1002/dvg.20508.

Liao, J. *et al.* (2015) 'Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells', *Nature genetics*, 47(5), pp. 469–478. Available at: https://doi.org/10.1038/ng.3258.

Liberzon, A. *et al.* (2015) 'The Molecular Signatures Database (MSigDB) hallmark gene set collection', *Cell systems*, 1(6), pp. 417–425. Available at: https://doi.org/10.1016/j.cels.2015.12.004.

Lienert, F. *et al.* (2011) 'Identification of genetic elements that autonomously determine DNA methylation states', *Nature Genetics*, 43(11), pp. 1091–1097. Available at: https://doi.org/10.1038/ng.946.

Linnekamp, J.F. *et al.* (2017) 'Clinical and biological effects of demethylating agents on solid tumours – A systematic review', *Cancer Treatment Reviews*, 54, pp. 10–23. Available at: https://doi.org/10.1016/j.ctrv.2017.01.004.

Linnenkamp, J.F. *et al.* (2021) 'Pre-Operative Decitabine in Colon Cancer Patients: Analyses on WNT Target Methylation and Expression', *Cancers*, 13, 2357. Available at: https://doi.org/10.3390/cancers13102357

Lister, R. *et al.* (2009) 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature*, 462(7271), pp. 315–322. Available at: https://doi.org/10.1038/nature08514.

Liu, W.H. *et al.* (2022) 'Discovery and Mechanism of Small Molecule Inhibitors Selective for the Chromatin-Binding Domains of Oncogenic UHRF1', *Biochemistry*, 61(5), pp. 354–366. Available at: https://doi.org/10.1021/acs.biochem.1c00698.

Liu, X. *et al.* (2013) 'UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9', *Nature Communications*, 4(1), p. 1563. Available at: https://doi.org/10.1038/ncomms2562.

Long, H.K., Blackledge, N.P. and Klose, R.J. (2013) 'ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection', *Biochemical Society Transactions*, 41(3), pp. 727–740. Available at: https://doi.org/10.1042/BST20130028.

Loo Yau, H. *et al.* (2021) 'DNA hypomethylating agents increase activation and cytolytic activity of CD8+ T cells', *Molecular Cell*, 81(7), pp. 1469-1483.e8. Available at: https://doi.org/10.1016/j.molcel.2021.01.038.

Lorincz, M.C. *et al.* (2002) 'DNA Methylation Density Influences the Stability of an Epigenetic Imprint and Dnmt3a/b-Independent De Novo Methylation', *Molecular and Cellular Biology*, 22(21), pp. 7572–7580. Available at: https://doi.org/10.1128/MCB.22.21.7572-7580.2002.

Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. Available at: https://doi.org/10.1186/s13059-014-0550-8.

Luger, K. (1997) 'Crystal structure of the nucleosome core particle at 2.8 A˚ resolution', 389.

Lukhele, S., Boukhaled, G.M. and Brooks, D.G. (2019) 'Type I interferon signaling, regulation and gene stimulation in chronic virus infection', *Seminars in immunology*, 43, p. 101277. Available at: https://doi.org/10.1016/j.smim.2019.05.001.

Lyko, F. (2018) 'The DNA methyltransferase family: a versatile toolkit for epigenetic regulation', *Nature Reviews Genetics*, 19(2), pp. 81–92. Available at: https://doi.org/10.1038/nrg.2017.80.

Macrae, T.A., Fothergill-Robinson, J. and Ramalho-Santos, M. (2023) 'Regulation, functions and transmission of bivalent chromatin during mammalian development', *Nature Reviews Molecular Cell Biology*, 24(1), pp. 6–26. Available at: https://doi.org/10.1038/s41580-022-00518-2.

Maiti, A. and Drohat, A.C. (2011) 'Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine', *The Journal of Biological Chemistry*, 286(41), pp. 35334–35338. Available at: https://doi.org/10.1074/jbc.C111.284620.

Maksakova, I.A., Mager, D.L. and Reiss, D. (2008) 'Endogenous retroviruses', *Cellular and Molecular Life Sciences*, 65(21), pp. 3329–3347. Available at: https://doi.org/10.1007/s00018-008-8494-3.

Mancini, M., Magnani, E., Macchi, F. and Bonapace, I.M. (2021) 'The multi-functionality of UHRF1: epigenome maintenance and preservation of genome integrity', *Nucleic Acids Research*, 49(11), pp. 6053–6068. Available at: https://doi.org/10.1093/nar/gkab293.

Mancini, M., Magnani, E., Macchi, F. and Bonapace, I. (2021) 'The multi-functionality of UHRF1: Epigenome maintenance and preservation of genome integrity', *Nucleic Acids Research*, 49. Available at: https://doi.org/10.1093/nar/gkab293.

Manzo, M. *et al.* (2017) 'Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands', *The EMBO Journal*, 36(23), pp. 3421–3434. Available at: https://doi.org/10.15252/embj.201797038.

Martire, S. and Banaszynski, L.A. (2020) 'The roles of histone variants in fine-tuning chromatin organization and function', *Nature Reviews Molecular Cell Biology*, 21(9), pp. 522–541. Available at: https://doi.org/10.1038/s41580-020-0262-8.

Masalmeh, R.H.A. (2019) 'Low levels of methylation are targeted to aberrantly methylated CGIs in human colorectal cancer cells'. Available at: https://era.ed.ac.uk/handle/1842/35826 (Accessed: 21 June 2024).

Masalmeh, R.H.A. *et al.* (2021) 'De novo DNA methyltransferase activity in colorectal cancer is directed towards H3K36me3 marked CpG islands', *Nature Communications*, 12(1), p. 694. Available at: https://doi.org/10.1038/s41467-020-20716-w.

McGarvey, K.M. *et al.* (2008) 'Defining a Chromatin Pattern That Characterizes DNA Hypermethylated Genes in Colon Cancer Cells', *Cancer research*, 68(14), pp. 5753–5759. Available at: https://doi.org/10.1158/0008-5472.CAN-08-0700.

Mehdipour, P. *et al.* (2020) 'Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency', *Nature*, 588(7836), pp. 169–173. Available at: https://doi.org/10.1038/s41586-020-2844-1.

Meir, Y.-J.J. *et al.* (2011) 'Genome-wide target profiling of piggyBac and Tol2in HEK 293: pros and cons for gene discovery and gene therapy', *BMC Biotechnology*, 11(1), p. 28. Available at: https://doi.org/10.1186/1472-6750-11-28.

Meissner, A. *et al.* (2008) 'Genome-scale DNA methylation maps of pluripotent and differentiated cells', *Nature*, 454(7205), pp. 766–770. Available at: https://doi.org/10.1038/nature07107.

Mikkelsen, T.S. *et al.* (2007) 'Genome-wide maps of chromatin state in pluripotent and lineage-committed cells', *Nature*, 448(7153), pp. 553–560. Available at: https://doi.org/10.1038/nature06008.

Milacic, M. *et al.* (2023) 'The Reactome Pathway Knowledgebase 2024', *Nucleic Acids Research*, 52(D1), pp. D672–D678. Available at: https://doi.org/10.1093/nar/gkad1025.

Ming, X. *et al.* (2020) 'Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration', *Cell Research*, 30(11), pp. 980–996. Available at: https://doi.org/10.1038/s41422-020-0359-9.

Moffat, J.G. *et al.* (2017) 'Opportunities and challenges in phenotypic drug discovery: an industry perspective', *Nature Reviews Drug Discovery*, 16(8), pp. 531–543. Available at: https://doi.org/10.1038/nrd.2017.111.

Mohn, F. *et al.* (2008) 'Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors', *Molecular Cell*, 30(6), pp. 755–766. Available at: https://doi.org/10.1016/j.molcel.2008.05.007.

Moore, L.D., Le, T. and Fan, G. (2013) 'DNA Methylation and Its Basic Function', *Neuropsychopharmacology*, 38(1), pp. 23–38. Available at: https://doi.org/10.1038/npp.2012.112.

Mordstein, C. *et al.* (2021) 'Transcription, mRNA Export, and Immune Evasion Shape the Codon Usage of Viruses', *Genome Biology and Evolution*, 13(9), p. evab106. Available at: https://doi.org/10.1093/gbe/evab106.

Morgensztern, D. *et al.* (2019) 'Multicenter phase II trial of RRx-001 in previously treated SCLC.', *Journal of Clinical Oncology*, 37(15_suppl), pp. e20104–e20104. Available at: https://doi.org/10.1200/JCO.2019.37.15_suppl.e20104.

Murakami, Y. (2013) 'Heterochromatin and Euchromatin', in W. Dubitzky et al. (eds) *Encyclopedia of Systems Biology*. New York, NY: Springer, pp. 881–884. Available at: https://doi.org/10.1007/978-1-4419-9863-7_1413.

Mutskov, V. and Felsenfeld, G. (2004) 'Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9', *The EMBO Journal*, 23(1), pp. 138–149. Available at: https://doi.org/10.1038/sj.emboj.7600013.

Myrianthopoulos, V. *et al.* (2016) 'Tandem virtual screening targeting the SRA domain of UHRF1 identifies a novel chemical tool modulating DNA methylation', *European Journal of Medicinal Chemistry*, 114, pp. 390–396. Available at: https://doi.org/10.1016/j.ejmech.2016.02.043.

Nady, N. *et al.* (2011) 'Recognition of Multivalent Histone States Associated with Heterochromatin by UHRF1 Protein *', *Journal of Biological Chemistry*, 286(27), pp. 24300–24311. Available at: https://doi.org/10.1074/jbc.M111.234104.

Natsume, T. *et al.* (2016) 'Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors', *Cell Reports*, 15(1), pp. 210–218. Available at: https://doi.org/10.1016/j.celrep.2016.03.001.

Neidhart, M. (2016) 'Chapter 1 - DNA Methylation – Introduction', in M. Neidhart (ed.) *DNA Methylation and Complex Human Disease*. Oxford: Academic Press, pp. 1–8. Available at: https://doi.org/10.1016/B978-0-12-420194-1.00001-4.

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

Nishimura, K. *et al.* (2009) 'An auxin-based degron system for the rapid depletion of proteins in nonplant cells', *Nature Methods*, 6(12), pp. 917–922. Available at: https://doi.org/10.1038/nmeth.1401.

Nishiyama, A. *et al.* (2020) 'Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation', *Nature Communications*, 11(1), p. 1222. Available at: https://doi.org/10.1038/s41467-020-15006-4.

Ohtani, H. *et al.* (2020) 'Activation of a Subset of Evolutionarily Young Transposable Elements and Innate Immunity Are Linked to Clinical Responses to 5-Azacytidine', *Cancer Research*, 80(12), pp. 2441–2450. Available at: https://doi.org/10.1158/0008-5472.CAN-19-1696.

Okano, M. *et al.* (1999) 'DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development', *Cell*, 99(3), pp. 247–257. Available at: https://doi.org/10.1016/S0092-8674(00)81656-6.

Okano, M., Xie, S. and Li, E. (1998) 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nature Genetics*, 19(3), pp. 219–220. Available at: https://doi.org/10.1038/890.

Olins, D.E. and Olins, A.L. (1978) 'Nucleosomes: The Structural Quantum in Chromosomes: Virtually all the DNA of eukaryotic cells is organized into a repeating array of nucleohistone particles called

nucleosomes. These chromatin subunits are close-packed into higher-order fibers and are modified during chromosome expression', *American Scientist*, 66(6), pp. 704–711.

Olivier, M. *et al.* (2014) 'Modelling mutational landscapes of human cancers in vitro', *Scientific Reports*, 4(1), p. 4482. Available at: https://doi.org/10.1038/srep04482.

Oronsky, B. *et al.* (2016) 'Rockets, radiosensitizers, and RRx-001: an origin story part I', *Discovery Medicine*, 21(115), pp. 173–180.

Oronsky, B. *et al.* (2019) 'Cardioprotective Effect of Phase 3 Clinical Anticancer Agent, RRx-001, in Doxorubicin-Induced Acute Cardiotoxicity in Mice', *Molecular Pharmaceutics*, 16(7), pp. 2929–2934. Available at: https://doi.org/10.1021/acs.molpharmaceut.9b00150.

Oronsky, B. *et al.* (2021) 'Discovery of RRx-001, a Myc and CD47 Downregulating Small Molecule with Tumor Targeted Cytotoxicity and Healthy Tissue Cytoprotective Properties in Clinical Development', *Journal of Medicinal Chemistry*, 64(11), pp. 7261–7271. Available at: https://doi.org/10.1021/acs.jmedchem.1c00599.

Ørskov, A.D. *et al.* (2015) 'Hypomethylation and up-regulation of PD-1 in T cells by azacytidine in MDS/AML patients: A rationale for combined targeting of PD-1 and DNA methylation', *Oncotarget*, 6(11), pp. 9612–9626.

Ostrowski, P.J. and Tatton-Brown, K. (1993) 'Tatton-Brown-Rahman Syndrome', in M.P. Adam et al. (eds) *GeneReviews®*. Seattle (WA): University of Washington, Seattle. Available at: http://www.ncbi.nlm.nih.gov/books/NBK581652/ (Accessed: 30 August 2024).

Otani, J. *et al.* (2009) 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain', *EMBO Reports*, 10(11), pp. 1235–1241. Available at: https://doi.org/10.1038/embor.2009.218.

Pappalardi, M.B. *et al.* (2021) 'Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia', *Nature cancer*, 2(10), pp. 1002–1017.

Pappalardo, X.G. and Barra, V. (2021) 'Losing DNA methylation at repetitive elements and breaking bad', *Epigenetics & Chromatin*, 14(1), p. 25. Available at: https://doi.org/10.1186/s13072-021-00400-z.

Park, C.W. *et al.* (2005) 'DNA methylation of Sleeping Beauty with transposition into the mouse genome', *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 10(8), pp. 763–776. Available at: https://doi.org/10.1111/j.1365-2443.2005.00875.x.

Parry, A.J. and Reik, W. (2022) 'DNMT3A binds ubiquitinated histones to regulate bivalent genes', *Nature genetics*, 54(5), pp. 537–538. Available at: https://doi.org/10.1038/s41588-022-01073-4.

Payer, B. and Lee, J.T. (2008) 'X Chromosome Dosage Compensation: How Mammals Keep the Balance', *Annual Review of Genetics*, 42(Volume 42, 2008), pp. 733–772. Available at: https://doi.org/10.1146/annurev.genet.42.110807.091711.

Peralta-Arrieta, I. *et al.* (2017) 'DNMT3B modulates the expression of cancer-related genes and downregulates the expression of the gene VAV3 via methylation', *American Journal of Cancer Research*, 7(1), pp. 77–87.

Pfeifer, G.P. (2023) '5-hydroxymethylcytosine stabilizes transcription by preventing aberrant initiation in gene bodies', *Nature Genetics*, 55(1), pp. 2–3. Available at: https://doi.org/10.1038/s41588-022-01253-2.

Píškala, A. and Šorm, F. (1964) 'Nucleic acids components and their analogues. LI. Synthesis of 1-glycosyl derivatives of 5-azauracil and 5-azacytosine', *Collection of Czechoslovak Chemical Communications*, 29(9), pp. 2060–2076. Available at: https://doi.org/10.1135/cccc19642060.

Platanias, L.C. (2005) 'Mechanisms of type-I- and type-II-interferon-mediated signalling', *Nature Reviews Immunology*, 5(5), pp. 375–386. Available at: https://doi.org/10.1038/nri1604.

Qin, W. *et al.* (2015) 'DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination', *Cell Research*, 25(8), pp. 911–929. Available at: https://doi.org/10.1038/cr.2015.72.

Raj, K. *et al.* (2007) 'CDKN2B methylation status and isolated chromosome 7 abnormalities predict responses to treatment with 5-azacytidine', *Leukemia*, 21(9), pp. 1937–1944. Available at: https://doi.org/10.1038/sj.leu.2404796.

Ramirez-Carrozzi, V.R. *et al.* (2009) 'A Unifying Model for the Selective Regulation of Inducible Transcription By CpG Islands and Nucleosome Remodeling', *Cell*, 138(1), pp. 114–128. Available at: https://doi.org/10.1016/j.cell.2009.04.020.

Ramos, J.A. *et al.* (2001) 'Rapid Degradation of Auxin/Indoleacetic Acid Proteins Requires Conserved Amino Acids of Domain II and Is Proteasome Dependent', *The Plant Cell*, 13(10), pp. 2349–2360. Available at: https://doi.org/10.1105/tpc.010244.

Reardon, E.S. *et al.* (2021) 'UHRF1 is a Novel Druggable Epigenetic Target in Malignant Pleural Mesothelioma', *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, 16(1), pp. 89–103. Available at: https://doi.org/10.1016/j.jtho.2020.08.024.

Reddington, J.P. *et al.* (2013) 'Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes', *Genome Biology*, 14(3), p. R25. Available at: https://doi.org/10.1186/gb-2013-14-3-r25.

Reid, T.R. *et al.* (2023) 'ROCKET: Phase II Randomized, Active-controlled, Multicenter Trial to Assess the Safety and Efficacy of RRx-001 + Irinotecan vs. Single-agent Regorafenib in Third/Fourth Line Colorectal Cancer', *Clinical Colorectal Cancer*, 22(1), pp. 92–99. Available at: https://doi.org/10.1016/j.clcc.2022.11.003.

Reik, W., Dean, W. and Walter, J. (2001) 'Epigenetic Reprogramming in Mammalian Development', *Science*, 293(5532), pp. 1089–1093. Available at: https://doi.org/10.1126/science.1063443.

Renbaum, P. *et al.* (1990) 'Cloning, characterization, and expression in Escherichia coli of the gene coding for the CpG DNA methylase from Spiroplasma sp. strain MQ1(M.SssI)', *Nucleic Acids Research*, 18(5), pp. 1145–1152. Available at: https://doi.org/10.1093/nar/18.5.1145.

Reynolds, G.A. *et al.* (1984) 'HMG CoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions', *Cell*, 38(1), pp. 275–285. Available at: https://doi.org/10.1016/0092-8674(84)90549-x.

Rhee, I. *et al.* (2000) 'CpG methylation is maintained in human cancer cells lacking DNMT1', *Nature*, 404(6781), pp. 1003–1007. Available at: https://doi.org/10.1038/35010000.

Rhee, I. *et al.* (2002) 'DNMT1 and DNMT3b cooperate to silence genes in human cancer cells', *Nature*, 416(6880), pp. 552–556. Available at: https://doi.org/10.1038/416552a.

Riggs, A.D. (1975) 'X inactivation, differentiation, and DNA methylation', *Cytogenetics and Cell Genetics*, 14(1), pp. 9–25. Available at: https://doi.org/10.1159/000130315.

Robert, M.-F. *et al.* (2003) 'DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells', *Nature Genetics*, 33(1), pp. 61–65. Available at: https://doi.org/10.1038/ng1068.

Robertson, K.D. *et al.* (1999) 'The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors.', *Nucleic Acids Research*, 27(11), pp. 2291–2298.

Rondelet, G. *et al.* (2016) 'Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B', *Journal of Structural Biology*, 194(3), pp. 357–367. Available at: https://doi.org/10.1016/j.jsb.2016.03.013.

Rose, N.R. and Klose, R.J. (2014) 'Understanding the relationship between DNA methylation and histone lysine methylation', *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms*, 1839(12), pp. 1362–1372. Available at: https://doi.org/10.1016/j.bbagrm.2014.02.007.

Rothbart, S.B. *et al.* (2012) 'Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation', *Nature Structural & Molecular Biology*, 19(11), pp. 1155–1160. Available at: https://doi.org/10.1038/nsmb.2391.

Roulois, D. *et al.* (2015a) 'DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts', *Cell*, 162(5), pp. 961–973. Available at: https://doi.org/10.1016/j.cell.2015.07.056.

Roulois, D. *et al.* (2015b) 'DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts', *Cell*, 162(5), pp. 961–973. Available at: https://doi.org/10.1016/j.cell.2015.07.056.

Rowe, H.M. and Trono, D. (2011) 'Dynamic control of endogenous retroviruses during development', *Virology*, 411(2), pp. 273–287. Available at: https://doi.org/10.1016/j.virol.2010.12.007.

Ruegger, M. *et al.* (1998) 'The TIR1 protein of Arabidopsis functions in auxin response and is related to human SKP2 and yeast Grr1p', *Genes & Development*, 12(2), pp. 198–207. Available at: https://doi.org/10.1101/gad.12.2.198.

Russo, V.E.A., Martienssen, R.A. and Riggs, A.D. (1996) *Epigenetic mechanisms of gene regulation*. Plainview, N.Y: Cold Spring Harbor Laboratory Press (Cold Spring Harbor monograph series, 32).

Saadeh, H. and Schulz, R. (2014) 'Protection of CpG islands against de novo DNA methylation during oogenesis is associated with the recognition site of E2f1 and E2f2', *Epigenetics & Chromatin*, 7(1), p. 26. Available at: https://doi.org/10.1186/1756-8935-7-26.

Sabag, O. *et al.* (2014) 'Establishment of methylation patterns in ES cells', *Nature Structural & Molecular Biology*, 21(1), pp. 110–112. Available at: https://doi.org/10.1038/nsmb.2734.

Sadeq, S. *et al.* (2021) 'Endogenous Double-Stranded RNA', *Non-Coding RNA*, 7(1), p. 15. Available at: https://doi.org/10.3390/ncrna7010015.

Sado, T. *et al.* (2000) 'X Inactivation in the Mouse Embryo Deficient for *Dnmt1:* Distinct Effect of Hypomethylation on Imprinted and Random X Inactivation', *Developmental Biology*, 225(2), pp. 294–303. Available at: https://doi.org/10.1006/dbio.2000.9823.

Salhab, A. *et al.* (2018) 'A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains', *Genome Biology*, 19, p. 150. Available at: https://doi.org/10.1186/s13059-018-1510-5.

Sandoval, J. and Serra, P.L. (2016) 'Chapter 13 - DNA Methylation Biomarkers in Lung Cancer', in J.L. García-Giménez (ed.) *Epigenetic Biomarkers and Diagnostics*. Boston: Academic Press, pp. 259–273. Available at: https://doi.org/10.1016/B978-0-12-801899-6.00013-9.

Santi, D.V., Garrett, C.E. and Barr, P.J. (1983) 'On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs', *Cell*, 33(1), pp. 9–10. Available at: https://doi.org/10.1016/0092-8674(83)90327-6.

Saxonov, S., Berg, P. and Brutlag, D.L. (2006) 'A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters', *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), pp. 1412–1417. Available at: https://doi.org/10.1073/pnas.0510310103.

Schermelleh, L. *et al.* (2005) 'Trapped in action: direct visualization of DNA methyltransferase activity in living cells', *Nature Methods*, 2(10), pp. 751–756. Available at: https://doi.org/10.1038/nmeth794.

Schindelin, J. *et al.* (2012) 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. Available at: https://doi.org/10.1038/nmeth.2019.

Schneider, K. *et al.* (2013) 'Dissection of cell cycle–dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling', *Nucleic Acids Research*, 41(9), pp. 4860–4876. Available at: https://doi.org/10.1093/nar/gkt191.

Schübeler, D. *et al.* (2000) 'Genomic Targeting of Methylated DNA: Influence of Methylation on Transcription, Replication, Chromatin Structure, and Histone Acetylation', *Molecular and Cellular Biology*, 20(24), pp. 9103–9112.

Schübeler, D. (2015) 'Function and information content of DNA methylation', *Nature*, 517(7534), pp. 321–326. Available at: https://doi.org/10.1038/nature14192.

Schulz, W.A., Steinhoff, C. and Florl, A.R. (2006) 'Methylation of Endogenous Human Retroelements in Health and Disease', in W. Doerfler and P. Böhm (eds) *DNA Methylation: Development, Genetic Disease and Cancer*. Berlin, Heidelberg: Springer, pp. 211–250. Available at: https://doi.org/10.1007/3-540-31181-5_11.

Sen, G.L. *et al.* (2010) 'DNMT1 maintains progenitor function in self-renewing somatic tissue', *Nature*, 463(7280), pp. 563–567. Available at: https://doi.org/10.1038/nature08683.

Sharif, J. *et al.* (2007) 'The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA', *Nature*, 450(7171), pp. 908–912. Available at: https://doi.org/10.1038/nature06397.

Sheaffer, K.L., Elliott, E.N. and Kaestner, K.H. (2016) 'DNA hypomethylation contributes to genomic instability and intestinal cancer initiation', *Cancer prevention research (Philadelphia, Pa.)*, 9(7), pp. 534–546. Available at: https://doi.org/10.1158/1940-6207.CAPR-15-0349.

Shen, L. *et al.* (2007) 'Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters', *PLOS Genetics*, 3(10), p. e181. Available at: https://doi.org/10.1371/journal.pgen.0030181.

Shen, L. *et al.* (2010) 'DNA Methylation Predicts Survival and Response to Therapy in Patients With Myelodysplastic Syndromes', *Journal of Clinical Oncology*, 28(4), pp. 605–613. Available at: https://doi.org/10.1200/JCO.2009.23.4781.

Shima, K. *et al.* (2011) 'Prognostic Significance of CDKN2A (p16) Promoter Methylation and Loss of Expression in 902 Colorectal Cancers: Cohort Study and Literature Review', *International journal of cancer. Journal international du cancer*, 128(5), pp. 1080–1094. Available at: https://doi.org/10.1002/ijc.25432.

Sinha, S. *et al.* (2008) 'Frequent deletion and methylation in SH3GL2 and CDKN2A loci are associated with early- and late-onset breast carcinoma', *Annals of Surgical Oncology*, 15(4), pp. 1070–1080. Available at: https://doi.org/10.1245/s10434-007-9790-0.

Smallwood, S.A. and Kelsey, G. (2012) 'De novo DNA methylation: a germ cell perspective', *Trends in Genetics*, 28(1), pp. 33–42. Available at: https://doi.org/10.1016/j.tig.2011.09.004.

Smeets, D.F.C.M. *et al.* (1994) 'ICF syndrome: a new case and review of the literature', *Human Genetics*, 94(3), pp. 240–246. Available at: https://doi.org/10.1007/BF00208277.

Smith, A.M. *et al.* (2021) 'Functional and epigenetic phenotypes of humans and mice with DNMT3A Overgrowth Syndrome', *Nature Communications*, 12(1), p. 4549. Available at: https://doi.org/10.1038/s41467-021-24800-7.

Smith, C.C. *et al.* (2018) 'Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma', *The Journal of Clinical Investigation*, 128(11), pp. 4804–4820. Available at: https://doi.org/10.1172/JCI121476.

Smith, C.L. and Peterson, C.L. (2004) 'ATP-Dependent Chromatin Remodeling', in *Current Topics in Developmental Biology*. Academic Press, pp. 115–148. Available at: https://doi.org/10.1016/S0070-2153(04)65004-6.

Smith, T., Heger, A. and Sudbery, I. (2017) 'UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy', *Genome Research*, 27(3), pp. 491–499. Available at: https://doi.org/10.1101/gr.209601.116.

Soboleski, M.R., Oaks, J. and Halford, W.P. (2005) 'Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells.', *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 19(3), pp. 440–442. Available at: https://doi.org/10.1096/fj.04-3180fje.

Song, J. *et al.* (2011) 'Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation', *Science (New York, N.Y.)*, 331(6020), pp. 1036–1040. Available at: https://doi.org/10.1126/science.1195380.

Spada, F. *et al.* (2007) 'DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells', *The Journal of Cell Biology*, 176(5), pp. 565–571. Available at: https://doi.org/10.1083/jcb.200610062.

Sperger, J.M. *et al.* (2003) 'Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors', *Proceedings of the National Academy of Sciences*, 100(23), pp. 13350–13355. Available at: https://doi.org/10.1073/pnas.2235735100.

Sproul, D. *et al.* (2011) 'Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer', *Proceedings of the National Academy of Sciences*, 108(11), pp. 4364–4369. Available at: https://doi.org/10.1073/pnas.1013224108.

Sproul, D. *et al.* (2012) 'Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns', *Genome Biology*, 13(10), p. R84. Available at: https://doi.org/10.1186/gb-2012-13-10-r84.

Stelzer, Y. *et al.* (2015) 'Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution', *Cell*, 163(1), pp. 218–229. Available at: https://doi.org/10.1016/j.cell.2015.08.046.

Stewart-Morgan, K.R. *et al.* (2023) 'Quantifying propagation of DNA methylation and hydroxymethylation with iDEMS', *Nature Cell Biology*, 25(1), pp. 183–193. Available at: https://doi.org/10.1038/s41556-022-01048-x.

Stomper, J. *et al.* (2021) 'Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: mechanisms of resistance and novel HMA-based therapies', *Leukemia*, 35(7), pp. 1873–1889. Available at: https://doi.org/10.1038/s41375-021-01218-0.

Stresemann, C. and Lyko, F. (2008) 'Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine', *International Journal of Cancer*, 123(1), pp. 8–13. Available at: https://doi.org/10.1002/ijc.23607.

Subramanian, A. *et al.* (2005) 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences*, 102(43), pp. 15545–15550. Available at: https://doi.org/10.1073/pnas.0506580102.

Suzuki, M.M. and Bird, A. (2008) 'DNA methylation landscapes: provocative insights from epigenomics', *Nature Reviews. Genetics*, 9(6), pp. 465–476. Available at: https://doi.org/10.1038/nrg2341.

Swinney, D.C. and Anthony, J. (2011) 'How were new medicines discovered?', *Nature Reviews. Drug Discovery*, 10(7), pp. 507–519. Available at: https://doi.org/10.1038/nrd3480.

Syeda, F. *et al.* (2011) 'The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1', *The Journal of Biological Chemistry*, 286(17), pp. 15344–15351. Available at: https://doi.org/10.1074/jbc.M110.209882.

Taglini, F. *et al.* (2024a) 'DNMT3B PWWP mutations cause hypermethylation of heterochromatin', *EMBO reports*, pp. 1–26. Available at: https://doi.org/10.1038/s44319-024-00061-5.

Taglini, F. *et al.* (2024b) 'DNMT3B PWWP mutations cause hypermethylation of heterochromatin', *EMBO Reports*, 25(3), pp. 1130–1155. Available at: https://doi.org/10.1038/s44319-024-00061-5.

Takeshita, K. *et al.* (2011) 'Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)', *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), pp. 9055–9059. Available at: https://doi.org/10.1073/pnas.1019629108.

Tatarakis, A., Saini, H. and Moazed, D. (2023) 'Requirements for establishment and epigenetic stability of mammalian heterochromatin'. bioRxiv, p. 2023.02.27.530221. Available at: https://doi.org/10.1101/2023.02.27.530221.

Teale, W.D., Paponov, I.A. and Palme, K. (2006) 'Auxin in action: signalling, transport and the control of plant growth and development', *Nature Reviews Molecular Cell Biology*, 7(11), pp. 847–859. Available at: https://doi.org/10.1038/nrm2020.

Teschendorff, A.E. *et al.* (2012) 'Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation', *Genome Medicine*, 4(3), p. 24. Available at: https://doi.org/10.1186/gm323.

Timp, W. *et al.* (2014) 'Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors', *Genome Medicine*, 6(8), p. 61. Available at: https://doi.org/10.1186/s13073-014-0061-y.

Toyota, M. *et al.* (1999) 'CpG island methylator phenotype in colorectal cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 96(15), pp. 8681–8686.

Trowbridge, J.J. *et al.* (2009) 'DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells', *Cell Stem Cell*, 5(4), pp. 442–449. Available at: https://doi.org/10.1016/j.stem.2009.08.016.

Troyanovsky, B. *et al.* (2016) 'The Functionality of Minimal PiggyBac Transposons in Mammalian Cells', *Molecular Therapy. Nucleic Acids*, 5(10), p. e369. Available at: https://doi.org/10.1038/mtna.2016.76.

Turpin, M. and Salbert, G. (2022) '5-methylcytosine turnover: Mechanisms and therapeutic implications in cancer', *Frontiers in Molecular Biosciences*, 9. Available at: https://doi.org/10.3389/fmolb.2022.976862.

Uren, A.G. *et al.* (2005) 'Retroviral insertional mutagenesis: past, present and future', *Oncogene*, 24(52), pp. 7656–7672. Available at: https://doi.org/10.1038/sj.onc.1209043.

Utsumi, T. *et al.* (2021) 'ANKRD22 is an N-myristoylated hairpin-like monotopic membrane protein specifically localized to lipid droplets', *Scientific Reports*, 11(1), p. 19233. Available at: https://doi.org/10.1038/s41598-021-98486-8.

Vilkaitis, G. *et al.* (2005) 'Processive Methylation of Hemimethylated CpG Sites by Mouse Dnmt1 DNA Methyltransferase*', *Journal of Biological Chemistry*, 280(1), pp. 64–72. Available at: https://doi.org/10.1074/jbc.M411126200.

Villanueva, L., Álvarez-Errico, D. and Esteller, M. (2020) 'The Contribution of Epigenetics to Cancer Immunotherapy', *Trends in Immunology*, 41(8), pp. 676–691. Available at: https://doi.org/10.1016/j.it.2020.06.002.

Vincent, F. *et al.* (2022) 'Phenotypic drug discovery: recent successes, lessons learned and new directions', *Nature Reviews Drug Discovery*, 21(12), pp. 899–914. Available at: https://doi.org/10.1038/s41573-022-00472-w.

Volff, J.-N. (2009) 'Cellular genes derived from Gypsy/Ty3 retrotransposons in mammalian genomes', *Annals of the New York Academy of Sciences*, 1178, pp. 233–243. Available at: https://doi.org/10.1111/j.1749-6632.2009.05005.x.

Waalwijk, C. and Flavell, R.A. (1978a) 'MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaII sites.', *Nucleic Acids Research*, 5(9), pp. 3231–3236.

Waalwijk, C. and Flavell, R.A. (1978b) 'MspI, an isoschizomer of hpaII which cleaves both unmethylated and methylated hpaII sites.', *Nucleic Acids Research*, 5(9), pp. 3231–3236.

Waddington, C.H. (1942) 'Canalization of Development and the Inheritance of Acquired Characters', *Nature*, 150(3811), pp. 563–565. Available at: https://doi.org/10.1038/150563a0.

Waddington, C.H. (2012) 'The Epigenotype', *International Journal of Epidemiology*, 41(1), pp. 10–13. Available at: https://doi.org/10.1093/ije/dyr184.

Walbott, H. *et al.* (2007) 'Cysteine of sequence motif VI is essential for nucleophilic catalysis by yeast tRNA m5C methyltransferase', *RNA*, 13(7), pp. 967–973. Available at: https://doi.org/10.1261/rna.515707.

Walsh, C.P., Chaillet, J.R. and Bestor, T.H. (1998) 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nature Genetics*, 20(2), pp. 116–117. Available at: https://doi.org/10.1038/2413.

Walter, M.J. *et al.* (2011) 'Recurrent DNMT3A mutations in patients with myelodysplastic syndromes', *Leukemia*, 25(7), pp. 1153–1158. Available at: https://doi.org/10.1038/leu.2011.44.

Wang, Q. *et al.* (2020) 'Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance', *Nature Genetics*, 52(8), pp. 828–839. Available at: https://doi.org/10.1038/s41588-020-0661-y.

Wang, W. *et al.* (2008) 'Chromosomal transposition of PiggyBac in mouse embryonic stem cells', *Proceedings of the National Academy of Sciences*, 105(27), pp. 9290–9295. Available at: https://doi.org/10.1073/pnas.0801017105.

Wang, Y. *et al.* (2023) 'UHRF1 inhibition epigenetically reprograms cancer stem cells to suppress the tumorigenic phenotype of hepatocellular carcinoma', *Cell Death & Disease*, 14(6), pp. 1–13. Available at: https://doi.org/10.1038/s41419-023-05895-w.

Watanabe, D. *et al.* (2002) 'Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis', *Mechanisms of Development*, 118(1), pp. 187–190. Available at: https://doi.org/10.1016/S0925-4773(02)00242-3.

Watanabe, D., Uchiyama, K. and Hanaoka, K. (2006) 'Transition of mouse de novo methyltransferases expression from Dnmt3b to Dnmt3a during neural progenitor cell development', *Neuroscience*, 142(3), pp. 727–737. Available at: https://doi.org/10.1016/j.neuroscience.2006.07.053.

Waterland, R.A. (2006) 'Epigenetic mechanisms and gastrointestinal development', *The Journal of Pediatrics*, 149(5 Suppl), pp. S137-142. Available at: https://doi.org/10.1016/j.jpeds.2006.06.064.

Watson, J.V., Chambers, S.H. and Smith, P.J. (1987) 'A pragmatic approach to the analysis of DNA histograms with a definable G1 peak', *Cytometry*, 8(1), pp. 1–8. Available at: https://doi.org/10.1002/cyto.990080101.

Watt, F. and Molloy, P.L. (1988) 'Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter.', *Genes & Development*, 2(9), pp. 1136–1143. Available at: https://doi.org/10.1101/gad.2.9.1136.

Weber, J. *et al.* (1994) 'Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine', *Cancer Research*, 54(7), pp. 1766–1771.

Weber, S., Jung, S. and Doerfler, W. (2016) 'DNA methylation and transcription in HERV (K, W, E) and LINE sequences remain unchanged upon foreign DNA insertions', *Epigenomics*, 8(2), pp. 157–165. Available at: https://doi.org/10.2217/epi.15.109.

Weinberg, D.N. *et al.* (2019) 'The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape', *Nature*, 573(7773), pp. 281–286. Available at: https://doi.org/10.1038/s41586-019-1534-3.

Weinberg, D.N. *et al.* (2021) 'Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands', *Nature Genetics*, 53(6), pp. 794–800. Available at: https://doi.org/10.1038/s41588-021-00856-5.

Weinberg, R.A. and Weinberg, R.A. (2013) *The Biology of Cancer*. New York: W.W. Norton & Company. Available at: https://doi.org/10.1201/9780203852569.

Weisenberger, D.J. *et al.* (2006) 'CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer', *Nature Genetics*, 38(7), pp. 787–793. Available at: https://doi.org/10.1038/ng1834.

Weon, J.L. and Potts, P.R. (2015) 'The MAGE protein family and cancer', *Current opinion in cell biology*, 37, pp. 1–8. Available at: https://doi.org/10.1016/j.ceb.2015.08.002.

Wilson, M.H., Coates, C.J. and George, A.L. (2007) 'PiggyBac Transposon-mediated Gene Transfer in Human Cells', *Molecular Therapy*, 15(1), pp. 139–145. Available at: https://doi.org/10.1038/sj.mt.6300028.

Worley, C.K. *et al.* (2000) 'Degradation of Aux/IAA proteins is essential for normal auxin signalling', *The Plant Journal: For Cell and Molecular Biology*, 21(6), pp. 553–562. Available at: https://doi.org/10.1046/j.1365-313x.2000.00703.x.

Wrangle, J. *et al.* (2013) 'Alterations of immune response of non-small cell lung cancer with Azacytidine', *Oncotarget*, 4(11), pp. 2067–2079.

Wu, B. and Hur, S. (2015) 'How RIG-I like receptors activate MAVS', *Current opinion in virology*, 12, pp. 91–98. Available at: https://doi.org/10.1016/j.coviro.2015.04.004.

Wu, S.C.-Y. *et al.* (2006) 'piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 103(41), pp. 15008–15013. Available at: https://doi.org/10.1073/pnas.0606979103.

Wu, X. and Zhang, Y. (2017) 'TET-mediated active DNA demethylation: mechanism, function and beyond', *Nature Reviews Genetics*, 18(9), pp. 517–534. Available at: https://doi.org/10.1038/nrg.2017.33.

Xue, B. *et al.* (2019) 'Epigenetic mechanism and target therapy of UHRF1 protein complex in malignancies', *OncoTargets and therapy*, 12, pp. 549–559. Available at: https://doi.org/10.2147/OTT.S192234.

Yamaguchi, K. *et al.* (2023) 'Non-canonical functions of UHRF1 maintain DNA methylation homeostasis in cancer cells'. bioRxiv, p. 2023.07.11.548318. Available at: https://doi.org/10.1101/2023.07.11.548318.

Yamaguchi, K. *et al.* (2024) 'Non-canonical functions of UHRF1 maintain DNA methylation homeostasis in cancer cells', *Nature Communications*, 15(1), p. 2960. Available at: https://doi.org/10.1038/s41467-024-47314-4.

Yang, A.S. *et al.* (2004) 'A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements', *Nucleic Acids Research*, 32(3), p. e38. Available at: https://doi.org/10.1093/nar/gnh032.

Yang, C.-R. *et al.* (2020) 'The RNA-binding protein DAZL functions as repressor and activator of mRNA translation during oocyte maturation', *Nature Communications*, 11, p. 1399. Available at: https://doi.org/10.1038/s41467-020-15209-9.

Yates, L.R. and Campbell, P.J. (2012) 'Evolution of the cancer genome', *Nature Reviews Genetics*, 13(11), pp. 795–806. Available at: https://doi.org/10.1038/nrg3317.

Yin, J. *et al.* (2017) 'ANKRD22 promotes progression of non-small cell lung cancer through transcriptional up-regulation of E2F1', *Scientific Reports*, 7(1), p. 4430. Available at: https://doi.org/10.1038/s41598-017-04818-y.

Yoder, J.A., Walsh, C.P. and Bestor, T.H. (1997) 'Cytosine methylation and the ecology of intragenomic parasites', *Trends in genetics: TIG*, 13(8), pp. 335–340. Available at: https://doi.org/10.1016/s0168-9525(97)01181-5.

Yusa, K., Takeda, J. and Horie, K. (2004) 'Enhancement of Sleeping Beauty Transposition by CpG Methylation: Possible Role of Heterochromatin Formation', *Molecular and Cellular Biology*, 24(9), pp. 4004–4018. Available at: https://doi.org/10.1128/MCB.24.9.4004-4018.2004.

Zeidan, A.M., Kharfan-Dabaja, M.A. and Komrokji, R.S. (2014) 'Beyond hypomethylating agents failure in patients with myelodysplastic syndromes', *Current opinion in hematology*, 21(2), pp. 123–130. Available at: https://doi.org/10.1097/MOH.0000000000000016.

Zhang, X. *et al.* (no date) 'Double-edged effects of interferons on the regulation of cancer-immunity cycle', *Oncoimmunology*, 10(1), p. 1929005. Available at: https://doi.org/10.1080/2162402X.2021.1929005.

Zhao, H. *et al.* (2017) 'The immunomodulatory anticancer agent, RRx-001, induces an interferon response through epigenetic induction of viral mimicry', *Clinical Epigenetics*, 9, p. 4. Available at: https://doi.org/10.1186/s13148-017-0312-z.

Zhao, H. *et al.* (2021) 'Cyclin G2, a novel target of sulindac to inhibit cell cycle progression in colorectal cancer', *Genes & Diseases*, 8(3), pp. 320–330. Available at: https://doi.org/10.1016/j.gendis.2020.11.006.

Zhao, R. *et al.* (2016) 'Implications of Genetic and Epigenetic Alterations of *CDKN2A* (p16INK4a) in Cancer', *EBioMedicine*, 8, pp. 30–39. Available at: https://doi.org/10.1016/j.ebiom.2016.04.017.

Zheng, Y. *et al.* (2023) 'Comprehensive analyses of partially methylated domains and differentially methylated regions in esophageal cancer reveal both cell-type- and cancer-specific epigenetic regulation', *Genome Biology*, 24(1), p. 193. Available at: https://doi.org/10.1186/s13059-023-03035-3.

Zhong, S. *et al.* (2007) 'Pharmacologic inhibition of epigenetic modifications, coupled with gene expression profiling, reveals novel targets of aberrant DNA methylation and histone deacetylation in lung cancer', *Oncogene*, 26(18), pp. 2621–2634. Available at: https://doi.org/10.1038/sj.onc.1210041.

Zhou, W. *et al.* (2018) 'DNA methylation loss in late-replicating domains is linked to mitotic cell division', *Nature Genetics*, 50(4), pp. 591–602. Available at: https://doi.org/10.1038/s41588-018-0073-4.

Zhu, J. *et al.* (2014) 'Antiviral activity of human oligoadenylate synthetases-like (OASL) is mediated by enhancing retinoic acid-inducible gene I (RIG-I) signaling', *Immunity*, 40(6), pp. 936–948. Available at: https://doi.org/10.1016/j.immuni.2014.05.007.