THE UNIVERSITY
of EDINBURGH

Low levels of methylation are targeted to aberrantly methylated CGIs in human colorectal cancer cells

Roza Masalmeh

THE UNIVERSITY
of EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2018

Declaration

I declare that the work presented in this PhD thesis is my own except where otherwise stated. This work has not been submitted for any other degree.

Roza Masalmeh

September 2018

Acknowledgements

Firstly, I wish to thank my supervisor Duncan Sproul for taking me as his first PhD student and for excellent scientific training under his supervision. His door was always open for my questions and through our discussions he trained me to think independently. At the start of my PhD, I was worried about making mistakes and he told me something that really made a difference. He said, “You can make as many mistakes as you want and these mistakes are important because you will learn from them”. Duncan, thank you for all your support, I couldn’t ask for more.

I also wish to thank all members of my lab Hazel Davidson-Smith, Cristina Ramon, Francesca Taglini, Jon Higham, Ioannis Kafetzopoulos and Nicholas Younger with whom I had great times discussing science and ate many tasty chocolate fondues. To my friends Charlene Lemaitre, Nefeli Dellepiane, Christine Mordstein, Heidi Mjoseng, Pragya Mittal, Shipra Bhatia, Yatendra Kumar, Toby Gurran, Nele Hug, Donncha Dunican, Shelagh Boyle, Abby Mann and Rob Illingworth; you made the IGMM my second home.

I wish to thank my thesis examiners Liz Patton and Gabi Ficz for the rich discussion and making my viva a very good experience. I wish to thank my PhD committee which consists of Wendy Bickmore, Bob Hill, Val Brunton and Ian Adam for their invaluable advice. I also wish to thank David Hay, Kathryn Jones, Ilya Flyamer and Nefeli Dellepiane for reading drafts of my thesis chapters and their feedback.

To my parents who raised me to think that I could do whatever I want and always supported me in doing so, thank you so much. There are no words to express how much I am appreciative of all that you have given me throughout the years.

Big thank you to the person who believed in me, always encouraged me to peruse my dreams, and supported me by all means in doing so, my husband and love, Menwer. Menwer even learnt how to cook in order for me to spend more time in the lab and come home to find a delicious meal waiting for me. To my little handsome man, my son, Amr who joined me in the lab for nine months and was much behaved. Amr, you enriched our lives and gave us love and happiness very generously. To my brothers, my in laws, and my friends in Palestine thank you for all the love and support you have been giving me throughout the years.

Abstract

Disruption of DNA methylation patterns is a primary hallmark of cancer. One main disruption is CpG islands (CGIs) aberrant hypermethylation which is associated with transcriptional repression of tumour suppressor genes such as *BRCA1*, *MLH1*, and *CDKN2A* (*p16/ARF*). However, the mechanism(s) underpinning this process are still unknown. One hypothesis is that high *de novo* methylation activity is targeted to CGIs that get aberrantly methylated in cancer. To investigate this, I used human colorectal cancer cell line (HCT116) as a model. Firstly, I have ectopically integrated representative aberrantly methylated CGIs randomly into the genome and assessed their methylation at ectopic loci. I show that integrated CGIs rarely gained methylation at ectopic locations. This suggested that the sequence doesn't program CGIs aberrant methylation at ectopic loci in HCT116 and that low *de novo* methylation is targeted to aberrantly methylated CGIs. Next, I wanted to confirm this result by integrating CGIs at a chosen genomic loci in order to exclude any effects from the CGI position on its methylation status. This was performed by recombinase mediated cassette exchange (RMCE). Successful targeting of a RMCE cassette to the desired chromosomal locus was achieved. However, isolating clonal cell lines with integrated CGIs was technically infeasible due to difficulty in using the thymidine kinase selection in HCT116. In order to investigate which CGIs have higher *de novo* methylation in HCT116 on a genome wide level, I restored DNMT3B expression in DNMTA1/DNMT3B double knockout (DKO) of HCT116. In agreement with previous reports, DNMT3B showed higher *de novo* methylation activity at CGIs overlapping H3K36me3. Most importantly, aberrantly methylated CGIs showed ~ 78.5% less methylation gain compared to normally methylated CGIs. To confirm results from this experiment, I treated HCT116 with 5-Aza-2′-deoxycytidine and then allowed cells to recover

methylation. The rate of methylation recovery of CGIs was assessed. This showed that aberrantly methylated CGIs recover slower than normally methylated CGIs, confirming that lower de novo methylation is targeted to aberrantly methylated CGIs compared to normally methylated ones. The work carried out during the course of this thesis provides novel insights into the process of *de novo* methylating CGIs in cancer on a genome wide level. It suggests that low levels, rather than high levels, of methylation are targeted to aberrantly methylated CGIs in colorectal tumours.

Lay Abstract

Every cell in our bodies contains ~ 20,000 genes, which hold instructions guiding the cells how to function. Each gene has a promoter acting like a switch to turn it on and off. Chemical marks are attached to DNA, affecting how these instructions are interpreted. The correct patterning of these marks is essential to keep cells in normal state. One of these marks is DNA methylation where a methyl group is attached to the DNA. When DNA methylation is attached to a gene promoter it associates with that gene being turned off. This mark is disrupted in cancer where many gene promoters gain methylation. One possible explanation to how this happen is that a protein in the cells responsible in depositing the methyl mark, called DNMT3B, is attracted to the promoter of these genes. In my PhD project, I am testing if this is true. I show that when promoters of genes that gain methylation in cancer are inserted in the DNA of a human cancer cell line, the inserted promoters do not get methylated. This suggests that the protein that deposits methylation is not highly attracted to these promoters. In order to comprehensively test if all promoters gaining methylation in cancer show this trend, I use cells where methylation is removed and ask if I introduce DNMT3B will it be highly attracted to these promoters. I show that DNMT3B is highly attracted to the promoters that are methylated in normal healthy conditions and that low methylation goes to the promoters that gain methylation in cancer. I next confirm this finding by first treating colon cancer cells with a drug that removes methylation and then allowing cells to rewrite methylation. In agreement with the previous experiment, the methylation goes back to normal methylated promoters quicker than it does to promoters methylated in cancer. This thesis shows that promoters that gain methylation in cancer have low attraction to proteins that deposit methylation in colorectal tumours. Studying the process of methylation gain at promoters further will help us better understand how some of the genes are turned off in cancer.

List of abbreviations

5-Aza 5-aza deoxycytodine

5mC 5-methyl cytosine

5hmC 5-hydroxymethyl cytosine

5fC 5-formylcytosine

5caC 5-carboxylcytosine

°C Degrees centigrade

A Adenine

BER Base excision repair

bp Base-pair

C Cytosine

cDNA Complementary deoxyribonucleic acid

CGI CpG island

CO2 Carbon dioxide

CpG Cytosine and guanine separated by a phosphate

CRISPR clustered regularly interspaced short palindromic repeats

C-terminal Carboxy-terminal

dH20 Distilled water

DKO Double knockout (DNMT1/DNMT3B knockout)

DMEM Dulbecco's modified eagle medium

DMSO Dimethyl sulphoxide

DMR Differentially methylated region

DNA Deoxyribonucleic Acid

DNase Deoxyribonuclease

Dnmt Dna methyltransferase

dNTP Deoxynucleotide triphosphate

Drosophila *Drosophila melanogaster*

EDTA Ethylenediamineotetra-acetic acid

ES Embryonic stem

FCS Foetal calf serum

G Guanine

GCV Ganciclovir

GFP Green fluorescent protein

HCT116 Human colorectal cancer cell line

H3K4me3 Histone H3 lysine 4 trimethylation

H3K9me3 Histone H3 lysine 9 trimethylation

H3K27ac Histone H3 lysine 27 acetylation

H3K27me3 Histone H3 lysine 27 trimethylation

HCl Hydrochloric acid

Kb Kilobase pairs of DNA

kDa KiloDaltons

KO Knockout

l liter

LTR Long terminal repeat

M Molar

mES Mouse embryonic stem (cell)

MDEM Dulbecco's Modified Eagle Medium

µg Microgram

mg Milligram

µl Microlitre

min Minute

mol mole

mRNA Messenger RNA

Mw Molecular weight

ng nanogram

nt Nucleotide

N-terminal N amino-terminal

Oligo Oligonucleotide

PAGE Polyacrylamide gel electrophoresis

PBS Phosphate buffered saline

PBST Phosphate buffered saline (plus 0.1% w/v Tween)

PCR Polymerase chain reaction

pen Penicillin

PFA Paraformaldehyde

PGC Primordial germ cell

PRC Polycomb repressive complex

qRT-PCR Quantitative reverse-transcriptase PCR

RNA Ribonucleic acid

RT-PCR Reverse transcriptase polymerase chain reaction

rpm Revolutions per minute

RRBS Reduced representation bisulfite sequencing

RT Room temperature

SDS Sodium dodecyl sulphate

strep Streptomycin

TBE Tris-EDTA boric acid buffer

TK Thymidine kinase

TKO Triple knockout

TSS Transcription start site

UTR Untranslated region

Vol Volumes

WT Wild-type

X-gal 5-Bromo-4-Chloro-3-indolyl B-D-galactopyranoside

Table of Contents

Chapter 1: Introduction 1

1.1 Epigenetics 1

1.1.1 DNA methylation 1

1.1.2 Histone modifications 3

1.2 DNMT enzymology 4

1.2.1 DNMT3 family 4

1.2.2 DNMT1 9

1.3 DNA demethylation 13

1.4 DNA methylation patterns in normal systems 14

1.5 The dynamics of methylation in mammalian embryo development 18

1.6 Cancer 20

1.6.1 Colorectal Cancer 23

1.6.2 DNA methylation changes in cancer and CRC 23

1.6.3 Proposed mechanisms of CGI hypermethylation in cancer 26

1.7 Aims of the thesis 30

Chapter 2: Materials and Methods 33

2.1 Reagents, Stock Solutions, and Buffers 33

2.2 Mammalian Cell Culture 35

2.2.1 Cell lines used 35

2.2.2 Thawing Cells from Storage 35

2.2.3 Maintenance of cell lines 35

2.2.4 Cell transfection 36

2.2.5 Cell selection 36

2.2.6 DNMT3B rescue in DKO cells 37

2.2.7 5-Aza-2'-deoxycytidine treatment 38

2.3 DNA extraction and manipulation 38

2.3.1 DNA isolation from cells 38

2.3.2 Resolution of DNA on Agarose Gels 40

2.3.3 Gel purification of DNA 40

2.3.4 Assessing quality and quantity of nucleic acids 40

2.3.5 Southern blotting 41

2.3.6 Bisulphite primer design 42

2.3.7 BS treatment 42

2.3.8 BS PCR and sequencing 43

2.3.9 Analysis of total CRISPR editing frequency by SURVEYOR nuclease assay 43

2.3.10 Analysis of homologous recombination by HindIII restriction digestion 44

2.3.11 Preparation of samples for mass spectroscopy and data analysis 44

2.3.12 RRBS samples preparation 45

2.4 RNA Extraction and Manipulation 46

2.4.1 RNA isolation and purification 46

2.4.2 cDNA synthesis 46

2.4.3 qRT-PCR 47

2.5 Protein Extraction and Manipulation 47

2.5.1 Protein isolation and purification 47

2.5.2 Western blotting 48

2.6 Bacterial Culture and Manipulation 49

2.6.1 Transformation of DH5-α E.Coli cells 49
2.6.2 Bacterial growth 49
2.6.3 Plasmid DNA isolation 50
2.6.4 PiggyBac constructs preparation 50
2.6.5 CRISPR targeting constructs preparation 51
2.6.6 L1-PuroTK-1L and L1-CGI-1L constructs preparation 51
2.6.7 Creating HCT116 cell line with L1-PuroTK-1L cassette at β-globin locus. 52
2.6.8 DNMT3B2 construct preparation 52
2.7 Bioinformatics analysis 53
2.7.1 RRBS data 53
2.7.2 TCGA data analysis 53
2.7.3 Heatmaps 55
Chapter 3: Ectopically integrated CpG islands are rarely hypermethylated in cancer cells 60
3.1 Experimental strategy 61
3.2 The majority of hypermethylated CGIs in colorectal cancer samples are also hypermethylated in HCT116 62
3.3 Selection of representative CGIs for my study 66
3.4 Experimental optimization 70
3.4.1 PiggyBac minimal terminal repeats are sufficient for transposition in HCT116. 70
3.4.2 Culturing cells for 4 weeks after transfection with piggyBac plasmid is sufficient for diluting the plasmid out 72
3.5 CGIs are often unmethylated at ectopic loci in HCT116 74
3.5.1 The full length MLH1 CGI maintains its unmethylated state at ectopic loci in HCT116 74
3.5.2 Truncated MLH1 CGI maintain their unmethylated status at ectopic loci in HCT116 76
3.5.3 Hypermethylated CGIs are rarely hypermethylated at ectopic loci in HCT116 78

3.6 Discussion 78

Chapter 4: Targeted Integration of CGIs 85

4.1 Experimental Design 85

4.2 Selecting a genomic locus near β-globin genes cluster for the targeted integration 88

4.3 Generating HCT116 clonal cell line with engineered RMCE cassette near the β-globin locus 90

4.4 Generation of HCT116 cell lines with integrated L1-PuroTK-1L into β -globin locus and α-globin locus. 92

4.5 Integration of CGIs into β -globin locus by RMCE 94

4.5.1 RMCE efficiency is very low in HCT116 99

4.6 Discussion 102

Chapter 5: Aberrantly methylated CGIs are not primary targets of DNMT3B 104

5.1 Experimental design 104

5.2 Generation of HCT116 model to track DNMT3B2 targeting 104

5.3 Assessing methylation changes by RRBS 108

5.4 A subgroup of CGIs in DKOs gained methylation after expressing DNMT3B2 111

5.5 CGIs overlapping H3K36me3 gain higher methylation after expressing DNMT3B2 in DKO cells 114

5.6 Aberrantly methylated CGIs showed less methylation compared to normally methylated CGIs 117

5.7 Discussion 124

Chapter 6: Low methylation levels are targeted to aberrantly methylated CGIs 128

6.1 Experimental design 128

6.2 Establishing HCT116 remethylation model utilizing 5-Aza-2′-deoxycytidine 130

6.3 Recovery after 5-Aza-2′-deoxycytidine treatment reflects remethylation dynamics 132

6.4 CGIs targeted by DNMT3B2 show higher remethylation rate in HCT116 after treatment with Aza-2′-deoxycytidine 138

6.5 Aberrantly methylated CGIs show lower recovery rate than CGIs overlapping H3K36me3 and normally methylated CGIs 140

6.6 No significant difference between remethylation rates of aberrantly methylated CGIs and CGIs overlapping H3K36me3 in the absence of DNMT3B 140

6.7 Methylation seeding can't explain the differences in remethylation rate of CGIs. 143

..... 144

6.8 Discussion 145

Chapter 7: Conclusions and Future Perspectives 147

7.1 Optimizing a hypomethylation tool for *in vivo* verification of the results 148

7.2 understanding the dynamics of aberrant hypermethylation in cancer 148

7.3 Is seeding methylation a prerequisite for aberrant hypermethylation of aberrantly methylated CGIs? 151

7.4 Do CGIs overlapping H3K36me3 and normally methylated CGIs become methylated at ectopic sites? 151

7.5 Therapeutic implications 152

Appendix 154

References 167

Chapter 1: Introduction

1.1 Epigenetics

1.1.1 DNA methylation

DNA is a polynucleotide made of four building blocks; adenine, thymine, cytosine and guanine. Cytosine and adenine are also found tagged with a methyl group (-CH3) coming from S-adenosylmethionine (SAM), giving rise to methylated variants of these bases. This chemical modification marks the genome without affecting the Watson–Crick base pairing of the underlying DNA nucleotides (Schübeler, 2015).

In prokaryotes, methylation can be found as 5-methylcytosine (5-mC) ((Hotchkiss, 1948), 4-methylcytosine (4-mC) (Janulaitis *et al.*, 1983) and 6-methyladenosine (6-mA)(Dunn and Smith, 1955) (figure1.1A). They are thought to be play an important role in protecting the genome of prokaryotes from endonucleases expressed as defense mechanism against invading viruses (Wilson and Murray, 1991).

5-mC is the predominant form of DNA methylation in eukaryotes. In most animals, this modification mainly occurs in a CpG dinucleotide context (a cytosine followed by guanine on the same DNA strand in the 5' to 3' direction). The symmetric methyl group sits in the major groove of DNA, and therefore has the potential to affect how proteins interact with DNA (figure 1.1B). Methylation of cytosine outside a CpG context was detected at low levels in ESC cells and neuronal cells (Ramsahoye *et al.*, 2000; Guo *et al.*, 2013; Habibi *et al.*, 2013; Lister *et al.*, 2013; Schultz *et al.*, 2015). 5-mC can be modified to 5-

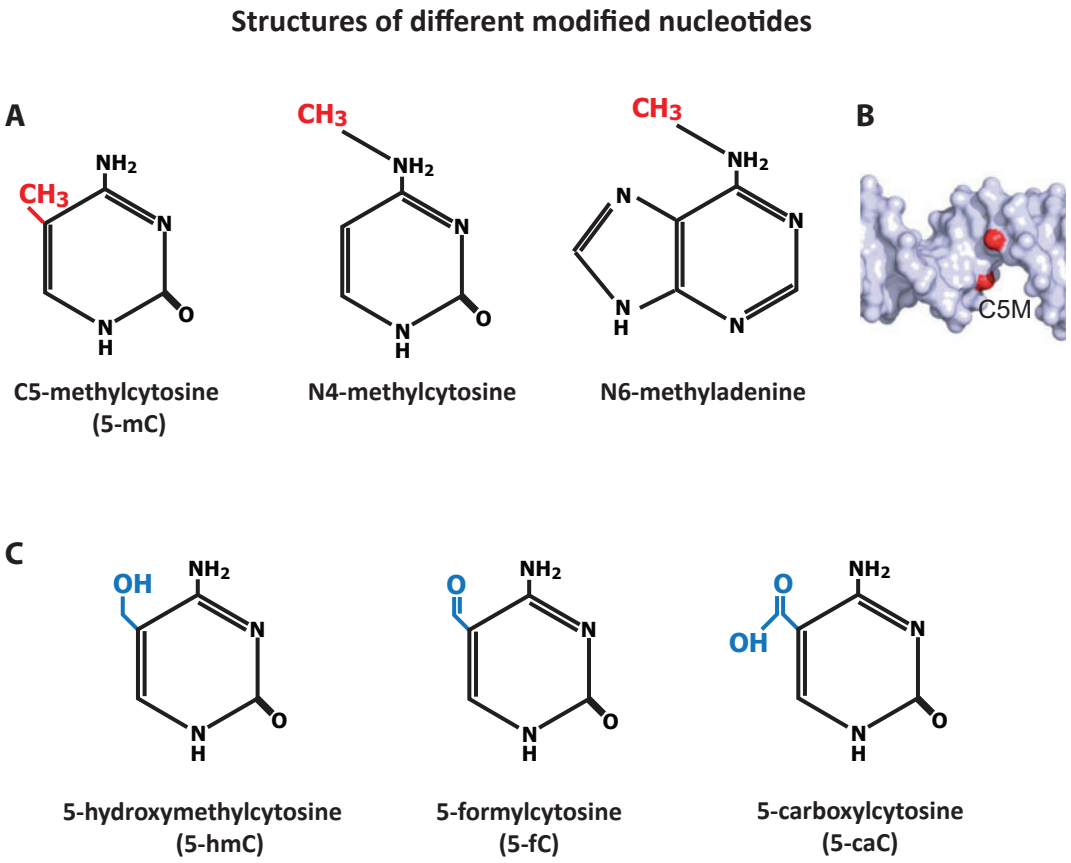


Figure 1.1: Chemical structures of different modified nucleotides
A. Chemical structures of three forms of methylated nucleotides: 5-methylcytosine (5-mC), 4-methylcytosine (4-mC) and 6-methyladenosine (6-mA).
B. Model of methylated DNA showing the symmetrically methylated CpGs with the methyl groups (in red) sitting in the major groove of DNA (adapted from Dantas Machado et al., 2015).
C. Chemical structures of other modified cytosines: 5-hydroxymethylcytosine (5-mhC), 5-formylcytosine (4-fC) and 5-carboxylcytosine (6-caC) found in vertebrates.

hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fc), and 5-carboxylcytosine (5-caC) (figure 1.1C) (Kriaucionis and Heintz, 2009; He *et al.*, 2011).hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fc), and 5-carboxylcytosine (5-caC) (figure 1.1C) (Kriaucionis and Heintz, 2009; He *et al.*, 2011).

Lower eukaryotes such as *Drosophila melanogaster* have low methylation levels, most of which takes place in CpA context (Jaenisch, Lyko and Ramsahoye, 2000; Krauss and Reuter, 2011). In plants and fungi, most methylation takes place in a CpH sequence, where H can be any nucleotide but G (Selker, 1993, 1997; Cokus et al., 2008; Lister et al., 2008; Jeon et al., 2015). N6-methyladenine was discovered in a number of eukaryotes including insects, nematodes and green algae (Fu et al., 2015; Greer et al., 2015; Zhang et al., 2015). Additionally, a recent study reported the detection of N6-methyladenine (6mA) in mouse embryonic stem cells and human cultured cells (Koziol et al., 2015; Wu et al., 2016). However this finding is controversial, as this modification was undetectable by quantitative LC-MS in ESCs (Schiffers et al., 2017). Moreover, a reassement of DNA immunoprecipitation-based genomic profie suggested that the detection of 5mA in ESCs by antibody-based enrichment methods was a false positive due to off-target binding inherent in all antibody based methods, and the detection of 5mA in one study by mass spectrometry could be false due to contamination with 6mA-rich bacteria such as *Mycoplasma* (Lentini *et al.*, 2018).

1.1.2 Histone modifications

In addition to DNA methylation, the other main epigenetic mark modulating chromatin states is modification of histones. Histones are small, conserved, and abundant proteins that, in humans, are used to package 2 meters of gDNA into a nucleus of ~ 6 micrometres (μm) diameter (Peterson and Laniel, 2004). DNA

fragments of 100-174bp length are wrapped, in “beads-on-a-string” form, around histone octamers made of two H2A-H2B dimers and a tetramer of H3-H4 forming nucleosome cores (Zlatanova, Leuba and van Holde, 1999). Histones have flexible N-terminal domains that protrude from the nucleosomes and has a tail-like structure serving as a platform for post-translational modifications such as methylation, acetylation, ubiquitination, phosphorylation, citrullination, and sumoylation (Schroth *et al.*, 1990; Peterson and Laniel, 2004). Post-translational modifications can also be found on the globular domain of histones (Kebede, Schneider and Daujat, 2015). These modifications play a role in gene regulation by dynamic modulation of chromatin (Bannister and Kouzarides, 2011).

1.2 DNMT enzymology

DNA methylation is written and maintained by a group of enzymes called DNA methyl transferase enzymes (hDNMTs in humans, and Dnmts or mDNMTs in mice). The methyl transferase core of these enzymes is widely conserved from bacteria to plants and mammals (Bestor, 2000; Cao *et al.*, 2000). Mammalian DNMTs are classified into two families; the DNMT3 family which mainly *de novo* methylate DNA, and DNMT1 which is the main maintenance methylation enzyme (Li, Okano and Xie, 1998; Okano *et al.*, 1999; Hata *et al.*, 2002).

1.2.1 DNMT3 family

In mammals, waves of DNA demethylation take place during gametogenesis and preimplantation development erasing most of the methylation patterns inherited from parents and allowing a totipotent state. Following that, during germ cell differentiation and at blastocyst implantation stage, DNA methylation is established by DNMT3 enzymes. The main active enzymes in the DNMT3 family are DNMT3A and DNMT3B. DNMT3-like (DNMT3L) lacks methylation activity

and is expressed during gametogenesis and embryonic development where it binds to DNMT3A and enhances its activity (Bourc'his *et al.*, 2001; Hata *et al.*, 2002; Bourc'his and Bestor, 2004). Several lines of evidence suggested that DNMT3 enzymes are the main *de novo* methylation enzymes. Firstly, mDNMT3A and mDNMT3B methylate DNA *in vitro* with no preference for hemi-methylated DNA (Li, Okano and Xie, 1998). Secondly, mouse embryonic stem cells (mESCs) lacking either mDNMT3A or mDNMT3B can methylate proviral DNA, whereas knocking out the two enzymes impairs *de novo* methylation activity (Okano *et al.*, 1999). Thirdly, [mDNMT3A -/-, mDNMT3B -/-] double mutant mouse embryos are incapable of re-establishing methylation at postimplantation stage after the global wave of demethylation (figure 1.2A).

The two enzymes have overlapping functions as suggested by the fact that the DNA of homozygous double mutant [mDNMT3B -/-, mDNMT3A -/-] mice is hypomethylated at endogenous C-type retroviral DNA and IAP repeats, whereas in mDNMT3A and mDNMT3B single mutant mice it retains methylation at these repeats. However, mDNMT3B has distinct targets as mDNMT3B -/-, but not mDNMT3A -/-, mutant mice are hypomethylated at the minor satellites and the single mutants show different phenotypes (Okano *et al.*, 1999). Additionally, hDNMT3B disruption in humans with ICF syndrome (Immunodeficiency, Centromere instability and Facial anomalies syndrome) is associated with hypomethylation of satellite II and III repeats, which are rich at pericentromeric regions (Ueda *et al.*, 2006; Heyn *et al.*, 2012). *De novo* methylation activity is high in postimplantation embryos, germ cells and ES cells, and is reduced in differentiated somatic cells (Li, Okano and Xie, 1998). Interestingly, a new member of the DNMT3 family has been recently identified in rodents, DNMT3C, which is responsible for methylating transposons in the male germ line (Barau *et al.*, 2016).

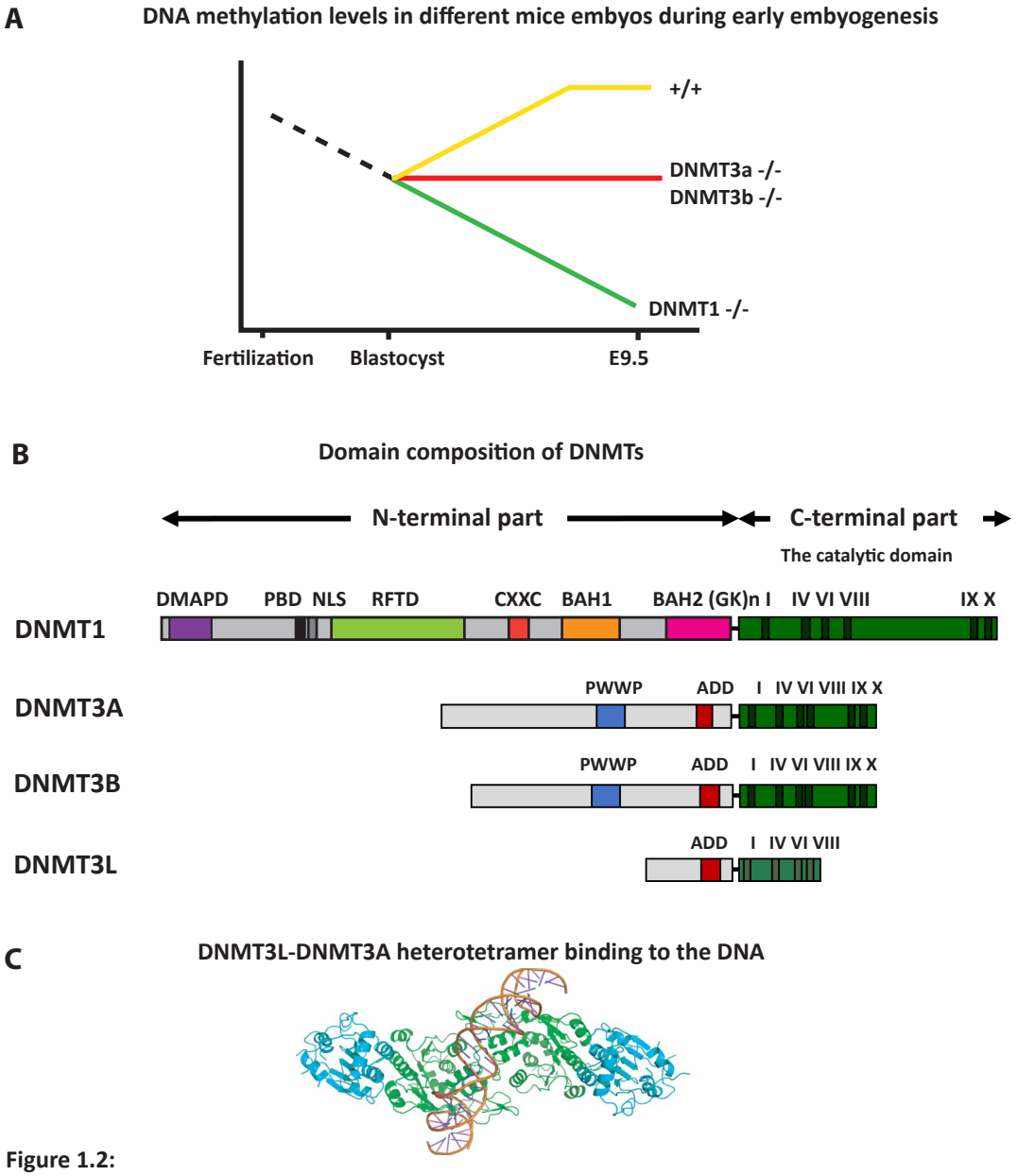


Figure 1.2:

A. Schematic diagram of global DNA methylation levels during mouse early embryogenesis. After fertilization, the genome undergoes a global wave of hypomethylation. Post-implantation, the genome of wild type embryo is de novo methylated (yellow line). Whereas, [DNMT3B -/-, DNMT3A -/-] double mutant embryos don't establish methylation, but retain some levels of methylation (red line). DNMT1 -/- mutant embryos lose methylation globally (adapted from Okano *et al.*, 1999).

B. Schematic representation of domain composition of DNA methyl transferase enzymes. All DNMTs are composed of two functional parts (the N-terminal and the C-terminal) linked through a fixable lysine-glycine repeats (GK)n linker. The C-terminal constitute the catalytic part of the enzymes, which is regulated by the different labelled domains in the N-terminal part (taken from Jeltsch and Jurkowska, 2016).

C. Structure of DNMT3L-DNMT3A C-terminal domains. DNMT3A binds DNMTL and forms a heterotetramer with DNMT3A dimer in the center and two DNMT3L at the edges. This enhances the binding of DNMT3A to the DNA (taken from Jia *et al.*, 2007).

The three enzymes of the DNMT3 family are composed of two parts, the N-terminus and the catalytic C-terminus (figure 1.2B). Targeting of DNMT3 enzymes is mainly controlled by their N-terminus part (Jeltsch and Jurkowska, 2016). This part contains an ATRX-DNMT3-DNMT3L (ADD) domain. The ADD domain tethers these enzymes to H3 tails that are unmethylated at Lys4 (H3K4) (Otani *et al.*, 2009). The ADD domain has a shallow groove where the H3 tail can fit and form hydrogen bonds. Methylation of H3K4 hinders this interaction (Guo *et al.*, 2015). The crystal structure of DNMT3A binding to H3 was resolved, and it showed that this interaction pulls the ADD domain away from the catalytic core allowing substrate DNA to bind. It is assumed that the same applies for DNMT3B; however, this hasn't been directly demonstrated for DNMT3B as no crystal structure for DNMT3B binding to the H3 tail has been resolved yet. The ADD domain is also implicated in targeting DNMT3A/B to heterochromatin through interacting with various histone marks and heterochromatin associated proteins (Fuks *et al.*, 2003; Smallwood *et al.*, 2007; Gazzar *et al.*, 2008).

Another important domain in the N-terminus part of DNMT3A and DNMT3B is the PWWP domain which is absent in DNMT3L. PWWP domains have the shape of a barrel with a hydrophobic core and a positive charge on the surface. This domain is implicated in the binding of DNMT3B, but not DNMT3A, to DNA nonspecifically (Chen, Tsujimoto and Li, 2004). Additionally, it was shown that it recruits DNMT3B *in vivo* to gene bodies by recognizing H3K36me3 (Baubec *et al.*, 2015; Rondelet *et al.*, 2016). The PWWP domain in DNMT3A has been shown to bind to H3K36me3 *in vitro* (Dhayalan *et al.*, 2010). However, in mESCs DNMT3A1 and DNMT3A2 were not localized with H3K36me3 despite possessing PWWP domains (Manzo *et al.*, 2017a). The PWWP domain was suggested to play a role in DNMT3A and DNMT3B recruitment at pericentromeric regions as well as their anchoring on nucleosomes (Chen, Tsujimoto and Li, 2004; Ge *et al.*, 2004). In ICF syndrome, mutations in the PWWP domain of DNMT3B are implicated in abolishing DNMT3B binding to

pericentromeric regions and consequently their hypomethylation (Shirohzu *et al.*, 2002; Ueda *et al.*, 2006). This notion needs further investigation, as ICF mutations in the PWWP domain are rare and most mutations are point mutations in the catalytic domain of DNMT3B (Weemaes *et al.*, 2013).

The sequence upstream of PWWP is different in DNMT3A and DNMT3B, and might have a role in their differential targeting. In DNMT3B, this sequence interacts with Centromere protein C (CENP-C) and thus was suggested to play a role in its tethering to centromeric and pericentromeric repeats (Gopalakrishnan *et al.*, 2009). Additionally, different isoforms of DNMT3A bind to different targets in a tissue-dependent manner (Manzo *et al.*, 2017b). The N-terminus of DNMT3A1 isoform was found to preferentially bind to shores of H3K27me3-positive CGIs.

DNMT3A and DNMT3B have some differences in the way they methylate DNA. DNMT3A has one polar interface that interacts with the same interface in another DNMT3A molecule to form a dimer. The other interface of DNMT3A is hydrophobic and binds DNMTL. DNMT3A and DNMT3L together form a heterotetramer with DNMT3A dimer in the center and two DNMT3L at the edges (figure 1.2C). This complex has a larger binding pocket and higher affinity for DNA than DNMT3A alone (Jia *et al.*, 2007). In the absence of DNMT3L, DNMT3A oligomerises forming a filament that is enriched on heterochromatin (Jurkowska *et al.*, 2011). A heterozygous mutation at the polar interface of DNMT3A exists in more than 50% of AML cases, and has been found to act as a dominant negative mutation against wild type DNMT3A, impeding homotetramerization (Russler-Germain *et al.*, 2014). DNMT3A methylates periodic CpGs (8 to 10) in a disrupted way, whereas DNMT3B methylates CpGs in a processive manner (Gowher and Jeltsch, 2001; Jeltsch and Jurkowska, 2016). Nevertheless, the exact differences between DNMT3A and DNMT3B are

still unclear as no similar heterotetramer or oligomers structures for DNMT3B have been resolved to date.

1.2.2 DNMT1

When cells divide the newly synthesized DNA strand is unmethylated. The process of copying the methylation pattern from the mother strand to the newly synthesized strand ensures methylation inheritance and is accomplished mainly by the maintenance methyltransferase DNMT1. It preferentially methylates hemi-methylated CpGs in *in vitro* assays (Bestor, Ingram and Buchanan, 1983; Hermann, Goyal and Jeltsch, 2004). Mouse and human DNMT1 knockouts show global loss of DNA methylation. DNMT1 loss in mouse ESCs reduces methylation globally to around 30% of wild type methylation levels (Li, Bestor and Jaenisch, 1992; Jackson *et al.*, 2004). In human ESCs, loss of DNMT1 causes global hypomethylation, arrests cells in G1, and subsequently triggers cell death (Liao *et al.*, 2015).This effect is due to the loss of the catalytic activity of DNMT1, as the hDNMT1 -/- ESCs can't be rescued by catalytically inactive hDNMT1(Liao *et al.*, 2015).

DNA methylation maintenance is not exclusively performed by DNMT1. DNMT [3A-/-,3B-/-]mESCs lose methylation passively over cell passages and have an estimated mDNMT1 maintenance fidelity of ~97.7-98.7% (55 passages, methylation assessed by nearest-neighbour analysis at the start and end points) (Jackson *et al.*, 2004) and hDNMT1 maintenance fidelity of ~ 99.7 - 99.97% (22 passages, methylation profiled over time using reduced- representation bisulfite sequencing (RRBS)) (Liao *et al.*, 2015). In wild type ES cells, *de novo* methyltransferase enzymes DNMT3B and DNMT3A cooperate with DNMT1 to maintain methylation at certain genomic features (Liang *et al.*, 2002; Liao *et al.*, 2015).

DNMT1 is abundant in actively replicating cells. It localizes to the replicating foci in early S phase and diffuses in the nucleus at the later stages of the cell cycle (Leonhardt *et al.*, 1992; Easwaran *et al.*, 2004). DNMT1 is the largest mammalian methyltransferase enzyme with two functional parts (the N-terminal and the C-terminal) linked through a fixable lysine-glycine repeat $(GK)_n$ linker (Figure 4). The N-terminal part contains DNA methyltransferase- associated protein 1 domain (DMAP1), the PCNA (proliferating cell nuclear antigen)-binding domain (PBD) (Chuang *et al.*, 1997; Egger *et al.*, 2006), replication foci-targeting domain (RFTD) (Leonhardt *et al.*, 1992), CxxC zinc finger domain, and poly bromo adjacent homology 1 and 2 domains (BAH1, and BAH2) (Liu *et al.*, 1998). The C-terminal part contains the methyltransferase catalytic domain. The activity of the catalytic domain is autoinhibited by the RFTD, which sits in the catalytic pocket of DNMT1 in its isolated form, acting as a safety lock to prevent substrate DNA from accessing this pocket (Song *et al.*, 2011; Takeshita *et al.*, 2011).

The activity, specificity, and targeting of DNMT1 maintenance is accomplished through several intermolecular protein-protein binding and intramolecular protein interactions. These interactions cooperate to ensure high methylation maintenance fidelity (Jeltsch and Jurkowska, 2016). At early S phase DNMT1 is recruited to the newly synthesized DNA through its PBD domain, which interacts with PCNA at the replication fork (Chuang *et al.*, 1997; Egger *et al.*, 2006) . This PCNA mediated recruitment of DNMT1 to early replicating domains is not strictly needed for methylation maintenance but ensures higher efficiency and fidelity of the process (Spada *et al.*, 2007; Schneider *et al.*, 2013). Moreover, during S phase DNMT1 is tethered to replicating heterochromatin by Ubiquitin-like containing PHD and ring finger domain 1 protein (UHRF1, also known as NP95 in mouse and ICBP90 in human) (Leonhardt *et al.*, 1992; Li, Bestor and Jaenisch, 1992; Easwaran *et al.*, 2004).

UHRF1 is the main regulator of methylation maintenance through targeting and stimulating of DNMT1. UHRF1 knockouts show the same phenotype as DNMT1 knockouts in mESCs (Muto *et al.*, 2002; Bostick *et al.*, 2007). UHRF1 harbors multi-domains that bind different epigenetic marks on the DNA and reinforce its localization to replicating heterochromatin (figure 1.3A). It binds specifically to hemi-methylated DNA by its SET and RING-associated (SRA) domain (Bostick *et al.*, 2007; Arita *et al.*, 2008; Avvakumov *et al.*, 2008; Hashimoto *et al.*, 2008), thus contributing largely to the preferential binding of DNMT1 to hemi-methylated CpGs (30 times greater than unmethylated CpGs) (Yoder *et al.*, 1997; Li, Okano and Xie, 1998). Additionally, it binds the trimethylated histone H3 at Lys 9 (H3K9me3) repressive heterochromatic mark by its tandem Tudor domain (TTD) (Nady et al., 2011; Rothbart et al., 2012). More recently, a study reported another way by which UHRF1 is recruited to the replicating foci (Ferry *et al.*, 2017). The study showed that UHRF1 binds to a replication protein, DNA ligase 1 (LIG1). This binding is mediated by a methylated H3K9 mimic in LIG1, which binds to the hydrophobic pocket in TDD domain of UHRF1. Moreover, it binds the unmethylated histone H3 at Arg 2 (H3R2) through a plant homeodomain (PHD) (Hu *et al.*, 2011; Rajakumara *et al.*, 2011). UHRF1 also mono-ubiquitylates histone 3 at Lys 14,18, and 23. Subsequently DNMT1 binds two mono-ubiquitylated H3 through its ubiquitin interacting motif (UIM) within RFTD (Qin *et al.*, 2015; Ishiyama *et al.*, 2017). This binding induces conformational change in DNMT1 leading to its enzymatic activation.

The direct interaction of DNMT1 with UHRF1 stimulates its methyltransferase activity. This interaction triggers conformational changes in DNMT1 where RFTD is evicted from the catalytic pocket allowing hemi-methylated CpGs to bind (Bashtrykov et al., 2014; Berkyurek et al., 2014). This collaboration between RFTD and UHRF1 ensures that DNMT1 methylates hemi-methylated CpGs specifically. Moreover, UHRF1 mediates (the) ubiquitination of DNMT1 through its E3 ubiquitin ligase activity (Citterio *et al.*, 2004). On the other hand, ubiquitin

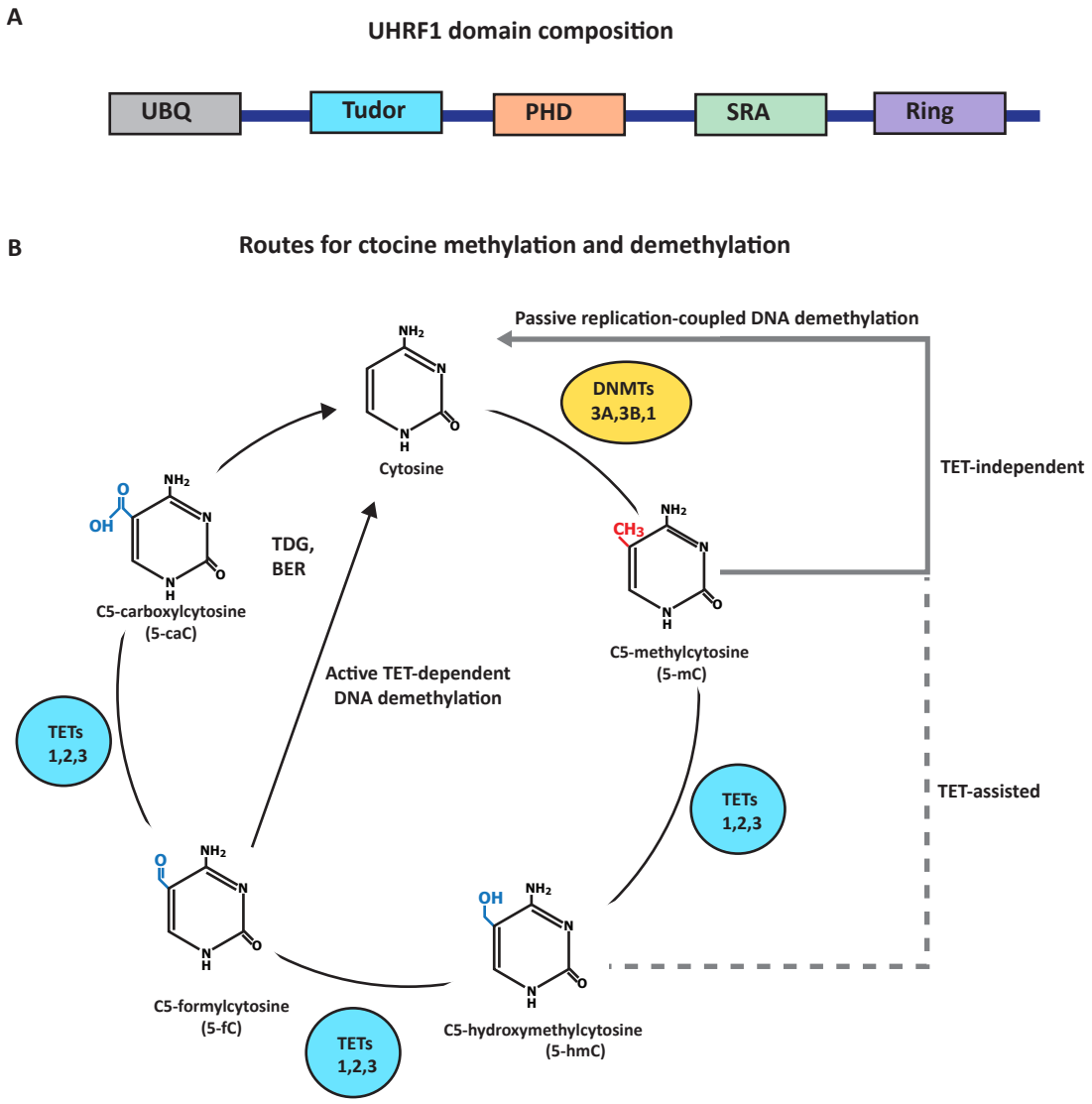


Figure 1.3:
A. UHRF1 protein diferent domains. UHRF1 harbors a ubiquitin-like (Ubl) domain, a tandem Tudor domain (TTD), a plant homeodomain (PHD) followed by a SET and RING-associated (SRA) domain and a really interesting new gene (RING) domain (adapted from Liu et al., 2013).

B. Cytosine methylation and demetylation pathways. In mammals, 5mC is canonically established and maintained by DNMT3 and DNMT1 family, respectively. It can be removed by either passive or active DNA demethylation pathways. Active DNA demethylation is mainly mediated via TET family which oxidizes 5mC through iterative oxidation giving rise to 5hmC and 5fC which can then be repaired to unmodified cytosine by BER (adapted from Kelsey and Feil, 2013).

specific peptidase 7 (Usp7) deubiquitinate DNMT1. Therefore, UHRF1 together with Usp7 regulate the stability of DNMT1 (Du *et al.*, 2010; Qin, Leonhardt and Spada, 2011).

From late S phase until the end of M phase, DNMT1 can also directly bind to heterochromatin through its RFT domain. This association is thought to give DNMT1 more time to methylate the CpG-rich heterochromatic regions (Easwaran *et al.*, 2004; Smets *et al.*, 2017).

1.3 DNA demethylation

As well as having DNA methylation writers and maintainers, cells harbor pathways antagonistic to DNA methylation. DNA demethylation is the process by which the methyl group is removed from 5-methyl cytosine to give unmodified cytosine. This can be achieved by one or a combination of passive and active DNA demethylation mechanisms. Passive demethylation is the loss of DNA methylation through dilution due to replication in the absence of DNA methylation maintenance. Active demethylation is the removal of the methyl group by a chemical reaction (Dean, 2016). This ability to erase DNA methylation is necessary to allow a totipotent state during early development as well as to unlock specific genes in response to stimulation at specific time points (Hajkova *et al.*, 2002; Morgan *et al.*, 2005).

Stripping away the methyl group from 5-mC is energetically unfavorable. Therefore removal of 5-mC is achieved by first converting 5-mC to 5hmC in the presence of O2, α-ketoglutarate (α-KG) and Fe(II) (Iyer *et al.*, 2009; Tahiliani *et al.*, 2009). In mammals, this reaction is catalyzed by enzymes called Ten-eleven Translocation (TET) enzymes which were first discovered as part of a MLL-TET1 fusion protein resulting from a common translocation between

chromosome 10 and 11 in leukemia (Lorsbach *et al.*, 2003). The TET family consists of three active enzymes TET1, TET2, and TET3, which are expressed at different stages in development and in different cells. TET enzymes can convert 5-mC to 5hmC, then to 5fC and then to 5caC (figure 1.3B) (Ito *et al.*, 2010; Ko *et al.*, 2010; Pastor *et al.*, 2011; Pastor, Aravind and Rao, 2013). Subsequently, the intermediates 5fC and 5caC can be converted to unmethylated cytosine by base excision repair.

Another pathway leading to active loss of methylation is deamination. This spontaneous reaction occurs in cells at a rate of 8bases/hour. 5-mC can undergo spontaneous deamination producing uracil. This creates a T-G mismatch, which is then repaired by BER generating unmethylated C (Zhang et al., 2012).

1.4 DNA methylation patterns in normal systems

CpG density is variable throughout the genome, where its levels are lower than expected in the bulk genome based on GC content but retained at the expected levels mainly near TSSs. The CpGs near TSSs are unmethylated possibly due to binding of transcription factors and chromatin-modifying factors containing CxxC domain (Blackledge *et al.*, 2010; Thomson *et al.*, 2010). The remaining 70-80% of CpGs distributed in the bulk genome mostly occur in repeats, introns, and intragenic regions, and are mostly methylated (Ehrlich *et al.*, 1982; M A Gama-Sosa *et al.*, 1983; Bird *et al.*, 1985). Methylated cytosine has a high tendency to be mutated to thymine. Consequently, it is thought that throughout evolution this caused the depletion of CpGs from the bulk genome, passively creating CpGs rich DNA stretches around TSS (Coulondre *et al.*, 1978; Bird, 1980). The small stiches of DNA of about 1kb in length with higher GC, and CpG content are called CpG islands (CGIs) (Bird *et al.*, 1985; Cross *et al.*, 1994; Illingworth *et al.*, 2010).

In human and mice, the majority of CGIs coincide with TSSs, where ~ 50% fall within annotated gene promoters and ~ 20 % are novel promoters (Illingworth *et al.*, 2010). It has been proposed that all CGIs might be promoters. CGIs are nucleosome-deficient, as suggested by *in vitro* analysis showing that CGIs have less propensity to assemble into nucleosomes (Ramirez-Carrozzi *et al.*, 2009). This possibly contributes to the creation of transcriptionally permissive environments at CGIs. Compatible with this, CGI promoters lack a TATA box and other promoter elements, and show distributed patterns of transcription initiation where transcription can be initiated from multiple sites within 50-100 bp regions (Reynolds *et al.*, 1984).

Unmethylated CpGs represent a target for proteins harbouring a zinc finger-CxxC (ZF-CxxC) domain, many of which are chromatin-modifying proteins that facilitate transcription initiation (Voo *et al.*, 2000). One of these proteins is CxxC finger protein 1(CFP1) which recruits Setd1 H3K4 methyltransferase leading to deposition of H3K4me3 (Lee and Skalnik, 2005). H3K4me3 is a modification that is enriched at CGIs regardless of gene expression state. Another chromatin modifying protein containing a CxxC domain is the H3K36 demethylase (KDM2A), which is enriched at more than 90% of the CGIs in mESCs. KDM2A has been suggested to cause H3K36me2 depletion at CGIs (Blackledge *et al.*, 2010). This histone mark is considered antagonistic to transcription (Strahl *et al.*, 2002; Youdell *et al.*, 2008; Li *et al.*, 2009).

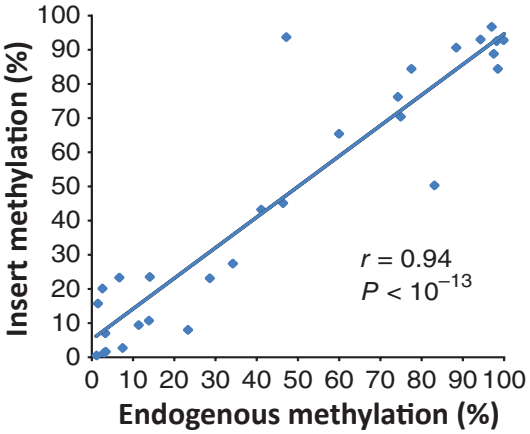
In stem cells, a proportion of repressed genes possess both active H3K4me3 and repressive H3K27me3 on their CGI promoters. H3K27me3 is deposited by polycomb repressive complexes (PRC) 2, which then recruits PRC1 repressor creating a repressive chromatin environment (Guenther et al., 2007; Mikkelsen et al., 2007; Pan et al., 2007; Zhao et al., 2007). H3K4me3 and H3K27me3 are frequently co-localized at CGI promoters of transcription factor, developmental and early differentiation genes. These genes are kept in poised state in ESCs to

be expressed later in differentiation (Boyer *et al.*, 2006; Lee *et al.*, 2006). During differentiation, transcription starts at some of these genes when all the transcription factors required are available, this is followed by eviction of polycomb proteins from CGIs by the new transcripts (Mohn *et al.*, 2008). A smaller fraction of CGI promoters (0.5% of genes) are methylated in ES and differentiated cells, these are enriched for germline associated genes (e.g., *Dazl*, *Tuba3*, and *Piwil*1) (Mohn *et al.*, 2008).

In differentiated cells, most CGI promoters remain unmethylated. However, germline genes, imprinted genes, and pluripotency genes are methylated (Farthing *et al.*, 2008; Meissner *et al.*, 2008; Mohn *et al.*, 2008). The methylation of these genes could stabilize their repression and guard the cellular identity. Tissue-specific genes are activated in the corresponding cell type and their CpG poor promoters become demethylated. The methylation of promoters in normal cells is mostly associated with CpG poor promoters rather than CGI promoters (Mohn *et al.*, 2008).

Factors protecting CGIs from methylation have been widely investigated. Their GC content, CpG density, and transcription factor binding are the main elements rendering CGIs refractory to methylation (Lienert *et al.*, 2011; Krebs, Dessus-Babus, Burger, Schübeler, *et al.*, 2014). Additionally, CGIs are binding sites for proteins which modulate the chromatin (e.g. CFP1 and KDM2A) and help CGIs evade methylation. Together, CGIs are protected from methylation by elements in cis. This view is supported by studies showing that CGIs are capable of establishing the correct methylation state when ectopically integrated in mESCs and through differentiation (Lienert *et al.*, 2011).

A % Methylation CGI fragments at ectopic vs. native loci in mouse stem cells



B

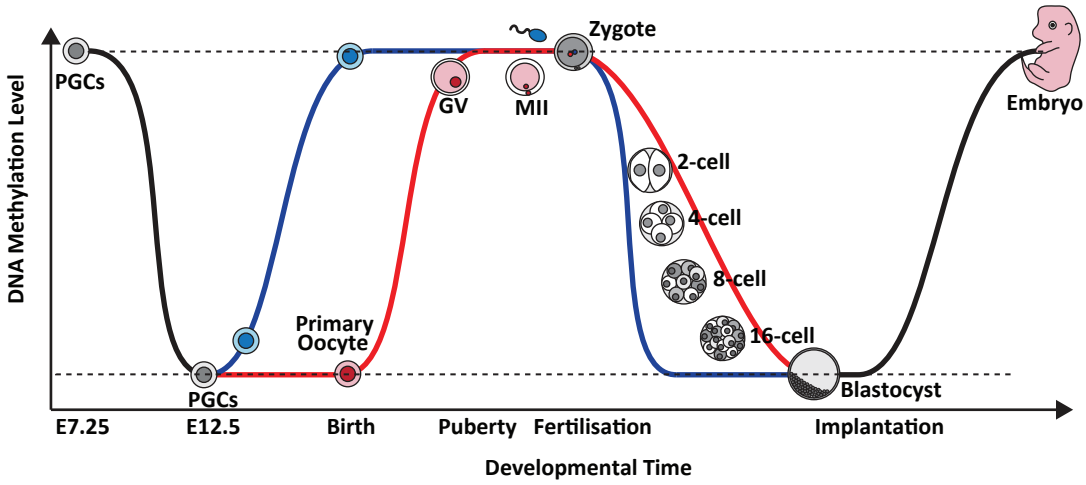


Figure 1.4:
A. CGIs recapitulate their methylation state at ectopic loci in mES cells. The methylation levels of endogenous (X-axis) and ectopically inserted (Y-axis) CGI promoters in ES cells. Methylation levels were determined by bisulphite PCR (taken from Lienert *et al.*, 2011).

B. Schematic representation of methylation dynamics in pre-implantation mouse embryo. The genome of PGCs is demethylated throughout their expansion and migration to the genital ridge (from E7.25 to E12.5). This is followed by a global wave of remethylation from E12.5 to birth in the sperm and from birth to puberty in the oocyte. Upon fertilization, a global wave of demethylation occurs in the zygotic paternal genome followed another one in the maternal genome. Some sequences such as imprinted DMR and repeats resist demethylation (upper dashed lines). Methylation is re-established during and after the implantation stage. Most CGI promoters are unmethylated throughout development.(lower dashed line) (adopted from Smallwood *et al.*,2011).

Other genomic features showing more variable methylation than CGIs and the bulk genome are CGI shores, Low methylation Regions (LMR) at enhancers, and a subset of DNA methylation valleys (DMV). CGI shores are regions flanking CGIs which show high variability in their methylation levels in different cell types (Doi et al., 2009; Irizarry et al., 2009). This variability might be due to their location at the boundary between hypomethylated CGIs and methylated flanking regions. LMRs are shorter than CGIs (median length ~ 250bp), have lower CpG density, and are distal to genes TSS (Stadler et al., 2011). Their methylation is dynamic and changes in a cell type-dependent manner. This is possibly due to differential transcription factor-binding in different cell types. Most of LMRs are thought to be enhancers. DMVs are large genomic regions about 5-68 kb long, showing depletion of methylation in stem cells (Xie et al., 2013). DMVs contain either single genes or cluster of genes that are enriched for transcription factor and developmental regulatory genes and are bivalently marked by H3K4me3 and H3K27me3. During cell differentiation, the majority of DMVs remain hypomethylated and only a small group of DMVs gain methylation (Bernstein et al., 2006; Laurent et al., 2010).

1.5 The dynamics of methylation in mammalian embryo development

Primordial germ cells (PGCs) are germ line precursors that differentiate to sperms in males and oocytes in females. When the mature sperm and oocyte are fused together a fertilized egg termed the zygote is produced. This zygote subsequently gives rise to an embryo. In order for PGCs and zygote to have the potential to give rise to mature germ line cells and all the different tissues in the embryo, parental epigenetic marks need to be reset (Morgan *et al.*, 2005; Hemberger, Dean and Reik, 2009). Reprogramming and differentiation involve many changes in gene expression states that are guided by transcription factors and assisted by epigenetic modifiers (Hemberger, Dean and Reik, 2009; Guo *et*

al., 2017). These changes are accompanied by global waves of hypomethylation followed by remethylation. DNA methylation is thought to act as a barrier to maintain the cellular identity at different stages through development, and to prevent the reversal of this process.

The first wave of hypomethylation takes place in the proliferating PGCs throughout their expansion and migration to the genital ridge, this is thought to occur via a combination of active and passive DNA demethylation (Hajkova *et al.*, 2002; Lee *et al.*, 2002; Yamazaki *et al.*, 2003). Next, remethylation takes place in sperm with maximum methylation reached at birth. Oocytes remain hypomethylated until birth and recover methylation from birth till puberty.

Sperm have higher overall methylation levels but lower number of methylated CGIs compared to mature oocytes (Smallwood *et al.*, 2011). After fertilization, the second wave of hypomethylation takes place in the paternal pronucleus and finishes within four hours before the first cell division, suggesting that active demethylation takes place at this stage (Mayer *et al.*, 2000; Oswald *et al.*, 2000). Following that, the maternal genome loses methylation in a replication dependent manner and with DNMT1 being excluded from the nucleus, suggesting that methylation is lost in a passive manner at this stage (Howell *et al.*, 2001; Hirasawa *et al.*, 2008) (figure 1.4). Some methylated sequences such as certain germline differentially methylated regions (DMRs) in imprinted loci (e.g. Snrpn, Nnat and Plagl1) and few somatic genes resist demethylation, thus passing on these marks from the gametes to the embryo (Borgel *et al.*, 2010).

At the epiblast stage (during and after Implantation), *de novo* methylation takes place at CGI promoters of germline genes (e.g. Dazl, Spo11, and Tex12) possibly by specific factors that recruit DNMT3B to these sequences (Borgel *et al.*, 2010; Auclair *et al.*, 2014). Additionally, tissue-specific genes, which are

usually CpG poor, are repressed and methylated. In E9.5 embryo, lineage specific genes are activated in a cell-type-specific manner and there promoters are consequently demethylated (Borgel *et al.*, 2010). It is worth noting that most CGI promoters maintain unmethylated throughout development.

1.6 Cancer

In healthy adults, cell division and growth are tightly regulated by several pathways to maintain normal tissue architecture. If a normal cell acquires mutations that disrupt these pathways and make it highly proliferative, then it becomes a tumor cell. The ability of tumor cells to hyperproliferate, however, is not enough to render them malignant cells. Usually some tumors will acquire other successive mutations that enable them to spread in nearby tissues and inseminate metastasis and these are termed malignant (Weinberg, 2007; Hanahan and Weinberg, 2011).

Cancer causing mutations can occur in oncogenes, tumor suppressor genes (TSGs) or repair genes, individually or in combination. Oncogenes are genes that when mutated produce a protein that increases the cell proliferation. An example is the proto-oncogene B-Raf (BRAF), where ~40% of melanoma cases have mutations in this gene producing a constitutively active form that contributes to proliferation signaling (Davies and Samuels, 2010). TSGs encode proteins that regulate cell cycle and can induce programmed cell death or cell cycle arrest in abnormally proliferating cells. A very famous example is the Tumor Protein P53 (TP53), which is responsible for sensing DNA damage and inducing one of the following pathways; DNA repair, cell cycle arrest, apoptosis, cell arrest, or senescence (Vogelstein, Lane and Levine, 2000). TP53 is mutated in up to half of all cancers leaving damaged DNA without repair and allowing cells to escape destruction or arrest. Repair genes are responsible for repairing mistakes in DNA, therefore disruption of these genes increases the chances of

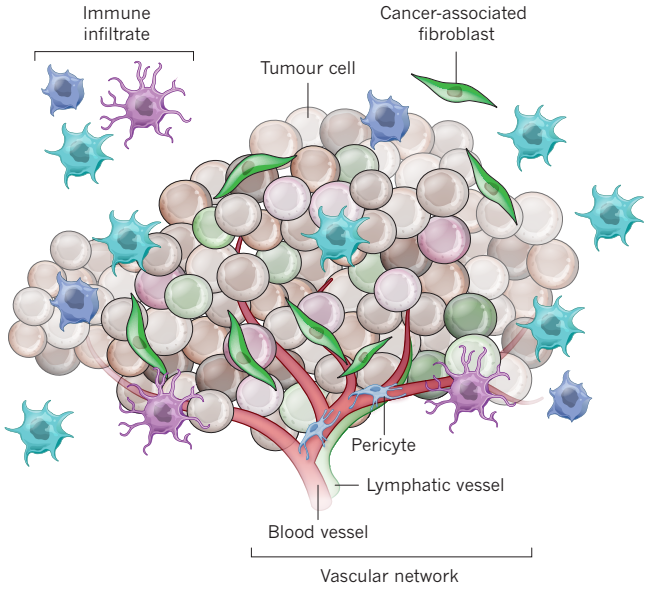
mutations in the cell. An example is the mismatch repair gene MutL Homolog 1 (MLH1), which is mutated in familial sporadic microsatellite unstable colon cancer (CRC) (Herman *et al.*, 1998).

The mixture of mutations a tumor acquires along the path of carcinogenesis gives rise to a heterogeneous population of cancer cells and increases its complexity. A further complication is the ability of solid tumors to recruit surrounding normal cells including fibroblast, myofibroblasts, and macrophages (figure 1.5A). These are termed tumor-associated stroma and play key role in supporting cancer cells ('Hallmarks of cancer: Interactions with the tumor stroma', 2010; Räsänen and Vaheri, 2010). Additionally, cancer cells can induce the formation of vascular network to ensure their access to the circular system in a process termed "angiogenesis" (Turner *et al.*, 2003).

Cancers are classified according to their tissue of origin. Accordingly, the main types of tumors are; carcinoma, sarcoma, hematopoietic and neuroectodermal tumors (Weinberg, 2007). Carcinoma occurs in epithelial tissues and represent ~80-90% of all cancers. Sarcoma forms in mesenchymal tissues and represent ~1% of all tumors. Hematopoietic cancer forms in circulatory and immune system, and neuroectodermal tumors occur in the nervous system.

Carcinoma is responsible for ~80% of tumor-associated death (Weinberg, 2007). It is further classified into subgroups: first, adenocarcinoma arising from secretory epithelial tissues such as breast, colon and ovaries epithelia; second, squamous cell carcinoma arising from non-secretory cells such as skin and lung

A **Tumour cellular heterogeneity**



B **Frequently mutated genes in CRC**

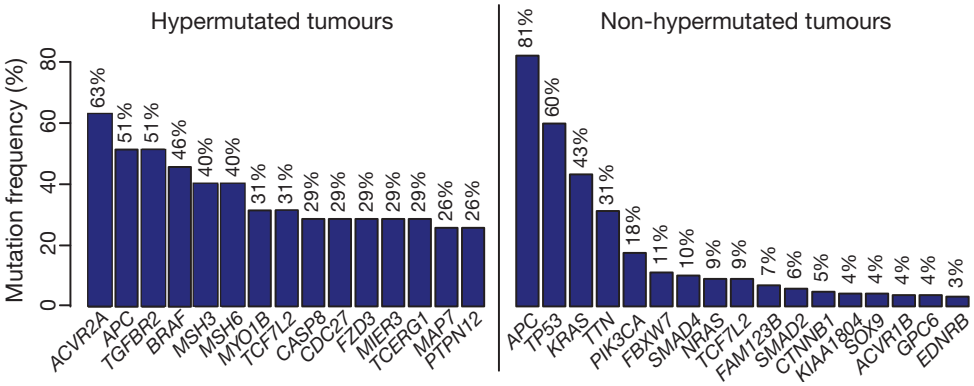


Figure 1.5:
A. The tumor heterogeneity. The tumor-associated stroma contains different types of cells recruited by the tumor. The Immune cells can be either promoting or antagonistic to carcinogenesis. The composition of the tumor-associated stroma changes as tumor progresses (taken from Junttila and de Sauvage, 2013).
B. Most frequent mutations in CRC by class. CRC are either hypermutated or non-hypermutated and each group share distinct frequent causing mutations. Bar graph showing percentage mutation frequency (Y-axis) vs. of the frequently mutated genes (X-axis) in hypermutated tumours (left panel) and non-hypermutated (right panel) CRC (adopted from Muzny *et al.*, 2012).

epithelia; third, other types of carcinoma such as renal cell carcinoma and hepatocellular carcinoma.

1.6.1 Colorectal Cancer

Colorectal cancer (CRC) is an adenocarcinoma of the colon or the rectum. Comprehensive molecular characterization of colorectal cancers showed that ~16% are hypermutated, mostly microsatellite instable (MSI), and possess hypermethylated MLH1 promoter (Muzny *et al.*, 2012). On the other hand, ~84% are non-hypermutated and are microsatellite stable. The two groups have different recurrent mutations, suggesting that each of them evolve through different order of genetic events (figure 1.5B).

A recent collaborative study stratified colorectal cancer based on gene expression into four consensus molecular subtypes. First, CMS1 representing ~14% of CRC cases which are hypermutated, microsatellite instable, associated with BRAF mutation. Second, CMS2 representing 37% of CRC cases and has activated WNT and MYC signaling. Third, CMS3 representing 13% of CRC cases which are epithelial. Fourth, CMS4 representing 23% of CRC and is mesenchymal, have prominent transforming growth factor β (TGF-β) activation, stromal invasion and angiogenesis. CMC1 and CMC4 are associated with worse survival. It remains to be seen whether this stratification will improve therapeutic decisions when translated to patient care.

1.6.2 DNA methylation changes in cancer and CRC

DNA methylation patterns are grossly altered in cancer, which is typified by aberrant hypermethylation of CGIs and loss of methylation from large genomic domains. The extensive hypermethylation of CGIs in cancer have been designated as the CpG island methylator phenotype (CIMP). CIMP was initially

characterized in colorectal cancer where samples were classified as CIMP-positive or CIMP negative according to the methylation status of 7 CGIs (Toyota et al., 1999). Subsequent work based on unsupervised hierarchical clustering using methylation status of 195 CGIs showed that CRC samples can be classified into CIMP-positive, CIMP-negative (Weisenberger et al., 2006). CIMP-positive tumours are associated with BRAF mutation, hypermethylation of the mismatch repair gene MLH1 and microsatellite instability. CIMP-negative do have hypermethylation at CGIs, but to less extent compared to CIMP-positive tumours. This subgroup is enriched for P53 mutations and high somatic copy-number alterations (Witte, Plass and Gerhauser, 2014). Another CIMP subgroup with intermediate methylation levels was stratified and termed CIMP-intermediate (or CIMP-low) which share KRAS mutation and chromosomal instability (Ogino *et al.*, 2007; Toshinori Hinoue *et al.*, 2012; Muzny *et al.*, 2012). CIMP was also characterized individually in different tumour tissues (Maruyama et al., 2001; Bae et al., 2004; Liu et al., 2008; Tanemura et al., 2009). Efforts toward identifying CIMP in cancer type-neutral way resulted in characterizing 89 genes that are can be used to identify CIMP in 5253 tumour samples across 15 solid human epithelial tumour types (Sánchez-Vega et al., 2015). This suggests that CIMP is a pan-cancer phenotype rather that tumour type specific.

The importance of DNA methylation changes in cancer is supported by the frequent mutations of methylation regulating genes. For example, DNMT3A is frequently mutated in AML and some solid tumours and TET2 enzyme is mutated in 6.7% of AML cases (Yan *et al.*, 2011; Gaidzik *et al.*, 2012). These mutations are implicated in changing DNA methylation patterns in cancer (*see section 1.6.3*).

The majority of hypermethylated CGIs in cancer are already silenced in the corresponding pre-cancerous tissue (Sproul *et al.*, 2012). However, hypermethylation of TSGs and DNA repair genes is thought to be particularly

important. There are several lines of evidence to support this idea. First, some of these genes are expressed in the normal tissues and are silenced and methylated in the corresponding cancer tissues, for example the DNA repair genes MLH1 in colorectal cancer (Cunningham *et al.*, 1998) and BRCA1 in breast and ovarian cancer (Fearon, 2000), and the cell-cycle regulator gene RB1 in retinoblastoma (Greger *et al.*, 1989). Additionally, senescence genes such as CDKN2A (p16/ARF) are usually polycomb-repressed in normal cells, and are activated in response to stress subsequently leading to apoptosis. The hypermethylation of CDKN2A (p16/ARF) in many tumour types prevents it's activation in cancer helping cancer cells escape apoptosis (Belinsky *et al.*, 1998; Alhejaily *et al.*, 2014; Zhang *et al.*, 2014). Moreover, targeted methylation of CDKN2A in breast primary cells, using DNMT3A catalytic domain and C-terminal domain of DNMT3L fused to catalytically dead Cas9, enabled cells to escape senescence and induced cancer like expression (Saunderson *et al.*, 2017). Most importantly, CDKN2A remained hypermethylated in the absence of the fusion protein. Finally, in human liver tumours ~12% of tumour specific expression changes are associated with aberrant DNA methylation changes (Acevedo *et al.*, 2008).

From another perspective, it has been suggested that methylation of the repressed promoters might block stochastic reactivation of these genes (Sproul and Meehan, 2013). This reactivation might be required for cancer cells to adapt to new conditions e.g. therapy or metastasizing. By restricting the epigenetic elasticity of cancer cells, methylation of certain genes might restrict tumour progression.

The mechanism underpinning the hypermethylation of CGI promoters is still unclear. Whether different mechanisms underpin the hypermethylation of TSGs and repair genes compared to the genes that are already repressed in normal

tissues, is unclear. Moreover, whether the methylation of TSGs and repair genes is the primary event in their repression is still an open question.

Another disruption in methylation patterns in cancer is 25-30% loss of total methylation (Miguel A. Gama-Sosa *et al.*, 1983). This hypomethylation is mostly confined within large blocks of about 1 MB, termed partially methylated domains (PMDs) (Hansen *et al.*, 2011; Gaidatzis *et al.*, 2014). The methylation of each CpG in these domains is variable and distributed across the range 0-100%, with average methylation 20-60%. These PMDs are a common feature of solid tumors, immortalized cells in culture but not hESC, and human placenta (Hansen *et al.*, 2014). PMDs cover ~ 40% of the genome, they are gene poor, and the majority of the genes within PMDs are repressed (figure 1.6) (Berman *et al.*, 2011; Hansen *et al.*, 2011). The mechanistic underpinnings for the formation of these domains and their contribution to carcinogenesis is still under investigation.

1.6.3 Proposed mechanisms of CGI hypermethylation in cancer

There are several proposed models for aberrant hypermethylation of CGI in cancer. Some of these mainly implicate the dysfunction or deregulation of methylation writers (DNMTs) or erasers (TET proteins) (You and Jones, 2012; Hamidi, Singh and Chen, 2015). For example, DNMT3A is frequently mutated in AML and myelodysplasia. However, the wild type DNMT3A is required for aberrant hypermethylation whereas dysfunction of DNMT3A was associated with impaired CGI hypermethylation (Spencer *et al.*, 2017). On the other hand, upregulation of DNMTs is associated with several cancers (el- Deiry *et al.*, 1991; Issa *et al.*, 1993). Although this apparent overexpression of DNMTs in cancer might be a consequence of increased cell division, it was correlated with aberrant CGI hypermethylation of several genes in colon cancer (Eads *et al.*, 1999; Robertson *et al.*, 2000; Ibrahim *et al.*, 2011).

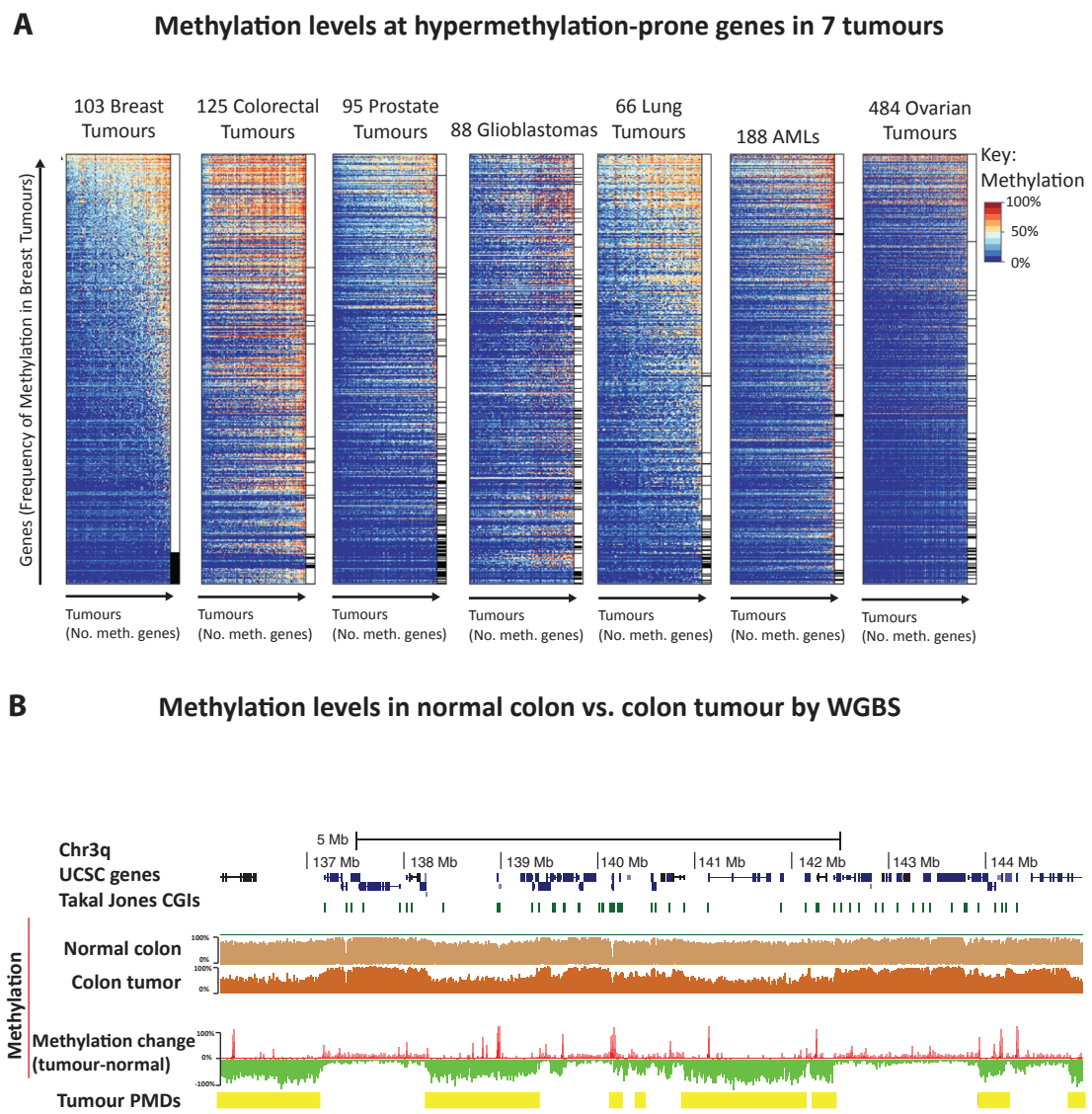


Figure 1.6: Methylation distruption is a hallmark of cancer.

A. Many CGIs gain aberrant hypermethylation in tumours. Shown are heatmaps of methylation levels at hypermethylation-prone genes in 7 tumour types determined by Illumina Infinium. CGIs are ordered according to their frequency of getting methylated in breast tumours and samples are ordered based on number of methylated CGIs (taken from Sproul et al., 2012).

B. Methylation disruptions in cancer are typified by aberrant hypermethylation at CGIs and loss of methylation at PMDs. UCSC Genome Browser plot of a representative genomic region of 8Mb on chr3q showing methylation levels as assayed by WGBS in normal colon (light brown), colon tumour (dark brown), methylation change between tumour and normal colon (gain in methylation coloured red, and loss of methylation coloured green), and PMDs defined in that tumour sample (yellow) (adapted from Berman et al., 2011).

The dysfunction of TET enzymes was proposed as a possible mechanism of CGI hypermethylation in cancer (Jin *et al.*, 2011; Huang *et al.*, 2013). This hypothesis was proposed based on the reduction in 5-hmC levels and frequent downregulation of TET1 in solid tumors (Jin *et al.*, 2011; Yang *et al.*, 2013; Müller *et al.*, 2012). Moreover, mutations impairing the enzymetic activity of TET2 are frequent in hematological malignancies (Figueroa *et al.*, 2010; Quivoron *et al.*, 2011), and mouse models deficient in TET2 recapitulate the hematological phenotype similar to that seen in human (Zhao and Chen, 2013). In a recent study, Thienpont *et al*. showed that hypoxia in cancer decreases TET activity, consequently decreasing 5hmC levels and increasing 5-mC at CGIs in cancer (Thienpont *et al.*, 2016). The group showed that this explains ~48% of CGI's hypermethylation events in tumors.

Aberrant hypermethylation takes place frequently at promoters marked by PRC2-deposited H3K27me3 during development (Ohm, McGarvey, *et al.*, 2007; Schlesinger *et al.*, 2007; Widschwendter *et al.*, 2007). DNA methylation and H3K27me3 are unlikely to occupy the same genomic loci and removal of DNA methylation results in accumulation of H3K27me3 at the previously methylated loci (Meissner *et al.*, 2008; Lister *et al.*, 2009; Hawkins *et al.*, 2010; Murphy *et al.*, 2013). Thus it can be expected that the loss of H3K27me3 mark at CGIs in cancer might lead to a passive accumulation of DNA methylation at polycomb-targeted CGIs. However, there is still no evidence to support this idea (Sproul and Meehan, 2013). It has been shown that upon PCR2 removal in mESC 860 genes gain methylation (Hagarman *et al.*, 2013). However, this group didn't examine whether these genes overlap with aberrantly methylated genes in cancer. It would be interesting to see if more targets gain methylation upon differentiating mESCs lacking PRC2. Additionally, it remains to be determined if the loss of H3K27me3 precedes the gain of aberrant methylation cancer.

Some sequence-specific binding factors have been proposed to play a role in promoting aberrant hypermethylation in cancer. One current model for hypermethylation of CIMP genes in CRC and melanoma, suggests that oncoproteins in these cancers upregulate transcriptional suppressors that recruit DNMTs at these genes (figure 1.7A). In BRAF-positive human CRC and melanoma, active BRAF/MEK/ERK signaling upregulates the transcriptional repressor v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog G (MAFG). MAFG binds CIMP genes and recruits corepressors including DNMT3B. CIMP genes, consequently, become silenced and hypermethylated (Fang *et al.*, 2014, 2016).

In KRAS- positive CRCs, the same group reported that the oncoprotein KRAS causes silencing and hypermethylation of CIMP genes via a similar transcriptional repression mechanism, albeit through a different repressor complex. The oncoprotein KRAS [G13D] upregulate the Serine/Threonine Kinase D1 (PRKD1), and the Ubiquitin Specific Peptidase 28 (USP28). PRKD1 phosphorylates USP28, which then deubiquitinate and stabilize Zinc Finger Protein 304 (ZNF 304). The transcriptional repressor ZNF304 binds to CIMP genes and CDKN2A (INK4-ARF) and recruits other corepressors including DNMT1. This leads to the silencing and hypermethylation of these genes (Serra *et al.*, 2014). Both transcriptional repressors MAFG and ZNF304 are sequence specific binding factors. This suggests that according to this model the hypermethylation of the CIMP genes in CRC and melanoma is encoded in their sequence.

Song *et al*. investigated factors behind the hypermethylation of the tumour suppressor gene glutathione-S-transferase gene (GSTP1) in prostate cancer (Song *et al.*, 2002). They showed that hypermethylation couldn't be recapitulated in cell lines by higher expression of DNMTs, deletion of SP1 binding site from the CGI, deletion of CGI boundaries, or silencing of the gene.

Instead, the combination of gene silencing and low methylation levels, termed seeding, triggered aberrant hypermethylation at *GSTP1*.

These proposed mechanisms might play a role in disrupting the normal methylation pattern in cancer. However, it's still unclear how much each mechanism contribute to the aberrant methylation of CGIs

1.7 Aims of the thesis

CGIs are mostly unmethylated in healthy cells irrespective of the expression state of their genes. Previous studies have shown that CpG density, through recruitment of CpG-binding factors, and transcription factor binding are the main elements rendering CGIs refractory to methylation (Lienert et al., 2011; Krebs, Dessus-Babus, Burger and Schübeler, 2014). This suggests that the methylation status of CGIs is largely programmed by their sequence. Many CGIs gain aberrant hypermethylation in cancer which is associated with transcriptional repression of tumour suppressor genes such as BRCA1, MLH1, and CDKN2A (p16/ARF). However, the mechanism(s) underpinning this process are still unknown. One group proposed that sequence-specific binding factors recruit corepressor complexes and DNMT enzymes to CGIs which drive their aberrant hypermethylation (Serra et al., 2014; Fang et al. 2016). Taken together, these studies imply that the sequence of the CGIs programs not only their normal methylation patterns but also their aberrant hypermethylation in cancer. However, this model has not been robustly tested experimentally. These studies were performed on methylated CGIs i.e. they were addressing what is required to maintain, rather than establish, aberrant methylation.

The aim of my project is to determine whether the sequence of CGIs programs their aberrant hypermethylation in cancer. Therefore, the first stage in the study investigates whether CGIs gain aberrant methylation if they were reintroduced ectopically into the genome of cancer cells. I addressed that by ectopically

integrating representative aberrantly methylated CGIs at several random genomic loci and determined if these CGIs re-gain aberrant methylation (Chapter 3). I also designed and tested a strategy for CGIs targeted integration into a single locus (chapter 4). Next, I asked whether aberrantly methylated CGIs in their native genomic context are primary targets of DNMTs in colorectal cancer cells. I addressed that by i) restoring the main *de novo* DNMT in a hypomethylated colorectal cancer cell line and asked which CGIs gain methylation (chapter 5) and ii) pharmaceutically hypomethylating a colorectal cancer cell line and assessing methylation recovery rate at CGIs (chapter 6).

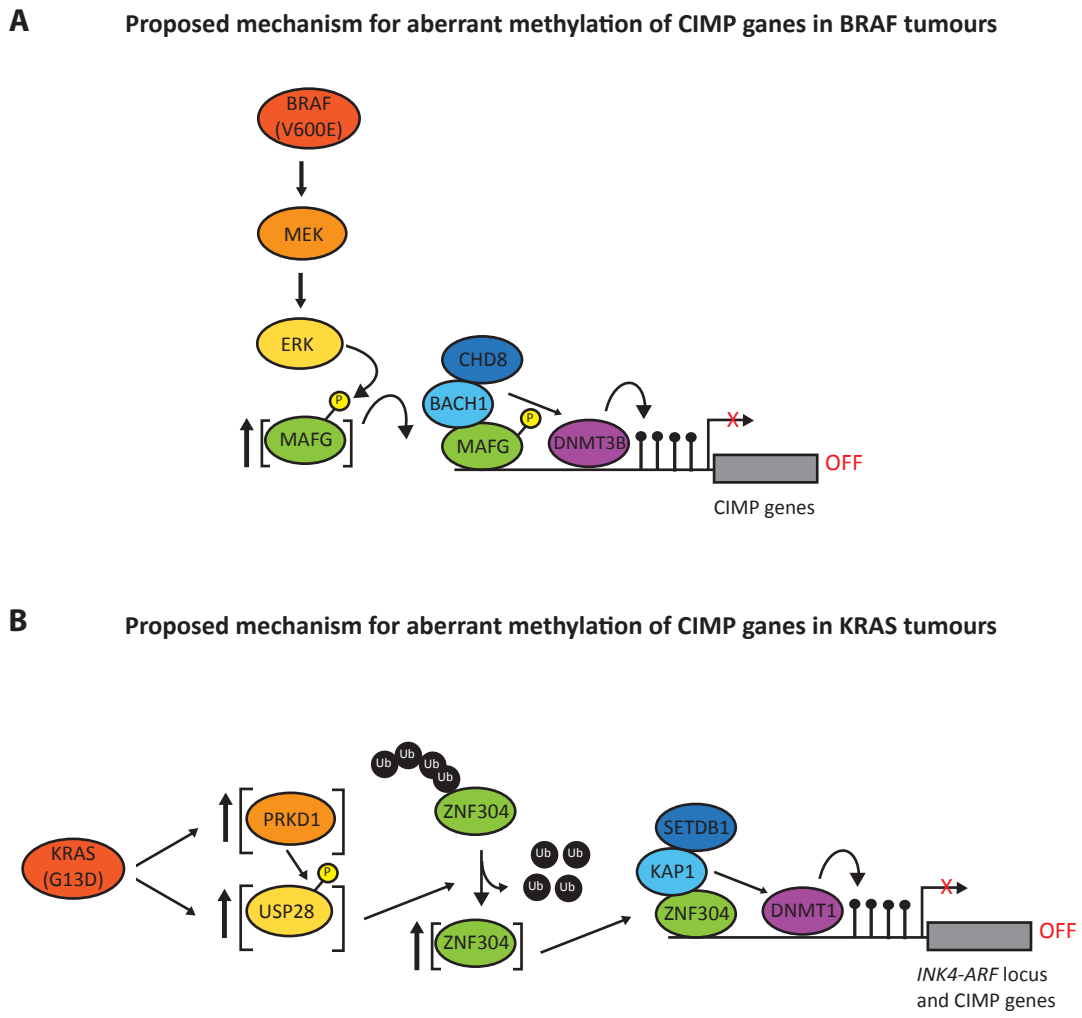


Figure 1.7: Schematic representation of proposed mechanisms of CGI hypermethylation of CIMP genes in BRAF and KRAS tumors.
A. In BRAF CRC and melanoma, the increase in BRAF/MEK/ERK signalling leads to the phosphorelation and stabilization of MAFG, which, in turn, binds and recruits a repressive complex containing BACH1, CHD8, and DNMT3B to CIMP genes. This leads to aberant hypermethylation and repression of CIMP and HIM CGI promoters (adapted from Fang *et al.*, 2016).

B. In KRAS CRC, the activated oncoprotein upregulates the deubiquinase USP28 transcriptionally and through upregulating the kinase PRKD1. USP28 stabilizes ZNF304 through its deubiquitinating. ZNF304 binds CIMP promoters and recruits silencing complex including KAP1, SETDB1 and DNMT1 leading to methylation associated silencing of these genes(adapted from Serra *et al.*, 2014).

Chapter 2: Materials and Methods

2.1 Reagents, Stock Solutions, and Buffers

Note: All solutions were made using deionised water unless otherwise stated.

Buffered Phenol-Chloroform (Sigma): a 25:24:1 mixture of Phenol, Chloroform, and Iso-Amyl Alcohol saturated with 10mM Tris (pH8.0) and 1mM EDTA.

Genomic Lysis Buffer: was prepared as a 2x stock solution: 300mM NaCl, 1% SDS (w/v) and 20mM EDTA. Samples were mixed at a 1:1 ratio (v/v) for use.

Phosphate Buffered Saline (PBS) (technical services, MRC Human Genetics Unit): was made from tablets purchased from Unipath (Oxford).

Proteinase K: Lyophilized powder was dissolved to a concentration of 20mg/ml (w/v) in 50mM Tris (pH 8.0), 1.5mM Calcium Acetate, and 50% Glycerol (v/v) and was stored at -20ºC.

SDS (technical services, MRC Human Genetics Unit): Was prepared as a 20% (w/v) stock solution.

20XSSC stock (technical services, MRC Human Genetics Unit): 3M NaCl, 0.3M tri-sodium citrate, pH7.

TE (technical services, MRC Human Genetics Unit): 10mM Tris HCl (pH7.6), 0.1mM EDTA.

20xTBE stock (technical services, MRC Human Genetics Unit): was prepared by dissolving 216g Tris Base, and 110g Boric Acid in 80ml of 0.5M EDTA (pH8.0) and 920ml of water.

5.0 M NaOH (technical services, MRC Human Genetics Unit): was made by dissolving 200.0g Sodium Hydroxide and 292.2g Sodium Chloride in 1 litre of water.

Penicillin/Streptomycin (technical services, MRC Human Genetics Unit): was made by dissolving 7.0g Penicillin and 13.0g Streptomycin in 1 litre of water.

Freezing Media (technical services, MRC Human Genetics Unit): was made by adding 480.0ml of f10 (Life Technologies) to 400ml of TPB and 120ml DMSO.

L-Broth (technical services, MRC Human Genetics Unit): was made by adding 10g Tryptone, 5g Yeast extract, 10g Sodium Chloride, and 1.0g Glucose to 1 litre of water.

2.2 Mammalian Cell Culture

2.2.1 Cell lines used

Human Colon cancer cell lines; HCT116, DKO (Dnmt1−/−, Dnmt3b−/−), 1KO (Dnmt1−/−), and 3BKO (Dnmt3b−/−) knockouts were obtained from R. Meehan's lab, MRC-HGU. Embryonic kidney cell line derived from HEK293 (HEK293T) was obtained from J. Caceres lab, MRC-HGU.

2.2.2 Thawing Cells from Storage

The cryovial containing the frozen cells was retrieved from liquid nitrogen and quickly thawed by placing it into a 37°C water bath with swirling for ~ 1 minute. In laminar flow hood, the cells were transferred to a 9 ml pre-warmed medium and centrifuged at 300× g for 5 minutes. Next, the cell pellet was resuspended in fresh medium, and plated in a T75 flask.

2.2.3 Maintenance of cell lines

HCT116 cells were cultured in McCoy's 5A (gibco) supplemented with 10% Fetal Calf Serum (FCS, Life technologies) and 1% Penicillin Streptomycin (Pen strep, technical services). HEK293T were cultured in Dulbecco's Modified Eagle Medium (DMEM, gibco), supplemented with 10% FCS and 1% Pen strep. All cells were cultured at 37ºC, 5% CO_2. Cells were passaged by first washing them with Phosphate Buffered Saline (PBS, technical services) to remove any residual medium. Next, cells were treated with 1X TrypLE Express (gibco) - 0.53 mM EDTA solution (Sigma) and placed at 37°C for 1-2 minutes to facilitate their detaching. Cells were subcultivated at 1:3 to 1:5 ratio.

2.2.4 Cell transfection

For HCT116: Per well in 6-well plate, 2µg of plasmid DNA in 100µl of Opti-MEM medium were mixed by vortexing. For a 3:1 FuGENE- HD Transfection Reagent: DNA ratio, 6µl of FuGENE HD was added directly to the Opti-MEM, and mixed immediately. The transfection mixture was incubated for 15 minutes at room temperature, and then added to $8X10^5$ cell/ml cell suspension. The cells and transfection mixture were mixed by pipetting. In order to spread the cells homogeneously, 1ml penstrip free medium was added to each well, and then cells were added on top of that. Cells were incubated at 37ºC. The medium was refreshed after 5 hours or next day. After 24–48 hours, transfection efficiency was checked under the microscope or by Fluorescence activated cell sorting (FACS). Typically, I observed that HCT116 transfection efficiency is >50% (figure 2.1).

For HEK293T: Per well in 6-well plate, 1.5µg of plasmid DNA was diluted in 150µl of Opti-MEM medium and mixed by vortexing. 6µl of Lipofectamin2000 (for a 4:1 Lipofectamin2000 transfection reagent: DNA ratio) was diluted in 150µl of Opti-MEM medium and mixed. The dilutions were incubated for 5 minutes at room temperature. Diluted DNA and Lipofectamin2000 were mixed. The transfection mixture was incubated for 15 minutes at room temperature, then added to the wells and swirled. The cells were added on the top of the transfection mixture and swirled. Cells were incubated at 37ºC and the medium was refreshed after 24 hours.

2.2.5 Cell selection

HCT116 cells transfected with Puromycin resistant gene for stable integration were selected with 2µg/ml Puromycin Dihydrochloride (Thermo Scientific) for 48hrs. If single clones are required, 500-1000 cells were seeded in 10cm cell

culture dishes under 1µg/ml Puromycin Dihydrochloride Puromycin for 8-10 days.

To determine the suitable concentration for Ganciclovir (VWR) selection, I treated HCT116 (thymidine kinase negative cells) and 2E8 cell line (thymidine kinase positive cells) with the following concentrations (0,0.5,1,1.5,2,2.5,3,3.5,4,4.5,5,5.5,6,6.5,10 µg/ml) of Ganciclovir. I found that 2µg/ml was the minimum concentration needed to kill thymidine kinase positive cells within, typically within 6 days, while HCT116 cells look healthy.

Hygromycin selection was performed at 250 µg/ml (Invitrogen), Neomycin positive expressing cells were selected with 500 2µg/ml G148 (Sigma-Aldrich) until all cells in untransfected control are dead. Cells expressing a florescent reporter gene were separated using a fluorescence-activated cell sorting (FACS) at technical services, MRC Human Genetic Unit by Elisabeth Freyer.

2.2.6 DNMT3B rescue in DKO cells

DNMT3B2 and DNMT3B2 CD were reintroduced in DKO8 cells using the PiggyBac transposon. Per well in 6-well plate, 2µg of DNMT3B2 or DNMT3B2-cd piggyBac plasmid and 1.5 µg of mPB transposase (gift from W. Akhtar, NKI) were mixed by vortexing in 100µl of Opti-MEM medium (Gibco). FuGENE HD (Promega) was added directly to Opti-MEM in a 4:1 FuGENE- HD Transfection Reagent: DNA ratio and mixed immediately. The transfection mixture was incubated for 15 minutes at room temperature, and then added to 8x10^5 DKO8 cells. The cells and transfection mixture were mixed by pipetting. Cells were plated in 6-well plate and incubated at 37ºC. The medium was refreshed next day. After 48 hours, cells stably expressing DNMT3B2 or DNMT3B2 CD stably were selected with 1.5 µg/ml puromycin (Thermo Scientific). Cells were

expanded in the presence of puromycin for 20 days. Next, cells were harvested and gDNA was extracted.

2.2.7 5-Aza-2'-deoxycytidine treatment

Cells were plated at ~20% confluent. Next day cells were treated with freshly prepared 1μM 5-Aza-2′-deoxycytidine (Sigma Aldrich). After 24 hours the drug was removed, cells were washed twice with PBS and supplemented with drug free media. For 48/72hrs treatment, the drug was refreshed every 24hrs.

2.3 DNA extraction and manipulation

2.3.1 DNA isolation from cells

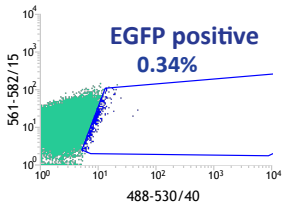
Cells were trypsinised and washed twice with PBS, all centrifugation are done at 1,200 rpm and room temperature for 4 minutes. Per well in 6-well plate, cell pellet was resuspend in 250μl PBS and mixed with an equal amount of 2xGLB to lyse cells. In order to break gDNA, the lysis mixture was syringed up and down using 2ml syringe and green needle (21G) 10-20 times. 20μl of pK (20mg/ml, Roche) were added and incubated at 55ºC overnight. Alternatively, they were incubated for three hours with interval shaking on thermomixer.

After that, 500μl buffered-phenol chloroform (BPC, pH8.0, Sigma) was added to the cell lysis, mixed, and centrifuged for 5 minutes. The upper layer was carefully taken, avoiding the white precipitate in the middle, and transferred to a new tube. BPC extraction was repeated once for crude DNA preparation. To remove remaining phenol, extraction was repeat once using pure chloroform. Next, 10μl of RNase A/T1 Cocktail (Ambion AM2286) were added and samples were incubated at 37ºC for 1hr to remove RNA. One further BPC extraction followed by one further chloroform extraction were done.

FACS analysis of HCT116 transfection effeciency

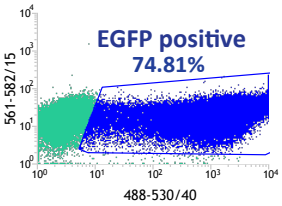
A

Untransfected HCT116

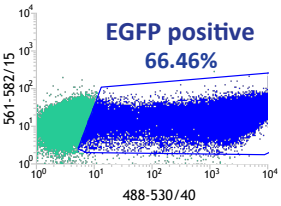


B

HCT116 transfected with vector expressing EGFP/ replicate #1



HCT116 transfected with vector expressing EGFP/ replicate #2



HCT116 transfected with vector expressing EGFP/ replicate #3

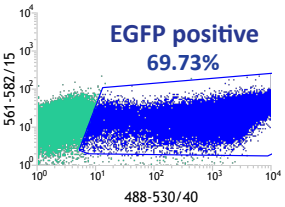


Figure 2.1: HCT116 cells are highly tranfectable.
FACS analysis plots showing % EGFP positive HCT116 cells (coloured blue) in a population of untransfected HCT116 (A), and three replicates of HCT116 transfected with an EGFP expressing construct (B). The plots show the cells counts at different wavelengths. The x-axis and the y-axis are labelled with the laser wavelengths used for excitation and detection bandbass filters. The laser wavelength used for excitation is 488nm(blue) and 561 (red fluorescence (RFP)). The detection bandpass filter is 530/40nm (for GFP) and 582/15 (RFP).

To precipitate DNA, 1/10th volume of NaAc and 1 volume of isopropanol were added and mixed by shaking. Samples were incubated for 5 minutes, centrifuged at 13,000rpm for 5-15 minutes at RT. Supernatant was removed and DNA pellet was washed with 70% ethanol (v/v) and left to dry with lid off for ~5mins at RT. Dry DNA pellet was resuspended in 25µl of 10mM Tris pH8 (or TE). DNA was quantified by Nanodrop and Qubit.

2.3.2 Resolution of DNA on Agarose Gels

Agarose gel with suitable percentage, 0.7%-2% depending on the size of DNA fragment visualised, was prepared by mixing Agarose (Life technologies) with 1X TBE buffer and microwaving it for 2-3min. Gels were stained after running the gel by incubating the gel in 1:20,000 ethidium bromide in TBE buffer. Alternatively, the gels were pre-stained by adding 1:20,000 ethidium bromide while making the gel. Purple loading dye 6X (NEB) was added to the DNA samples. 500ng of 1kb plus (NEB), 1kb (NEB) or 100bp (NEB) were used as markers, depending on the size of DNA visualised.

2.3.3 Gel purification of DNA

The DNA were run on the gel until good separation between the desired and undesired DNA was obtained. The DNA was visualised using UV. The desired DNA fragments were excised using clean razorblade. The DNA was purified using the QIAQuick Gel Extraction Kit (QIAGen) following manufacturers protocol.

2.3.4 Assessing quality and quantity of nucleic acids

Nanodrop spectrophotometer (Nanodrop ND-1000, Thermo Scientific) was used to quantify gDNA. Absorption at 260nm was used to measure DNA

concentration. 260/280 and 260/230 ratios were used as an indication of DNA quality. For a more accurate DNA concentration measurement, Qubit fluormeter (Invitrogen) with Qubit broad range or Qubit High sensitivity DNA reagents were used.

2.3.5 Southern blotting

To visualise DNA methylation levels at human satellite II by southern analysis, 1-3 µg DNA was digested with methylation sensitive enzyme (BstBI, NEB) at 65 ºC overnight in a 40 µl reaction. The DNA fragment were separated according to size by gel electrophoresis. Low percentage agarose gel (0.8% in TBE buffer) was used to enhance the DNA transfer. To obtain a better separation of DNA, 40 Volts were initially applied for 15 minutes, and then increased to 80 volts until the purple loading dye had migrated ¾ length of the gel. To visualise the DNA, the gel was stained with ethidium Bromide for 15 minutes and quickly photographed under UV. To denature the DNA, the gel was submerged in denaturation solution (0.5M NaOH, 1.5M NaCl) twice for 15 minutes at room temperature and then washed with ddH_2O. Next, the gel was submerged in neutralisation solution (0.5 Tris Ph7.5. 1.5M NaCl) twice for 15 minutes at room temperature. Afterward, the gel was equilibrated for at least 10 minutes with 20X SSC. DNA was transferred to a positively charged membrane using an upward-transfer method.

Next day, the DNA was fixed on the membrane by subjecting it to UV irradiation at 0.15J. The probe labelling, hybridization and detection was done using Amersham AlkPhos Direct Labelling and Detection system (GE Healthcare-Life Sciences) according to the manufacturer protocol. Briefly, the membrane was pre-hybridized by incubating with hybridization buffer supplemented with 0.5 M NaCl and 4% blocking reagent at 55 ºC with rotation for 30 minutes after which the labelled probe for satellite II was added to the mix at final concentration 5

ng/ml for overnight incubation. Next, the membrane was washed with pre-warmed primary wash buffer twice for 10 minutes at 55 ºC and with secondary wash buffer twice for 5 minutes at room temperature with rocking at 20 rpm. The signals were detected with CDP-Star and membrane was imaged by ImageQuant LAS 4000 (GE Health care Life Sciences) using the chemiluminescent setting.

Southern blotting to determine location and number of RMCE cassette integrations was performed following the above method with some modifications on the amount of input DNA, the restriction enzyme used and the probe. Specifically, 10 µg DNA were digested with AseI (NEB) overnight at 37 ºC and the full Puro-TK DNA fragment (~3.9 Kb long) was used as a probe. In another southern, DNA was digested with AseI and BglII (NEB) overnight at 37ºC.

2.3.6 Bisulphite primer design

MethPrimer software was used to convert the sequence of the target region of interest into bisulfite (BS) converted sequence and to design BS primers. Primers were 20-30bp long and the size of the amplicon was 200-350bp. Subsequently, the designed primers and the theoretical sequence of BS converted target were used as input in Primer3 software to calculate each primer melting temperature.

2.3.7 BS treatment

Genomic DNA was bisulfite converted by Zymo Research EZ DNA Methylation-Gold Kit (Cat No: D5005) with slight modifications. All centrifugation were performed at RT, full speed for 30 sec, followed by discarding flow-through unless otherwise stated. Briefly, DNA was sonicated on Bioruptor for 3 cycles/ 30 secs/ high settings to shear DNA and possibly increase conversion efficiency.

To convert unmethylated cytosines to uracil, 500ng DNA of each sample was diluted to 20µl, added to 130µl prepared conversion reagent, mixed, and briefly centrifuged. Samples were incubated in PCR block at 98ºC for 10 minutes to denature DNA followed by 64ºC for 2.5 hrs for the deamination of unmethylated cytosines. 600µl M-Binding buffer was added to Zymo Spin IC Column, followed by loading the samples, mixing and centrifugation. Columns were washed with 100µl M-Wash Buffer and centrifuged. 200µl M-Desulphonation buffer was added to the column and left at RT for 20 minutes to allow the removal of the sulphite moiety and generate uracil. The columns were washed twice by adding 200µl M-Wash Buffer, followed by centrifugation. Converted DNA was eluted by addition of 10µl of M-Elution buffer, incubation at RT for 5 minutes, and centrifugation for 1min. Elution was repeated with further 10µl to increase the yield. Bisulphite converted DNA was stored at -20ºC.

2.3.8 BS PCR and sequencing

To perform BS PCR, the genomic region of interest was amplified using a 10µM primer mix (0.5µM) and FastStart Taq DNA Polymerase (Roche) according to the manufacturer's instructions. Briefly, a master mix containing; 1/8th of 500ng bisulfite conversion reaction, 1 unit Taq DNA Polymerase, 25mM $MgCl_2$ (1mM), 10mM dNTPs (0.2mM), and 10x Buffer were used in a 25µl. Subsequently, the PCR product of the correct band size was purified, cloned into PGEM-T Easy (Promega) and transformed into Subcloning Efficiency DH5α competent cells (Invitrogen). Next, white clones were picked with clean tips and grown in LB overnight and sent for miniprep and sequencing using T7 primer at technical services. Sequences were analysed and lollipop diagrams representing the methylation state generated by BiQ analyser software (Bock *et al.*, 2005).

2.3.9 Analysis of total CRISPR editing frequency by SURVEYOR nuclease assay

When DNA cleavage is repaired, insertions/deletions (indels) can be introduced during repair. To estimate CRISPR editing frequency, SURVEYOR nuclease digestion was performed (as described by Ran *et al.*, 2013) with slight modifications. Briefly, targeted sites were PCR amplified form heterogeneous CRISPR-targeted cell population. Amplicons were then purified with QIAQuick PCR purification kit. DNA heteroduplex were formed by first denaturing DNA followed by ramping down temperature for gradual DNA annealing. Digestion with the mismatch-specific nuclease SURVEYOR nuclease (IDT) was performed.

Digestion product were run on 1.3% agarose gel and stained with 1:10000 SYBR gold (Life technologies) while covered with aluminium foil to protect it from light. Bands intensities were quantified using ImageJ software and percentage of total editing was calculated according to this formula: (1 – (1 – (b + c / a + b +c))/2) x 100, where “a” and “b” represent the intensity of the cleavage bands and “c” represents the intensity of the DNA substrate.

2.3.10 Analysis of homologous recombination by HindIII restriction digestion

200 ng of PCR purified DNA were digested with HindIII-HF (NEB) at 37 ºC for 15 minutes according the manufacturer’s instructions. Digestion product were run on 1.3% agarose gel and stained with SYBR gold (Life technologies), bands intensities were quantified as described above. The % HR was calculated according to the following formula: (b + c / a + b +c) x 100, where “a”, “b” and “c” are as described above (Lin *et al.*, 2014).

2.3.11 Preparation of samples for mass spectroscopy and data analysis

Genomic DNA was extracted from cells using DNeasy Blood & Tissue Kit according to the manufacturer's protocol with some modifications. The kit co-purify DNA and RNA, therefore RNAse was added to proteinaseK and were incubated for 1 hr at room temperature to remove RNA. At the final step, DNA was eluted with ddH2O instead of AE buffer; this is compatible with the following steps and enables you to concentrate the DNA by vacuum pump if needed. 1µg gDNA was fragmented by denaturing DNA at 95°C for 10 min in 17.5µL water. Next, DNA was digested by overnight incubation at 37°C with T7 DNA polymerase (Thermo Scientific, Cat #EP0081).

The reaction was inactivated by incubating at 75°C for 10 min. Sample were then centrifuged for 45min at > 12,000 × g. The supernatant was then transferred into new tubes and submitted to the mass spectrometry facility at IGMM. For each sample, files were opened using Xcalibur (Thermo Scientific) and the peaks were visualised to make sure a single peak, rather than multiple peaks, is detected for each nucleotide. A custom Xcalibur method (mass spectrometry facility, IGMM) was used to analyse mass spectrometry results. % 5-mC was calculated using the following equation: %5-mC = (Area under 5-mC /Area under guanine)*100%.

2.3.12 RRBS samples preparation

RRBS libraries were prepared according to the Ovation RRBS Methyl-Seq System protocol and was performed by the Edinburgh Clinical Research Facility. Briefly, 200ng of high-quality human genomic DNA was used as input. To enrich for CpG-dense regions, gDNA was digested by a methylation-insensitive restriction enzyme (MspI) that is specific for CpG containing motifs. Fragments with the smaller sizes (40–220 bp) are enriched for CGIs and were selected by Ampure beads. Adaptors were ligated to the digestion product. The adaptors contain 6nt index to identify samples that are sequenced together, followed by

6nt unique molecular identifier for identification of any PCR duplicates. Adapters ligated DNA is bisulphite converted using EpiTect Bisulfite Kit (QIAGEN) and PCR amplified. Samples were sequenced from opposite directions. Data were generated by 75bp paired end sequencing.

2.4 RNA Extraction and Manipulation

2.4.1 RNA isolation and purification

From $5X10^5$ cells, RNA was extracted using RNAeasy plus kit (QIAGEN) according to the manufacturer protocol. All centrifugation were performed at ≥ 8000g at RT as instructed by the protocol to avoid salt precipitation. Briefly, frozen cell pellets were thawed on ice, resuspended in 350µl RLT plus lysis buffer, and homogenised by syringing up and down with 2ml syringe (21G Green Needle). To remove DNA, samples were loaded on gDNA Eliminator spin column, centrifuged and the flow through containing total RNA was kept. RNA was precipitated using an ethanol-containing buffer, loaded into RNA binding column, and centrifuged. RNA was washed on the column twice with 700µl Buffer RW1 and 500µl Buffer RPE, respectively. The membrane was dried by spinning the column for 5 min with the lid open. RNA was then eluted in 30µl elution buffer and stored at -80ºC.

2.4.2 cDNA synthesis

cDNA was synthesized using QuantiTect Reverse Transcription Kit (QIAGEN) according to the manufacturer protocol. Briefly, 1µg RNA was diluted in 12µl H2O, mixed with 2µl gDNA Wipeout buffer, incubated at 92ºC for 2 min, and placed on ice afterwards. In order to synthesize cDNA with high transcript coverage, a master mix containing 1µl RT random Primer Mix,1µl Quantiscript Reverse Transcriptase, and 4µl Quantiscript RT Buffer per reaction was

prepared and distributed in PCR tubes. A master mix without Reverse Transcriptase was also prepared as a negative control to verify that there is no DNA contamination. For first-strand cDNA synthesis total RNA from previous step was added into this mix and incubated at 42°C for 15min. Samples were incubated for 3min at 95°C to deactivate Quantiscript Reverse Transcriptase, and subsequently stored at -20ºC.

2.4.3 qRT-PCR

qRT-PCR were prepared with LightCycler480 SYBR Green Master and run on Roche LightCycler 480. A standard curve for each primer pair was performed. cDNA reactions were diluted 1:5 and 4µl of that was used as input in 20µl qPCR mix with 0.25µM gene-specific primers. Samples were run in triplicates and for each sample, water control, and no reverse transcriptase control. Data were analysed according to the ΔC_T method. The relative expression of each sample calculated by normalising it [2^(C_T Reference-C_T Target)]. The ratio of expression was calculated by dividing target expression over control expression.

DNA was first denatured for 5min at 95°C before entering a cycle (50x) of denaturing at 90C for 10sec, annealing for 7sec at 55-60°C (depending on primers used), extension for 10sec at 72°C and data acquisition. DNA was then gradually heated up by 2.20°C/s from 65 to 95°C for 5sec each and data continuously collected (for melting curve analysis).

2.5 Protein Extraction and Manipulation

2.5.1 Protein isolation and purification

Samples and reagents were kept on ice. All centrifugations were performed at 4ºC. For each sample, $5X10^5$ Cells were trypsinized, and washed with PBS. A

lysis master mix was prepared by mixing 200µl RIPA buffer, 2µl Halt™ protease and phosphatase inhibitor cocktail (Thermo Scientific™ 78441) and 8µl cOmplete™ Protease Inhibitor Cocktail (prepared by dissolving one tablet in 2ml ddH_2O) (Sigma-Aldrich). The inhibitors are added to prevent protein degradation and stabilize phosphorylation. This mixture was added to cell pellets, mixed by pipetting until the pellet was properly dissolved and incubated on ice for 10 minutes. For DNA removal, 1µl of Benzonase® Nuclease (Sigma-Aldrech) was added. Samples were sonicated on the Bioruptor (On/ Off 30 sec/ 30 sec, HIGH setting) to shear chromatin followed by Centrifugation for 10min full speed. Supernatant was transferred to new tubes and protein was quantified using the Pierce BCA kit. Samples were aliquited 30µg each in 20µl total volume, 5µl 4X sample buffer and 2µl reducing agent were added. Samples were denatured on hot block at 70°C for 5min and then stored at -80ºC.

2.5.2 Western blotting

For blotting DNMT3B tagged with T7 30µg and 60µg of each sample were loaded in NuPAGE™ 3-8% Tris-Acetate (invitrogen) and run in a XCell SureLock gel tank (Thermo Scientific) filled by tris-acetate running buffer (invitrogen?) on 100 Volts for ~1:30 hours. Protein ladder Kaleidoscope™ Prestained Protein (Bio-rad) was run alongside the samples. Proteins were transferred using Criterion Blotter (Bio-Rad) according to the manufacteror instructions at 800mA for 60 min. Protein transfer was confirmed by staining the blot with Ponceau s solution (Sigma). The blot was blocked with 10% western blocking reagent (Roche) for 1hr at room temperature with shaking at 25rpm. The primary antibody (T7-Tag D9E1X0XP® Rabbit mAB/ Cell Signalling Technology) was diluted in 5% western blocking reagent in PBS. The membrane was incubated with the primary antibody overnight at 4ºC on a tubes roller.

The membrane was washed three times for 5min with PBST (0.1% tween in PBS). Then goat anti-rabbit IgG (H+L) secondary antibody (HRP conjugate, Invitrogen) was diluted 1:10,000 in 5% western blocking reagent in PBS and incubated with the membrane for 1hr at room temperature with shaking 25rpm. The membrane was washed 3 times with PBST and the proteins were detected using Pierce ECL western blotting substrate (Thermo Scientific). Signals were detected using ImageQuant LAS 4000.

2.6 Bacterial Culture and Manipulation

2.6.1 Transformation of DH5-α E.Coli cells

DH5-α Competent E. coli cells (invitrogen) were thawed on ice for 10min, gently mixed by flicking, and aliquoted into 50µl into tubes on ice. 10 ng DNA was added to the cells, gently mixed by flicking, and placed on ice for 30 minutes to maximise transformation efficiency. Cells were heat shocked by placing the tubes in a water bath at exactly 42°C for exactly 20 seconds. Cells were quickly placed on ice for 2min. 950µl of room temperature LB was added into the mixture and placed at 37°C for 60min with shaking at 250 rpm to allow cells to recover and express the antibiotic resistance protein. Luria-Bertani (LB) broth plates supplemented with 1.5% agar and with the suitable selection marker (Ampicillin 50mg/ml or Kanamycin 50mg/ml) were pre-warmed at 37°C. 100µl and 200µl of each sample mix were spread on the pre-warmed LB plates and incubated at 37°C for overnight growth. If selection of white/blue colonies was required, plates were spread with X-gal (50µg/ml) and left to dry in the hood before spreading the bacteria.

2.6.2 Bacterial growth

To prepare DNA from bacteria, colonies were picked using plastic sterilised tips. Each colony was inoculated in 3-5ml LB medium supplemented with the suitable selective antibiotic in falcon round bottom tubes. Cultures were incubated at 37°C for overnight growth with shaking at 225 rpm.

2.6.3 Plasmid DNA isolation

Plasmids were isolated using Plasmid Mini kit (QIAGEN) according to the manufacturer protocol. Briefly, bacteria culture was harvested, resuspend in 250ul buffer PI and lysed by adding 250ul buffer P2 and inverting the tube 4-6 times until the solution becomes clear. The lysate was neutralised by adding 350ul of buffer N3 followed by mixing. The neutral lysis was centrifuged full speed for 10 minutes. DNA was bound to a column matrix by loading the supernatant into the supplied column and centrifuging. Matrix was washed with washing buffer 1 and 2. DNA was eluted in TE, quantified by Nanodrop, and stored at -20ºC.

2.6.4 PiggyBac constructs preparation

Primers specific for the selected CGI models were designed using Primer3 (Untergasser et al., 2012). In-Fusion Cloning Primer Design Tool was used to design 15bp overhangs that should be added to the CGI primers to facilitate the cloning. pEGFP-N2 plasmid (Clontech) was modified by removing its multiple cloning site (MCS) by BamHI-NheI double digestion, followed by ends blunting by DNA Polymerase I, Large (Klenow) fragment (NEB) and ligation with T4 (Life technologies) creating N2_NO_MCS vector. A new MCS with flanking minimum TRs was commercially synthesized (IDT) and inserted in the modified pEGFP-N2 to create pEGFP-N2-min. CGI models were PCR amplified from the genome

of HCT116 and cloned into pEGFP-N2-min between the minimum TR sites by In_fusion HD cloning (Takarabio) following the manufacturer instructions. The transposase (mPB, from W. Akhtar, NKI) expressing mCherry was co-transfected with the piggyBac transposon into HCT116 (Akhtar *et al.*, 2014) in 1:1 ratio. After 48 hrs, cells were sorted by FACS and expanded for 1 month to dilute out the plasmid. Methylation of CGI models at native and ectopic locations were assessed by BS PCR using primers specific for each set.

2.6.5 CRISPR targeting constructs preparation

pSpCas9(BB)-2A-GFP (PX458) (Addgene plasmid # 48138) and pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene plasmid # 62988) were gifts from Feng Zhang. PX458 and PX459 V2.0 targeting DYRK1, EMX1, CXCR4, and β-globin loci were constructed (as described by Ran et al., 2013). The CRISPR plasmids were validated by double digest with SacI and BbSI (NEB) 1 hour at 37°C and were subsequently validated by sequencing from U6 promoter.

2.6.6 L1-PuroTK-1L and L1-CGI-1L constructs preparation

A double stranded oligo composed of MCS flanked by LoxP and 15bp homology arms to facilitate cloning was synthesized (IDT). This oligo was cloned into N2_NO_MCS vector by In_fusion HD cloning (takarabio) according to manufacturer instructions to create N2-LoxP-MCS vector. CGIs were PCR amplified with primers containing digestion site overhangs (AgeI and PacI). The amplified CGIs and the N2-LoxP-MCS vector were double digested with AgeI-HF (NEB) at 37ºC in CutSmart buffer for 2 hrs, followed by AgeI heat inactivation, then digestion with PacI-HF (NEB) at 37ºC for 1hr. The vector and each CGI were ligated by T4 (Life technologies), and transfected into DH5α. Correct clones were identified by double digest with AgeI-HF and Bam-HI in CutSmart buffer and verified by sequencing.

To create L1-PuroTK-1L construct, a Puromycin-thymidine kinase fusion (Puro-TK) gene was amplified from pLCA.66/2272 vector (Addgene plasmid number #22733, a gift from Mark Magnuson) by Expand Long Template PCR System (Roche) with primers containing 15bp overhangs to facilitate cloning. The amplicon was cloned into downstream CMV promoter in N2_NO_MCS vector by In_fusion HD cloning. A cassette containing LoxP-CMV-PuroTK-LoxP was created by PCR using primers with LoxP overhangs and Expand Long Template PCR System. The amplicon was cloned into PGEM-T Easy. Finally, L1-PuroTK-1L cassette was obtained with 18 and 15 additional bps on both sites by digestion with NotI-HF (NEB) in CutSmart at 37C for 1hr, followed by blunting the ends with DNA Polymerase I, Large (Klenow) fragment (NEB).

2.6.7 Creating HCT116 cell line with L1-PuroTK-1L cassette at β-globin locus.

In order to integrate L1-PuroTK-1L cassette at β-globin locus, $8x10^5$ HCT116 cells were seeded. Next day, cells were transfected with 3μg CRISPR/Cas9 construct targeting β-globin site and 200 ng of the cassette. After 48hrs, cells were plated at low density (104-103cell/ml) in 15cm plates and were treated with puromycin at 2μg/ml to grow clones. Clones were picked and grown in 2 x96-well plates. Clones with L1-PuroTK-1L cassette at β-globin locus were identified by PCR-screening puromycin resistant clones for β-globin-cassette junction. Of the puromycin-resistant clones, 9% had RMCE construct inserted at β-globin locus with both orientations of the cassette represented.

2.6.8 DNMT3B2 construct preparation

This construct was cloned by F. Taglini as the following: DNMT3B2 sequence from pcDNA3/Myc-DNMT3B2 plasmid (Addgene plasmid # 36942, a gift from Arthur Riggs) was cloned into pCG plasmid, downstream the T7 tag with XbaI-

BamHI double digestion (Chen *et al.*, 2005). T7-DNMT3B2 was cloned into pB530A-puroVal2 with BamHI-EcoRI double digestion. pB530A-puroVal2 was created by substituting the copEGFP in pB530A plasmid (System Biosciences) with Puromycin resistance gene.

2.7 Bioinformatics analysis

2.7.1 RRBS data

RRBS libraries from each lane were demultiplexed based on 6nt indices using a custom python script (D. Sproul). The quality of sequences was checked using FastQC. Random 3nt before each adapter added for nucleotide diversity at the PCR step were removed using a python script supplied by Ovation. Sequences were aligned to the human genome (GRCh38-hg38) and their cytosine methylation states determined using Bismark-0.16.3 program which utilizes the BOWTIE algorithm (version 2.2.6) (Krueger and Andrews, 2011). PCR duplicates were filtered out using a script that compares unique molecular identifier added before PCR amplification. Aligned sequences were converted to bedGraph files using Bismark-0.16.3 then further processed using custom scripts (AWK/Python) to produce a file with total reads and methylated read counts for CpGs in each sample. Bedtools (version 2.23.0) package (Quinlan and Hall, 2010) was used to intersect the output file with CGIs locations (Illingworth *et al.*, 2010). R-3.4.3 was subsequently used to calculate the weighted mean methylation of each CGI (R Core Team, 2017).

2.7.2 TCGA data analysis

To identify aberrantly and normally methylated CGIs in CRC, I utilized Illumina Infinium HumanMethylation450 data processed by D. Sproul for 273 primary colorectal tumour samples and 38 normal colon samples from The Cancer

Genome Atlas (TCGA) (Muzny *et al.*, 2012). Beta values were used as a measure of the methylation level at a given CpG probe as derived from the intensity of the methylated and unmethylated allele probes. The data from 450K arrays was normalized to correct for systematic biases between the two types of probes present on these arrays as described (Pabst *et al.*, 2009). Beta values were then filtered to remove unreliable data points based on the detection P value from the Infinium arrays (threshold 0.01) by setting their values to NA. Bedtools package (version 2.23.0) (Quinlan and Hall, 2010) was used to map the probes to their corresponding CGI using published CGI locations (Illingworth *et al.*, 2010). R-3.4.3 program was subsequently used to calculate the mean of probes methylation for each CGI (R Core Team, 2017).

All CGIs showing less than 30% methylation across all normal samples and showing more than 30% methylation in at least 20% of the CRC samples were considered as aberrantly methylated in CRC. All CGIs showing more than 70% methylation in all normal colon and CRC samples, with no change in median methylation between CRC samples and normal samples, were considered as normally methylated CGIs. The thresholds were chosen based on previous work showing probes with beta values <0.3 represents unmethylated regions (Sproul *et al.*, 2011). In cell lines probes with beta values >0.7 represent fully methylated regions. In tumour samples probes with beta values >0.3 we considered as gaining aberrant methylation if their beta values <0.3 in normal samples.

To validate our subset of aberrantly methylated CGIs, I calculated their enrichment in H1 ES ChromHMM annotations (Ernst *et al.*, 2011). ChromHMM annotations for H1 ES were downloaded as BEDfiles using the UCSC table browser. Next, Bedtools package (version 2.23.0) was used to overlap them with our defined set of aberrantly methylated CGIs. Enrichment of aberrantly methylated CGIs (against all CGIs in the analysis) in different ChromHMM features was performed using R-3.4.3.

2.7.3 Heatmaps

Heatmaps were generated using Image function in R-3.4.3 to visualise DNA methylation levels at CGIs (generated from Infinium HumanMethylation450 data) in 273 primary colorectal tumour samples and 38 normal colon samples. CGIs were ordered according to their frequently of methylation and samples were ordered based on number of methylated CGIs.

Table 2.1: Primer sequences for different genes or genomic loci used in bsPCR, qPCR, and RT-qPCR.

6.5	Forward primer	Reverse primer	Purpose
Native CDH7	AAGGAATTTATTTGTATAGATT**Y**GTTAGG	AACTCCTCCTCCAATCACAAC	BS primers
Native CDH7	AGTGGGTTGTGATTGGAGG	CACTCACCTAAAACCTC**R**CTA	BS primers
Native CDH7	GTTGTGATTGGAGGAGGAGTT	CAACAAAAAC**R**CACATCTACC	BS primers
Native CDH13	TTTTGTAAAGTTATTGGTTTTGGTTATTA	CAACCCCTCTTCCCTACCT	BS primers
Native CDH13	AGGTAGGGAAGAGGGGTTG	CCTCTAACTAAATCTTTCTCCAATACA	BS primers
Native CDH13	TTTTTTGTTTTAGGTAGGGAAGAGG	AAAAACCAAAATTACCCCACTTAATATAAA	BS primers
Native BUB1	ATATAAGGTTTAGGATTTTTGTTAGGTTG	ACCTACCTTCTTTTACCCCCT	BS primers
Native BUB1	AGGGGGTAAAAGAAGGTAGGT	CCTAAATAACTAAAACAAAAAAACCAATTC	BS primers
Native GATA4	GGATGAGGATTATAGGAAGGGG	AACCCTACCTACTAAACCTAAAAATTC	BS primers
Native GATA4	GAATTTTTAGGTTTAGTAGGTAGGGTT	AAAAACAAAACAAAAATTAAACAATTTCR	BS primers
Native ZFP42	GGTTTAAAAGGGTAAATGTGATTATATTTA	ATCTAATCAAACTACAACCACCCA	BS primers
Native ZFP42	AGTTGATGGGTGGTTGTAGTT	ACACATTCCAAATAATAAACAACACAA	BS primers
Native DAZL	TGTGGGTTATGGTTGTGGTG	CTTCATCTTTAACTCCTTTAACCACT	BS primers
Native DAZL	AGTGGTTAAAGGAGTTAAAGATGAAG	**R**CCTTCCTAAAACTAAAACACC	BS primers
Native SFRP1	TTTTGTYGTAAATTTTTAGGGATTTT	CACTCCAACCCTACAACCTC	BS primers
Native SFRP1	GAGGTTGTAGGGTTGGAGTG	TCTAATTCTAATAAACCRAACC	BS primers
Native EPHB1	GTTGTTGTTT**Y**GGTTTGGTTT	ACTCCCAAACCTCCTTACCC	BS primers
Native MLH1	GCCTCGTCGACTTCCATCTTGCT	CGAATAACCCCTGCCACGAACGA	BS primers
Native MLH1	TTAGTAATTTATAGAGTTGAGAAATTTGA	ATCATCTCTTTAATAACATTAACTAACC	BS primers
Native CDKN2A	GTTTT**Y**GTTGTAGATTTTTTATTTATTTGGAT	ATCCCTCCAAAAAATTTAAAAAACAAAAT	BS primers
Native CDKN2A	ATTTTGTTTTTTAAATTTTTTGGAGGGAT	CCAATCCTCCTTCCTTACCAA	BS primers
Integrated CDH7	GATAGTTTGCGTAAAATTGACGTATG	TCCTCCAATCACAACCCACT	BS primers
Integrated CDH7	GTTGTGATTGGAGGAGGAGTT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated CDH7	GTTGTGATTGGAGGAGGAGTT	TTAACCCTAAAAAAATAATCATATTATAACRTA	BS primers
Integrated CDH13	TTAATTTTAGAAAGATAGTTTGYGTAAAATTG	CAACCCCTCTTCCCTACCT	BS primers
Integrated CDH13	AGGTAGGGAAGAGGGGTTG	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated CDH13	AGGTAGGGAAGAGGGGTTG	TTAACCCTAAAAAAATAATCATATTATAACRTA	BS primers
Integrated CDH13	GATAGTTTGCGTAAAATTGACGTATG	CCCCAAACCTACCTTCTTTTAC	BS primers

Gene	Forward primer	Reverse primer	Purpose
Integrated ZFP1	GATAGTTTGCGTAAAATTGACGTATG	ATCTAATCAAACTACAACCACCCA	BS primers
Integrated ZFP1	TGGGTGGTTGTAGTTTGATTAGAT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated ZFP1	AGTTGATGGGTGGTTGTAGTT	TTAACCCTAAAAAAAATAATCATATTATAACRTA	BS primers
Integrated GATA4	GATAGTTTGCGTAAAATTGACGTATG	AACCCTACCTACTAAACCTAAAAATTC	BS primers
Integrated GATA4	TTAATTTTAGAAAGATAGTTTGYGTAAAATTG	CTCCCCCAACAAACAAAATCC	BS primers
Integrated GATA4	GAATTTTTAGGTTTAGTAGGTAGGGTT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated GATA4	GGATTTTGTTTGTTGGGGGAG	TTAACCCTAAAAAAAATAATCATATTATAACRTA	BS primers
Integrated EPHB1	GGTTTTGGATTATTTATTATTGTTTTTTTTGG	GATAGTTTGCGTAAAATTGACGTATG	BS primers
Integrated EPHB1	GAGTAAGTTGGTTGTTGGTGT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated EPHB1	GTTGTTGTTT**Y**GGTTTGGTTT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated SFRP1	GATAGTTTGCGTAAAATTGACGTATG	CACTCCAACCCTACAACCTC	BS primers
Integrated SFRP1	GATAGTTTGCGTAAAATTGACGTATG	CCTACAACCTCC**R**AAATCAATAC	BS primers
Integrated SFRP1	GAGGTTGTAGGGTTGGAGTG	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated SFRP1	GTATTGATTTYGGAGGTTGTAGG	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated DAZL	GATAGTTTGCGTAAAATTGACGTATG	CCTCACAACAACCCCAAAAATAA	BS primers
Integrated DAZL	AGTGGTTAAAGGAGTTAAAGATGAAG	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated DAZL	TTATTTTTGGGGTTGTTGTGAGG	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated MLH1	GATAGTTTGCGTAAAATTGACGTATG	TTCAACCAATCACCTCAATACCT	BS primers
Integrated MLH1	GATAGTTTGCGTAAAATTGACGTATG	ATCATCTCTTTAATAACATTAACTAACC	BS primers
Integrated MLH1	AGGTATTGAGGTGATTGGTTGAA	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated MLH1	GGTTAGTTAATGTTATTAAAGAGATGA	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated CDKN2A	GATAGTTTGCGTAAAATTGACGTATG	ATCCCTCCAAAAAATTTAAAAAACAAAAT	BS primers
Integrated CDKN2A	GATAGTTTGCGTAAAATTGACGTATG	AAACAACATAAAACCTTC**R**ACTAACTA	BS primers
Integrated CDKN2A	ATTTTGTTTTTTAAATTTTTTGGAGGGAT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated CDKN2A	TAGTTAGT**Y**GAAGGTTTTATGTTGTTT	TAAAAATAATCATACGTAAAATTAACGCATA	BS primers
Integrated BUB1	GATAGTTTGCGTAAAATTGACGTATG	CCCCAAACCTACCTTCTTTTAC	BS primers

Primer	Primer sequence	Purpose
VWA1_forward	TTGACGCATGGCTAGGCTACCACGTGCAGTTCG	Cloning CGIs in N2-min using Infusion HD cloning
VWA1_reverse	AATTGACGCATGTTAAATTATCACCCACGGACCC	
TNFRSF1A_forward	TTGACGCATGGCTAGCCGCACGAATTCCTTCCAGC	Cloning CGIs in N2-min using Infusion HD cloning
TNFRSF1A_reverse	AATTGACGCATGTTACGCCACCTTCTCTCTTTCAGGG	
CDH7_forward	TTGACGCATGGCTAGTTCGCGTTCTTTTCCAGTAGCCC	Cloning CGIs in N2-min using Infusion HD cloning
CDH7_reverse	AATTGACGCATGTTACGTGTGCATCCGAAAGAACGTG	
CDH13_forward	TTGACGCATGGCTAGGCCTCTACCCAATGCTTTCGTGA	Cloning CGIs in N2-min using Infusion HD cloning
CDH13_reverse	AATTGACGCATGTTAAGCTCTCTCCTCCTCCCGTTAAC	
CDKN2A_forward	TTGACGCATGGCTAGGCCTCCGACCGTAACTATTCGG	Cloning CGIs in N2-min using Infusion HD cloning
CDKN2A_reverse	AATTGACGCATGTTAGAGCCCAGTCCTCCTTCCTTG	
BUB1_forward	TTGACGCATGGCTAGCTAACGAATTATCCAGATTGCTCCA	Cloning CGIs in N2-min using Infusion HD cloning
BUB1_reverse	AATTGACGCATGTTAGGCCAGGTTTCGGTTCAAC	
ZFP42_Forward	TTGACGCATGGCTAGTGATTACACCCACGCGTATTTGT	Cloning CGIs in N2-min using Infusion HD cloning
ZFP42_reverse	AATTGACGCATGTTACATGAAACAAGACTCACCCCTGTT	
EPHB1_forward	TTGACGCATGGCTAGTAGCTAGCAATGTGACACCAGGA	Cloning CGIs in N2-min using Infusion HD cloning
EPHB1_reverse	AATTGACGCATGTTAGCAAAGCAGCCAGAGGAACC	
DAZL_forward	TTGACGCATGGCTAGGCCGAGTTTCACCCACGAGTGAA	Cloning CGIs in N2-min using Infusion HD cloning
DAZL_reverse	AATTGACGCATGTTAGACTGAGGCACCGGACCTGC	
SFRP1_forward	TTGACGCATGGCTAGCGTCCTGCCGCAAACTTCCAG	Cloning CGIs in N2-min using Infusion HD cloning
SFRP1_reverse	AATTGACGCATGTTATCTCCCCTTGTCTCTTTCCTCCT	
MLH1_forward	TTGACGCATGGCTAGGACGAAGAGACCCAGCAACCCAC	Cloning CGIs in N2-min using Infusion HD cloning
MLH1_reverse	AATTGACGCATGTTAAGAAACACACGGTCTGCGGAAAA	
GATA4_forward	TTGACGCATGGCTAGCGGGGAGGGAGAAAGGGAAC	Cloning CGIs in N2-min using Infusion HD cloning
GATA4_reverse	AATTGACGCATGTTAGTTGGGCAATTTCGGTGAAGTGA	
CDH7-TR_forward	TGACGCATGGCTAGTTCGCGT	qPCR
CDH7-TR_reverse	GCCCAGGGAGCCACTGAGAC	
CDH13_forward	GCCATGCAAAACGAGGGAGC	qPCR
CDH13_reverse	CCAGCCAGCCAACTTCCCAA	
Kanamycin_forward	CCCATTCGACCACCAAGCGA	qPCR
Kanamycin_everse	GCCCCTGATGCTCTTCGTCC	
DNMT3B2_forward	CTGCAGGCCTTCTTCACCAGT	RT-qPCR
DNMT3B2_reverse	CCTTTATGCCCAACTCTTTGAGGA	
BRCA2_forward	AGGGCCGTACACTGCTCAAATC	RT-qPCR
BRCA2_reverse	GGCTGAGACAGGTGTGGAAACA	

Table 2.2: Cloning oligo's sequences.

Oligo	Sequence	Purpose
Min-requirement-top oligo	ATGCATTAGTTATTATTAACCCTAGAAAGATAGTCTGCGTAAAATTGAC GCATGGCTAGCACCGGTACTAGTTTAATTAACATGCGTCAATTTTACG CATGATTATCTTTAACGTACGTCACAATATGATTATCTTTCTAGGGTTA ATAATAGTAATCAATT	Creating N2-min vector
Min- requirement -bottom oligo	AATTGATTACTATTATTAACCCTAGAAAGATAATCATATTGTGACGTAC GTTAAAGATAATCATGCGTAAAATTGACGCATGTTAATTAAACTAGTAC CGGTGCTAGCCATGCGTCAATTTTACGCAGACTATCTTTCTAGGGTTA ATAATAACTAATGCAT	Creating N2-min vector
B-globin HDR oligo	GAGGTCATTCTGACCTCAGAACATAGGAGAGATGGATATTGTACACCA CAGAGTCCAGGATGACTAGCAGTGAAGGATTCAAATGCCTGGGCTAG CAAGCTTAAATCTCCTATACAACCGGAGAGACCCAAGTCCCCAAAGAC AAGAAGGGATGATGAAATGGCCTCTGCATGACATGACTTG	HDR template for Beta globin locus
B-globin gRNA top oligo	CACCGTTGTATAGGAGATTTCGCC	Cloning into PX459_V2 plasmid
B-globin gRNA bottom oligo	AAACGGCGAAATCTCCTATACAAC	Cloning into PX459_V2 plasmid

Chapter 3: Ectopically integrated CpG islands are rarely hypermethylated in cancer cells

I hypothesised that if the sequence of CGI programs their aberrant methylation, then CGIs should be able to recapitulate their methylation status when ectopically integrated into the genome regardless of their positions. Integrating a desired sequence can be performed by several techniques. One way is direct integration to a single site by recombinase mediated cassette exchange (Feng *et al.*, 1999). Another way is introducing the desired sequence to the cells on an episomal vector (Inoue *et al.*, 2017). Alternatively, random integration by transposons or retroviruses can be used (de Jong *et al.*, 2014).

In order to test our hypothesis, it is preferable to integrate CGIs at a number of different locations. So, if we were to use direct integration by RMCE, we would need to perform the experiment in two steps. Firstly, establishing a number of cell lines with RMCE cassettes at different loci. Then, using these cell lines to integrate the desired sequence. However, using random integration techniques, a desired sequence is integrated into different genomic locations in the same population in one step.

On the other hand, using an episomal vector to introduce the desired sequence to the cell, will maintain that sequence on the episome and will not allow its integration into different genomic loci. Moreover, episomal vectors are not representative of physiological chromatin due to differences in histone stoichiometry, cooperativity between transcription factors and nucleosome positioning compared to endogenous chromosomes (Hebbar and Archer, 2007, 2008). Importantly, studies comparing episomal integration of enhancers and

genomic integration showed that activities of enhancers integrated into the genome correlate better with their native counterparts than those introduced to the cell on episomal vectors (Inoue *et al.*, 2017). Therefore, performing random integration of CGIs was our method of choice.

Random integrations can be achieved by transposons or retroviruses. Transposons are reportedly less expected to interfere with the chromatin than retroviruses as retroviruses can activate genes located in the vicinity of their integration sites up to many mega bases distant (Uren *et al.*, 2005). DNA transposons have been used widely and successfully to integrate DNA of interest into chromosomal DNA (Ding *et al.*, 2005; Akhtar *et al.*, 2014; Stelzer *et al.*, 2015). They are composed of two specific DNA sequences termed terminal repeats (TRs) flanking the desired sequence. These TRs are recognized by the transposase that is usually co-transfected with the transposon which relocates them into chromosomal DNA (Ding *et al.*, 2005). While both Sleeping Beauty and piggyBac are effective DNA transposon systems, for our experiment piggyBac is preferable to Sleeping Beauty because piggyBac is footprint free (i.e. the DNA sequence is unchanged when piggyBac remobilises from its integration site to another location) (Ding *et al.*, 2005; Li *et al.*, 2011). Most importantly, piggyBac has been utilized to stably integrate CGIs in mESCs (Stelzer *et al.*, 2015). These CGIs were capable of recapitulating the correct methylation status suggesting that piggyBac doesn't interfere or hinder methylation.

3.1 Experimental strategy

To perform CGI random integration we chose to use piggyBac transposons containing our CGIs flanked by inverted TRs. I designed the construct to have a GFP reporter in its backbone. Additionally, I chose an inducible transposase plasmid that encodes mCherry reporter (a kind gift from Dr. Waseem Akhtar,

Netherlands Cancer Institute). The mCherry and GFP reporters will help with enriching for cells that have been transfected. Tamoxifen inducible transposase should help in controlling the time window for transposase expression, which when restricted to 24 hours can decrease the remobilization of the piggyBac and thus simplify the identification of the integration sites. However, in my hands, the transposase was found to be functional in HCT116 cells regardless of the presence of tamoxifen.

After transfecting the cells with these two plasmids, the cells were FACS sorted for the expression of GFP and mCherry (figure 3.1A). FACS-sorted cells were expanded for 4 weeks to minimize the presence of the piggyBac plasmids in the population and their interference with the results. Next, gDNA was extracted and bisulfite-treated. bsPCR was performed using primers specific to the inserted CGIs (the primers amplify from the PiggyBac TRs) to assess their methylation state irrespective of their locations. Moreover, bsPCR was also performed to assess the methylation at the native CGIs.

3.2 The majority of hypermethylated CGIs in colorectal cancer samples are also hypermethylated in HCT116

After choosing piggyBac as our method of integrating CGIs, I next wanted to choose a cell line model for my study. CGI aberrant hypermethylation is a common alteration in all cancer types; hence any tumour type can be used as a model (Witte, Plass and Gerhauser, 2014). However the extent of CGIs hypermethylation vary between different tumours. Colorectal cancer (CRC) shows a widespread aberrant CGI hypermethylation, and is one of the tumours where CGI hypermethylation has been widely investigated. It was the first tumour type in which CIMP was characterised (Toyota *et al.*, 1999; Lao and Grady, 2011). Thus, CRC represents a good model to investigate aberrant CGI hypermethylation.

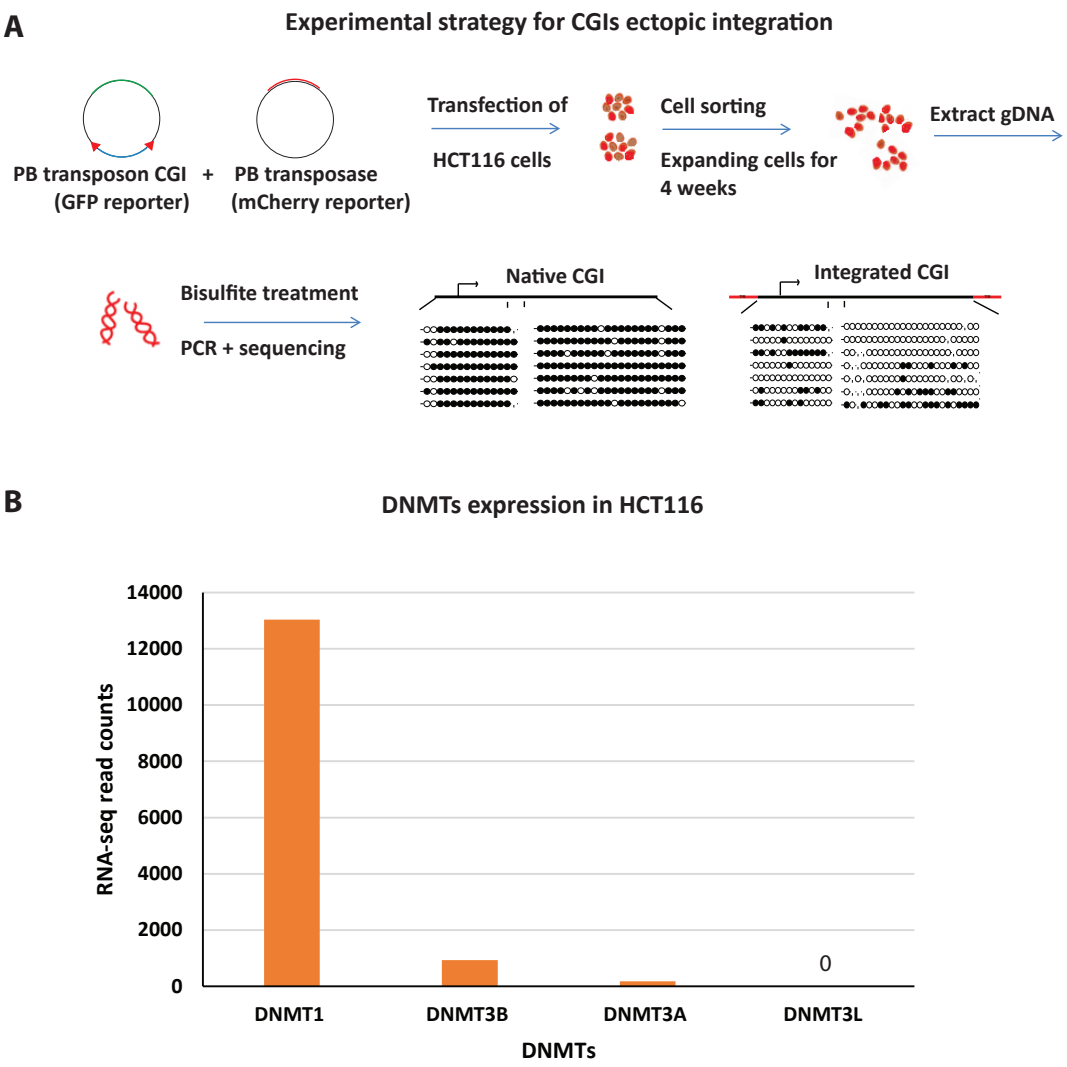


Figure 3.1:

A. Experimental strategy for ectopic integration of CGIs using the piggyBac transposon. HCT116 were transfected with transposase vector having mcherry gene and piggyBac construct having the CGI, 24 hrs after transfection, mCherry positive cells were sorted, seeded and expanded for 4 weeks. Subsequently, gDNA was extracted and the methylation status of integrated CGIs was assessed from the pool of cells by clonal bisulfite sequencing (BS).

B. HCT116 cells express DNMT1, DNMT3B and low amounts of DNMT3A. Expression levels of DNMTs in HCT116 represented as RNA-seq read counts. RNA-seq data were analysed by Catherine Doust.

A well characterised human CRC cell line is HCT116 whose DNA methylome has been studied widely (Schuebel *et al*., 2000; Rhee *et al*., 2002; De Carvalho *et al*., 2012; Roulois *et al*., 2015). Whole genome bisulphite (WGBS) and expression data for this cell line are available in our lab. Additionally, DNMTs knockout cell lines and their WGBS data are also available. Our RNA-seq data shows that HCT116 expresses DNMT1, DNMT3B and low levels of DNMT3A (figure 3.1B). Most importantly, HCT116 is one of the CRC cell lines with *de novo* methylation activity, which is a key player in CGI aberrant hypermethylation as these CGIs are unmethylated in normal cells and gain aberrant *de novo* methylation in cancer (Lengauer, Kinzler and Vogelstein, 1997). Furthermore, my initial characterisation of HCT116 cell line indicated that it is a robust cell line and has transfection efficiency of ≥50% (see, section 2.2.5).

To test the extent to which CGI hypermethylation in HCT116 represents hypermethylation in CRC, I wanted to examine what proportion of hypermethylated CGIs in CRC are also hypermethylated in HCT116. Therefore, I first needed to identify the aberrantly hypermethylated CGIs in CRC. To achieve this, I utilized Illumina Infinium HumanMethylation450 data for 273 primary colorectal tumour samples and 38 normal colon samples from The Cancer Genome Atlas (TCGA) (Muzny *et al.*, 2012).

I defined aberrantly methylated CGIs in CRC as CGIs showing less than 30% methylation across all normal samples and showing more than 30% methylation in at least 20% of the CRC samples as previously published (Sproul *et al.*, 2012). Heatmaps of methylation levels of the CGIs I identified as aberrantly methylated in CRC show their unmethylated state across all normal colon samples, whereas in colorectal tumour samples they gain methylation in at least one fifth of the samples (figure 3.2A).

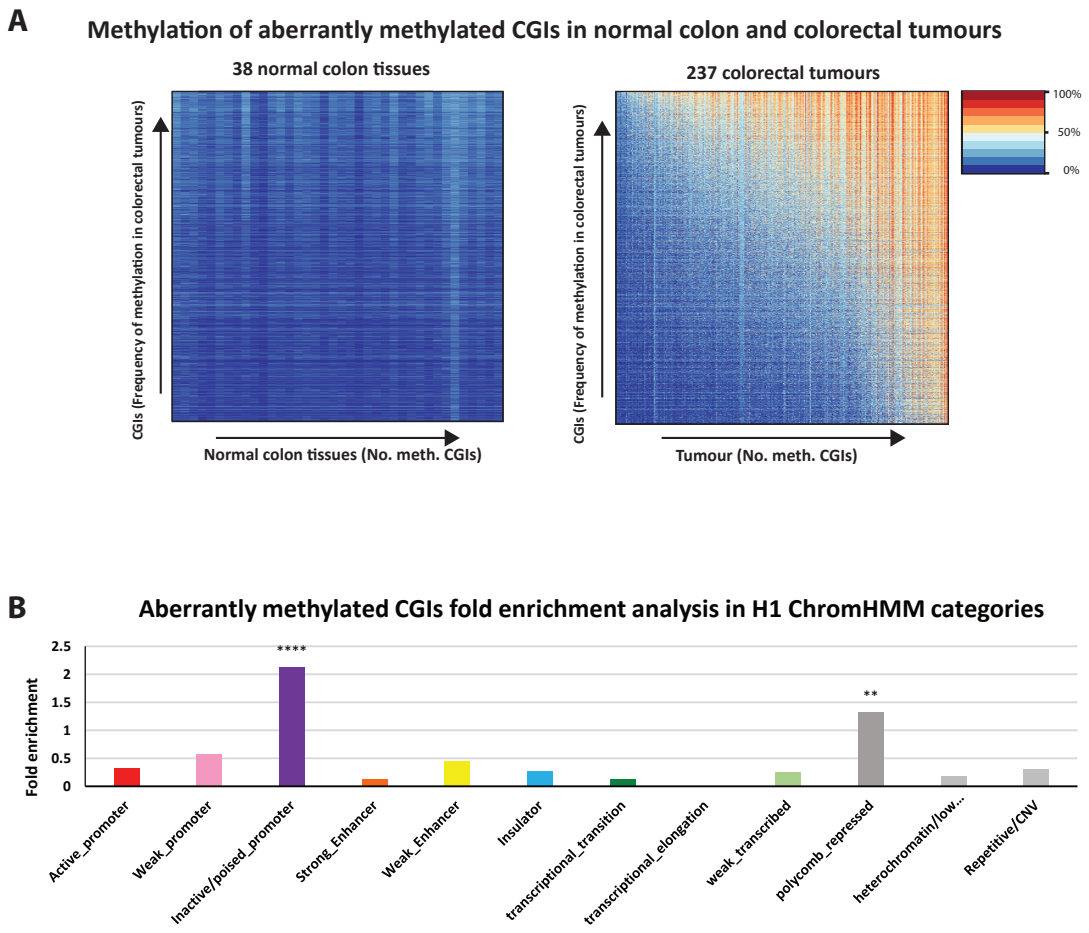


Figure 3.2:

A. Aberrantly methylated CGIs are hypomethylated in normal colon samples (left panel) and hypermethylated in at least 20% of colorectal tumor samples (right panel). Heat maps showing the methylation levels of aberrantly methylated CGIs in 38 normal colon tissues (left panel) and in 273 colorectal tumours samples tissues (right panel) from their Infinium DNA methylation data. CGIs are ordered according to their frequently of methylation in colorectal tumours and samples are ordered based on number of methylated CGIs.

B. Aberrantly methylated CGIs are significantly enriched in inactive and poised promoters and polycomb repressed promoters in H1 cells. Shown is the fold enrichment of aberrantly methylated CGIs in H1 ChromHMM categories.

Values of P<0.0001 are marked with four asterisks (****) and values of P<0.001 are marked with three asterisks (***), Fisher's exact test.

Many of the aberrantly methylated CGIs in cancer are bivalently marked by H3K4me3 and H3K27me3 in human ES cells, a mark characteristic of poised promoters (Ohm, Mcgarvey, *et al.*, 2007). In order to further validate my subset of aberrantly methylated CGIs, I calculated their enrichment in H1 ES ChromHMM annotations (Ernst *et al.*, 2011). These ChromHMM annotations represent segmentation of the genome into different chromatin states (e.g. active promoters, strong/weak enhancers, heterochromatin) generated by integrating nine chromatin marks using a Hidden Markov model. Supporting my approach in identifying aberrantly methylated CGIs, this analysis showed these CGIs were enriched at inactive/poised and polycomb repressed promoters' annotations from H1 ES cells (figure 3.2B).

Among the aberrantly methylated CGIs in CRC ~60% were found highly methylated in HCT116 (their mean methylation by WGBS in HCT116>=70%). This suggested that hypermethylation in HCT116 is a good representative of CGI hypermethylation in CRC, and HCT116 is a suitable cell line for investigating the mechanism through which CGIs become hypermethylated in cancer.

3.3 Selection of representative CGIs for my study

In order to address my study question, I picked CGIs that frequently gain aberrant methylation in CRC, have a potential clinical value as suggested from literature , and have sizes ranging from ~ 500 to 2000bp (in case I needed to integrate the full length CGI). *SFRP1, ZFP42*, *GATA3*, *CDH7*, *CDH13*, *MLH1*, *CDKN2A* and *EPHB1* were picked as models for CGIs that frequently gain aberrant methylation in CRC compared to normal epithelial tissues (figure 3.3A). Among these, *CDKN2A* is a TSGs that have one allele mutated and the other allele methylated in a proportion of colorectal cancer cell lines (Myöhänen, Baylin and Herman, 1998a). *MLH1*, *CDKN2A*, *SFRP1,* and *EPHB1* are

methylation markers for CRC subtype characterization (Serra *et al.*, 2014). *BUB1* is a housekeeping gene that is unmethylated in normal colon epithelial tissues and CRC samples, and therefore was picked as a model for CGIs that are always unmethylated. *DAZL* is methylated in healthy and tumour somatic tissue and was picked as a model for normally methylated CGIs (figure 3.3A).

My analysis of TCGA methylation data shows that the CGIs of *SFRP1* and *ZFP42* genes are methylated in some normal colon epithelial samples (albeit to a lower extent than in the primary colorectal tumour samples), which could be due to methylation associated with greater age (figure 3.3A) (Toyota *et al.*, 1999). Subsequently, I examined this in normal samples and found that the percentage mean methylation of both *SFRP1* and *ZFP42* CGIs correlates with the age of patients when the samples were taken (figure 3.3B).

I next tested the ten CGIs for their methylation status in HCT116 from WGBS data (figure 3.4A), and these findings were confirmed by bsPCR. WGBS and bsPCR showed the CGIs of *SFRP1, ZFP42, GATA3, CDH7*, *CDH13*, *EPHB1*, and *DAZL* are hypermethylated whereas *BUB1* and *MLH1* are unmethylated in HCT116. *CDKN2A* showed percentage mean methylation of 50%, which is supported by previous reports identifying one *CDKN2A* allele mutated and the other allele methylated at the CGI in HCT116 (Myöhänen, Baylin and Herman, 1998b). The methylation state of CGI models in HCT116 is reflective of their methylation state in CRC since all of the methylated CGI models in CRC samples are methylated in HCT116 except MLH1 (figure 3.4B). This is expected since MLH1 is methylated in BRAF-positive CRCs, and HCT116 is BRAF-negative (Fang *et al.*, 2016). However, in HCT116 the methylated CGIs tend to have higher levels of methylation compared to CRC. This is a characteristic of methylation in transformed cell lines, where a 5-fold increase in the levels of hypermethylation has been observed in cancer cell lines compared to their primary corresponding tumours (Smiraglia *et al.*, 2001). This increase in the

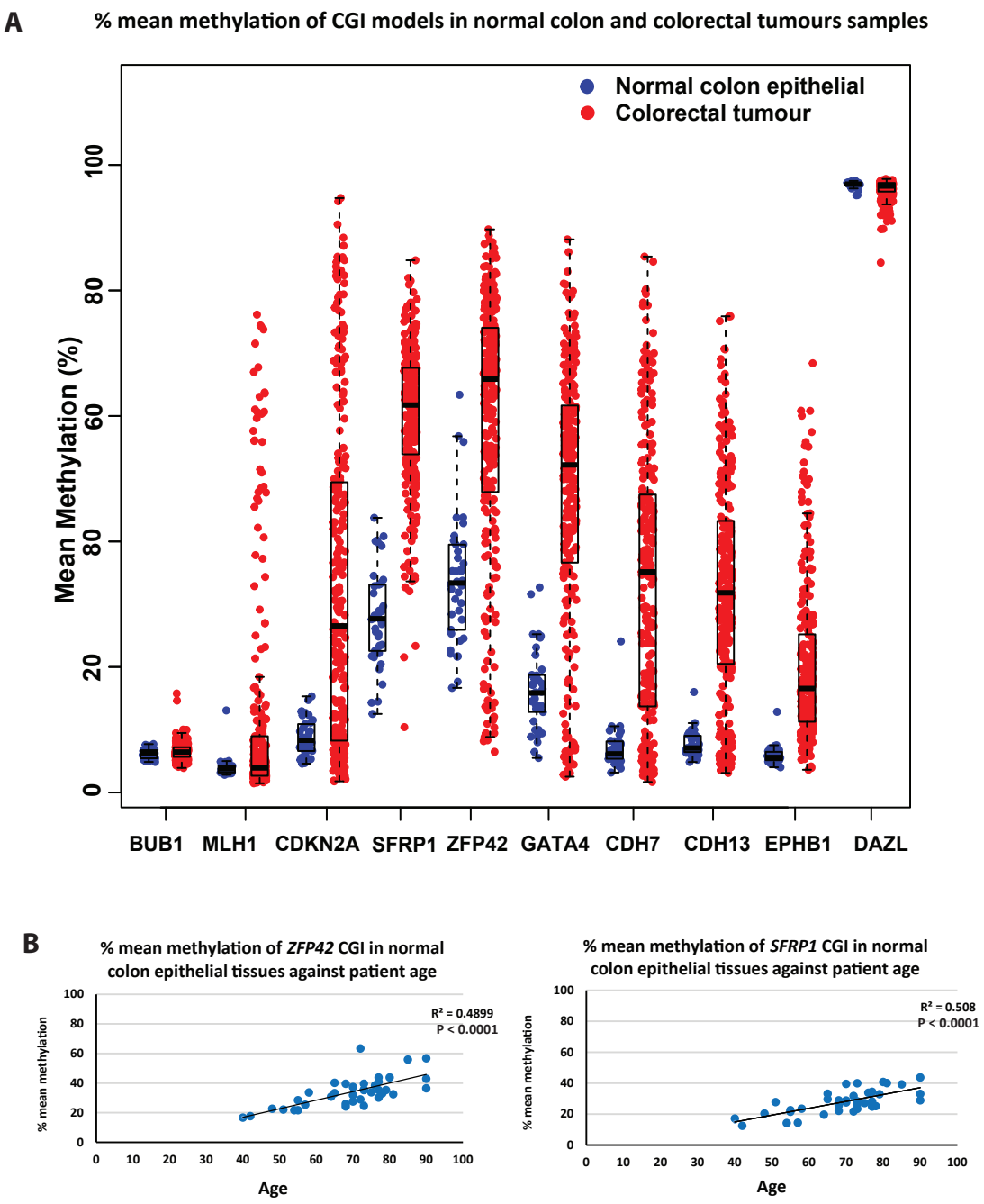


Figure 3.3: CGIs methylation status in primary colorectal tumor and normal colon epithelial samples

A. Distribution of mean methylation of CGI models in normal colon epithelial tissues (blue) and primary colorectal tumor samples (red). Each point represents the % mean methylation of the probes in a CGI for one sample. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the % change in mean methylation at a single CGI. Horizontal line within the box represents the % median methylation of a CGI across samples.

B. *SFRP1* and *ZFP42* CGIs methylation in normal colon tissues is corrolated with patient age. % Mean methylation of *SFRP1* CGI (left panel) and *ZFP42* CGI (right panel) in normal colon epithelial tissues against age of patients at the time the samples were taken.

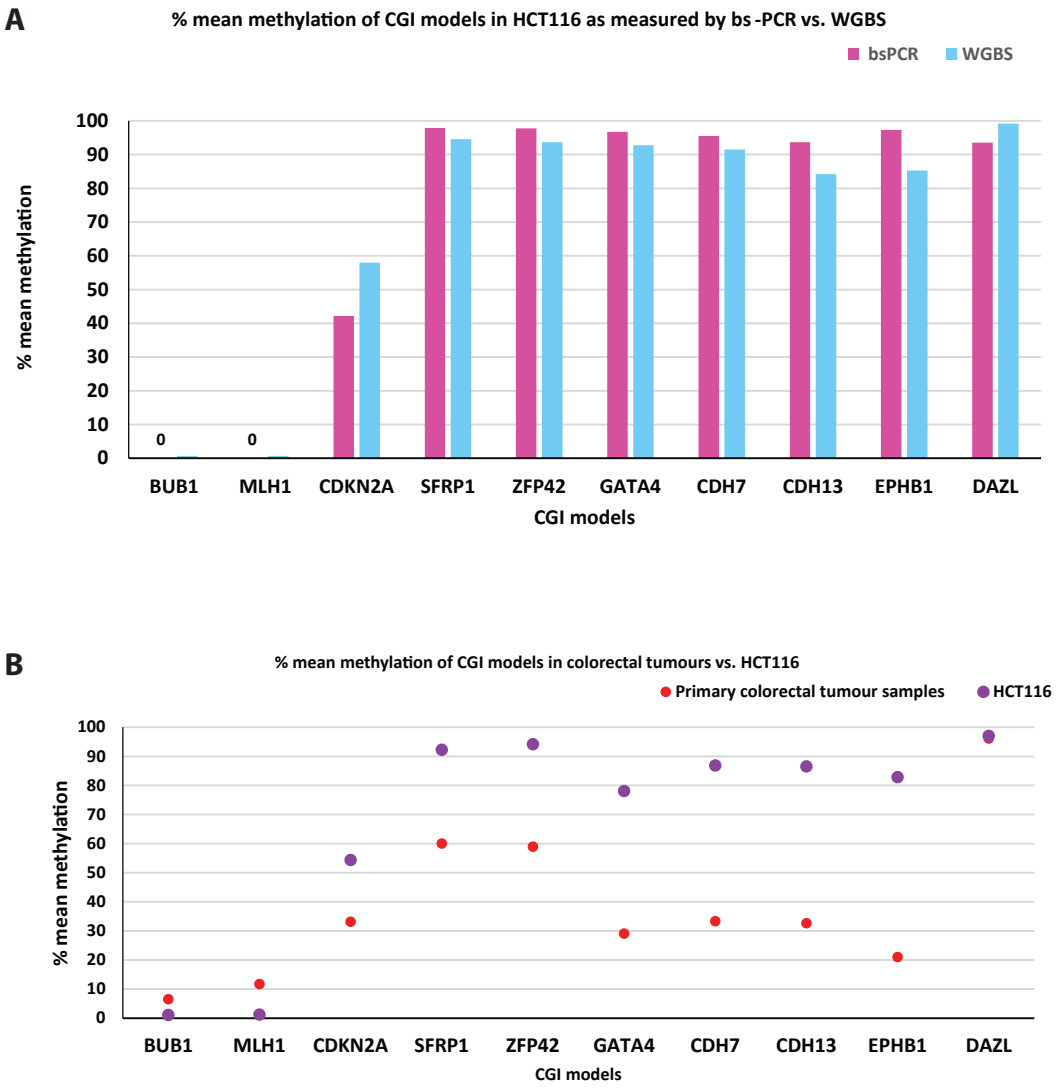


Figure 3.4: Charaterising methylation state of CGI models in HCT116 and colon tumour samples.
A. Bar graph showing the percentage mean methylation of CGI models in HCT116 as detected by clonal bsPCR (blue), and whole genome bisulfite sequencing (pink). The % mean methylation for MLH1 and BUB1 CGIs as measured by bsPCR was zero.

B. Aberrantly methylated CGIs in colorectal tumours are also methylated in HCT116. Percentage mean methylation of CGI models detected by clonal bisulfite sequencing in HCT116 (purple), and primary colorectal tumor samples (red).

levels of hypermethylation might be due to the immortalization process or because cancer cell lines are not contaminated with normal cells.

Next, I checked the expression status of our representative genes in 26 pairs of primary colorectal tumours (CIMP-H and non CIMP tumours). The gene expression profiling for these samples was performed using the Illumina Ref-8 whole-genome expression BeadChip (T. Hinoue *et al.*, 2012). In these 26 samples*, CDKN2A*, *CDH7*, and *DAZL* were inactive in both normal and tumour tissues (figure 3.5A). *BUB1*, *MLH1*, *SFRP1*, *ZFP42*, *CDH7*, *CDH13*, *EPHB1* were expressed to different extents in normal colon epithelial and primary colorectal tumour samples. BUB1 was overexpressed in tumour samples (two fold increase). *SFRP1* is down regulated in tumour (about eight fold decrease). *SFRP1* down regulation in tumour samples is concomitant with its aberrant hypermethylation, suggesting that its methylation might be important during colorectal tumorigenesis.

The fact that the aberrant methylation of most of these genes is not affecting their expression is expected since most of the genes that gain aberrant methylation in cancer are already repressed in their normal precursors (Sproul *et al.*, 2011). On the other hand, there are few genes that become both aberrantly methylated and repressed in cancer. SFRP1 was reported to be one of these genes (T. Hinoue *et al.*, 2012).

3.4 Experimental optimization

3.4.1 PiggyBac minimal terminal repeats are sufficient for transposition in HCT116.

Wild type PiggyBac TRs are 313bp and 235bp long at the 3`and 5`ends, respectively. However minimal sequence of only 40bp 3`and 67bp 5`TRs were

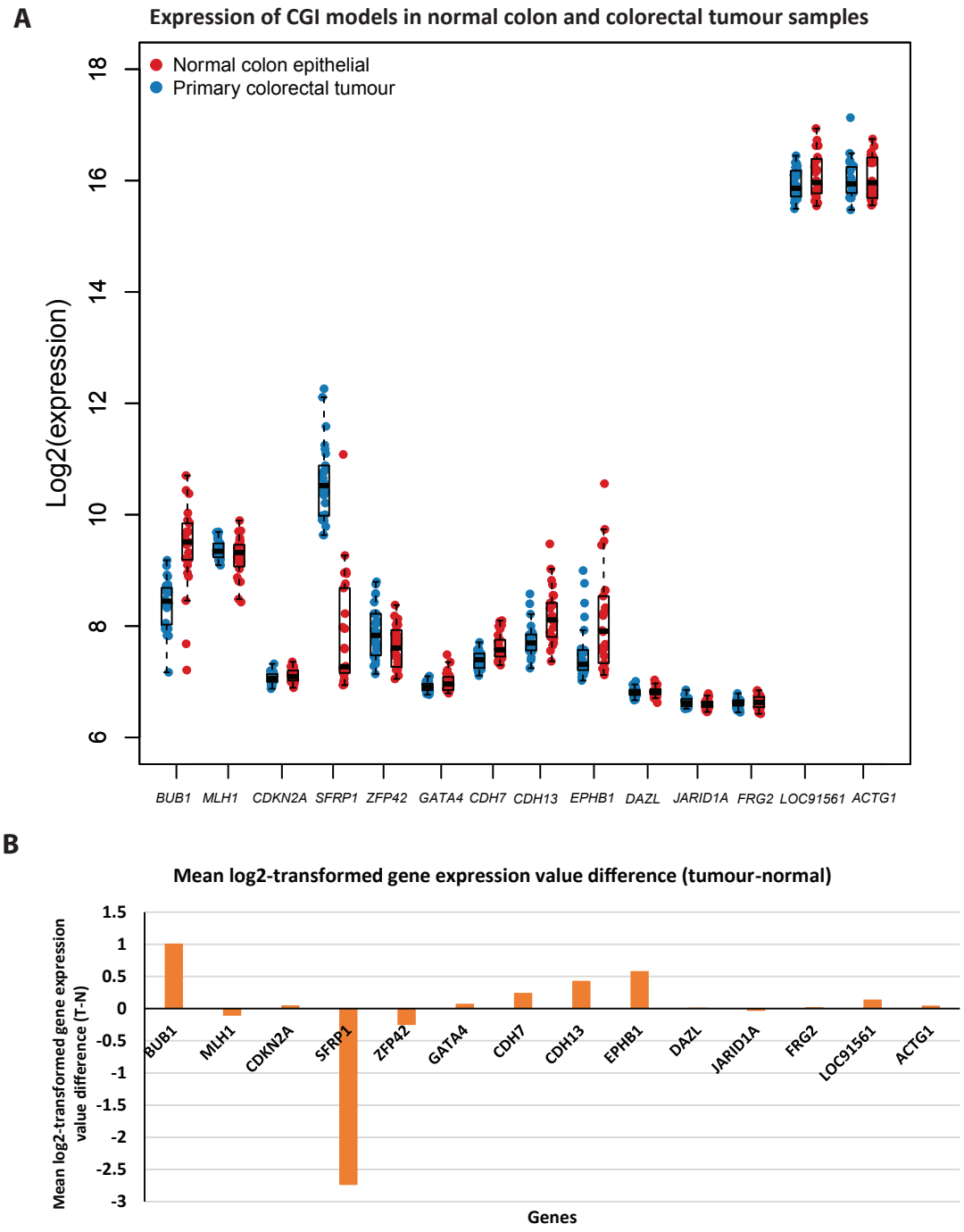


Figure 3.5: Expression of CGI models in normal colon and colorectal tumour samples
A. Distribution of Log2 Expression of CGI models in normal colon (blue) and colorectal tumour samples (red) form TCGA data. The two genes with the lowest expression (JARID1A and FRG2) and the two genes with the highest expression (LOC91561 and ACTG1) in the data set are plotted as a reference.
B. Mean Log2-transformed gene expression difference between normal colon and colorectal tumour samples from TCGA data.

reported as sufficient for successful transposition in D. melanogaster w1118 flies and in HEK 293 cells (Li *et al.*, 2005; Meir *et al.*, 2011). Utilizing piggyBac with minimal TRs instead of the long wild type TRs would make it easier to assess the methylation of CGIs at a specific location by designing a primer binding to the flanking genomic region and a primer binding to the integrated CGI.

Moreover, due to the smaller size of minimal TRs they are less likely to insulate the integrated CGIs from the flanking genomic context relative to the wild type TRs. Therefore, I adopted the minimal TRs for my experiments. Therefore, I tested both the wild type and the minimal TRs in HCT116 cells by colony formation assay. Transposition by the minimal TRs was evident albeit with lower transposition activity (10-fold reduction) compared to WT TRs (figure 3.6), as previously reported (Li *et al.*, 2005).

3.4.2 Culturing cells for 4 weeks after transfection with piggyBac plasmid is sufficient for diluting the plasmid out

In order to assess the methylation state of integrated CGIs irrespective of their locations, bsPCR was performed using primers specific to integrated CGIs (the primers amplify from the PiggyBac TRs). The presence of the plasmid should be minimized in order to minimize its amplification when bsPCR is performed.

Previous reports suggested that cells should be expanded for a minimum of three weeks for the plasmid to be diluted out (Akhtar *et al.*, 2014). To test whether expanding the cell population for 4 weeks in culture was long enough, I performed qPCR to determine the plasmid copy number (for qPCR primers standard curves, see figure S3.1) . qPCR showed that the average plasmid copy number was one copy per 400 cells (figure 3.6B). This represented both passively integrated plasmids and remaining unintegrated plasmids in the cells. On the other hand, the integrated average copy number was 2.5 copy/cell.

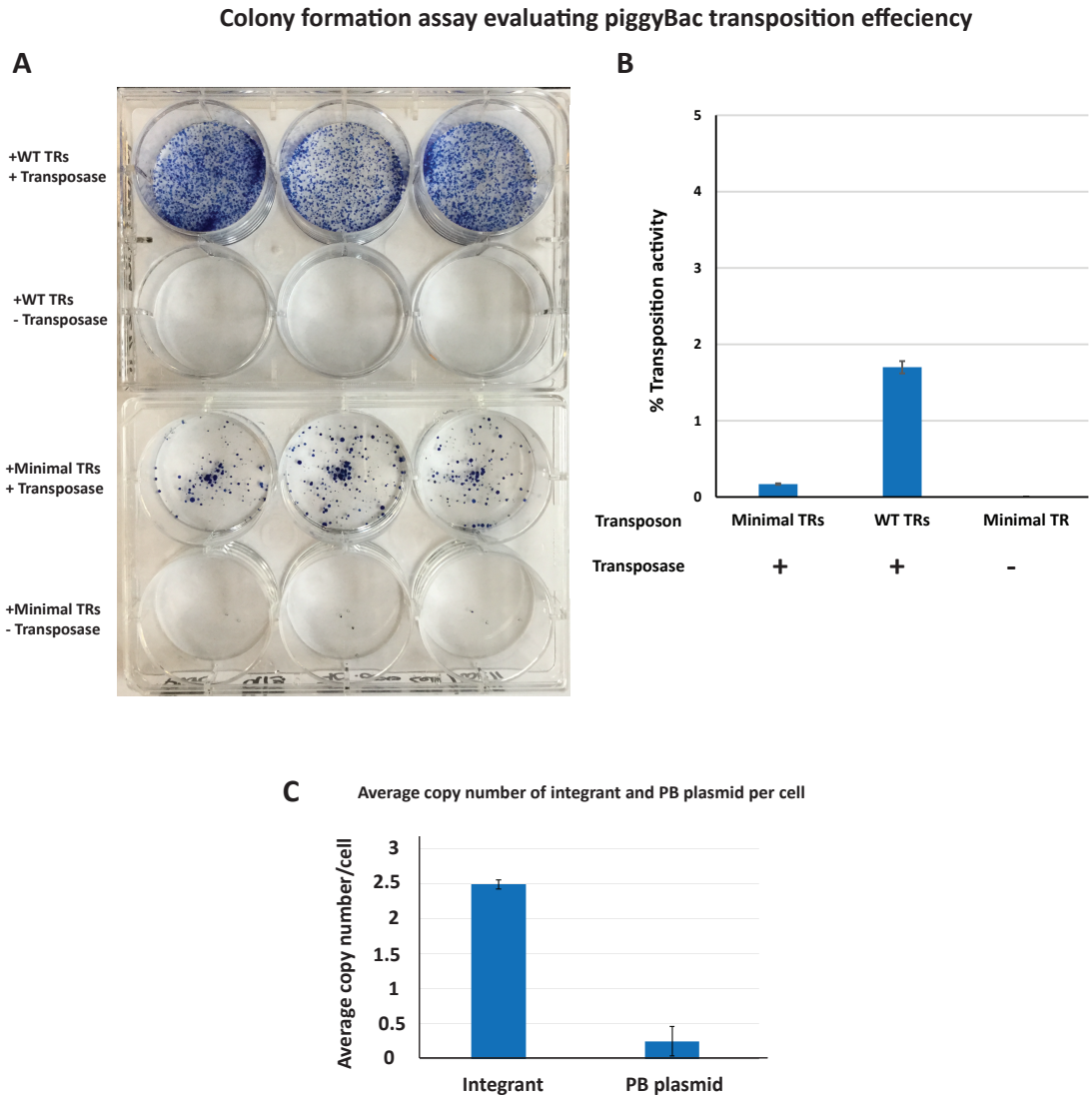


Figure 3.6: Optimization of piggyBac transposition experiment.
A. Minimal terminal repeats of piggyBac are sufficient for transposition in HCT116. Colony formation assay for HCT116 cells transfected with piggyBac vector that have wild type TRs with and without transposase (top), and piggyBac vector that have minimal TRs with and without transposase (bottom). The piggyBac has puromycin resistance gene and cells were selected with puromycin for 10 days and subsequently stained with methylene blue.
B. Percentage transposition activity of WT and minimal PB TRs, calculated from colony formation assay. Error bars indicate the standard deviation of three replicates.
C. Average copy number of integrant and piggyBac plasmid per cell, estimated from qPCR measurements. Error bars indicate the standard deviation of three replicates.

This indicates that 4 weeks is a sufficient time scale to insure that most of the plasmid has degraded.

3.5 CGIs are often unmethylated at ectopic loci in HCT116

One month after integrating the CGIs in unmethylated form in HCT116 using PiggyBac transposon, their methylation status was assessed in all loci and compared to the methylation status of their native corresponding CGIs.

3.5.1 The full length MLH1 CGI maintains its unmethylated state at ectopic loci in HCT116

MLH1 is hypermethylated and silenced in ~12% of CRC (Muzny *et al.*, 2012). MLH1 hypermethylation is frequently associated with BRAF mutations. However, in HCT116, which is BRAF-negative, MLH1 is expressed and unmethylated. I ectopically integrated the full size *MLH1* CGI (967bp long). The MLH1 CGI remained unmethylated at ectopic locations (figure 3.7A). This suggests that the sequence of unmethylated CGIs is sufficient to maintain their correct methylation state.

It is worth mentioning that I could only assess the methylation of ~300bp fragments on each side of the integrated *MLH1* CGI, since the preferable amplicon size for bsPCR is ~300bp. This is because gDNA becomes fragmented after BS treatment. Since one of the integrated CGI specific primers have to bind to the piggyBac TR, I couldn't assess the methylation of the middle part of the integrated *MLH1* without interference from the native CGI.

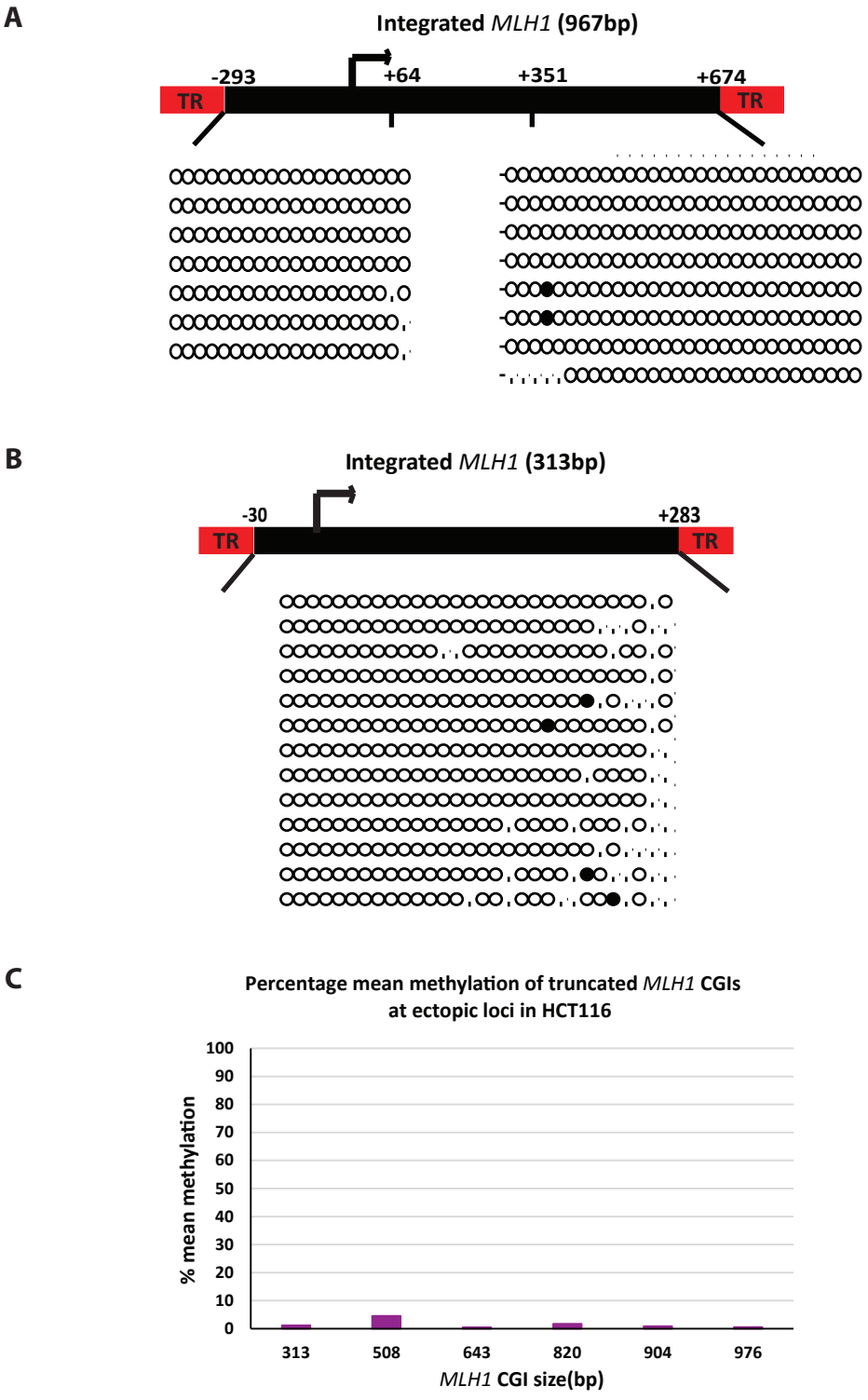


Figure 3.7: The *MlH1* CGI remained unmethylated at ectopic loci in HCT116.
Clonal bisulphite sequencing for ectopically integrated (A) 967bp long and (B) 313bp long truncated *MLH1* CGIs. Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. Empty spaces reflect ambiguous base calls in Sanger sequencing reads.
C. Bar graph showing percentage mean methylation(y-axis) of truncated *MLH1* CGIs as assessed by bsPCR.

3.5.2 Truncated MLH1 CGI maintain their unmethylated status at ectopic loci in HCT116

For the aforementioned reason, I wanted to test if smaller fragments of CGIs are reflective of the full size CGIs. To do this, truncated *MLH1* CGI constructs ranging from 300bp to 900bp were integrated using piggyBac transposon. Clonal bsPCR shows that truncated *MLH1* CGIs behaved similarly to the full length CGI and remained unmethylated at ectopic loci in HCT116 (figure 3.7B and C, figure 3.8C). This result suggested that CGI fragments of ~500bp flanking transcription start site (TSS) can be used as representatives of the full length CGIs.

I next asked whether truncated CGIs of ~500bp represent their native corresponding CGIs in their CpG and GC content. This is important as in normal cells CpG density and GC content significantly anti-correlate with methylation (Lienert *et al.*, 2011; Krebs, Dessus-Babus, Burger and Schübeler, 2014). CpG density is usually assessed as the ratio of the observed CpGs over the expected CpGs by GC content (Gardiner-Garden and Frommer, 1987). CpG observed-to-expected ratio (obs./exp. CpG), is calculated as follows:

obs./exp. CpG = Number of CpG * N / (Number of C * Number of G)

Where N = length of sequence. Comparing obs./exp. CpG and GC content of ~500bp fragments of CGI models to their native corresponding CGIs showed that most CGI fragments have similar CpG and GC content to their native corresponding CGIs (figure 3.8A and B). Although some fragments had higher or lower CpG and GC content, the differences were subtle.

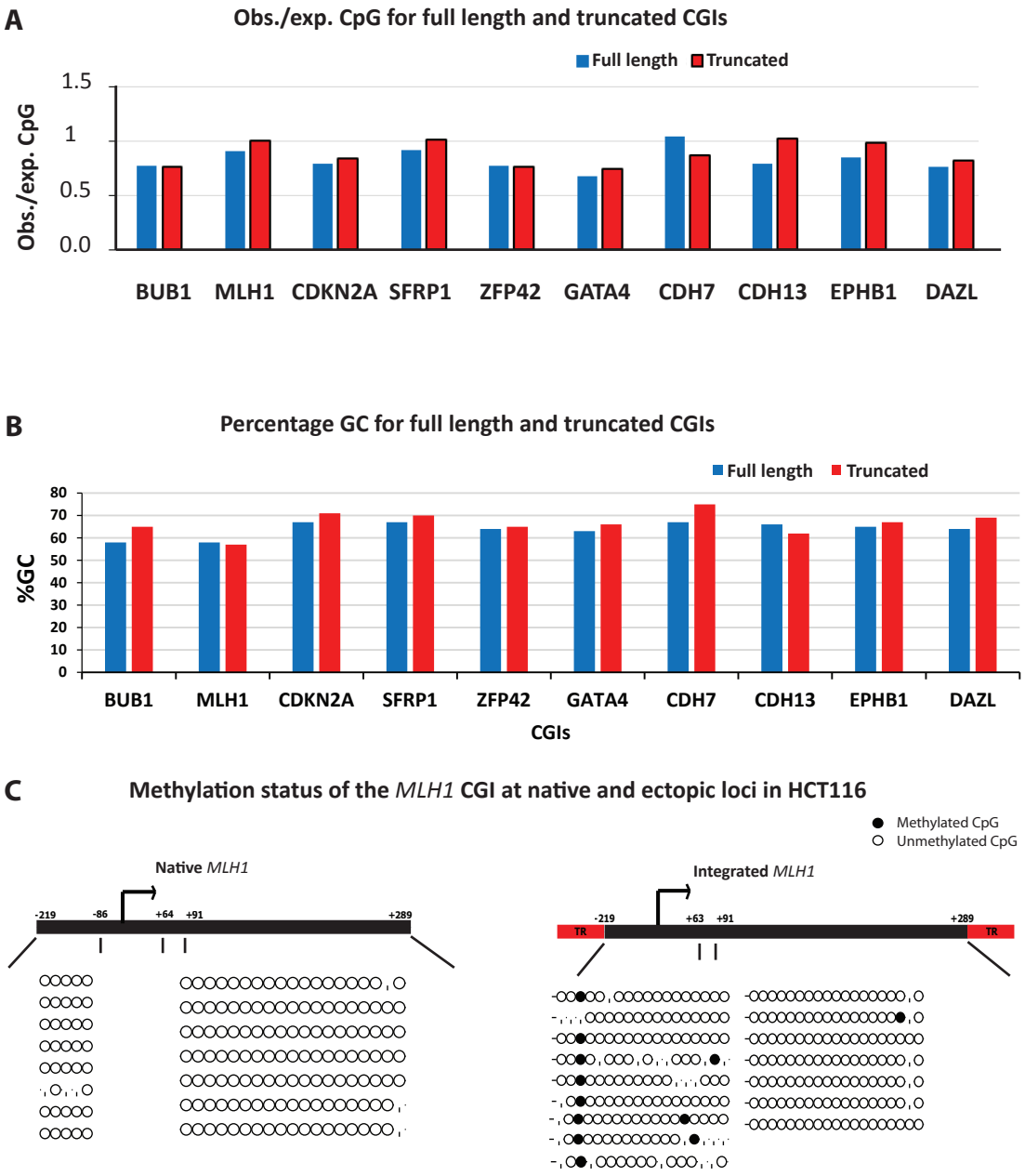


Figure 3.8:

A./B. CpG and GC content of truncated CGI models is similar to that of full length corresponding CGIs. Bar graph showing (A) obs./exp. CpG and (B) percentage GC of the truncated CGIs and their full length corresponding CGIs.

C. The *MLH1* CGI maintained unmethylated state at ectopic loci in HCT116. Clonal bisulphite sequencing for *MLH1* at native and ectopic loci. Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. Empty spaces reflect ambiguous base calls in Sanger sequencing reads.

Next I wanted to test another unmethylated CGI other than MLH1. The mitotic checkpoint serine/threonine kinase gene BUB1 is unmethylated in HCT116 and in healthy and tumour colon tissues. BUB1 CGI behaved similarly to MLH1 and maintained its unmethylated state (Appendix figure S3.1).

3.5.3 Hypermethylated CGIs are rarely hypermethylated at ectopic loci in HCT116

I next tested CGIs that are methylated in HCT116 and prone to hypermethylation in colorectal cancer. Therefore seven aberrantly methylated CGIs (*SFRP1, ZFP42, GATA3, CDH7*, *CDH13, CDKN2A* and *EPHB1*) were ectopically integrated in HCT116 using PiggyBac and their methylation was assessed one month later. I found that these methylated CGIs were mainly unmethylated at their ectopic loci in HCT116 with some variations, however a number of methylated clones were observed for *CDH13* and *CDH7* (figure 3.9, figure 3.10). Finally, I wanted to test if normally methylated CGIs would show the same trend. Therefore, I integrated *DAZL*, and found that it also remained mostly unmethylated at ectopic locations (figure 3.11).

3.6 Discussion

In this chapter I investigated whether the sequence of CGIs is the major determinant of their methylation state in HCT116 cell line. Data presented here shows that *MLH1* CGI fragments with variable lengths (900-300) bp are capable of recapitulating the hypomethylated state of the full size *MLH1* CGI (967bp). However, aberrantly hypermethylated CGIs are rarely methylated at their ectopic locations suggesting that the sequence of these CGIs is not enough to drive their aberrant methylation.

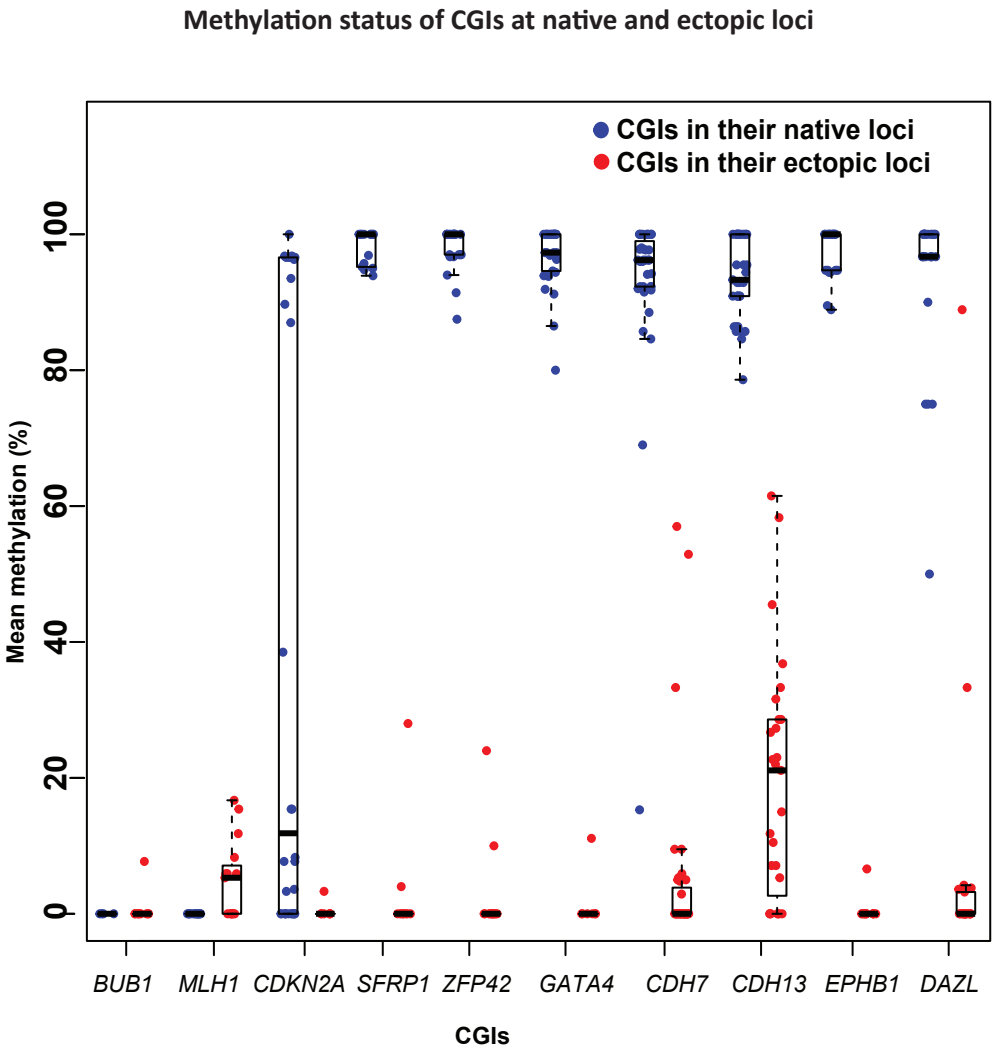


Figure 3.9: Integrated CGIs remained mostly unmethylated at ectopic loci in HCT116. Box plots showing the percentage mean methylation (y-axis) of CGIs (x-axis) at their native (blue) and ectopic loci (red). Each dot represent the percentage methylation of a single allelic CGI as assessed by clonal bsPCR. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Horizontal line within the box represents median.

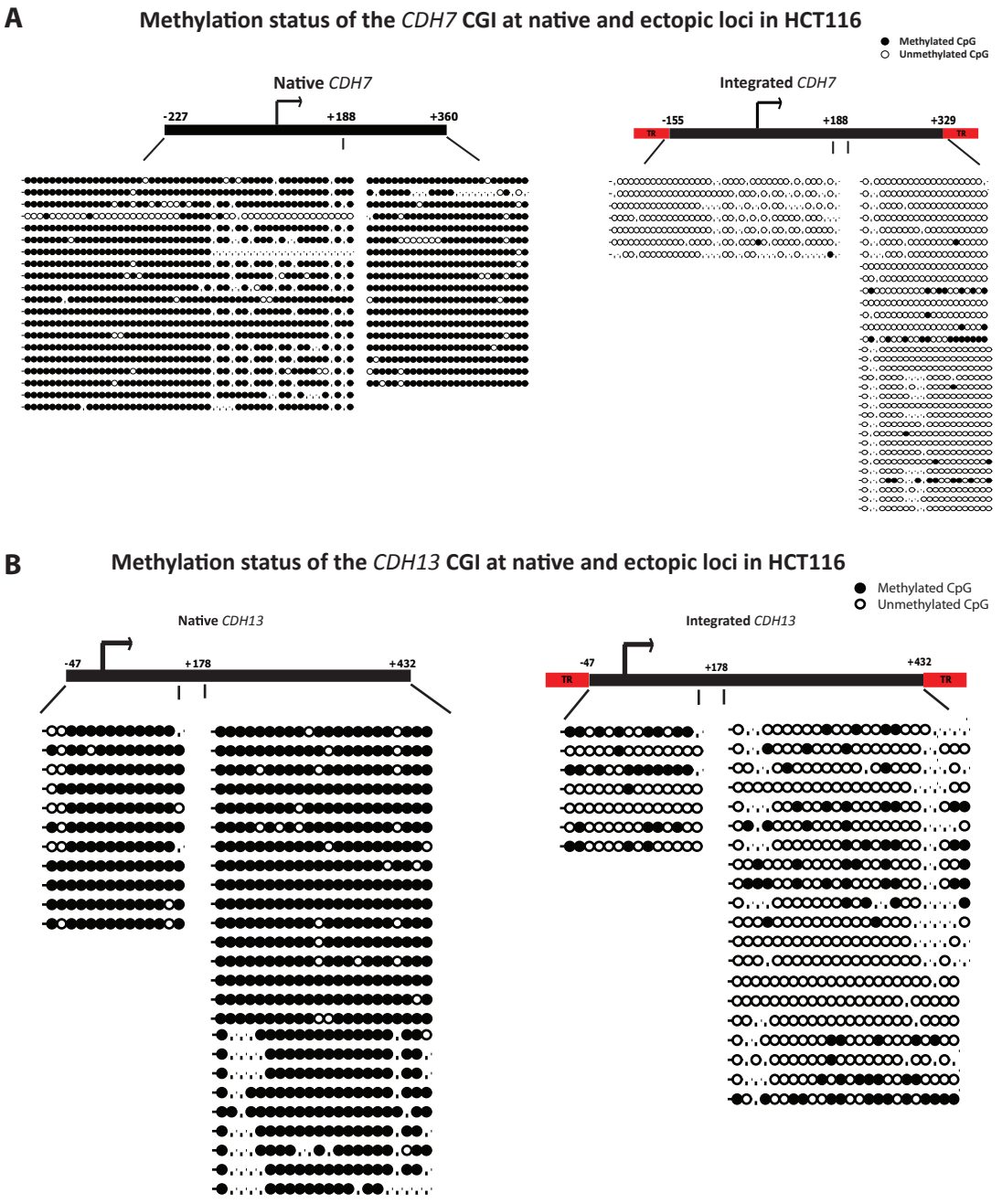


Figure 3.10: Aberrantly methylated CGIs were mostly unmethylated at ectopic loci in HCT116. Clonal bisulphite sequencing for native and ectopically integrated CDH7 (A) and CDH13 (B). Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. Empty spaces reflect ambiguous base calls in Sanger sequencing reads.

Methylation status of *DAZL* at native and ectopic loci in HCT116

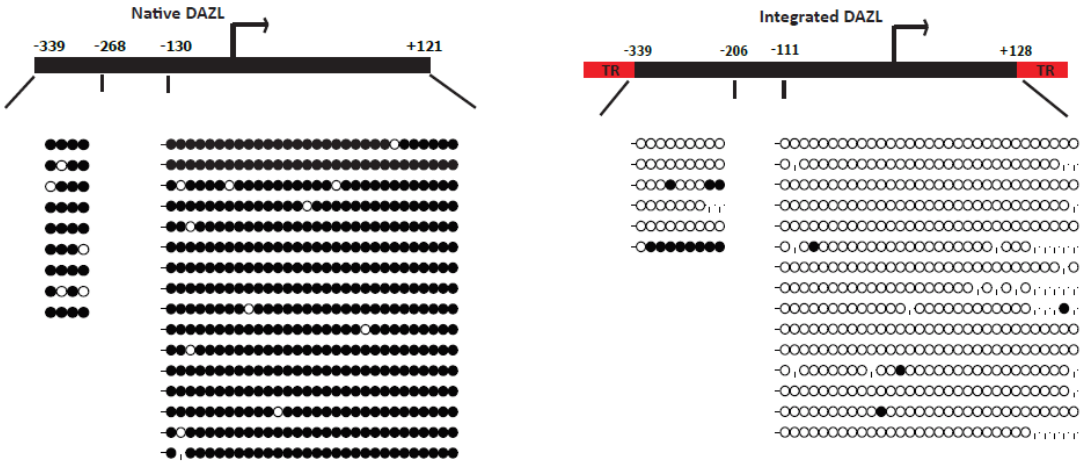


Figure 3.11: *DAZL* remained mostly unmethylated at ectopic loci in HCT116. Clonal bisulphite sequencing for native and ectopically integrated *DAZL*. Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. Empty spaces reflect ambiguous base calls in Sanger sequencing reads.

Truncated *MLH1* CGIs recapitulated the unmethylated state of the full length CGI. This result is in agreement with previous work done by Krebs *et al*., 2014. Their work showed that among 400 tested CGIs fragments of different sizes (100-400 bp), 63.5% of them could establish a methylation states similar to their native counterparts when integrated at an ectopic location in ES cells (Krebs, Dessus-Babus, Burger, Schübeler, *et al.*, 2014).

Moreover, my data shows that hypomethylated CGIs, *BUB1* and *MLH1*, remained unmethylated at ectopic locations thereby recapitulating the methylation state of their native counterparts. This agrees with the model proposed to explain methylation state of CGIs in normal cells, where the DNA binding factors, specifically transcription factors are the main determinants of its methylation state (Lienert *et al.*, 2011; Takahashi *et al.*, 2017). Both *BUB1* and *MLH1* genes are active in HCT116, suggesting that the transcription factors that bind these CGIs are expressed in HCT116. Therefore, it is expected that these DNA binding factors can still bind to the CGIs at ectopic loci and protect them from methylation.

Interestingly, aberrantly methylated CGIs were not methylated at most of their ectopic locations. This suggests that the sequence of these CGIs doesn't program their aberrant methylation. My results contrast with a study by Serra et al. who proposed that a zinc finger protein ZNF304 recruits a corepressor complex and DNMT1 to *CDKN2A* and other CGIs in KRAS-positive CRC causing their aberrant hypermethylation (Serra *et al.*, 2014). If this were the case, we would expect to *CDKN2A* CGI to be hypermethylated at ectopic locations in HCT116 which is KRAS-positive.

The same study demonstrates that *CDKN2A* is repressed in stem cells and ZNF304 is required for its repression. However, *CDKN2A* is not

hypermethylated in that system, rather it has an enrichment of H3K27me3 at its promoter. Moreover, the same corepressor complex including DNMT1 exists at the *CDKN2A* locus in ES cells (Serra *et al.*, 2014). Their findings argue that although ZNF304 is repressing *CDKN2A*, a different mechanism is driving the hypermethylation of the repressed gene in cancer. A possible explanation for the difference in our results, is that the binding site of ZNF304 occurs outside the annotated the *CDKN2A* CGI and was not included within the fragment of *CDKN2A* CGI used in my experiment. One way to check whether ZNF304 binding could cause hypermethylation in HCT116, would be to ectopically integrate the genomic fragment covering the *CDKN2A* CGI and the ZNF304 binding site and see if hypermethylation of the integrated fragment occurs.

The normally methylated CGI, *DAZL*, was mostly hypomethylated at ectopic locations. *DAZL* is a germline associated gene, whose methylation is needed for its repression (Maatouk *et al.*, 2006). It is *de novo* methylated during implantation in epiblast cells. The unmethylated state of integrated *DAZL* implicate that the sequence of the integrated fragment is not sufficient to recruit factors required for its *de novo* methylation or that these factors are not present in HCT116.

Exceptions to the hypomethylation of integrated CGIs were observed. For example, *CDH7* and *CDH13* CGIs were methylated (methylation >30%) at 3, and 6 genomic locations, respectively. This confirms that HCT116 has *de novo* methylation activity in line with what's previously published (Lengauer, Kinzler and Vogelstein, 1997). The fact that the same sequence can be methylated or unmethylated at different locations, suggests that some CGIs are landing in genomic sites that can recruit DNMTs.

Since the transposase was functional in HCT116 without induction, the expression of the transposase couldn't be restricted to a short time window. From previous reports, this is expected to increase the remobilization of the integrated CGIs. For example, for a HCT116 cell with a CGI integrated at a certain site, the progenitor cells will not necessarily maintain the CGI at the same site as long as the transposase exists in the cells. Thus each integration site will be a rare event. I was able to identify the integration sites of CGIs in the population, however due to this limitation I couldn't assess the methylation of that CGI at a specific locus.

Overall, results presented in this chapter suggest that the sequence of CGIs doesn't program their aberrant methylation in cancer. However, this finding needs to be confirmed.

Chapter 4: Targeted Integration of CGIs

In order to confirm results from the previous chapter, I designed and tested a strategy for targeted integration into a chosen genomic locus in HCT116. The main advantage that the targeted integration strategy has over the random integration is that it minimizes the possible effects of the context on the methylation status of the integrated CGIs. However, clonal cell lines with the integrated CGIs at a chosen genomic locus could not be generated due to a combination of high bystander effect and low RMCE efficiency in HCT116.

4.1 Experimental Design

Targeting the integration of exogenous DNA into a specific locus in the genome of mammalian cells can be achieved through recombination between homologous sequences in the exogenous and endogenous DNA, a process termed homologous recombination (HR) (Smithies *et al.*, 1985; Thomas, Folger and Capecchi, 1986; Doetschman *et al.*, 1987; Thomas and Capecchi, 1987). However, the frequency of this process is very low and can range from one event in 10^3 to one in 10^7 cells (Vasquez *et al.*, 2001).

Higher integration frequency can be achieved using a nuclease to induce a DNA break at a desired locus (Bachu, Bergareche and Chasin, 2015). The DNA double strand break (DSB) can stimulate repair pathways in the cells during which the DNA template gets inserted and ligated around the breakage site. There are three major groups of targetable nucleases which can be used to induce DNA DSB; zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the clustered regularly interspaced short

palindromic repeat (CRISPR)/Cas9 (Jinek *et al.*, 2013). CRISPR/Cas9 takes less time to engineer and is the most straightforward strategy among the three

groups of nucleases. Engineered CRISPR/Cas9 is composed of Cas9 protein and a guide RNA (gRNA) that guides the nuclease to a certain genomic locus. The 20nt sequence at the 5'-end of the gRNA determines the targeting specificity by binding to a unique complementary genomic locus (Ran *et al.*, 2013a). Thus, the same nuclease can be used to target various loci using different gRNAs. On the other hand, both ZFNs and TALENs require engineering a pair of nucleases for each desired genomic loci. Therefore, I used CRISPR/Cas9 to perform the DSB to facilitate targeted integration in my experiment (figure 4.1).

The two main pathways through which a DSB can be repaired are HR, and non-homologous end joining (NHEJ). NHEJ pathway is faster and more frequent than HR in somatic cells (Mao *et al.*, 2008). Recent studies have successfully perform targeted integrations in mammalian using a process termed *in vivo* blunt end ligation (Maresca *et al.*, 2013; Bachu, Bergareche and Chasin, 2015; Geisinger *et al.*, 2016). In this method, cells were provided with an exogenous DNA template without flanking homology arms which got integrated, possibly by NHEJ, at higher frequencies than those reported by HR. Therefore, I designed my experiment to utilize *in vivo* blunt end ligation to perform the targeted integration. Nevertheless, the frequency of the insertion using this method is still low and selection is usually used to enrich for the correct insertion. For my experiment it is preferable to avoid using a positive marker in the vicinity of the integrated CGIs, since it might interfere with their methylation state. Therefore, I chose Recombinase Mediated Cassette Exchange (RMCE) to perform targeted integrations. This technique can be used with a negative selection strategy to perform targeted integration. RMCE has been used successfully in previous

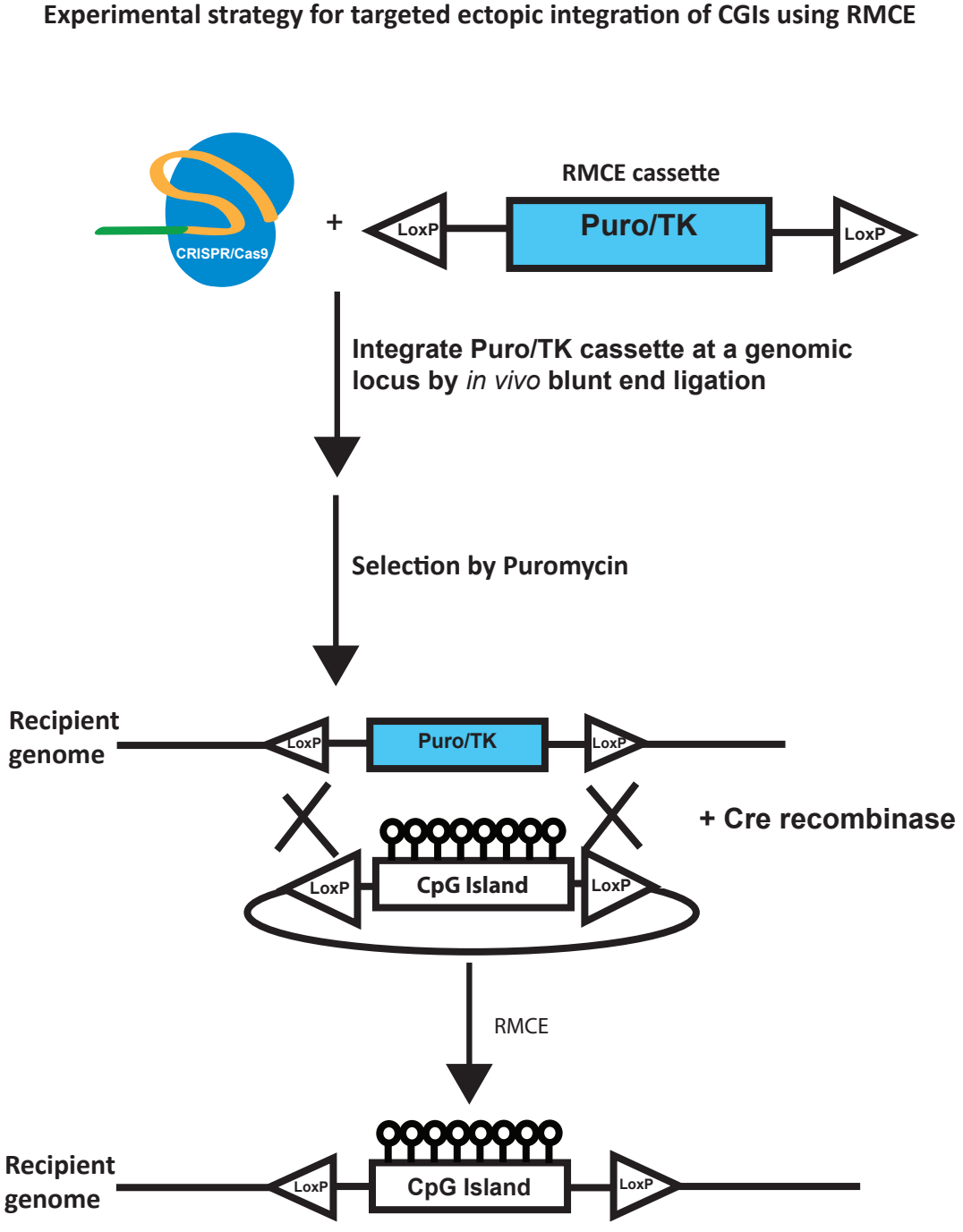


Figure 4.1: Overview of an experimental strategy for targeted ectopic integration of CGIs by RMCE. In the first step, the loxP-flanked Puro/TK cassette is introduced into a selected genomic site by *in vivo* blunt end ligation through NHEJ to establish HCT116 cell line with a landing site. Puromycin allows positive selection in this step. The donor cassette carries the loxP-flanked CGI. Next, a Cre recombinase mediate the recombination of the CGI into the recipient genome replacing the Puro/TK cassette. Cells with integrated CGIs are selected by ganciclovir.

studies to integrate CGIs in mouse embryonic stem cells and in mouse models (Feng *et al.*, 1999; Lienert *et al.*, 2011; Lynch *et al.*, 2011; Krebs *et al.*, 2014).

I used CRISPR/Cas9 to induce a DSB at a chosen genomic locus and subsequently introduced a RMCE targeting cassette at that locus by *in vivo* blunt end ligation. This enabled me to establish a clonal HCT116 cell line with engineered targeting site (figure 4.1). The targeting cassette is composed of a positive/negative selection gene (puromycin/thymidine kinase fusion gene, "PuroTK") flanked by an inverted loxP and thus positive selection was applied at this step to enrich for the cassette integration. Next, CGIs were integrated at the engineered locus by transfecting the cells with a donor vector containing a CGI with flanking loxP sites and a Cre-recombinase expressing vector. Recombination occurred between the loxP integrated in the genome and the loxP site in the donor vector. As a result, the selection cassette was replaced by the CGI from the donor vector. A negative selection can be applied at this step to kill the cells that still contain the selection cassette and to enrich for cells with the CGI insertion.

4.2 Selecting a genomic locus near β-globin genes cluster for the targeted integration

First of all, I needed to select a suitable genomic locus for testing the methylation of ectopically integrated CGIs without any interference from the surrounding genomic environment. Previously, Lienert *et al.* had selected the β-globin locus to perform targeted integrations of CGIs in mESC (Lienert *et al.*, 2011). In their study CGIs integrated in the β-globin locus were capable of recapitulating the correct methylation state. This suggests that the β-globin locus might represent a suitable site for genomic integration of CGIs. Therefore, I wanted to check whether the β-globin locus in HCT116 is transcriptionally inert. This is important to avoid any interference of the transcription machinery at the

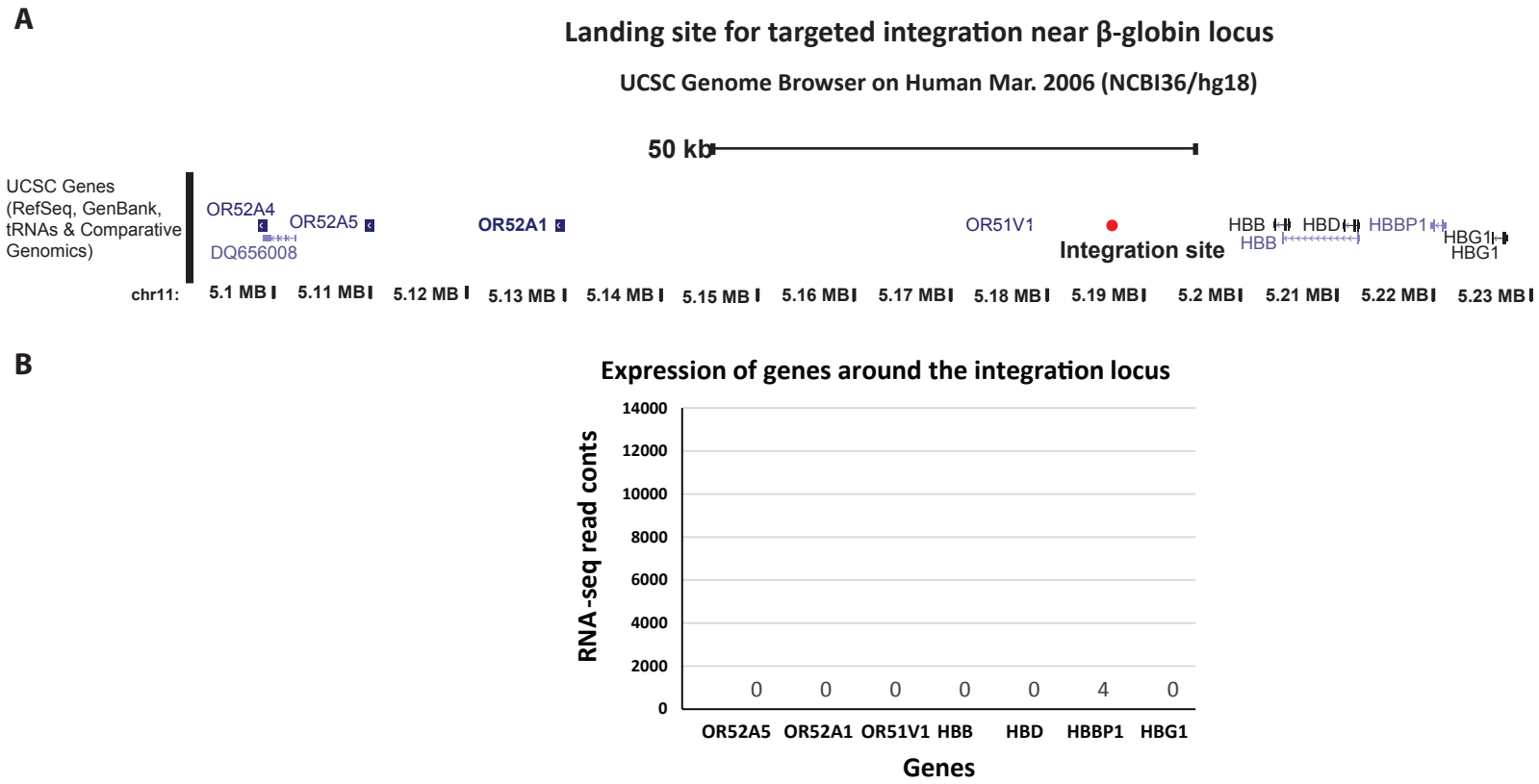


Figure 4.2: β-globin locus is located in a transcriptionally inactive region on chromosome 11.
A. The selected integration site is located in a 20 Kb gene desert betweet the olfactory receptor genes and β-globin genes cluster. UCSC Genome Browser plot of 130kb surrounding the selected integration site near β-globin locus (represented by the red circle).
B. The genes flanking the integration site are repressed. Expression levels of genes surrounding the chosen integration site represented as RNA -seq read counts. (RNA-seq data for HCT116 cell line were processed by Catherine Doust).

integration site with methylation of the integrated CGIs. The genomic region between the human Beta globin gene cluster and the olfactory receptor gene cluster on chromosome 11 is a gene desert of ~ 25kb (Figure 4.2A). The flanking olfactory receptor gene and β-globin gene are primarily expressed in olfactory receptor neurons and erythroid cells, respectively. The function of these genes is not required in colorectal cancer cells and their expression is undetectable in HCT116 (Figure 4.2B). Thus, this region represents a transcriptionally inert environment in HCT116 and was therefore chosen as a test region for methylation. For validating my results, I choose the human α-globin gene cluster, which is located on chromosome 16 and extends over about 30 kb, as another site to perform targeted integration.

4.3 Generating HCT116 clonal cell line with engineered RMCE cassette near the β-globin locus

In order to establish a clonal cell line with integrated RMCE cassette at the β-globin locus, I needed to introduce a Cas9-induced DSB at the β-globin site. Firstly, I wanted to have a positive CRISPR/Cas9 editing control. Therefore, I adopted a target from previous published work, DYRK1, which showed up to 32% total editing frequency in HEK293T cells as detected by T7 endonuclease I assay (Lin *et al.*, 2014). I targeted DYRK1 in HCT116, this resulted in editing efficiency of about 33%, as detected by the SURVEYOR assay (Figure 4.3A). Both T7 endonuclease I assay and SURVEYOR assay detects errors in DNA repair and are widely used as surrogates for Cas9-induced DSB (Lin *et al.*, 2014) (Ran *et al.*, 2013a). DYRK1 was therefore chosen as a positive CRISPR/Cas9 editing control.

Next, I wanted to design sgRNA which will target CRISPR/Cas9 to the β-globin locus. Since CRISPR/Cas9 sgRNAs are usually designed to target short regions within genes, most of the sgRNA design tools pick targets based on standards

related to this goal. For my experiment, I need to pick an intergenic target site within a ~25kb genomic region. Therefore, three web based tools were utilized: CRISPR-screen, chopchop (Montague et al., 2014) and CRISPR design tool (Ran *et al.*, 2013a). CRISPR-screen was first used to select the target sites within the 25kb region with minimum potential off-targets. Subsequently, 250bp around each target site were supplied to chopchop and CRISPR design tool. The target sites that were the consensus of the three tools were annotated in the UCSC genomic browser and those that are not within repetitive elements and do not have single nucleotide polymorphisms (SNP) in the protospacer adjacent motif (PAM) site or within the 12 nucleotide next to PAM (termed seeding region) were chosen. Mismatches in the PAM or the seeding sequences have been reported to affect the gRNA targeting efficiency substantially (Cong *et al.*, 2013).

Next, I cloned the top four sgRNAs targeting the β-globin site, and assessed their targeting efficiency. One gRNA showed 34% total editing efficiency as assessed by SURVEYOR and was chosen to target the β-globin locus in the following experiments (Figure 4.3A).

Using my CRISPR/Cas9 construct that successfully targets the β-globin locus, I wanted to assess HR mediated integration frequency at that site. Therefore, HCT116 cells were transfected with the CRISPR/Cas9 targeting construct and an exogenous DNA template composed of a 12 bp containing a HindIII site flanked by 90bp homology arms. 48hrs after transfection, gDNA was extracted from the population of cells and the targeted site was PCR amplified. HindIII digestion of the amplified region indicated that HR was achieved at this site (Figure 4.3B). To estimate editing efficiency I cloned and sequenced the amplified region. One clone out of 96 was detected to contain the correct insertion (Figure 4.3C). Consistent with previous studies, this indicated that the frequency of HR mediated integration in somatic cells was low. Moreover, a

further reduction in HR mediated integration frequency is expected with a larger DNA template. In my experiment, the DNA templates are 500-1000 long.

4.4 Generation of HCT116 cell lines with integrated L1-PuroTK-1L into β -globin locus and α-globin locus.

L1-PuroTK-1L a 2.5 kb long positive/negative-selection construct, composed of a Puromycin-thymidine kinase fusion (PuroTK) gene flanked by inverted loxP sites, was cloned (see section 2.6.7). HCT116 cells were transfected with the CRISPR/cas9 construct targeting the β-globin locus and the L1-PuroTK-1L cassette. 48 hrs after transfection, cells were plated at low density (500-1000 cells/ 10 cm cell culture dish) to grow single clones. Puromycin resistant clones were screened by amplifying the β-globin-cassette junction. Out of 194 puromycin-resistant clones screened, 18 clones had the RMCE construct inserted at the β-globin locus in either orientation (Figure 4.4A).

Errors at the genomic/cassette junctions, such as small deletions or insertions, have been reported in literature (Geisinger *et al.*, 2016). Therefore, the clonal cell lines were further screened to check that they have correct LoxP sequence. The cassette junctions were PCR amplified, cloned and sequenced. Two clones (2E8 and 2B7) had precise LoxP sequences (figure 4.4B). Some of the other clones had deletions within the LoxP site and therefore can't be used for RMCE (figure 4.5A). The other clones couldn't be verified by sequencing because the sequences stop before covering the LoxP sequence.

Ideally the RMCE clonal lines should have only one copy of the cassette, as additional copies can cause complications in the negative selection. Additionally, having several copies of the cassette might lead to having several copies of the integrated CGI that might create competition between them for binding factors and thus influence their methylation state. Therefore, southern analyses were

Testing CRISPR\Cas9 gRNAs targeting at B globin locus

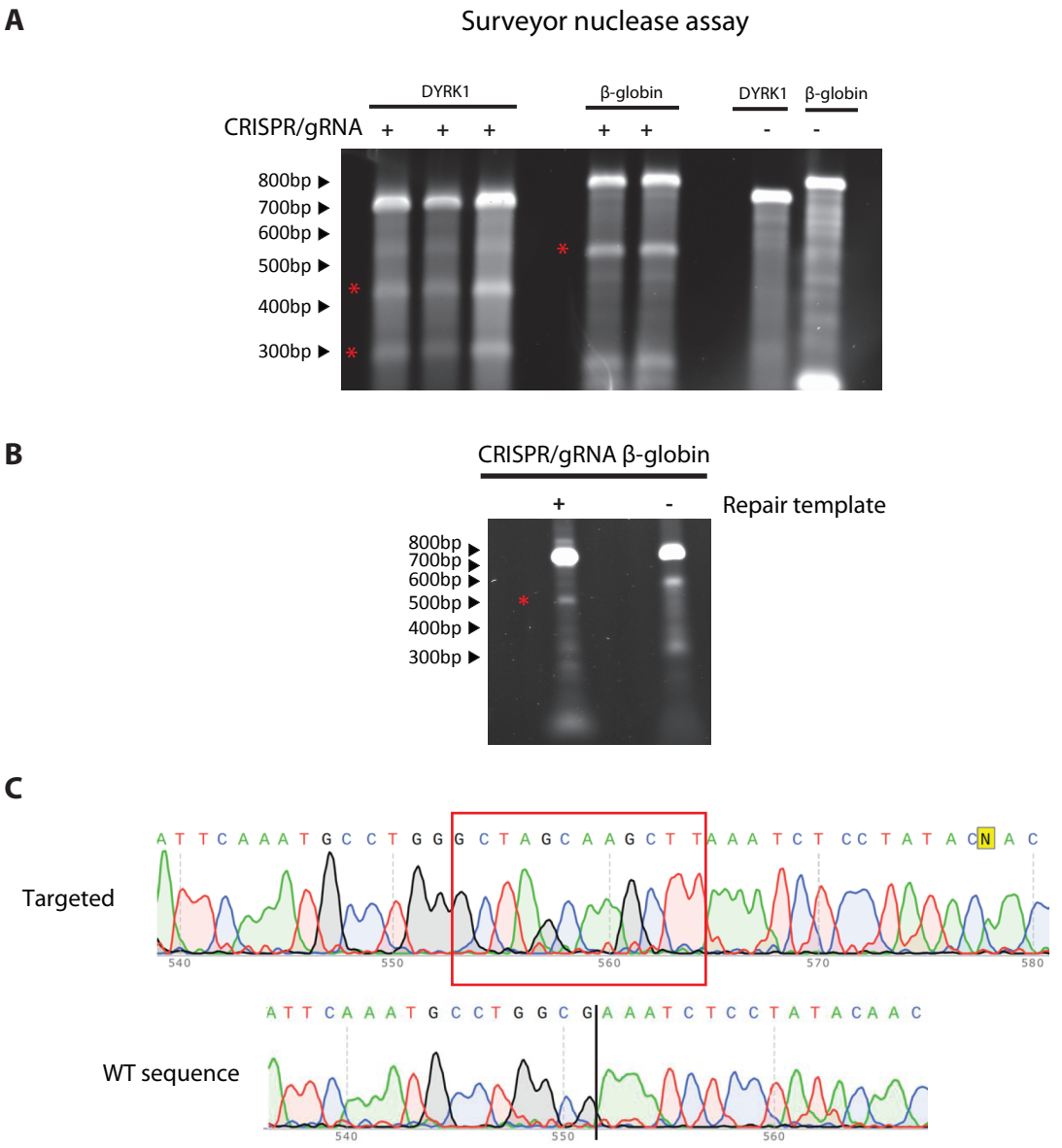


Figure 4.3: CRISPR/Cas9 induced NHEJ and HDR at the β-globin locus in HCT116.
A. SURVEYOR nuclease assay performed on genomic DNA form HCT116 cells after transfection with CRISPR and sgRNAs targeting the genomic loci next to β-globin locus and DYRK1 (positive control). Red asterisks mark the digestion products that are indicative of correct editing.
B. HindIII digestion of genomic DNA extracted from HCT116 after transfection with CRISPR , sgRNA targeting the genomic region next to the β-globin locus, and a repair template with 12bp containing a HindIII site. Red asterisks mark the digestion products that are indicative of correct editing.
C. Sequence traces showing the 12 bp integrated DNA fragment (Top sequence, outlined with a red box) at the β-globin locus compared to the corresponding wild type locus (lower sequence). CRISPR/Cas9 cutting site is marked by a solid black line.

performed using 760bp fragment of the cassette as a probe for RMCE cassette (figure 4.5A). Southern blot showed that the 2E8 clone had one band of the correct size verifying that the cassette is at the β-globin locus (figure 4.5B, right panel). On the other hand, the 2B7 clone gave a fainter band of greater size.

To assess if the RMCE cassette was integrated at one allele or both alleles, the targeted site at β-globin was PCR amplified using primers specific for the β-globin locus spanning the targeted site. The presence of the wild type allele is expected to give a band of 237bp and 257bp length using P1 and P2 set of primers, respectively. The other allele with the integrated RMCE cassette is expected to give a band of 4.2kb. The amplicon from the 2E8 clonal cell line showed a band of the expected size for wild type allele and no band was observed at 4.2kb possibly due to its large size (figure 4.6B). The amplicon from the 2B7 clonal cell line didn't show any band. Therefore, I selected the 2E8 clone to perform RMCE.

4.5 Integration of CGIs into β -globin locus by RMCE

HCT116 cells were transfected with a CGI sequence flanked by inverted LoxP. Negative selection with 2μg/ml ganciclovir (GCV) was applied for 7 days. Correct integration of the CGI at the β-globin locus was confirmed by PCR amplifying the genome-CGI junction, followed by cloning and sequencing the PCR product (figure 4.7A).

GCV is a prodrug that gets activated inside the cells possessing the TK gene. Early studies in the field of gene suicide therapy had reported that GCV treatment not only kills cells with TK genes, but also neighbouring cells lacking this gene through bystander effect (van Dillen *et al.*, 2002). This effect is

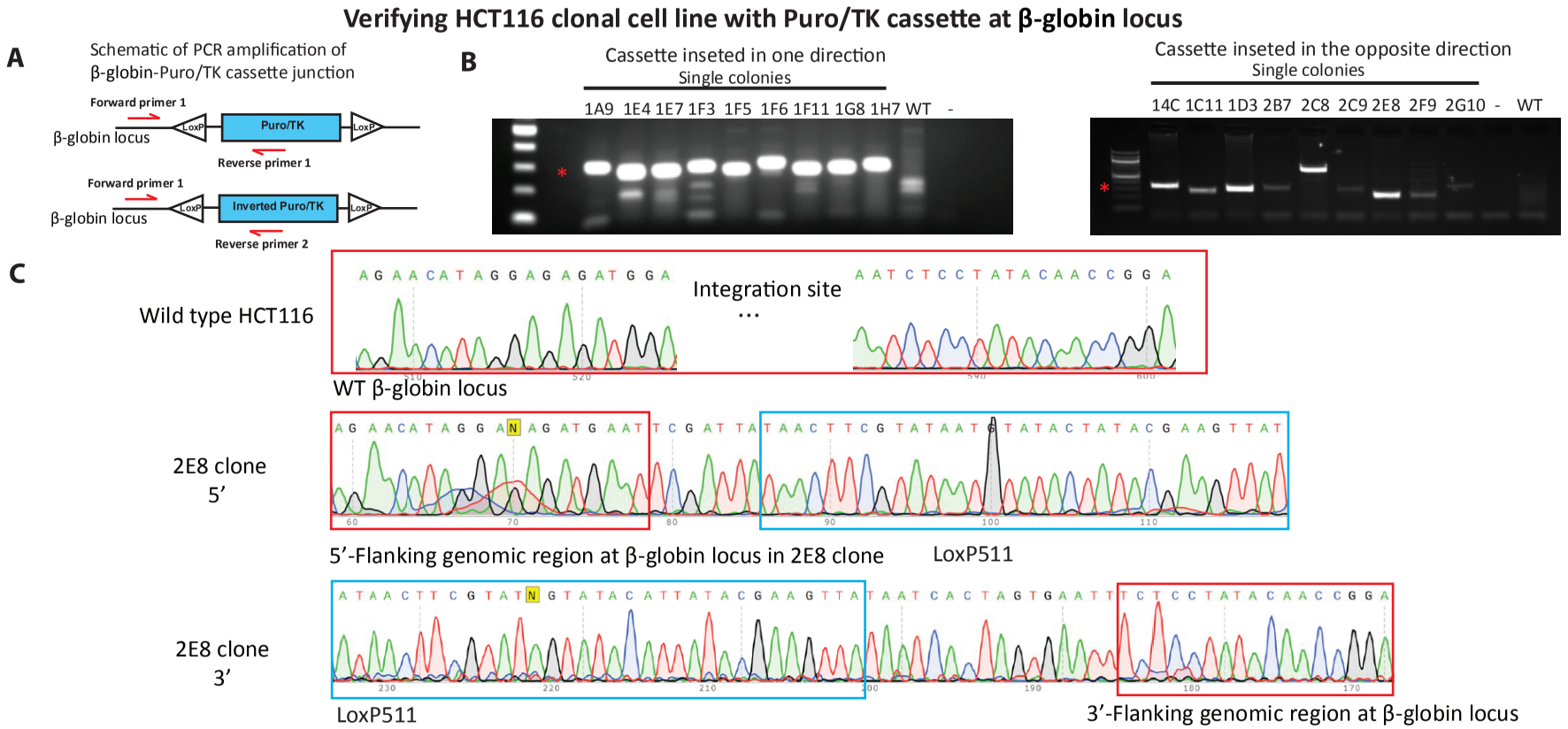


Figure 4.4: Verifying HCT116 clonal cell line with Puro/TK cassette at β-globin locus

A. Schematic of PCR amplification of β-globin-Puro/TK cassette junction integrated in one direction (top) and in the opposite direction (bottom). One primer is specific for the β-globin locus and the other is specific for puro/TK cassette.

B. PCR amplification of β-globin-Puro/TK cassette junction integrated in one direction in 9 isolated HCT116 clones and integrated in the opposite direction in another 9 HCT116 clones. The correct PCR product size is marked by red asterisk. Some amplicons have either larger or smaller band size due to insertions or deletions at the integration junction (as verified by sequencing).

B. Sequence traces showing β-globin-Puro/TK cassette junctions in the 2E8 HCT116 clone compared to the wild type targeted sequence at β-globin locus. The sequence of LoxP511 and the flanking genomic region are outlined with blue and red boxes, respectively.

Southern blotting of Puro/TK cassette in 2E8 HCT116 clonal cell line

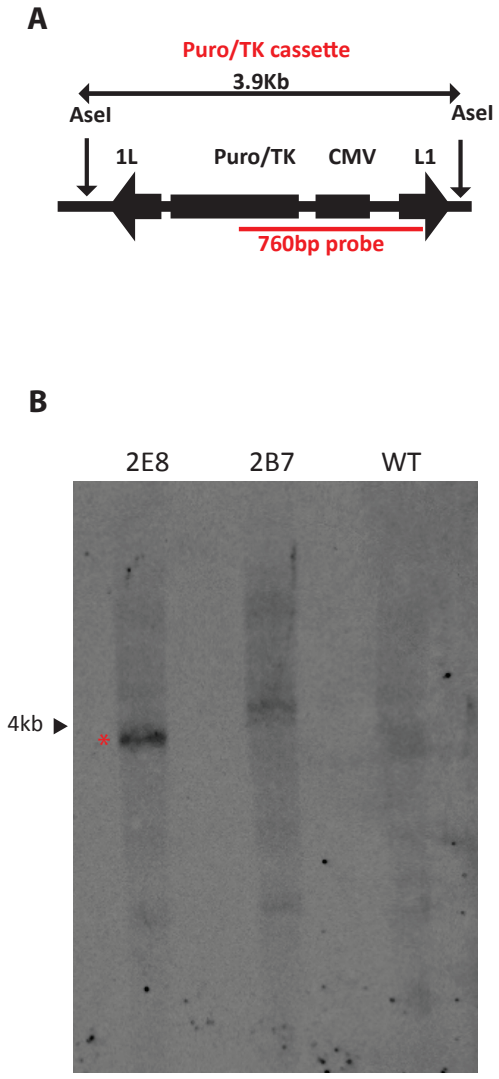


Figure 4.5: The 2E8 clone has a single correct integration of the Puro/TK cassette.
A. Schematic, outlining a 760 bp long probe used to detect the puro/TK cassette in southern blot. The gDNA was digested with AseI. The presence of the puro/TK cassette at the β-globin site should be detected as a single band of 3.9kb and is labeled with an asterisk.
B. Southern blotting analysis to identify copy number and verify integration site of the Puro/TK cassette in the 2E8 clone. 10 µg AseI digests of gDNA from the 2E8 clone, the 2B7 clone and WT HCT116 cell line were run on a gel, transfared to a Hybond N+ membrane and hybridised to the indicated probe.

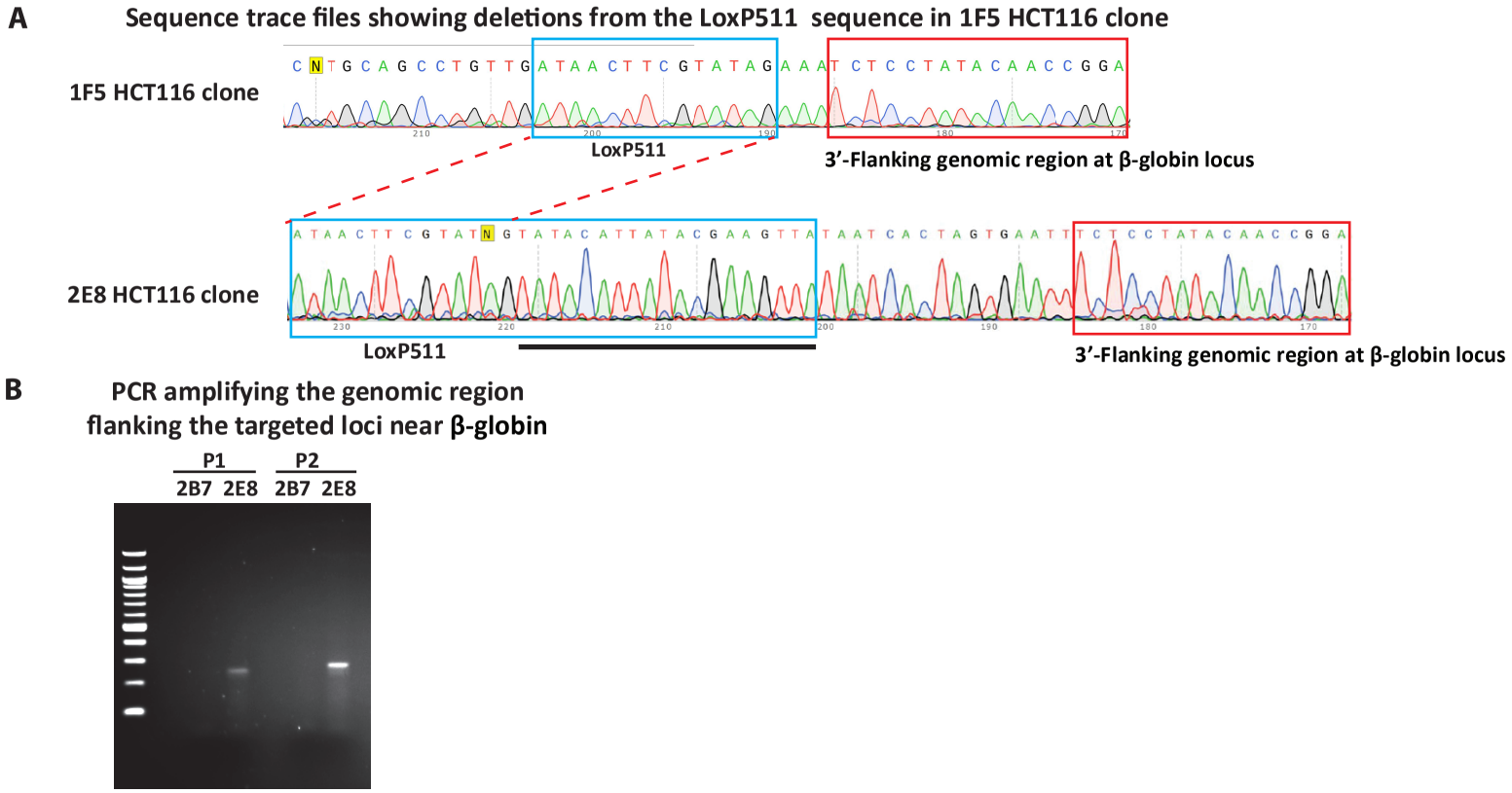


Figure 4.6:

A. An example of a HCT116 clonal cell line with deletions in the LoxP511 sequence. Sequence traces showing β-globin-Puro/TK cassette junctions in 1F5 HCT116 clone with deletions from the LoxP sequence compared to the 2E8 HCT116 clone with the correct LoxP511 sequence. The deleted sequence is underlined in the alignment. The sequences of LoxP511, and the flanking genomic region are outlined with blue and red boxes, respectively.

B. 2E8 HCT116 clone has a single integration at the β-globin locus. PCR amplification of the genomic region flanking the targeted loci at β-globin locus from 2E8 and 2B7 HCT116 clone. The presence of the wild type allele in 2E8 clone is seen as bands of 237bp and 257 bp length amplified by P1 and P2, respectively. 2B7 HCT116 clone didn't produce any amplicons with primers P1 and P2.

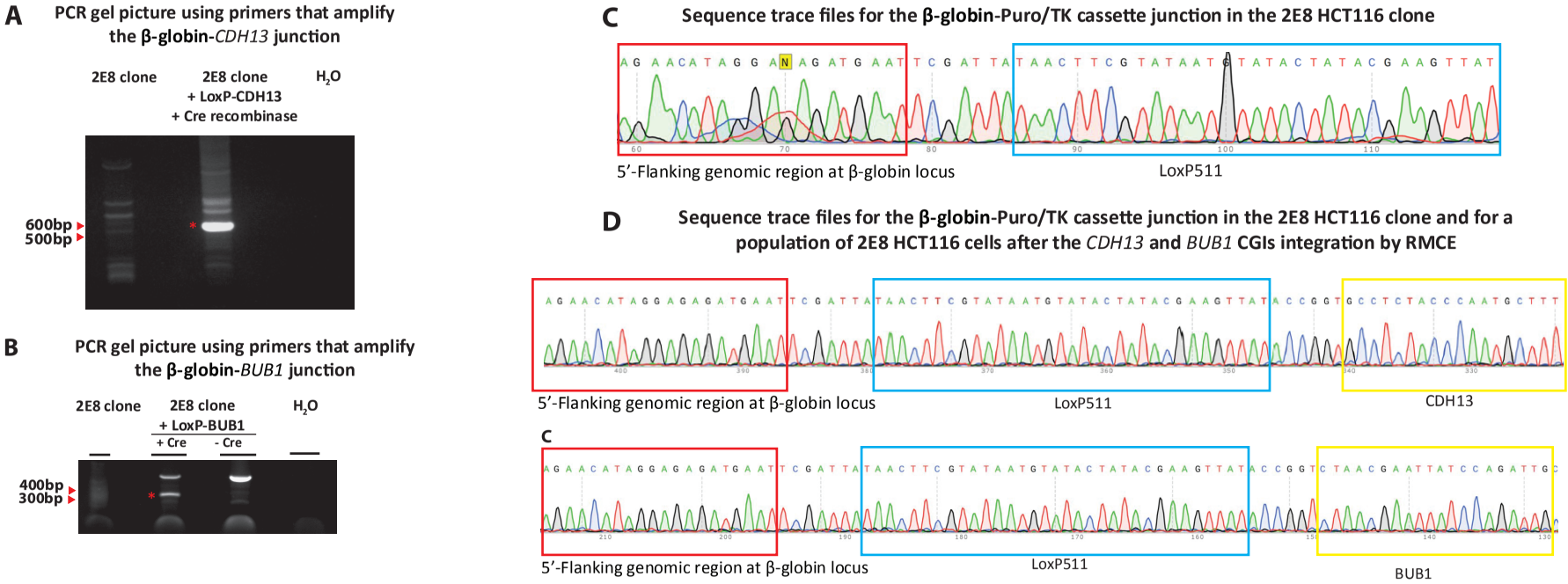


Figure 4.7: *CDH13* **and** *BUB1* **CGIs were successfully integrated at the β-globin locus by RMCE .**
A. A PCR gel picture using primers to amplify the β-globin-*CDH13* junction from gDNA of untransfected HCT116 (first lane), HCT116 transfected with a LoxP-CDH13 construct and a Cre recombinase (second lane), and a no gDNA negative contol (third lane). The correct band size is 599bp and is labeled with an asterisk.
B. A PCR gel picture using primers to amplify the β-globin-BUB1 junction from gDNA of untransfected HCT116 (first lane), HCT116 transfected with a LoxP-BUB1 construct and a Cre recombinase (second lane), and a no gDNA negative contol (third lane). The correct band size is 342bp and is labeled with an asterisk.
C. Sequence traces showing the β-globin-Puro/TK cassette junction in the 2E8 HCT116 clone.
D. Sequence traces showing the β-globin-*CDH13* junction (top panel) and the β-globin-*BUB1* junction (lower panel).
The sequences of CGIs, LoxP511, and the flanking genomic region are outlined with yellow, blue and red boxes, respectively.
The nucleotides between the boxes are additional bases that were included during cloning the constructs.

achieved by communicating activated GCV through cell-cell junctions and/or its release into the medium. The degree of this effect varies substantially between cell types. Therefore, I needed to test whether the HCT116 cells display a high level of bystander effect. If bystander effect is indeed impeding isolating clonal cell lines then the selection could potentially be adjusted to overcome this effect.

Equal ratios of HCT116 (TK-negative) and 2E8 (TK-positive) were mixed (figure 4.8A). The representation of TK gene in this mixed population resembles that of cells with 50% RMCE efficiency. Following that, $4X10^6$ cells of 2E8: HCT116, HCT116, and 2E8 were seeded in 10cm plates. 24 hrs later the cells were subjected to GCV selection. Next, the cells were diluted 1:2000 and seeded to form clones (figure 4.8B). The HCT116 cells survived the GCV selection and formed colonies. If there was no bystander effect, it would be expected that 2E8: HCT116 population should give as much as half the number of the clones arising from the HCT116 cells. However, most of the cells in 2E8: HCT116, and 2E8 died (figure 4.8B and D). This shows that 50% TK-positive cells were enough to kill the remaining 50% TK-negative cells, indicative of high bystander effect in HCT116. Consequently, a new selection method should be used to establish clonal cell lines.

4.5.1 RMCE efficiency is very low in HCT116

If colonies were picked, expanded in separate wells, and then subjected to GCV selection, colonies would bypass the bystander effect. Therefore, I picked colonies and expanded them in two 96-well plates. Five colonies survived GCV selection. Nevertheless, PCR amplification of the genome-integrant junction showed that the five colonies were false positives. This suggests that RMCE is lower than 2.5% in HCT116.

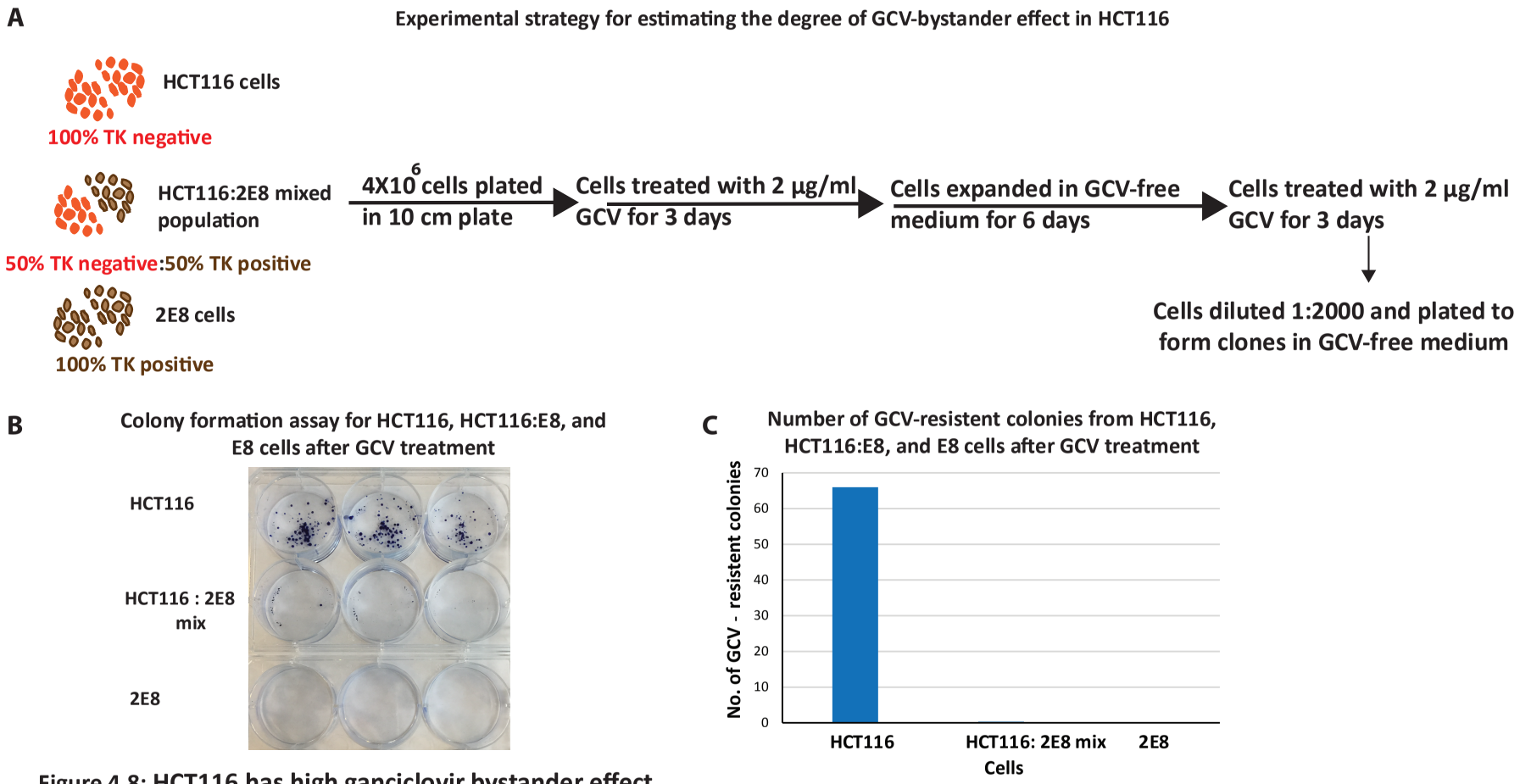


Figure 4.8: HCT116 has high ganciclovir bystander effect.

A. Experimental strategy for testing GCV-bystander effect in HCT116.

B. Colony formation assay of HCT116, HCT116:2E8, and 2E8 cells. After GCV treatment, Cells were diluted 1:2000 and plated to in trplicates in 6-wells plate to form clones in a GCV-free medium.

C. Number of ganciclovir resistant colonies of HCT116, HCT116:2E8 mix, and 2E8 cells in the colony formation assay.

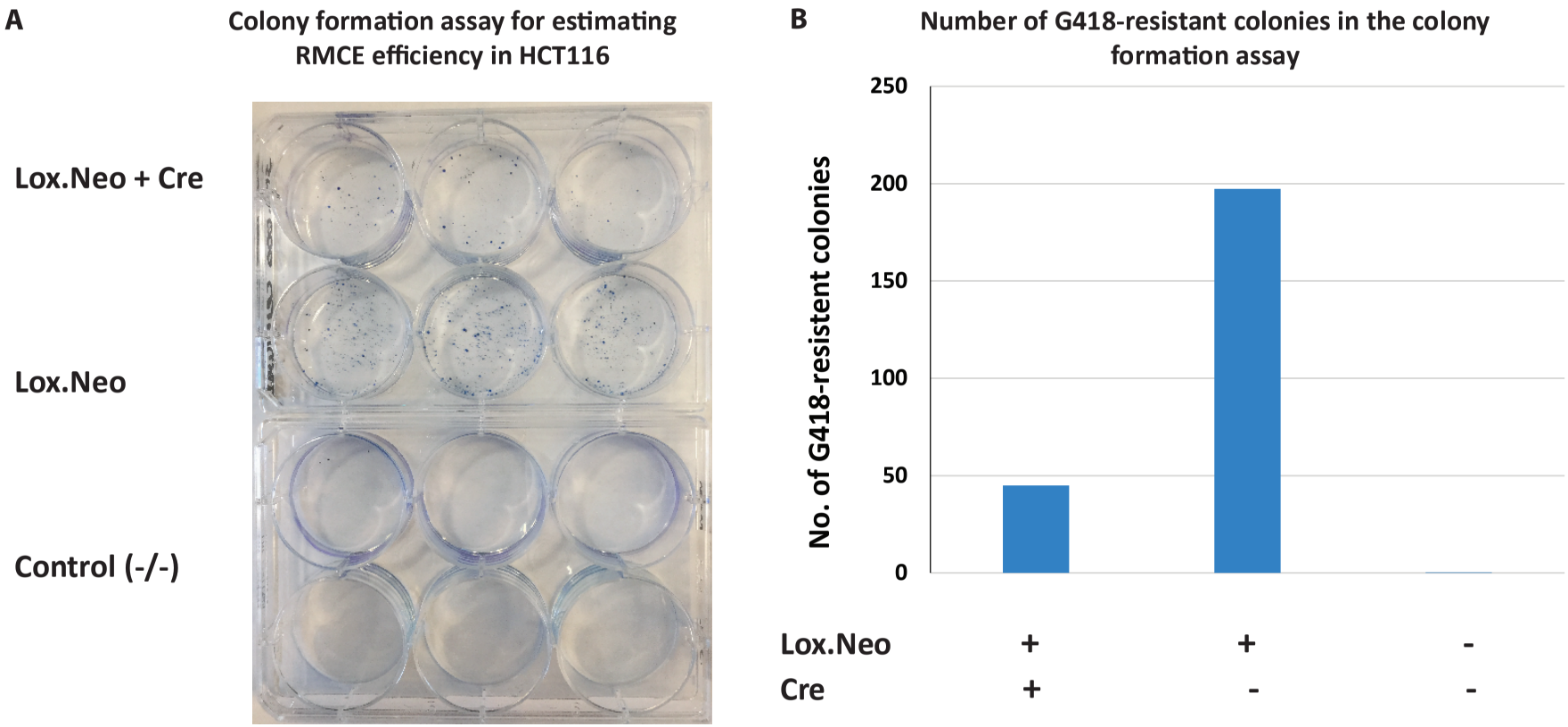


Figure 4.9: RMCE efficiency is very low in HCT116

A. Colony formation assay to estimate RMCE frequency in HCT116. 2E8 clonal cell line (30,000 cells/well) was transfected with a vector containing neomycin-resistant gene flanked by inverted LoxP511 , with and without Cre recombinase vecto. Cells were selected with G418 for 2 weeks.

B. Bar grapg showing number of G418 resistant clones in the colony formation assay.

In order to determine RMCE efficiency in HCT116, a colony formation assay under positive selection is required. HCT116 were transfected with a construct containing neomycin-resistance gene flanked by inverted LoxP sequences (LoxP-Neo) with and without Cre (figure 4.9). G418 selection was administrated two days after transfection. The colony formation assay indicated that the efficiency of RMCE in HCT116 is very low (~.007%) (figure 4.9B). The low efficiency of Cre-mediated RMCE in HCT116 makes generating a clonal cell line with integrated CGIs technically infeasible.

4.6 Discussion

In this chapter, I aimed to verify that the sequence of the CGIs doesn't program their aberrant hypermethylation in cancer by targeted integration of CGIs in a predetermined locus. RMCE was chosen to perform targeted integration because it is has been widely used for this purpose in ES cells and other cell types.

This approach involved two main steps. In the first step, HCT116 cell line with integrated RMCE cassette at the β-globin locus was successfully generated. The intact integration was verified by sequencing and southern blotting analysis. At the second step, integration of CGIs at the β-globin locus by RMCE was confirmed by sequencing the genome-CGI junction. Nevertheless, isolating a clonal cell line was infeasible due to two technical issues; GCV-bystander effect, and the low efficiency of RMCE in HCT116.

An alternative negative selection method that can be tested in HCT116, is using Casp8-FKBP fusion gene (Pajvani *et al.*, 2005; Baker *et al.*, 2011). Casp8 is an

inactive proenzyme, and when fused to Phe36Val mutant FKBP domain it homodimerizes and induces cell apoptosis upon administration of a synthetic drug AP20187. Most importantly, I did not find any report of the “Bystander effect” using this strategy. I designed and ordered a cassette containing Casp8-FKBP, however time did not allow testing it.

A previous study reported generating clonal HCT116 cell line using RMCE strategy (Lee *et al.*, 2013). However, they used Flp-recombination system rather than Cre-recombination. No details about the efficiency of Flp-recombination in HCT116 were mentioned. In a follow up study by the same group, one HCT116 clone was isolated for GCV resistance, but it is unknown how this was done and how many colonies they screened (Ballikaya *et al.*, 2014). The results of these studies suggests that creating a HCT116 clonal cell line by Flp-recombination might be feasible.

Alternatively, the phage integrases Bxb1 and phiC31 can be used to mediate efficient unidirectional recombination between each enzyme own attP and attB sites with 100% recombination specificity after screening and selection (Zhu *et al.*, 2014).

Chapter 5: Aberrantly methylated CGIs are not primary targets of DNMT3B

In chapter 3, I have shown that when the sequences of aberrantly methylated CGIs are integrated into ectopic loci, they rarely gain methylation. This suggests that aberrantly methylated CGIs doesn't program their aberrant hypermethylation in HCT116. However, in that experiment it is possible that sequences outside the integrated CGIs contain binding site for factors (e.g., ZNF403) that can recruit DNMTs. In this chapter, I assess the targeting of DNMT3B to aberrantly methylated CGIs on a genome wide level at their native loci.

5.1 Experimental design

To globally study whether aberrantly methylated CGIs are targeted by a *de novo* DNMTs, a model system is required where the genome is hypomethylated at the starting point following which methylation is established and monitored. One way this can be achieved is by introducing the *de novo* DNMT enzyme of interest into a DNMT deficient cell line and ask where methylation is gained. Therefore, I have stably restored DNMT3B expression in DNMT1/DNMT3B double knockout (DKO) of HCT116 cells (figure 5.1A). To assess methylation gain in this system, single-base resolution DNA methylation profiling was generated using reduced representation bisulfite sequencing (RRBS).

5.2 Generation of HCT116 model to track DNMT3B2 targeting

The main DNMT enzymes expressed in HCT116 are the maintenance methylation enzyme DNMT1, and the *de novo* methylation enzyme DNMT3B

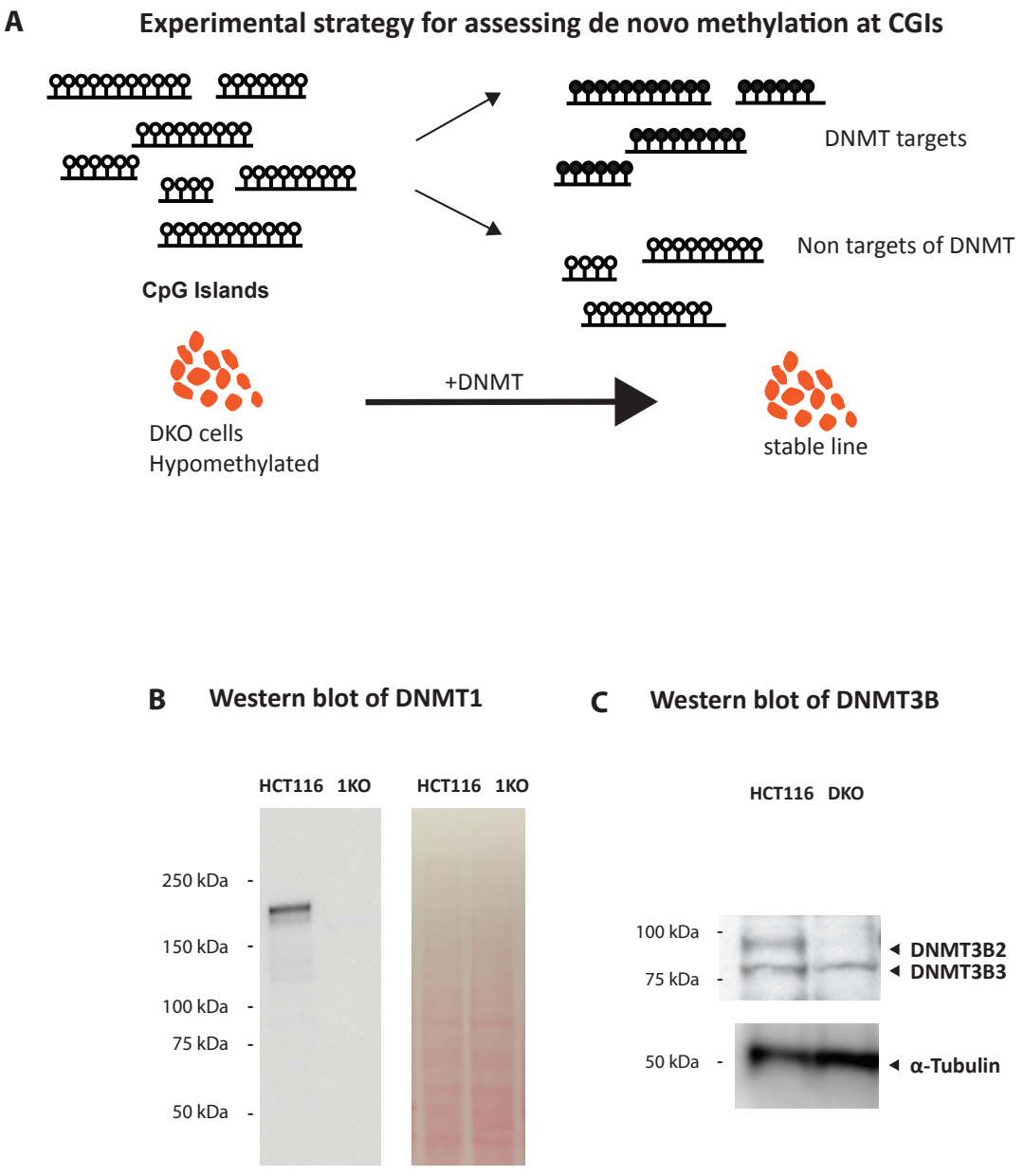
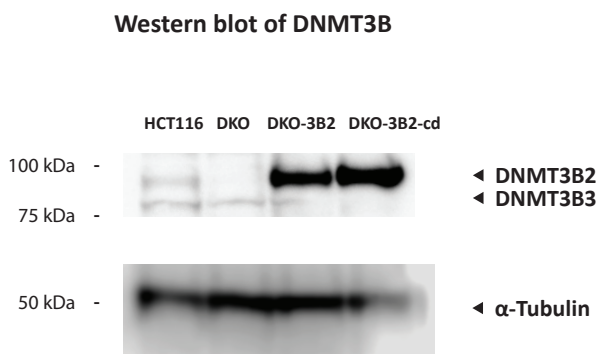


Figure 5.1:
A. Experimental strategy to investigate *de novo* methylation at CGIs. DNMT expression was restored in a hypomethylated cell line (DKO) cells then CGIs targeted by the introduced DNMT were identified.
B/C. The main DNMTs expressed in HCT116 are DNMT1 and DNMT3B.
B. Western blot for DNMT1 in HCT116 and 1KO cells. DNMT1 is seen at 184KD in HCT116 and is absent in 1KO cells. Ponceau staining of the blot is also shown (left). Western blot was performed by Hazel Davidson-Smith.
C. Western blot for DNMT3B in HCT116 and DKO cells. The band at 97KD represents DNMT3B-2. Lower faint band represents DNMT3B-3 (identified by Portia McGhan - from RT-PCR and RNA-seq data for HCT116 cells). The band at 52 kDa represents α-Tubulin and serves as a loading control. Western blot was performed by Cristina Rubio.

(figure 5.2). DNMT3B isoforms expressed in HCT116 are the somatic catalytically active isoform DNMT3B2 and the in-active isoform DNMT3B3 (figure 5.2B) (Weisenberger *et al.*, 2004). DKO (DNMT1−/−, DNMT3B−/−) HCT116 cells express hypomorphic DNMT1 (Egger *et al.*, 2006). Global methylation levels of our DKO cells are ~ 43% less than the wild type HCT116 as measured by mass spectrometry. A previous report by the group that derived the DKO cell line stated global methylation levels of DKOs measured by reversed-phase HPLC are ~ 58% less than the wild type HCT116 (Rhee *et al.*, 2002).

I stably expressed DNMT3B2 in DKO cells using a piggyBac construct containing the cDNA of the isoform and a puromycin resistance gene. A polyclonal population of DKO cells stably expressing DNMT3B2 was selected for puromycin resistance. The population was expanded for 30 days under puromycin selection in order to get sufficient cells for analysis. DNMT3B2 expression was detected by western (figure 5.2A). Compared to wild type cells, mass spectrometry for DKO cells expressing DNMT3B2 shows that 20% of the wild type methylation levels are restored (figure 5.2B). The global level of methylation in DKO cells expressing DNMT3B2 is similar to global methylation levels of DNMT1 knockout (1KO) cells. This indicates that integrated DNMT3B2 is *de novo* methylating the genome of DKOs, suggesting that this system can be used as a model to ask which CGIs it targets.

A



B

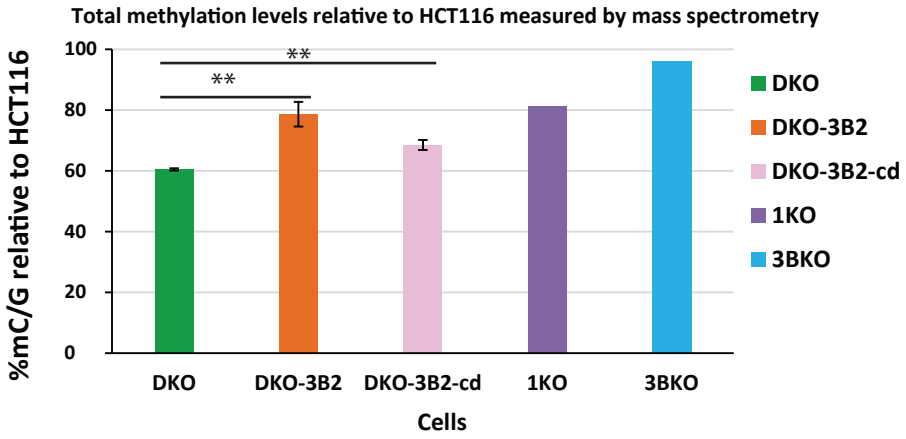


Figure 5.2: Stable expression of exogenous DNMT3B2 in DKO cells restores~ 20% of the wild type total methylation levels.

A. Western blot for DNMT3B in HCT116, DKO, DKO cells with exogenous DNMT3B2, DKO cells with exogenous catalytic-dead DNMT3B2. The band at 97KD represents DNMT3B-2. Lower faint band represents DNMT3B-3 (identified by Portia McGhan - from RT-PCR and RNA-seq data for HCT116 cells). The band at 52 kDa represents α-Tubulin and serves as a loading control. Western blot was performed by Cristina Rubio.

B. Bar graph showing total methylation levels measured by mass spectroscopy (y-axis) in DKO, DKO cells with exogenous DNMT3B2, DKO cells with exogenous catalytic-dead DNMT3B2, and HCT116 cells (x-axis). ** P <0.01, Student's t-Test.

5.3 Assessing methylation changes by RRBS

CGIs are the main focus of my study. Reduced Representation Bisulphite Sequencing (RRBS), which is designed to enrich for CpG rich regions, was therefore selected to assess methylation changes at CGIs (Smith *et al.*, 2009). RRBS enrichment at CpG rich DNA fragments is achieved by digesting gDNA with a methylation insensitive restriction enzyme specific to a CpG containing motif, followed by selecting small DNA fragments. Consequently, this enrichment yields a high coverage for CGIs at lower sequencing depth, compared to whole genome sequencing methods. Unlike bisulfite treatment-based methylation microarray methods, RRBS covers more CpGs within a CGI, providing more accurate read out of its methylation status.

RRBS data for 5 samples was generated, with at least 25 X mean coverage for a mean of 9143298 CpGs in each sample. Table 5.1 shows the alignment statistics of RRBS data for the five cell lines.

The subset of CpGs overlapping CGIs was identified, this covered 24542 CGIs out of 25472 CGIs identified using CXXC Affinity Purification plus deep sequencing (CAP-seq) (Illingworth *et al.*, 2010). Weighted mean methylation of CGIs was calculated by dividing the number of methylated CpGs in that CGI over total number of CpGs.

Mean methylation of CGI = ∑ methylated CpGs reads ÷ ∑ Total CpGs reads.

By calculating the weighted mean methylation of CGIs, the contribution of each CpG to the mean methylation is based on its coverage. After setting a threshold for CGIs to have ≥ 20 coverage, we obtained 17947 CGIs. Out of these, 7287 CGIs were methylated in HCT116 (>=50% methylation).

RRBS data allignment statistics for five cell lines

Sample	Total Reads (millions)	Mean CG coverage	Reads After Duplicate UMI (millions)	CG Observed (millions)	% CG Observed	CG % Meth	CHG % Meth	CHH % Meth	Conversion effeciency
DKO	56.164	26.535	30.605	4.583	16.45%	30.56%	0.50%	0.32%	99.9969
DKO_3B2	56.12	29.105	32.214	4.7	16.87%	43.38%	0.58%	0.39%	99.9966
DKO_3B2_cd	52.628	25.186	28.31	4.499	16.15%	36.22%	0.58%	0.41%	99.9962
1KO	61.574	28.233	33.503	4.677	16.79%	49.34%	0.56%	0.36%	99.9966
3BKO	48.41	25.042	27.879	4.651	16.70%	58.83%	0.58%	0.38%	99.9966
HCT116	57.98	26.826	31.217	4.564	16.38%	60.58%	0.59%	0.37%	99.9967

Table 5.1: RRBS data allignment statistics for five cell lines.
Table showing number of total RRBS reads, mean CG coverage, reads passing quality control, alligned reads CG observed and conversion effeciency. Conversion effeciency is calculated as 1-Lambda Meth CG (%).

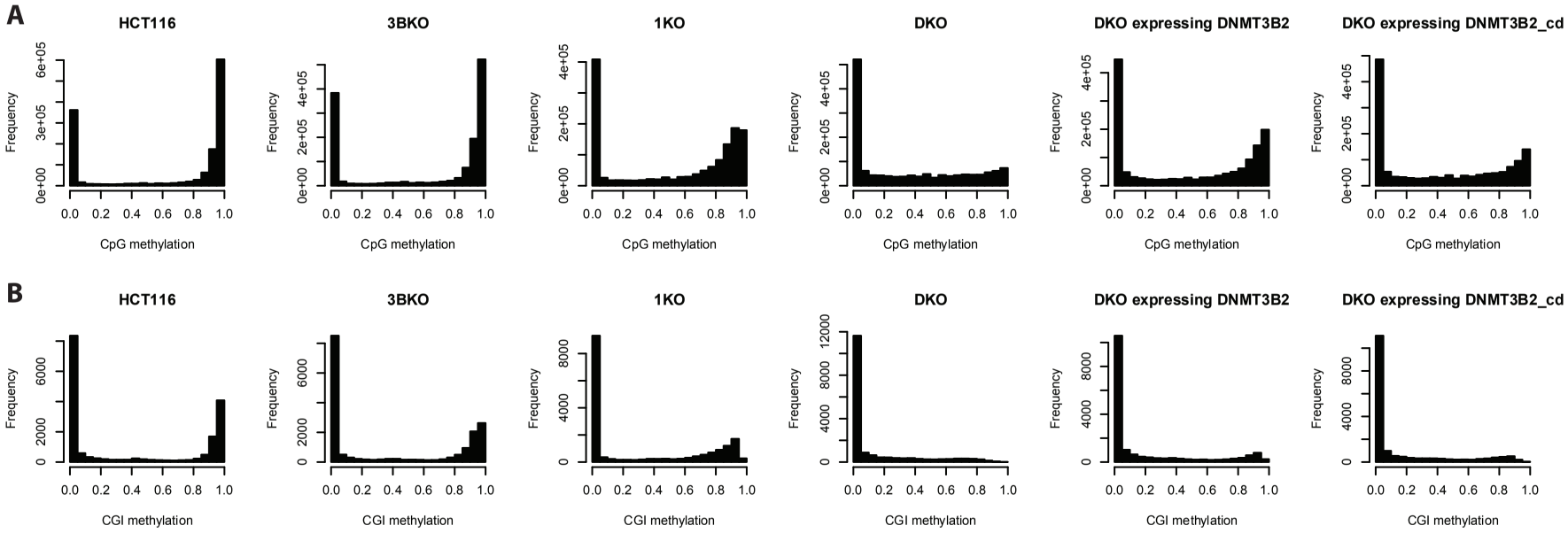


Figure 5.3: Visualizing RRBS data as methylation histograms for five cell lines.
A. Histograms for CpGs methylation in five cell lines. Only CpGs with greater than 5 reads coverage were included.
B. Histograms for CGIs methylation in five cell lines. Only CGIs with greater than 20 reads coverage were included.

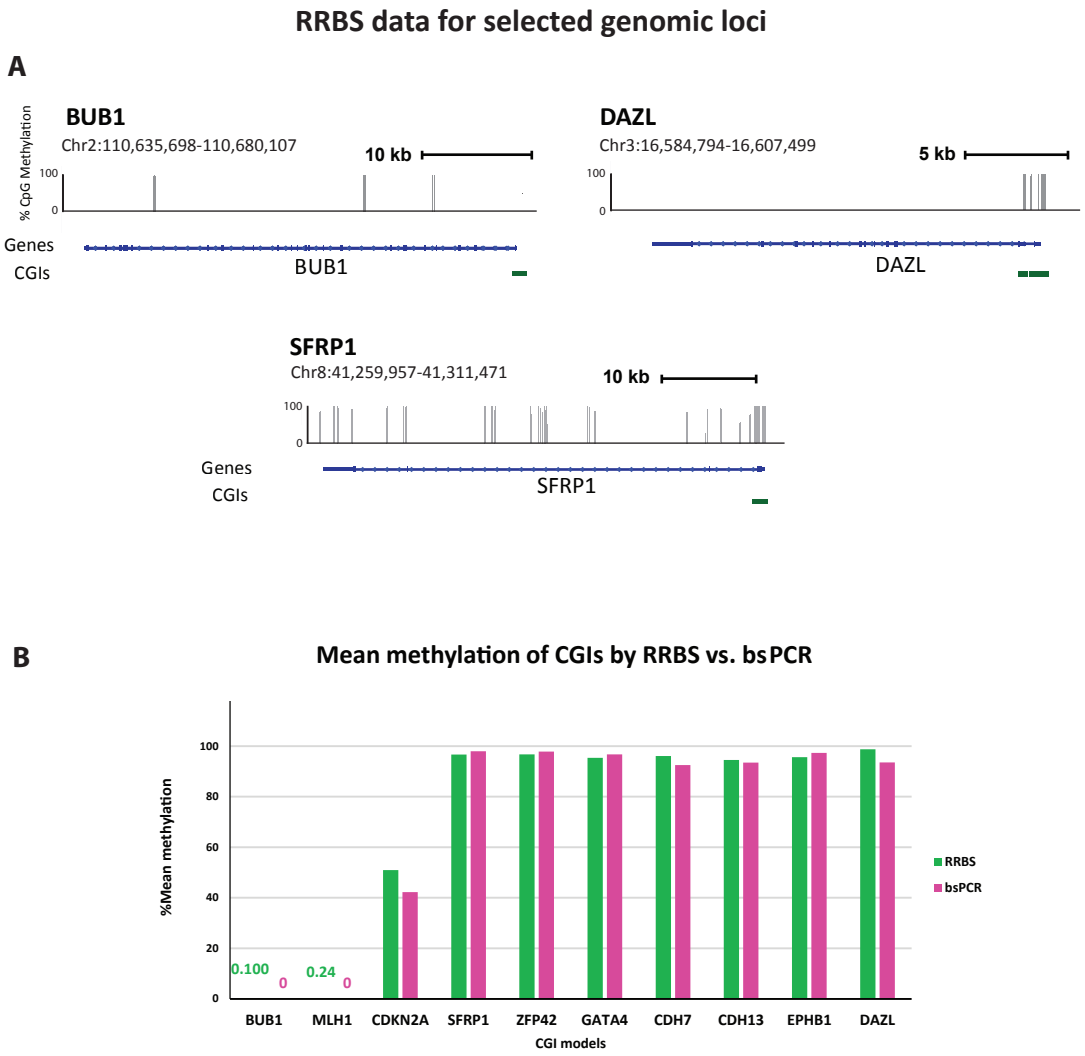


Figure 5.4: Mean CGIs methylation as measured by RRBS is consistent with BS PCR for representative CGIs.

A. Browser tracks of RRBS data at selected genomic loci

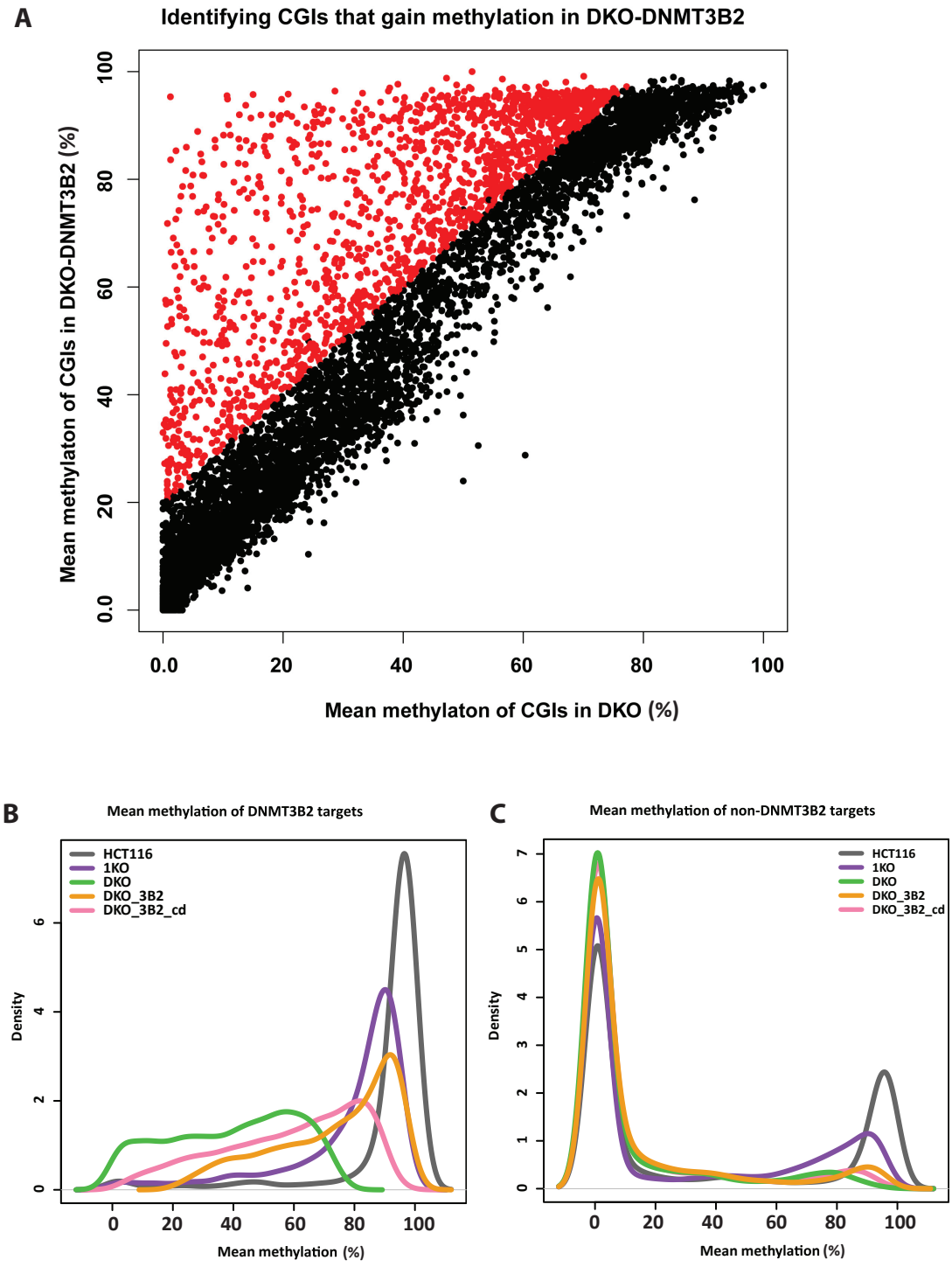
RRBS methylation data in HCT116 reflects the correct methylation states of three selected CGIs; BUB1 (unmethylated), Dazl (methylated) and SFRP1 (methylated). Genes are shown in blue. CGIs are shown as dark green bars. The reference genome used is human genome build hg38/GRCh38.

B. % Mean methylation of CGI representatives as measured by RRBS (green) and bsPCR from chapter 3 (pink).

Visualizing our RRBS as histograms of CpGs methylation (figure 5.3A) or CGI methylation (figure 5.3B) illustrates the hypermethylation of HCT116 compared to the hypomethylated DKO cells. Additionally, it shows the gain of methylation in DKO cells expressing DNMT3B2. The fidelity of our RRBS data was tested by examining the methylation status at CGIs with known methylation status. For example, RRBS reflected the unmethylated state of BUB1, and the methylation of DAZL and SFRP1 (figure 5.4A and B).

5.4 A subgroup of CGIs in DKOs gained methylation after expressing DNMT3B2

Whether all CGIs methylated in HCT116 will equally gain methylation after stably re-introducing DNMT3B2 in DKO cells or a subset of CGIs will be preferentially targeted by DNMT3B2 is unknown. To identify which CGIs are significantly gaining methylation following DNMT3B2 rescue, the probability that there is no significant change in CGI methylation after expressing DNMT3B2 was calculated by performing Fisher's exact test on the raw methylation counts summed across all CpGs within each CGI. To exclude any p-values that might by chance be less than 0.05, multiple testing correction using Benjamini & Hochberg method was performed (Benjamini and Hochberg, 1995). 48% of the CGIs were identified as significantly gaining methylation (FDR<0.05). However, many of these CGIs showed a small gain in methylation. Therefore, from this set of CGIs only those showing gain in methylation greater than 20% were identified as primary targets of DNMT3B (figure 5.5A).This subset contains 1653 CGIs (~9% of all CGIs in the analysis) with 29.3% median methylation gain. Figure 5.5B shows that these CGIs are hypomethylated in DKO and are recovering methylation upon DNMT3B2 expression, close to their methylation levels in 1KO cells. This subset of CGIs are also gaining methylation, albeit to a lesser extent, in DKOs expressing DNMT3B2-cd (17% median methylation gain). All non-DNMT3B targets are not showing substantial change in methylation in DKO cells



5.5: A subset of CGIs gain methylation after DNMT3B2 expression in DKOs.

A. Scatter plot comparing mean methylation of 24,542 CGIs in DKO cells expressing DNMT3B2 (y-axis) versus DKO cells (x-axis). 1653 CGIs (red) gained >20% methylation.

Density histograms showing CGIs methylation in DKO (green line), DKO expressing ectopic DNMT3B2 (orange line), DKO expressing ectopic DNMT3B2-cd (pink line), 1KO (purple line), and HCT116 (grey line) for DNMT3B2 targets (B) and non-DNMT3B2 targets (C). Band width used= 0.04.

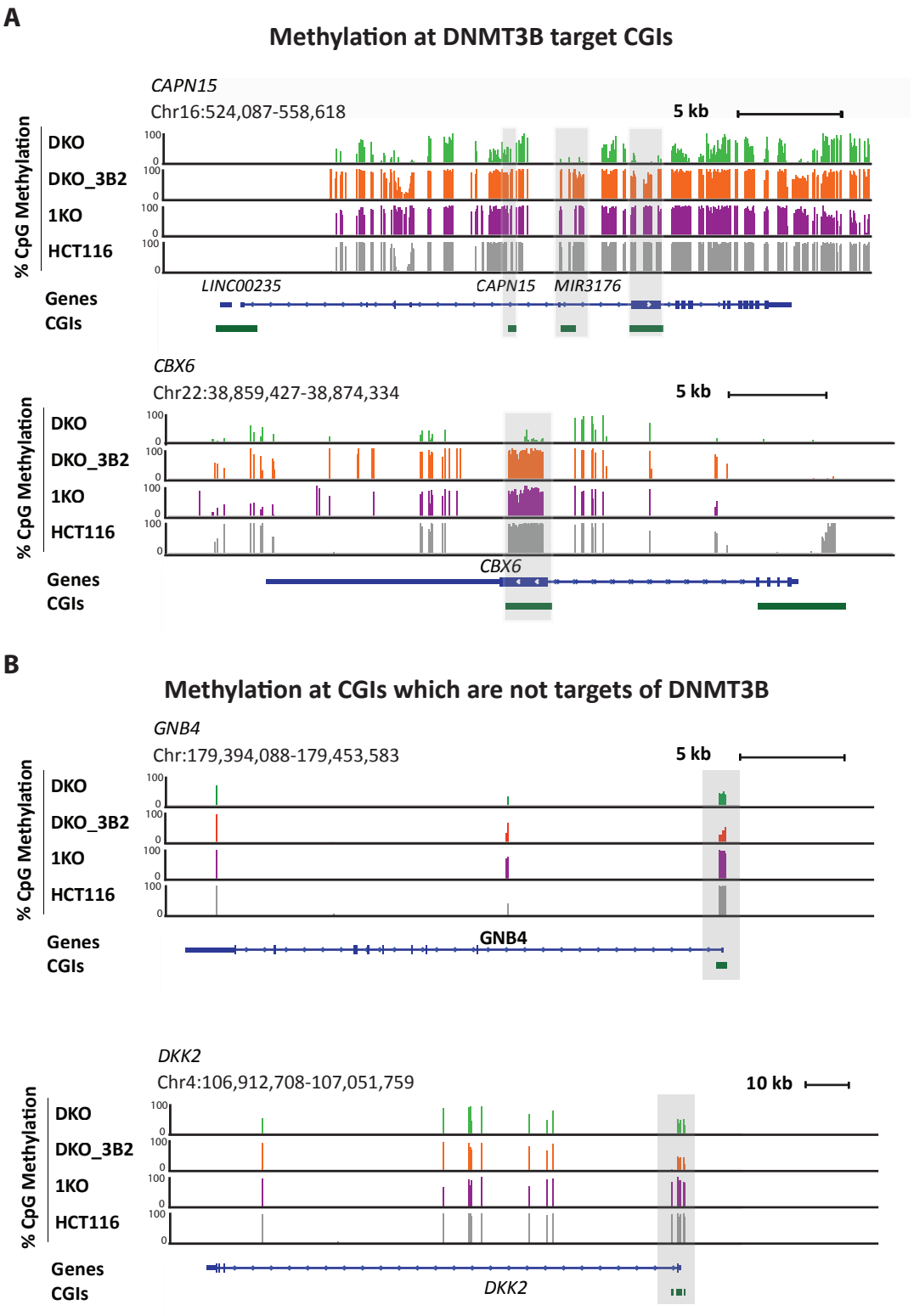


Figure 5.6: Examples of CGIs targeted and non targeted by DNMT3B2.
Browser plots showing % CpG methylation as measured by RRBS at DNMT3B2 target CGIs (A) and non-targets (B) (highlighted in grey) in DKO (light green), DKO expressing DNMT3B2 (orange), 1KO (purple), and HCT116 (grey). Genes are shown in blue. CGIs are shown as dark green bars. The reference genome used is human genome build hg38/GRCh38.

expressing either DNMT3B2 or DNMT3B2-cd (0.42% and 0.25% median methylation gain, respectively) (figure 5.5C). Examples of CGIs targeted by DNMT3B2 and non DNMT3B targets are shown in figure 5.6.

5.5 CGIs overlapping H3K36me3 gain higher methylation after expressing DNMT3B2 in DKO cells

Previous studies have reported links between certain histone marks and DNA methylation. I therefore wanted to examine if our model reproduces these links calculating the enrichment of different histone peaks at primary DNMT3B targets. To achieve this, I identified how many primary DNMT3B2 targets are overlapping histone peaks in HCT116 cells from ENCODE data (ENCODE Project Consortium, 2012). Next, the fold enrichment was calculated by dividing the overlap counts compared to a background expected by chance (all CGIs in the analysis) (figure 5.7A). Statistical testing was performed by Fisher's exact test on the number of CGIs overlapping a certain histone mark versus the number of CGIs non-overlapping that mark.

As shown in figure 5.7A, DNMT3B2 targeted CGIs are ~ 4-fold enriched in H3K36me3 peaks. To visualize DNA methylation gain at CGIs overlapping H3K36me3 the change in mean methylation after re-expressing DNMT3B2 was plotted for CGIs overlapping and non-overlapping H3K36me3 (figure 5.7B). Notably, the mean methylation change at CGIs overlapping H3K36me3 is ~ 4.2 fold higher compared to CGIs non-overlapping H3K36me3 (Student's t-Test, p-value < $2.2X10^{-16}$). This result is in agreement with previous studies suggesting that DNMT3B interacts with H3K36me3 mark, a histone mark enriched in actively transcribed gene bodies (Baubec et al., 2015). Notably, most CGIs overlapping H3K36me3 were hypomethylated in DKO cells compared to HCT116 and 1KO, and their methylation levels reaches similar levels as in 1KO after expressing DNMT3B2 (figure 5.7C). Figure 5.8 shows two examples of

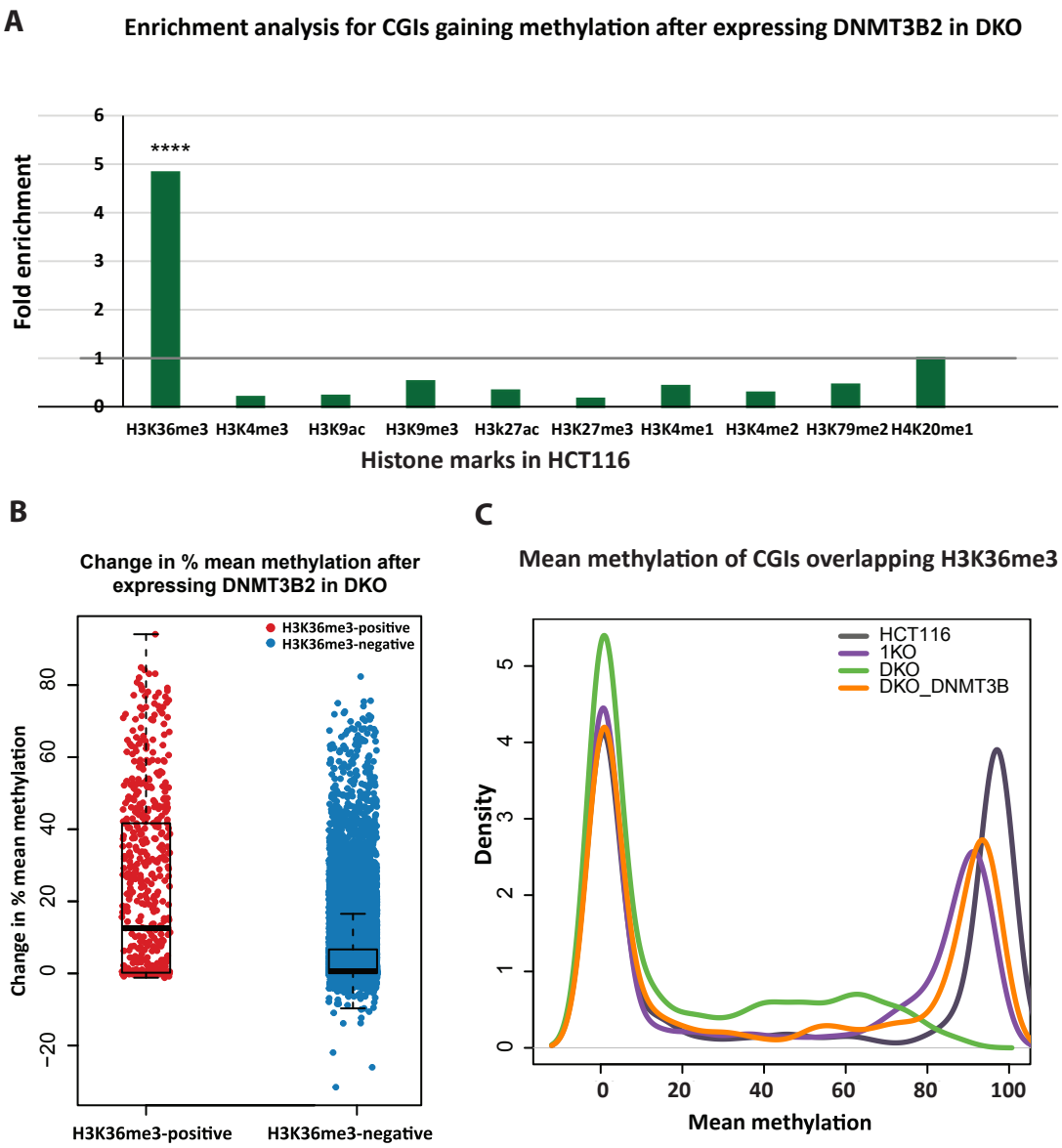


Figure 5.7: CGIs overlapping H3K36me3 peaks gain high methylation upon DNMT3B2 expression in DKO cells.

A. Bar graph showing the fold enrichment of CGIs gaining methylation by DNMT3B2 (>=20% change in mean methylation) (y-axis) at histone peaks identified in HCT116 (x-axis). CGIs overlapping H3K36me3 peaks are ~4- folds enriched within CGIs gaining methylation by DNMT3B-2. **** P<0.0001, Fisher's exact test.

B. Box plots showing % change in mean methylation at CGIs overlapping H3K36me3 (red) compared to CGIs not overlapping H3K36me3 (blue) after restoring DNMT3B2 expression in DKO cells. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the % change in mean methylation at a single CGI. Horizontal line within the box represents median.

C. Density histogram of methylation of CGIs overlapping H3K36me3 in DKO (green), DKO expressing ectopic DNMT3B2 (orange), 1KO (purple), and HCT116 (grey). Band width used is 0.04.

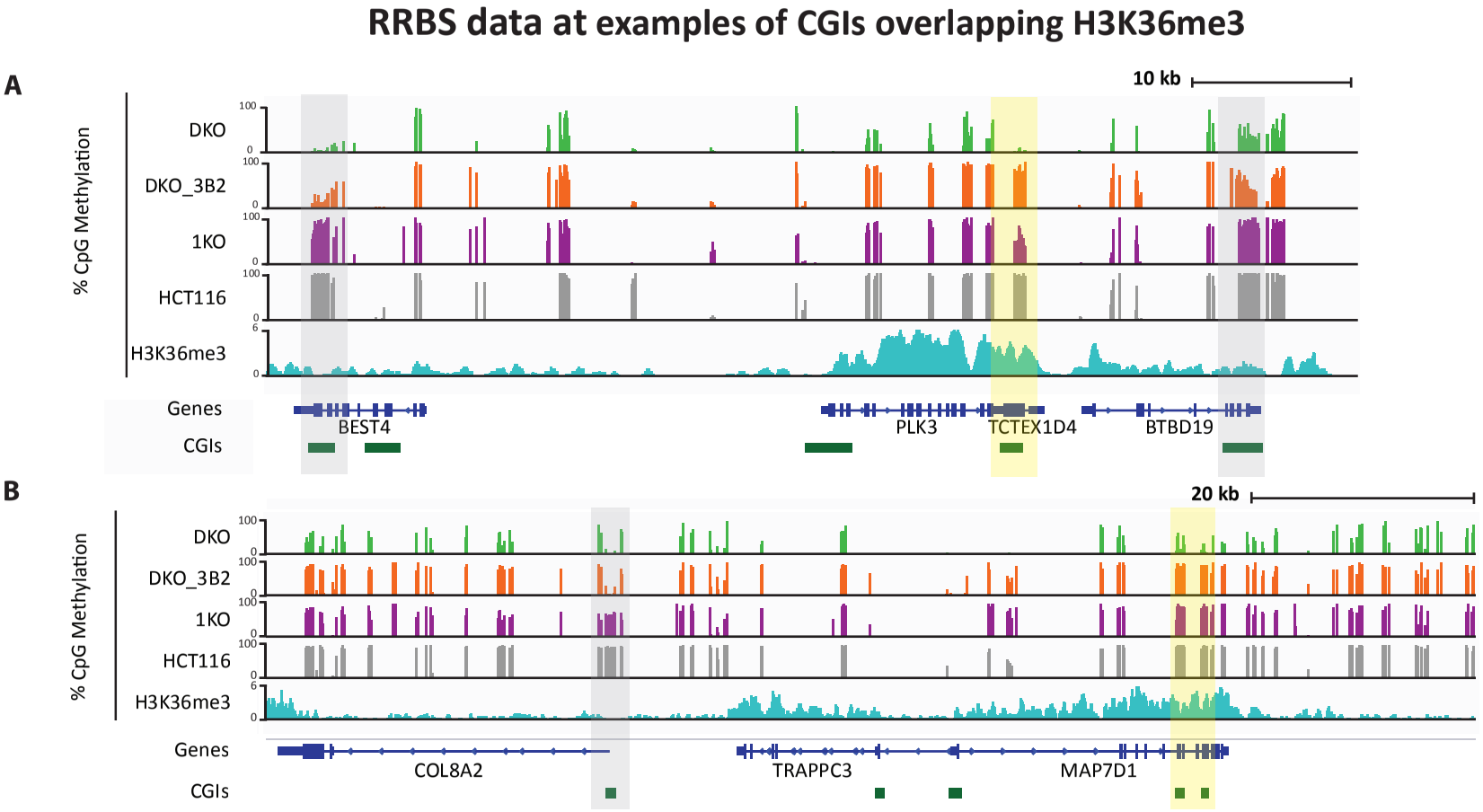


Figure 5.8: High methylation gain at CGIs overlapping H3K36me3.
% CpG methylation as measured by RRBS at two different genomic loci containing CGIs marked by H3K36me3 (highlighted in yellow) and CGIs nonoverlapping H3K36me3 (highlighted in grey) in DKO (light green), DKO expressing DNMT3B2 (orange), 1KO (purple) and HCT116 (grey).
H3K36me3 ChIP-seq data in HCT116 from ENCODE (ENCODE Project Consortium, 2012)**. Genes are shown in blue. CGIs are shown as dark green bars** (Illingworth *et al.*, 2010)**. The reference genome used is hg38/GRCh38.**

genomic loci containing CGIs overlapping and non-overlapping H3K36me3. The CGIs overlapping H3K36me3 (highlighted yellow) are gaining more methylation upon expressing DNMT3B2 compared to CGIs non-overlapping H3K36me3 (highlighted in grey).

Moreover, DNMT3B2 mediated methylation was found significantly depleted at CGIs overlapping histone marks known to be associated with active gene promoters and enhancers (e.g.H3K4me3, H3K4me2, and H3K27ac) and at polycomb repressed CGIs marked by H3K27me3 in HCT116. It is previously known that the presence of one or more methyl groups on H3K4 hinders the binding of the ADD domain in DNMT3A, and presumably DNMT3B, to the H3 tail at lysine 4 (Otani *et al.*, 2009). Collectively, our data shows that this system successfully reproduces results previously reported in the literature suggesting that it is suitable as a model for testing whether aberrantly methylated CGIs in cancer are highly targeted by *de novo* methylation.

5.6 Aberrantly methylated CGIs showed less methylation compared to normally methylated CGIs

In order to assess the levels of methylation at aberrantly methylated CGIs versus CGIs normally methylated in the colon, I first went to identify normally methylated CGIs. I utilized Illumina Infinium HumanMethylation450 data for 273 primary colorectal tumor samples and 38 normal colon samples from The Cancer Genome Atlas (TCGA) (Muzny *et al.*, 2012). Normally methylated CGIs were defined as those showing more than 70% methylation in all normal colon and CRC samples, with no change in median methylation between CRC samples and normal samples. The methylation status of these CGIs is visualized in heatmaps of their methylation levels in normal colon and CRC samples (figure 5.7). Of this group, I took only CGIs with more than 70%

methylation in HCT116. This resulted in 615 CGIs identified as normally methylated CGIs in colon and HCT116.

It is known that CGIs methylated in normal cells are enriched at promoters of germline specific genes, and CGIs overlapping transcriptionally active gene bodies marked with H3K36me3 (Illingworth *et al.*, 2010; Jeziorska *et al.*, 2017). To check if the subset of CGIs identified is in line with the literature, I tested their enrichment in colonic mucosa ChromHMM annotations (Kundaje *et al.*, 2015). These ChromHMM annotations represents segmentation of the genome into different chromatin states generated by integrating six chromatin marks using Hidden Markov Model (HMM). The bar graph (figure 5.9) summarizes the results of this analysis and shows that my identified set of normally methylated CGIs is enriched for CGIs associated with transcription, genic enhancers, Zinc finger (ZNF) genes and CGIs in heterochromatin.

According to ChromHMM annotations, the genic enhancer state overlaps H3K36me3 and H3K4me1. Genomic regions assigned as heterochromatin in ChromHMM annotations are defined as regions enriched for H3K9me3 and depleted for the other histone marks. The quiescent state is composed of CGIs overlapping genomic segments depleted for the six histone marks. Zinc finger (ZNF) genes and repeats are identified as genomic regions overlapping both H3K36me3 and H3K9me3 marks. These enrichments are in line with characteristics known for normally methylated CGIs.

Aberrantly methylated CGIs are CGIs methylated in CRC and HCT116 but not in normal epithelial colon tissues and were identified previously (Section 3.2). Comparing the gain of methylation at CGIs from the two groups in DKO experiment shows aberrantly methylated CGIs have ~ 79.7% less median

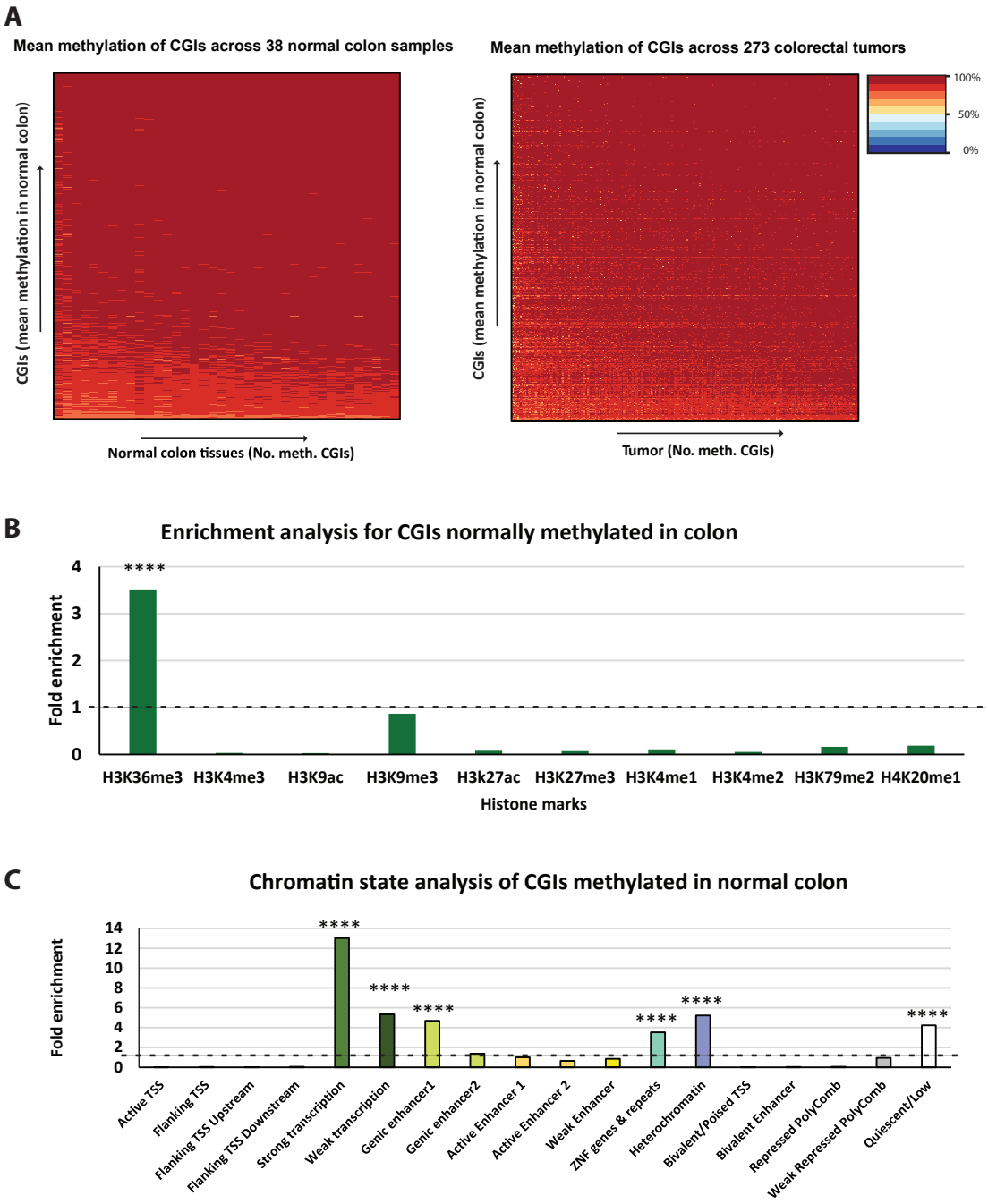


Figure 5.9: Identifying CGIs methylated in normal colon.

A. Our set of normally methylated CGIs are hypermethylated in all normal colon and colorectal tumor samples. Heat maps showing the methylation levels of normally methylated CGIs in 38 normal colon tissues (left panel) and in 273 colorectal tumor samples (right panel). Methylation of CGIs calculated from Infinium DNA methylation data. CGIs are ordered according to their frequently of methylation in normal colon samples.

B. Normally methylated CGIs are significantly enriched at H3K36me3 peaks in HCT116 cells. Shown is the fold enrichment of normally methylated CGIs at ten histone peaks in HCT116. Values of P<0.0001 are marked with four asterisks (****), Fisher's exact test.

C. Normally methylated CGIs are significantly enriched in transcription, genic enhancer, znf genes and repeats, and heterochromatin categories in normal colonic mucosa. Shown is the fold enrichment of normally methylated CGIs in H1 ChromHMM categories. **** P<0.0001, Fisher's exact test.

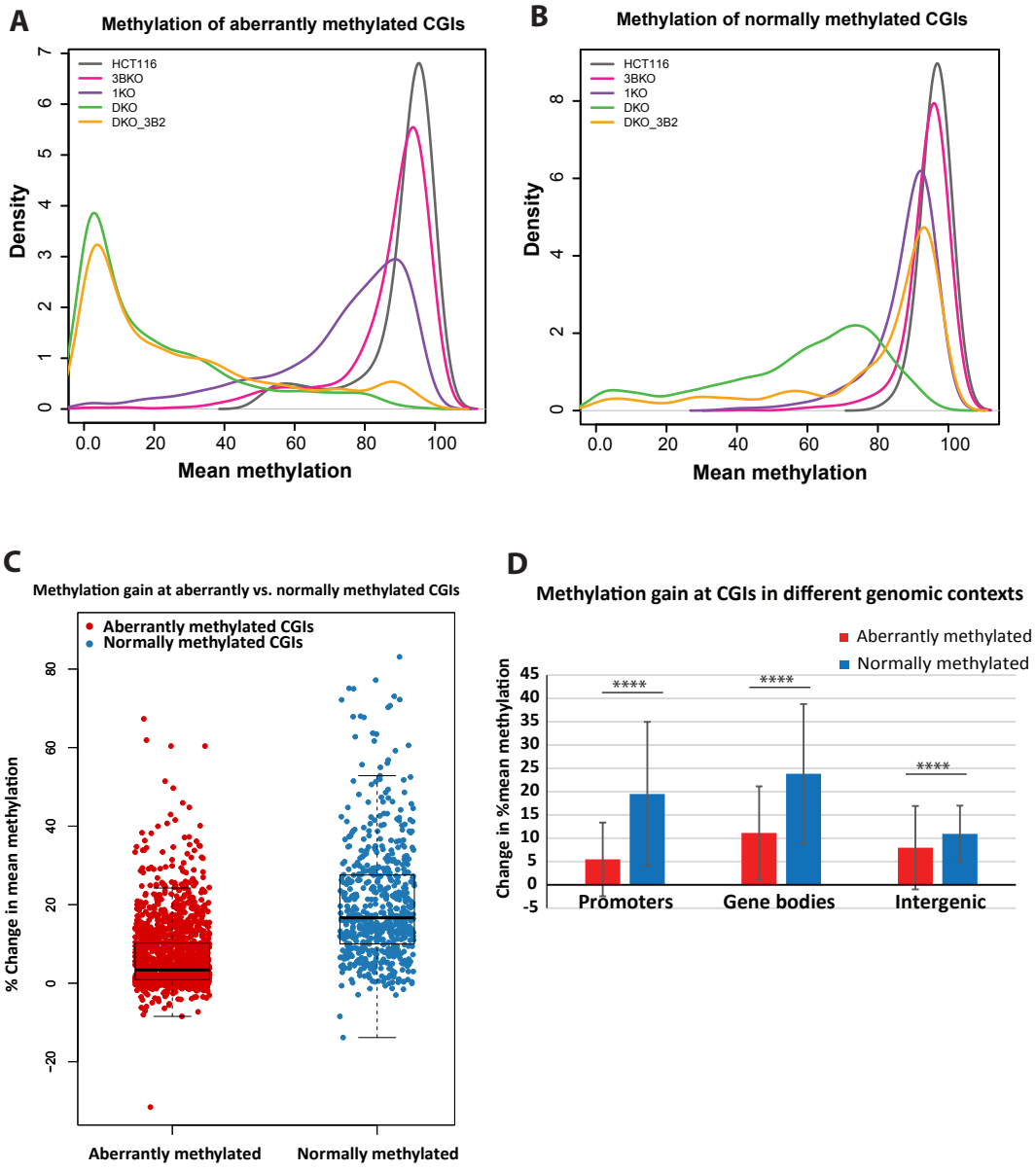


Figure 5.10: Aberrantly methylated CGIs gain less methylation than normally methylated CGIs.
Density histograms showing mean methylation in DKO (green line), DKO expressing ectopic DNMT3B2 (orange), 3BKO (pink), 1KO (purple), and HCT116 (grey) at abrrantly methylated (A) and normally methylated (B) CGIs.
C. Box plots showing % change in mean methylation of aberrantly methylated CGIs (red) compared to the normally methylated CGIs (blue) after restoring DNMT3B2 expression in DKO cells. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the % change in mean methylation at a single CGI. Horizontal line within the box represents median.
D. Bar graph showing for each category of CGIs (x-axis), the mean of %change in mean methylation (y-axis) aberrantly methylated CGIs (red) and was compared to normally methylated CGIs (blue). Error bars represents standard deviation. **** P<0.0001, Student's t-Test.

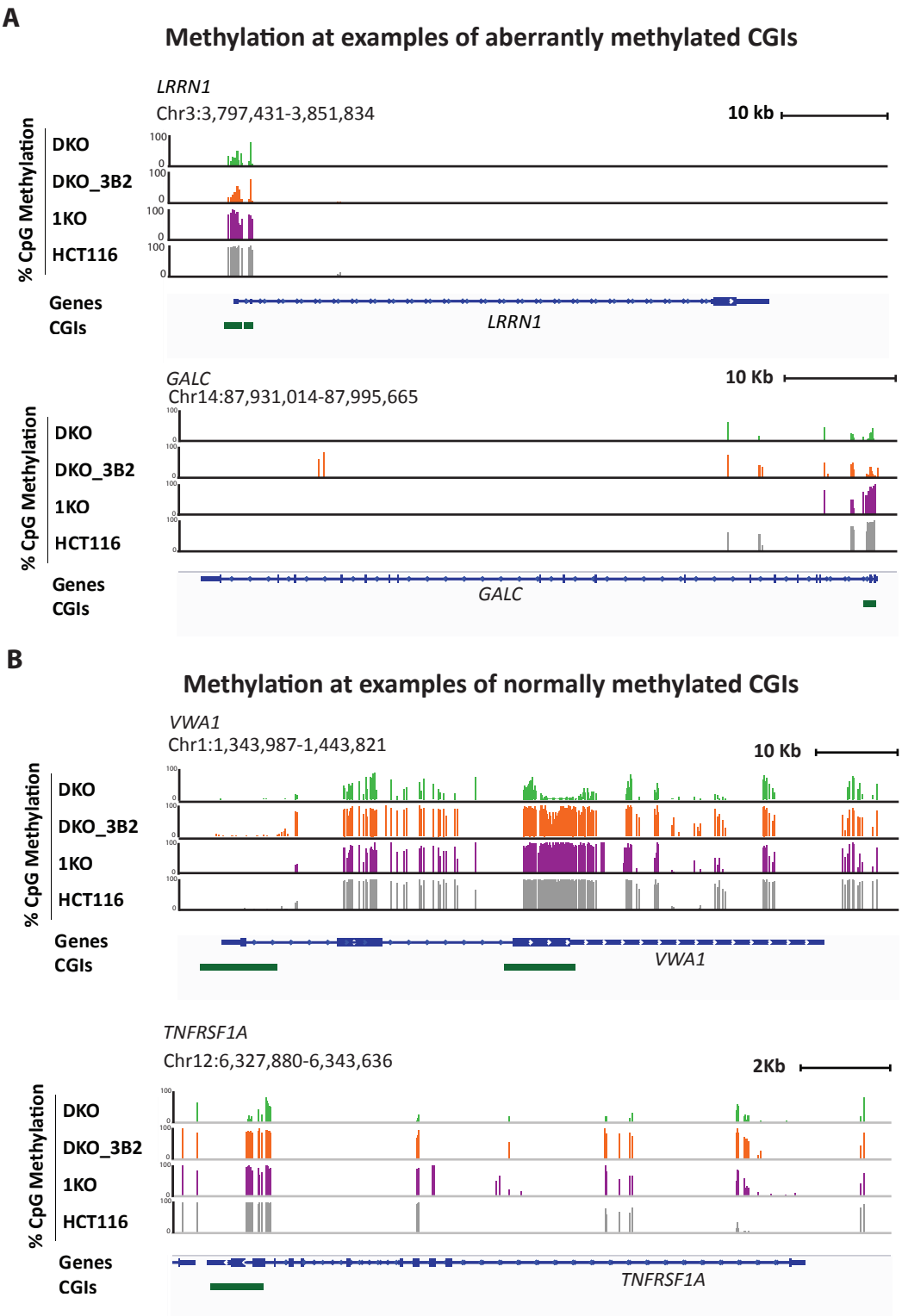


Figure 5.11: Aberrantly methylated CGIs gain less methylation than normally methylated CGIs. Browser plots showing % CpG methylation as measured by RRBS at examples of aberrantly methylated (A) and normally methylated (B) CGIs in DKO (light green), DKO expressing DNMT3B2 (orange), 1KO (purple), and HCT116 (grey). Genes are shown in blue. CGIs are shown as dark green bars. The reference genome used is hg38/GRCh38.

methylation gain compared to normally methylated CGIs (normally methylated CGIs median gain=~16.655%, aberrantly methylated CGIs median gain= ~3.376%, Student's t-Test, p-value < 2.2 $X10^{-16}$;figure 5.10A, and figure 5.11). This suggests that, lower methylation levels are targeted to aberrantly methylated CGIs compared to normally methylated ones after expressing DNMT3B2 in DKO cells.

In the absence of DNMT1, aberrantly methylated CGIs are more affected by DNMT3B loss than normally methylated CGIs suggesting a role for DNMT3B in maintaining methylation at aberrantly methylated CGIs in 1KO cells (figure 5.10A and B). However, when rescuing DNMT3B expression in the absence of DNMT1, higher gain of methylation is observed at normally methylated CGIs compared to aberrantly methylated CGIs (figure 5.10A, B, and C). This raised the question of whether methylation gain by DNMT3B is affected by the methylation levels of the CGIs. It has been reported that having sparse methylated CpGs within GSTP1 CGI combined with GSTP1 gene silencing trigger hypermethylation of the CGI in prostate cancer cells (Song *et al.*, 2002). These random methylated CpGs were proposed to act as "seeds" that trigger subsequent hypermethylation of the adjacent unmethylated CpGs. To test whether the methylation gain by DNMT3B is affected by "methylation seeding", I asked if the methylation gain by DNMT3B correlates with CGIs methylation levels in DKO cells.

In order to address this question, I stratified aberrantly methylated CGIs, normally methylated CGIs and CGIs overlapping H3K36me3 into ten groups depending on their methylation levels in DKO cells. Next, I plotted the gain of methylation by DNMT3B in each group. Figure 5.12 shows, firstly, the methylation gain by DNMT3B is proportionally correlated with the methylation level of the aberrantly methylated CGIs and normally methylated CGIs. However, for CGIs overlapping H3K36me3 the methylation gain by DNMT3B is

inversely correlated with the methylation level of these CGIs. This later observation might be due to the fact that H3K36me3-positive CGIs with methylation levels ranging from 10 to 20 gain high levels of methylation (median methylation gain =70.5). Therefore, as we move to H3K36me3-positive CGIs with higher methylation levels there is less capacity for these CGIs to gain methylation due to the total methylation limit of 100%. Secondly, figure 5.12 shows that aberrantly methylated CGIs gain less methylation by DNMT3B compared to CGIs overlapping H3K36me3 and normally methylated CGIs in most of the CGIs groups.

Overall, these results suggest that methylation levels of aberrantly methylated CGIs and normally methylated CGIs might enhance the methylation gain by DNMT3B. However, this can't totally explain the observation that aberrantly methylated CGIs are less efficiently targeted by DNMT3B compared to normally methylated CGIs and H3K36me3-positive CGIs.

Since CGIs in different contexts might be methylated through different mechanisms, I stratified the CGIs based on their genomic context (i.e. promoter, gene-body and intergenic) using Goldmine package in R and analyzed each class separately (Bhasin and Ting, 2016). Comparing aberrantly methylated CGIs to normally methylated ones showed that in all contexts methylation gain was substantially lower at aberrantly methylated CGIs (figure 5.10D). Mean methylation gain of aberrantly methylated promoters is 72% less than normally methylated promoters (Student's t-Test, p-value < $2.2X10^{-16}$). Median methylation gain of aberrantly methylated gene bodies is 53% less than for normally methylated gene-bodies (Student's t-Test, p-value < $2.2X10^{-16}$). Median methylation gain of aberrantly methylated intergenic CGIs is 27% less than for normally methylated ones (Student's t-Test, p-value = $2.525X10^{-6}$). Taken together, these results suggest that aberrantly methylated CGIs in all genomic contexts are not primary targets of DNMT3B. Notably, 11.6% of normally

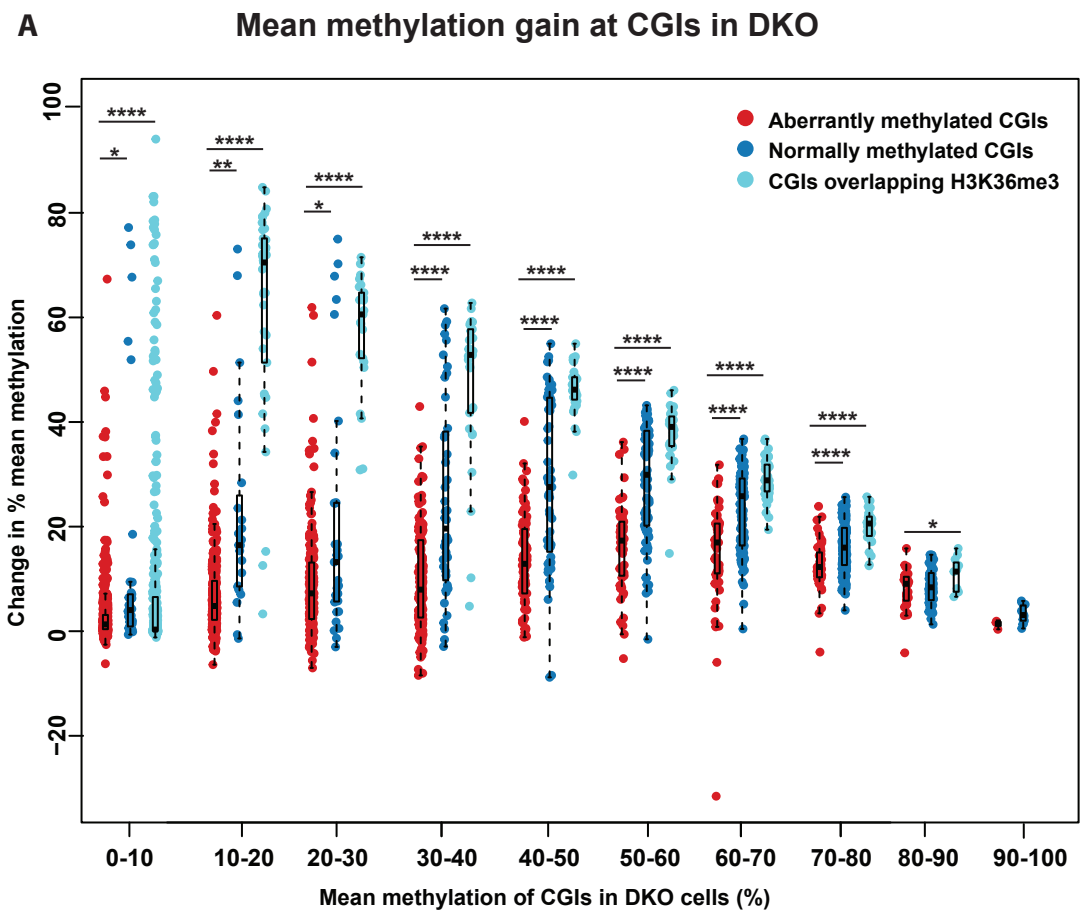


Figure 5.12: The methylation levels of aberrantly methylated CGIs and normally methylated CGIs might enhance the methylation gain by DNMT3B. Box plots showing % change in mean methylation of aberrantly methylated CGIs (red) compared to the normally methylated CGIs (blue), and CGIs overlapping H3K36me3 (cyan) after restoring DNMT3B2 expression in DKO cells. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the % change in mean methylation at a single CGI. Horizontal line within the box represents median. Error bars represents standard deviation. **** P<0.0001, ** P<0.01, * P<0.05, Mann-Whitney-Wilcoxon test.

methylated CGIs overlap H3K36me3 peaks in HCTT16, whereas only 0.6% of aberrantly methylated CGIs overlap H3K36me3 (figure 5.12A).

5.7 Discussion

DNMT3B had been implicated in causing aberrant methylation in colorectal tumors. Many previous studies investigated the correlation between DNMT3B levels, as well as other DNMTs, and levels of aberrant methylation at CGIs and reported conflicting results leaving the answer to this question unclear (Robertson, 2001; Linhart *et al.*, 2007; Ibrahim *et al.*, 2011; Kobayashi *et al.*, 2011; Joensuu *et al.*, 2015). However, in this study we are asking, are aberrantly methylated CGIs the main targets of DNMT3B in colorectal cancer? In this chapter, the DNMT3B-mediated methylation targeted to CGIs was assessed on a genome wide level. The methylation targeted to aberrantly methylated CGIs was lower than for CGIs overlapping H3K36me3 and for normally methylated CGIs. This suggests that aberrantly methylated CGIs are not main targets of DNMT3B2.

Exogenous DNMT3B2 expression was restored in DKO cells and caused greater than 20% methylation gain at ~9% of CGIs. High DNMT3B-mediated methylation was detected at CGIs overlapping H3K36me3 (12.55% median methylation gain) compared to CGIs non-overlapping H3K36me3 (0.6% median methylation gain). This is in line with a previous report, showing H3K36me3 enhances DNMT3B localization to actively

A

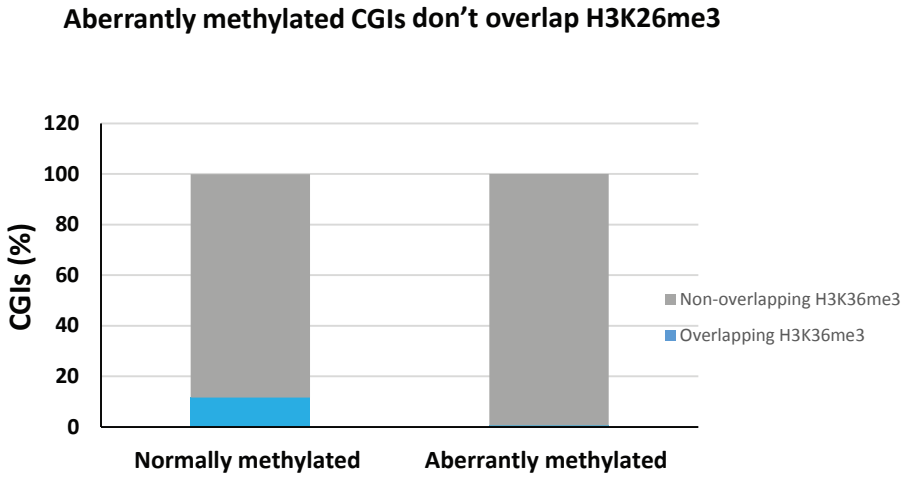


Figure 5.13:
Higher number of normally methylated CGIs overlap H3K36me3 peaks in HCT116 compared to aberrantly methylated CGIs. Bar graph showing the fraction of normally methylated and aberrantly methylated (CGIs overlapping H3K36me3 (cyan). Fraction of CGIs non-overlapping H3K36me3 are coloured grey.

transcribed gene bodies that are marked with H3K36me3 *in vivo* (Baubec *et al.*, 2015).

Most importantly, aberrantly methylated CGIs showed only 3.4% median methylation gain. Aberrantly methylated CGIs gained less methylation than normally methylated CGIs and CGIs overlapping H3K36me3 in another independent biological replicate with higher levels of DNMT3B expression (experiment performed by Cristina Rubio, and RRBS data analyzed by Duncan Sproul).

Given that DKO cells expresses a hypomorph of DNMT1, it is possible that preferential targeting of DNMT1 hypomorph also contributes to the observed gain of methylation at CGIs overlapping H3K36me3. One experiment to test that would be to perform ChIP-seq for DNMT3B2 in DKO cells, to assess the binding sites of the protein and how much they overlap our identified primary targets of DNMT3B2. Another way to test that, would be to restore DNMT1 expression following DNMT3B restoration, and assess how this will affect methylation gain at CGIs. Additionally, it would be interesting to restore DNMT1 alone. This experiment has been performed in the past, however methylation increase was assessed by measuring total methylation levels and bsPCR at only 3 CGIs (Jair et al., 2006). In this experiment an increase in total methylation levels and hypermethylation of one of the CGIs was observed.

In a previous study, Duymich *et al* have reintroduced different active and inactive DNMT3B isoforms into DKO cells, and showed that DNMT3B isoforms lacking catalytic activity stimulated DNA methylation at gene bodies (Duymich *et al.*, 2016). This was suggested to happen through recruitment of DNMT3A by inactive DNMT3B isoforms. Our results confirm this finding, as DNMT3B2-cd resulted in 17% median methylation gain at DNMT3B2 targets. This suggests

that aberrantly methylated CGIs are also not primary targets of DNMT3A. It is worth mentioning that Duymich *et al* study is different from mine as they have restored the embryonic stem cell form (DNMT3B1) whereas I restored the main active *de novo* DNMT in somatic cells and in HCT116 cell line (DNMT3B2) and its catalytic dead version.

Chapter 6: Low methylation levels are targeted to aberrantly methylated CGIs

In the previous chapter, levels of DNMT3B-mediated methylation was assessed at aberrantly methylated CGIs genome-wide in their native loci in a knockout cell system. Cooperativity between DNMTs in maintaining methylation at certain genomic contexts has been reported (Kim *et al.*, 2002; Liang *et al.*, 2002; Rhee *et al.*, 2002). Therefore, in this chapter I aimed to assess the endogenous levels of methylation targeted to aberrantly methylated CGIs in the presence of all DNMTs that are normally expressed in HCT116 system as well as in the absence of each DNMT enzyme individually.

6.1 Experimental design

In order to assess methylation levels at different CGIs, I de-methylated WT HCT116 cells and DNMT knockout derivatives by treatment with 5-Aza-2′-deoxycytidine. Subsequently, the cells were allowed to recover methylation, and I assessed the rate of methylation recovery of CGIs (Figure 6.1). Hypomethylating the cells and allowing them to recover methylation enables us to identify which CGIs show a higher remethylation rate compared to all CGIs. This reflects which CGIs are the main targets of DNMT enzymes in HCT116 cells. The knockout cell lines help to dissect the effect of each of the DNMT enzymes on methylation recovery and complements my other results.

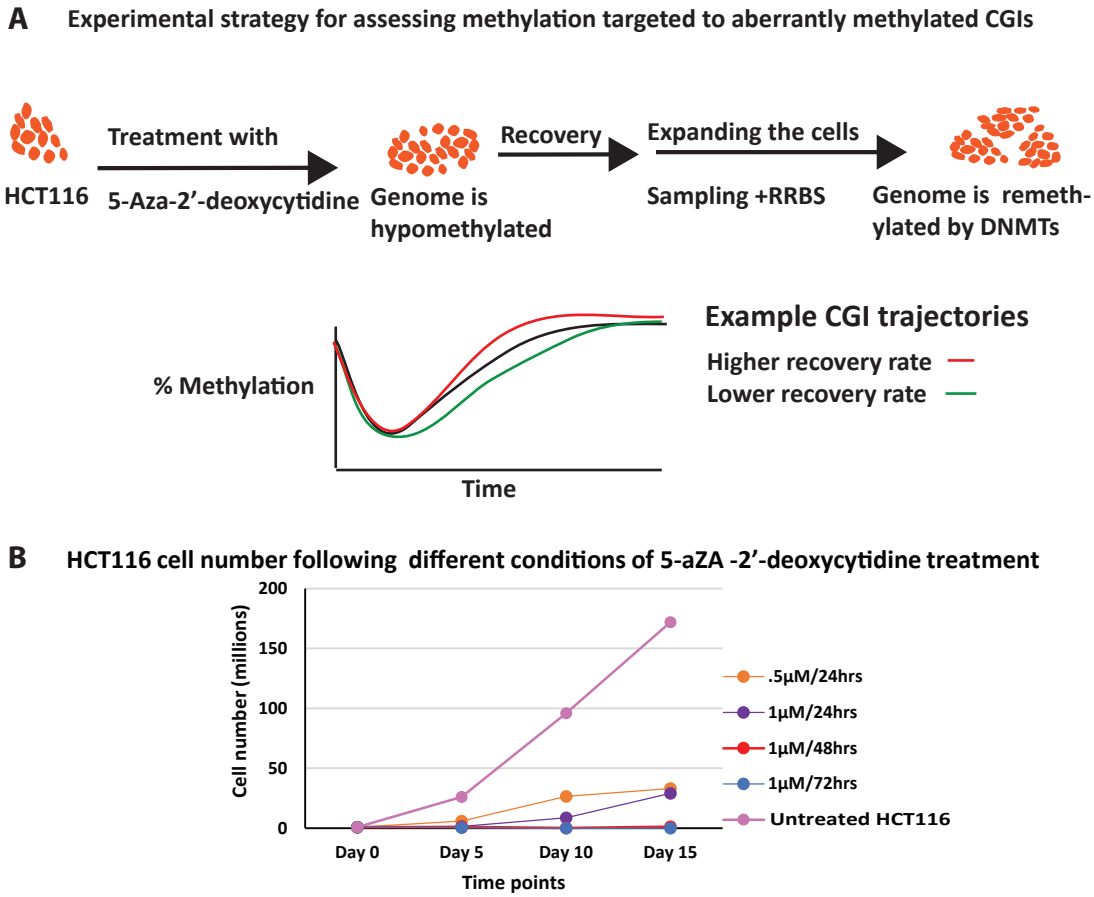


Figure 6.1: Establishing HCT116 re-methylation model utilizing 5-Aza-2'-deoxycytidine.

A. Experimental strategy to investigate methylation targeted to aberrantly methylated CGIs. The genomes of HCT116, 1KO, and 3BKO were hypomethylated by treatment with 5-Aza-2'-deoxycytidine. The cells were subsequently expanded to allow remethylation. Samples were collected at different time points to assess the remethylation rate of CGIs.

B. 5-Aza-2'-deoxycytidine treatment for one day at either .5µM or 1µM leaves live cells. Line graph showing cell number X10^6 (y-axis) at different time points (x-axis) for four 5-Aza-2'-deoxycytidine treatment conditions: 0.5µM for 24hrs (orange), 1µM for 24hrs (purple), 1µM for 48hrs (red), 1µM for 72hrs (blue), and untreated control (pink).

C. HCT116 cells recover their ability to proliferate by day 10 following 5-Aza-2'-deoxycytidine treatment for 1µM for 24hrs. Colony formation assay (top panel) and phase contrast pictures (lower panel) of HCT116 cells at different time points following 5-Aza-2'-deoxycytidine treatment at 1µM for 24hrs. 1000 cell/well of HCT116 were plated for colony formation assay. Colonies were stained 10 days later.

6.2 Establishing HCT116 remethylation model utilizing 5-Aza-2′-deoxycytidine

5-Aza-2′-deoxycytidine is a cytidine analog that incorporates into newly replicated DNA. Upon binding of 5-Aza-2′-deoxycytidine to DNMT1, stable adducts are formed (Patel *et al.*, 2010). These adducts induce DNMT1 degradation leading to passive demethylation of the genome as cells divide. Additionally, 5-Aza-2′-deoxycytidine triggers the P53 DNA damage pathway causing cell toxicity (Karpf *et al.*, 2001). In order to obtain treatment conditions that will produce efficient hypomethylation with minimum Aza-induced cell death, I tested 4 conditions some of which were reported in the literature (Sproul *et al.*, 2011; De Carvalho *et al.*, 2012).

HCT116 were treated with 0.5μM for 24hrs, 1μM for 24hrs, 1μM for 48hrs and 1μM for 72hrs. For the two latter conditions, the drug was refreshed every 24hrs. Subsequently, cell numbers were assessed before adding the drug (day 0) and at day 5, 10, and 15 after adding the drug. After 5-Aza-2′-deoxycytidine treatment, HCT116 divided and increased their numbers when treated with 0.5μM or 1μM for 24hrs conditions (figure 6.1B). On the other hand, 5-Aza-2′-deoxycytidine treatment at 1μM for either 48hrs or 72hrs induced high cell toxicity as evident by no significant increase in HCT116 cell number until day 15. A very small number of cells survived the latter two conditions which would prevent continuing the experiment.

To assess the extent of hypomethylation in HCT116 under the studied treatment conditions, methyl sensitive southern blotting was performed. Genomic DNA from HCT116 was digested with the methylation sensitive enzyme BstBI and probed with a labelled DNA fragment of human satellite II. The extent of hypomethylation can be visualized by the decrease in the intensity of large-size DNA fragments (figure 6.2A). Aza-2′-deoxycytidine treatment at 1μM for either

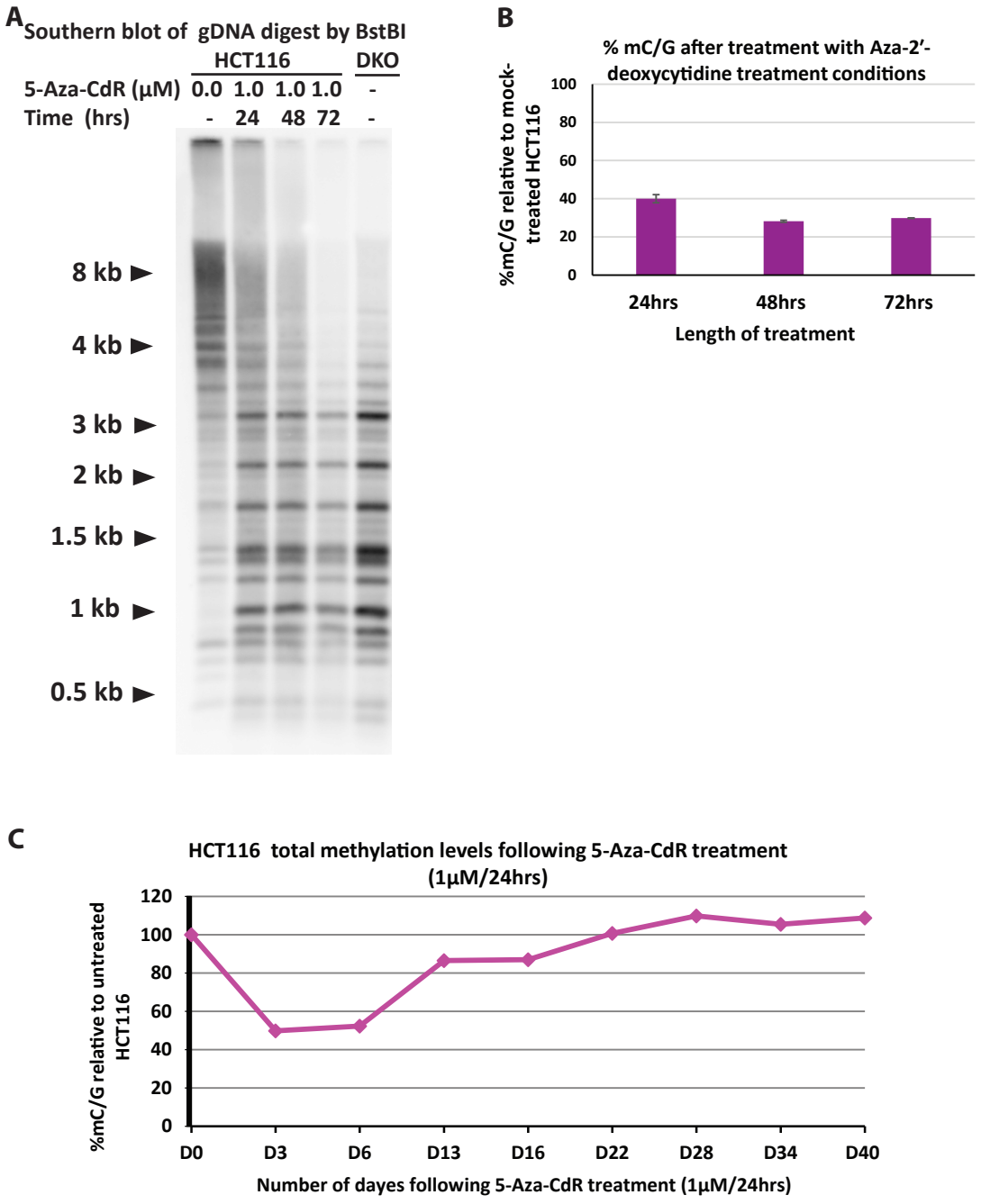


Figure 6.2: Optimizing Aza-2'-deoxycytidine treatment conditions.

A. Aza-2'-deoxycytidine induced hypomethylation under the three tested treatment conditions. Southern blot analysis of HCT116 cells to detect methylation status using the methylation sensitive restriction enzyme BstBI. Genomic DNA (1µg) from HCT116 cells treated with 1µM 5-Aza-2'-deoxycytidine for different lengths of time and untreated DKO cells was digested with BstBI and probed with satellite II labeled DNA fragment.

B. Bar graph showing % total methylation levels (y-axis) at different Aza-2'-deoxycytidine treatment conditions (x-axis). Aza-2'-deoxycytidine treatment at 1µM for 24 hrs induced ~60% reduction in global methylation levels, whereas treatment for either 48hrs or 72hrs induced ~ 70% reduction. The measurements were taken from 3 technical replicates. Error bars represents standard deviation.

C. Substantial changes in HCT116 % methylation levels take place between day 0 and day 22 following Aza-2'-deoxycytidine treatment. Line graph showing % methylation levels of HCT116 cells (y-axis) at different time points following administration of Aza-2'-deoxycytidine at 1µM for 24 hrs (x-axis).

24 hrs, 48hrs or 72hrs results in hypomethylation compared to untreated control. Next, to assess the extent of hypomethylation more quantitatively and globally, methylation levels were measured using mass spectrometry (Figure 6.2B). Aza-2′-deoxycytidine treatment at 1μM for 24 hrs induced ~60% reduction in global methylation levels, and treatment for either 48hrs or 72hrs induced ~ 70% reduction. Overall, Aza-2′-deoxycytidine treatment at 1μM for 24 hrs induces substantial hypomethylation and induce less cell toxicity compared to longer treatments.

I treated HCT116 cells with Aza-2′-deoxycytidine at 1μM for 24 hrs. Subsequently, medium was refreshed and cells were expanded for 40 days. To determine which time points to consider for RRBS, global methylation levels were measured by mass spectrometry at various time points up to 40 days. Figure 6.2C shows that the substantial changes in methylation take place between day 0 and day 22, therefore all samples collected within this interval were subjected to RRBS. To ensure our RRBS data covers the whole remethylating process, samples collected at day 28 and day 40 were also included in RRBS analysis.

6.3 Recovery after 5-Aza-2′-deoxycytidine treatment reflects remethylation dynamics

For 5-Aza-2′-deoxycytidine to hypomethylate the cells, it first needs to be incorporated into the DNA. Next, cell divisions are required so that the genome gets passively hypomethylated. In order to reduce number of cells escaping the treatment, I plated the cells at low density (~20% cell confluency) to give cells space to divide and get hypomethylated.

Subsequently, I tested if the methylation recovery seen by mass spectrometry is indeed due to remethylation of hypomethylated cells rather than outgrowth of

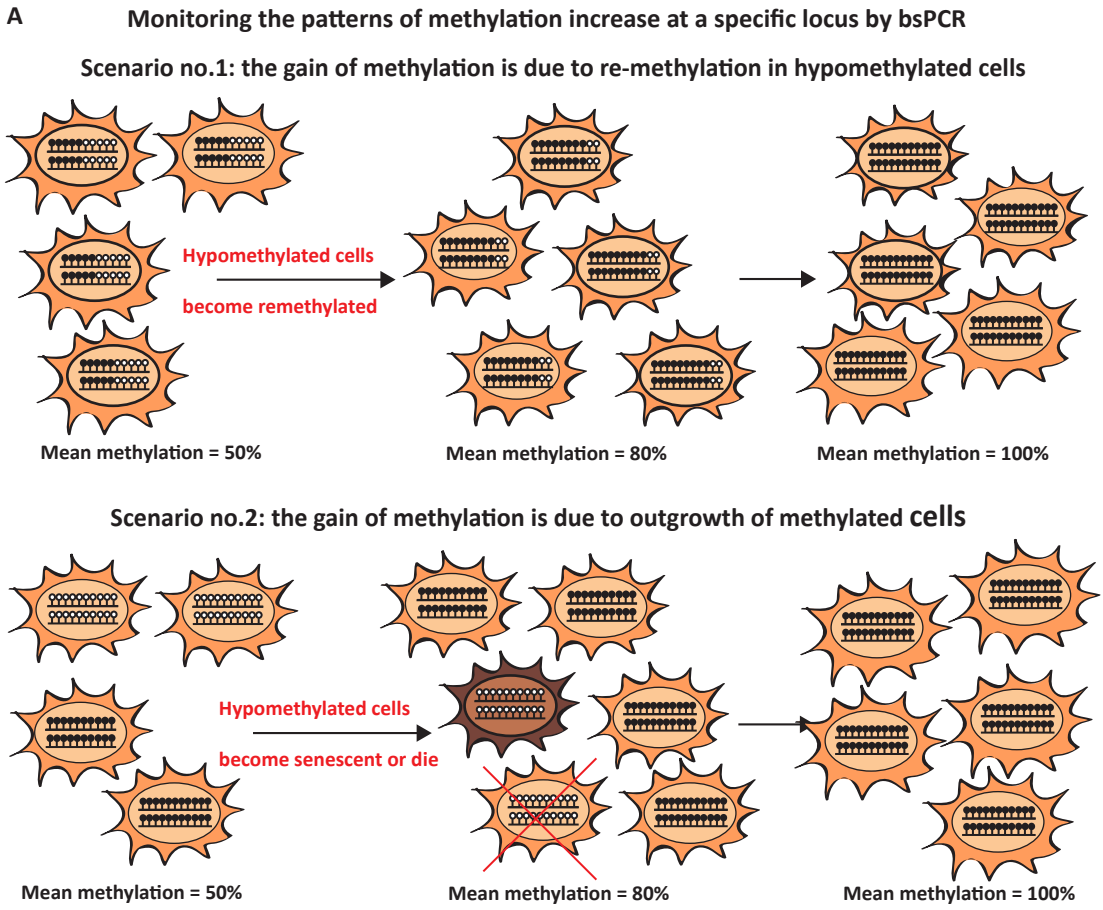


Figure 6.3: Ways of validating re-methylations of cells following 5-Aza-2'-deoxycytidine treatment.
A. Schematic for monitoring methylation at a certain CGI by bsPCR. If the increase in methylation is due to re-methylation, an increase in the methylation of the demethylated alleles will be observed. If the increase in methylation is due to outgrowth, only an increase in the number of fully methylated alleles will be observed. This is might happen if hypomethylated cells die or become senescent. Dividing cells are coloured orange, senescent cells are couloured brown, and dying cells are crossed in red.
B. Schematic illustration of modeling the expected gain of methylation in re-methylation vs. outgrowth scenarios. If the gain of methylation is due to remethylation, CGIs will gain methylation at different rates. If the gain in methylation is due to outgrowth of methylated cells, then CGIs will gain methylation at an identical rate.

cells that escaped 5-Aza-2′-deoxycytidine treatment. This was done by monitoring the patterns of increase in methylation at CDH7 locus by bsPCR (Figure 6.3). CDH7 CGI is one of the representative CGI I studied in chapter 3. At day 0, as expected, the locus was hypermethylated (97% mean methylation). At day 3, the locus showed demethylation, and the maximum demethylation was reached at day 6 (59% mean methylation). The unmethylated CpGs were dispersed (or scattered) within the demethylated CDH7 alleles. Notably, some alleles stayed fully methylated suggesting some cells were not demethylated. The most drastic increase in mean methylation at CDH7 locus took place between day 6 (59% mean methylation) and day 13 (90% mean methylation).

If the gain of methylation is due to re-methylation, then an increase in methylation would be observed within the demethylated alleles (figure 6.3A). On the other hand, if the increase in methylation observed is due to outgrowth of cells that escaped 5-Aza-2′-deoxycytidine treatment, then only an increase in the number of fully methylated alleles would be observed (figure 6.3A). To test this, I quantified % mean methylation of de-methylated alleles (i.e. alleles with at least 3 unmethylated CpGs) (Figure 6.4A). Corresponding to the 31% increase in total mean methylation observed between day 6 and 13, the mean methylation of de-methylated alleles also increased 30%. This suggests we see remethylation of HCT116 cells after Aza-2′-deoxycytidine induced demethylation. Additionally, the increase in methylation within demethylated alleles between day 6 and 13 can be visualized in Figure 6.4B.

Having established that a substantial part of methylation recovery after 5-Aza-2′-deoxycytidine treatment reflects remethylation dynamics, RRBS data for HCT116 samples collected at several time points before and after 5-Aza-2′-deoxycytidine treatment were subsequently generated with at least 20x coverage for a mean of 9,143,298 CpGs in each sample. The subset of CpGs

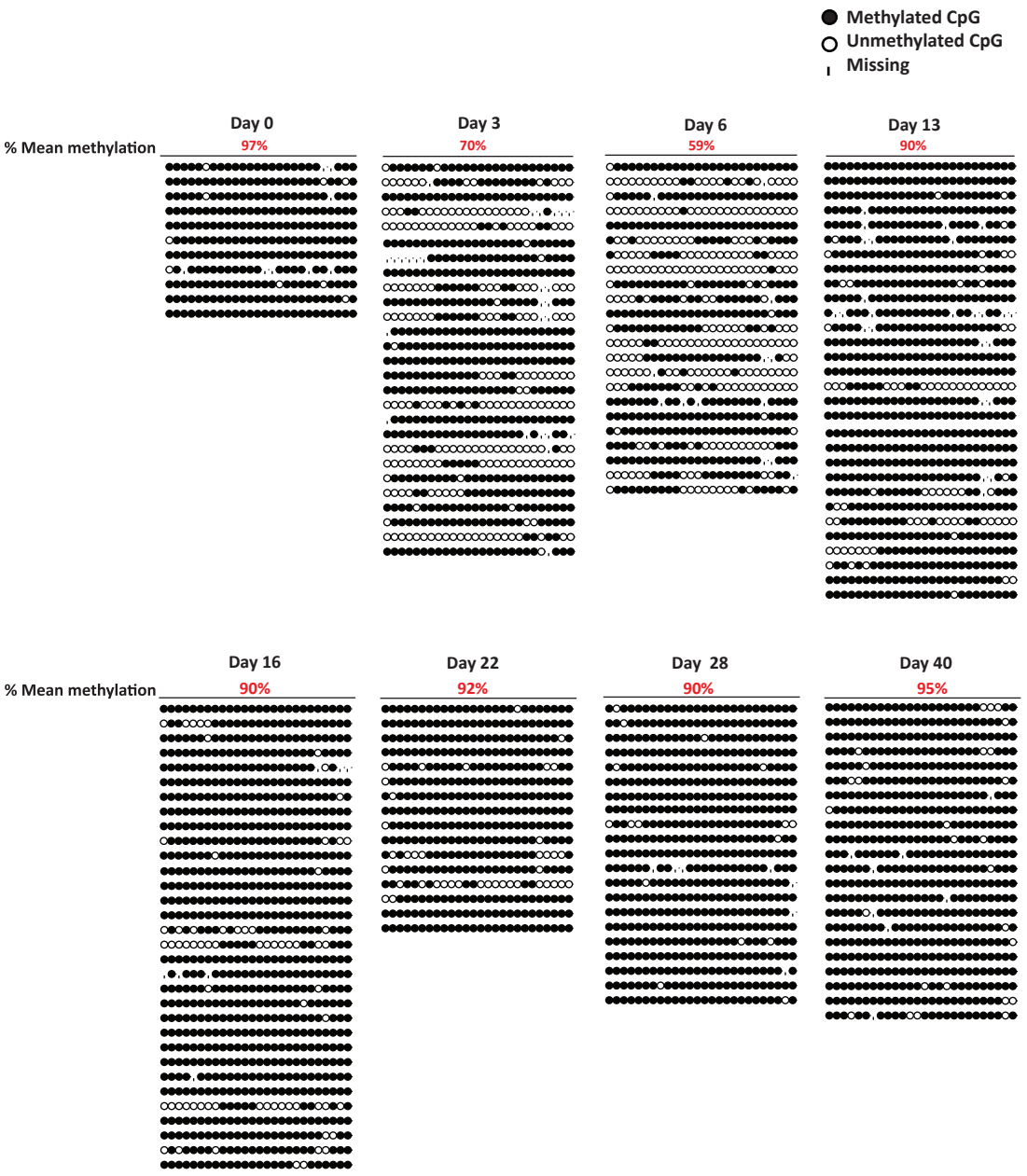


Figure 6.4: CDH7 CGIs recovers most of its methylation between day 6 and day 13. Bisulfite sequencing analysis of CDH7 CGI at different time points after Aza-2'-deoxycytidine treatment quantified as mean methylation at each time points (represented in red). Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. Empty spaces reflect ambiguous base calls in Sanger sequencing reads.

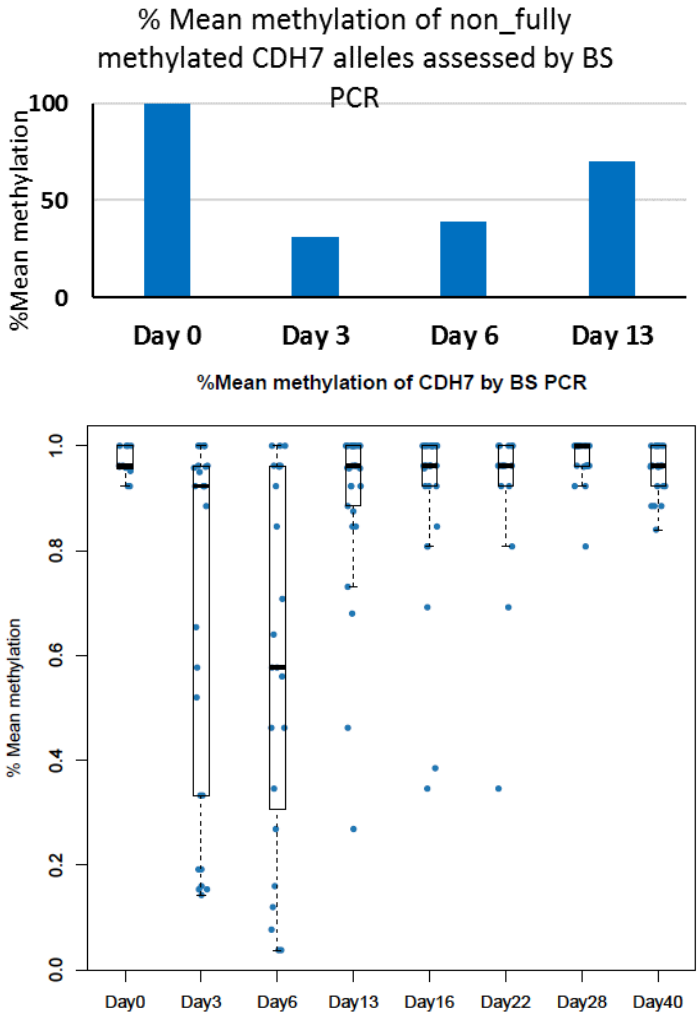


Figure 6.5: Remethylation is observed within de-methylated CDH7 alleles. A. Bar graph showing % mean methylation of demethylated CDH7 alleles (y-axis) at different time points (x-axis) following Aza-2′-deoxycytidine treatment as measured by BS PCR. Mean methylation of de-methylated alleles increased by 30% between day 6 and 13.

B. Box plots showing % mean methylation (y-axis) for each CDH7 allele (blue dots) at different time points (x-axis) following Aza-2′-deoxycytidine treatment as measured by BS PCR. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Horizontal line within the box represents median.

described in chapter 5 producing 17,947 CGIs with ≥ 20 coverage. For looking at remethylation rates, only CGIs highly methylated in HCT116 (>= 70% mean methylation) were considered in the analysis. This included 6772 CGIs. The alignment statistics of RRBS data for these samples is shown in the supplementary data (Table 6.1).

Next, I used RRBS data to perform a second test to validate that the increase in methylation observed substantially reflects remethylation rather than outgrowth of methylated cells. Assuming two possible cases, first, if the increase in methylation is due to outgrowth of methylated cells, then all CGIs would have the same rate of gain provided that methylation of CGIs at every time point is normalized by dividing over initial methylation levels. Conversely, if the gain of methylation is due to re-methylation, then differences in the rate of methylation gain for different CGIs are expected.

In order to address whether different CGIs have different recovery rate, I modeled the two possible scenarios. In the first scenario, methylation is a function of time and CGIs have different recovery rate. In the second scenario, methylation is a function of time and CGIs have the same recovery rate. These cases were formulated using a mixed linear model (MLM) which can distinguish real changes from changes caused by noise in the assay (figure 6.2B).

Next, the probability that each CGI is gaining methylation with identical rate was calculated by performing the likelihood ratio of the two models using Anova. The probability that methylation of each CGI is increasing with identical rate (i.e. outgrowth) is $6.71X10^{-20}$. This indicates that a significant portion of the gain in methylation is due to re-methylation. However, this might be partially masked by some non-hypomethylated cells outgrowing.

6.4 CGIs targeted by DNMT3B2 show higher remethylation rate in HCT116 after treatment with Aza-2′-deoxycytidine

In order to test whether I can use remethylation rate after treatment with Aza to confirm the findings from chapter 5, I tested whether DNMT3B targets, identified in the previous chapter (see, *section 5.4*), show higher remethylation rate in this experiment. Mean methylation of DNMT3B targets against time is illustrated in figure 6.6A. The median methylation for all DNMT3B targets was calculated at each time point and represented by the yellow line and for non-DNMT3B targets by the pink line. DNMT3B-targeted CGIs showed slightly greater remethylation rate as shown from the yellow line compared to the pink line (figure 6.6A). To compare the two groups of CGIs in a more quantitative way, the rate of recovery of each CGI was calculated from the slope of a linear model fitted to the data between day 6 and day 22. DNMT3B-targeted CGIs showed 9.48% higher median methylation recovery rate than the CGIs that didn't get methylated after expressing DNMT3B2 in DKOs (Student's t Test, p-value < $2.2X10^{-16}$) (figure 6.6B).

Next, I examined the remethylation rate at CGIs overlapping H3K36me3 peaks. CGIs overlapping H3K36me3 peaks showed 7.79% higher median recovery rate than CGIs non-overlapping H3K36me3 (Student's t-Test, p-value = $8.0X10^{-16}$) (figure 6.6C and 6.6D). These result indicate that, in HCT116, differences in remethylation rate after treatment with Aza between different CGI groups are subtle. However, these subtle differences are significant and agree with the literature, suggesting that remethylation rate can be used to confirm the findings from chapter 5.

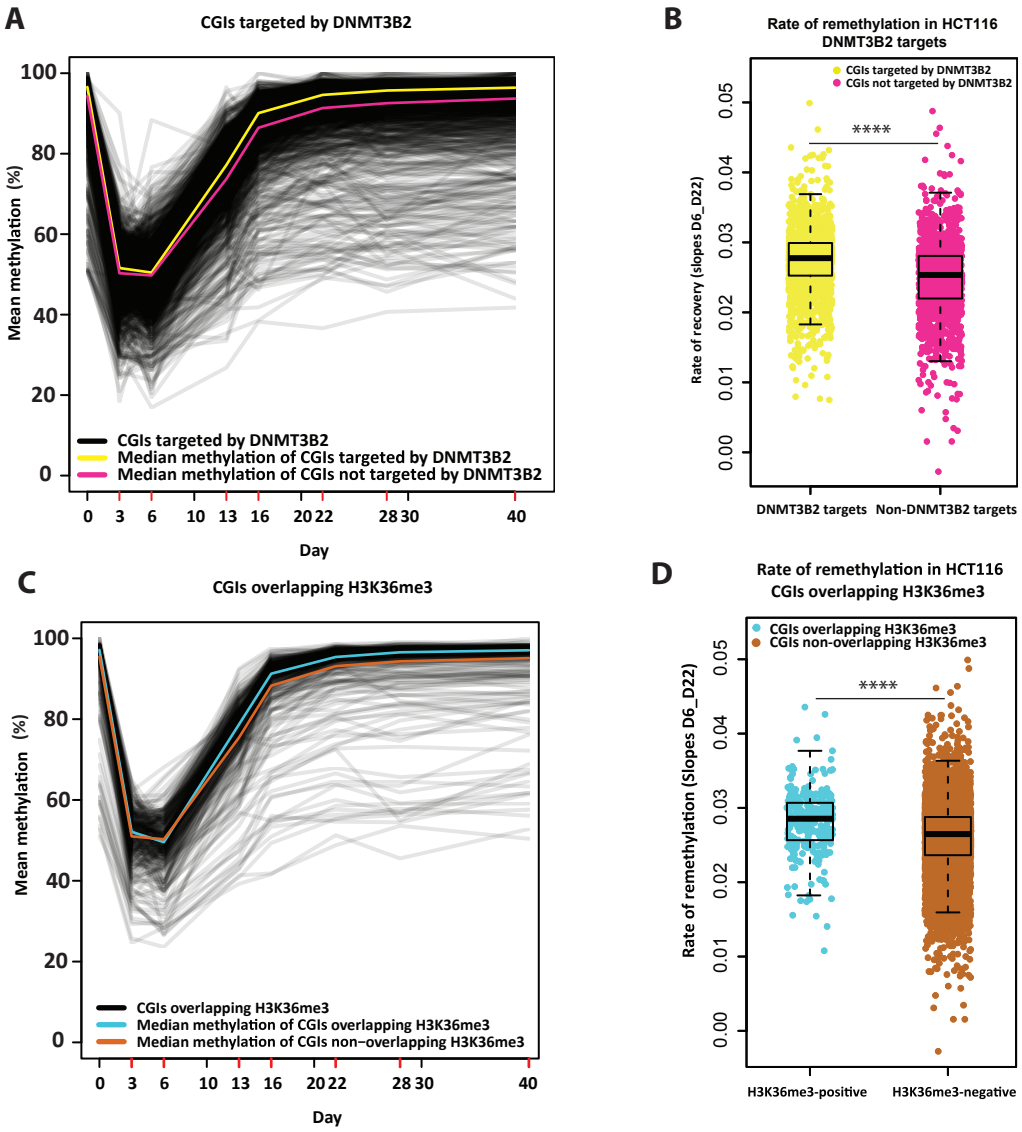


Figure 6.6: DNMT3B2 targets show greater remethylation rate following Aza-2′-deoxycytidine treatment than CGIs which are not targeted by DNMT3B2.

A. Line graph showing mean methylation in HCT116 of DNMT3B2 targeted CGIs (y-axis) at different time points following Aza-2′-deoxycytidine treatment (x-axis) as measured by RRBS. Mean methylation of each CGI is represented by a black line. Median methylation of CGIs methylated by DNMT3B2 in DKO cells (yellow) and for CGIs not gaining methylation by DNMT3B2 (pink). Red ticks on the x-axis represent time points at which methylation was measured.

B. Box plots showing rate of remethylation following Aza-2′-deoxycytidine treatment (y-axis) for CGIs targeted by DNMT3B-2(colored yellow) compared to CGIs that didn't gain methylation by DNMT3B2 (colored pink). Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the rate of remethylation for a single CGI. Horizontal line within the box represents median. **** p<0.0001, Student's t test.

C. Line graph showing mean methylation of CGIs marked by H3K36me3 in HCT116 (y-axis) at different time points following Aza-2′-deoxycytidine treatment (x-axis) as measured by RRBS. Mean methylation of each CGI is represented by a black line. Median methylation of CGIs overlapping H3K36me3 (cyan) and for CGIs non-overlapping H3K36me3 (brown). Red ticks on the x-axis represent time points at which methylation was measured.

B. Box plots showing rate of remethylation following Aza-2′-deoxycytidine treatment for CGIs overlapping H3K36me3 (cyan) compared to CGIs non-overlapping H3K36me3 (brown).

**** p<0.0001, Student's t test.**

6.5 Aberrantly methylated CGIs show lower recovery rate than CGIs overlapping H3K36me3 and normally methylated CGIs

Our main result from the previous chapter suggests that aberrantly methylated CGIs are not main targets of DNMT3B. I next aimed to confirm this result and extend it to a system where all HCT116 DNMTs are present. Compared to CGIs overlapping H3K36me3, aberrantly methylated CGIs had 6.09% lower median remethylation rate (p-value = 2.02 $X10^{-7}$) (Figure 6.7). The recovery rate of aberrantly methylated CGIs is also 2.58% lower than normally methylated CGIs (p-value = 0.00271). This implies that lower methylation levels are targeted to aberrantly methylated CGIs compared to CGIs overlapping H3K36me3 in WT HCT116.

6.6 No significant difference between remethylation rates of aberrantly methylated CGIs and CGIs overlapping H3K36me3 in the absence of DNMT3B

In this section, I wanted to dissect how each DNMT enzyme present in HCT116 is distributing its methylation across CGIs, Therefore, I compared the remethylation rate of the three groups of CGIs in 1KO cells and 3BKO. Figure 6.9A shows that in 1KO cells, at day 40, the aberrantly methylated CGIs don't fully recover their initial methylation, whereas CGIs overlapping H3K36me3 and normally methylated CGIs recover more of their initial methylation presumably because they are DNMT3B targets. Moreover, the remethylation rate of aberrantly methylated CGIs is significantly lower than normally methylated CGIs and CGIs overlapping H3K36me3 (Figure 6.8B).

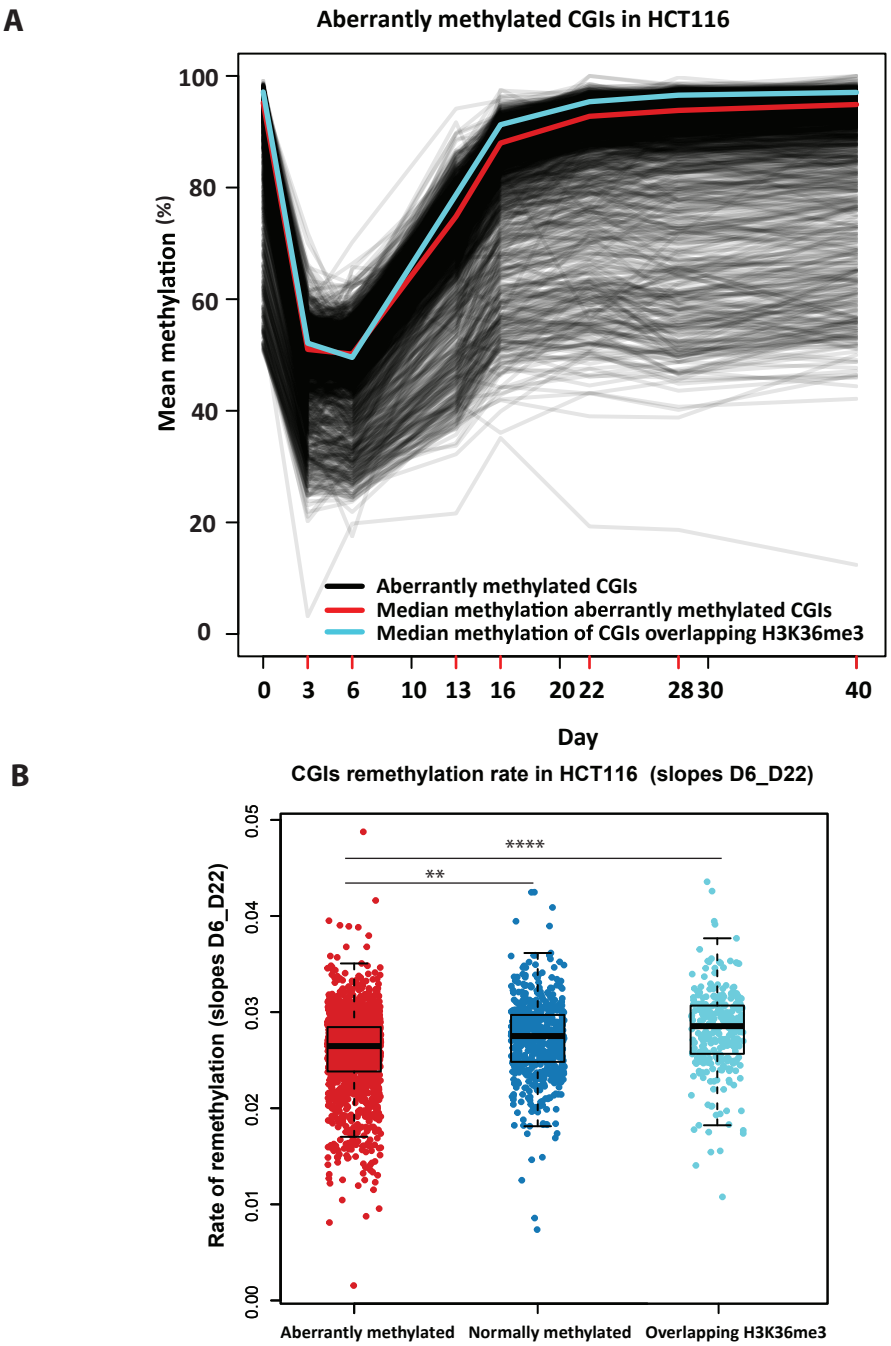


Figure 6.7: Aberrantly methylated CGIs show lower recovery rate than CGIs overlapping H3K36me3 and normally methylated CGIs.

A. Line graph showing mean methylation of aberrantly methylated CGIs in HCT116 (y-axis) at different time points following Aza-2′-deoxycytidine treatment (x-axis) as measured by RRBS. Mean methylation of each CGI is represented by a black line. Median methylation of aberrantly methylated CGIs (red) and for CGIs overlapping H3K36me3 (cyan). Red ticks on the x-axis represent time points at which methylation was measured.

B. Box plots showing remethylation rate following Aza-2′-deoxycytidine treatment (y-axis) for aberrantly methylated CGIs (red) compared to normally methylated CGIs (colored blue) and CGIs overlapping H3K36me3 (cyan). Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the rate of remethylation for a single CGI. Horizontal line within the box represents median. **** p<0.0001, Student's t test.

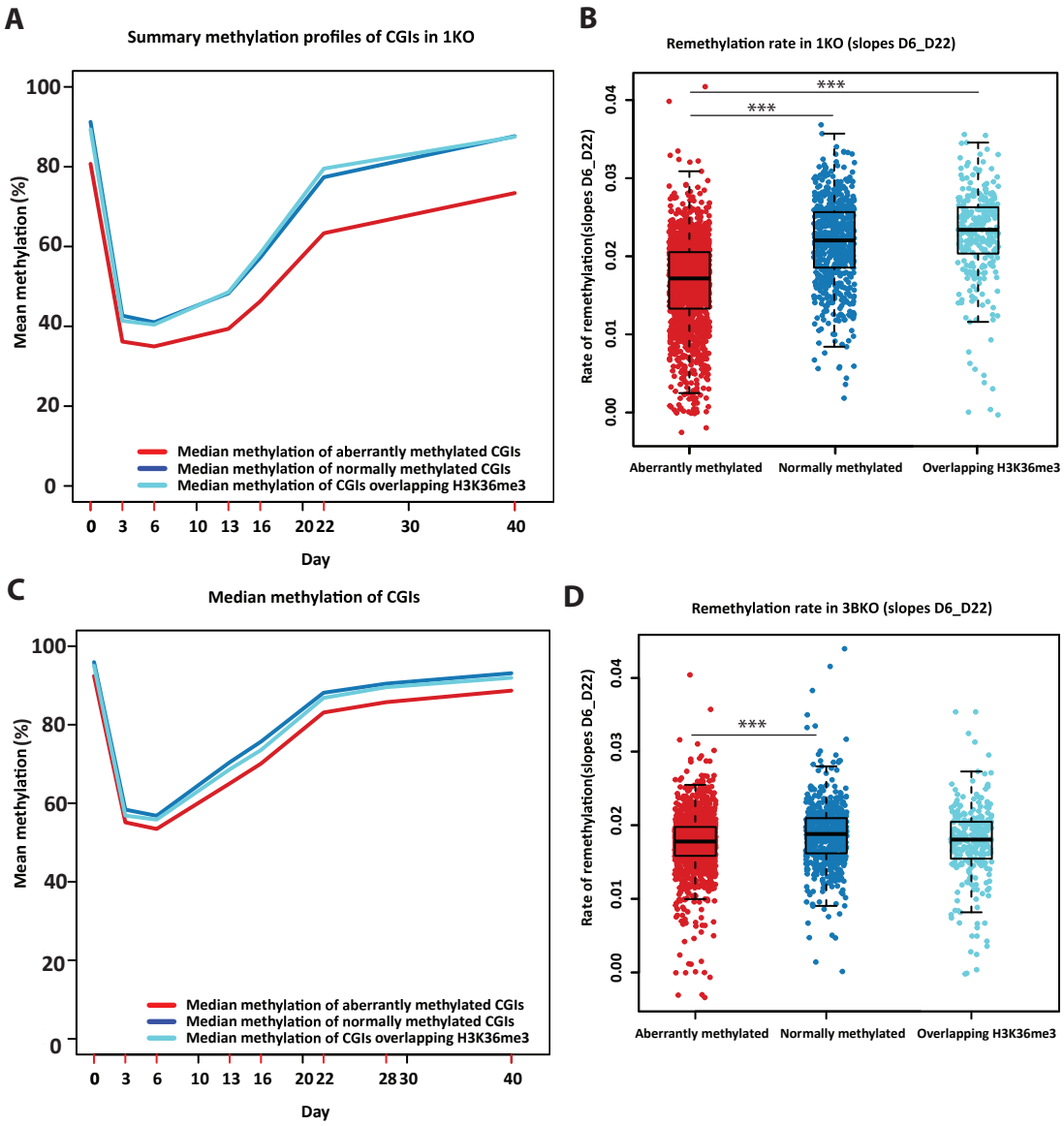


Figure 6.8: No significant difference in CGI remethylation rates in the absence of DNMT3B.

Line graph showing median mean methylation of CGIs (y-axis) In 1KO (A) and 3BKO (C) at different time points following Aza-2′-deoxycytidine treatment (x-axis) as measured by RRBS. Median methylation of aberrantly methylated CGIs (red) and CGIs overlapping H3K36me3 (cyan). Red ticks on the x-axis represent time points at which methylation was measured.

Box plots showing remethylation rate following Aza-2′-deoxycytidine treatment (y-axis) for aberrantly methylated CGIs (red) compared to normally methylated CGIs (blue) and CGIs overlapping H3K36me3 (cyan) in 1KO (B) and 3BKO (D). Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the rate of remethylation for a single CGI. Horizontal line within the box represents median. *** p<0.001, Student's t test.

Conversely, in 3BKO cells, there is no significant difference between the remethylation rate of aberrantly methylated CGIs, normally methylated CGIs and CGIs overlapping H3K36me3 (Figure 6.8C and D). This suggests that DNMT1 mediated methylation is equally distributed over the three groups of CGIs. This goes in line with the main role of DNMT1 in maintaining methylation.

6.7 Methylation seeding can't explain the differences in remethylation rate of CGIs.

In chapter 5, I have shown that the methylation levels of aberrantly methylated CGIs and normally methylated CGIs might enhance the methylation gain by DNMT3B. Therefore, I asked here whether CGIs methylation levels affect the remethylation rate following 5-Aza-2′-deoxycytidine treatment. Since CGIs start getting remethylated after day 6, I stratified the CGIs into ten groups depending on their methylation levels on day 6 and plotted their remethylation rate. Figure 6.9A and B shows that in 1KO and 3BKO cells the remethylation rate is not proportionally correlated with the CGIs mean methylation on day 6.

In HCT116, the remethylation rate of aberrantly methylated CGIs is proportionally correlated with CGIs mean methylation levels (in the methylation range 30- 60%) and becomes inversely correlated with CGIs methylation levels (in the methylation range 60-100%). However, this trend is not consistent between aberrantly methylated, normally methylated and H3K36me3 positive CGIs. Notably, in HCT116, aberrantly methylated CGIs still have lower remethylation rate compared to normally methylated CGIs and CGIs overlapping H3K36me3 in all CGIs groups. Overall, this data suggests that there isn't a consistent effect of the CGIs methylation levels on their remethylation rate and that "methylation seeding" can't explain the differences in remethylation rate of CGIs.

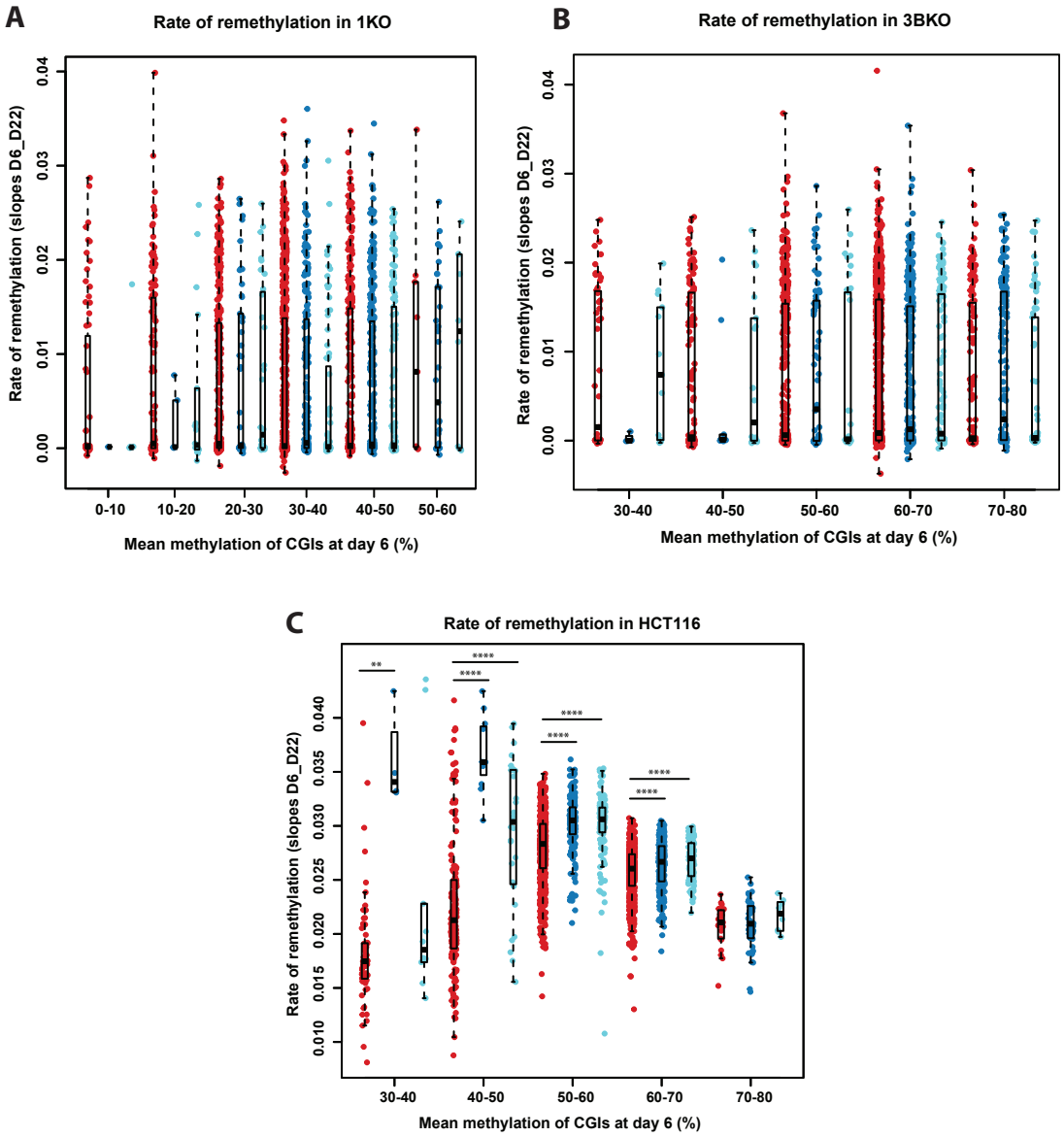


Figure 6.9: CGIs methylation levels has no consistent effect on the remethylation rate.
Box plots showing remethylation rate following Aza-2'-deoxycytidine treatment (y-axis) for aberrantly methylated CGIs (red) compared to normally methylated CGIs (blue) and CGIs overlapping H3K36me3 (cyan) in 1KO (A), 3BKO (B), and HCT116 (C). The CGIs were stratified into groups based on their methylation levels on day 6. Only groups containing sufficient number of CGIs in them (for performing a statistical test) are displayed. Boxes denote the inter-quartile range (IQR) and whiskers 1.5 × IQR. Each dot represents the rate of remethylation for a single CGI. Horizontal line within the box represents median. Error bars represents standard deviation. **** P<0.0001, ** P<0.01, Mann-Whitney-Wilcoxon test.

6.8 Discussion

In this chapter, I addressed how much methylation is targeted to aberrantly methylated CGIs in WT HCT116, 1KO, and 3BKO. To do this, HCT116 cells and its DNMT knock out derivatives were treated with Aza-2′-deoxycytidine and subsequently allowed to recover methylation. The rate of remethylation was assessed at CGIs across the genome. DNMT3B2 targets, identified in the previous chapter, had higher recovery rate than non-DNMT3B2 targets. Additionally, CGIs overlapping H3K36me3 recovered at higher rate than CGIs non-overlapping H3K36me3. This later result, is consistent with previous study investigating the effect of gene body demethylation by Aza-2′-deoxycytidine on their expression (Yang et al., 2014). Together, these results suggested we can use remethylation rate after Aza-2′-deoxycytidine treatment to corroborate results from my experiment in DKO cells (chapter 5).

Aberrantly methylated CGIs recover slower than normally methylated CGIs and CGIs overlapping H3K36me3 in WT HCT116 cells. Since WT HCT116 cells mainly express DNMT1 and DNMT3B, this suggests that these two DNMTs or at least one of them is targeting CGIs overlapping H3K36me3 more than normally methylated CGIs and targeting both of these groups more than aberrantly methylated CGIs. This result holds true in 1KO cells, with more obvious differences in the remethylation rate between the three groups compared to WT HCT116. 1KO cells express only DNMT3B, DNMT3A, and hypomorphic DNMT1. One possible explanation for the greater differences in remethylation rates is that DNMT1 might be preferentially maintaining methylation at aberrantly methylated CGIs. Another possible explanation might be that hypomorphic DNMT1 is targeting aberrantly methylated CGIs less than CGIs overlapping H3K36me3 and normally methylated CGIs. To address which one of these possible cases is taking place I can re-express DNMT1 alone and in

combination with DNMT3B2 in DKO cells as suggested in the discussion of the previous chapter.

In 3BKO, the differences in remethylation rate between the three groups of CGIs are abolished, suggesting that in the absence of DNMT3B, the methylation is equally distributed across the three CGIs groups. These results reinforce that aberrantly methylated CGIs are not main targets of DNMT3B. Moreover, DNMT3B is responsible for the higher re-methylation rates of H3K36me3 and normally methylated CGIs compared to aberrantly methylated CGIs. The 3BKO cells express DNMT1 and DNMT3A, this suggests that in the absence of DNMT3B, DNMT3A doesn't have a preferential targeting of any of the three studied CGI groups.

Since the 5-Aza-2-deoxycytidine treatment induced only partial hypomethylation, the remaining methylated CpGs may act as a substrate for re-methylation by DNMT3B. Therefore, I tested if the CGIs methylation levels have an effect on the remethylation rate in HCT116 and 1KO and found no consistent effect. However, this doesn't exclude the possibility that the methylated CpGs act as substrate for remethylation without affecting the dynamics of that process (i.e., the rate of remethylation). Therefore, it would be interesting to further test the methylation seeding model by ectopically integrating CGIs with sparse methylated CpGs within them and see if this triggers their hypermethylation in HCT116 cells.

In summary, results from this chapter suggests that aberrantly methylated CGIs are not primary targets of *de novo* DNMT in HCT116 cells, in agreement with the results from the previous chapters. Future studies to confirm this finding *in vivo* are now needed.

Chapter 7: Conclusions and Future Perspectives

In this thesis, I have used different strategies to ask if the aberrant methylation of CGIs is programmed by their sequence and if the aberrantly methylated CGIs are direct targets of DNMTs. To address the first question I ectopically integrated CGIs into the genome and assessed their methylation state. I found that integrated aberrantly methylated CGIs didn't gain aberrant methylation at most ectopic locations in HCT116 (chapter3, figure 3.8). This finding is surprising and suggests that, unlike normal methylation in stem cells, aberrant methylation in cancer cells is not programmed by CGIs' sequence (Lienert *et al.*, 2011). Additionally, it suggests that aberrantly methylated CGIs are not main targets of DNMTs. However, one caveat of this experiment is that I am assessing methylation at the integrated CGIs out of their native context. Moreover, it is possible that the integrated truncated CGIs are missing a sequence that is needed to recruit factors that causes their aberrant methylation.

To overcome these two caveats, I next asked whether aberrantly methylated CGIs in their native loci are main targets of DNMTs. To address that I re-expressed DNMT3B in the hypomethylated HCT116 derivative DKO. I found that aberrantly methylated CGIs gained substantially less methylation than normally methylated CGIs and CGIs overlapping H3K36me3. This suggested that aberrantly methylated CGIs are not main targets of DNMT3B.

To validate this result, I hypomethylated the genome of HCT116 by treatment with 5-Aza-2′-deoxycytidine and subsequently expanded the cells to allow remethylation. Cells were collected at different time points to assess the remethylation rate of CGIs. Aberrantly methylated CGIs recover slower than CGIs overlapping H3K36me3 and normally methylated CGIs. However, the differences between the three of groups of CGIs were subtle. Taken together

these results suggest that the aberrantly methylated CGIs are not directly targeted by DNMTs.

7.1 Optimizing a hypomethylation tool for *in vivo* verification of the results

In my study, I hypomethylated the genome of HCT116 by treatment with 1μM 5-Aza-2′-deoxycytidine for 24hrs. However, this induced partial loss of methylation (50-60% of wild type methylation) and bsPCR of CDH7 CGI showed that some CDH7 alleles remained methylated at day 3 and 6 implicating that some cells remained methylated (chapter6, figure 6.2B, figure 6.4). This might mask some of the differences in CGIs remethylation rate and make these differences appear subtle. Therefore it would be of particular interest to optimize tools to induce further hypomethylation without inducing cytotoxicity and utilize these tools to verify my results *in vivo*.

7.2 understanding the dynamics of aberrant hypermethylation in cancer

The majority of the work in this thesis and the study done by (Serra et al., 2014; Fang et al. 2016) did not directly address the actual dynamics of aberrant hypermethylation *in vivo*. However, work done by (Serra et al., 2014; Fang et al. 2016) implies quick gain of methylation at early stage and active maintenance of that methylation during cancer progression. This active methylation maintenance is thought to occur through recruiting DNMTs to CGIs by sequence specific proteins (figure 7.1A, top panel). My data doesn't support this model as we have shown that integrated aberrantly methylated CGIs remained mostly unmethylated.

My work implies that low levels of methylation is targeted to aberrantly methylated CGIs in colorectal tumour resulting in a slow accumulation of aberrant methylation at CGIs as tumour progresses (figure 7.1, middle panel). If

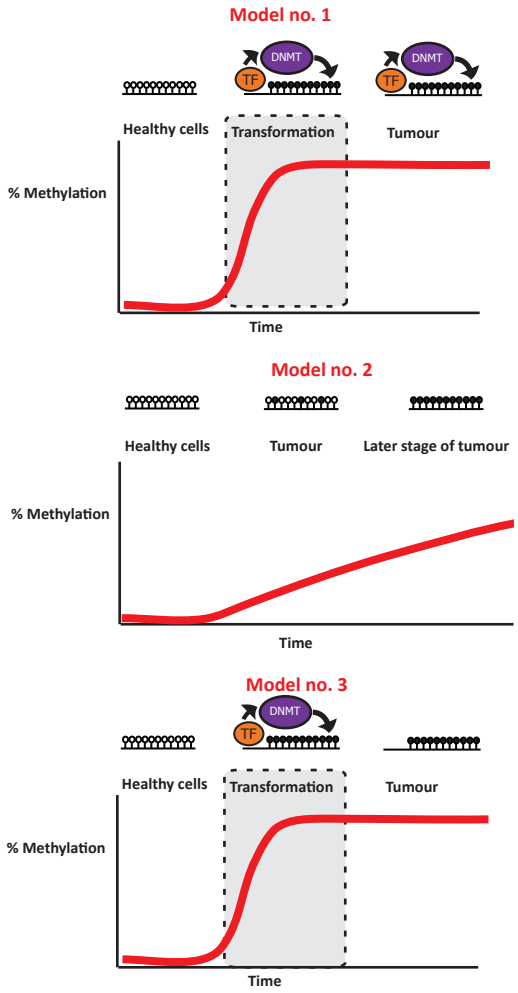
this model is true, the methylation level at CGIs will likely correlate with the tumor age. Another possible model is that DNMTs are targeted to aberrantly methylated CGIs by sequence specific proteins at a specific tumour stage resulting in quick gain of methylation. Aberrant methylation is subsequently passively maintained during tumour progression in the absence of the sequence specific proteins (figure 7.1A, lower panel).

The ideal way of testing these models would be to assess methylation levels at CGIs at different stages of the tumour in the same patient. However, this is not feasible in solid tumours as taking biopsy is invasive and these tumours are usually treated and/or removed by surgery after detection. Alternatively, and assuming that I can extend my results to other cancer types, blood samples are collected more often from leukemia patients after detection and thus can be used to assess methylation at different time points in a patient. One way this can be tested is by assessing the correlation between methylation levels and tumour stage.

Cai *et al.* have recently examined the methylation changes at four stages of tumour progression in the MMTV-PyMT mouse model (Cai *et al.*, 2018). They performed Enhanced RRBS to assess methylation and defined differentially methylated regions of ~182bp long and showed that the number of hypermethylated DMRs increases with tumour progression (figure 7.1B). It would be useful to use their data to calculate % mean methylation levels at CGIs and examine if that correlates with tumour stage.

Teasing out the exact dynamics of methylation gain at CGIs and possibly configuring the difference between it and the dynamics of normal methylation is challenging. One exciting experiment that can address this would be adopting the DNA methylation reporter established by Stelzer *et al.* to trace methylation changes at a CGI on a single cell resolution throughout tumour progression in vivo (Stelzer *et al.*, 2015). Tracing tumour cells by live-cell imaging would also

A Dynamics of aberrant hypermethylation in cancer



B Percentage of hypermethylated DMR at different satges of tumour progression in MMTV-PyMT mouse model

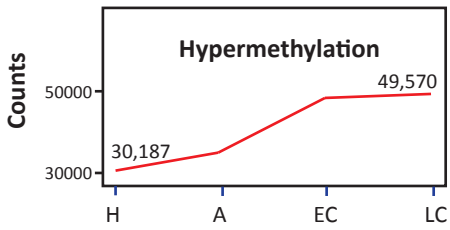


Figure 7.1:
A. Models for the dynamics of aberrant hypermethylation in cancer. In the first model, DNMTs are recruited to CGIs by sequence specific proteins upon transformation and throughout progression. This leads to quick gain of aberrant methylation at transformation which is actively maintained during tumour progression. In the second model, aberrant methylation slowly accumulaties at CGIs as tumour progrsses. In the third model, quick gain of methylation takes place upon transformation because DNMTs are recruited by sequence specific proteins. Subsequently, aberrant methylation is passively maintained by DNMTs in the absence of the sequence specific proteins.
B. The number of DMRs increases with tumour progression in PyMT mouse model. Line graph showing the hypermethylated DMRs counts (y-axis) at four different stages of tumour progression: hyperplasia (H), adenoma/MIN (A), early carcinoma (EC) and late carcinoma (LC). Percentage is shown as dashed line and counts are shown as solid line (adapted from Cai *et al*., 2018).

de-convolute if the gain in methylation will take place in observed in all tumour cells in or in a subpopulation.

7.3 Is seeding methylation a prerequisite for aberrant hypermethylation of aberrantly methylated CGIs?

In prostate cancer, prior silencing and seeding of random CpGs methylation within GSTP1 promoter was sufficient to drive its aberrant hypermethylation (Song *et al.*, 2002). In line with this, my data from chapter 5 suggests that the methylation levels of CGIs may enhance deposition of methylation by DNMT3B2 in DKO cells. It would be interesting to investigate whether integrating CGIs with low methylation levels will trigger their aberrant hypermethylation in HCT116. That would support the "methylation seeding" model proposed previously (Clark and Melki, 2002). According to this model, random methylation takes place at CGIs in both normal and cancer cells. In normal cells, the active transcription of the gene promotes demethylation of its lowly methylated CGI. In cancer cells, the gene become silenced and in the absence of active transcription the methylation seeds are followed by a "snow storm" of methylation. This model proposes active transcription as the main protector against methylation. However, the majority of the genes that get aberrantly hypermethylated in cancer are already repressed in the pre-cancerous tissue (Sproul *et al.*, 2012). This suggests that the part of the "methylation seeding" model that implicate active transcription as the main protecting factor against hypermethylation, may apply only on a few genes.

7.4 Do CGIs overlapping H3K36me3 and normally methylated CGIs become methylated at ectopic sites?

In chapter 3, aberrantly methylated CGIs remained mainly unmethylated when ectopically integrated using piggyBac. Results from chapter 5 and 6, suggest that CGIs overlapping H3K36me3 and normally methylated CGIs are more

efficiently targeted by DNMT3B than aberrantly methylated CGIs. Taking this finding in consideration, these two groups of CGIs can be tested as positive controls that piggyBac minimal repeats don't prevent deposition of methylation at the integrated CGIs. Moreover, these two groups of CGIs may be an additional positive control for HCT116 *de novo* methylation activity. However, having these as positive biological controls has the caveat that these CGIs are usually found in gene bodies whereas aberrantly methylated CGIs are in promoters.

Another possible control to overcome this caveat, would be integrating the aberrantly methylated CGIs with methylation seeds as suggested in the previous section. If these CGIs become hypermethylated that would verify that aberrantly methylated CGIs don't become methylated at ectopic loci unless they have methylation seeds.

7.5 Therapeutic implications

One exciting aspect of this study was uncovering that CGIs remain mostly unmethylated after their reintroduction into the genome of cancer cells. This suggests that reintroducing unmethylated version of tumour suppressor genes that have been hypermethylated and suppressed to the genome of cancer cells might be a potential therapeutic opportunity.

5-Aza-2′-deoxycytidine has been reported to induce an interferon response in cancer by demethylating the endogenous retrovirus (ERV) leading to their expression (Chiappinelli *et al.*, 2015). ERV expression triggers the dsRNA response leading to interferon response and apoptosis. Treatment with 5-Aza-2′-deoxycytidine thereby sensitizes to immunotherapy in a melanoma mouse model. However in the clinic, 5-Aza-2′-deoxycytidine is given at low dose to avoid DNA damage and its cytotoxic effects (Tsai *et al.*, 2012). The low dose of 5-Aza-2′-deoxycytidine triggers slow antitumor effects. A drug that can trigger higher degree of hypomethylation without DNA damage and cytotoxicity might

be more efficient in sensitizing to immunotherapy. Future studies testing a non-cytidine analogue in a cancer mouse model will help test this hypothesis.

Appendix

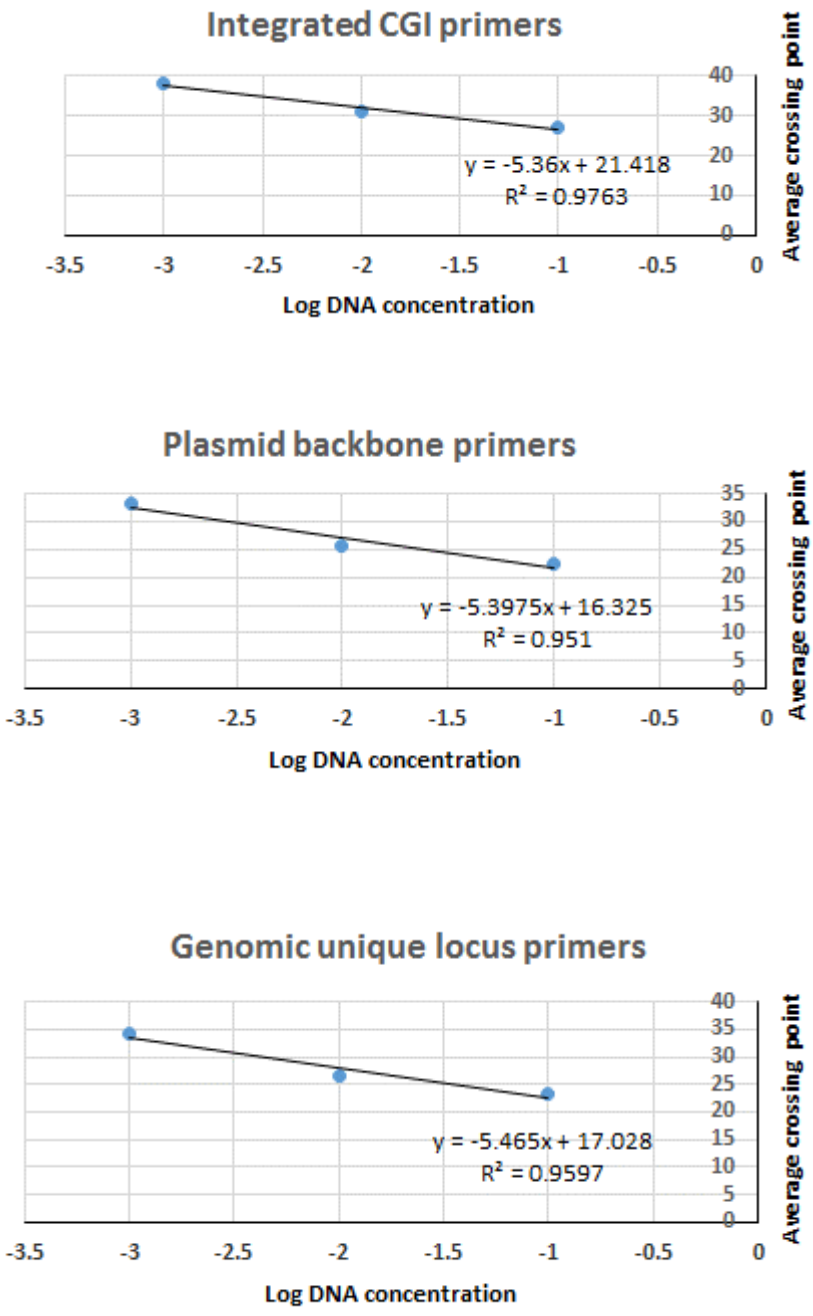


Figure S3.1: Standard curves for qPCR primers. Average crossing point vs. log DNA concentration for primers amplifying an integrated CGI by piggyBac (top panel), the piggyBac plasmid backbone (middle panel), and a unique genomic sequence (bottom panel).

Methylation status of truncated *MLH1* CGIs at ectopic loci in HCT116

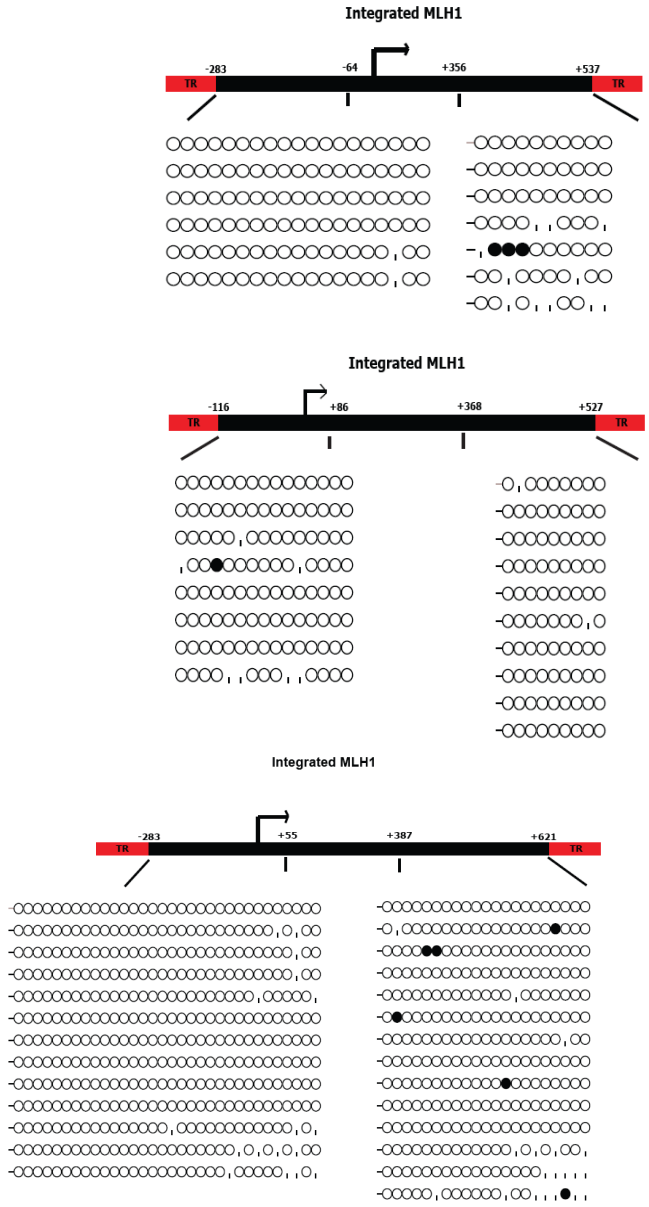
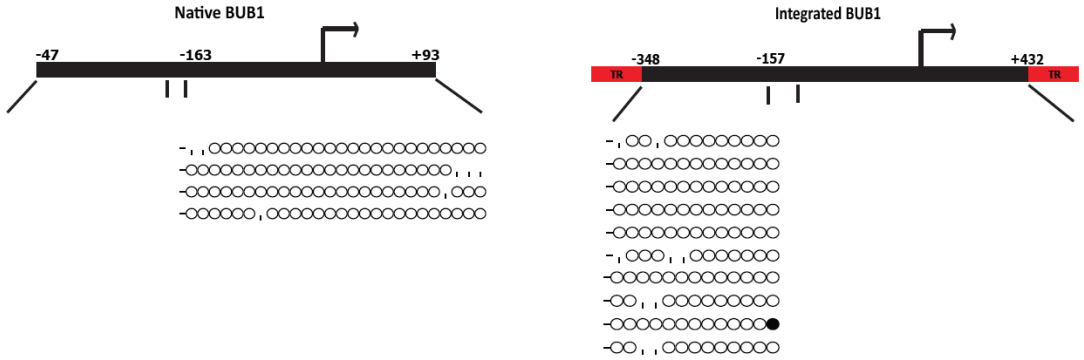


Figure S3.2: Truncated *MLH1* CGIs remained unmethylated at ectopic locations. Clonal bisulphite sequencing for 643bp (top panel), 820bp (middle panel), and 904bp (bottom panel) ectopically integrated *MLH1 CGI* fragments. Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively. Empty spaces reflect ambiguous base calls in Sanger sequencing reads.

A Methylation status of the *BUB1* CGI at native and ectopic loci in HCT116



B Methylation status of the *ZFP42* CGI at native and ectopic loci in HCT116

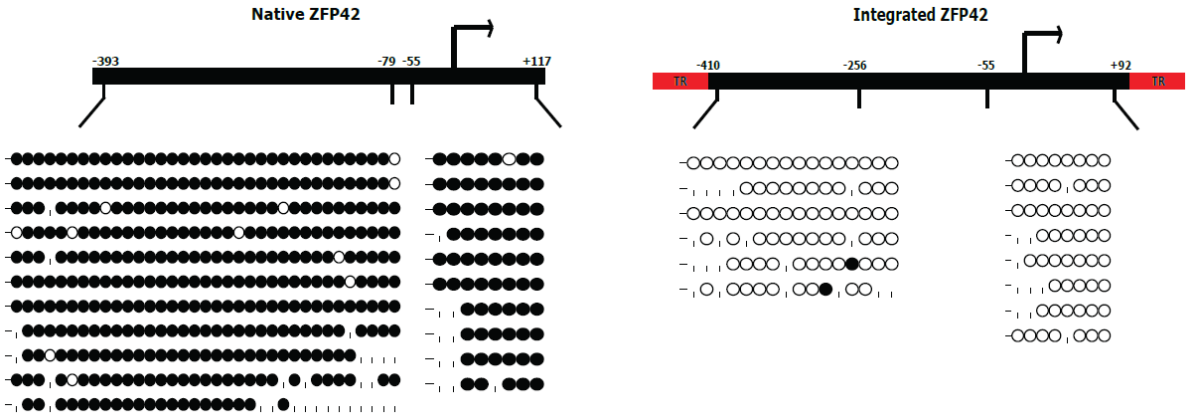
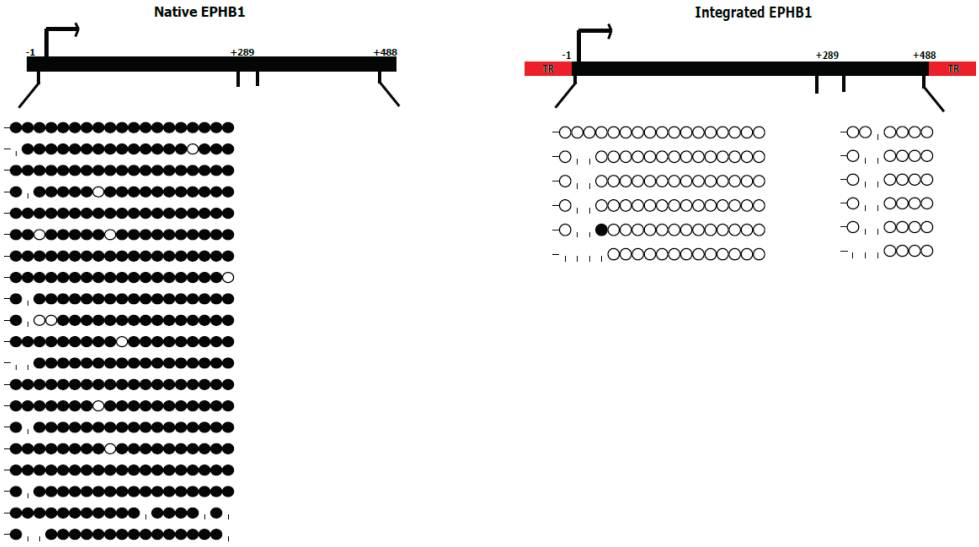


Figure S3.3: Integrated *BUB1* and *ZFP42* CGIs remained unmethylated at ectopic loci in HCT116. Clonal bisulphite sequencing for the native and ectopically integrated *BUB1* (A) and *ZFP42* (B). Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively.

A Methylation status of the *EPHB1* CGI at native and ectopic loci in HCT116



B Methylation status of the *GATA4* CGI at native and ectopic loci in HCT116

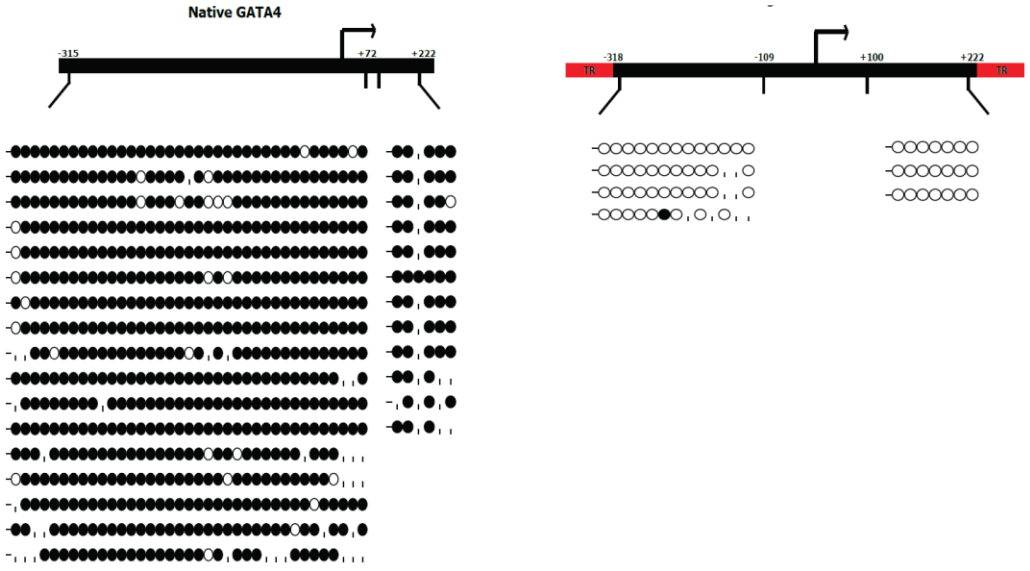
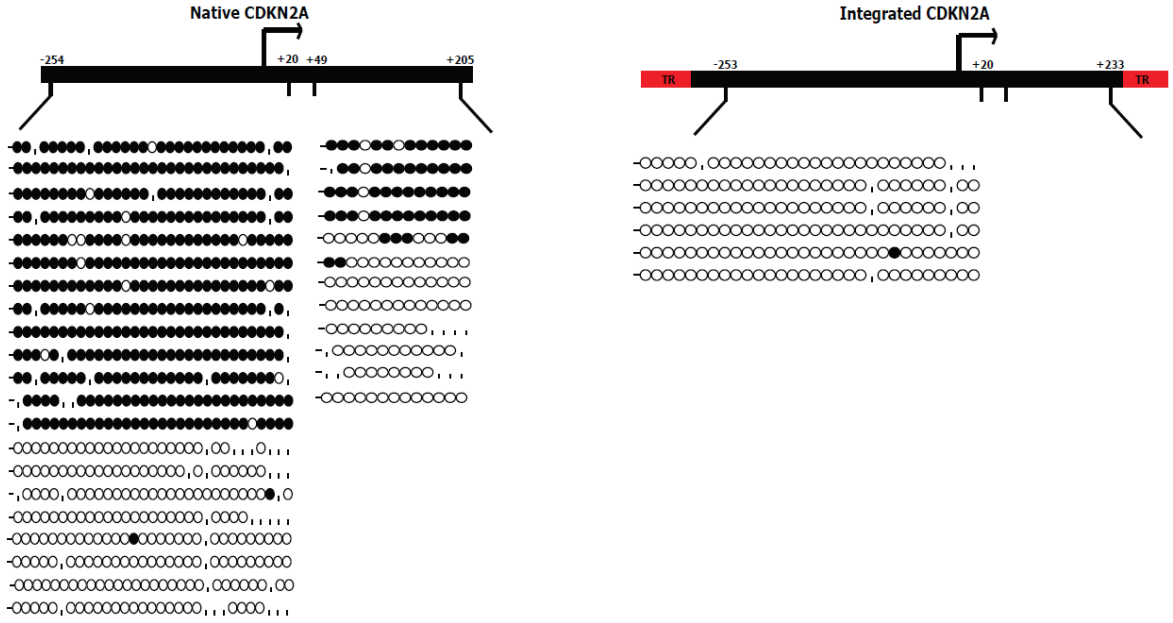


Figure S3.4: Integrated *EPHB1* and *GATA4* CGIs remained unmethylated at ectopic loci in HCT116. Clonal bisulphite sequencing for the native and ectopically integrated *EPHB1* (A) and *GATA4* (B). Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively.

A Methylation status of the *CDKN2A* CGI at native and ectopic loci in HCT116



B Methylation status of the *SFRP1* CGI at native and ectopic loci in HCT116

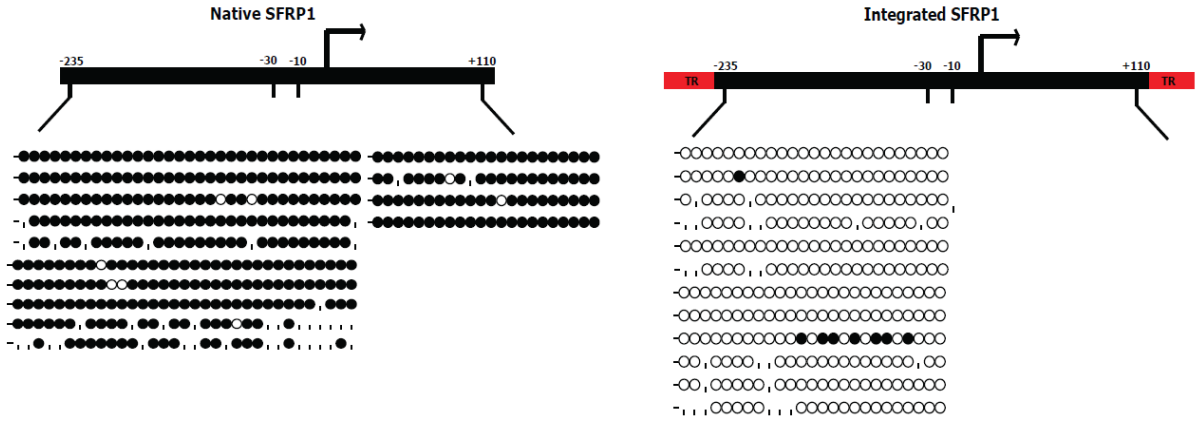


Figure S3.5: Integrated *CDKN2A* and *SFRP1* CGIs remained unmethylated at ectopic loci in HCT116. Clonal bisulphite sequencing for the native and ectopically integrated *CDKN2A* (A) and *SFRP1* (B). Each horizontal line represents a single allele. CpG dinucleotides are represented by circles. Black and white circles indicate methylated and unmethylated CpG, respectively.

R scripts

Preparing RRBS BEDgraph file

Removing the file header

tail -n +2 Roza_RRBS_151217_DEDUP_allSAMPLE_BEDgraph.bed > Roza_RRBS_151217_DEDUP_allSAMPLE_BEDgraph_noheader.bed

Compressing the file:

gzip Roza_RRBS_151217_DEDUP_allSAMPLE_BEDgraph_noheader.bed

Intersecting the RRBS file with CGIs coordinates file (illingworth38)

bedtools intersect -a illingworth38.bed -b Roza_RRBS_151217_DEDUP_allSAMPLE_BEDgraph_noheader.bed.gz -wa -wb > Roza_RRBS_all_CGI_110118.bed

Summing the methylation reads and total reads of all CpGs within a CGI

Groupby the first 4 columns

bedtools groupby -i Roza_RRBS_all_CGI_110118.bed -g 1-4 -c 8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67 -o sum > RRBS_ALL_CGI_GB_11012018.bed

#Writing back the header of the RRBS file

All<- read.table("RRBS_ALL_CGI_GB_11012018.bed")

header<-c("seqnames","start", "end" , "CGI","X3BKO_D0.T","X3BKO_D0.M","HCT116_D13_AZA.T" ,"HCT116_D13_AZA.M" ,"HCT116_D28_AZA.T","HCT116_D28_AZA.M","X1KO_D40_cont.T",

"X1KO_D40_cont.M", "X3BKO_D_40_cont.T", "X3BKO_D_40_cont.M" ,"DKO_3B2.T" , "DKO_3B2.M" , "X1KO_AZA_D3.T" ,"X1KO_AZA_D3.M" ,"X1KO_D6_AZA.T" ,"X1KO_D6_AZA.M" , "X3BKO_D6_AZA.T" , "X3BKO_D6_AZA.M" , "X1KO_D13_AZA.T" ,"X1KO_D13_AZA.M" ,"X1KO_D40_AZA.T" , "X1KO_D40_AZA.M" , "DKO_3B2_cd.T" , "DKO_3B2_cd.M" , "X1KO_D0.T" ,"X1KO_D0.M" ,"HCT116_AZA_D3.T" , "HCT116_AZA_D3.M" , "X3BKO_D13_AZA.T", "X3BKO_D13_AZA.M" , "X1KO_D22_AZA.T" ,"X1KO_D22_AZA.M" , "X1KO_D28_AZA.T" ,"X1KO_D28_AZA.M" , "X3BKO_D28_AZA.T" , "X3BKO_D28_AZA.M" , "HCT116_D0.T" ,"HCT116_D0.M" ,"HCT116_D16_AZA.T" , "HCT116_D16_AZA.M" , "X1KO_D16_AZA.T" , "X1KO_D16_AZA.M" , "HCT116_D22_AZA.T" ,"HCT116_D22_AZA.M" ,"X3BKO_D22_AZA.T" , "X3BKO_D22_AZA.M" , "DKO.T" ,"DKO.M", "X3BKO_AZA_D3.T", "X3BKO_AZA_D3.M" , "HCT116_D6_AZA.T" , "HCT116_D6_AZA.M" , "X3BKO_D16_AZA.T" , "X3BKO_D16_AZA.M"

```
, "HCT116_D40_AZA.T", "HCT116_D40_AZA.M", "X3BKO_D40_AZA.T" ,"X3BKO_D40_AZA.M"
,  "HCT116_D40_cont.T", "HCT116_D40_cont.M")

 colnames(All)<- header

#Filtering out CGIs with less than 20 reads

ALL_filtered<-
All[apply(All[c(5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,5
9,61,63)],1,function(z) !any(z<20)),]

RRBS_CGI_GB<- ALL_filtered

#Taking only HCT116 data

HCT_CGI<-
RRBS_CGI_GB[,c("seqnames","start","end","CGI","HCT116_D0.T","HCT116_D0.M","HCT116_
AZA_D3.T","HCT116_AZA_D3.M","HCT116_D6_AZA.T","HCT116_D6_AZA.M","HCT116_D13_
AZA.T","HCT116_D13_AZA.M","HCT116_D16_AZA.T","HCT116_D16_AZA.M","HCT116_D22_
AZA.T","HCT116_D22_AZA.M","HCT116_D28_AZA.T","HCT116_D28_AZA.M","HCT116_D40_
AZA.T","HCT116_D40_AZA.M","HCT116_D40_cont.T", "HCT116_D40_cont.M")]

#Calculating %meth for each CGIs

HCT_CGI<-
transform(HCT_CGI,HCT116_D0=HCT116_D0.M/HCT116_D0.T,HCT116_AZA_D3=HCT116_A
ZA_D3.M/HCT116_AZA_D3.T,HCT116_D6_AZA=HCT116_D6_AZA.M/HCT116_D6_AZA.T,HC
T116_D13_AZA
=HCT116_D13_AZA.M/HCT116_D13_AZA.T,HCT116_D16_AZA=HCT116_D16_AZA.M/HCT11
6_D16_AZA.T,HCT116_D22_AZA=HCT116_D22_AZA.M/HCT116_D22_AZA.T,HCT116_D28_
AZA=HCT116_D28_AZA.M/HCT116_D28_AZA.T,HCT116_D40_AZA=HCT116_D40_AZA.M/H
CT116_D40_AZA.T,HCT116_D40_cont=HCT116_D40_cont.M/HCT116_D40_cont.T)

#Taking only CGIs with methylation => .5

HCT_meth<- HCT_CGI[HCT_CGI$HCT116_D0>=.5,]

nrow(HCT_meth)

7287

#calculating the chance that the increase in methylation is due to outgrowth using mixed linear
model

library(lme4)

library(reshape)

rm(list=ls())

setwd("C:/MyWorkspace ")
```

```
#Change this to whatever days you want to use

daysToUse <- c(3,6,13,16,22)


HCT116_AllCGI_Methylation_Data <-
read.table("RRBS_ALL_CGI_GB_HCT_Eddie_060218.bed", header=TRUE)

#Taking only CGIs with methylation >= 50%

HCT116_MethylatedCGI_Methylation_Data <-
HCT116_AllCGI_Methylation_Data[HCT116_AllCGI_Methylation_Data$HCT116_D0 >= 0.5,]


methyNorm <- HCT116_MethylatedCGI_Methylation_Data[,23:30]

colnames(methyNorm) <- c(0,3,6,13,16,22,28,40)


#Calculating the relative change in methylation from day 0

for(i in dim(methyNorm)[2]:2){

  methyNorm[,i] <- methyNorm[,i] / methyNorm[,"0"]

}

methyNorm <- methyNorm[, colnames(methyNorm) %in% daysToUse]


#Melt the data into a shape for the MLM

methyNorm <- melt(methyNorm)

methyNorm <- cbind(methyNorm, rep(HCT116_MethylatedCGI_Methylation_Data[,4],
length(daysToUse)))

colnames(methyNorm) <- c("Time", "Methylation", "CGI")

methyNorm$Time <- scale(as.numeric(levels(methyNorm$Time))[methyNorm$Time])


#Calculate mixed linear model

mlm.model <- lmer(Methylation ~ Time + (1+Time|CGI), data=methyNorm, REML=FALSE)
```

```
#Calculate a simple null model

null.model <- lm(Methylation~ Time ,data=methyNorm)



#Test the difference between the two

anova.test <- anova(mlm.model,null.model)

cat("The chance that each CpG is changing with identical rate (i.e. outgrowth) is",
anova.test$`Pr(>Chisq)`[2])



#DNMT3B experiment

All<- read.table("RRBS_ALL_CGI_GB_11012018.bed")

header<-c("seqnames","start", "end" ,
"CGI","X3BKO_D0.T","X3BKO_D0.M","HCT116_D13_AZA.T" ,"HCT116_D13_AZA.M"
,"HCT116_D28_AZA.T","HCT116_D28_AZA.M","X1KO_D40_cont.T", "X1KO_D40_cont.M",
"X3BKO_D_40_cont.T",  "X3BKO_D_40_cont.M" ,"DKO_3B2.T"     ,    "DKO_3B2.M"       ,
"X1KO_AZA_D3.T"     ,"X1KO_AZA_D3.M"     ,"X1KO_D6_AZA.T"    ,"X1KO_D6_AZA.M"    ,
"X3BKO_D6_AZA.T" ,   "X3BKO_D6_AZA.M"  ,  "X1KO_D13_AZA.T"    ,"X1KO_D13_AZA.M"
,"X1KO_D40_AZA.T" , "X1KO_D40_AZA.M"   , "DKO_3B2_cd.T" ,    "DKO_3B2_cd.M"     ,
"X1KO_D0.T"         ,"X1KO_D0.M"         ,"HCT116_AZA_D3.T"  , "HCT116_AZA_D3.M"  ,
"X3BKO_D13_AZA.T",   "X3BKO_D13_AZA.M" ,  "X1KO_D22_AZA.T"    ,"X1KO_D22_AZA.M"
,  "X1KO_D28_AZA.T" ,"X1KO_D28_AZA.M" , "X3BKO_D28_AZA.T",  "X3BKO_D28_AZA.M" ,
"HCT116_D0.T","HCT116_D0.M","HCT116_D16_AZA.T", "HCT116_D16_AZA.M" ,
"X1KO_D16_AZA.T" ,   "X1KO_D16_AZA.M"  ,  "HCT116_D22_AZA.T"
,"HCT116_D22_AZA.M"  ,"X3BKO_D22_AZA.T", "X3BKO_D22_AZA.M" , "DKO.T" ,"DKO.M" ,
"X3BKO_AZA_D3.T"   , "X3BKO_AZA_D3.M" , "HCT116_D6_AZA.T" , "HCT116_D6_AZA.M"
,"X3BKO_D16_AZA.T", "X3BKO_D16_AZA.M","HCT116_D40_AZA.T","HCT116_D40_AZA.M",
"X3BKO_D40_AZA.T", "X3BKO_D40_AZA.M","HCT116_D40_cont.T","HCT116_D40_cont.M")

> colnames(ALL)<- header

ALL_filtered<-
All[apply(All[c(5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,5
9,61,63)],1,function(z) !any(z<20)),]

#Taking only cell lines and DNMT3B  data

celllines<-
ALL_filtered[,c("seqnames","start","end","CGI","HCT116_D0.T","HCT116_D0.M","X1KO_D0.T","
X1KO_D0.M","X3BKO_D0.T","X3BKO_D0.M","DKO.T","DKO.M",
```

```
"DKO_3B2.T","DKO_3B2.M","DKO_3B2_cd.T","DKO_3B2_cd.M")]

#Calculating %meth for each CGIs

celllines<-
transform(celllines,HCT116_D0=HCT116_D0.M/HCT116_D0.T,X1KO_D0=X1KO_D0.M/X1KO_
D0.T,X3BKO=X3BKO_D0.M/X3BKO_D0.T,DKO=DKO.M/DKO.T,

DKO_3B2=DKO_3B2.M/DKO_3B2.T,DKO_3B2_cd=DKO_3B2_cd.M/DKO_3B2_cd.T)


#To load the file

celllines<- read.table("RRBS_ALL_CGI_GB_celllines_Eddie_270218.bed", header=TRUE)


#Calculating which CGIs gained methylation in DNMT3B by fisher test

#Taking total and methylated reads of DKO, DKO_DNMT3B and DNMT3B_cd:

celllines<- transform(celllines,DKO_3B2_diff=DKO_3B2-DKO,DKO_3B2_cd_diff=DKO_3B2_cd-
DKO)

DNMT3B<-
celllines[,c("seqnames","start","end","CGI","DKO.T","DKO.M","DKO_3B2.T","DKO_3B2.M","DKO
_3B2_cd.T","DKO_3B2_cd.M")]

DNMT3B<- transform(DNMT3B,DKO.U=DKO.T-DKO.M,DKO_3B2.U=DKO_3B2.T-
DKO_3B2.M,DKO_3B2_cd.U=DKO_3B2_cd.T-DKO_3B2_cd.M)


#Calculating P-value by fisher test:

DNMT3B_sub<- DNMT3B[,c("DKO_3B2.M","DKO_3B2.U","DKO.M","DKO.U")]

DNMT3B_sub<- transform(DNMT3B_sub, p_value=apply(DNMT3B_sub,1,function(x)
fisher.test(matrix(x, nrow=2, ncol=2),alternative = "greater")$p.value))

DNMT3B<- cbind(celllines,DNMT3B_sub)

DNMT3B[,25:28]<- NULL

#Calculating p-values adjusted by Benjamini and Hochberg method “BH”.

DNMT3B$p_adj <- p.adjust(DNMT3B$p_value, method="BH")

DNMT3B_gain_adjp<- DNMT3B[apply(DNMT3B[c(26)],1,function(z) any(z<.05)),]
```

```
#CGIs only those showing gain in methylation greater than 20%

DNMT3B_PADJ_.2<- DNMT3B_gain_adjp[DNMT3B_gain_adjp$DKO_3B2_diff>0.2,]

nrow(DNMT3B_PADJ_.2)

# 1653

#Scatter plot for all CGIs in 1KO vs DKO_3B2, highlighting CGIs gaining methylation:

#prepare data

DNMT3B_PADJ_.2_id<-DNMT3B_PADJ_.2$CGI

celllines$colour="black"

celllines$colour[celllines$CGI%in%DNMT3B_PADJ_.2_id]="red"

pdf("Scatter_plots_CGIs in DKOvs DKO_DKO3B2_060518.pdf", useDingbats=FALSE)

plot(celllines$X1KO_D0,celllines$DKO_3B2,pch=20,col=celllines$colour,xlab="Mean methylaton
of CGIs in 1KO",ylab="Mean methylaton of CGIs in DKO expressing DNMT3B2", main="CGIs
that gain methylation in DKO cells expressing DNMT3B2")

dev.off()
```

```
# Plotting gain of methylation by DNMT3B2 at CGIs overlapping H3K36me3

CGIs H3K36me3_meth<- celllines[celllines$CGI%in%H3K36me3_CGI,]

CGIs_overlapping_H3K36me3<- H3K36me3_meth[,23]*100

CGIs_non_overlapping_H3K36me3<- celllines[!celllines$CGI%in%H3K36me3_CGI,]

CGIs_non_overlapping_H3K36me3<- CGIs_non_overlapping_H3K36me3[,23]*100

g<-list(CGIs_overlapping_H3K36me3,CGIs_non_overlapping_H3K36me3)

names(g)<- c("CGIs_overlapping_H3K36me3","CGIs_non_overlapping_H3K36me3")

pdf("H3K36me3_meth_change_DKO_DNMT3B_boxplot.pdf", useDingbats=FALSE)

par(cex.axis=.7)
```

```
stripchart(g,  method="jitter", vertical=TRUE,  jitter=0.1,pch = 20,
col=c("#D60000","#1F78B4"),at=c(.2,.9),xlim=c(0,2.1),main="Change in % mean methylation
after expressing DNMT3B-2 in DKO")

boxplot(g, add=TRUE, outline=FALSE, boxwex = 0.2, at=c(.2,.9), ylab= "Change in % mean
methylation")

legend(1.1,80,legend=c("CGIs_overlapping_H3K36me3","CGIs_non_overlapping_H3K36me3"),

col=c("#D60000","#1F78B4"),pch=c(19,19), bty = "n",cex=.8, pt.cex = 1)

dev.off()

#Calculating the methylation gain median and performing Student's t-test

summary(CGIs_overlapping_H3K36me3)

summary(CGIs_non_overlapping_H3K36me3)

t.test(CGIs_overlapping_H3K36me3,CGIs_non_overlapping_H3K36me3, alternative =
c("greater"))

**********

#Heat map and line chart and slopes for H3K36me3

#Preparing the data

HCT<- read.table("RRBS_ALL_CGI_GB_HCT_Eddie_060218.bed", header=TRUE)

HCT<-HCT[complete.cases(HCT),]

#Taking only CGIs with methylation => .5

HCT_meth<- HCT[HCT$HCT116_D0>=.5,]

nrow(HCT_meth)

#7287

HCT_meth_D0TO40<- HCT_meth[,c(4,23:30)]

rownames(HCT_meth_D0TO40)<- HCT_meth_D0TO40[,1]

HCT_meth_D0TO40$CGI<- NULL


#Calculating the slope from day 6 to 22

  NumberOfIslands = 7287

NumberOfSamples = 1
```

```
slopes_CGI_D6_D22<- matrix(data=NA,nrow=NumberOfIslands, ncol=NumberOfSamples)
for(row in 1:NumberOfIslands){
     y <-  as.numeric(HCT_meth_D0TO40[row,3:6])
    x <- as.vector(c(6,13,16,22))
    if(sum(is.nan(x)+is.na(x))==0){
 lmdata <- lm(y ~ x)
      slopes_CGI_D6_D22[row] <- summary(lmdata)$coefficients[2,1]
   }
 }
```

References

Acevedo, L. G. *et al.* (2008) 'Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors.', *Cancer research*. American Association for Cancer Research, 68(8), pp. 2641–51. doi: 10.1158/0008-5472.CAN-07-5590.

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

Alhejaily, A. *et al.* (2014) 'Inactivation of the CDKN2A Tumor-Suppressor Gene by Deletion or Methylation Is Common at Diagnosis in Follicular Lymphoma and Associated with Poor Clinical Outcome', *Clinical Cancer Research*, 20(6), pp. 1676–1686. doi: 10.1158/1078-0432.CCR-13-2175.

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

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. doi: 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, Published online: 03 September 2008; | doi:10.1038/nature07273*. Nature Publishing Group, 455(7214), p. 822. doi: 10.1038/nature07273.

Bachu, R., Bergareche, I. and Chasin, L. A. (2015) 'CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining', *Biotechnology and Bioengineering*, 112(10), pp. 2154–2162. doi: 10.1002/bit.25629.

Bae, Y. K. *et al.* (2004) 'Hypermethylation in Histologically Distinct Classes of Breast Cancer', *Clinical Cancer Research*, 10(18), pp. 5998–6005. doi: 10.1158/1078-0432.CCR-04-0667.

Baker, D. J. *et al.* (2011) 'Clearance of p16 Ink4a -positive senescent cells delays ageing-associated disorders'. doi: 10.1038/nature10600.

Ballikaya, S. *et al.* (2014) 'De Novo proteome analysis of genetically modified tumor cells by a metabolic labeling/azide-alkyne cycloaddition approach.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 13(12), pp. 3446–56. doi: 10.1074/mcp.M113.036665.

Bannister, A. J. and Kouzarides, T. (2011) 'Regulation of chromatin by histone modifications', *Cell Research*. Nature Publishing Group, 21(3), pp. 381–395. doi: 10.1038/cr.2011.22.

Barau, J. *et al.* (2016) 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity.', *Science (New York, N.Y.)*. American Association for the Advancement of

Science, 354(6314), pp. 909–912. doi: 10.1126/science.aah5143.

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*. American Society for Biochemistry and Molecular Biology, 289(7), pp. 4106–15. doi: 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. 243–7. doi: 10.1038/nature14176.

Belinsky, S. A. *et al.* (1998) 'Aberrant methylation of p16(INK4a) 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*. National Academy of Sciences, 95(20), pp. 11891–6. doi: 10.1073/PNAS.95.20.11891.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society. Series B (Methodological)*. WileyRoyal Statistical Society, pp. 289–300. doi: 10.2307/2346101.

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.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 289(1), pp. 379–86. doi: 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. doi: 10.1038/ng.969.

Bernstein, B. E. *et al.* (2006) 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells', *Cell*, 125(2), pp. 315–326. doi: 10.1016/j.cell.2006.02.041.

Bestor, T. H. (2000) 'The DNA methyltransferases of mammals', *Human Molecular Genetics*. Oxford University Press, 9(16), pp. 2395–2402. doi: 10.1093/hmg/9.16.2395.

Bestor, T. H., Ingram, V. M. and Buchanan, J. M. (1983) 'Two DNA methyltransferases from murine erythroleukemia cells: Purification, sequence specificity, and mode of interaction with DNA (facilitated transfer)', 80, pp. 5559–5563. Available at: http://www.pnas.org.ezproxy.is.ed.ac.uk/content/80/18/5559.long (Accessed: 16 October 2017).

Bhasin, J. M. and Ting, A. H. (2016) 'Goldmine integrates information placing genomic ranges into meaningful biological contexts', *Nucleic Acids Research*, 44(12), pp. 5550–5556. doi: 10.1093/nar/gkw477.

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

Bird, A. P. (1980) 'DNA methylation and the frequency of CpG in animal DNA.', *Nucleic acids research*. Oxford University Press, 8(7), pp. 1499–504. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6253938 (Accessed: 4 December 2017).

Blackledge, N. P. *et al.* (2010) 'CpG Islands Recruit a Histone H3 Lysine 36 Demethylase', *Molecular Cell*, 38(2), pp. 179–190. doi: 10.1016/j.molcel.2010.04.009.

Bock, C. *et al.* (2005) 'BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing', *Bioinformatics*, 21(21), pp. 4067–4068. doi: 10.1093/bioinformatics/bti652.

Borgel, J. *et al.* (2010) 'Targets and dynamics of promoter DNA methylation during early mouse development', *Nature Publishing Group*, 42. doi: 10.1038/ng.708.

Bostick, M. *et al.* (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells.', *Science (New York, N.Y.)*, 317(5845), pp. 1760–4. doi: 10.1126/science.1147939.

Bourc'his, D. *et al.* (2001) 'Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos.', *Current biology : CB*, 11(19), pp. 1542–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11591324 (Accessed: 9 November 2017).

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

Boyer, L. A. *et al.* (2006) 'Polycomb complexes repress developmental regulators in murine embryonic stem cells', *Nature*, 441(7091), pp. 349–353. doi: 10.1038/nature04733.

Cai, Y. *et al.* (2018) 'Epigenetic alterations to Polycomb targets precede malignant transition in a mouse model of breast cancer', *Scientific Reports*. Nature Publishing Group, 8(1), p. 5535. doi: 10.1038/s41598-018-24005-x.

Cao, X. *et al.* (2000) 'Conserved plant genes with similarity to mammalian de novo DNA methyltransferases.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 97(9), pp. 4979–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10781108 (Accessed: 15 December 2017).

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

De Carvalho, D. D. *et al.* (2012) 'DNA methylation screening identifies driver epigenetic events of cancer cell survival.', *Cancer cell*. Elsevier, 21(5), pp. 655–67. doi: 10.1016/j.ccr.2012.03.045.

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*. American Society for Microbiology, 24(20), pp. 9048–58. doi: 10.1128/MCB.24.20.9048-9058.2004.

Chen, Z.-X. *et al.* (2005) 'Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family', *Journal of Cellular Biochemistry*, 95(5), pp. 902–917. doi: 10.1002/jcb.20447.

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. doi: 10.1016/j.cell.2015.07.011.

Chuang, L. S. *et al.* (1997) 'Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 277(5334), pp. 1996–2000. doi: 10.1126/SCIENCE.277.5334.1996.

Citterio, E. *et al.* (2004) 'Np95 is a histone-binding protein endowed with ubiquitin ligase activity.', *Molecular and cellular biology*, 24(6), pp. 2526–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14993289 (Accessed: 17 December 2017).

Clark, S. J. and Melki, J. (2002) 'DNA methylation and gene silencing in cancer: which is the guilty party?', *Oncogene*. Nature Publishing Group, 21(35), pp. 5380–5387. doi: 10.1038/sj.onc.1205598.

Cokus, S. J. *et al.* (2008) 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature*. Nature Publishing Group, 452(7184), pp. 215–219. doi: 10.1038/nature06745.

Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas Systems', *Science*, 339(6121), pp. 819–823. doi: 10.1126/science.1231143.

Coulondre, C. *et al.* (1978) 'Molecular basis of base substitution hotspots in Escherichia coli.', *Nature*, 274(5673), pp. 775–80. Available at: http://www.ncbi.nlm.nih.gov/pubmed/355893 (Accessed: 4 December 2017).

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

Cunningham, J. M. *et al.* (1998) 'Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability.', *Cancer research*, 58(15), pp. 3455–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9699680 (Accessed: 30 November 2017).

Davies, M. A. and Samuels, Y. (2010) 'Analysis of the genome to personalize therapy for melanoma', *Oncogene*. Nature Publishing Group, 29(41), pp. 5545–5555. doi: 10.1038/onc.2010.323.

Dean, W. (2016) 'Pathways of DNA Demethylation', in *Advances in experimental medicine and biology*, pp. 247–274. doi: 10.1007/978-3-319-43624-1_11.

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. 26114–20. doi: 10.1074/jbc.M109.089433.

van Dillen, I. J. *et al.* (2002) 'Influence of the bystander effect on HSV-tk/GCV gene therapy. A review.', *Current gene therapy*, 2(3), pp. 307–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12189718 (Accessed: 1 January 2018).

Ding, S. *et al.* (2005) 'Efficient Transposition of the piggyBac (PB) Transposon in Mammalian Cells and Mice', *Cell*, 122(3), pp. 473–483. doi: 10.1016/j.cell.2005.07.013.

Doetschman, T. *et al.* (1987) 'Targetted correction of a mutant HPRT gene in mouse embryonic stem cells', *Nature*, 330(6148), pp. 576–578. doi: 10.1038/330576a0.

Doi, A. *et al.* (2009) 'Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts', *Nature Genetics*. Nature Publishing Group, 41(12), pp. 1350–1353. doi: 10.1038/ng.471.

Du, Z. *et al.* (2010) 'DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination.', *Science signaling*, 3(146), p. ra80. doi: 10.1126/scisignal.2001462.

Dunn, D. B. and Smith, J. D. (1955) 'Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli.', *Nature*, 175(4451), pp. 336–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13235889 (Accessed: 6 November 2017).

Duymich, C. E. *et al.* (2016) 'DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells', *Nature Communications*. Nature Publishing Group, 7, p. 11453. doi: 10.1038/ncomms11453.

Eads, C. A. *et al.* (1999) 'CpG island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression', *Cancer Research*, 59(10), pp. 2302–2306.

Easwaran, H. P. *et al.* (2004) 'Replication-independent chromatin loading of Dnmt1 during G2 and M phases.', *EMBO reports*. EMBO Press, 5(12), pp. 1181–6. doi: 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*. National Academy of Sciences, 103(38), pp. 14080–5. doi: 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. 2709–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7079182 (Accessed: 18 December 2017).

el- Deiry, W. S. *et al.* (1991) 'High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer', *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), pp. 3470–3474.

ENCODE Project Consortium (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. doi: 10.1038/nature11247.

Ernst, J. *et al.* (2011) 'Mapping and analysis of chromatin state dynamics in nine human cell types', *Nature*. Nature Publishing Group, 473(7345), pp. 43–49. doi: 10.1038/nature09906.

Fang, M. *et al.* (2014) 'The BRAF Oncoprotein Functions through the Transcriptional Repressor MAFG to Mediate the CpG Island Methylator Phenotype', *Molecular Cell*, 55(6), pp. 904–915. doi: 10.1016/j.molcel.2014.08.010.

Fang, M. *et al.* (2016) 'Common BRAF(V600E)-directed pathway mediates widespread epigenetic silencing in colorectal cancer and melanoma', *Proceedings of the National Academy of Sciences*, 113(5), pp. 1250–1255. doi: 10.1073/pnas.1525619113.

Farthing, C. R. *et al.* (2008) 'Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes', *PLoS Genetics*. Edited by W. N. Frankel. Public Library of Science, 4(6), p. e1000116. doi: 10.1371/journal.pgen.1000116.

Fearon, E. R. (2000) 'BRCA1 and E-Cadherin Promoter Hypermethylation and Gene Inactivation in Cancer--Association or Mechanism?', *Journal of the National Cancer Institute*. Oxford University Press, 92(7), pp. 515–517. doi: 10.1093/jnci/92.7.515.

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

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), p. 550–565.e5. doi: 10.1016/j.molcel.2017.07.012.

Figueroa, M. E. *et al.* (2010) 'Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation', *Cancer Cell*, 18(6), pp. 553–567. doi: 10.1016/j.ccr.2010.11.015.

Fu, Y. *et al.* (2015) 'N6-Methyldeoxyadenosine Marks Active Transcription Start Sites in Chlamydomonas', *Cell*, 161(4), pp. 879–892. doi: 10.1016/j.cell.2015.04.010.

Fuks, F. *et al.* (2003) 'The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase.', *Nucleic acids research*. Oxford University Press, 31(9), pp. 2305–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12711675 (Accessed: 15 December 2017).

Gaidatzis, D. *et al.* (2014) 'DNA Sequence Explains Seemingly Disordered Methylation Levels in Partially Methylated Domains of Mammalian Genomes', *PLoS Genetics*. Edited by A. J. Sharp. Public Library of Science, 10(2), p. e1004143. doi: 10.1371/journal.pgen.1004143.

Gaidzik, V. I. *et al.* (2012) 'TET2 Mutations in Acute Myeloid Leukemia (AML): Results From a Comprehensive Genetic and Clinical Analysis of the AML Study Group', *Journal of Clinical Oncology*, 30(12), pp. 1350–1357. doi: 10.1200/JCO.2011.39.2886.

Gama-Sosa, M. A. *et al.* (1983) 'Tissue-specific differences in DNA methylation in various mammals.', *Biochimica et biophysica acta*, 740(2), pp. 212–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6860672 (Accessed: 18 December 2017).

Gama-Sosa, M. A. *et al.* (1983) 'Tissue-specific differences in DNA methylation in various mammals', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. Elsevier, 740(2), pp. 212–219. doi: 10.1016/0167-4781(83)90079-9.

Gardiner-Garden, M. and Frommer, M. (1987) 'CpG islands in vertebrate genomes.', *Journal of molecular biology*, 196(2), pp. 261–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3656447 (Accessed: 25 July 2018).

Gazzar, M. El *et al.* (2008) 'G9a and HP1 Couple Histone and DNA Methylation to TNFα Transcription Silencing during Endotoxin Tolerance', *Journal of Biological Chemistry*, 283(47), pp. 32198–32208. doi: 10.1074/jbc.M803446200.

Ge, Y.-Z. *et al.* (2004) 'Chromatin targeting of de novo DNA methyltransferases by the PWWP domain.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 279(24), pp. 25447–54. doi: 10.1074/jbc.M312296200.

Geisinger, J. M. *et al.* (2016) '*In vivo* blunt-end cloning through CRISPR/Cas9-facilitated non-homologous end-joining', *Nucleic Acids Research*. Oxford University Press, 44(8), pp. e76–e76. doi: 10.1093/nar/gkv1542.

Gopalakrishnan, S. *et al.* (2009) 'DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions.', *Human molecular genetics*, 18(17), pp. 3178–93. doi: 10.1093/hmg/ddp256.

Gowher, H. and Jeltsch, A. (2001) 'Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites', *Journal of Molecular Biology*, 309(5), pp. 1201–1208. doi: 10.1006/jmbi.2001.4710.

Greer, E. L. *et al.* (2015) 'DNA Methylation on N6-Adenine in C. elegans', *Cell*, 161(4), pp. 868–878. doi: 10.1016/j.cell.2015.04.005.

Greger, V. *et al.* (1989) 'Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma.', *Human genetics*, 83(2), pp. 155–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2550354 (Accessed: 30 November 2017).

Guo, G. *et al.* (2017) 'Epigenetic resetting of human pluripotency.', *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 144(15), pp. 2748–2763. doi: 10.1242/dev.146811.

Guo, J. U. *et al.* (2013) 'Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain', *Nature Neuroscience*. Nature Publishing Group, 17(2), pp. 215–222. doi: 10.1038/nn.3607.

Guo, X. *et al.* (2015) 'Structural insight into autoinhibition and histone H3-induced activation of DNMT3A', *Nature*, 517(7536), pp. 640–644. doi: 10.1038/nature13899.

Habibi, E. *et al.* (2013) 'Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells', *Cell Stem Cell*. Cell Press, 13(3), pp. 360–

369. doi: 10.1016/J.STEM.2013.06.002.

Hagarman, J. A. *et al.* (2013) 'Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells', *PLoS ONE*. Edited by J. G. Knott. Public Library of Science, 8(1), p. e53880. doi: 10.1371/journal.pone.0053880.

Hajkova, P. *et al.* (2002) 'Epigenetic reprogramming in mouse primordial germ cells', *Mechanisms of Development*, 117(1–2), pp. 15–23. Available at: https://ac-els-cdn-com.ezproxy.is.ed.ac.uk/S0925477302001818/1-s2.0-S0925477302001818-main.pdf?_tid=53352966-cbbf-11e7-871b-00000aacb361&acdnat=1510941005_182aa1b6965341101ba919f43b6332ca (Accessed: 17 November 2017).

'Hallmarks of cancer: Interactions with the tumor stroma' (2010) *Experimental Cell Research*. Academic Press, 316(8), pp. 1324–1331. doi: 10.1016/J.YEXCR.2010.02.045.

Hamidi, T., Singh, A. K. and Chen, T. (2015) 'Genetic alterations of DNA methylation machinery in human diseases', *Epigenomics*, 7(2), pp. 247–265. doi: 10.2217/epi.14.80.

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

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

Hansen, K. D. *et al.* (2014) 'Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization.', *Genome research*. Cold Spring Harbor Laboratory Press, 24(2), pp. 177–84. doi: 10.1101/gr.157743.113.

Hashimoto, H. *et al.* (2008) 'The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix', *Nature*, 455(7214), pp. 826–829. doi: 10.1038/nature07280.

Hata, K. *et al.* (2002) 'Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice.', *Development (Cambridge, England)*, 129(8), pp. 1983–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11934864 (Accessed: 9 November 2017).

Hawkins, R. D. *et al.* (2010) 'Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.', *Cell stem cell*. Elsevier, 6(5), pp. 479–91. doi: 10.1016/j.stem.2010.03.018.

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. doi: 10.1126/science.1210944.

Hebbar, P. B. and Archer, T. K. (2007) 'Chromatin-dependent cooperativity between site-specific transcription factors in vivo.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 282(11), pp. 8284–91. doi: 10.1074/jbc.M610554200.

Hebbar, P. B. and Archer, T. K. (2008) 'Altered histone H1 stoichiometry and an absence of

nucleosome positioning on transfected DNA.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 283(8), pp. 4595–601. doi: 10.1074/jbc.M709121200.

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. doi: 10.1038/nrm2727.

Herman, J. G. *et al.* (1998) 'Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma', *Genetics*, 95, pp. 6870–6875. Available at: http://www.pnas.org/content/95/12/6870.full.pdf (Accessed: 25 May 2017).

Hermann, A., Goyal, R. and Jeltsch, A. (2004) 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites.', *The Journal of biological chemistry*, 279(46), pp. 48350–9. doi: 10.1074/jbc.M403427200.

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

Hinoue, T. *et al.* (2012) 'Genome-scale analysis of aberrant DNA methylation in colorectal cancer.', *Genome research*. Cold Spring Harbor Laboratory Press, 22(2), pp. 271–82. doi: 10.1101/gr.117523.110.

Hinoue, T. *et al.* (2012) 'Genome-scale analysis of aberrant DNA methylation in colorectal cancer', *Genome Research*, 22(2), pp. 271–282. doi: 10.1101/gr.117523.110.

Hirasawa, R. *et al.* (2008) 'Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development.', *Genes & development*. Cold Spring Harbor Laboratory Press, 22(12), pp. 1607–16. doi: 10.1101/gad.1667008.

Hotchkiss, R. D. (1948) 'The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography.', *The Journal of biological chemistry*, 175(1), pp. 315–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18873306 (Accessed: 6 November 2017).

Howell, C. Y. *et al.* (2001) 'Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene.', *Cell*, 104(6), pp. 829–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11290321 (Accessed: 17 November 2017).

Hu, L. *et al.* (2011) 'Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2', *Cell Research*. Nature Publishing Group, 21(9), pp. 1374–1378. doi: 10.1038/cr.2011.124.

Huang, H. *et al.* (2013) 'TET1 plays an essential oncogenic role in MLL-rearranged leukemia', *Proceedings of the National Academy of Sciences*, 110(29), pp. 11994–11999. doi: 10.1073/pnas.1310656110.

Ibrahim, A. E. K. *et al.* (2011) 'Sequential DNA methylation changes are associated with DNMT3B overexpression in colorectal neoplastic progression', *Gut*, 60(4), pp. 499–508. doi:

10.1136/gut.2010.223602.

Illingworth, R. S. *et al.* (2010) 'Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome', *PLoS Genetics*. Edited by W. Reik. Public Library of Science, 6(9), p. e1001134. doi: 10.1371/journal.pgen.1001134.

Inoue, F. *et al.* (2017) 'A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity', *Genome Research*, 27(1), pp. 38–52. doi: 10.1101/gr.212092.116.

Irizarry, R. A. *et al.* (2009) 'The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores', *Genome Biology*. BioMed Central, 41(2), pp. 178–186. doi: 10.1038/ng.298.

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*. Elsevier, 68(2), p. 350–360.e7. doi: 10.1016/j.molcel.2017.09.037.

Issa, J. P. *et al.* (1993) 'Increased cytosine DNA-methyltransferase activity during colon cancer progression', *Journal of the National Cancer Institute*, 85(15), pp. 1235–1240.

Ito, S. *et al.* (2010) 'Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification', *Nature*. Nature Publishing Group, 466(7310), pp. 1129–1133. doi: 10.1038/nature09303.

Iyer, L. M. *et al.* (2009) 'Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids', *Cell Cycle*. Taylor & Francis, 8(11), pp. 1698–1710. doi: 10.4161/cc.8.11.8580.

Jackson, M. *et al.* (2004) 'Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells.', *Molecular and cellular biology*. American Society for Microbiology, 24(20), pp. 8862–71. doi: 10.1128/MCB.24.20.8862-8871.2004.

Jaenisch, R., Lyko, F. and Ramsahoye, B. H. (2000) 'Development: DNA methylation in Drosophila melanogaster', *Nature*. Nature Publishing Group, 408(6812), pp. 538–540. doi: 10.1038/35046205.

Jair, K.-W. *et al.* (2006) 'De novo CpG island methylation in human cancer cells.', *Cancer research*. American Association for Cancer Research, 66(2), pp. 682–92. doi: 10.1158/0008-5472.CAN-05-1980.

Janulaitis, A. *et al.* (1983) 'Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine.', *FEBS letters*, 161(1), pp. 131–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6884523 (Accessed: 6 November 2017).

Jeltsch, A. and Jurkowska, R. Z. (2016) 'Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm.', *Nucleic acids research*. Oxford University

Press, 44(18), pp. 8556–8575. doi: 10.1093/nar/gkw723.

Jeon, J. *et al.* (2015) 'Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, Magnaporthe oryzae', *Scientific Reports*. Nature Publishing Group, 5(1), p. 8567. doi: 10.1038/srep08567.

Jeziorska, D. M. *et al.* (2017) 'DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 114(36), pp. E7526–E7535. doi: 10.1073/pnas.1703087114.

Jia, D. *et al.* (2007) 'Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation', *Nature*, 449(7159), pp. 248–251. doi: 10.1038/nature06146.

Jin, S.-G. *et al.* (2011) '5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations', *Cancer Research*, 71(24), pp. 7360–7365. doi: 10.1158/0008-5472.CAN-11-2023.

Jinek, M. *et al.* (2013) 'RNA-programmed genome editing in human cells', *eLife*. eLife Sciences Publications Limited, 2, p. e00471. doi: 10.7554/eLife.00471.

Joensuu, E. I. *et al.* (2015) 'Methyltransferase expression and tumor suppressor gene methylation in sporadic and familial colorectal cancer', *Genes, Chromosomes and Cancer*, 54(12), pp. 776–787. doi: 10.1002/gcc.22289.

de Jong, J. *et al.* (2014) 'Chromatin Landscapes of Retroviral and Transposon Integration Profiles', *PLoS Genetics*. Edited by M. J. Roth. Public Library of Science, 10(4), p. e1004250. doi: 10.1371/journal.pgen.1004250.

Jurkowska, R. Z. *et al.* (2011) 'Oligomerization and Binding of the Dnmt3a DNA Methyltransferase to Parallel DNA Molecules', *Journal of Biological Chemistry*, 286(27), pp. 24200–24207. doi: 10.1074/jbc.M111.254987.

Karpf, A. R. *et al.* (2001) 'Activation of the p53 DNA damage response pathway after inhibition of DNA methyltransferase by 5-aza-2'-deoxycytidine.', *Molecular pharmacology*. American Society for Pharmacology and Experimental Therapeutics, 59(4), pp. 751–7. doi: 10.1124/MOL.59.4.751.

Kebede, A. F., Schneider, R. and Daujat, S. (2015) 'Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest', *FEBS Journal*, 282(9), pp. 1658–1674. doi: 10.1111/febs.13047.

Kim, G.-D. *et al.* (2002) 'Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases.', *The EMBO journal*. European Molecular Biology Organization, 21(15), pp. 4183–95. doi: 10.1093/EMBOJ/CDF401.

Ko, M. *et al.* (2010) 'Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2', *Nature*. Nature Publishing Group, 468(7325), pp. 839–843. doi: 10.1038/nature09586.

Kobayashi, Y. *et al.* (2011) 'DNA methylation profiling reveals novel biomarkers and important roles for DNA methyltransferases in prostate cancer.', *Genome research*. Cold Spring Harbor Laboratory Press, 21(7), pp. 1017–27. doi: 10.1101/gr.119487.110.

Koziol, M. J. *et al.* (2015) 'Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications', *Nature Structural & Molecular Biology*. Nature Research, 23(1), pp. 24–30. doi: 10.1038/nsmb.3145.

Krauss, V. and Reuter, G. (2011) 'DNA Methylation in Drosophila—A Critical Evaluation', *Progress in Molecular Biology and Translational Science*. Academic Press, 101, pp. 177–191. doi: 10.1016/B978-0-12-387685-0.00003-2.

Krebs, A. R., Dessus-Babus, S., Burger, L., Schübeler, D., *et al.* (2014) 'High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions', *eLife*. eLife Sciences Publications Limited, 3, p. e04094. doi: 10.7554/eLife.04094.

Krebs, A. R., Dessus-Babus, S., Burger, L. and Schübeler, D. (2014) 'High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions', *eLife*, 3. doi: 10.7554/eLife.04094.

Kriaucionis, S. and Heintz, N. (2009) 'The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.', *Science (New York, N.Y.)*. Howard Hughes Medical Institute, 324(5929), pp. 929–30. doi: 10.1126/science.1169786.

Krueger, F. and Andrews, S. R. (2011) 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*. Oxford University Press, 27(11), pp. 1571–1572. doi: 10.1093/bioinformatics/btr167.

Kundaje, A. *et al.* (2015) 'Integrative analysis of 111 reference human epigenomes', *Nature*. Nature Publishing Group, 518(7539), pp. 317–330. doi: 10.1038/nature14248.

Lao, V. V. and Grady, W. M. (2011) 'Epigenetics and colorectal cancer', *Nature Reviews Gastroenterology & Hepatology*. Nature Publishing Group, 8(12), pp. 686–700. doi: 10.1038/nrgastro.2011.173.

Laurent, L. *et al.* (2010) 'Dynamic changes in the human methylome during differentiation', *Genome Research*, 20(3), pp. 320–331. doi: 10.1101/gr.101907.109.

Lee, J.-H. and Skalnik, D. G. (2005) 'CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 280(50), pp. 41725–31. doi: 10.1074/jbc.M508312200.

Lee, J. *et al.* (2002) 'Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells.', *Development (Cambridge, England)*, 129(8), pp. 1807–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11934847 (Accessed: 17 November 2017).

Lee, J. *et al.* (2013) 'Transforming Growth Factor Beta Receptor 2 (TGFBR2) Changes Sialylation in the Microsatellite Unstable (MSI) Colorectal Cancer Cell Line HCT116', *PLoS ONE*. Edited by

C.-H. Tang. Public Library of Science, 8(2), p. e57074. doi: 10.1371/journal.pone.0057074.

Lee, T. I. *et al.* (2006) 'Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells', *Cell*. Cell Press, 125(2), pp. 301–313. doi: 10.1016/J.CELL.2006.02.043.

Lengauer, C., Kinzler, K. W. and Vogelstein, B. (1997) 'DNA methylation and genetic instability in colorectal cancer cells', *Medical Sciences*, 94, pp. 2545–2550. Available at: http://www.pnas.org/content/94/6/2545.full.pdf (Accessed: 15 June 2017).

Lentini, A. *et al.* (2018) 'A reassessment of DNA-immunoprecipitation-based genomic profiling', *Nature Methods*. Nature Publishing Group, 15(7), pp. 499–504. doi: 10.1038/s41592-018-0038-7.

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

Li, B. *et al.* (2009) 'Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 284(12), pp. 7970–6. doi: 10.1074/jbc.M808220200.

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

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

Li, M. A. *et al.* (2011) 'Mobilization of giant piggyBac transposons in the mouse genome', *Nucleic Acids Research*. Oxford University Press, 39(22), pp. e148–e148. doi: 10.1093/nar/gkr764.

Li, X. *et al.* (2005) 'piggyBac internal sequences are necessary for efficient transformation of target genomes', *Insect Molecular Biology*, 14(1), pp. 17–30. doi: 10.1111/j.1365-2583.2004.00525.x.

Liang, G. *et al.* (2002) 'Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements.', *Molecular and cellular biology*. American Society for Microbiology, 22(2), pp. 480–91. doi: 10.1128/MCB.22.2.480-491.2002.

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

Lienert, F. *et al.* (2011) 'Identification of genetic elements that autonomously determine DNA methylation states.', *Nature genetics*, 43(11), pp. 1091–7. doi: 10.1038/ng.946.

Lin, S. *et al.* (2014) 'Enhanced homology-directed human genome engineering by controlled

timing of CRISPR/Cas9 delivery', *eLife*. eLife Sciences Publications Limited, 3, p. e04766. doi: 10.7554/eLife.04766.

Linhart, H. G. *et al.* (2007) 'Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing.', *Genes & development*. Cold Spring Harbor Laboratory Press, 21(23), pp. 3110–22. doi: 10.1101/gad.1594007.

Lister, R. *et al.* (2008) 'Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis', *Cell*. Elsevier, 133(3), pp. 523–536. doi: 10.1016/j.cell.2008.03.029.

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

Lister, R. *et al.* (2013) 'Global epigenomic reconfiguration during mammalian brain development.', *Science (New York, N.Y.)*, 341(6146), p. 1237905. doi: 10.1126/science.1237905.

Liu, Y. *et al.* (1998) 'Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci.', *Nucleic acids research*. Oxford University Press, 26(4), pp. 1038–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9461465 (Accessed: 4 November 2017).

Liu, Z. *et al.* (2008) 'CpG island methylator phenotype involving tumor suppressor genes located on chromosome 3p in non-small cell lung cancer', *Lung Cancer*, 62(1), pp. 15–22. doi: 10.1016/j.lungcan.2008.02.005.

Lorsbach, R. B. *et al.* (2003) 'TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23)', *Leukemia*. Nature Publishing Group, 17(3), pp. 637–641. doi: 10.1038/sj.leu.2402834.

Lynch, M. D. *et al.* (2011) 'An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment', *The EMBO Journal*, 31, pp. 317–329. doi: 10.1038/emboj.2011.399.

Maatouk, D. M. *et al.* (2006) 'DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages', *Development*, 133(17). Available at: http://dev.biologists.org/content/133/17/3411 (Accessed: 23 April 2017).

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

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

Mao, Z. *et al.* (2008) 'Comparison of nonhomologous end joining and homologous recombination in human cells.', *DNA repair*. NIH Public Access, 7(10), pp. 1765–71. doi:

10.1016/j.dnarep.2008.06.018.

Maresca, M. *et al.* (2013) 'Obligate Ligation-Gated Recombination (ObLiGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining', *Genome Research*, 23(3), pp. 539–546. doi: 10.1101/gr.145441.112.

Maruyama, R. *et al.* (2001) 'Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features.', *Cancer research*, 61(24), pp. 8659–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11751381 (Accessed: 7 July 2018).

Mayer, W. *et al.* (2000) 'Demethylation of the zygotic paternal genome.', *Nature*, 403(6769), pp. 501–2. doi: 10.1038/35000654.

Meir, Y.-J. J. *et al.* (2011) 'Genome-wide target profiling of piggyBac and Tol2 in HEK 293: pros and cons for gene discovery and gene therapy.', *BMC biotechnology*, 11(1), p. 28. doi: 10.1186/1472-6750-11-28.

Meissner, A. *et al.* (2008) 'Genome-scale DNA methylation maps of pluripotent and differentiated cells', *Nature*. Nature Publishing Group, 454(7205), p. 766. doi: 10.1038/nature07107.

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. doi: 10.1016/j.molcel.2008.05.007.

Montague, T. G. *et al.* (2014) 'CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing', *Nucleic Acids Research*, 42(W1), pp. W401–W407. doi: 10.1093/nar/gku410.

Morgan, H. D. *et al.* (2005) 'Epigenetic reprogramming in mammals', *Human Molecular Genetics*. Oxford University Press, 14(suppl_1), pp. R47–R58. doi: 10.1093/hmg/ddi114.

Müller, T. *et al.* (2012) 'Nuclear exclusion of TET1 is associated with loss of 5-hydroxymethylcytosine in IDH1 wild-type gliomas', *The American Journal of Pathology*, 181(2), pp. 675–683. doi: 10.1016/j.ajpath.2012.04.017.

Murphy, P. J. *et al.* (2013) 'Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells', *Proceedings of the National Academy of Sciences*, 110(19), pp. 7772–7777. doi: 10.1073/pnas.1218495110.

Muto, M. *et al.* (2002) 'Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 277(37), pp. 34549–55. doi: 10.1074/jbc.M205189200.

Muzny, D. M. *et al.* (2012) 'Comprehensive molecular characterization of human colon and rectal cancer', *Nature*, 487(7407), pp. 330–337. doi: 10.1038/nature11252.

Myöhänen, S. K., Baylin, S. B. and Herman, J. G. (1998a) 'Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia.', *Cancer research*. American Association for

Cancer Research, 58(4), pp. 591–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9485004 (Accessed: 31 May 2017).

Myöhänen, S. K., Baylin, S. B. and Herman, J. G. (1998b) 'Hypermethylation Can Selectively Silence Individual pltfnk4A AlÃ-elesin Neoplasia1', *CANCER RESEARCH*, 58, pp. 591–593. Available at: http://cancerres.aacrjournals.org/content/canres/58/4/591.full.pdf (Accessed: 25 May 2017).

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. doi: 10.1016/j.ejmech.2016.02.043.

Nady, N. *et al.* (2011) 'Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 286(27), pp. 24300–11. doi: 10.1074/jbc.M111.234104.

Ogino, S. *et al.* (2007) 'Evaluation of Markers for CpG Island Methylator Phenotype (CIMP) in Colorectal Cancer by a Large Population-Based Sample', *The Journal of Molecular Diagnostics*, 9(3), pp. 305–314. doi: 10.2353/jmoldx.2007.060170.

Ohm, J. E., McGarvey, K. M., *et al.* (2007) 'A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing', *Nature Genetics*, 39(2), pp. 237–242. doi: 10.1038/ng1972.

Ohm, J. E., Mcgarvey, K. M., *et al.* (2007) 'A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing'. doi: 10.1038/ng1972.

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. doi: 10.1016/S0092-8674(00)81656-6.

Oswald, J. *et al.* (2000) 'Active demethylation of the paternal genome in the mouse zygote.', *Current biology : CB*, 10(8), pp. 475–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10801417 (Accessed: 17 November 2017).

Otani, J. *et al.* (2009) 'Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain.', *EMBO reports*. EMBO Press, 10(11), pp. 1235–41. doi: 10.1038/embor.2009.218.

Pabst, T. *et al.* (2009) 'Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis', *British Journal of Cancer*, 100(8), pp. 1343–1346. doi: 10.1038/sj.bjc.6604977.

Pajvani, U. B. *et al.* (2005) 'Fat apoptosis through targeted activation of caspase 8: a new mouse model of inducible and reversible lipoatrophy', *Nature Medicine*. Nature Publishing Group, 11(7), pp. 797–803. doi: 10.1038/nm1262.

Pastor, W. A. *et al.* (2011) 'Genome-wide mapping of 5-hydroxymethylcytosine in embryonic

stem cells', *Nature*. Nature Publishing Group, 473(7347), pp. 394–397. doi: 10.1038/nature10102.

Pastor, W. A., Aravind, L. and Rao, A. (2013) 'TETonic shift: biological roles of TET proteins in DNA demethylation and transcription.', *Nature reviews. Molecular cell biology*, 14(6), pp. 341–56. doi: 10.1038/nrm3589.

Patel, K. *et al.* (2010) 'Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme.', *Nucleic acids research*. Oxford University Press, 38(13), pp. 4313–24. doi: 10.1093/nar/gkq187.

Peterson, C. L. and Laniel, M.-A. (2004) 'Histones and histone modifications.', *Current biology : CB*. Elsevier, 14(14), pp. R546-51. doi: 10.1016/j.cub.2004.07.007.

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

Qin, W., Leonhardt, H. and Spada, F. (2011) 'Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1', *Journal of Cellular Biochemistry*, 112(2), pp. 439–444. doi: 10.1002/jcb.22998.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*. Oxford University Press, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Quivoron, C. *et al.* (2011) 'TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis', *Cancer Cell*, 20(1), pp. 25–38. doi: 10.1016/j.ccr.2011.06.003.

R Core Team (2017) *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Available at: https://www.r-project.org/ (Accessed: 13 January 2018).

Rajakumara, E. *et al.* (2011) 'PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression.', *Molecular cell*. NIH Public Access, 43(2), pp. 275–284. doi: 10.1016/j.molcel.2011.07.006.

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. 114–28. doi: 10.1016/j.cell.2009.04.020.

Ramsahoye, B. H. *et al.* (2000) 'Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), pp. 5237–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10805783 (Accessed: 19 August 2017).

Ran, F. A. *et al.* (2013a) 'Genome engineering using the CRISPR-Cas9 system.', *Nature protocols*. NIH Public Access, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Ran, F. A. *et al.* (2013b) 'Genome engineering using the CRISPR-Cas9 system', *Nature Protocols*. Nature Publishing Group, 8(11), pp. 2281–2308. doi: 10.1038/nprot.2013.143.

Räsänen, K. and Vaheri, A. (2010) 'Activation of fibroblasts in cancer stroma', *Experimental Cell Research*, 316(17), pp. 2713–2722. doi: 10.1016/j.yexcr.2010.04.032.

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

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

Robertson, K. D. *et al.* (2000) 'Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G0/G1 to S phase transition in normal and tumor cells', *Nucleic Acids Research*, 28(10), pp. 2108–2113.

Robertson, K. D. (2001) 'DNA methylation, methyltransferases and cancer', *Oncogene*. Nature Publishing Group, 20(24), pp. 3139–3155. doi: 10.1038/sj.onc.1204341.

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. doi: 10.1016/j.jsb.2016.03.013.

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

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

Russler-Germain, D. A. *et al.* (2014) 'The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers', *Cancer Cell*, 25(4), pp. 442–454. doi: 10.1016/j.ccr.2014.02.010.

Sánchez-Vega, F. *et al.* (2015) 'Pan-cancer stratification of solid human epithelial tumors and cancer cell lines reveals commonalities and tissue-specific features of the CpG island methylator phenotype', *Epigenetics & Chromatin*. BioMed Central, 8(1), p. 14. doi: 10.1186/s13072-015-0007-7.

Saunderson, E. A. *et al.* (2017) 'Hit-and-run epigenetic editing prevents senescence entry in primary breast cells from healthy donors', *Nature Communications*. Nature Publishing Group, 8(1), p. 1450. doi: 10.1038/s41467-017-01078-2.

Schiffers, S. *et al.* (2017) 'Quantitative LC-MS Provides No Evidence for m ^{6} dA or m ^{4} dC in the Genome of Mouse Embryonic Stem Cells and Tissues', *Angewandte Chemie International Edition*, 56(37), pp. 11268–11271. doi: 10.1002/anie.201700424.

Schlesinger, Y. *et al.* (2007) 'Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer', *Nature Genetics*, 39(2), pp. 232–236. doi: 10.1038/ng1950.

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. 4860–76. doi: 10.1093/nar/gkt191.

Schroth, G. P. *et al.* (1990) 'A NMR study of mobility in the histone octamer', *FEBS Letters*, 268(1), pp. 117–120. doi: 10.1016/0014-5793(90)80987-T.

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

Schuebel, K. E. *et al.* (2000) 'CpG methylation is maintained in human cancer cells lacking DNMT1', *Nature*. Nature Publishing Group, 404(6781), pp. 1003–1007. doi: 10.1038/35010000.

Schultz, M. D. *et al.* (2015) 'Human body epigenome maps reveal noncanonical DNA methylation variation', *Nature*. Nature Publishing Group, 523(7559), pp. 212–216. doi: 10.1038/nature14465.

Selker, E. U. (1993) 'Control of DNA methylation in fungi', in *DNA Methylation*. Basel: Birkhäuser Basel, pp. 212–217. doi: 10.1007/978-3-0348-9118-9_10.

Selker, E. U. (1997) 'Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion?', *Trends in genetics : TIG*, 13(8), pp. 296–301. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9260514 (Accessed: 4 September 2017).

Serra, R. W. *et al.* (2014) 'A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype', *eLife*, 2014(3), p. e02313. doi: 10.7554/eLife.02313.

Shirohzu, H. *et al.* (2002) 'Three novelDNMT3B mutations in Japanese patients with ICF syndrome', *American Journal of Medical Genetics*, 112(1), pp. 31–37. doi: 10.1002/ajmg.10658.

Smallwood, A. *et al.* (2007) 'Functional cooperation between HP1 and DNMT1 mediates gene silencing.', *Genes & development*. Cold Spring Harbor Laboratory Press, 21(10), pp. 1169–78. doi: 10.1101/gad.1536807.

Smallwood, S. A. *et al.* (2011) 'Dynamic CpG island methylation landscape in oocytes and preimplantation embryos', *Nature Publishing Group*, 43(8). doi: 10.1038/ng.864.

Smets, M. *et al.* (2017) 'DNMT1 mutations found in HSANIE patients affect interaction with UHRF1 and neuronal differentiation.', *Human molecular genetics*. Oxford University Press, 26(8), pp. 1522–1534. doi: 10.1093/hmg/ddx057.

Smiraglia, D. J. *et al.* (2001) 'Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies.', *Human molecular genetics*, 10(13), pp. 1413–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11440994 (Accessed: 31 May 2017).

Smith, Z. D. *et al.* (2009) 'High-throughput bisulfite sequencing in mammalian genomes', *Methods*, 48(3), pp. 226–232. doi: 10.1016/j.ymeth.2009.05.003.

Smithies, O. *et al.* (1985) 'Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination.', *Nature*, 317(6034), pp. 230–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2995814 (Accessed: 30 December 2017).

Song, J. *et al.* (2011) 'Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation', *Science*, 331(6020), pp. 1036–1040. Available at: http://science.sciencemag.org.ezproxy.is.ed.ac.uk/content/sci/331/6020/1036.full.pdf (Accessed: 9 July 2017).

Song, J. Z. *et al.* (2002) 'Hypermethylation trigger of the glutathione-S-transferase gene (GSTP1) in prostate cancer cells', *Oncogene*. Nature Publishing Group, 21(7), pp. 1048–1061. doi: 10.1038/sj.onc.1205153.

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*. Rockefeller University Press, 176(5), pp. 565–71. doi: 10.1083/jcb.200610062.

Spencer, D. H. *et al.* (2017) 'CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression.', *Cell*, 168(5), p. 801–816.e13. doi: 10.1016/j.cell.2017.01.021.

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. doi: 10.1073/pnas.1013224108.

Sproul, D. *et al.* (2012) 'Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns', *Genome Biology*, 13(R84). Available at: http://download.springer.com/static/pdf/367/art%253A10.1186%252Fgb-2012-13-10-r84.pdf?originUrl=http%3A%2F%2Fgenomebiology.biomedcentral.com%2Farticle%2F10.1186%2Fgb-2012-13-10-r84&token2=exp=1495713390~acl=%2Fstatic%2Fpdf%2F367%2Fart%25253A10.1186%25252F (Accessed: 25 May 2017).

Sproul, D. and Meehan, R. R. (2013) 'Genomic insights into cancer-associated aberrant CpG island hypermethylation', *Briefings in Functional Genomics*, 12(3), pp. 174–190. doi: 10.1093/bfgp/els063.

Stadler, M. B. *et al.* (2011) 'DNA-binding factors shape the mouse methylome at distal regulatory regions.', *Nature*, 480(7378), pp. 490–5. doi: 10.1038/nature10716.

Stelzer, Y. *et al.* (2015) 'Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution', *Cell*. Elsevier, 163(1), pp. 218–29. doi: doi:10.1016/j.cell.2015.08.046.

Strahl, B. D. *et al.* (2002) 'Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression.', *Molecular and cellular biology*. American Society for Microbiology, 22(5), pp. 1298–306. doi: 10.1128/MCB.22.5.1298-1306.2002.

Tahiliani, M. *et al.* (2009) 'Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1', *Science*, 324(5929). Available at: http://science.sciencemag.org/content/324/5929/930/tab-pdf (Accessed: 7 July 2017).

Takahashi, Y. *et al.* (2017) 'Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells', *Science*, 356(6337), pp. 503–508. doi: 10.1126/science.aag3260.

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*. National Academy of Sciences, 108(22), pp. 9055–9. doi: 10.1073/pnas.1019629108.

Tanemura, A. *et al.* (2009) 'CpG Island Methylator Phenotype Predicts Progression of Malignant Melanoma', *Clinical Cancer Research*, 15(5), pp. 1801–1807. doi: 10.1158/1078-0432.CCR-08-1361.

Thienpont, B. *et al.* (2016) 'Tumour hypoxia causes DNA hypermethylation by reducing TET activity', *Nature*. Nature Publishing Group, 537(7618), pp. 63–68. doi: 10.1038/nature19081.

Thomas, K. R. and Capecchi, M. R. (1987) 'Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells.', *Cell*, 51(3), pp. 503–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2822260 (Accessed: 30 December 2017).

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986) 'High frequency targeting of genes to specific sites in the mammalian genome.', *Cell*, 44(3), pp. 419–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3002636 (Accessed: 30 December 2017).

Thomson, J. P. *et al.* (2010) 'CpG islands influence chromatin structure via the CpG-binding protein Cfp1', *Nature*, 464(7291), pp. 1082–1086. doi: 10.1038/nature08924.

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*. National Academy of Sciences, 96(15), pp. 8681–6. doi: 10.1073/PNAS.96.15.8681.

Tsai, H.-C. *et al.* (2012) 'Transient Low Doses of DNA-Demethylating Agents Exert Durable Antitumor Effects on Hematological and Epithelial Tumor Cells', *Cancer Cell*, 21(3), pp. 430–446. doi: 10.1016/j.ccr.2011.12.029.

Turner, H. E. *et al.* (2003) 'Angiogenesis in Endocrine Tumors', *Endocrine Reviews*, 24(5), pp. 600–632. doi: 10.1210/er.2002-0008.

Ueda, Y. *et al.* (2006) 'Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome', *Development*, 133(6), pp. 1183–1192. doi: 10.1242/dev.02293.

Untergasser, A. *et al.* (2012) 'Primer3--new capabilities and interfaces.', *Nucleic acids research*. Oxford University Press, 40(15), p. e115. doi: 10.1093/nar/gks596.

Uren, A. G. *et al.* (2005) 'Retroviral insertional mutagenesis: past, present and future',

Oncogene. Nature Publishing Group, 24(52), pp. 7656–7672. doi: 10.1038/sj.onc.1209043.

Vasquez, K. M. *et al.* (2001) 'Manipulating the mammalian genome by homologous recombination.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 98(15), pp. 8403–10. doi: 10.1073/pnas.111009698.

Vogelstein, B., Lane, D. and Levine, A. J. (2000) 'Surfing the p53 network', *Nature*, 408(6810), pp. 307–310. doi: 10.1038/35042675.

Voo, K. S. *et al.* (2000) 'Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1.', *Molecular and cellular biology*. American Society for Microbiology, 20(6), pp. 2108–21. doi: 10.1128/MCB.20.6.2108-2121.2000.

Weemaes, C. M. *et al.* (2013) 'Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects', *European Journal of Human Genetics*, 21(10), pp. 1219–1225. doi: 10.1038/ejhg.2013.40.

Weinberg, R. A. (Robert A. (2007) *The biology of cancer*. Garland Science. Available at: https://books.google.co.uk/books/about/The_Biology_of_Cancer.html?id=i0Dfq8pzll8C&source=kp_cover&redir_esc=y (Accessed: 4 December 2017).

Weisenberger, D. J. *et al.* (2004) 'Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation.', *Molecular cancer research : MCR*. American Association for Cancer Research, 2(1), pp. 62–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14757847 (Accessed: 8 November 2017).

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. doi: 10.1038/ng1834.

Widschwendter, M. *et al.* (2007) 'Epigenetic stem cell signature in cancer', *Nature Genetics*, 39(2), pp. 157–158. doi: 10.1038/ng1941.

Wilson, G. G. and Murray, N. E. (1991) 'Restriction and Modification Systems', *Annual Review of Genetics*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 25(1), pp. 585–627. doi: 10.1146/annurev.ge.25.120191.003101.

Witte, T., Plass, C. and Gerhauser, C. (2014) 'Pan-cancer patterns of DNA methylation.', *Genome medicine*. BioMed Central, 6(8), p. 66. doi: 10.1186/s13073-014-0066-6.

Wu, T. P. *et al.* (2016) 'DNA methylation on N6-adenine in mammalian embryonic stem cells', *Nature*. Nature Research, 532(7599), pp. 329–333. doi: 10.1038/nature17640.

Xie, W. *et al.* (2013) 'Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells', *Cell*, 153(5), pp. 1134–1148. doi: 10.1016/j.cell.2013.04.022.

Yamazaki, Y. *et al.* (2003) 'Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning.', *Proceedings*

of the National Academy of Sciences of the United States of America, 100(21), pp. 12207–12. doi: 10.1073/pnas.2035119100.

Yan, X.-J. *et al.* (2011) 'Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia', *Nature Genetics*, 43(4), pp. 309–315. doi: 10.1038/ng.788.

Yang, H. *et al.* (2013) 'Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation', *Oncogene*, 32(5), pp. 663–669. doi: 10.1038/onc.2012.67.

Yang, X. *et al.* (2014) 'Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer', *Cancer Cell*. Elsevier, 26(4), pp. 577–590. doi: 10.1016/j.ccr.2014.07.028.

Yoder, J. A. *et al.* (1997) 'DNA (cytosine-5)-methyltransferases in Mouse Cells and Tissues. Studies with a Mechanism-based Probe'. Available at: https://ac-els-cdn-com.ezproxy.is.ed.ac.uk/S0022283697911256/1-s2.0-S0022283697911256-main.pdf?_tid=e494b59c-b4c8-11e7-bdf5-00000aab0f6c&acdnat=1508416238_2cc6b9e1a2ae714ff23c6632d9a75c96 (Accessed: 19 October 2017).

You, J. S. and Jones, P. A. (2012) 'Cancer genetics and epigenetics: two sides of the same coin?', *Cancer Cell*, 22(1), pp. 9–20. doi: 10.1016/j.ccr.2012.06.008.

Youdell, M. L. *et al.* (2008) 'Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36.', *Molecular and cellular biology*. American Society for Microbiology, 28(16), pp. 4915–26. doi: 10.1128/MCB.00001-08.

Zhang, G. *et al.* (2015) 'N6-Methyladenine DNA Modification in Drosophila', *Cell*, 161(4), pp. 893–906. doi: 10.1016/j.cell.2015.04.018.

Zhang, J.-C. *et al.* (2014) 'Promoter hypermethylation of p14 ARF , RB, and INK4 gene family in hepatocellular carcinoma with hepatitis B virus infection', *Tumor Biology*, 35(3), pp. 2795–2802. doi: 10.1007/s13277-013-1372-0.

Zhang, L. *et al.* (2012) 'Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA', *Nature Chemical Biology*. doi: 10.1038/nchembio.914.

Zhao, H. and Chen, T. (2013) 'Tet family of 5-methylcytosine dioxygenases in mammalian development', *Journal of Human Genetics*, 58(7), pp. 421–427. doi: 10.1038/jhg.2013.63.

Zhu, F. *et al.* (2014) 'DICE, an efficient system for iterative genomic editing in human pluripotent stem cells', *Nucleic Acids Research*. Oxford University Press, 42(5), pp. e34–e34. doi: 10.1093/nar/gkt1290.

Zlatanova, J., Leuba, S. H. and van Holde, K. (1999) 'Chromatin structure revisited.', *Critical reviews in eukaryotic gene expression*, 9(3–4), pp. 245–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10651241 (Accessed: 15 December 2017).