Novel insights into the effects of manipulating DNA methyltransferase levels on the imprinted and non-imprinted regions of the genome

Sarah-Jayne Mackin

BSc (Hons) Genetics

PgCert Stem Cell Biology

MSc Biomedical Science

Research conducted within

Genomic Medicine Research Group

School of Biomedical Sciences

Faculty of Life and Health Sciences

Ulster University

A thesis submitted for the degree of

Doctor of Philosophy

April 2019

I confirm that the word count of this thesis is less than 100,000

Note on access to contents

I hereby declare that with effect from the date on which the thesis is deposited in the Research Office of the University of Ulster, I permit;

1. the Librarian of the University to allow the thesis to be copied in whole or in part without reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library.

2. the thesis to be made available through the Ulster Institutional Repository and/or EThOS under the terms of the Ulster eTheses Deposit Agreement which I have signed.

Contents

Acknowledgements

Without hesitation, I would firstly like to offer a special thanks to my supervisor Professor Colum Walsh, who took a chance on me before a PhD was even on the horizon. Without the opportunity to complete my MSc project in his research group, I never stood a chance of ever successfully applying for a PhD and for this, I am eternally grateful. I am inspired by your considered approach to your work and those around you- especially your annual brick wall words of encouragement and the patience you have given me throughout this write up.

My PhD journey started on the right path with a special thanks to Dr Karla M. O'Neill and her nurturing and empathetic approach to me day-to-day. On Karla's departure Dr. Declan McKenna became an incredible inspiration to me with his balanced approach to work and showing me how to use the PhD process to become a more rounded researcher. Your advice has never let me down and I have been very lucky to have you as a second supervisor throughout this process.

A special thanks for the encouragement an unending support from my housemates Dr. Rachelle Irwin, Dr. Michelle Rudden and Dr. Katie Christie! I loved sharing this journey with you all, and hope it continues.

To the past members of the Walsh group- Dr Avinash Thakur for all of our laughing over setting up experiments during my MSc to having the pleasure of finishing some data for your paper, you were such a happy and kind person to share the lab with and a great mentor. Dr Hazel Mangan- for painstakingly trying to explain anything COBRA related many many times before it finally sunk in. Dr Seodhna Lynch- for our many years of friendship and your kindness. Thanks too to the current members of the Walsh lab- Sara-Jayne Thursby, Gareth Pollin and Catherine Scullion for the giggles in lab meeting, chocolates, baking and laughs. Hope you enjoy the remainder of your PhDs as much as I have. Laura Mairs you have been my partner in crime from nearly the beginning- here is to more trips away together when we are finished.

To my best friends Trudy, Aileen, Lizzie and Rebecca whom I have seen such little of since this process started, thank you for your patience and understanding of how much completing my PhD meant to me. Dr Mary Bonner- thank you for your wisdom, words of encouragement, and efforts to help me get to where I am today.

To my partner Niall, who maybe unwittingly never knew what moving in with a panicking academic, would do to his sanity! Meeting you during my PhD has been a blessing, and you have kept me grounded, fed, watered and loved. I cannot wait to see where this next step takes our future together (and thanks for always asking about the CpGs).

To my Mum and Dad, and my guardian angel Aaron whose belief in me never ever feigned. Without you, I would never have come to have the strength, motivation and support to come this far. I am sorry for leaving you both all those years ago to move to Glasgow but thank you encouraging me to become the best version of myself I can be, whichever path that takes me down. Thank you teaching me to be proud of me and trying to remove as much financial and emotional stress from me as you can. Thank for you for always visiting me, your care packages and the countless hours chatting. It has been paramount in my life so far and will be no doubt be to come.

Abstract

DNA methylation or 5-methylcytosine (5mC) is an important epigenetic mark integral to appropriate genomic imprinting, X chromosome inactivation and in the silencing of developmentally important genes. There remains to be a lot unknown about how DNA methylation is programmed and reprogrammed once lost particulary in early mammalian development. In untransformed hTERTs and in mouse embryonic stem cells (mESCs) we were able to investigate how methylation can or cannot be recovered following the removal and reintroduction of key methyltransferases, by using a range of wet-lab and bioinformatic approaches to assay methylation at individual CpG sites. Indeed, this thesis has shown that methylation, once lost is able to be recovered in nearly all of the known and putative imprints assayed in the mESCs through rescue with DNMT3A2. Not only does the *de novo* methyltransferase-3a assist with the recovery of methylation but it also plays a role in maintenance methylation.

A reduction of methylation through a stable loss of DNMT1 in the hTERTs however highlighted enriched groups of hypomethylated genes such as the Cancer Testis Antigen (CTA) genes and the Protocadherins (PCDH) that are sentistive to this loss, and in some instances associated with polycomb marks. Interestingly a transient loss of methylation however in the same system with pharmacological 5-aza-2dC (Aza) showed not only widespread methylation, but also some gains of methylation at selected loci such as the the GPCRs known to act as oncogenes and tumour supressors. In conclusion, through examining the effects of the loss of DNA methylation in both human and mouse model systems, this thesis has successfully identified gene classes sensitive to a transient and stable loss of DNA methylation and identified where this methylation can be recovered.

Abbreviations

AA	Amino Acid
AML	Acute Myeloid Leukaemia
ART	Assisted Reproductive Technologies
Blimp1	B lymphocyte induced maturation protein 1
BMIQ	Beta Mixture Quantile Dilation
BMP	Bone morphogenetic protein
BS	Bisulfite treated
BWS	Beckwith Weidemann Syndrome
CGI	CpG Islands
CHARM	Comprehensive High Throughput Arrays for Relative Methylation
ChiP	Chromatin Immunoprecipitation
CTA	Cancer Testis Antigen Genes
CTCF	CCCTC Binding Factor
DAVID	Database for Annotation, Visualisation and Integrated Discovery
ddNTPs	Dideoxynucleotides
DDR	DNA Damage Response
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyltransferase
DMR	Differentially Methylated Region
dsRNA	Double Stranded RNA
ERV	Endogenous Retrovirus
ESCs	Embryonic Stem Cells
EZH2	Enhancer of Zeste 2
GEO	Gene Expression Omnibus
gDMR	gametic Differentially Methylated Region
HMM	Hidden Markov Model

IAP	Intracisternal Alpha Particles
ICR	Imprint Control Region
IDAT	Intensity Data Files
KD	Knock Down
KO	Knock Out
LINE	Long Interspersed Nuclear Element
L1	Line-1
LncRNA	Long non-coding RNA
LOF	Loss of Function
LOI	Loss of Imprinting
LTR	Long Terminal Repeats
MBDCap-Seq	Methyl CpG Binding Domain based Capture and Sequencing
MDS	Myelodysplastic Syndromes
MeDIP	Methylated DNA Immunoprecipitation
MHD	MAGE Homology Domain
MRE-seq	Methylation Sensitive Restriction Enzyme Sequencing
ncRNA	Non- Coding Ribonucleic Acid
NIAD	National Institute of Allergy and Infectious Disease
NIH	National Institute of Health
NuRD	Nucleosome Remodelling and Deacetylation Complex
PANTHER	Protein Analysis through Evolutionary Relationships
PCNA	Proliferating Cell Nuclear Antigen
PGC	Primordial Germ Cell
PCR	Polymerase Chain Reaction
PHD	Plant homeodomain
PRC	Polycomb Repressive Complex
OOP	Object Orientated Programming
QC	Quality Control

RBP	RNA binding protein
RING	Really Interesting New Gene
RNA	Ribonucleic Acid
RRBS	Reduced Representation Bisulfite Sequencing
RTEs	Retrotransposable Elements
SAM	5-adenosyl methionine
scRRBS	Single Cell Reduced Representation Bisulfite Sequencing
scWGBS	Single Cell Whole Genome Bisulfite Sequencing
SBS	Sequencing by Synthesis
SINE	Short Interspersed Nuclear Element
SRA	Set and Ring Associated Domain
TAA	Tumour Associated Antigens
TAB-seq	TET-assisted Bisulfite Sequencing
TKO	Triple Knock Out
TSG	Tumour Supressor Genes
TSS	Transcription Start Site
TTD	Tandem Tudor Domain
WGBS	Whole Genome Bisulfite Sequencing
WHO	World Health Organisation
XCI	X-chromosome Inactivation
Xi	Inactivated X-chromosome
XIC	X-chromosome inactivation centre
Xist	X-inactive-specific transcript
XML	eXtensible Markup Language
5aC	5-carboxylcytosine
5fC	5-formlycytosine
5mC	5-methylcytosine
5hmC	5-hydroxymethylcytosine

General Introduction

DNA methylation

DNA methylation involves the transfer of a methyl group from 5-adenosylmethionine to the 5'carbon of a cytosine residue within CpG dinucleotides, converting it to 5- methylcytosine (5mC) (Gibney and Nolan, 2010). DNA methylation is essential for differentiation and normal development, due to its effects on long-term stable gene repression (Cedar and Bergman, 2009, Lindroth et al., 2001). 5mC is the most prevalent covalent epigenetic mechanism of genomic DNA in eukaryotic cells and it is responsible for tightly regulating gene expression (Zingg and Jones, 1997) especially in CG dense regions i.e. at CpG dinucleotides. Stirzaker et al (2014) suggests that there are close to 28 million CpG dinucleotides in the human genome. These CpGs are not distributed evenly throughout the mammalian genome; they cluster in fact as CpG Islands (CGI). Housekeeping genes usually have CGI present at their promoter regions (Zhu et al., 2008) that remain unmethylated and are often associated with active gene expression, particularly in development (Messerschmidt et al., 2014). Conversely, methylated CGI located at promoters of such genes are associated with gene repression (Jeziorska et al., 2017). DNA methylation is present at not only CGI but across less dense regions of CG such as intragenic regions and gene bodies themselves. Unlike CGI promoters, methylation at these regions is usually associated with active gene expression.

It is thought that 5mC DNA methylation is the form of methylation transmitted inter-generationally. The DNA methyltransferase (DNMT) family of enzymes are responsible for the maintenance of this methylation, which can in turn be removed in both an active and passive manner (Gibney and Nolan, 2010).

5mC is a widely documented mechanism for transcriptional silencing, important for the regulation of germline-specific genes, imprinted genes, of DNA repeat elements and genes on the inactive X chromosome (Smith & Meissner, 2013). Exactly how DNA methylation changes as cells differentiate (Meissner, 2011), and its co-operative action with other chromatin modifiers remains unclear.

Main functions of DNA methylation

DNA methylation is one of the predominant epigenetic players in the regulation of gene expression, doing so without causing any changes to the underlying DNA sequence. DNA methylation is an essential control mechanism particularly during mammalian embryonic development helping to direct cells towards their future lineages (Messerschmidt et al., 2014). Acting as an epigenetic barrier DNA methylation can also restrict differentiation (and reversal to an undifferentiated state) as originally postulated by Riggs (1975) and Holliday & Pugh (1975). This epigenetic regulator also plays an essential role in sex chromosome dosage compensation, protection against selfish elements such as transposons, and the appropriate expression of the imprints (Subramaniam et al., 2014). Methylation of DNA is catalysed by the DNA methyltransferase enzymes through the attachment of a $–CH_3$ group that has been donated by SAM (S-adenosyl methionine) to the 5' position of a cytosine base (preceding a guanine) at CG dinucleotides helping to ensure the faithful propagation of DNA methylation patterns between mitotic cellular divisions. Four DNA methyltransferases have been identified so far- the maintenance methyltransferase DNMT1 and the *de novo* methyltransferases, DNMT3A, DNMT3B and DNMT3C, as well as the catalytically inactive cofactor DNMT3L (Hata et al., 2002; Kaneda et al., 2004).

CpG islands - protection from DNA methylation

In mammalian genomes DNA methylation is frequently found at CG dinucleotides or CpGs where Cytosine (C) is followed by Guanine (G) with an intermediary phosphate (denoted by the 'p,' Ziller et al., 2011). The current human genome assembly (hg38) has ~3×10^7 CpG dinucleotides (Edwards et al., 2017). CGI are dense clusters of CG dinucleotides found at 70% of gene promoters- (namely of housekeeping genes and those important for development) (Saxonov et al., 2006), they are generally GC-rich in nature (50% or higher) and are 1000bp in length. CGI encompass 1% of the human genome (Ehrlich et al., 1982).

These CGI frequently have unmethylated cytosines unlike the bulk of chromosomal DNA, and as such have a level of protection against epigenetic reprogramming. This is in part due to interaction with histones and a range of transcription factors including CTCF (CCCTC-binding factor). The CGI are responsible for regulating and initiating transcription through the selective binding of transcription factors that affect the chromatin structure and ultimately influence the

genomic structure where they are localised to further regulate gene activity (Deaton and Bird, 2011). A class of CGI promoters have unique protection from *de novo* methylation by FBXL10- bound by the polycomb repressive complexes PRC1 and -2 (Boulard et al., 2015). Originally there was thought to be little homology between the number of CGIs between mouse and human, with the mouse genome considered to have fewer CGI (Gardiner-Garden and Frommer, 1987). The additional 'orphan' CGI in the human genome however frequently occupy additional intergenic and intragenic regions (Smith and Meissner, 2013; Deaton and Bird, 2011).

Methylated cytosines can be found outside of the CGI in areas called shores, shelves, open sea, UTR and gene bodies. Shores are usually depicted as the region from 0kb to2kb away from CGI (Figure 1), shelves from 2-4kb away from the CGI and open sea (Sandoval et al., 2011) are outside these dimensions usually with isolated CpGs (Rechache et al., 2012).

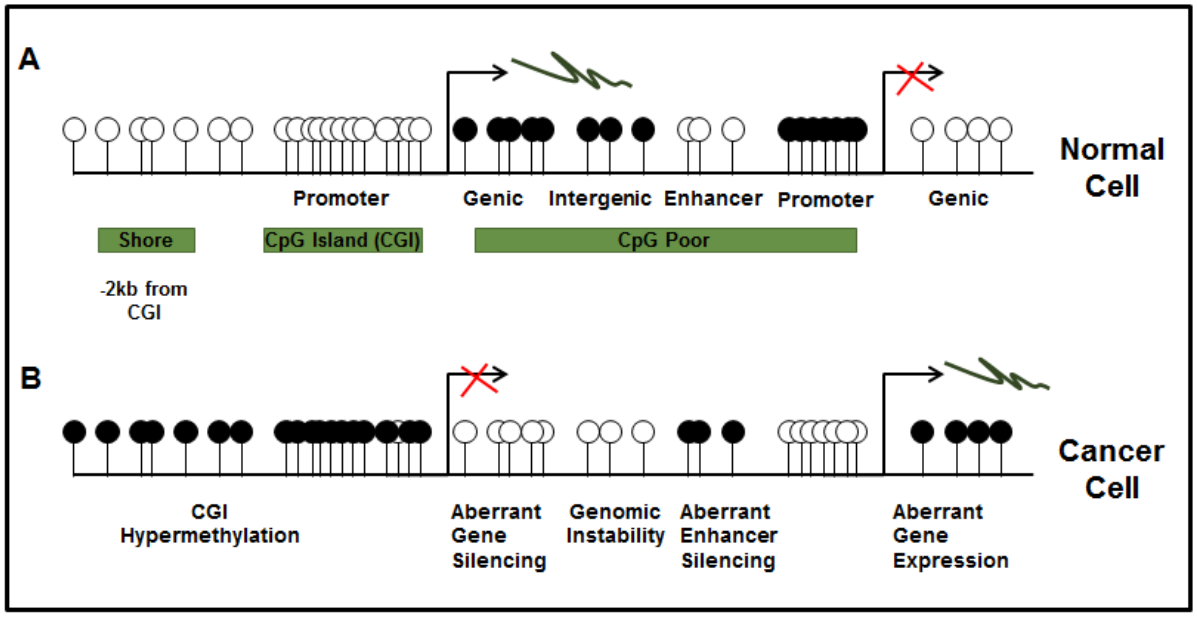


Figure 1. Distribution of CpG dinucleotide methylation and the effect on regulation across the genomes of both healthy and cancerous cells. A) In a healthy cell when the CGI shore and promoter are unmethylated (as represented by the empty circles) there is transcription, and the converse is true when the promoter is methylated (depicted by black filled circles). However, in a cancerous cell B) hypermethylation of the CGI leads to aberrant gene silencing, inappropriate gene expression and the silencing of enhancer elements. Adapted and redrawn from Stirzaker *et al* (2014). Copyright Clearance Centre license number 4270160962178.

Additional cytosine modifications

As well as 5mC, there is an additional cytosine modification called 5-hydroxymethylcytosine (5hmC). 5hmC is the first oxidative product generated when 5mC undergoes active demethylation (Kohli and Zhang, 2013), and it was first discovered in 1952 by Wyatt & Cohen in a T-even bacteriophage (Wyatt & Cohen, 1952) before discovery in mammalian DNA in 1972 (Penn et al.,). This base was also rediscovered in extracts from brain tissue, where it is most abundant (Tahiliani et al., 2009). More robust experimental validation however was achieved in recent experiments in mouse embryonic stem cells (mESCs) (Tahiliani et al., 2009), demonstrated that 5hmC is an intermediate in the demethylation of 5mC (Tahiliani et al., 2009), supported by low levels in the mESC genome at around 10% 5mC and 0.4% of all the cytosine residues (Branco et al., 2011).

Non-cytosine methylation

Methylated cytosines are not only found at CpG dinucleotides, but can also be found at CpA, -T and -C sites in the human genome inclusive of repetitive sequences and gene bodies. This non-CpG methylation was originally documented in the plant genome (Lindroth et al., 2001). Non-cytosine methylation is enriched in particular cell types in the human genome such as the oocytes (Guo et al., 2014), neurons (Lister et al., 2013) and pluripotent stem cells (Lister et al., 2011). CpH (H= A/C/T nucleotides) methylation is crucial for both cellular differentiation and development and in adult somatic cells, whilst only accounting for 0.02% of the overall 5mCs (Jang et al., 2017).

The mechanisms of non-CpG methylation are still poorly understood but it is widely accepted that this form of methylation is catalysed by the *de* novo methyltransferases DNMT3A and -3B. However, both the maintenance methyltransferase DNMT1 and the *de novo* methyltransferase DNMT3A and -3B have been shown *in vitro* to actively methylate non-CpG dinucleotides. CpH methylation is established in a *de novo* manner and DNMT3A is thought to play a role in the maintenance of these CpH methylation patterns however the exact mechanisms responsible are still poorly elucidated (Jang et al., 2017).

Maintenance methylation- DNMT1

The heritability of DNA methylation patterns provides an epigenetic marking of the genome that is stable through multiple cell divisions. Originally discovered in 1988 by Bestor and colleagues (Bestor et al., 1988), through cloning and sequencing the cDNA encoding DNMT1 in mouse cells, DNMT1 was reported to have a strong resemblance to the bacterial cytosine methyltransferase, in particular the catalytic and DNA binding domain. These regions are highly conserved between human and mouse, with 80% homology between proteins in the mammals (Yen et al., 1992).

DNMT1 was the first mammalian DNA methyltransferase identified to faithfully maintain and propagate DNA methylation marks to daughter strands post-replication (Chuang et al., 1997). The DNA methylation marks were originally set during embryonic development by the *de novo* methyltransferase 3A and 3B. The maintenance methyltransferase DNMT1 therefore drives a mechanism that is responsible for securing a stable epigenetic marking of the genome and in doing so creating a cellular memory that is stable through multiple divisions. Incorrect maintenance of DNA methylation has serious implications for the imprints especially (biallelic expression of imprints), X-inactivation on both copies of the X chromosome and activation of selfish elements (Walsh et al., 1998).

The expression of Dnmt1 is known to be activated by cell cycle dependent transcription factors in the 'S phase' of cell division and as such it is highly expressed in mitotic cells (Kishikawa et al., 2003). DNMT1 is located close to the replication foci (Leonhardt et al., 1992), attracted by its PCNA (Proliferating Cell Nuclear Antigen) binding partner NP95 (Messerschmidt et al., 2014) where it preferentially binds to hemi methylated DNA as shown later in 1997 (Yoder et al.,) where it restores methylation (Leonhardt et al., 1992). NP95 has been shown to specifically direct DNMT1 to the parental, methylated strand, where DNMT1 is orientated in such a way that its activity is directed towards the newly synthesised unmethylated daughter strand (Sharif et al., 2007).

In mice, $Dnmt1^{-/-}$ is embryonically lethal and as such the null embryos fail to survive post gastrulation due to the consequences of global hypomethylation (Yoder et al., 1997; Arand et al., 2012) and die from e9.5 (Li et al., 2004). However, in mESCS such a deletion of Dnmt1 causes a widespread hypomethylation but not comparable to the low levels observed in the null embryos. Interestingly, the proliferation and morphology in the mESCs shows negligible

changes, however there is an impaired functionality with respect to differentiation and such cells were observed to undergo apoptosis (Jackson-Grusby et al., 2001) including cells in tissues such as the brain (Chen et al., 1998).

Maintenance methyltransferase- UHRF1

DNA methylation is a fundamental epigenetic mark essential for the propagation of methylation marks through each round of cell division from parent to daughter strand, in doing so it is important in the epigenetic regulation of gene expression. The ubiquitin ligase multidomain protein UHFR1 (ubiquitin-like, containing PHD and RING finger domains) is an essential accessory protein for this process (Hashimoto et al., 2008).

In mammals, UHRF1 has been shown to be essential for maintaining DNA methylation as homozygous deletions phenocopied DNMT1 mutations (Sharif et al., 2007; Bostick et al., 2007). UHRF1 recruits DNMT1 to hemimethylated replication forks (Bostick et al., 2007) through its ability to recognise and cooperatively bind hemi methylated CpGs (and H3K9me2/3) on newly synthesised DNA daughter strands via its SET and RING (SRA) associated domain- a methyl binding domain (Avvakumov et al., 2008; Hashimoto et al., 2008; Lallous et al., 2011). In doing so, UHRF1 can increase DNMT1's activity by ~5 fold (Bashtrykov et al., 2014).

UHRF1 has additional domains which offer additional functionality, the tandem tudor domain (TTD) recognises H3K9me3, and unmethylated H3K4 (Nady et al., 2011; Avvakumov et al., 2008). UHRF1 co-localises with DNMT1 during the synthesis phase of the cell cycle (Liu et al., 2013) and is involved in the formation of closed conformation heterochromatin, facilitated by binding to H3K9me3. This interaction is also important for DNA methylation, Rothbart et al. (2012) showed that a H3K9me3 defective UHRF1 mutant was incapable of restoring methylation levels in UHRF1 knock down (KD) cells. A further domain- the plant homeodomain (PHD) of UHRF1 binds H3 tails that are unmodified at residue arginine 2 (H3R2). The PHD domain has been shown to assist in the reorganisation of pericentromeric heterochromatin in DNA replication (Papait et al., 2008).

UHRF1 is capable of binding both the H3R2 and H3K9me3 of one H3 tail simultaneously through its PHD domain via the PHD TTD respectively (Xie et al., 2012; Arita et al., 2012).

This coordinated recognition of H3R3 and H3K9me3 is required for DNA methylation maintenance (Rothbart et al., 2012).

A further domain belonging to UHRF1 is the RING (really interesting new gene) domain, responsible for the E3 ubiquitin ligase activity that ubiquitinates both histones and DNMT1; therefore, helping to regulate the chromatin structure and stability of DNMT1 (Citterio et al., 2004).

UHRF1 can also be localised to euchromatin, where it is responsible for the regulation of gene expression, notably TSGs (tumour suppressor genes) such as BRCA1. The effects on methylation of UHRF1 deficient cells is similar to that of DNMT1 knockout in mice and mESC cells (Li et al., 1992) whereby global hypomethylation is observed. In *Uhrf1*$^{-/-}$ ESCs major and minor satellite repeats in addition to Intracisternal Alpha Particles (IAP) elements were most demethylated in comparison to other genomic elements (Bostick et al., 2007). Sharif et al (2007) also showed how the effects of hypomethylation in *Uhrf1*$^{-/-}$ ESCs affected ICRs including Igf2-H19m and Dlk1-Gtl2.

The association of USP7 with UHRF1 and DNMT1

USP7 (ubiquitin specific peptidase 7) is a de-ubiquitinating enzyme (Qin et al., 2011) that acts as a regulator for the DNMT1 dependent methylation at hemimethylated DNA sites by actively stimulating the enzymatic activity of the DNMT1 and USP7 complex (Felle et al., 2011). The UHRF1 protein directs the dimeric complex to sites of methylation, and in doing so forms a trimeric complex in conjunction with chromatin. USP7 is therefore acknowledged as playing the role of the middle man between UHRF1 and DNMT1, regulating their interaction and maintaining the stability of DNMT1 by its deubiquitination (Qin et al., 2011). This regulation is possible due to two domains on USP7; the C-terminal and the DRAF domain which interact with their corresponding domains of UHRF1, capable of stimulating both the de novo and maintenance methylation properties of DNMT1 both in vivo and in vitro (Felle et al., 2011).

De novo methyltransferases – DNMT3A and DNMT3B

In the early 2000s, Dnmt1 had still not been proven to initiate *de novo* methylation in vivo despite showing activity in vitro (Chen et al., 2003). Yet in contrast, Dnmt3a and Dnmt3b had been shown to be essential for the *de novo* methylation where they are strongly expressed in ESCs, early/post implantation embryos, developing germ cells (and imprints of) (Okano et al., 1999) but conversely Dnmt3a and -3b are both expressed at low levels in differentiated somatic cells (Chen et al., 2003). Studies of Dnmt3a and -3b inactivation in ESCs provided also evidence that these genes encode the proteins needed to *de novo* methylate proviral genomes and repetitive elements (Okano et al., 1999).

Dnmt3a loss of function (LOF) phenotypes usually present with postnatal lethality between 4-6 weeks and a failure to establish the methylation pattern of imprints in male and female germ cells (Li et al., 2004). A Dnmt3b LOF phenotype however causes embryonic lethality at e14.5 due to vascular and liver complications, in addition there is demethylation of minor satellite DNA. For those null embryos that lack both Dnmt3a and -3b, *de novo* methylation is not initiated after implantation and consequently the deficient embryos die at e9.5 (Li et al., 2004). Dnmt3a and 3-b have also been shown to play a role in the maintenance of DNA methylation patterns (Chen et al., 2003).

DNMT3L

The zinc finger containing gene DNMT3L (DNA cytosine-5-methyltransferase-like) was originally isolated in the human system by Aapola et al., (2000) where it was observed to retain a high degree of homology with human and mouse *de novo* DNMT3A and -3B (Xie et al., 1999; Okano et al., 1998).

DNMT3L has since been identified to be important in the establishment of methylation at gametic differentially methylated regions (gDMRs) in mouse oocytes (Kelsey and Feil, 2013), maturing sperm (Kaneda et al., 2004) and it is highly expressed in both the ovary and testis. *De novo* methylation by DNMT3A is stimulated by DNMT3L, which acts as an essential cofactor in the germline (Chedin et al., 2002). Kaneda et al. (2004) reported on observations that offspring derived from Dnmt3a conditional mutant female mice are not viable in utero. These offspring were identified to have altered methylation levels in addition to allele-specific expression at all of the maternally imprinted loci examined (Kaneda et al., 2004). On the other hand, *Dnmt3a* conditional mutant males were observed to show impairment in the process of

spermatogenesis whilst lacking methylation at two of three paternally imprinted loci examined in spermatogonia. Interestingly the *Dnmt3b* conditional mutants showed no germline phenotype, though they died shortly after birth (Kaneda et al., 2004).

Binding of both DNMT3A in oocytes (Kaneda et al., 2004) and its cofactor DNMT3L is sensitive to the modification state (Liu et al., 2013) of the N-terminal tail of H3. Histone tail binding by DNMT3L and -3A is hindered by methylation at lysine 4 (Ooi et al., 2007) while the PWWP domain of DNMT3A specifically binds H3K3me3 (Zhang et al., 2010; Ooi et al., 2007).

Silencing of retrotransposons and other repetitive DNA elements

Repetitive DNA elements make up ~55% of the human genome as revealed when the human genome was originally sequenced in 2001 (Lander et al., 2001). These repetitive DNA elements are transposable and as such are able to relocate within the genome by jumping from one chromosome to another. These elements and their abundance in the human genome can be further classified as simple sequence repeats, tandem repeats, satellite DNA and segmental duplications and processed pseudogenes (~10%) with the remaining ~45% made up of DNA transposons and retrotransposable elements (RTEs) (Criscione et al., 2014).

Retrotransposition can be described as the transcription of an mRNA intermediate followed by its reverse transcription into cDNA. This process causes insertional mutagenesis and as such disrupts genes in the human genome. Retrotransposition has been shown to happen in the germ-line in a *de novo* fashion, resulting in single gene mutations such as that of e.g. haemophilia A (Hancks and Kazazian, 2012). As such despite playing a role in regulation gene expression transposable elements are frequently silenced to prevent the elements relocating and inserting to the coding region of a gene.

RTEs have the most prevalent transposable element in the human genome and they can be further subdivided into long terminal repeat (LTR) elements, and non-LTR elements. LTRs are similar to retroviruses with respect to their structure and mechanism of retrotransposition, whilst the non-LTRs are akin to integrated mRNAs capable of inserting into the genome in a cut and paste style mechanism that involves the reverse transcription of an RNA intermediate and insertion of its cDNA copy at a new chromosomal location (Goodier, 2016). The non-LTR elements can be further divided into LINES (Long Interspersed Nuclear Elements) or SINES

(Short Interspersed Nuclear Elements) (Levin and Moran, 2011) as represented in Fig. 2 (Wolff et al., 2017) and are as old as the oldest multi-cellular organism that has existed on earth.

The LTR containing elements are predominantly represented by the Line-1 (L1) families and ERVs (endogenous retrovirus transposons), and the non-LTR elements by the Alu families (Criscione et al., 2014). The most prominent group of LTR containing elements however are the ERVs making up 8% of the human genome (McCarthy and McDonald, 2004). Along with the non-LTR LINES, both of these elements employ an autonomous method of transposition (Figure 2) whereby they are capable of active transposition. Non-LTR SINE elements however use non-autonomous retrotransposition and retrotranspose in-*trans* with use of the protein machinery belonging to L1.

A class of LTR-containing ERVs are the IAPs which have flanking LTR sequences which are generally heavily methylated (Rebollo et al., 2012), and a *pol* gene which encodes the reverse transcriptase. When these IAPs become unmethylated in for e.g. *Dnmt1*-deficent mouse embryos a significant number of transcripts are observed to accumulate when DNA maintenance methylation activity is reduced showing that Dnmt1 is essential for the silencing of IAPs in mouse. This is true too for the silencing of L1 elements (Walsh et al., 1998; Bourc'his and Bestor, 2004). Generally, CGI are widely acknowledged to have low methylation despite their high density of CG dinucleotides. CpG sites however outside the typical CG dense regions such as promoters have been found to be highly methylated. Highly methylated CGI include those of selfish DNA elements which act to suppress retrotransposons that could potentially have deleterious effects on the genome (Walsh et al., 1998). Other epigenetic players can be utilised to induce the silencing of potentially deleterious transposable elements such as H3K4me3 in mESCs (Karimi et al., 2011) and CTCF that act to maintain ERVs in an unmethylated state (Rebollo et al., 2012).

Many of retroviral insertions into the human genome are as a result of evolutionary insertions that have been later mutated or disintegrated and as such have become 'junk DNA' (Wolff et al., 2017). There are exceptions to this and some ERVs are capable of being re-transcribed and incorporated into the genome (Kassiotis and Stoye, 2016) where they play a fundamental role in gene expression through cis-acting transcriptional influences (Rowe et al., 2013) and gene regulation (Oliver and Greene, 2011).

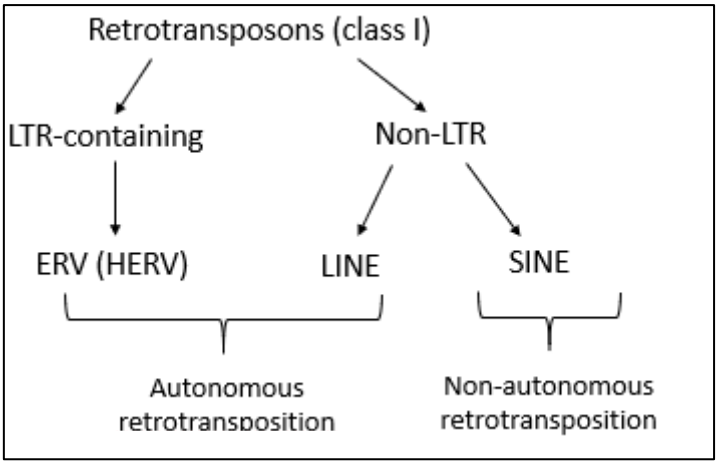


Figure 2- Summary schematic of retrotransposon taxonomy. Class 1 retrotransposons can be delineated into LTR containing, and Non-LTR containing elements. The most highly documented LTR containing retrotransposons are the ERVs. The ERVs along with the non-LTR LINE elements deploy a method of autonomous retrotransposition. This is in direct contrast to the SINE elements such as Alu that are unable to perform autonomous retrotransposition despite their potential to move with the help of the LINE elements. Figure was adapted from Wolff et al. (2017) under the terms of Creative Commons Attribution 4.0 International License.

Mutations in the DNA methyltransferases and their linked syndromes

Mutations in each of the methyltransferases DNMT1, -3A and -3B have been found in human diseases (Edwards et al., 2017). Some of these diseases and associated pathogenic mutations are discussed further below;

ICF syndrome and DNMT3B mutations

The first documented mutation in the human DNA methyltransferase DNMT3B (Hansen et al., 1999) was identified in patients with ICF syndrome (discriminated usually by a variable immunodeficiency, centromeric instability and mild facial anomalies) as one of the two main causes of the syndrome. ICF syndrome (OMIM 602900) is a rare autosomal recessive disorder originally described in the 1970's in patients across Europe (Tiepolo et al., 1979) with an incidence rate of (<1 /1,000,000). The condition typically features centromeric instability (decondensation) of Chr1, Chr16 and infrequently Chr9 (Ehrlich et al., 2006) in addition to aberrant hypomethylation of CpG sites in the cognate pericentromeric satellite repeats (Hansen et al., 1999). The condition is characterised by a range of clinical features that include facial dimorphisms such as a flat nasal bridge, and macroglossia in addition to mental retardation and skin and digestive system infections. ICF can be further subdivided into five clinical groups

dependent on the underlying causative genetic defect- ICF1, ICF2, ICF3, ICF4 and ICFX (Thijssen et al., 2015).

The enzymatic activity of DNMT3B is frequently perturbed in ICF patients due to a single amino acid substitution (A766P and R840Q) in its conserved catalytic domain (Xie et al., 2006). In addition to this conserved catalytic domain found in DNMT3B (and 3A) the *de novo* methyltransferases also contain an N- terminal regulatory region that houses several conserved domains. These domains include the PWWP domain that is 100-150 amino acids (aa) in length and it is frequently associated with chromatin associated proteins, and in particular pericentric heterochromatin (Chen et al., 2004). A missense mutation in the PWWP domain of DNMT3B can cause ICF (particularly ICF1) (Weemaes et al., 2013) and a loss of chromatin targeting capabilities (Ge et al., 2004) in around half of the ICF patients diagnosed.

DNMT3B is responsible for establishing DNA methylation *de novo* early in mammalian embryonic development, and experiments performing site-directed DNMT3B mutations associated with ICF syndrome show that DNMT3B requires a genetic co-factor to bind to in a stable manner (Moarefi & Chedin, 2011). Indeed the mutations in DNMT3B linked with ICF are associated not just with the catalytic domain, but to regions affecting homo-oligerisation, SAM binding and utilisation, and DNA binding (Moarefi & Chedin, 2011). It is important to emphasise that Dnmt3b knockout (KO) in mice is lethal (Okano et al., 1999) proving the essentiality of this catalytically active enzyme for downstream functions (Ehrlich et al., 2006).

Aside from the causative mutations in DNMT3B, Weemaes (Weemaes et al., 2013) reported that 30% of a cohort of 44 patients clinically diagnosed with ICF had mutations in ZBTB24 (zinc-finger-and BTB domain containing 24 gene) (Nitta et al., 2013) and this is referred to as ICF2. Interestingly loss of functional Zbtb24 has been shown to cause embryonic lethality in mouse (Wu et al., 2016). The subtypes ICF3 and -4 have been recently linked to mutations in cell cycle CDAC7 and HELLS respectively. For those clinical diagnoses not stratified into the subtypes 1-4, these are referred to as ICFX (Thijssen et al., 2015).

Leukaemia and DNMT3A mutations

Interest in the *de novo* methyltransferase DNMT3A with respect to leukemic conditions has gained momentum due to the incidence of mutations of this enzyme in an array of hematologic neoplasms (Roller et al., 2013; Yang et al., 2015). Currently all the contributory mutations to Acute Myeloid Leukaemia (AML) in DNMT3A may not be defined (Cancer Genome Atlas Research Network et al., 2013), but overexpression of DNMT3A has been frequently reported in many cancers (Mizuno et al., 2001; Rahman et al., 2010). In fact, heterozygous somatic mutations found in the *de novo* methyltransferase DNMT3A are present in nearly 15% of confirmed patients (Ley et al., 2010) with AML (OMIM 601626) and in some myelodysplastic syndromes (Edwards et al., 2017). The successful use of pharmacological hypomethylating treatments such as 5-aza-2'-deoxycytosine in malignancies further supports the oncogenic overexpression of DNA methyltransferases in tumour cells as one of the most frequent observations (Santini et al., 2001).

Some advances have been made in attempting to identify these mutations by Ley et al (2010) who identified 18 somatic missense mutations of DNMT3A in AML patients using massively parallel DNA sequencing. The median overall survival among patients with such mutations is significantly lower (12.3 months, $p < 0.001$) than among AML sufferers without such mutations (41.4 months) (Ley et al., 2010). In most instances, the mutations affect the R282 codon causing a C → T mutation at a methylated CpG site ultimately converting the codon to one of cysteine if the top strand is mutated, and conversely a histidine codon if the bottom strand is affected (Edwards et al., 2017). This has lately been supported by Brunetti et al. (2017) who state that the mutations in DNMT3A which occur early in leukemogenesis correlate (Yang et al., 2015; Shlush et al., 2014) with a poor prognosis for AML patients. As such mutations in DNMT3a offer potential targets for the development of new therapeutic targets (Brunetti et al., 2017)

Tatton–Brown–Rahman syndrome (TBR)

TBR is an overgrowth syndrome (OMIM 615879) caused by a heterozygous germline mutation in DNMT3A on Chr2p23. In a study of 55 individuals diagnosed with the overgrowth syndrome (Tatton-Brown et al., 2014) over 80% of individuals presented with both an overgrowth and intellectual disability, in addition to a lower incidence of joint hypermobility, obesity, hypotonia and prominent facial features that include low set heavy eyebrows and prominent upper central incisors (Tatton-Brown et al., 2018).

TET proteins

The TET family of enzymes are responsible for the oxidation of 5mC and its subsequent demethylation at select genomic loci. The first of the TETs to be discovered was TET1 (Tahiliani et al., 2009) and it is frequently grouped together with TET2 and -3. This family of proteins are multidomain enzymes around 180-230kDA in size and contain a conserved double stranded β helix, binding sites for cofactors 2-oxoglutarate and Fe (II) that form part of the core-catalytic region in the C-terminus, and a cysteine rich region (An et al., 2017). The core catalytic regions preferentially bind cytosine rich CpG, whilst TET1 and -3 have an additional N-terminal Zinc finger domain (CXXC) that can bind DNA (Frauer et al., 2011) (Figure 3).

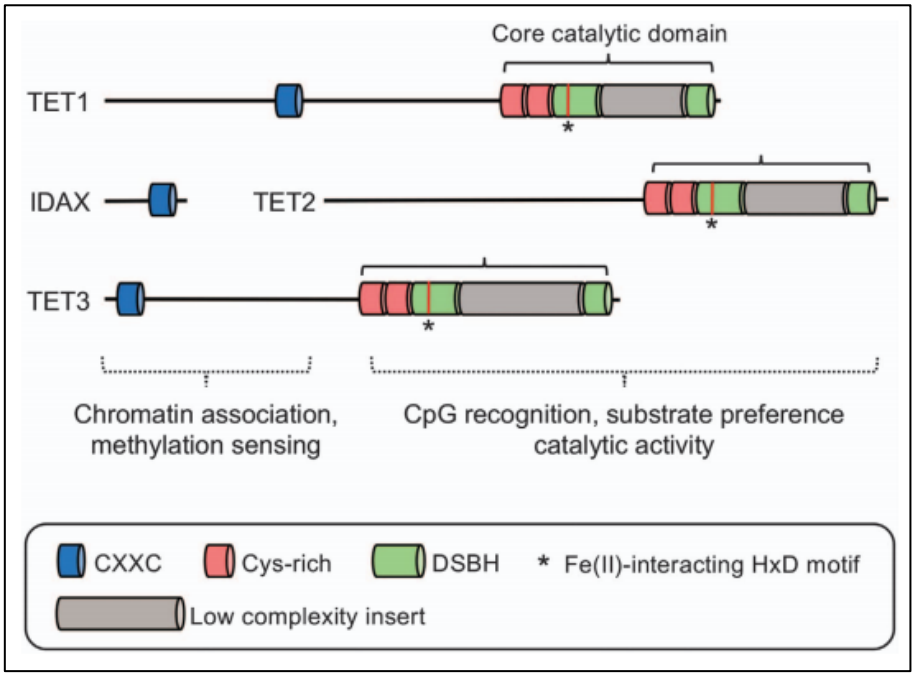


Figure 3: Domain structure of TET proteins. The carboxyl-terminal core catalytic domain is conserved within the TET family. The core catalytic domain is made up of a DSBH and a cysteine rich domain. The cysteine rich domain has two subdomains and is responsible for the modulation of the chromatin targeting functionality of TET proteins. Whilst the DSBH domain has a set of catalytic motifs e.g HxD motif that is capable of interacting with Fe (II) and 2OG. The large low complexity insert highlighted in grey is found in the DSBH domain but no function has yet been assigned to it. Image permitted for reuse under Commercial-Share Alike 4.0 International (CC BY-NC-SA 4.0) from An et al. (2017).

5hmC levels in somatic tissue are relatively low, with the exception of brain tissue where high levels have been reported (Tahiliani et al., 2009). The high levels in ESCs decline steadily as cell lineages progress through differentiation (Ko et al., 2011). This family of enzymes catalyse the oxidation of 5mC to 5hmC in addition to both 5-formylcytosine (5fC), and 5-carboxylcytosine (5aC) (Ito et al., 2011). Functionally however the TET enzymes play a role in the regulation of DNA methylation patterns through 'Active' and 'Passive' DNA demethylation mechanisms. Passive DNA demethylation refers to the lack of maintenance of DNA methylation patterns through cell divisions. This is thought to cause a replication-dependent dilution of 5mC (Rasmussen and Helin, 2016), whereas active DNA demethylation is an enzymatic process whereby the 5mC bases along with their oxidised intermediates such as 5fC are replaced with unmodified cytosines in a replication-independent manner i.e. as catalysed by TETs.

The TET family of enzymes are implicated in Cancer by acting as tumour suppressors- TET2 commonly appears as a mutated gene in hematopoietic malignancies. TET1, -2 and -3 all appear to be mutated in various cancer types showing reduced levels of expression (Rasmussen and Helin, 2016). Through their primary function, TETs play a role in the regulation of DNA methylation patterns and in the progression of normal development, through ensuring protection against anomalous cellular transformation (Rasmussen and Helin, 2016).

Additionally, the deposition of 5hmC may induce passive DNA demethylation. Studies have demonstrated that in vitro DNMT1 activity can be reduced up to 60-fold on a DNA substrate containing 5hmC (Valinluck and Sowers, 2007). Therefore, such TET-mediated 5hmC deposition could prompt passive DNA demethylation on the opposite DNA strand, which could be important to hinder the accumulation of aberrant DNA methylation patterns over cellular generations.

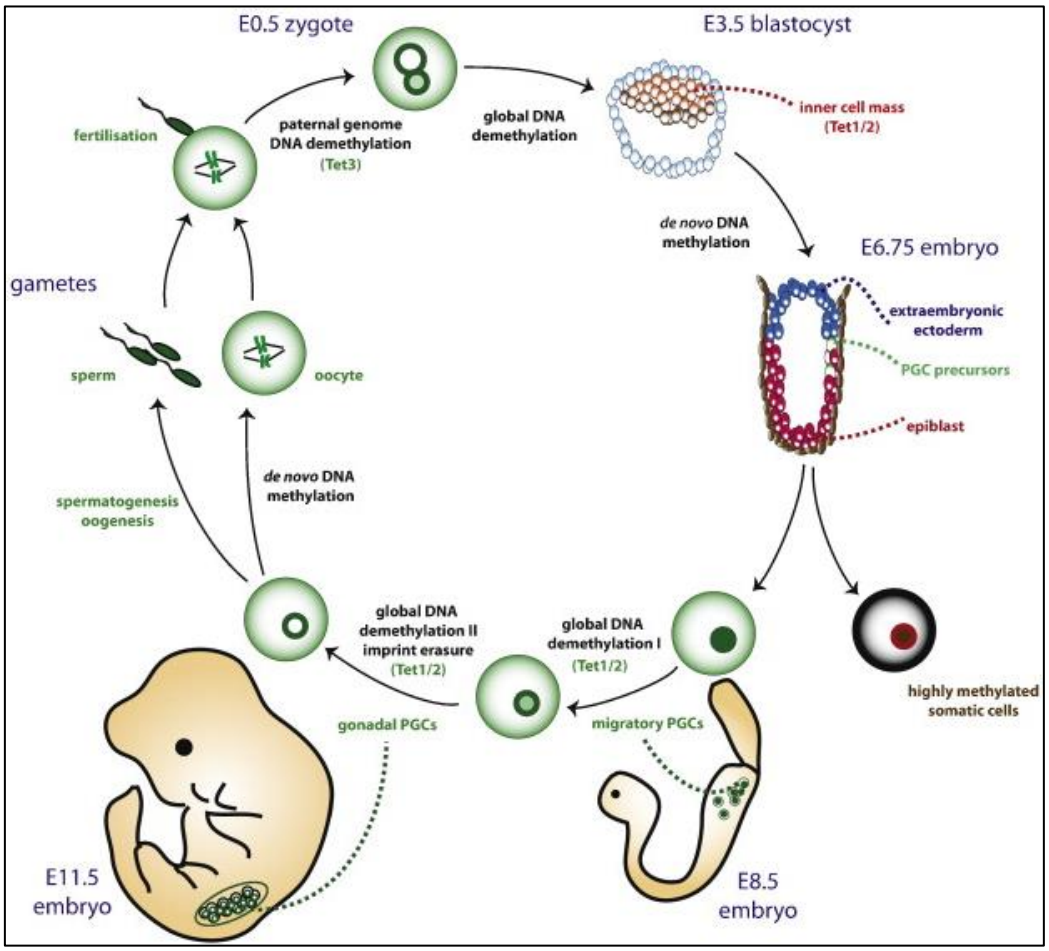


Figure 4: Genome wide DNA demethylation during mouse developmental stages mouse with the relevant Tet protein expression. Epigenetic reprogramming occurs in the early zygote and in the PGCs of a developing embryo. At E.05 the paternal genome of the zygote is decondensed and quickly demethylated. Until the E3.5 blastocyst stage the maternal and paternal experience a loss of methylation, prior to implantation. DNA methylation levels are re-established by the *de* novo methyltransferases following implantation in the epiblast. The second demethylation event in the PGCs occurs is a marker of germline development. This round of demethylation is when the methylation of genomic imprints and repetitive elements is reduce thanks to the Tets. This hypomethylated phase lasts from E8.5 through to post-natal development of the female germ cells during oocyte growth whilst the male germ cells regain their DNA methylation levels quickly following the global demethylation event. Figure sourced from Hill et al. (2014). Reuse was obtained through Copyright Clearance Centre under licence #4270180488769.

In recent years it has been shown that DNMT1 and its recruiting enzyme UHRF1 (and UHFR2) are capable of binding 5hmC DNA (Frauer et al., 2011; Spruijt et al., 2013; Iurlaro et al., 2013) and both UHRF1 and UHRF2 have a role in the maintenance of DNA methylation through recruiting DNMT1 to regions of hemi-methylated DNA. In addition, the two isoforms of the *de novo* methyltransferases DNMT3A2 and DNMT3B2 have also been implicated in the

maintenance of DNA methylation (Chen et al., 2003) in addition to the establishment of DNA methylation on previously unmethylated CG dinucleotides.

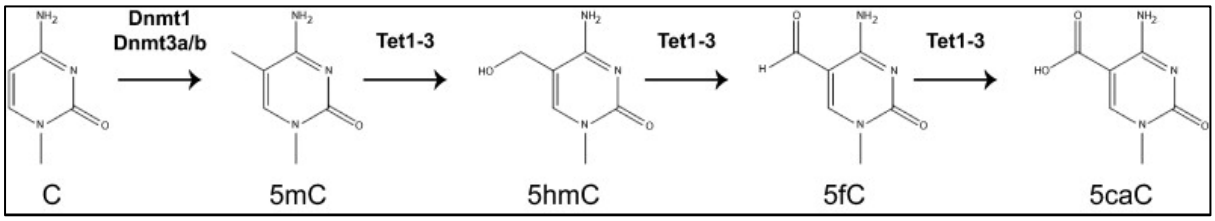


Figure 5: **Common cytosine modifications in mammalian DNA.** Genomic methylation events are frequently catalysed by both the maintenance and *de novo* methyltransferases, and the Tet family of enzymes. The Tet family discovered in 2009 are associated with active and passive mechanisms of DNA demethylation. Figure sourced from Hill et al. (2014). Re-use obtained through Copyright Clearance Centre under licence number 4270180488769

The developmental cycle of DNA methylation

Methylation marks are globally erased following fertilisation on both the maternal and paternal genomes as represented in Fig.4. Following implantation however the methylation marks are reset by the *de novo* methyltransferases DNMT3A and -3B (Okano et al., 1999b). This sequence of methylation changes is fundamental to ensure the appropriate differentiation of the organism. DNA methylation patterns are further erased in the diploid primordial germ cells (PGCs) to allow the sex dependent methylation marks to be reset, maintenance of which is carried out by DNMT1.

Germ-cell development

At an early stage of development, there are a number of cells allocated to form oocytes and spermatozoa and these are referred to as PGCs. The PGCs will migrate to the posterior endoderm and onto the genital ridge, which forms the location of the respective developing gonad, and then the surrounding tissue will form the somatic cells of the gonad around them. These germ cells are precursors to the gametes that are responsible for establishing the next generation in the form of a zygote thanks to their totipotent potential with the inheritance of both genetic and epigenetic information from each parent (Nikolic et al., 2016; Hill et al., 2018). PGCs undergo a wave of demethylation during migration to allow parental specific methylation marks to be reset (Hajkova et al., 2002; Lees-Murdock et al., 2003; Li et al., 2004).

In the mouse genome, germline specification begins its initial stage at gastrulation around E6.25 in the proximal epiblast (Nikolic et al., 2016). As the PGCs migrate into the respective developing gonad they undergo a wave of epigenetic reprogramming E10-E11.5 (Guibert et al., 2012) through a genome-wide loss of 5mC (Hackett et al., 2012). The events leading to global demethylation (in an attempt to restore the epigenome for totipotency) are yet to be fully defined but are considered to be as a result of a combination of both active and passive demethylation (Hackett et al., 2013), and other mechanisms working in parallel such as that of the interplay with chromatin, and Tet dioxygenases (Hackett et al., 2012). It should be noted however that some IAPs, LTR-ERV1 retroelements and single-copy sequences are capable of resisting global demethylation events in PGCs (Hajkova et al., 2002; Lees-Murdock et al., 2003; Li et al., 2004).

X chromosome inactivation

X-chromosome inactivation (XCI) is a method of dosage compensation in mammal genomes whereby one X chromosome is transcriptionally silenced in the female (Lyon, 1961). This heterochromatinisation of the one X chromosome is important (Slotkin and Martienssen, 2007) to make sure that both females (XX) and males (XY) have equivalent X-linked gene dosage (Lyon, 1961; Wutz, 2003). X-inactivation occurs at the X-chromosome inactivation centre (XIC) in the embryo early in development– a *cis* acting sequence, initially before the effects are distributed across the chromosome into adjacent chromatin where some regions are more efficiently inactivated than others on the X chromosome (Panning and Jaenisch, 1996).

Xist is the lncRNA (long non-coding RNA) necessary for selective silencing of one X chromosome per cell (Zlotorynski, 2015), it is expressed from the inactive X chromosome in female somatic cells (Borsani et al., 1991) and continues to be expressed throughout subsequent cell divisions where it is thought to recruit additional silencing factors necessary for inducing a stable heterochromatic conformation (Ng et al., 2007). Beard et al. (1995) reported observations that when Dnmt1 expression is disrupted, Xist becomes hypomethylated resulting in an increased levels of Xist RNA in the male Dnmt1 mutant ESCs and embryos; and therefore, showing a functional role of DNA methylation in regulating Xist expression appropriately in the differentiation of murine somatic cells (Beard et al., 1995). As such DNMT3A is essential for methylation of the *Xist* promoter of the X chromosome and is therefore important for X-inactivation (Chen et al., 2003). This is echoed in a review by Barbara Panning (Panning, 2008) who discusses the role of DNA methylation in X-inactivation.

Xist promoter methylation has been shown to correlate with *Xist* expression in ESCs (Nesterova et al., 2008). There are two regions that flank the Xist transcription start site (TSS) which exhibit high levels of DNA methylation. In fact, Nesterova et al. (2008) showed in two XY ESC lines that had mutations in the Xist promoter mutations that Xist expression increased in accordance with DNA hypomethylation at these locations. The methylation of the Xist promoter occurs before the onset of X-chromosome inactivation and is affected by the levels of sense and antisense transcription, independent of the RNAi pathway (Nesterova et al., 2008) (as shown in Figure 6).

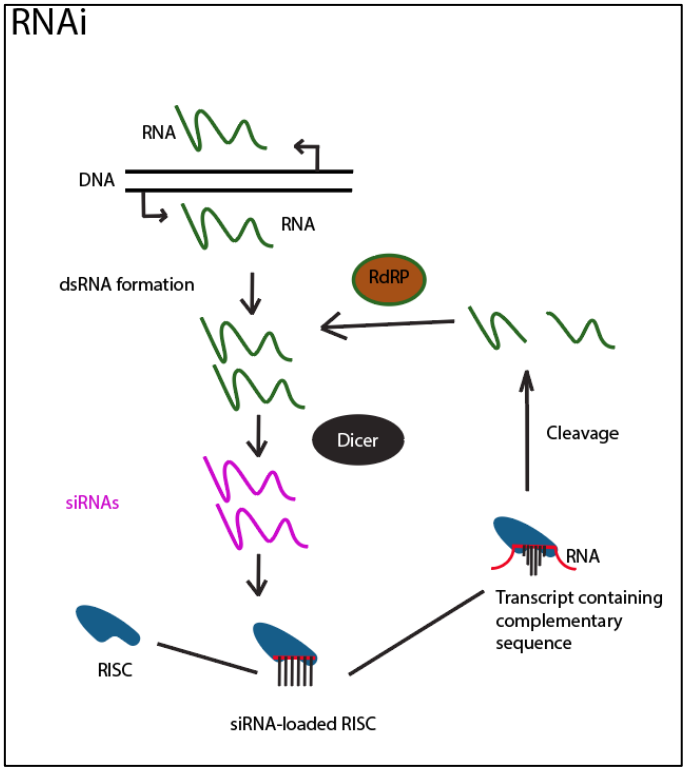


Figure 6: RNAi mechanism of transposable element silencing. Post transcriptionally double stranded RNA (dsRNA) is cleaved by the dicer proteins into small interfering RNAs (siRNA). These siRNAS become incorporated into the RISC complex. Which then cleaves complementary transcripts to the loaded siRNA. Image was redrawn from Slotkin & Martienssen (2007). Reuse was obtained through the Copyright clearance centre under licence number 4271071318628.

DNA methylation of germline specific genes

DNA methylation acts as an epigenetic barrier to the expression of many gene classes, namely the germline genes expressed in the mammalian germ cells (Borgel et al., 2010). The expression of these germline specific genes are predominately controlled by DNA methylation in a tissue specific manner.

In early mammalian development, the genome undergoes extensive reprogramming, for which DNA methylation plays a fundamental role. Typically, the CGI that are located at promoters of germline specific genes become *de novo* methylated in mammalian germ cells in an attempt to maintain their suppression. This however is not true of all promoters; some repressed genes do have such methylated CGI at their promoter to encourage repression where long-term stabilisation of their repressed state is preferred. Such classes of genes include the imprinted genes, those found on the X chromosome in addition to those only expressed from the germ cells. The repression of the germline specific genes is important in somatic tissue where their expression is unsuitable (Jones, 2012). This phenomenon is important in the epigenetic reprogramming of PGCs, particularly in the instance where if the germline specific genes were to be inappropriately expressed ectopically there could be deleterious downstream effects (Maatouk et al., 2006; Borgel et al., 2010). For example, Dnmt3l KO mice are viable but due to the diminished ability to assist in the *de novo* methylation of germline genes, as well as inappropriate expression of RTEs, the knockout can cause male sterility and embryonic lethality of the maternal null derived embryos (Bourc'his et al., 2001; Bourc'his and Bestor, 2004). Dnmt3l is an important cofactor of the de novo methyltransferases Dnmt3a/b and plays an important role in methylation despite its poor catalytic activity (Chedin et al., 2002).

DNA methylation andtTranscription

The epigenetic mechanism of DNA methylation involves the addition of a methyl group to the C5 position of cytosine to form 5mC. This mechanism is important to regulate gene expression through the recruitment of proteins that are involved in expression or alternatively through the inhibition of a variety of transcription factors. During mammalian development, DNA methylation undergoes a wave of changes due to the interaction of *de novo* and maintenance methyltransferases and is important in downstream processes such as X-inactivation, silencing of transposable elements and genomic imprinting (Smith and Meissner, 2013). To promote

active transcription, DNA methylation in the gene body can enhance gene expression. This is converse to methylation at the gene promoter where DNA methylation is associated with gene repression in a tissue specific manner (Moore et al., 2013) e.g. germline genes in somatic tissues. The gene body frequently refers to the genomic region following the first genic exon (Moore et al., 2013). The exact mechanism by which DNA methylation of the gene body can contribute to gene regulation is yet to be fully elucidated.

Cells treated with the cytosine analogue 5-aza-2′-deoxycytidine reactivate genes and decrease expression of genes such as those genes regulated by c-MYC e.g. genes involved in metabolic processes. DNA demethylation of the gene bodies causes down-regulation and the expression of genes is restored by the *de novo* methyltransferases. Given the effect of methylation at the gene body this may offer a therapeutic target for alternative DNA methylation inhibitors that can assist in the normalisation of gene overexpression that is a consequence of carcinogenesis (Yang et al., 2014).

DNA methylation and cancer generally

The initial observation of a link between DNA methylation and cancer was published in 1983 showing how cancers have a tendency to exhibit global hypomethylation on comparison with normal human controls (Feinberg and Vogelstein, 1983; Gama-Sosa et al., 1983). Indeed, aberrant DNA methylation is a prominent event in carcinogenesis and the associated loss of methylation was soon shown to be an early event in tumorigenesis (Feinberg et al., 1988) resulting in tumour progenitor cells.

With this knowledge in hand, cancer research at that time reverted to focusing on dense regions of hypermethylation (predominantly at promoters) with the idea that the epigenetic silencing of TSGs could act as a “second hit,” thereby essentially nullifying the need to inactivate the genetic pathway as reviewed in Baylin et al. (1997).

In normal cells, CGIs are surprisingly rarely methylated with some exceptions such as those on the inactivated X chromosome and at imprinted genes. Aberrant methylation of CGI is not limited to cultured cells but can occur during aging and tumorigenesis (Baylin and Jones, 2011). Observations indicate that unmethylated CGI are associated with a variety of genes that have either become partially or fully methylated in tumours cells and are able to undergo reactivation by 5-aza-2’-deoxycytidine. Some of the methylation at CGI has little to no effect

on downstream gene activity simply because these are CGI not associated with the regulatory regions of genes (Jones et al., 1998).

In 2001, a study by Esteller et al. (2001) showed that in over 600 primary tumour samples tested the promoters associated with CGI in at least three or more genes (from a panel of 12 known TSGs) were hypermethylated in 5–10% of the samples tested. Of these hypermethylated TSG at least one CGI was methylated in 80% of samples for each tumour type tested (Esteller et al., 2001). Using methods that allow genome wide screening of CpG islands, it was observed that around 1% of CGI in DNA from tumour tissues are abnormally methylated whilst demonstrating tumour specific patterns of methylation (Costello et al., 2000).

Imprinting

Imprinting refers to the process of allelic restriction of the expression of genes depending upon the parent of origin (Docherty et al., 2014). Imprinting is maintained through various epigenetic mechanisms such as DNA methylation. DNA methylation involves the transfer of a methyl group from 5-adenosylmethionine to 5'carbon of a cytosine residue, converting it to 5-methylcystosine (Gibney and Nolan, 2010) and it is essential for differentiation and normal development, due to its effects on long term stable gene repression (Cedar and Bergman, 2009). In addition to DNA methylation, the effects of the presence of particular repressive or active histone marks at a site of differential methylation may affect the levels of DNA methylation in an individual. As a more complete review of imprinting during development was conducted as part of the Review article (Paper-IV) I will here concentrate on aspects not covered there, namely correlations between imprinting and cancer.

Cancer and Imprinting Syndromes

A loss of imprinting (LOI) is frequently observed in human cancers and it is often caused by misregulation of epigenetically controlled genes (Ogawa et al., 2003; Cui et al., 2003; Holm et al., 2005). Monk (2010) suggests that there is in fact a higher frequency of epigenetic changes than DNA mutations in cancer. For example, LOI can include the reactivation of the typically silenced copy of the imprinted growth promoting IGF2 or silencing of the active parental copy of p57KIP2 (a growth inhibitory gene) (Diaz-Meyer et al., 2003). LOI is so prevalent that is

has been reported in 100% of all documented cases of chronic myeloid leukaemia (Randhawa et al., 1998), in addition to 70% of Wilm's tumours (Mummert et al., 2005) and even 80% of ovarian tumours (Kamikihara et al., 2005).

The imprinted control region of the growth-related locus *IGF2/H19* ICR in Wilm's tumour is a well-documented and common cancer that is found in patients (mostly children) who have the overgrowth imprinted disorder Beckwith-Wiedemann (BWS). The degree of risk of a child with BWS developing a Wilm's tumour is dependent on the causative molecular mechanism, with some mechanisms thought to be lower risk than others e.g. loss of methylation at the centromeric imprinting centre (IC2) (Brzezinski et al., 2017). Despite the efforts made to stratify Wilm's tumours the Weksberg lab still recommends that all children diagnosed with BWS continue to be screened for Wilm's tumours until the age of eight (Brzezinski et al., 2017). As a consequence of having this disorder, there is biallelic expression of the growth promoting gene *IGF2*, which has the effect of promoting the growth of cancer cells (Ogawa et al., 1993).

Cancerous cells in humans generally undergo global DNA hypomethylation with locus specific hypermethylation. These events have the consequence of initially causing a decondensation of chromatin, activation of certain retroviral elements and can cause chromosomal rearrangements due to the inherent instability caused (Monk, 2010). Hypomethylation of genes is usually paired with the derepression of e.g. oncogenes (e.g. c-Myc and c-Ha-Ras) (Cheah et al., 1984) and MAGE-A1 or-A3 germline genes (De Smet et al., 2004). Conversely, hypermethylation of DNA promoters can cause gene-specific silencing, and in some cases these genes can be tumour suppressors for e.g. non-small cell lung cancer (Liu et al., 2016). Genic misregulation can also be caused by aberrant methylation through atypical DNMT expression. If there is an overexpression of the maintenance methyltransferase DNMT1 hypermethylation might result (Etoh et al., 2004), whereas global hypomethylation appears to correlate with the expression of an abnormal isoform of DNMT3B (DNMT3B4) which does not have the normally conserved methylation motifs and so cannot participate in sufficient maintenance or *de novo* methylation activity (Saito et al., 2002).

Pharmacological inhibition of DNMT1- Mechanism of Action

The pharmacological DNMT1 inhibitors 5-azacytidine (5-aza) and 5-aza-2′-deoxycytidine (5-aza-dC) were originally synthesised in 1964 (Sorm et al., 1964). 5-aza in particular was developed as a nucleoside antimetabolite for AML, where it could be activated to the nucleoside triphosphate. As a result, it was capable of being incorporated into the replicating DNA and preferentially into the RNA (Herman et al., 1998).

The hypomethylating effect of 5-aza-dC (or Decitabine as it is clinically known) however is thanks to the DNMT1 inhibitors innate ability to act as cytosine analogue (Patel et al., 2010). 5-aza-dC was initially developed as a cytostatic agent (Sorm et al., 1964), but treatment of human cell lines with the agent showed it was capable of inhibiting DNA methylation activity and thus the DNMT1 inhibitor stimulated the development of such azanucleosides as potential epigenetic drugs (Stresemann and Lyko, 2008). The DNMT1 inhibitor and cytosine analogue 5-aza-dC is preferentially incorporated into DNA. This is despite 5-aza-dC having ten times more cytotoxicity than 5-aza (Flatau et al., 1984; Momparler et al., 1984).

When 5-aza-dC is incorporated into DNA, an irreversible covalent bond is formed between 5-aza-dC and a cysteine residue located within the active site of DNMT1 (Seelan et al., 2017). This binding renders the DNMT1 ineffective and unable to operate its full methylation maintenance activities at CG dinucleotides of freshly synthesised DNA.

The treatment dose levels of 5-aza and 5-aza-dC are low enough to avoid causing cell death, as they incorporate into the replicating DNA of cells in culture, ultimately leading to a rapid loss of DNA methyltransferase activity as the enzyme becomes irreversibly bound to the cytosine residues in the DNA (Fig.7). Consequently, the amount of hemi-methylated DNA available falls according the amount of cytosine that is methylated. In Fig. 7B treatment also causes the 5-aza-dC incorporated DNMT1 to be targeted for proteosomal degradation (Fig.7B.2) (Ghoshal et al., 2005). Such degradation of unbound DNMT1 reduces the DNMT1 available to bind to the hemi methylated DNA and prevents it from being remethylated (Fig.7B.3), as such this contributes to local hypomethylation at this location.

The extent of the hypomethylation caused by 5-aza-dC is more noticeable at low concentrations where the formation of 5-aza-dC-DNMT adducts is restricted as to not hinder normal DNA

synthesis. At higher concentrations of the inhibitor however, a larger number of adducts are formed which has the effect of restricting DNA polymerase from carrying out its normal function. This ultimately causes a change in gene expression, decondensation of chromatin (Jones, 1985) growth arrest and in many instances cell death (Kuo et al., 2007). Although it is worth noting that the inhibition of DNA synthesis does not take effect until at least two cycles due to the lack of repair or the DNMT1- 5'aza-dC adducts (Davidson et al., 1992).

Interestingly there is evidence indicating that administering 5-aza-dC treatment may also act by enabling the reactivation of ERVs that are typically found to be silenced through epigenetic mechanisms (reviewed in Wolff et al., 2017). Pharmacological agents such as the DNMT inhibitor 5-aza-dC has been reported to upregulate immune signalling in cancerous cells through the viral defence pathway (Chiappinelli et al., 2015). Indeed, viral defence genes that include IFI27, ILI44 and IFNB1 have been found to be upregulated in primary ovarian tumour samples sourced from 19 patients and this significantly correlated with the high expression of the levels of ERV transcript ($p = 0.000141$).

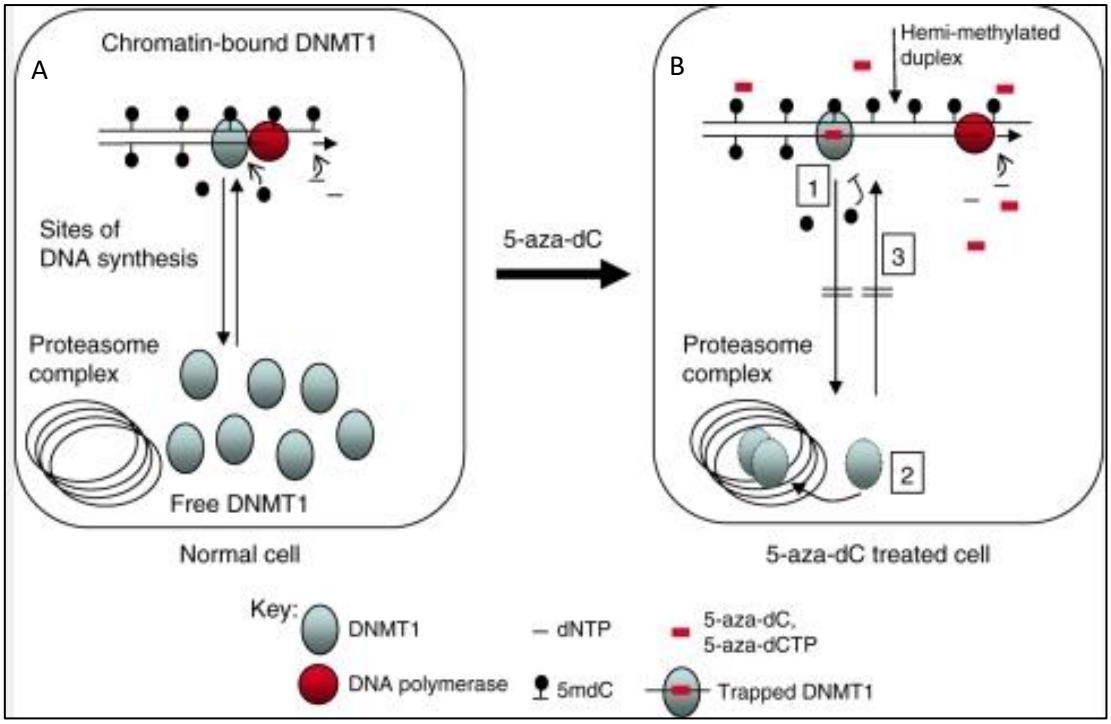


Figure 7- Impact of treatment with 5-aza-dC on DNA methylation and subsequent breakdown of DNMT1. DNMT1 selectively binds to the hemi-methylated DNA created as part of the usual DNA synthesis process under the action of DNA polymerase in actively dividing healthy somatic human cells. At this stage, there is a free interchange of the maintenance methyltransferase and the chromatin as shown in A) by the arrow. In such cells, there is a minimal amount of hemi-methylated DNA available due to efficient maintenance methylation being carried out by DNMT1 (mostly). Upon addition and incorporation of 5-aza-dC, in B) DNMT1 tries to methylate the analogue by flipping out the base but is unable to complete the reaction and so becomes trapped in a dosage dependent manner (as shown by the red dashes) (1). Figure was sourced from Patel et al. (2010): Copyright Clearance Centre #4270170712545.

Main clinical uses- Effectiveness of DAC

A range of nucleoside inhibitors such as 5-aza-dC, Decitabine (FDA) (Kaminskas et al., 2005) or Dacogen (NICE) -have been approved for the treatment Myelodysplastic syndromes (MDS) and for AML under the condition that the marrow blast counts are between 20-30% in addition to fulfilling the criteria for AML as per the World Health Organisation (WHO) classification of the condition (Nieto et al., 2016).

The main uses of nucleoside analogues 5-Azacytidine and 5-aza-2′-deoxycytidine primarily reside in their role as therapeutic agents for the treatment of certain cancers through epigenetic silencing of important regulatory genes. The inhibitors can do this through hypomethylation of aberrantly methylated CpG promoters that leads to re-expression of e.g. TSGs (Wolff et al., 2017). This is frequently referred to as ‘Epigenetic Therapy’ (Li et al., 2016). When 5-aza-2′-

dC was administered to AML patients aged over 60 in a study by Yan et al., (2012) global methylation was significantly reduced when compared with pre-treatment baseline in cells of AML patients.

Yang (2014) reports that 5-aza dC treatment successfully down regulates genes due to DNA hypermethylation of the gene bodies and transient exposure to this agent can cause also demethylation of CGI promoters (Tsai et al., 2012). Both effects exhibit anti-tumour capabilities. Aza's widespread demethylating effects have been shown in a range of cell lines by the Walsh lab (Irwin et al., 2014; Rutledge et al., 2014) and previously by others (Jones and Taylor, 1980).

Such pharmacological loss of DNMT1 is problematic for normal non-cancerous cells. DNMT1 is essential for cell viability, $Dnmt1^{-/-}$ mouse embryos are embryonically lethal and fail to develop throughout the gestation period (Li et al., 1992). On the other hand ESCs can continue to proliferate without DNA methylation (Tsumura et al., 2006) but *Dnmt1* KO ESCs die when promoted to differentiate (Li et al., 1992). The appropriate maintenance of DNMT1 is of crucial importance for development as it is used to silence genes at developmentally important milestones in embryo development. Post development, DNMT1 has been associated with cell viability with evidence in a range of adult differentiated cells which have diminished viability when DNMT1 expression has been reduced (Jackson-Grusby et al., 2001; Chen et al., 2007).

Loss of DNMT1 has also been implicated in the activation of the DNA damage response (DDR) pathway. The DDR is employed to protect the integrity of the human genome and its mechanisms predominately include cell-cycle checkpoints, the detection of DNA lesions and DNA repair as nicely reviewed in Giglia-Mari et al., (2011). DNMT1 deficient hypomorphic systems such as those sequenced in O'Neill et al., (Paper II) offer an opportunity to look at which enriched genes/proteins of the DDR become up/down regulated when the DDR is activated and so identify those that affect cell viability. Many studies into the effects of acute depletion of DNMT1 have been performed in hypermethylated cancer cell lines where the effects make it more difficult to identify the triggers for DDR and cell death (Loughery et al., 2011). DNA repair pathways and 5-aza dC-mediated DNMT1 depletion correlates with the active recruitment of DNMT1 to sites of potential DNA repair in cells that have undergone cell cycle arrest where it can restore methylation marks (Easwaran et al., 2004). The major checkpoint mediator p53 has been identified to be activated in cell death following the pharmacological depletion on DNMT1 with the creation of 5-aza dC adducts in DNA (Karpf

et al., 2001); for example, the mediator p53 is required to safeguard G2/M arrest seen under 5-aza-dC treatment of the colon-cancer HCT116 cell line (Schneider-Stock et al., 2005). A PARP-mediated cell killing mechanism involving PARylation of the DNA and apoptosis-inducing factor 1 (AIF1) translocation, and which is at least partially independent of p53 but sensitive to PARP inhibitors, is also involved in DDR in DNMT1-depleted hTERT-1604 and HT29 cancer cells (Loughery et al, 2011).

DNA packaging into chromatin

Chromatin allows for the packaging of DNA into a dynamic structure within the nucleus of a cell in such a way that the DNA can be propagated and transmitted within cell division. However, DNA must be packaged appropriately for it to remain accessible to the protein machinery required for transcription, DNA repair and even recombination (Rothbart and Strahl, 2014). Chromatin is a complex of DNA and histone proteins, the basic unit of which is a nucleosome. A nucleosome contains 145-147bp of DNA wrapped around a histone octamer (Lawrence et al., 2016). The histone octamers are made up of two copies of H2A, H2B, H3 and H4 each, resulting in a nucleosome core particle (Kornberg and Lorch, 1999) forming the basis of chromatin. Chromatin can take the form of euchromatin and heterochromatin; euchromatin has a looser packing structure and its more open conformation is more accessible to transcription machinery. Heterochromatin on the other hand is associated with a closed chromatin conformation due to its condensed packing and it is considered to be repressive to active transcription (Li and Reinberg, 2011).

The organisation and packing of chromatin is dependent on factors which include; the action of linker histone H1, chromatin remodellors and histone chaperones among others. The histone H1 in particular regulates the creation of higher order chromatin states (Harshman et al., 2013) through its 50bp linker region and the associated ability to tether to DNA on the outside of the nucleosome allowing nucleosomes to interact together. Ultimately this allows for the formation of a spiral of 6-8 nucleosomes/turn creating a 30nm fibre- a solenoid. Heterochromatin results when the solenoids are folded to form the dense transcriptionally repressive chromatin (Li and Reinberg, 2011) and a resultant metaphase chromosome (Annunziato, 2013).

The ordered structural organisation of chromatin is key for the appropriate regulation of transcription, DNA repair and the process of replication and so too are the building blocks of chromatin- nucleosomes. Nucleosomes are known to vary in their composition of the histone protein element due to the incorporation of a range of histones and the post-translational

modifications of their amino acid side chains (Luger et al., 2012). These chains can be targeted by the chromatin remodelling complexes, and through the hydrolysis of ATP the nucleosomes can be mobilised using DNA translocase. This allows the DNA to unwind as nucleosomes are evicted to improve accessibility to the transcriptional machinery through such alternation of the chromatin structure (Venkatesh and Workman, 2015).

Histone modifications

As discussed previously the nucleosome units have amino acid side chains which extend from the nucleosome. These side chains can be modified post-translationally mostly at the N-terminal of the tails, but also some at the C-terminal. Some modifications however can take place in the histone folds (globular domains) responsible for regulating interactions between histone-histone and histone-DNA (Cosgrove et al., 2004). The lateral surfaces of these globular domains is the outer region of the histone octamer and so they have direct contact with the DNA. Recently mass-spectrometry has revealed novel modifications in this area as described in Lawrence (Lawrence et al., 2016) and Tropberger (Tropberger and Schneider, 2013).

The post translational modification of histones is important for the regulation of processes such as replication, transcription and DNA repair and for nucleosomal architecture (Tessarz and Kouzarides, 2014). Covalent modifications of the histones can alter not only the organisation of chromatin, but also chromatin function. The covalent modifications are laid down by chromatin modifying enzymes where they act as docking sites for specific chromatin readers. This has the effect of recruiting chromatin modifiers and other such remodelling factors (Tessarz and Kouzarides, 2014). Examples of post-translational modifications include methylation, acetylation, phosphorylation, ubiquitinylation and sumoylation (Lawrence et al., 2016). Many of these modifications are carried out by specialised enzymes such as the histone methyltransferases, acetyltransferases and deacetylases. Recently there have been additional modifications that have been described - propionylation and butyrylation (Kebede et al., 2015).

Some of the modifications found on the N-terminal of the side chains are able to directly affect the nucleosome-nucleosome interactions. This is evident in H4K16ac which has the effect of causing a reduction in the level of chromatin compaction and increasing transcription both in vitro and in vivo (Shogren-Knaak et al., 2006). Conversely the tail modifications can cause increasing chromatin condensation in vitro e.g. H4K20me2/me3 (Lu et al., 2008). Furthermore, post-translational histone modifications can recruit effector proteins which

activate downstream signalling (Wysocka et al., 2006), in addition to blocking the access of chromatin remodelling complexes. Additionally, such modifications can also affect the recruitment of both transcription factors and chromatin modifiers. It is important to note however despite the effect of such modifications on nucleosomal activity, the tails and their modifications can be removed whilst having minimal effect on nucleosomal stability and integrity (Ausio et al., 1989).

The majority of post-translational histone modifications are reversible to provide a level of regulation over how long a mark can reside at a particular genomic location endeavouring to maintain genomic instability in addition to offering a mechanism to remove any aberrantly placed marks; for e.g. acetylation of H2A.Z (a variant of H2A) by the promoter bound acetyltransferase acts to prevent its removal from the promoter by the INO80 remodeller complex (Lawrence et al., 2016).

Histone methylation is one of the most well described post translational modifications in the literature and some of the most extensively studied histone methylation sites include H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20 (Greer and Shi, 2012). The addition of a methyl group is catalysed by the enzyme histone methyltransferase which uses SAM as a methyl group donor. Histone methylation can occur at residues including arginine, lysine and histidine, whilst the removal of methyl groups can be facilitated by the histone demethylases. The lysine residue can be monomethylated, di- or trimethylated on their ε-amine group, whilst arginine can be monomethylated, symmetrically dimethylated or asymmetrically dimethylated on their guanidinyl group. Histidines on the other hand have been reported to be monomethylated but this methylation is rare with little investigation to further characterise it, as reviewed in Greer & Shi (2012).

Chromatin remodelling complexes

As discussed previously, the amino acid side chains of the nucleosomes can be targeted by chromatin remodelling complexes. For the most part, these large macromolecular complexes are ATP-dependent, using energy from ATP hydrolysis to structurally alter the nucleosomes by sliding, disassembling or other means (Luger et al., 2012). The remodelling complexes are found at a ratio of 1:10 to the nucleosomes (Langst and Manelyte, 2015). Such ATP-dependent chromatin remodelling complexes have been identified as novel targets for cancer therapy (as described in (Mayes et al., 2014). These complexes possess 2 to 20 non-catalytic subunits

essential for the targeting and regulation of nucleosome positioning activities of the remodelling complexes. Thus, they are very important for the determining gene expression and the cell fate.

The ATP-dependent complexes contain an ATPase subunit, part of the SNF2 superfamily. The SNF3 superfamily can be subdivided into three groups; SWI1/SNF2, the imitation group SWI (ISWI), and the ATP-dependent complexed which contain a Snf2-like ATPase. This third class has also been shown to exhibit deacetylase activity (Vignali et al., 2000). The SWI/SNF family has an N-terminal located HSA (helicase-SANT) domain which recruits actin and other such related proteins, and at the C-terminal there is a bromo domain used for binding the acetylated lysines of histones (Filippakopoulos and Knapp, 2012). The ATPases of the ISWI family on the other hand have a C-terminal SANT domain beside their SLIDE domain (SANT-like ISWI). These components form a nucleosome recognition unit that is able to bind to DNA and unmodified H4 (Clapier and Cairns, 2009).

There is another class of remodelling complexes- the ATP independent complexes. The rearrangement of the nucleosomes has a direct impact on the transcription levels as accessibility of the DNA is affected by the chromatin packing and organisation (Gibney and Nolan, 2010). Most of the known mutations in chromatin remodelling complexes identified in a variety of cancers can be found in the SWI1/SNF2 complex. The mutations within this particular remodelling complex can be found in ~20% of cancers, most of which have been identified to be inactivating mutations and as such suggest a tumour suppressive function (Kadoch et al., 2013).

Polycomb Interaction

Histone modifications can also be influenced by the polycomb group of proteins, originally identified in 1985 in the species *Drosophila melanogaster* (Jurgens, 1985). The polycomb proteins are a set of transcription regulatory factors and can be stratified into functionally distinct groups- PRC1, and PRC2. PRC1 has E3 ligase activity, and its substrate is the monoubiquitinated form of H2A at lysine 119. PRC2 on the other hand has methyltransferase activity and generates H3K27me2/me3 (Di Croce and Helin, 2013). One of the main PRC2 enzymes is EZH2 (enhancer of zeste 2) which contains a SET domain responsible for the methyltransferase activity of the complex (Aranda et al., 2015). Both PRC1 and -2 further

interact with additional factors to regulate enzymatic activity and in which fashion the chromatin is recruited to the complex (Aranda et al., 2015). Moreover, PRC2-Ezh1 and –Ezh2 exhibit distinct chromatin binding properties, as illustrated by the specific chromatin compaction property of PRC2-Ezh1 (Margueron et al., 2008).

The activity of polycomb proteins during development is critical for successful gestation: deletions of some of the polycomb proteins such as Eed, Ezh2 or Suz12 results in embryonic lethality in mice (Schuettengruber and Cavalli, 2009). The abundance of polycomb proteins in development is also depicted by Mohn who reported that PRC2 targets represent ~10% of the genes in ESCs (Mohn et al., 2008).

Epigenetic mechanisms such as histone modifications and chromatin remodelling are key for the appropriate regulation of gene expression. Post-translational modifications to the side chains extruding from the nucleosomes include histone acetylation and methylation. Such modifications impact on the packing and structure of chromatin controlling its accessibility to transcriptional machinery. The polycomb group of proteins are important part of silencing through its influence on histone modification in a cell and developmental stage specific manner (Gunster et al., 2001). These proteins are responsible for the repression of gene expression through histone methylation e.g. H3K27 and H3K9 (Lindroth et al., 2004), and also the ubiquitination of H2A (Wang et al., 2004).

Relationship between post-translational histone modifications and DNA methylation

DNA methylation and post-translational histone modifications are two fundamental mechanisms used to regulate transcription amongst other functions (Hu et al., 2009). There are a number of post-translational modifications involved in the recruitment and utilisation of the DNA methylation machinery. In particular, the histone modifications can play a fundamental role in shaping the DNA methylome (Stewart et al., 2015). However, the mechanisms responsible for targeting the DNA methylation machinery at key developmental milestones are poorly understood.

H3K4 and DNMT3L

DNMT3L is a regulatory co-factor for the de-novo methyltransferases DNMT3A and -3B (Chen et al., 2005). The co-factor DNMT3L binds to the tail of H3K4 in vitro. Ooi & Bestor discovered that it is the N terminus of H3 that is specifically required for *de novo* methylation, which occurs in dense regions of heterochromatin that lack H3K4 (Ooi et al., 2007). Dnmt3a's methylation levels are dependent on the PHD domain of Dnmt3L's ability to recognise the tail of H3 with its unmethylated lysine 4 which was shown through a functional analysis of DNM3L in mouse ESCs.

Further evidence to support the role of *de novo* methylation is supported by observations through the targeted mutagenesis of Dnmt3L in mouse. Such observations have revealed that 3L is required for the methylation of essentially all sequences, including imprinted loci in both the male and female germ cells (Hata et al., 2002; Kaneda et al., 2004), retrotransposons and other sequences, consistent with Dnmt3L's ability to co-localize and co-immunoprecipitate with Dnmt3a and -3b (Hata et al., 2002).

CpG islands experiencing active DNA methylation have been shown to have reduced levels of H3K4me2 and H3K4me3 in oocytes. In particular methylome profiling of oocytes that were deficient in H3K4 demethylase KDM1A or KDM1B showed that a lack of H3K4 methylation is fundamental for the correct methylation establishment at CpG islands (Stewart et al., 2015). Indeed, trimethylation and dimethylation of H3K4 (K4me3/K4me2) is uniformly associated with sites of transcription initiation (Deaton and Bird, 2011). It follows suit that the promoters of ESCs that are dense with CpG are associated with H3K4me3 which is supported by the observations that *de novo* DNA methylation is selectively inhibited by methylation of H3K4 (Ooi et al., 2007; Otani et al., 2009; Williams et al., 2011).

H3K27 and -3A

H3K27 is associated with regions of the genome silenced by polycomb transcriptional repression, catalysed in particular by the Ezh1/2 components of PRC2 (Rose and Klose, 2014). Conversely to H3K4 and its association with a state of active transcription, the trimethylation of H3K27 is usually associated with transcriptional repression. However bivalent domains have been shown to exist in many developmental genes in ESCs. Bivalent domains are unique in that they contain both H3K4me3 and H3K27me3 and are thought to poise genes for future

activation or repression. Through stimuli, which promote differentiation the bivalent domains, are able to settle into either the active H3K4me3 or repressive H3K27me3 monovalent chromatin structures (Kong et al., 2016). The repressive H3K27me3 is not specifically localised to a particular element of the genome but instead is found widely distributed throughout. Interestingly the PRC2 complex, which methylates H3K27me3 is primarily localised to dense regions of unmethylated CpGs and afore mentioned bivalent domains reside here (Kong et al., 2016). The opposite is true however in somatic and cancerous cell lines where the repressive H3K27 is not strictly localised to CGI.

Genome-wide techniques to study the epigenome

History of technologies for assaying methylation and transcription

Since the accelerated development of sequencing technologies following the first large-scale practical method coined by Sanger (Sanger et al., 1992) further development of the sequencing field was needed to promote automation, develop high throughput sequencing approaches and cut the cost of sequencing per base exponentially (as reviewed in Heather and Chain, 2016). In doing so massive parallelisation was achieved (Anderson and Schrijver, 2010). These early sequencing techniques are often referred to as first generation sequencing and were particularly important for the requirements of the large-scale collaborative Human Genome Project which was ultimately completed in 2003, with an initial draft published in 2001 (Lander et al., 2001).

Throughout the 1980s and 1990s many research groups explored alternatives to electrophoretic sequencing which ultimately did not pay off until the human genome project. NGS technologies differ mostly from these techniques in their ability to multiplex, and cyclical biochemistry e.g. polymerase mediated incorporation of fluorescently labelled nucleotides and imaging- that is Sequencing by Synthesis (SBS).

SBS methods are at the fore-front of these parallel DNA sequencing technologies. The principle of SBS is based upon the incorporation of four base-specific fluorescent dyes. The dyes are specific for each of the organic bases guanine, adenine, cytosine and thymine, and are used instead of the more traditional radioactive labels (Ju et al., 2006). The replacement of the laborious gel electrophoresis with automated capillary electrophoresis has also helped to promote parallelisation. Examples of these SBS techniques include pyrosequencing (Ronaghi et al., 1998), the sequencing of individual DNA molecules (Braslavsky et al., 2003) and polymerase colonies (Mitra et al., 2003).

Post-Sanger the development of DNA sequencing technologies enabled the explosion in capacity and the exponential drop in sequencing costs/base through massive parallelisation. A fragmented input sample of genomic DNA or cDNA (depending on the application of the array- can be whole genome or expression) is captured on an array in such a way that each well or feature is populated by a single target molecule (Trevino et al., 2007).

Current methods for genome-wide assessment of DNA methylation

Techniques currently available to assay methylation on a genome wide scale at single nucleotide resolution differ in respect to the amount of input DNA, resolution and probe coverage of the genome and the ease of downstream bioinformatic analysis of the respective platforms (Yong et al., 2016). Such third-generation experimental approaches to study epigenetic profiling of the methylome can be categorised into the following experimental approaches (Figure 8);

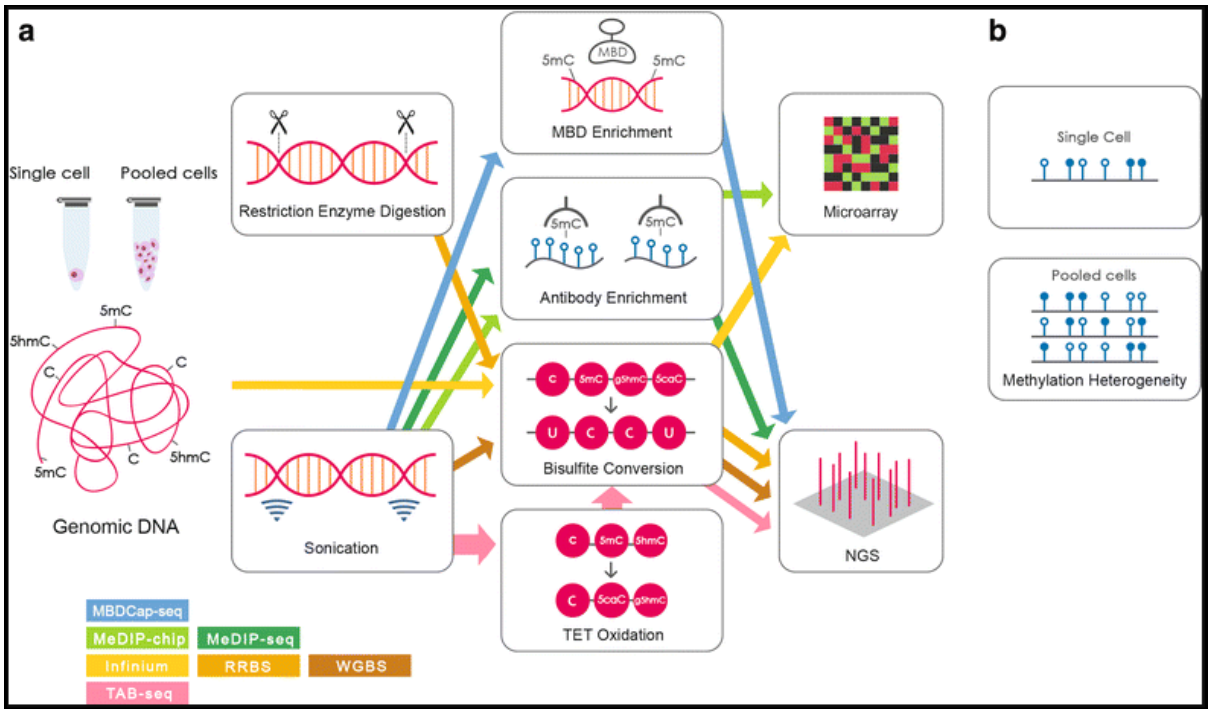


Figure 8- Diagram showing commonality of genome-wide techniques to analyse methylation. A) Fragmentation of DNA is often generated through sonication or digestion by restriction enzymes. From here the genomic segments of DNA can undergo MBD (Methyl Binding Domain) or antibody enrichment, bisulfite conversion (for the detection of 5mC or 5hmc as depicted by the circles) or TET oxidation. A microarray or NGS platform then analyses the enriched fragments. B) The methylation analysis of single cells permits the assessment of methylation heterogeneity in whichever populations of cells being tested. Contrarily, it is possible to assess methylation in pooled heterogeneous cell populations. The filled blue dots represent 5mC, whilst unfilled represent cytosine. Each track is one read. Figure sourced from Yong et al. (2016) with re-use allowed under the Creative Commons Attribution 4.0 International License.

i) Reduced-representation bisulfite sequencing (RRBS)

Developed by Meissner et al (Meissner et al., 2008), this technique applied the reduced representation shotgun sequencing approaches used for SNP discovery to assay methylation. Here the genome is subjected to restriction enzyme digestion using one or more enzymes, fragments in a certain size category are selected then bisulfite conversion is followed by next-generation sequencing for the downstream analysis of the patterns of specific fragments. While the digested fragments only represent typically 5-10% of the genome, they were found to capture 85% CGI when they were of 40-220bps in size. These CGIs mostly lay in promoter regions, which had the effect of limiting the genome coverage, as promoters are only found in 1-3% of the genome; thus there is a lack of coverage at intergenic and some distal regulatory elements. This however is a more cost-effective approach than whole-genome shotgun sequencing (WGBS) yet retains high sensitivity and single base resolution.

ii) TET-assisted bisulfite sequencing (TAB-seq)

This expensive approach to sequencing allows the creation of 5hmC profiles at a single nucleotide resolution in both Human and mouse ESCs. In this instance the 5hmC is protected from TET-mediated oxidation through the addition of glucose to the 5hmC with the use of β-glucosyltransferase (β-GT) to form β-glucosyl-5 hydroxymethylcytosine (Yong et al., 2016). Tet1 oxidises the 5mC to 5caC and this and the unmethylated C are susceptible to bisulfite conversion. Consequently, they are sequenced as T, and 5hmC is sequenced as C. There are complications with TAB-seq whereby the Tet enzyme is used at a low efficiency and as such may omit some methylated cytosines. This can be overcome by using more processive Tet enzymes with conversion rates of more than 96% (Yu et al., 2012).

iii) Comprehensive High Throughput Arrays for Relative Methylation (CHARM)

This approach uses a restriction enzyme *McrBC* to fractionate the input DNA, followed by hybridisation to an array (Irizarry et al., 2008). *McrBC* is responsible for cleaving half of the methylated DNA, and all the methylated CGIs (Sutherland et al., 1992), and after size selection hybridises the fragments to an array. The cost-effective method has a moderate resolution as the coverage is limited to sites that are close to the recognition sites of the restriction sites, however this does have the advantage that the CGs assayed are not selective to only CGI and promoter regions and so in this respect offers genome wide coverage. DMRs outside CGI

shores have been detected using this method, which is not possible with the immunoprecipitation-based methods like MeDIP.

iv) Methylation DNA Immunoprecipitation (MeDIP)

As a cost-effective method MeDIP uses an anti-methylcytosine antibody to immunoprecipitate DNA with CG sites that are methylated (Zhao et al., 2014). The enriched fractions can then be further analysed using either MeDIP-chip or MeDIP-seq. MeDIP-seq has a typical resolution of 100-300bp and is particularly sensitive in areas of low CG coverage. 1x coverage has been suggested by Taiwo et al (Taiwo et al., 2012) to cover up to 70% of all CpGs in human, and this can even be achieved with as little as 1ng of genomic DNA starting material. This proves useful for methylation analyses in small rare samples and micro-dissected tissue samples.

v) Whole Genome Bisulfite Sequencing (WGBS)

This approach uses purified and sheared genomic DNA fragments. The fragments are end repaired, a poly-A tail is added, followed by the ligation of methylated adapters (Urich et al., 2015). Prior to bisulfite conversion and polymerase chain reaction (PCR), the fragments are size selected, and a library is sequenced. WGBS is an expensive method of sequencing, but it does offer nearly full resolution of all the CG sites (even at lower density CG regions and distal regulatory elements), despite not being able to discriminate between 5mC and 5hmC at these locations. WGBS has been employed by the large epigenome consortiums such as ENCODE (ENCODE Project Consortium, 2012) and BLUEPRINT (Adams et al., 2012).

vi) Methylation sensitive Restriction Enzyme Sequencing (MRE-Seq)

MRE based sequencing techniques rely on enzymes that include BstUI, HpaII, NotI and SmaI. MREs cleave at the unmethylated CpG, leaving the methylated DNA intact. Indeed, this is the underlying principle of MRE-seq, whereby upon enzyme mediated cutting of the unmethylated site, the fragmented DNA is selected upon size and sequenced (Li et al., 2015). Sequencing this way shows the MRE recognition sites within the unmethylated DNA fragment, and in this way DNA methylation can be estimated but at a low resolution due to the limiting factor of the recognition sites containing CpGs (Li et al., 2015).

vii) Illumina Infinium Methylation Array- from Generation 27k to EPIC

The Infinium family of methylation arrays from Illumina was adapted from its BeadArray technology that was originally designed for genotyping to quantify methylation at individual

CG dinucleotides at defined probes across the human genome (Bibikova et al., 2011). Prior to the launch of the HumanMethylation27 BeadChip (27k), the existing platforms required to assay methylation relied upon tiling microarrays, or large scale bisulfite-genome DNA sequencing for example. Both techniques relied upon a large amount of input material and were labour intensive with complex downstream bioinformatic analysis. These drawbacks rendered the use of such platforms useless where there were only small amounts of precious material in a large study.

The 27k was released in 2008 (later phased out) and had the capacity to quantify DNA methylation at 27,578 CG dinucleotides over a region of over 14,000 genes in the human genome. Only 1µg of genomic DNA is required as input material with an availability of 12 samples per chip to promote high throughput and comparable analyses.

The 27k and later generations of the BeadChip use Infinium technology. Infinium technology is based upon the bisulfite conversion of genomic DNA, such that the unmethylated cytosines are deaminated to uracil whilst the methylation status of the methylated cytosines is maintained. Post conversion the individual samples are whole genome amplified and fragmented using enzymes. It is these bisulfite converted fragments that are purified before their addition to their respective BeadChips.

There are two bead types- one type for the methylated cytosine, and the other for the unmethylated thymine at the CpG locus. Primers anneal in an allele specific fashion, before the single base extension facilitated by DNP- and Biotin-labelled dideoxynucleotides (ddNTP). Each bead type for the same CG dinucleotide can incorporate the same DNP/Biotin- labelled ddNTP, and this is incorporation is dependent on the base that is before the cytosine of interest at the CG dinucleotide for which methylation is to be measured. Both types can be measured in the same colour channel.

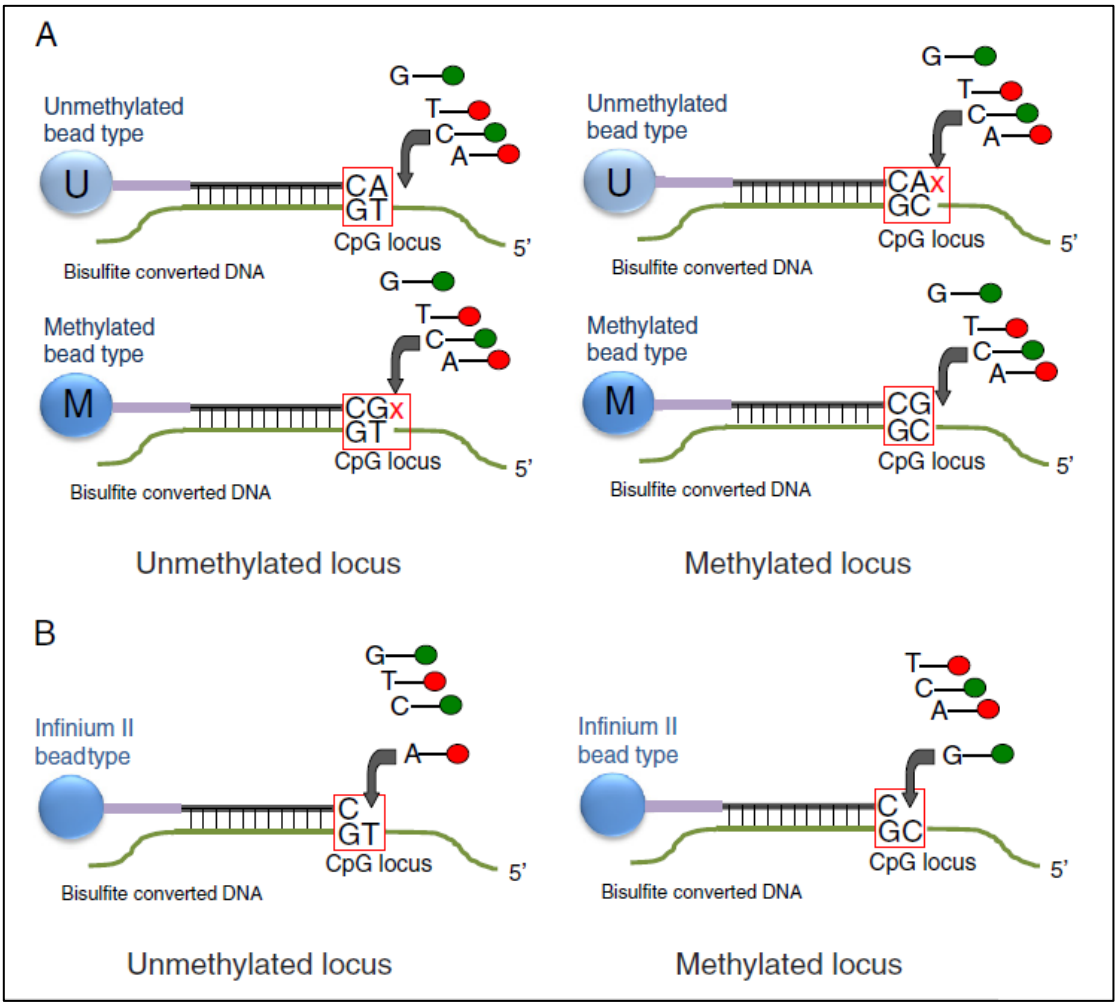


Figure 9: Schematic depicting the mechanism of each Infinium bead type in Illumina's Methylation BeadChip Arrays. A- Infinium I assay) Each CpG locus has two correspondent beads, one for methylated cytosine, and one for unmethylated thymine at that dinucleotide. The same labelled nucleotide is associated with both bead types for that locus, and this is determined by the cytosine that precedes the cytosine at the CG dinucleotide being assayed for methylation; thus both values are recorded in the same channel. B- Infinium II assay) In this instance only one bead type correspondents to each CpG locus, not two as per the Infinium I assay. The probe contains 1-3 underlying CpG sites, in addition to a degenerate R base. This R base corresponds to cytosine at the CG dinucleotide. Methylation levels are assayed through single-base extension, and the locus is detected in two colours. Methylation state is detected by single-base extension. Infinium Image taken from Bibikova et al. (2011) according to the Copyright Clearance Centre under order number 4395961226402.

After each subsequent extension the array is stained fluorescently and through scanning of this florescence the intensities of both the methylated and unmethylated beads is measured before downstream analysis via e.g. Illumina's BeadStudio software. DNA methylation is measured on a β scale from 0-1 that is from being an unmethylated CpG site to a fully methylated CpG

site (Barrera and Peinado, 2012). The beta score is derived from a ratio of the intensity of the methylated bead to the combined locus intensity. Initially the process preparing for and completing 27k was up to one week.

The 27k was superseded in 2011 by the IlluminaHumanMethylation450 ('450k') array that had the ability to profile the methylation status of over 450,000 CG dinucleotides. The 450k probes cover 99% of RefSeq genes, spanning 17,755 unique CpG Islands with additional coverage in areas identified as CpG shores and miRNA promoters (Morris and Beck, 2015). The design of the 27k array uses Infinium I probe chemistry (two beads per probe- one bead in the red channel, and one in the green channel) unlike the 450k array that uses Infinium II. Infinium II was designed to allow more probes to be accommodated on the array, whilst retaining 30% of the Infinium probes (Morris and Beck, 2015). The genomic locations of probe types in the 450k array can be seen in Table 1. The presence of more than one Infinium probe type however can cause type II bias during analysis (Teschendorff et al., 2013), which needs to be corrected downstream through bioinformatic curation using normalisation strategies such as BMIQ (Beta Mixture Quantile dilation) (Teschendorff et al., 2013).

Region Type	**Probe Design**		**Total**
	I	**II**	
CpG Island	77,764	72,850	150,524
CpG Island Shore	22,371	89,696	112,067
CpG Island Shelf	6,913	40,231	47,144
Open Sea	28,518	147,529	176,047
Total	135,476	350,036	485,512

Table 1- Number of 450k array loci stratified by probe design and genomic region. Different proportions of both probe types are aligned to delineated genomic regions ranging from the CG dense CGI to the sparse regions of the open sea. The proportions of probes reflect this density. The total number of probes on the 450k array were totalled at 485,512. Table adapted from Aryee et al. (2014). Reuse of this table was approved using Copyright Clearance Centres RightsLink® Service (Order Number 4317021403928).

The 450k BeadChip however despite its improved coverage lacked probes over distal regulatory elements such as the enhancers, distal regulatory elements and DNase hypersensitivity sites. To address this, Illumina developed the MethylationEPIC (EPIC) BeadChip in late 2015. The EPIC BeadChip has an improved coverage compared to that of the 450k, with probes for over 850,000 CG dinucleotides (with a concordant overlap of 90% with the 450k probes) with probes distributed across many genomic regions as shown in Table 2. It

is important to note however that all generations of the Infinium arrays contain probes specifically designated to determining the quality of the data generated, and some sample independent and –dependent controls (Illumina, 2011). Illumina responded to users of the 450k BeadChip for requests for improved coverage at regions such as enhancers (ENCODE) and DNase hypersensitivity which designers addressed. The EPIC array also covers the enhancers as identified by FANTOM5 across multiple human tissue types and miRNA promoter regions among others.

Feature Type	No Covered	% Covered	Avg No Loci/Feature
Island	26,000	>95	6
North Shore	25,000	>90	3.5
South Shore	25,000	>90	3.5
North Shelf	22,000	>80	2
South Shelf	22,000	>80	2

Table 2: Distribution of EPIC array probes across multiple gene regions. The probes of EPIC area are distributed and spread across genomic features at a coverage level of 80-95% with 2-6 loci covered per feature (Illumina, 2015).

BeadChip Array IDATs

Intensity Data Files (IDAT) are used to store BeadChip array data generated by the Illumina suite of platforms that offer genome wide profiling e.g. the Illumina iScan system which can be used for epigenome wide 450k and 850k BeadChip methylation profiling (Smith et al., 2013). IDATs are used by the respective scanner (e.g. iScan) to store intensity values for each of the array probes compactly. These IDATs can then be analysed using the Illumina licenced Genome Studio and/or a range of open source platforms and pipelines such as the R packages illuminaio or RnBeads- both of which can be used in RStudio. However, it is important to note that the Genome Studio software supported by Illumina fails to extract all of the available information that can be extracted from IDATs such as the control probe intensities and meta-information that provides information about the physical scanning process and software versions.

The type of array platform used dictates the type of IDAT that is formatted and used as the output. The IDATs generated in the process of methylation array profiling using the BeadChips

are binary files – in that the red and green channels have distinguishable and individual IDATs. These binary files contain an identification of the bead type from whichever array has been used, their mean and standard deviation intensities and the number of each type of bead present. Metadata including details about the array are also included such as the type of BeadChip and specific software used in the generation of the respective output IDAT. On the other hand, gene expression BeadArray data is usually presented in encrypted XML (eXtensible Markup Language) format with ten fields of data, compared to four fields for BeadArray Methylation data.

Illuminaio was one of the first free packages that offered a user-friendly approach to extracting data for downstream analysis from the IDATs from Illumina BeadChip sequencing platforms whilst opening the possibility of creating an open-source sharing of array data in open access databases. This in turn has facilitated the use of publicly available data between multiple research groups on an international scale. As of 2013 (Hansen) 1.5% of the Illumina BeadChip array data submitted to GEO have included these raw IDATs as supplemental material

RnBeads

As one of the freely available software tools used to analyse 450k (Assenov et al., 2014) and more recently 850k IDATs, it is arguably one of the most comprehensive packages that currently exists to manage high resolution Infinium DNA methylation array data. RnBeads offers a manageable approach to allow the user the opportunity to avail of a completely integrated pipeline capable of the running and tailoring of multiple analytical modules. Some of these workflow modules include normalisation, quality control (QC), exploratory analysis, differential methylation and others specific to the user's defined analysis aim, all from one individual function rnb.run.analysis(). Some of the modules in the workflow such as normalisation can be run with various options depending on the nature of the study by using SWAN (Maksimovic et al., 2012), BMIQ (Teschendorff et al., 2013), watermelon (Pidsley et al., 2013), or the methylumi noob method (Davis et al., 2007). RnBeads was designed based on a review of other software tools that previously existed to analyse DNA methylation microarray data; with this in mind the developers of RnBeads based their development on several key essential features (Table 3).

Key Feature	Rationale
Support	Single nucleotide resolution from genome wide methylation assays
Functionality	Offers data visualisation of DNA methylation data, QC, exploratory analysis, differential methylation analysis.
Interaction	RnBeads can generate reports that are both interactive and adjustable within defined parameters.
Standardised	standardised yet configurable workflow outputs
Flexibility	RnBeads does not require high-spec computers or cluster and can even run through an online cloud service.
Performance	RnBeads maintains the ability to analyse not only 450k and 850k data but that of WGBS and RRBS methylation data.
Reproducibility	The pipeline generates replicable outputs when using comparable input parameters and analytical options.

Table 3: Fundamental features of RnBeads are based on seven key principles, for which a rationale is offered for each. This table is amended from text in Assenov et al. (2014).

R and Object Orientated Programming

R is an extensible programming language (XML) capable of introducing different approaches to object orientated programming (OOP). OOP itself refers to a type of programming language model that combines data and functionality into an object. These objects are arranged into classes and interact with each other. For those who were responsible for developing R the two approaches to OOP were S3 and S4 class objects.

S3 class objects allow computations to be carried out via a generic function that decides what type of method to call. These generic methods are simply ordinary functions. S3 is a casual class object, easily manipulated and informal by nature and as such is widely implemented. On the other hand, S4 is a more rigorous and formal class system with special helper functions for defining generics and methods. S4 is capable of multiple dispatch permitting genetic functions to pick methods based on the class of multiple arguments and is frequently employed by Bioconductor in the analysis of large microarray datasets. S4 class objects offer lower error rates due to the rigorous definition of the object, and as such when a function acts on this type of object the function is clear as to what information is contained within that object. S4 objects

allow ease of coordination for multiple contributions to large datasets and projects (Gentleman et al., 2004).

Bioconductor

The Bioconductor project is an open source collaborative suite of bioinformatic software tools for the analysis of high throughput genomic data that primarily runs on the R programming language. The goals of the Bioconductor project are three-fold in that the collaborative effort aims to 1) nurture the on-going development of innovative software, 2) diminish potential barriers to initiating interdisciplinary scientific research, and 3) to bolster the success of remote reproducibility of research results for the wider scientific community (Gentleman et al., 2004). To comfortably use Bioconductor the user requires some background knowledge in biology, computer science and in statistics however many users are not aligned to be comfortable with all three disciplines. In this instance Bioconductor produces support documentation accessible to all users.

The Bioconductor project attempts to simplify the statistical inferences researchers are required to perform when analysing large genomic data sets by providing a set of statistical methods that are easy to interact with. This is implemented by placing emphasis on data management and transformation, machine learning, and the continued development of data modelling strategies that are complementary to computational biology and bioinformatics. In addressing these areas of emphasis, Bioconductor offers a software environment that promotes transparency, reproducibility and efficiency of development.

The team behind the development of Bioconductor had an interest in the data management, downstream analysis and problems therein of data contained in DNA microarrays. To suitably address these concepts, a programming environment was required that offers complex numerical analyses, potential for data visualisation, opportunity to access databases and provides a comprehensive range of statistical and mathematical algorithms.

Minfi

Minfi is a package implemented in the Bioconductor suite designed for the analysis of DNA methylation array (Aryee et al., 2014), and the Infinium 450k BeadChip array in particular from Illumina (Bibikova et al., 2011). Modifications have been made to the Minfi package since the release the Illumina HumanMethylationEPIC array (Fortin et al., 2017). Minfi functions through modular representations of the 450k array data by using S4 classes. Minfi

begins with the raw IDAT files that hold both the red and green channel intensities. From there, the 'RGChannelSet' class sorts the intensities into an object that can be accessed by multiple methods. Once the data is converted to methylation measurements it can be stored as the following four classes that each can form part of the pre-processing data;

- MethylSet
- GenomicMethylSet
- RatioSet and,
- GenomicRatioSet.

'Genomic' refers to the non-reversible association of the methylation loci with a genomic location. Whereas the two classes MethylSet and GenomicMethylSet represent the raw Methylated and Unmethylated measurements. These two classes accommodate pre-processing routines that are responsible for the final measurements in these channels. These final measurements include normalisation that is currently included with Illumina's GenomeStudio Software. The RatioSet and GenomicRatioSet classes represent the data as beta values or M-values (log ratios of beta values). The ratioConvert function converts MethylSet to RatioSet and GenomicMethylSet to GenomicRatioSet. This design provides a flexible framework for method development and downstream analyses, the natural starting point for which tends to be the GenomicRatioSet class.

Minfi as a package simplifies the import of raw data from the IDAT files or pre-processed data from GenomeStudio (Illumina's default software), or indeed the public deposited and available array data in NCBI GEO- from there the data can also be conveniently stored in GenomicRatioSet objects.

Tertiary Analysis

The resultant genomic data generated from processing of methylation array data is subject to downstream tertiary analysis. This analysis uses a variety of databases and reference systems to translate the output data to gene lists to allow functional annotation e.g. GO and KEGG classifications. However, whilst there are a growing number of databases for the functional annotation of genome information (such as PFAM, REPBASE, and ENSEMBL), I will concentrate here on those which were used most in the work presented in this thesis.

Galaxy

The Galaxy web-based platform was initially launched in 2005 to offer scientists with minimal levels of computational expertise an opportunity to perform free informatic analysis of large datasets. The focus of the Galaxy online platform, which can be found on a public server at http://usegalaxy.org, is to offer a range of tools to aid downstream data manipulation and visualisation. The Galaxy software framework is an open-source application, with which the user interacts by uploading their data and analysing through servers. The platform is free and offers substantial CPU and disk space.

The Galaxy platform offers an efficient approach to attacking large datasets and integrating these resultant gene lists using analysis tools and offers functionality to visualise interactive user defined tracks on genome browsers such as UCSC genome browser (Kent et al., 2002) or IVG (Robinson et al., 2011). The three-fold goals of the Galaxy project are to;

- Increase access to complex computational analysis for those with limited computational knowledge.
- Create workflows that are automated, multi-step and available for multiple users to promote analysis automation.
- Provide collaborative analyses through enabling Galaxy users to share and publish their analyses online

PANTHER Classification System

Protein Analysis through Evolutionary Relationships (PANTHER) is a free online software for denoting gene function based upon evolutionary relationships, and to classify proteins for high-through put analysis (Mi et al., 2010). PANTHER is financially supported through research grants from the National Human Genome Research Institute and the National Science Foundation and it is maintained at the University of Southern California (the Thomas lab). The latest version (13.1) dated 03/02/2018 contains 15524 protein families, divided into 79562 functionally distinct protein subfamilies (http://www.pantherdb.org/about.jsp).

Throughout the continued development of PANTHER, the primary function remains to be the accurate inference of gene and/or protein function from sizeable online sequence databases, using e.g. phylogenetic trees to further extrapolate from the minimal experimental information

generated in a limited range of model organisms (Mi et al., 2013). PANTHER is made up of two components- the PANTHER library and the PANTHER index. The library holds a collection of protein families in the form of a multiple sequence alignment, and a Hidden Markov Model (HMM) in addition to a family tree. The PANTHER index covers ontology and subsequent molecular functions and biological processes associated with the families (or sub-families) (Thomas et al., 2003).

DAVID

The Database for Annotation, Visualisation and Integrated Discovery (DAVID) is an online bioinformatics resource with built in tools used to functionally analyse large genome datasets particularly those derived from microarray analyses such as methylation or transcriptional array data. The database is a freely available online interface maintained by the National Institute of Allergy and Infectious Disease (NIAD), part of the National Institute of Health (NIH) and can quickly and comprehensively analyse sizeable uploaded gene lists. Thus, DAVID offers assistance in the interpretation of genome wide array derived datasets that can help bridge the gap between arbitrary gene lists to some translational biological meaning where themes can be derived. The processed data is presented in the forms of exploratory visualisation through functional classification, biochemical pathway maps, and conserved protein domain architectures. DAVID is updated weekly through an automated VBA procedure from multiple sources that come from the annotation data shown in Table 4.

Resource	URL
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html
UniGene	http://www.ncbi.nlm.nih.gov/UniGene/
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/
LocusLink	http://www.ncbi.nlm.nih.gov/LocusLink/
KEGG	http://www.genome.ad.jp/kegg/
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Gene Ontology	http://www.geneontology.org/
University of Michigan	http://dot.ped.med.umich.edu:2000/ourimage/pub/shared/JMR_pub_affyannot.html
NetAffx	http://www.affymetrix.com/analysis/index.affx

Table 4 Annotation resources used complied into David. DAVID is updated on a weekly basis using Visual Basic (VB) to call a set of Perl and Java applications that are capable of downloading public data through a process of file transfer protocols. These are unpacked with any relevant annotation data; from here tab-delimited files are created for database import onto a regional database management system. Table 4 represents a list of the key resources from which data is imported (Dennis et al., 2003).

DAVID is composed of four main modules

1. Annotation Tool,
2. Go Charts,
3. Kegg Charts, and
4. Domain Charts.

The Annotation Tool is an automated method for the functional annotation of gene lists. Any combination of annotation data can be chosen from options including GenBank, Unigene, LousLink, Refseq and others. The annotations are added to the user defined submitted gene list to return a HTML table containing the user's original list of identifiers and the chosen functional annotations. The next module GoCharts is responsible for a graphical display of the distribution of differentially expressed and/or methylated genes among functional categories using the controlled vocabulary of the Gene Ontology Consortium (GO)

The third module Kegg Charts display the spread of differentially expressed and/or methylated genes among KEGG biochemical pathways in the form of graphs. Each pathway is linked to a KEGG pathway map, wherein differentially expressed/methylated genes from the original list are highlighted in red to allow individual gene delineation. Finally, the DomainCharts module displays the distribution of differentially expressed genes among PFAM protein domains

(Dennis et al., 2003). Each domain designation is linked to the Conserved Domain Database (CDD) of the National Centre for Biotechnology Information (NCBI).

Concluding Remarks

DNA methyltransferases play a critical role in the regulation of gene expression at key development stages and continuously during mammalian life. DNA methylation interacts with additional epigenetic modifiers such as the polycomb and the histone proteins to do this. A number of gene classes are regulated by DNA methylation and are especially sensitive to global methylation changes, namely the imprints and those on the X chromosome. In this thesis I describe global demethylation changes caused by transient and stable reductions of DNA methyltransferases in human and mouse model systems and how we identified gene classes that are particularly sensitive to such methylation events.

Aims of the Study

Through the use of both bioinformatic and wet-lab experimental approaches in human and/or mouse model systems, I aim to;

1) Identify whether a range of known and putative imprinted regions in mESCs are capabale of recovery of methylation once lost,

2) Determine whether the *de novo* methyltransferases Dnmt3a/3b have a role in the maintenance of methylation in the mESCs,

3) Identify if there are any enriched loci in the human genome that are particulary sensitive to a stable loss of maintenance methylation and whether this correlates with any other mark,

4) Determine if there is any difference in the response of non-transformed hTERTs cells to a transient loss of methylation through DNMT1 targeting siRNA or treated with 5-aza-2'dC.

References

Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., *et al.* (2000) Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family, *Genomics,* 65 (3): 293-298.

Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., *et al.* (2012) BLUEPRINT to decode the epigenetic signature written in blood, *Nature Biotechnology,* 30 (3): 224-226.

Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., *et al.* (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens, *Nucleic Acids Research,* 37 (Database issue), D816-9.

An, J., Rao, A. & Ko, M. (2017) TET family dioxygenases and DNA demethylation in stem cells and cancers, *Experimental & Molecular Medicine,* 49 (4), e323.

Anderson, M.W. & Schrijver, I. (2010) Next generation DNA sequencing and the future of genomic medicine, *Genes,* 1 (1): 38-69.

Annunziato, A.T. (2013) Assembling chromatin: the long and winding road, *Biochimica Et Biophysica Acta,* 1819 (3-4): 196-210.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., *et al.* (2012) In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases, *PLoS Genetics,* 8 (6), e1002750.

Aranda, S., Mas, G. & Di Croce, L. (2015) Regulation of gene transcription by Polycomb proteins. *Science Advances,* 1 (11), e1500737.

Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., *et al.* (2012) Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1, *Proceedings of the National Academy of Sciences of the United States of America,* 109 (32): 12950-12955.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D. & Irizarry, R.A. (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays, *Bioinformatics (Oxford, England),* 30 (10): 1363-1369.

Assenov, Y., Muller, F., Lutsik, P., Walter, J., Lengauer, T. & Bock, C. (2014) Comprehensive analysis of DNA methylation data with RnBeads, *Nature Methods,* 11 (11): 1138-1140.

Ausio, J., Dong, F. & van Holde, K.E. (1989) Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome, *Journal of Molecular Biology,* 206 (3): 451-463.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., *et al.* (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, *Nature,* 455 (7214): 822-825.

Barbarino, J.M., Haidar, C.E., Klein, T.E. & Altman, R.B. (2014) PharmGKB summary: very important pharmacogene information for UGT1A1, *Pharmacogenetics and Genomics,* 24 (3): 177-183.

Barrera, V. & Peinado, M.A. (2012) Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale, *Nucleic Acids Research,* 40 (22): 11490-11498.

Bartlett, M.G. & Gourley, G.R. (2011) Assessment of UGT polymorphisms and neonatal jaundice, *Semin. Perinatol.* 35 (3): 127-133.

Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S. & Jeltsch, A. (2014) The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism, *The Journal of Biological Chemistry,* 289 (7): 4106-4115.

Baylin, S.B. & Jones, P.A. (2011) A decade of exploring the cancer epigenome - biological and translational implications, *Nature Reviews.Cancer,* 11 (10): 726-734.

Baylin, S.B., Herman, J., G, Graff, J.R., Vertino, P.M. & Issa, J. (1997) Alterations in DNA Methylation: A Fundamental Aspect of Neoplasia, *Advances in Cancer Research,* 72, 141-182-182a, 182b, 183-196.

Beard, C., E.Li & Jaenisch, R. (1995) Loss of methylation activates Xist in somatic but not in embryonic cells, *Genes and Development,* 9: 2325-2334.

Belanger, A., Tojcic, J., Harvey, M. & Guillemette, C. (2010) Regulation of UGT1A and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells, *BMC Molecular Biology,* 11 (9): 1-11.

Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases, *Journal of Molecular Biology,* 203 (4): 971-983.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., *et al.* (2011) High density DNA methylation array with single CpG site resolution, *Genomics,* 98 (4): 288-295.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., *et al.* (2010) Targets and dynamics of promoter DNA methylation during early mouse development, *Nature Genetics,* 42 (12): 1093-1100.

Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., *et al.* (1991) Characterization of a murine gene expressed from the inactive X chromosome, *Nature,* 351 (6324): 325-329.

Bostick, M., Kyong Kim, J., Esteve, P., Clark, A. & Pradhan, S. (2007) UHRFI plays a role in maintaining DNA methylation in mammalian cells, *Science,* 317: 1760-1764.

Boulard, M., Edwards, J.R. & Bestor, T.H. (2015) FBXL10 protects Polycomb-bound genes from hypermethylation, *Nature Genetics,* 47 (5): 479-485.

Bourc'his, D. & Bestor, T.H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L, *Nature,* 431 (7004): 96-99.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. & Bestor, T.H. (2001) Dnmt3L and the establishment of maternal genomic imprints, *Science (New York, N.Y.),* 294 (5551): 2536-2539.

Branco, M.R., Ficz, G. & Reik, W. (2011) Uncovering the role of 5-hydroxymethylcytosine in the epigenome, *Nature Reviews.Genetics,* 13 (1): 7-13.

Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S.R. (2003) Sequence information can be obtained from single DNA molecules, *Proceedings of the National Academy of Sciences of the United States of America,* 100 (7): 3960-3964.

Brunetti, L., Gundry, M.C. & Goodell, M.A. (2017) DNMT3A in Leukemia, *Cold Spring Harbor Perspectives in Medicine,* 7 (2), 10.1101/cshperspect.a030320.

Brzezinski, J., Shuman, C., Choufani, S., Ray, P., Stavropoulos, D.J., Basran, R., *et al.* (2017) Wilms tumour in Beckwidth-Wiedemann Syndrome and loss of methylation at imprinting centre 2:revisiting tumour surveillance guidelines, *European Journal of Human Genetics: EJHG,* 25: 1031-1039.

Cancer Genome Atlas Research Network, Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., *et al.* (2013) Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia, *The New England Journal of Medicine,* 368 (22): 2059-2074.

Cedar, H. & Bergman, Y. (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews.Genetics,* 10 (5): 295-304.

Cheah, M.S., Wallace, C.D. & Hoffman, R.M. (1984) Hypomethylation of DNA in human cancer cells: a site-specific change in the c-myc oncogene. *Journal of the National Cancer Institute,* 73 (5): 1057-1065.

Chedin, F., Lieber, M.R. and Hsieh, C.L. (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a, *Proceedings of the National Academy of Sciences of the United States of America,* 99 (26), 16916-16921.

Chen, Zhao-Xia., Mann, J. R., Hsieh, C.H., Diggs, A.D. & Chedin, F. (2005) Physical and functional interactions between the human DNMT3L protein and members of the *de novo* methyltransferase family, *Journal of Cellular Biochemistry,* 95 (5): 902-917.

Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L. & Jaenisch, R. (1998) DNA hypomethylation leads to elevated mutation rates, *Nature,* 395 (6697): 89-93.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y. & Li, E. (2007) Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells, *Nature Genetics,* 39 (3): 391-396.

Chen, T., Tsujimoto, N. & Li, E. (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin, *Molecular and Cellular Biology,* 24 (20): 9048-9058.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. (2003a) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b, *Molecular and Cellular Biology,* 23 (16): 5594-5605.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. (2003b) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b, *Molecular and Cellular Biology,* 23 (16): 5594-5605.

Chen, W.V., Alvarez, F.J., Lefebvre, J.L., Friedman, B., Nwakeze, C., Geiman, E., *et al.* (2012) Functional significance of isoform diversification in the protocadherin gamma gene cluster. *Neuron,* 75 (3): 402-409.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., *et al.* (2015) Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses, *Cell,* 162 (5): 974-986.

Chomez, P., De Backer, O., Bertrand, M., De Plaen, E., Boon, T. & Lucas, S. (2001) An overview of the MAGE gene family with the identification of all human members of the family, *Cancer Research,* 61 (14): 5544-5551.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G. & Li, B.F. (1997) Human DNA-(cytosine5) methyltransferase-PCNA complex as a target for p21WAF1, *Science,* 277: 1996-2000.

Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P. & Bonapace, I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity, *Molecular and Cellular Biology,* 24 (6): 2526-2535.

Clapier, C.R. & Cairns, B.R. (2009) The biology of chromatin remodeling complexes, *Annual Review of Biochemistry,* 78: 273-304.

Cosgrove, M.S., Boeke, J.D. & Wolberger, C. (2004) Regulated nucleosome mobility and the histone code, *Nature Structural & Molecular Biology,* 11 (11): 1037-1043.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., *et al.* (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns, *Nature Genetics,* 24 (2): 132-138.

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., *et al.* (2014) Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of

human imprinting and suggests a germline methylation-independent mechanism of establishment, *Genome Research,* 24 (4): 554-569.

Criscione, S.W., Zhang, Y., Thompson, W., Sedivy, J.M. & Neretti, N. (2014) Transcriptional landscape of repetitive elements in normal and cancer human cells, *BMC Genomics,* 15, 583-2164-15-583.

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., *et al.* (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk, *Science (New York, N.Y.),* 299 (5613): 1753-1755.

Cunningham, D., Maroun, J., Vanhoefer, U. & Van Cutsem, E. (2001) Optimizing the use of irinotecan in colorectal cancer, *The Oncologist,* 6 Suppl 4: 17-23.

Davidson, S., Crowther, P., Radley, J. & Woodcock, D. (1992) Cytotoxicity of 5-aza-2'-deoxycytidine in a mammalian cell system, *European Journal of Cancer,* 28 (2): 362-368.

Davis, S., Du, P., Bilke, S., Triche, J.T. & Bootwalla, M. (2007) *methylumi: Handle Illumina methylation data.* . 3.7th ed.Bioconductor.Available from: https://www.bioconductor.org/packages/release/bioc/html/methylumi.html [Accessed 31/07/2018].

De Smet, C., Loriot, A. & Boon, T. (2004) Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells, *Molecular and Cellular Biology,* 24 (11): 4781-4790.

Deaton, A.M. & Bird, A. (2011) CpG islands and the regulation of transcription, *Genes & Development,* 25 (10), 1010-1022.

Dennis, G.,Jr, Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. & Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery, *Genome Biology,* 4 (5), P3.

Di Croce, L. & Helin, K. (2013) Transcriptional regulation by Polycomb group proteins, *Nature Structural & Molecular Biology,* 20 (10): 1147-1155.

Diaz-Meyer, N., Day, C.D., Khatod, K., Maher, E.R., Cooper, W., Reik, W., *et al.* (2003) Silencing of CDKN1C (p57KIP2) is associated with hypomethylation at KvDMR1 in Beckwith-Wiedemann syndrome, *Journal of Medical Genetics,* 40 (11): 797-801.

Docherty, L.E., Rezwan, F.I., Poole, R.L., Jagoe, H., Lake, H., Lockett, G.A., *et al.* (2014) Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes, *Journal of Medical Genetics,* 51 (4): 229-238.

Easwaran, H.P., Schermelleh, L., Leonhardt, H. & Cardoso, M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases, *EMBO Reports,* 5 (12): 1181-1186.

Edwards, J.R., Yarychkivska, O., Boulard, M. & Bestor, T.H. (2017) DNA methylation and DNA methyltransferases, *Epigenetics & Chromatin,* 10, 23-017-0130-8. eCollection 2017.

Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little, *Oncogene*, 21 (35): 5400-5413.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A. *et al.* (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells, *Nucleic Acids Research,* 10 (8): 2709-2721.

Ehrlich, M., Jackson, K. & Weemaes, C. (2006) *ICF Syndrome.* The portal for rare diseases and orphan drugs. Available from: https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=GB&Expert=2268 [Accessed 06/08/2018].

El Hajj, N., Dittrich, M. and Haaf, T. (2017) Epigenetic dysregulation of protocadherins in human disease, *Seminars in Cell & Developmental Biology,* 69: 172-182.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome, *Nature,* 489 (7414): 57-74.

Esteller, M., Corn, P.G., Baylin, S.B. & Herman, J.G. (2001) A gene hypermethylation profile of human cancer, *Cancer Research,* 61 (8): 3225-3229.

Etoh, T., Kanai, Y., Ushijima, S., Nakagawa, T., Nakanishi, Y., Sasako, M., *et al.* (2004) Increased DNA methyltransferase 1 (DNMT1) protein expression correlates significantly with poorer tumor differentiation and frequent DNA hypermethylation of multiple CpG islands in gastric cancers, *The American Journal of Pathology,* 164 (2): 689-699.

Feinberg, A.P., Gehrke, C.W., Kuo, K.C. & Ehrlich, M. (1988) Reduced genomic 5-methylcytosine content in human colonic neoplasia, *Cancer Research,* 48 (5): 1159-1161.

Feinberg, A.P. & Vogelstein, B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts, *Nature,* 301 (5895): 89-92.

Felle, M., Joppien, S., Nemeth, A., Diermeier, S., Thalhammer, V., Dobner, T., *et al.* (2011) The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Research,* 39 (19), 8355-8365.

Filippakopoulos, P. & Knapp, S. (2012) The bromodomain interaction module, *FEBS Letters,* 586 (17): 2692-2704.

Flatau, E., Gonzales, F.A., Michalowsky, L.A. & Jones, P.A. (1984) DNA methylation in 5-aza-2'-deoxycytidine-resistant variants of C3H 10T1/2 C18 cells, *Molecular and Cellular Biology,* 4 (10): 2098-2102.

Fortin, J.P., Triche, T.J.,Jr & Hansen, K.D. (2017) Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi, *Bioinformatics (Oxford, England),* 33 (4): 558-560.

Frank, M., Ebert, M., Shan, W., Phillips, G.R., Arndt, K., Colman, D.R. & Kemler, R. (2005) Differential expression of individual gamma-protocadherins during mouse brain development, *Molecular and Cellular Neurosciences,* 29 (4): 603-616.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., *et al.* (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1, *PloS One,* 6 (2), e16627.

Gagne, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G. & Guillemette, C. (2002) Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38), *Molecular Pharmacology,* 62 (3): 608-617.

Gagnon, J.F., Bernard, O., Villeneuve, L., Tetu, B. & Guillemette, C. (2006) Irinotecan inactivation is modulated by epigenetic silencing of UGT1A1 in colon cancer, *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research,* 12 (6): 1850-1858.

Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W. & Ehrlich, M. (1983) The 5-methylcytosine content of DNA from human tumors, *Nucleic Acids Research,* 11 (19): 6883-6894.

Gardiner-Garden, M. & Frommer, M. (1987) CpG Islands in vertebrate genomes, *Journal of Molecular Biology,* 196: 261-282.

Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A. & Xu, G.L. (2004) Chromatin targeting of de novo DNA methyltransferases by the PWWP domain, *The Journal of Biological Chemistry,* 279 (24): 25447-25454.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics, *Genome Biology,* 5 (10), R80-2004-5-10-r80. Epub 2004 Sep 15.

Ghoshal, K., Datta, J., Majumder, S., Bai, S., Kutay, H., Motiwala, T., *et al.* (2005) 5-Aza-Deoxycytidine Induces Selective Degradation of DNA Methyltransferase 1 by a Proteasomal Pathway That Requires the KEN Box, Bromo-Adjacent Homology Domain, and Nuclear Localization Signal, *Molecular and Cellular Biology,* 25 (11): 4727.

Gibney, E.R. & Nolan, C.M. (2010) Epigenetics and gene expression, *Heredity,* 105 (1): 4-13.

Giglia-Mari, G., Zotter, A. & Vermeulen, W. (2011) DNA damage response, *Cold Spring Harbor Perspectives in Biology,* 3 (1), a000745.

Gjerstorff, M.F., Andersen, M.H. & Ditzel, H.J. (2015) Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget,* 6 (18): 15772-15787.

Gjerstorff, M.F. & Ditzel, H.J. (2012) Limited SP17 expression within tumors diminishes its therapeutic potential. *Tissue Antigens,* 80 (6): 523-527.

Goodier, J.L. (2016) Restricting retrotransposons: a review, *Mobile DNA*, 16-016-0070-z. eCollection 2016.

Greer, E.L. & Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance, *Nature Reviews.Genetics*, 13 (5), 343-357.

Guibert, S., Forne, T. & Weber, M. (2012) Global profiling of DNA methylation erasure in mouse primordial germ cells, *Genome Research*, 22 (4): 633-641.

Gunster, M.J., Raaphorst, F.M., Hamer, K.M., den Blaauwen, J.L., Fieret, E., Meijer., *et al.* (2001) Differential expression of human Polycomb group proteins in various tissues and cell types, *Journal of Cellular Biochemistry.Supplement*, Suppl 36: 129-143.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., *et al*. (2014) The DNA methylation landscape of human early embryos, *Nature*, 511 (7511): 606-610.

Habano, W., Kawamura, K., Iizuka, N., Terashima, J., Sugai, T. & Ozawa, S. (2015) Analysis of DNA methylation landscape reveals the roles of DNA methylation in the regulation of drug metabolizing enzymes, *Clinical Epigenetics*, 7, 105-015-0136-7. eCollection 2015.

Habano, W., Gamo, T., Sugai, T., Otsuka, K., Wakabayashi, G. & Ozawa, S. (2009) CYP1B1 but not CYP1A1 is downregulated by promoter methylation in colorectal cancers, *J Oncology*, 34 , 1085-1091.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A. & Surani, M.A. (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine, *Science (New York, N.Y.)*, 339 (6118): 448-452.

Hackett, J.A., Zylicz, J.J. & Surani, M.A. (2012) Parallel mechanisms of epigenetic reprogramming in the germline, *Trends in Genetic*, 28 (4): 164-174.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., *et al.* (2002) Epigenetic reprogramming in mouse primordial germ cells, *Mechanisms of Development*, 117 (1-2): 15-23.

Hancks, D.C. & Kazazian, H.H.,Jr (2012) Active human retrotransposons: variation and disease, *Current Opinion in Genetics & Development*, 22 (3): 191-203.

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M. *et al.* (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome, *Proceedings of the National Academy of Sciences of the United States of America*, 96 (25): 14412-14417.

Harshman, S.W., Young, N.L., Parthun, M.R. & Freitas, M.A. (2013) H1 histones: current perspectives and challenges, *Nucleic Acids Research*, 41 (21): 9593-9609.

Hashimoto, H., Horton, J.R., Zhang, X., Bostick, M., Jacobsen, S.E. & Cheng, X. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix, *Nature*, 455 (7214): 826-829.

Hata, K., Okano, M., Lei, H. & Li, E. (2002) Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice, *Development (Cambridge, England),* 129 (8): 1983-1993.

Heather, J.M. & Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA, *Genomics,* 107 (1): 1-8.

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., *et al.* (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma, *Proceedings of the National Academy of Sciences of the United States of America,* 95 (12): 6870-6875.

Hill, P.W., Amouroux, R. & Hajkova, P. (2014) DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story, *Genomics,* 104 (5): 324-333.

Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., *et al.* (2018) Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature,* 555 (7696): 392-396.

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

Holm, T.M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout, W.M., *et al*. (2005) Global loss of imprinting leads to widespread tumorigenesis in adult mice, *Cancer Cell,* 8 (4), 275-285.

Hu, J., Zhou, R., Zhang, R., Zhang, K., Zhou, J. & Xu, G. (2009) The N-terminus of histone H3 is required for de novo DNA methylation in chromatin, *Proceedings of the National Academy of Sciences of the United States of America,* 106 (52): 22187-22192.

Hulpiau, P. & van Roy, F. (2011) New insights into the evolution of metazoan cadherins, *Molecular Biology and Evolution,* 28 (1), 647-657.

Illumina, I. (2015) *Infinium MethylationEPIC BeadChip Data Sheet.* Illumina Online Content.: Illumina. Available from: https://emea.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/humanmethylationepic-data-sheet-1070-2015-008.pdf [Accessed 25/07/2018].

Illumina, I. (2011) *GenomeStudio/ BeadStudio Software Methylation Module.* Illumina Online Content: Ilumina. Available from: https://www.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/genomestudio/genomestudio-2011-1/genomestudio-methylation-v1-8-user-guide-11319130-b.pdf [Accessed 25/07/2018].

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., *et al*. (2008) Comprehensive high-throughput arrays for relative methylation (CHARM), *Genome Research,* 18 (5): 780-790.

Irwin, R., Thakur, A., O'Neill, K.M. & Walsh, C.P. (2014) 5-Hydroxymethylation marks a class of neuronal gene regulated by intragenic methylcytosine levels, *Genomics,* 104 (5): 383-392.

Isbel, L., Prokopuk, L., Wu, H., Daxinger, L., Oey, H., Spurling, A., *et al.* (2016) Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse, *eLife,* 5 , 10.7554/eLife.15082.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., *et al*. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science (New York, N.Y.),* 333 (6047): 1300-1303.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E., Bachman, M., Booth, M.J., *et al*. (2013) A screen for hydroxymethylcytosine and formyl cytosine binding proteins suggests functions in transcription and chromatin regulation, *Genome Biology,* 14, R119.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., *et al.* (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation, *Nature Genetics,* 27 (1): 31-39.

Jang, H.S., Shin, W.J., Lee, J.E. & Do, J.T. (2017) CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes,* 8 (6), 10.3390/genes8060148.

Jeziorska, D.M., Murray, R.J.S., De Gobbi, M., Gaentzsch, R., Garrick, D., Ayyub, H., *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,* 114 (36), E7526-E7535.

Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond, *Nature Reviews.Genetics,* 13 (7): 484-492.

Jones, P.A. (1985) Altering gene expression with 5-azacytidine, *Cell,* 40 (3): 485-486.

Jones, P.A. & Taylor, S.M. (1980) Cellular differentiation, cytidine analogs and DNA methylation, *Cell,* 20 (1): 85-93.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., *et al.* (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription, *Nature Genetics,* 19 (2): 187-191.

Ju, J., Kim, D.H., Bi, L., Meng, Q., Bai, X., Li, Z., *et al.* (2006) Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators, *Proceedings of the National Academy of Sciences of the United States of America,* 103 (52): 19635-19640.

Jurgens, G. (1985) A group of genes controlling the spatial expression of the bithorax complex in Drosophila, *Nature,* 316: 153-155.

Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J. & Crabtree, G.R. (2013) Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy, *Nature Genetics,* 45 (6): 592-601.

Kamikihara, T., Arima, T., Kato, K., Matsuda, T., Kato, H., Douchi, T., *et al.* (2005) Epigenetic silencing of the imprinted gene ZAC by DNA methylation is an early event in the progression of human ovarian cancer. *International Journal of Cancer,* 115 (5): 690-700.

Kaminskas, E., Farrell, A.T., Wang, Y.C., Sridhara, R. & Pazdur, R. (2005) FDA drug approval summary: azacitidine (5-azacytidine, Vidaza) for injectable suspension, *The Oncologist,* 10 (3): 176-182.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E. & Sasaki, H. (2004) Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting, *Nature,* 429 (6994): 900-903.

Karpf, A.R., Lasek, A.W., Ririe, T.O., Hanks, A.N., Grossman, D. & Jones, D.A. (2004) Limited gene activation in tumor and normal epithelial cells treated with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine, *Molecular Pharmacology,* 65 (1): 18-27.

Karpf, A.R., Moore, B.C., Ririe, T.O. & Jones, D.A. (2001) Activation of the p53 DNA damage response pathway after inhibition of DNA methyltransferase by 5-aza-2'-deoxycytidine, *Molecular Pharmacology,* 59 (4): 751-757.

Kawaguchi, M., Toyama, T., Kaneko, R., Hirayama, T., Kawamura, Y. & Yagi, T. (2008) Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster, *The Journal of Biological Chemistry,* 283 (18):12064-12075.

Kebede, A.F., Schneider, R. & Daujat, S. (2015) Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *The FEBS Journal,* 282 (9): 1658-1674.

Kelsey, G. & Feil, R. (2013) New insights into establishment and maintenance of DNA methylation imprints in mammals, *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences,* 368 (1609), 20110336.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., *et al.* (2002) The human genome browser at UCSC, *Genome Research,* 12 (6): 996-1006.

Kishikawa, S., Murata, T., Ugai, H., Yamazaki, T. & Yokoyama, K.K. (2003) Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner, *Nucleic Acids Research.Supplement (2001),* (3): 307-308.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., *et al.* (2011) Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice, *Proceedings of the National Academy of Sciences of the United States of America,* 108 (35): 14566-14571.

Kohli, R.M. & Zhang, Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation, *Nature,* 502 (7472): 472-479.

Kong, L., Tan, L., Lv, R., Shi, Z., Xiong, L., Wu, F., *et al.* (2016) A primary role of TET proteins in establishment and maintenance of *De Novo* bivalency at CpG islands, *Nucleic Acids Research,* 44 (18): 8682-8692.

Kornberg, R.D. & Lorch, Y. (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome, *Cell,* 98 (3), 285-294.

Koslowski, M., Bell, C., Seitz, G., Lehr, H., Roemer, K., Munteferling, H., *et al.* (2004) Frequent Nonrandom Activation of Germ-Line Genes in Human Cancer, *Cancer Res.,* 64: 5988-5993.

Kuo, H.K., Griffith, J.D. & Kreuzer, K.N. (2007) 5-Azacytidine induced methyltransferase-DNA adducts block DNA replication in vivo, *Cancer Research,* 67 (17), 8248-8254.

Laduron, S., Deplus, R., Zhou, S., Kholmanskikh, O., Godelaine, D., De Smet, C., *et al.* (2004) MAGE-A1 interacts with adaptor SKIP and the deacetylase HDAC1 to repress transcription, *Nucleic Acids Research,* 32 (14): 4340-4350.

Lallous, N., Legrand, P., McEwen, A.G., Ramon-Maiques, S., Samama, J.P. & Birck, C. (2011) The PHD finger of human UHRF1 reveals a new subgroup of unmethylated histone H3 tail readers, *PloS One,* 6 (11), e27599.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., *et al.* (2001) Initial sequencing and analysis of the human genome, *Nature,* 409 (6822): 860-921.

Langst, G. & Manelyte, L. (2015) Chromatin Remodelers: From Function to Dysfunction, *Genes,* 6 (2): 299-324.

Lawrence, M., Daujat, S. & Schneider, R. (2016) Lateral Thinking: How Histone Modifications Regulate Gene Expression, *Trends in Genetics,* 32 (1): 42-56.

Lees-Murdock, D.J., De Felici, M. & Walsh, C.P. (2003) Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage, *Genomics,* 82 (2): 230-237.

Leonhardt, H., Page, A.W., Weier, H.U. & Bestor, T.H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei, *Cell,* 71 (5): 865-873.

Levin, H.L. & Moran, J.V. (2011) Dynamic interactions between transposable elements and their hosts, *Nature Reviews.Genetics,* 12 (9): 615-627.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., *et al.* (2010) DNMT3A mutations in acute myeloid leukemia, *The New England Journal of Medicine,* 363 (25), 2424-2433.

Li, D., Zhang, B., Xing, X. & Wang, T. (2015) Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation, *Methods (San Diego, Calif.),* 72: 29-40.

Li, E., Bestor, T.H. & Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, *Cell,* 69 (6): 915-926.

Li, G. & Reinberg, D. (2011) Chromatin higher-order structures and gene regulation, *Current Opinion in Genetics & Development,* 21 (2): 175-186.

Li, J.Y., Lees-Murdock, D.J., Xu, G.L. & Walsh, C.P. (2004) Timing of establishment of paternal methylation imprints in the mouse, *Genomics,* 84 (6): 952-960.

Li, X., Qin, B. & Liu, B.O. (2016) Delineating the effect of demethylating agent 5-aza-2'-deoxycytidine on human Caco-2 colonic carcinoma cells, *Oncology Letters,* 12 (1): 139-143.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S. & Jacobsen, S.E. (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation, *Science (New York, N.Y.),* 292 (5524): 2077-2080.

Lindroth, A.M., Shultis, D. & Jasencakova, Z. (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with chromomethylase. *Embo J.,* 23: 4146-4155.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., *et al. (*2013) Global epigenomic reconfiguration during mammalian brain development, *Science (New York, N.Y.),* 341 (6146), 1237905.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., *et al.* (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells, *Nature,* 471 (7336): 68-73.

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., *et al*. (2013) UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature Communications,* 4: 1563-1565.

Liu, Y., An, S., Ward, R., Yang, Y., Guo, X.X., Li, W., *et al.* (2016) G protein-coupled receptors as promising cancer targets, *Cancer Letters,* 376 (2): 226-239.

Loughery, J.E., Dunne, P.D., O'Neill, K.M., Meehan, R.R., McDaid, J.R. & Walsh, C.P. (2011) DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response, *Human Molecular Genetics,* 20 (16): 3241-3255.

Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M. & Luger, K. (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure, *Nature Structural & Molecular Biology,* 15 (10): 1122-1124.

Luger, K., Dechassa, M.L. & Tremethick, D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Reviews.Molecular Cell Biology,* 13: 436-447.

Lyon, M.F. (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.), *Nature,* 190: 372-373.

Maatouk, D.M., Kellam, L.D., Mann, M.R., Lei, H., Li, E., Bartolomei, M.S. & Resnick, J.L. (2006) DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages, *Development (Cambridge, England),* 133 (17): 3411-3418.

Maksimovic, J., Gordon, L. & Oshlack, A. (2012) SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips, *Genome Biology,* 13 (6), R44-2012-13-6-r44.

Margueron, R., Li, G., Sarma, K., Blais, A., Zavadil, J., Woodcock, C.L., *et al.* (2008) Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms, *Molecular Cell,* 32 (4): 503-518.

Mayes, K., Qiu, Z., Alhazmi, A. & Landry, J.W. (2014) Chapter Five- ATP-Dependent Chromatin Remodelling Complexes as Novel Targets for Cancer Therapy, *Advances in Cancer Research.,* 121: 183-233.

McCarthy, E.M. & McDonald, J.F. (2004) Long terminal repeat retrotransposons of Mus musculus, *Genome Biology,* 5 (3), R14-2004-5-3-r14. Epub 2004 Feb 13.

Meissner, A. (2011) Guiding DNA methylation, *Cell Stem Cell,* 9 (5): 388-390.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., *et al*. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells, *Nature,* 454 (7205): 766-770.

Messerschmidt, D.M., Knowles, B.B. & Solter, D. (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos, *Genes & Development,* 28 (8): 812-828.

Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S. & Thomas, P.D. (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium, *Nucleic Acids Research,* 38 (Database issue), D204-10.

Mi, H., Muruganujan, A. & Thomas, P.D. (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research,* 41 (Database issue), D377-86.

Mitra, R.D., Shendure, J., Olejnik, J., Edyta-Krzymanska-Olejnik & Church, G.M. (2003) Fluorescent in situ sequencing on polymerase colonies. *Analytical Biochemistry,* 320 (1): 55-65.

Mizuno, S., Chijiwa, T., Okamura, T., Akashi, K., Fukumaki, Y., Niho, Y.& Sasaki, H. (2001) Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia, *Blood,* 97 (5):1172-1179.

Moarefi, A.H. & Chedin, F. (2011) ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated *de novo* DNA methylation, *J Mol Bio*, 409 (5): 758-772.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., *et al.* (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors, *Molecular Cell*, 30 (6): 755-766.

Momparler, R.L., Rossi, M., Bouchard, J., Vaccaro, C., Momparler, L.F. & Bartolucci, S. (1984) Kinetic interaction of 5-AZA-2'-deoxycytidine-5'-monophosphate and its 5'-triphosphate with deoxycytidylate deaminase, *Molecular Pharmacology*, 25 (3): 436-440.

Monk, D. (2010) Deciphering the cancer imprintome, *Briefings in Functional Genomics*, 9 (4): 329-339.

Moore, L.D., Le, T. & Fan, G. (2013) DNA methylation and its basic function, *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 38 (1): 23-38.

Morishita, H. & Yagi, T. (2007) Protocadherin family: diversity, structure, and function. *Current Opinion in Cell Biology*, 19 (5): 584-592.

Morris, T.J. & Beck, S. (2015) Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data, *Methods (San Diego, Calif.)*, 72: 3-8.

Mummert, S.K., Lobanenkov, V.A. & Feinberg, A.P. (2005) Association of chromosome arm 16q loss with loss of imprinting of insulin-like growth factor-II in Wilms tumor, *Genes, Chromosomes & Cancer*, 43 (2): 155-161.

Nady, N., Lemak, A., Walker, J.R., Avvakumov, G.V., Kareta, M.S., Achour, M., *et al.* (2011) Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein, *The Journal of Biological Chemistry*, 286 (27): 24300-24311.

Nesterova, T.B., Popova, B.C., Cobb, B.S., Norton, S., Senner, C.E., Tang, Y.A., *et al.* (2008) Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a, *Epigenetics & Chromatin*, 1 (1), 2-8935-1-2.

Ng, K., Pullirsch, D., Leeb, M. & Wutz, A. (2007) Xist and the order of silencing, *EMBO Reports*, 8 (1): 34-39.

Nieto, M., Demolis, P., Behanzin, E., Moreau, A., Hudson, I., Flores, B., *et al.* (2016) The European Medicines Agency Review of Decitabine (Dacogen) for the Treatment of Adult Patients With Acute Myeloid Leukemia: Summary of the Scientific Assessment of the Committee for Medicinal Products for Human Use, *The Oncologist*, 21 (6): 692-700.

Nikolic, A., Volarevic, V., Armstrong, L., Lako, M. & Stojkovic, M. (2016) Primordial Germ Cells: Current Knowledge and Perspectives, *Stem Cells International*, 2016, 1741072.

Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., *et al.* (2013) Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients, *Journal of Human Genetics,* 58 (7): 455-460.

Ogawa, H., Ono, Y., Shimozawa, N., Sotomaru, Y., Katsuzawa, Y., Hiura, H., *et al.* (2003) Disruption of imprinting in cloned mouse fetuses from embryonic stem cells, *Reproduction (Cambridge, England),* 126 (4): 549-557.

Ogawa, O., Eccles, M.R., Szeto, J., McNoe, L.A., Yun, K., Maw, M.A., *et al.* (1993) Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour, *Nature,* 362 (6422): 749-751.

Okano, M., Bell, D.W., Haber, D.A. & Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development, *Cell,* 99 (3): 247-257.

Okano, M., Xie, S. & Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nature Genetics,* 19 (3): 219-220.

Oliver, K.R. & Greene, W.K. (2011) Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates, *Mobile DNA,* 2 (1), 8-8753-2-8.

O'Neill, K.M., Irwin, R.E., Mackin, S.J., Thursby, S.J., Thakur, A., Bertens, C., *et al.* (2018) Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression, *Epigenetics & Chromatin,* 11 (1), 12-018-0182-4.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., *et al*. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA, *Nature,* 448 (7154): 714-717.

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M. & Shirakawa, M. (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain, *EMBO Reports,* 10 (11): 1235-1241.

Panning, B. (2008) X-chromosome inactivation: the molecular basis of silencing, *Journal of Biology,* 7 (8), 30.

Panning, B. & Jaenisch, R. (1996) DNA hypomethylation can activate Xist expression and silence X-linked genes, *Genes & Development,* 10 (16): 1991-2002.

Papait, R., Pistore, C., Grazini, U., Babbio, F., Cogliati, S., Pecoraro, D., *et al.* (2008) The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin, *Molecular Biology of the Cell,* 19 (8): 3554-3563.

Patel, K., Dickson, J., Din, S., Macleod, K., Jodrell, D. & Ramsahoye, B. (2010) Targeting of 5-aza-2'-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme, *Nucleic Acids Research,* 38 (13): 4313-4324.

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. (1972) The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid, *The Biochemical Journal*, 126 (4): 781-790.

Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J. & Schalkwyk, L.C. (2013) A data-driven approach to preprocessing Illumina 450K methylation array data, *BMC Genomics*, 14, 293-2164-14-293.

Qin, W., Leonhardt, H. & Spada, F. (2011) Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1, *Journal of Cellular Biochemistry*, 112 (2): 439-444.

Rahman, M.M., Qian, Z.R., Wang, E.L., Yoshimoto, K., Nakasono, M., Sultana, R., *et al.* (2010) DNA methyltransferases 1, 3a, and 3b overexpression and clinical significance in gastroenteropancreatic neuroendocrine tumors, *Human Pathology*, 41 (8): 1069-1078.

Randhawa, G.S., Cui, H., Barletta, J.A., Strichman-Almashanu, L.Z., Talpaz, M., Kantarjian, H., *et al.* (1998) Loss of imprinting in disease progression in chronic myelogenous leukemia. *Blood*, 91 (9): 3144-3147.

Rasmussen, K.D. & Helin, K. (2016) Role of TET enzymes in DNA methylation, development, and cancer, *Genes & Development*, 30 (7): 733-750.

Rebollo, R., Miceli-Royer, K., Zhang, Y., Farivar, S., Gagnier, L. & Mager, D.L. (2012) Epigenetic interplay between mouse endogenous retroviruses and host genes, *Genome Biology*, 13 (10), R89-2012-13-10-r89.

Rechache, N.S., Wang, Y., Stevenson, H.S., Killian, J.K., Edelman, D.C., Merino, M., *et al.* (2012) DNA methylation profiling identifies global methylation differences and markers of adrenocortical tumors, *The Journal of Clinical Endocrinology and Metabolism*, 97 (6), E1004-13.

Riggs, A. (1975) X inactivation, differentiation, and DNA methylation, *Cytogenetics and Cell Genetics*, 14 (1): 9-25.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P. (2011) Integrative genomics viewer, *Nature Biotechnology*, 29 (1): 24-26.

Roller, A., Grossmann, V., Bacher, U., Poetzinger, F., Weissmann, S., Nadarajah, N., *et al.* (2013) Landmark analysis of DNMT3A mutations in hematological malignancies, *Leukemia*, 27 (7): 1573-1578.

Ronaghi, M., Uhlen, M. & Nyren, P. (1998) A sequencing method based on real-time pyrophosphate. Science (New York, N.Y.), 281 (5375), 363, 365.

Rose, N.R. & Klose, R. J. (2014) Understanding the relationship between DNA methylation and histone lysine methylation, *Biochimica Et Biophysica Acta*, 1839 (12): 1362-1372.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., *et al.* (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation, *Nature Structural & Molecular Biology,* 19 (11): 1155-1160.

Rothbart, S.B. & Strahl, B.D. (2014) Interpreting the language of histone and DNA modifications, *Biochimica Et Biophysica Acta,* 1839 (8): 627-643.

Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., *et al.* (2015) DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts, *Cell,* 162 (5): 961-973.

Rowe, H.M., Friedli, M., Offner, S., Verp, S., Mesnard, D. *et al.* (2013) *De novo* DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET, *Development (Cambridge, England),* 140 (3): 519-529.

Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Mannepalli, S., *et al.* (2015) Molecular logic of neuronal self-recognition through protocadherin domain interactions, *Cell,* 163 (3): 629-642.

Rutledge, C.E., Thakur, A., O'Neill, K.M., Irwin, R.E., Sato, S., Hata, K. & Walsh, C.P. (2014) Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes, *Development (Cambridge, England),* 141 (6): 1313-1323.

Saito, Y., Kanai, Y., Sakamoto, M., Saito, H., Ishii, H. & Hirohashi, S. (2002) Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis, *Proceedings of the National Academy of Sciences of the United States of America,* 99 (15): 10060-10065.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M. & Esteller, M. (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome, *Epigenetics,* 6 (6), 692-702.

Sanger, F., Nicklen, S. & Coulson, A.R. (1992) DNA sequencing with chain-terminating inhibitors, 1977. *Biotechnology (Reading, Mass.),* 24: 104-108.

Santini, V., Kantarjian, H.M. & Issa, J.P. (2001) Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications, *Ann Intern Med,* 134: 573-586.

Satie, A.P., Rajpert-De Meyts, E., Spagnoli, G.C., Henno, S., Olivo, L., Jacobsen, G.K., *et al.* (2002) The cancer-testis gene, NY-ESO-1, is expressed in normal fetal and adult testes and in spermatocytic seminomas and testicular carcinoma in situ, *Laboratory Investigation; a Journal of Technical Methods and Pathology,* 82 (6): 775-780.

Saxonov, S., Berg, P. & Brutlag, D.L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters, *Proceedings of the National Academy of Sciences of the United States of America,* 103 (5): 1412-1417.

Schneider-Stock, R., Diab-Assef, M., Rohrbeck, A., Foltzer-Jourdainne, C., Boltze, C., Hartig, R., *et al.* (2005) 5-aza-cytidine is a potent inhibitor of DNA methyltransferase 3a and induces apoptosis in HCT-116 colon cancer cells via Gadd45-and p53-dependent mechanisms, *Journal of Pharmacology,* 312 (2): 525-536.

Schuettengruber, B. & Cavalli, G. (2009) Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice, *Development (Cambridge, England),* 136 (21): 3531-3542.

Seelan, R.S., Mukhopadhyay, P., Warner, D.R., Smolenkova, I.A., Pisano, M.M. & Greene, R.M. (2017) Determinants of orofacial clefting II: Effects of 5-Aza-2'-deoxycytidine on gene methylation during development of the first branchial arch, *Reproductive Toxicology (Elmsford, N.Y.),* 67: 100-110.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., *et al.* (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, *Nature,* 450 (7171): 908-912.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., *et al.* (2014) Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia, *Nature,* 506 (7488): 328-333.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R. & Peterson, C.L. (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions, *Science (New York, N.Y.),* 311 (5762): 844-847.

Sigalotti, L., Coral, S., Nardi, G., Spessotto, A., Cortini, E., Cattarossi, I., *et al.* (2002) Promoter methylation controls the expression of MAGE2, 3 and 4 genes in human cutaneous melanoma, *J Immunotherapy,* 25 (1): 16-26.

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T. & Old, L.J. (2005) Cancer/testis antigens, gametogenesis and cancer, *Nature Reviews.Cancer,* 5 (8): 615-625.

Slotkin, R.K. & Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome, *Nature Reviews.Genetics,* 8 (4): 272-285.

Smith, M.L., Baggerly, K.A., Bengtsson, H., Ritchie, M.E. & Hansen, K.D. (2013) illuminaio: An open source IDAT parsing tool for Illumina microarrays, *F1000Research,* 2, 264-264.v1. eCollection 2013.

Smith, Z.D. & Meissner, A. (2013) DNA methylation: roles in mammalian development, *Nature Reviews.Genetics,* 14 (3):204-220.

Sorm, F., Piskala, A., Cihak, A. & Vesely, J. (1964) 5-Azacytidine, a new, highly effective cancerostatic, *Experientia,* 20 (4): 202-203.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, *Cell,* 152 (5): 1146-1159.

Stewart, K.R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S., *et al.* (2015) Dynamic changes in histone modifications precede de novo DNA methylation in oocytes, *Genes & Development,* 29 (23): 2449-2462.

Stirzaker, C., Taberlay, P.C., Statham, A.L. & Clark, S.J. (2014) Mining cancer methylomes: prospects and challenges, *Trends in Genetic,* 30 (2): 75-84.

Stresemann, C. & Lyko, F. (2008) Modes of action of the DNA methyltransferase inhibitors azacytidine and Decitabine, *International Journal of Cancer,* 123 (1): 8-13.

Subramaniam, D., Thombre, R', Dhar, A., & Anant, S. (2014) DNA Methyltransferases: A Novel Target for Prevention and Therapy, *Frontiers of Oncology,* 4 (80).

Sutherland, E., Coe, L. & Raleigh, E.A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease, *Journal of Molecular Biology,* 225 (2): 327-348.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, *Science (New York, N.Y.),* 324 (5929): 930-935.

Taiwo, O., Wilson, G.A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., *et al.* (2012) Methylome analysis using MeDIP-seq with low DNA concentrations, *Nature Protocols,* 7 (4): 617-636.

Takeichi, M. (2007) The cadherin superfamily in neuronal connections and interactions, *Nature Reviews.Neuroscience,* 8 (1): 11-20.

Taniura, H., Kobayashi, M. & Yoshikawa, K. (2005) Functional domains of necdin for protein-protein interaction, nuclear matrix targeting, and cell growth suppression, *Journal of Cellular Biochemistry,* 94 (4): 804-815.

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., *et al.* (2014) Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability, *Nature Genetics,* 46 (4): 385-388.

Tatton-Brown, K., Zachariou, A., Loveday, C., Renwick, A., Mahamdallie, S., Aksglaede, L., *et al.* (2018) The Tatton-Brown-Rahman Syndrome: A clinical study of 55 individuals with de novo constitutive DNMT3A variants, *Wellcome Open Research,* 3: 46.

Teschendorff, A.E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D. & Beck S. (2013) A beta-mixture quantile normalisation method for correcting probe design bias in Illumina Infinium 450k DNA methylation data, *Bioinformatics,* 29 (2): 189-196.

Tessarz, P. & Kouzarides, T. (2014) Histone core modifications regulating nucleosome structure and dynamics. *Nature Reviews.Molecular Cell Biology,* 15 (11): 703-708.

Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., *et al.* (2015) Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome, *Nature Communications,* 6: 7870.

Thomas, P.D., Campbell, M., J, Kejariwal, A., Mi, H., Brian, K., Daverman, R., *et al.* (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by Function, *Genome Res.,* 13: 2129.

Thomas, R., Al-Khadairi, G., Roelands, J., Hendrickx, W., Dermime, S., Bedognetti, D., *et al.* (2018) NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives, *Frontiers in Immunology,* 9: 947.

Tiepolo, L., Maraschio, P., Gimell, G., Cuoco, C., Gargani, G.F. & Romano, C. (1979) Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency, *Human Genetics,* 51: 127-137.

Timp, W. & Feinberg, A.P. (2013) Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host, *Nature Reviews.Cancer,* 13 (7): 497-510.

Toyoda, S., Kawaguchi, M., Kobayashi, T., Tarusawa, E., Toyama, T., Okano, M., *et al.* (2014) Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity, *Neuron,* 82 (1): 94-108.

Trevino, V., Falciani, F. & Barrera-Saldana, H.A. (2007) DNA microarrays: a powerful genomic tool for biomedical and clinical research, *Molecular Medicine (Cambridge, Mass.),* 13 (9-10): 527-541.

Tropberger, P. & Schneider, R. (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications, *Nature Structural & Molecular Biology,* 20 (6): 657-661.

Tsai, H.C., Li, H., Van Neste, L., Cai, Y., Robert, C., Rassool, F.V., *et al.* (2012) Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells, *Cancer Cell,* 21 (3): 430-446.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, *et al.* (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b, *Genes to Cells:Devoted to Molecular & Cellular Mechanisms,* 11 (7): 805-814.

Urich, M.A., Nery, J.R., Lister, R., Schmitz, R.J. & Ecker, J.R. (2015) MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing, *Nature Protocols,* 10 (3): 475-483.

Valinluck, V. & Sowers, L.C. (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1, *Cancer Research,* 67 (3): 946-950.

van Kruijsbergen, I., Hontelez, S. & Veenstra, G.J. (2015) Recruiting polycomb to chromatin, *The International Journal of Biochemistry & Cell Biology,* 67: 177-187.

Venkatesh, S. & Workman, J.L. (2015) Histone exchange, chromatin structure and the regulation of transcription, *Nature Reviews Molecular Cell Biology,* 16 (3): 178-189.

Vignali, M., Hassan, A.H., Neely, K.E. & Workman, J.L. (2000) ATP-dependent chromatin-remodeling complexes, *Molecular and Cellular Biology,* 20 (6): 1899-1910.

Vispe, S., Deroide, A., Davoine, E., Desjobert, C., Lestienne, F., Fournier, L., *et al.* (2015) Consequences of combining siRNA-mediated DNA methyltransferase 1 depletion with 5-aza-2'-deoxycytidine in human leukemic KG1 cells, *Oncotarget,* 6 (17): 15265-15282.

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., *et al.* (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation, *Nucleic Acids Research,* 44 (12): 5615-5628.

Walsh, C.P., Chaillet, J.R. & Bestor, T.H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation, *Nature Genetics,* 20 (2): 116-117.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., *et al.* (2004) Role of histone H2A ubiquitination in Polycomb silencing, *Nature,* 431 (7010): 873-878.

Wang, X., Weiner, J.A., Levi, S., Craig, A.M., Bradley, A. & Sanes, J.R. (2002) Gamma protocadherins are required for survival of spinal interneurons, *Neuron,* 36 (5): 843-854.

Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., *et al.* (2013) Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects, *European Journal of Human Genetics: EJHG,* 21 (11): 1219-1225.

Weiser, T.S., Guo, Z.S., Ohnmacht, G.A., Parkhurst, M.L., Tong-On, P., Marincola, F.M., *et al.* (2001) Sequential 5-Aza-2 deoxycytidine-depsipeptide FR901228 treatment induces apoptosis preferentially in cancer cells and facilitates their recognition by cytolytic T lymphocytes specific for NY-ESO-1, *Journal of Immunotherapy (Hagerstown, Md.: 1997),* 24 (2): 151-161.

Whitehurst, A.W. (2014) Cause and consequence of cancer/testis antigen activation in cancer, *Annual Review of Pharmacology and Toxicology,* 54: 251-272.

Williams, K., Christensen, J. & Helin, K. (2011) DNA methylation: TET proteins-guardians of CpG islands? *EMBO Reports,* 13 (1): 28-35.

Wolff, F., Leisch, M., Greil, R., Risch, A. & Pleyer, L. (2017) The double-edged sword of (re)expression of genes by hypomethylating agents: from viral mimicry to exploitation as priming agents for targeted immune checkpoint modulation, *Cell Communication and Signaling: CCS,* 15 (1), 13-017-0168-z.

Woodfine, K., Huddleston, J.E. & Murrell, A. (2011) Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue, *Epigenetics & Chromatin,* 4 (1), 1.

Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., *et al.* (2016) Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals, *Human Molecular Genetics,* 25 (18): 4041-4051.

Wutz, A. (2003) RNAs templating chromatin structure for dosage compensation in animals, *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 25 (5): 434-442.

Wyatt, G.R. & Cohen, S.S. (1952) A New Pyrimidine Base from Bacteriophage, *Nucleic Acids*, 170 (4339): 1072-1073.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., *et al.* (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling, *Nature*, 442 (7098): 86-90.

Xie, S., Jakoncic, J. & Qian, C. (2012) UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail. *Journal of Molecular Biology*, 415 (2): 318-328.

Xie, S., Wang, Z., Xie, S., Nogami, M., Li, Y., *et al.* (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family, *Gene*, 236 (1): 87-95.

Xie, Z.H., Huang, Y.N., Chen, Z.X., Riggs, A.D., Ding, J.P., Gowher, H., *et al.* (2006) Mutations in DNA methyltransferase DNMT3B in ICF syndrome affect its regulation by DNMT3L, *Human Molecular Genetics*, 15 (9): 1375-1385.

Xu, G.L. &.W, C.P. (2014) Enzymatic DNA oxidation: mechanisms and biological significance, *BMB Reports*, 47 (11): 609-618.

Yan, P., Frankhouser, D., Murphy, M., Tam, H.H., Rodriguez, B., Curfman, J., *et al.* (2012) Genome-wide methylation profiling in decitabine-treated patients with acute myeloid leukemia, *Blood*, 120 (12): 2466-2474.

Yang, L., Rau, R. & Goodell, M.A. (2015) DNMT3A in haematological malignancies. *Nature Reviews.Cancer*, 15 (3): 152-165.

Yang, N., Sun, R., Liao, X., Aa, J. & Wang, G. (2017) UDP-glucuronosyltransferases (UGTs) and their related metabolic cross-talk with internal homeostasis: A systematic review of UGT isoforms for precision medicine, *Pharmacological Research*, 121: 169-183.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A. & Liang, G. (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer, *Cancer Cell*, 26 (4): 577-590.

Yen, R.W., Vertino, P.M., Nelkin, B.D., Yu, J.J., el-Deiry, W., Cumaraswamy, A., *et al.* (1992) Isolation and characterization of the cDNA encoding human DNA methyltransferase, *Nucleic Acids Research*, 20 (9): 2287-2291.

Yoder, J.A., Walsh, C.P. & Bestor, T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites, *Trends in Genetics*, 13 (8): 335-340.

Yong, W.S., Hsu, F.M. & Chen, P.Y. (2016) Profiling genome-wide DNA methylation, *Epigenetics & Chromatin*, 9, 26-016-0075-3. eCollection 2016.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., *et al.* (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome, *Cell,* 149 (6): 1368-1380.

Zhang, T., Termanis, A., Ozkan, B., Bao, X.X., Culley, J., de Lima Alves, F., Rappsilber, J., *et al.* (2016) G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells, *Cell Reports,* 15 (1): 77-85.

Zhang, T., Haws, P. & Wu, Q. (2004) Multiple variable first exons: a mechanism for cell- and tissue- specific gene regulation, *Genome Res.*, 14: 79-84.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., *et al. (*2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail, *Nucleic Acids Research,* 38 (13): 4246-4253.

Zhao, M.T., Whyte, J.J., Hopkins, G.M., Kirk, M.D. & Prather, R.S. (2014) Methylated DNA immunoprecipitation and high-throughput sequencing (MeDIP-seq) using low amounts of genomic DNA, *Cellular Reprogramming,* 16 (3): 175-184.

Zhu, J., He, F., Hu, S. & Yu, J. (2008) On the nature of human housekeeping genes. *Trends in Genetics,* 24 (10): 481-484.

Ziller, M.J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., *et al.* (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types, *PLoS Genetics,* 7 (12), e1002389.

Zingg, J.M. & Jones, P.A. (1997) Genetic and epigenetic aspects of DNA methylation on genome expression, evolution, mutation and carcinogenesis, *Carcinogenesis,* 18 (5): 869-882.

Zlotorynski, E. (2015) Non-coding RNA: X-chromosome inactivation unravelled, *Nature Reviews.Molecular Cell Biology,* 16 (6): 325.

PAPER -1

Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2.

Avinash Thakur, Sarah-Jayne Mackin, Rachelle E. Irwin, Karla M. O'Neill, Gareth Pollin and Colum P. Walsh.

Epigenetics Chromatin. 2016; 9: 53.

doi: 10.1186/s13072-016-0104-2

The main aims of this paper were to:

1. Compare the roles of DNMT1 and the DNMT3A/B enzymes in maintaining methylation levels at imprinted genes in mouse ESC.
2. Obtain more quantitative and systematic data on all gametic differentially methylated regions (gDMR) which control imprinted loci, including new putative gDMR recently identified.
3. To re-examine the dependence of imprinted regions on germline passage to restore methylation, once lost.

CONTRIBUTION

I carried out the bioinformatic mining of Wang et al., (2014) (Reference 15) that allowed me to design custom pyrosequencing assays in-house to provide a quantative measurement of methylation at both the known and recently delineated imprinted regions. I also carried out wet-lab work that included maintaining all of the ES cell lines in culture, extracting DNA from all five ES cell lines, bisulfite converting this DNA, completing the pyrosequencing for these samples in all of the pyrosequecing assyas I designed in at least triplicate. I optimised the RT-PCR primers and completed a range of statistical analyses in addition to figure assembly and contributed to the writing of the manuscript and edits.

Summary of the major findings;

1. A loss of DNMT1 and DNMT3A/B created similar reductions in methylation levels at imprinted gDMRs.

2. Recovery of methylation at imprints is possible when 3abKO cells are rescued with DNMT3a2. Methylation cannot be recovered however at imprints in the rescued 1KO+1 ES cells.

3. It was observed at three gDMRS once methylation was lost; it was unable to be recovered.

Thakur *et al. Epigenetics & Chromatin (2016) 9:53*
DOI 10.1186/s13072-016-0104-2

Epigenetics & Chromatin

RESEARCH

Open Access

Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2

Avinash Thakur[1,2], Sarah-Jayne Mackin[1], Rachelle E. Irwin[1], Karla M. O'Neill[1,3], Gareth Pollin[1] and Colum Walsh[1*]

Abstract

Background: Imprinted loci are paradigms of epigenetic regulation and are associated with a number of genetic disorders in human. A key characteristic of imprints is the presence of a gametic differentially methylated region (gDMR). Previous studies have indicated that DNA methylation lost from gDMRs could not be restored by DNMT1, or the de novo enzymes DNMT3A or 3B in stem cells, indicating that imprinted regions must instead undergo passage through the germline for reprogramming. However, previous studies were non-quantitative, were unclear on the requirement for DNMT3A/B and showed some inconsistencies. In addition, new putative gDMR has recently been described, along with an improved delineation of the existing gDMR locations. We therefore aimed to re-examine the dependence of methylation at gDMRs on the activities of the methyltransferases in mouse embryonic stem cells (ESCs).

Results: We examined the most complete current set of imprinted gDMRs that could be assessed using quantitative pyrosequencing assays in two types of ESCs: those lacking DNMT1 (1KO) and cells lacking a combination of DNMT3A and DNMT3B (3abKO). We further verified results using clonal analysis and combined bisulfite and restriction analysis. Our results showed that loss of methylation was approximately equivalent in both cell types. 1KO cells rescued with a cDNA-expressing DNMT1 could not restore methylation at the imprinted gDMRs, confirming some previous observations. However, nearly all gDMRs were remethylated in 3abKO cells rescued with a DNMT3A2 expression construct (3abKO + 3a2). Transcriptional activity at the *H19/Igf2* locus also tracked with the methylation pattern, confirming functional reprogramming in the latter.

Conclusions: These results suggested (1) a vital role for DNMT3A/B in methylation maintenance at imprints, (2) that loss of DNMT1 and DNMT3A/B had equivalent effects, (3) that rescue with DNMT3A2 can restore imprints in these cells. This may provide a useful system in which to explore factors influencing imprint reprogramming.

Keywords: Imprinting, DNA methylation, Reprogramming, ESC

Background

In mouse, DNA methylation is found predominantly at cytosine when followed by guanine (CpG) and is associated with various biological functions including the regulation of gene expression, X chromosome inactivation, silencing of retrotransposons and imprinting [1]. Many CpGs are protected from methylation by being clustered into CpG islands (CGI), which are commonly found near the transcriptional start sites of genes and are normally unmethylated, except for CGI on the inactive X or on inactive imprinted alleles. DNMT1, a maintenance methyltransferase [2], is crucial to ensure the regular propagation of DNA methylation patterns to the daughter strand during replication [3]. This enzyme is predominantly found near replication foci [4] and preferentially targets hemi-methylated DNA [4–6] suggesting its main functions as a maintenance methyltransferase [7–9]. The

*Correspondence: cp.walsh@ulster.ac.uk
[1] Genomic Medicine Research Group, Biomedical Sciences Research Institute, Centre for Molecular Biosciences, University of Ulster, Coleraine BT52 1SA, UK
Full list of author information is available at the end of the article

addition of methylation to an unmethylated template (de novo) is carried out by DNMT3A and DNMT3B, with the former responsible for most de novo activity in germ cells [10], while the latter predominates in somatic tissues [11]. However, in addition to their de novo methylation activity, several reports on DNMT3A and DNMT3B indicate a role in methylation maintenance in embryonic stem cells (ESCs), although the extent of their requirement at imprinted loci remains unclear [12, 13].

Once established on a DNA duplex, methylation is stably maintained through most of life [14, 15], but during certain developmental stages undergoes large-scale changes [11–13, 16]. Methylation patterns inherited from the sperm and oocyte are remodelled during pre-implantation development, when the paternal and maternal genomes of the embryo undergo widespread active demethylation involving the TET enzymes as well as passive demethylation via replicative dilution [15, 17]. The blastocyst stage sees methylation reach its nadir, but following implantation, a wave of de novo methylation occurs causing overall global hypermethylation at most non-island CpG in the adult tissues [18]. This de novo activity is present at high levels in ESCs [5], developing germ cells and early post-implantation embryos [19] but is present at lower levels in somatic cells [20, 21]. The presence of de novo activity in ESCs makes these cells a suitable model to study the mechanism of de novo methylation in mammals.

One group of genes that largely escapes global methylation remodelling during somatic development is the imprinted genes [14, 15]. These are a group of genes which exhibit expression from one parental allele only [22, 23]. Regulation of imprinting has biological significance as imprinted genes are important for embryonic development and their dysregulation leads to embryonic death in mouse and to various disease syndromes in human [22]. Initiation of allele-specific gene methylation patterns starts in the male and female germline during gametogenesis [24]. For imprinted genes, one of the parental alleles acquires DNA methylation at certain locations, and these are detected as differentially methylated regions (DMRs) in somatic cells [25]. Due to their origin in the germline, they are known as the gametic differentially methylated regions (gDMRs) [26], to distinguish them from other types of DMR such as tissue-specific DMR. Some of the gDMR are at *cis*-acting regulatory regions and are known to control monoallelic expression of more than one linked gene: where this has been proven by experimentation the DMRs are called imprinting control regions (ICRs) [27–31]. Methylation at gDMRs is established in the germline largely by the de novo methyltransferase DNMT3A with the aid of the essential cofactor DNMT3L [32–35]. The gDMRs at imprinted regions exhibit the property of being able to resist the processes of active and passive DNA demethylation during the pre-implantation stages of mammalian development or iPS formation [14, 18, 36, 37].

Loss of imprinting is thought to be irreversible and requires germline passage for its recovery due to the presence of essential factors and de novo methyltransferases needed for imprint establishment there [38]. Previous work has shown that rescuing DNA methyltransferase activity in *Dnmt1*$^{-/-}$ (1KO) cells by adding back a cDNA expressing the enzyme failed to restore methylation at paternal and maternal ICRs [38]. Other laboratories confirmed this but reported, however, that the paternally imprinted *H19* gDMR regained methylation in *Dnmt3a*$^{-/-}$; *Dnmt3b*$^{-/-}$ double knock-out (3abKO) cells rescued with a DNMT3A2 expression plasmid [39], suggesting that some imprints could be somatically reprogrammed. As well as these differing results, the early studies were carried out on a very limited number of gDMRs using qualitative approaches, which had limited resolution. Given the important implications somatic resetting could have for imprinted disease syndromes as well as cellular reprogramming generally, we wished to re-examine whether methylation at gDMRs could be established outside of the germline. Recent work has delineated the gDMRs far more sharply since the original studies were carried out, and more quantitative techniques are now available. We aimed to investigate (1) whether deletion of *Dnmt3ab* gives comparable methylation loss at imprinted loci to *Dnmt1* mutated cells; (2) whether imprints can be restored in 3abKO cells, unlike 1KO ESCs; (3) does loss of methylation result in dysregulated expression of imprinted genes; and (4) are there any exceptional imprinted gDMRs that do not regain methylation in rescued cells?

Methods

Statistical analysis

All laboratory experiments were carried out in triplicate with at least one biological repeat, with one or two exceptions as noted. Pyrosequencing, bisulfite sequencing and RT-qPCR data are represented as graphs, where error bars represent standard error of the mean (s.e.m). Statistical analysis was carried out using EXCEL and GraphPad PRISM software; for pyrosequencing data were compared by *t* test and Kruskal–Wallis, and bisulfite clonal analysis comparison was made using the χ^2 test.

Cells

All cell culture media components were purchased from Invitrogen (Paisley, UK). *Dnmt1* KO and *Dnmt3a/3b* double KO cells with matching WT were kind gifts from

Dr. Masaki Okano (RIKEN Center for Developmental Biology, Kobe, Japan). ESCs were maintained on Nunc plates (Davidson & Hardy, Belfast, UK) treated with 0.1% gelatin (Sigma-Aldrich, Dorset, UK) and cultured in Knockout DMEM plus 15% knockout serum replacement, 1% ESC-qualified Foetal Bovine Serum, 1× NEAA, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol (Sigma-Aldrich, Dorset, UK) and 1000U/ml LIF (Merck Millipore, Hertfordshire, UK).

Animal work

Tissues of interest were derived from outbred TO mice (Harlan, Huntingdon, UK). Sperm collection was carried out as previously described [13].

RNA extraction, cDNA synthesis and RT-qPCR

RNA was extracted using the RNeasy kit (Qiagen, Crawley, UK), according to the manufacturer's instructions. For cDNA synthesis, 300–500 ng RNA was used in combination with 0.5 µg random primers (Roche, West Sussex, UK), 40 U RNaseOUT 0.5 µM dNTPs (Invitrogen, Paisley, UK) 1× RT Buffer (Fermentas, Cambridge, UK) and RevertAid reverse transcriptase (Fermentas, Cambridge, UK) made up to a final volume of 20 µl using RNase-free water (Qiagen, Crawley, UK). Reactions were carried out in a thermocycler with conditions—25 °C for 10 min, 42 °C for 60 min and 70 °C for 10 min. One microlitre cDNA per well on a 96-well plate (Roche) was used for RT-qPCR with SYBR Green reagent and remaining cDNA stored at −80 °C. RT-qPCRs were performed using a LightCycler 480 Instrument II (Roche, West Sussex, UK). Gene expression was normalised to *Hprt* and relative expression calculated by the $\Delta\Delta C_T$ method [40]. Each RT-qPCR contained 1× buffer, 0.4 mM dNTPs, 50 µM primers (Additional file 1: Table S1), 0.01 U Taq DNA polymerase (Invitrogen, Paisley, UK) and nuclease-free water (Qiagen, Crawley, UK). Four primer sets for *Dnmt1*, *Dnmt3a*, *Dnmt3b* [47] and *Hprt* [13] were used. The general thermocycler conditions are as follows—94 °C for 3 min, followed by 30 cycles of 94 °C for 30 s, 63 °C for 1 min, 72 °C for 1 min with a final elongation step of 72 °C for 4 min.

Protein analysis

Protein was extracted from cells growing in log phase using protein extraction buffer (50 mM Tris–HCl, 150 mM NaCl, 1% Triton-X, 10% glycerol, 5 mM EDTA; all Sigma-Aldrich) and 0.5 µl protease inhibitor mix (Sigma-Aldrich, Dorset, UK). For Western blotting, 30 µg protein was denatured in the presence of 5 µl 4× LDS sample buffer (Invitrogen, Paisley, UK) and 2 µl 10× reducing agent (Invitrogen) in a total volume of 20 µl nuclease-free water (Qiagen, Crawley, UK) via incubation at 70 °C. Proteins were fractionated on a 4–12% SDS-PAGE gel, then electroblotted onto a nitrocellulose membrane (Invitrogen, Paisley, UK) and blocked in 5% non-fat milk for 1 h at room temperature (RT). Membranes were incubated with anti-DNMT1 (ab87654, Abcam), anti-DNMT3A (clone 64B1446, Novus Biologicals, Abingdon, UK), anti-GAPDH (clone 14C10, Cell Signalling Technologies, Leiden, Netherlands) or anti-β-actin (clone AC-15, Sigma-Aldrich) overnight at 4 °C, followed by incubation with the relevant HRP-conjugated secondary antibody (Sigma-Aldrich, Dorset, UK) for 1 h at RT and then visualised using ECL (ThermoFisher Scientific, Loughborough, UK).

DNA isolation and bisulfite conversion

DNA extraction from sperm and tissues was as previously described [41]. All ESCs were pelleted and incubated overnight at 55 °C in lysis buffer (50 mM Tris pH 8, 0.1 M EDTA, 0.5% SDS (all from Sigma-Aldrich, Dorset, UK), 0.2 mg/ml proteinase K (Roche, West Sussex, UK) with rotation. DNA was extracted next day using the phenol/chloroform/isoamylalcohol (25:24:1, pH 8.0; Sigma-Aldrich, Dorset, UK) extraction method. The integrity of the DNA was checked on a 1% agarose gel (Eurogentec, Southampton, UK) and quality and quantity checked using a NanoDrop UV spectrophotometer (Labtech International, Ringmer, UK). For bisulfite conversion, 500 ng of DNA was processed with the EpiTect Bisulfite Kit (Qiagen, Crawley, UK) or EZ DNA Methylation Kit (Zymo, Cambridge, UK) according to manufacturer's instructions.

Methylation analysis

Bisulfite-converted DNA was PCR amplified in a reaction containing 1 µM primers, 1× buffer and 0.4 mM dNTPs, with $MgCl_2$ at a concentration specific to the primer set and 0.01U *Taq* DNA polymerase (all reagents from Invitrogen, Paisley, UK). Combined bisulfite restriction analysis (COBRA) on genes was carried out as previously described [13] using *Taqα*I enzyme for *H19* and *KvDMR* and *BstU*I for *Snrpn* (both New England Biolabs, Hitchin, UK). Clonal analysis of bisulfite-converted PCR-amplified products in pJET1.2 vector (Fermentas, Cambridge, UK) was carried out using a PRISM 3130 Genetic Analyzer (Applied Biosystems, Paisley, UK). All pyrosequencing assays (Additional File 2: Table S2) were designed in-house using PyroMark (V2.0) assay design software (Qiagen, Crawley, UK). The PyroMark PCR Kit (Qiagen, Crawley, UK) was used to amplify genes using a thermocycler (Techne, Stone, UK) with conditions: 95 °C, 15 min; followed by 45 cycles of 94 °C for 30 s, 56 °C for 30 s and 72 °C for 30 s; with final elongation at 72 °C for 10 min. Subsequent pyrosequencing was carried out

using Pyromark reagents as per manufacturer's recommendations (Qiagen, Crawley, UK); 2 M NaOH was from Sigma-Aldrich (Dorset, UK) and Sepharose beads from GE Healthcare (Chalfont St. Giles, UK).

The Luminometric Methylation Assay (LUMA) using 300 ng/μl of genomic DNA from the respective cell lines was carried out exactly as described previously [13, 49]. HCT116 WT DNA (hypermethylated) and DKO DNA (hypomethylated) samples were used as a control (data not shown).

Optimising primer alignment with galaxy user-defined tracks

Wang et al. [15] provided chromosomal coordinates for numerous known and putative germline imprints as part of their supplemental material. The coordinates delineated for each imprint were used as a tool to define the minimal gDMR regions, from which the respective genomic sequence was extracted by visualising these regions on UCSC genome browser. The extracted genomic sequence was used to promote specificity in the design of pyrosequencing assays. BED files were generated using these chromosomal coordinates and uploaded through the Galaxy interface [42] as user-defined tracks visible on UCSC genome browser. The genomic sequence of interest generated from each respective imprint primer set created was matched using the BLAT tool at UCSC against the user-defined track to confirm the positions of the assays (Additional file 1: Table S1).

Results

Initial gDMR examined and regions assayed

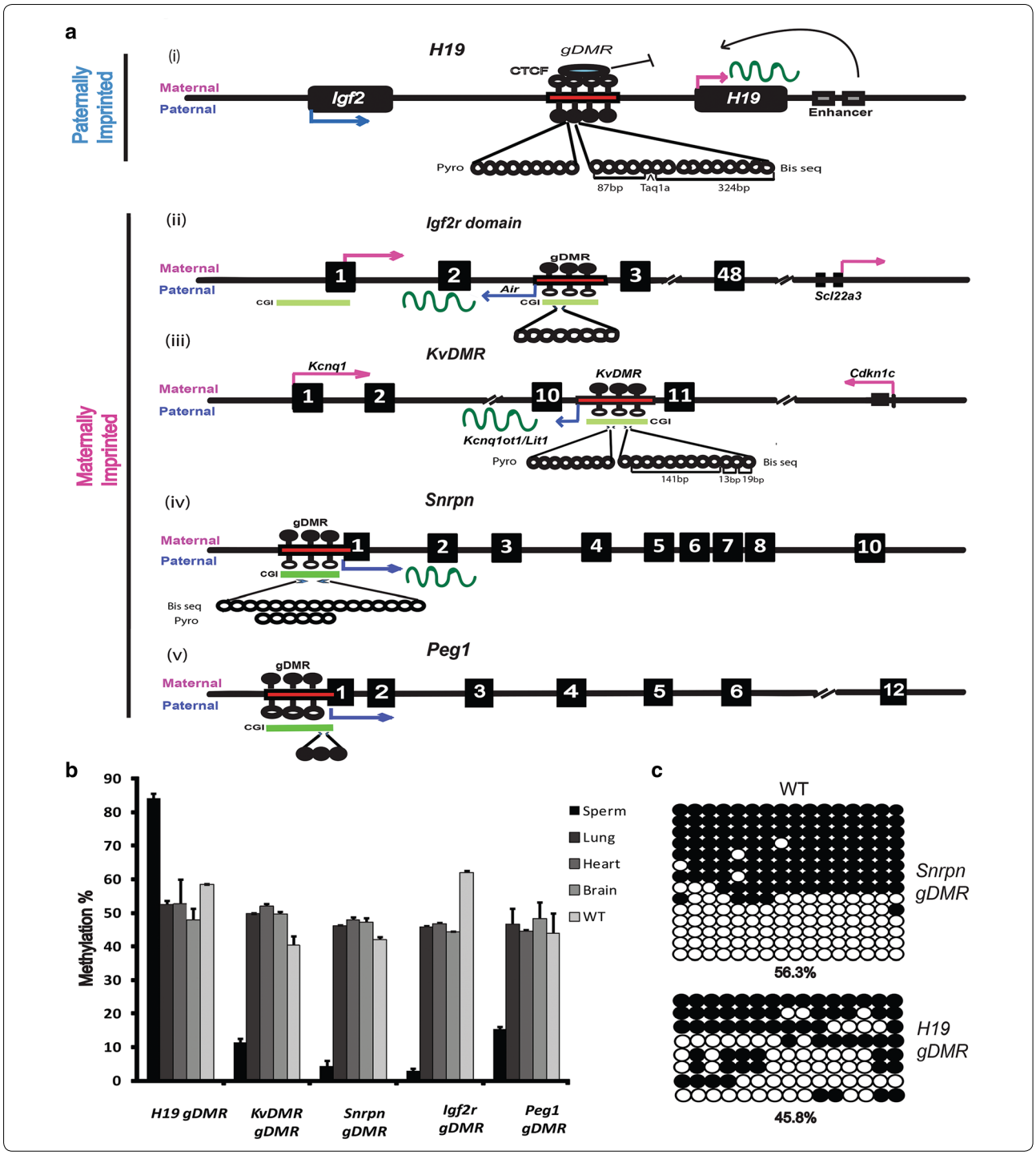
We began our study by designing and validating pyrosequencing assays, as it is crucial that the designed primers cover the right regions at imprinted loci where methylation, once established, remains unchanged throughout development. To validate the approach, we initially chose five of the best-characterised imprinted loci for which extensive data on the gametic differentially methylated regions (gDMR) are available and which are representative of the different kinds of imprinted locus. The positioning of the gDMRs at these five imprinted loci is shown in Fig. 1. The paternally imprinted *H19* gDMR controls a small cluster of genes including *Igf2* [Fig. 1a(i)] and represents an insulator model of imprinting. On the maternal chromosome, CCCTC-binding protein (CTCF) binds to the gDMR, located intergenically, and forms an insulator to stop the interaction of the enhancer with the *Igf2* promoter. Such binding results in the silencing of *Igf2* on the maternal allele but allows the enhancers to activate *H19* (bent arrow) on the same allele. On the paternal chromosome, the ICR is methylated which prevents CTCF from binding; therefore, the enhancers can interact with *Igf2*, resulting in its transcription. The two parts of the intergenic gDMR covered by our pyrosequencing assay and by the clonal analysis/COBRA are also shown [Fig. 1a(i)]. Current indications are that many other imprinted genes seem likely to follow a non-coding RNA (ncRNA)-mediated model for regulation of imprinting. Two examples of this class are the maternally imprinted loci controlled by the *Igf2r* [Fig. 1a(ii)] [22] and *KvDMR* gDMRs [Fig. 1a(iii)], both located intragenically in introns. *Igf2r* and its neighbouring genes show maternal expression, and the *Igf2r* gDMR generates a paternally expressed non-coding transcript *Air* [Fig. 1a(ii)]. The full-length *Air* ncRNA and its transcription are required for the silencing of *Igf2r* and other neighbouring genes [29, 43]. *KvDMR* is the origin of a paternally expressed long ncRNA *Kcnq1ot1/Lit1* [Fig. 1a(iii)], which regulates imprinting at the *Kcnq1* locus. Truncation of *Kcnq1ot1* results in a loss of imprinting [44, 45]. Maternally imprinted *Snrpn* and *Peg1* represent another type of imprinted loci, where the gDMR is located directly at the promoter region of a gene, rather than intra- or intergenically. At these loci, methylation directly controls transcription [Fig. 1a(iv–v)].

Methylation at imprinted gDMRs in WT ESCs is similar to that in normal tissues

Using pyrosequencing assays designed to match the known gDMR, we found that the *H19* gDMR (paternally imprinted) was hypermethylated (84.2%) in sperm samples, while all maternally imprinted gDMRs displayed very low methylation (Fig. 1b), as expected. All the gDMRs assayed also showed methylation around 50% (range normally observed 40–60% [46]) in somatic tissues and WT ESCs, although the level of *Igf2r* methylation was reproducibly higher in WT ESC. Further, we confirmed a normal level of methylation for *H19* and *Snrpn* gDMRs in WT ESCs by clonal analysis (Fig. 1c), which at 56.3 and 45.8%, respectively, was very comparable to that seen by pyrosequencing (58.5 and 42%) (Fig. 1b). These data: (1) indicated that the regions assayed by pyrosequencing showed the expected levels of methylation in somatic tissue, validating these assays, and (2) that the parental ESCs from which all the subsequent knockouts were derived had relatively normal levels of methylation at the gDMR.

Comparable demethylation at imprinted loci in cells lacking DNMT3A/B or DNMT1

Cells lacking DNMT1 (1KO) and as well as a recued cell line expressing a DNMT1 cDNA from an integrated transgene (1KO + 1) have been previously described [39]: the same authors describe cells lacking both DNMT3A and DNMT3B (3abKO) or rescued with DNMT3A2, and both protein levels and mRNA levels of


a
Paternally Imprinted
Maternally Imprinted
(i)
H19
gDMR
CTCF
Maternal
Paternal
Igf2
H19
Enhancer
Pyro
Bis seq
87bp
Taq1a
324bp
(ii)
Igf2r domain
Maternal
Paternal
gDMR
Air
CGI
Scl22a3
(iii)
KvDMR
Kcnq1
KvDMR
Cdkn1c
Kcnq1ot1/Lit1
141bp
13bp 19bp
(iv)
Snrpn
gDMR
Bis seq
Pyro
(v)
Peg1
b
Methylation %
Sperm
Lung
Heart
Brain
WT
H19 gDMR
KvDMR gDMR
Snrpn gDMR
Igf2r gDMR
Peg1 gDMR
c
WT
Snrpn gDMR
56.3%
H19 gDMR
45.8%

(See figure on previous page.)
Fig. 1 Five canonical imprinted regions in mouse and validation of gDMR methylation assays in WT ESCs. **a** Schematic showing the main features of the imprinted domains examined, along with positioning of imprinted gDMRs (*narrow rectangles*): (*i*) *H19/Igf2* region; (*ii*) *Igf2r* region; (*iii*) *KvDMR/Kcnq1* region; (*iv*) *Snrpn*; and (*v*) *Peg1*. Exons are represented by *dark-filled boxes*, and expression of the genes from maternal or paternal alleles is indicated by *bent arrows above* or *below* the *line*, respectively. *Wavy lines* and *narrow bars below* the *line* represent long non-coding RNA and CpG islands (CGIs), respectively. The number of CpG sites (*circles*) covered by pyrosequencing or clonal analysis, and enzyme restriction site positions and fragment sizes for COBRA are all shown below the gDMRs. *Filled circles* represent methylated sites, while *empty circles* are unmethylated. CCCTC-binding factor (CTCF) binds to *H19/Igf2* ICR on the maternal chromosome only (*i*) to block access of the downstream enhancers to *Igf2*. On the paternal chromosome, CTCF binding is blocked by methylation and the enhancers preferentially interact with *Igf2* rather than *H19*. **b** Methylation level of imprinted gDMRs in various mouse tissues and WT ESCs was quantified by pyrosequencing. Sperm was used as a control: all maternally imprinted genes exhibited low levels (<25%) of methylation and the paternally imprinted gene *H19* showed high levels (>75%) of methylation here, as expected. All maternally and paternally imprinted genes show ~50% methylation at gDMRs in lung, heart and brain as assessed by pyrosequencing. In WT ESC, methylation levels were very comparable, with only *Igf2r* a little high. *Error bars* indicate s.e.m. **c** Clonal analysis of *Snrpn* and *H19* gDMRs in parental WT ESCs. *Each circle* represents a CpG site assessed by bisulfite clonal analysis. Percentage total methylation is indicated at the *bottom of each panel*

the various proteins have been verified [39, 47]. Nevertheless, to ensure the cells have remained stable we verified the correct patterns of loss and rescue in the various cell lines using both westerns and RT-PCR (Additional file 3: Fig. S1). While previous studies have indicated a role for DNMT3A/B proteins in maintenance methylation at some repeats, and possibly at some other sequences in ESC, a potential maintenance role at imprinted gDMRs has not previously been examined in detail. All five gDMRs were found to be severely hypomethylated in 1KO and 3abKO cells. For the paternally imprinted *H19* gDMR (Fig. 2a), significant loss of methylation from 52.19% to less than 10% was observed for both cell types compared to WT ESCs (p value <0.05 for WT ESCs vs. 1KO and 3abKO). All maternal gDMRs also showed significant decreases in methylation in both KO cell types with p value <0.001 for WT ESCs versus either KO for all genes except *Peg1*, where p values were <0.01 and <0.05 for both 1KO and 3abKO respectively. At the *Igf2r* locus, methylation is almost completely lost in both types of knockout line compared to WT. We observed nevertheless a larger decrease in methylation for *Snrpn* in 1KO cells than 3abKO (8.7 vs. 26.1%) cells (Fig. 2b). To check this, we used clonal analysis and could confirm that the methylation level at *Snrpn* was lower in 1KO (1.4%) compared to 3abKO (24.16%) samples (Fig. 2c). The *H19* gDMR was equally hypomethylated in 1KO and 3abKO cells as shown by clonal analysis (p value <0.001 for WT versus 1KO or 3abKO) (Fig. 2c).

Methylation can be restored following loss of DNMT3A/B, but not DNMT1

As DNMT1 and DNMT3A/B appeared to have broadly similar roles in maintaining methylation marks at imprinted gDMRs in this system, we further aimed to investigate whether the loci responded in the same way to restoration of the respective enzymes. We first investigated the methylation levels of these gDMRs in 1KO ES cells and in those rescued with a cDNA expressing the full-length DNMT1 protein (1KO + 1) [48]. As previously reported by ourselves and others, imprinted gDMRs failed to restore methylation to normal WT levels (Fig. 3a). We also compared these results with 3abKO rescued cells (3abKO + 3a2) expressing the full-length DNMT3A2 protein [48]. Surprisingly, the maternally methylated *Igf2r* gDMR displayed complete recovery of methylation in 3abKO rescued cells, while for *Peg1*, gain of methylation in rescued cells (3abKO + 3a2) brought the levels to somewhat higher than WT level. There was in addition very substantial recovery of methylation at the paternally methylated *H19* gDMR (38.14%) as well as the *Snrpn* gDMR (38.88%) although recovery was not fully restored to normal WT levels (WT was 52.19% for *H19* and 42% *Snrpn*, respectively). The increase in methylation for all gDMRs in 3abKO + 3a2 cells was also very significant when compared to 1KO + 1 (Fig. 3a).

As methylation at gDMR was not completely restored to WT levels in the 3abKO + 3a2, we compared whole-genome methylation levels using LUMA [49], a bisulfite-independent quantitative assay that uses methylation-sensitive and methylation-insensitive enzymes to estimate total genomic methylation levels. LUMA analysis of 1KO and 3abKO cells shows clear loss of methylation compared to WT (set to 100%). While both 1KO + 1 and 3abKO + 3a2 cells showed gain in methylation, neither fully recovered to WT levels (Fig. 3b). HCT116 WT and DKO were used to confirm the LUMA was working: as expected, HCT116WT was found to be hypermethylated and methylation drops significantly for DKO cells lacking DNMT1 and DNMT3B (data not shown). To further confirm our results with respect to the imprinted gDMR, we carried out COBRA on *Snrpn* and *H19* gDMRs, which indicated clear losses of methylation in both KO cell types, no recovery in 1KO + 1 and almost complete methylation restoration in 3abKO + 3a2 samples (Fig. 3c). We also used clonal analysis for these gDMRs, which gave similar results (Fig. 3d, e).

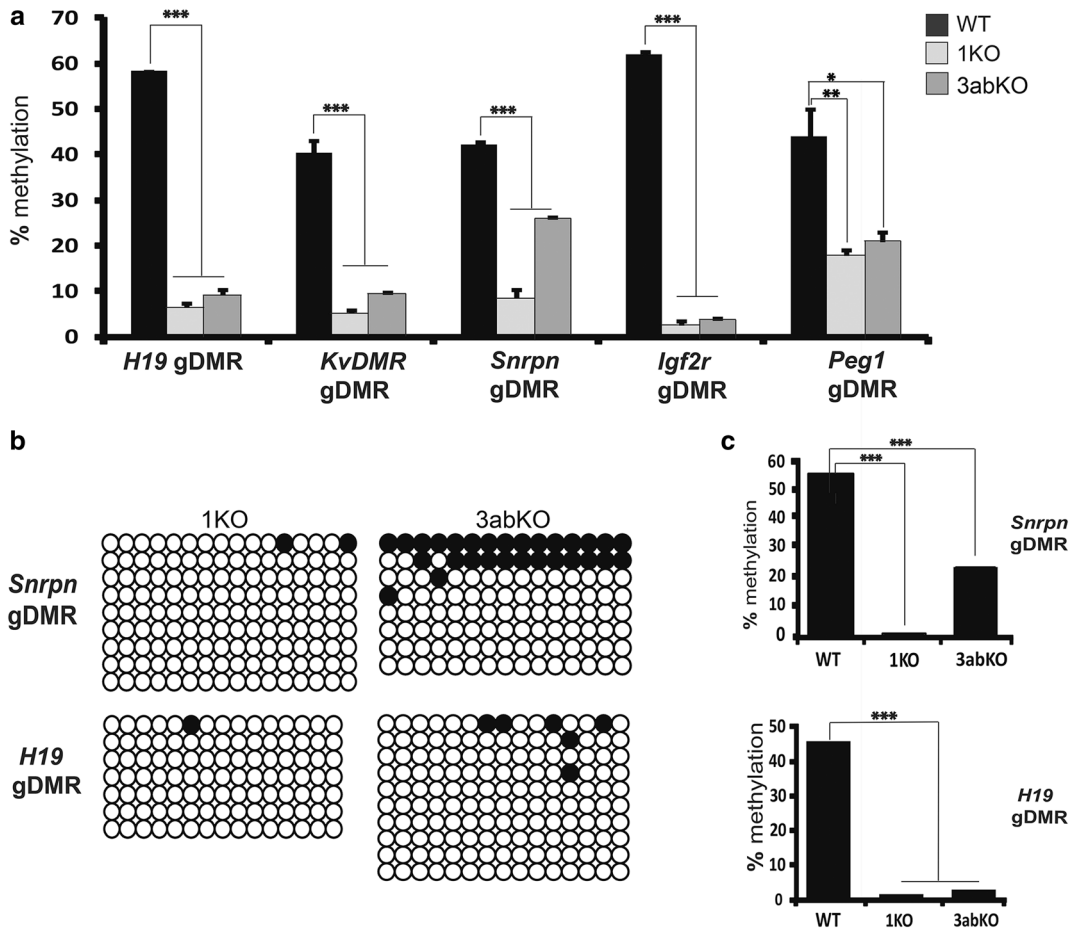


Fig. 2 Mutation of *Dnmt3a/b* or *Dnmt1* gives comparable methylation loss at imprinted loci. **a** Pyrosequencing displaying methylation loss at various imprinted loci in DNMT1-deficient cells (1KO) and cells lacking both DNMT3A and 3B (3abKO). The methylation levels drop significantly for 1KO and 3abKO for all gDMRs assayed. **b** Clonal methylation analysis of imprinted gDMRs (*Snrpn* and *H19*) in 1KO and 3abKO ESCs. **c** Summary of methylation levels from the clonal data in **b**. There was a significant decrease in methylation for *H19* and *Snrpn* in all KO cell types compared to WT. *Error bars* indicate s.e.m. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Transcriptional activity tracks with methylation at the *H19* locus

To test whether the loss of these methyltransferases and their recovery are associated with abnormal expression, we carried out RT-qPCR on a number of imprinted loci. While most imprinted genes tested were not transcribed at significant levels in these ESC, precluding assessment of response, we did find that the expression level of *H19* was significantly higher in 1KO and 3abKO cells as compared to WT, consistent with biallelic expression of *H19* in those cell lines (Fig. 3f). Rescuing 1KO cells with DNMT1 did not restore repression but rescuing 3abKO cells with DNMT3A significantly reduced ($p < 0.05$) levels of *H19* mRNA (Fig. 3f). In keeping with the regulatory mechanism in place at this locus [Fig. 1a(i)], *Igf2* showed the opposite pattern, with repression in 1KO and 3abKO cells and greatest recovery in 3abKO + 3a2 cells (Fig. 3f), though this did not reach significance because of the lower transcription levels. In addition, the *Peg1* mRNA, which is mildly depressed in the 3abKO line, showed a significant repressive effect of adding back in the DNMT3A2 enzyme (Fig. 3f).

Failure to restore methylation at *KvDMR*

The *KvDMR* gDMR showed a normal level of methylation in WT ESCs, loses methylation in 1KO (5.3%) and does not regain methylation in 1KO + 1 cells (Fig. 4a) as for the other imprinted loci. While this gDMR also showed a very low level of methylation in 3abKO (9.5%), unlike the other four benchmark loci examined, rescue with DNMT3A2 in 3abKO cells failed to reinstate methylation at this maternally imprinted locus (Fig. 4a). Differences in methylation were significant for WT versus 1KO, 1KO + 1, 3abKO and 3abKO + 3a2 (p value

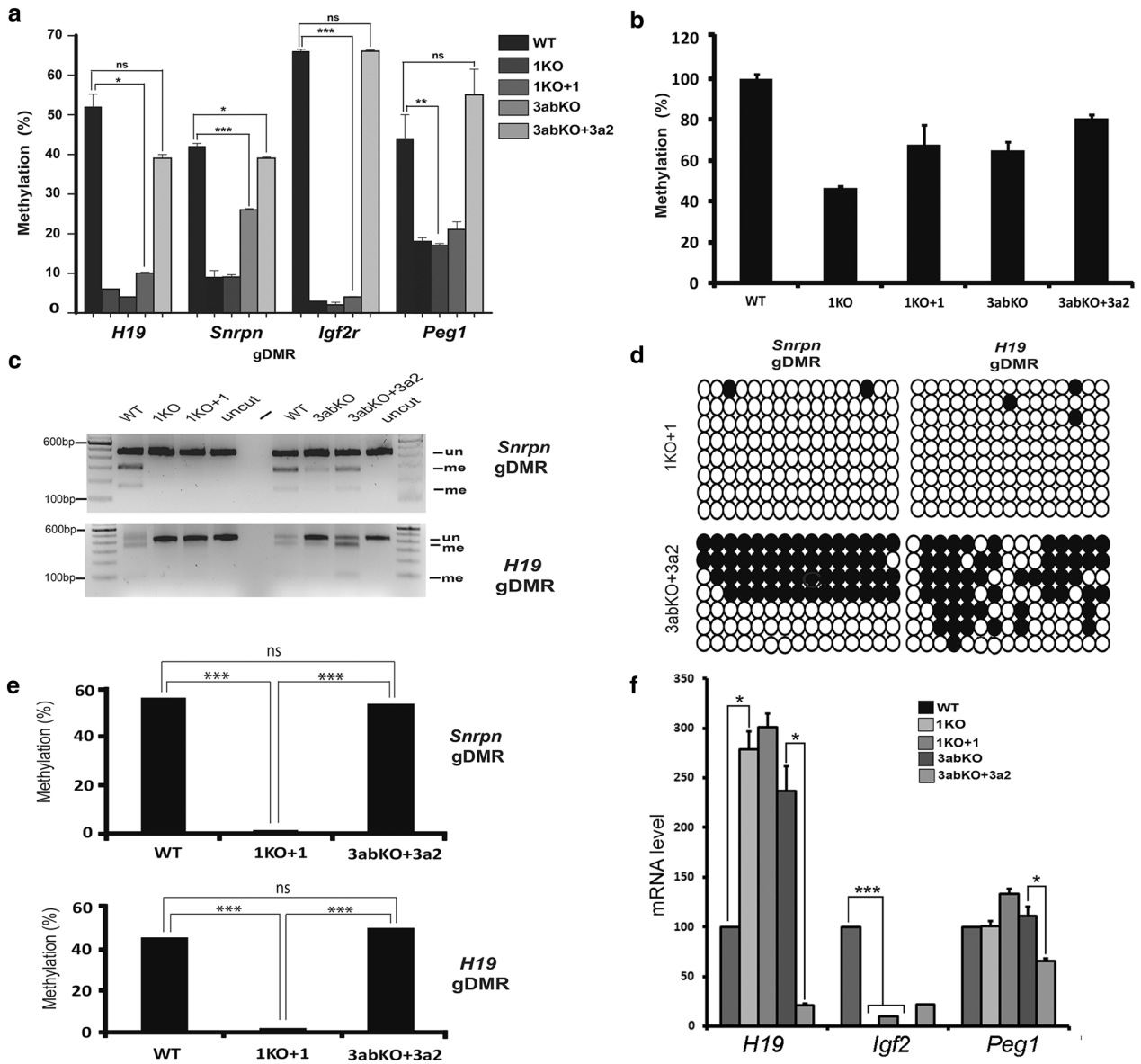


Fig. 3 Despite similar levels of methylation, imprints can be restored only in 3abKO cells. **a** Methylation level assessed by pyrosequencing for parental (WT), DNMT1 rescued (1KO + 1) and 3abKO rescued (3abKO + 3a2) cells. Compared to WT, the methylation level remains significantly lower for all gDMRs in 1KO + 1 cells. All gDMRs show significant increase in methylation in 3abKO + 3a2 cells as compared to 1KO + 1. **b** Global DNA methylation in WT cells, 1KO, 3abKO and 3abKO + 3a2 rescued cells estimated by LUMA. **c** COBRA for two representative imprinted gDMRs, *H19* and *Snrpn*. Smaller fragments represent methylated DNA (me) and can be clearly seen in the WT and 3abKO + 3a2 lanes, but not in the 1KO + 1 lanes: un, unmethylated (D) clonal methylation analysis for *Snrpn* and *H19* showing methylation restoration in 3abKO + 3a2 cells and no recovery for 1KO + 1 cells. **e** Graphical summary of clonal data in **d**. **f** RT-qPCR for *H19*, *Igf2* and *Peg1* in the different ES lines. For *H19*, a significant increase in expression was observed in 1KO and 3abKO cells compared to WT (*p* value <0.05 for all). Rescue with *Dnmt3A2* (3abKO + 3a2), but not *Dnmt1* (1KO + 1), caused a significant decrease in transcription again (*p* value <0.05 3abKO + 3a2 vs. 3abKO cells). *Igf2* transcription is inversely linked to *H19* [Fig. 1a(i)] and as expected decreases significantly (*p* value <0.001 WT vs. 1KO and 3abKO) on loss of methylation in 1KO and 3abKO lines. Of the two rescue lines, 3abKO + 3a2 shows a greater recovery of transcription, though it fails to reach statistical significance. At the *Peg1* locus, there is an increase in transcription on loss of methylation (n.s.) and a significant decrease on reintroduction of DNMT3A2. *Error bars* indicate s.e.m in all panels: only significant changes are shown

<0.001) and non-significant for 1KO versus 1KO + 1, 1KO versus 3abKO and 1KO + 1 versus 3abKO + 3a2 (Fig. 4a). A small but significant decrease in methylation was observed in 3abKO + 3a2 cells compared to 3abKO cells, with a *p* value <0.05. We further verified these results overall using COBRA; a clear loss of methylation

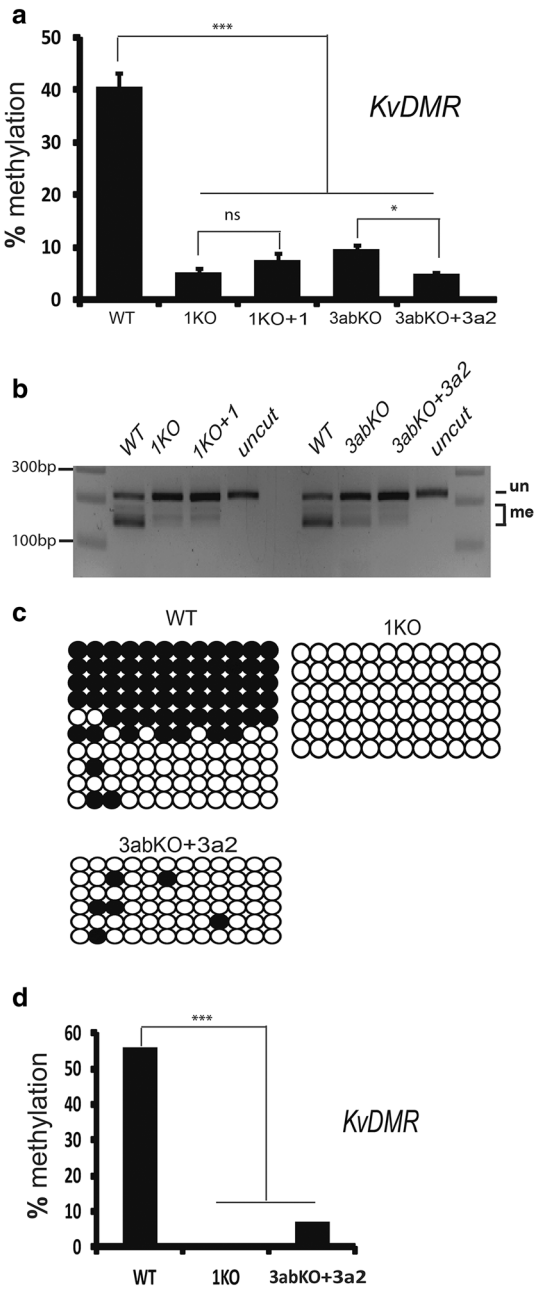


Fig. 4 Rescued 3abKO cells are inefficient at re-establishing imprints on the KvDMR gDMR suggesting a unique regulatory mechanism associated with this imprinted locus. **a** Methylation level of *KvDMR* in various ES cell lines by pyrosequencing. The gDMR exhibits significant loss of methylation in 1KO and 3abKO cells, and methylation remains unchanged in both rescued lines. **b** Validation of pyrosequencing results by COBRA; *TaqaI* was used to digest the PCR product: me, fragments resulting from cleavage of methylated product; un, fragments when unmethylated. There is no increase in methylated fragments from 3abKO to 3abKO + 3a2 samples. **c** Clonal analysis for *KvDMR* in WT, 1KO and 3abKO + 3a2 cells. **d** Graphical summary of clonal data from **c**. *Error bars* indicate s.e.m

was observed in 1KO and 3abKO cells with no methylation restoration for either rescued cell type (Fig. 4b). *KvDMR* was also found via clonal analysis to be 56, 0 and 7% methylated in WT, 1KO + 1 and 3abKO + 3a2 cells, respectively (Fig. 4c), with differences in methylation between WT and rescued cell lines remaining very significant (p value <0.001) (Fig. 4c, d). These results suggest that a unique mechanism may be associated with this maternally imprinted locus.

Loss of methylation of remaining known and putative gDMR in Dnmt3ab KO cells

To determine whether recovery of methylation in 3abKO + 3a2 rescue cells is a general phenomenon for imprinted loci, we wished to extend our study to include all known imprints as well as the putative imprinted loci recently identified by Wang et al. [15]. To this end, we designed pyrosequencing assays for all the remaining known and putative gDMR based on the coordinates indicated in the latter. Assays were excluded which (1) gave low scores in the design software and poor peaks on assaying; (2) displayed methylation values outside ±1.5 times the SD from the 50% methylation expected in a range of mouse somatic tissues [46]; and (3) did not show >75% (paternal) or <25% (maternal) methylation in sperm samples. The results for the remaining five known loci we could assay in somatic tissues (*Plagl1*, *Innp5f*, *Grb10*, *Rasgrf1* and *Dlk1-Gtl2 IG*) are shown in Fig. 5a and displayed relatively tight clustering around the median, which was noticeably higher in heart tissue.

Assays designed to cover four novel putative gDMRs described in Wang et al. [15] (*Neurog3* upstream, *FR149454* promoter, *Pvt1* promoter and *6330408a02Rik* 3′ end) also validated in a range of mouse somatic tissue (Fig. 5b), with average values for the assays falling within the threshold criteria as per Fig. 5a. Notably, there was greater variation in the methylation values for these loci than for the known gDMR. There was less deviation from the median methylation value in liver than heart tissue, which has consistently exhibited a wider spread of results (Fig. 5a, b). The results for the known and putative imprints in mouse sperm samples are shown in Fig. 5c, where they displayed low (22.8% or below) methylation for maternally methylated gDMRs, or conversely hypermethylation (78%) at the respective paternally methylated regions.

With our validated assays, we next examined these nine additional known or putative gDMRs in the mouse ESCs. All of these loci showed decreases in methylation in 3abKO cells when compared to WT (Fig. 5d), though differences at some (*Innp5f*, *Dlk1-Gtl2 IG* and *Pvt1* promoter) did not reach statistical significance.

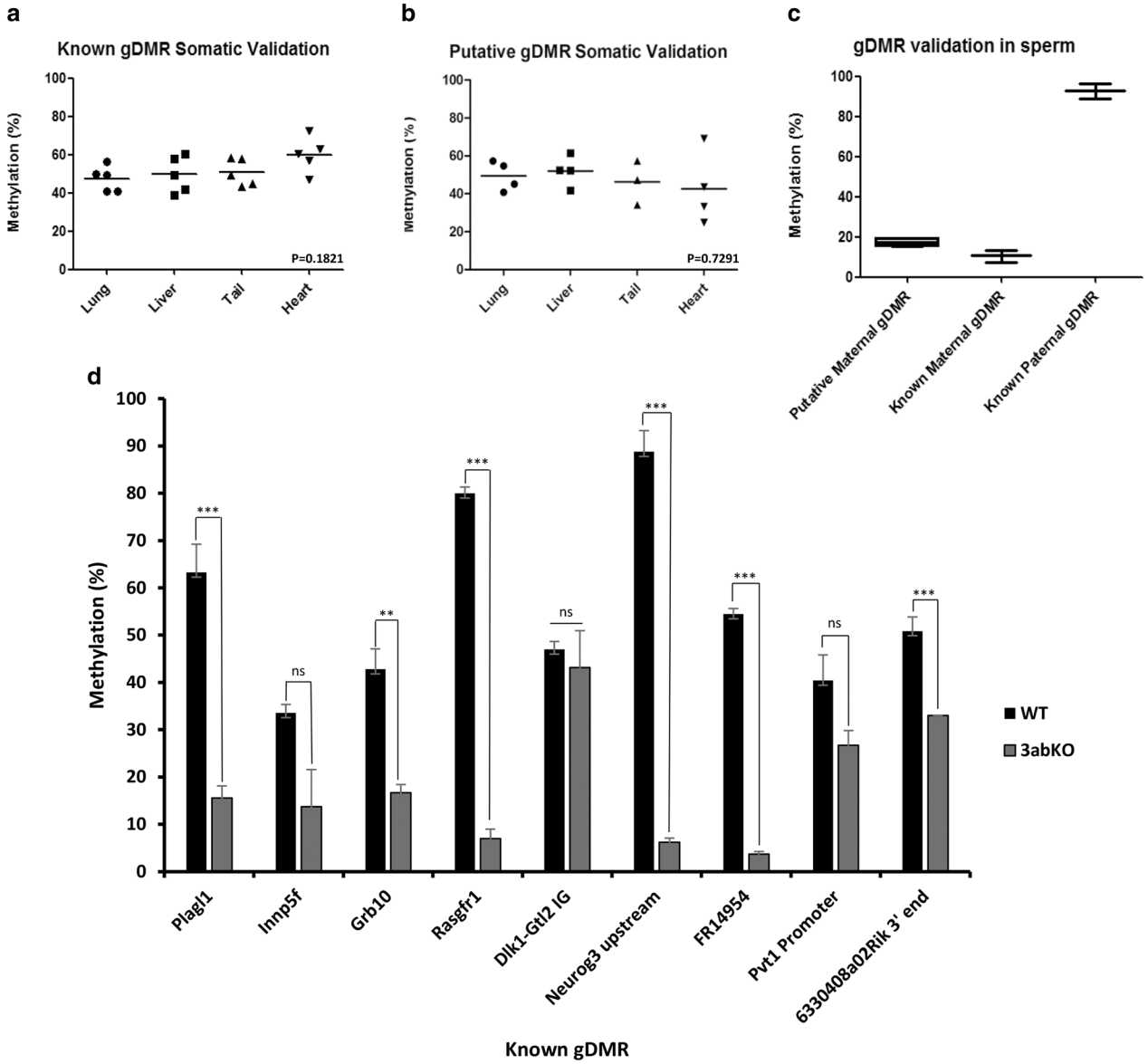


Fig. 5 *Examination by pyroassay of remaining known and putative imprinted gDMR*. **a** Validation of pyrosequencing assays for remaining known imprinted gDMRs. Average methylation levels across all CpG in each assay were plotted. The median for all assays is indicated by a *horizontal line*, and these did not significantly differ from one another using Kruskal–Wallis (*p* value 0.1821). **b** Results for putative imprinted gDMR, plotted as in **a**. While differences between medians are not significant (Kruskal–Wallis *p* value 0.7291), a greater variance can be seen, particularly in heart. **c** Verification that pyrosequencing assays for both known and putative maternal gDMR showed low methylation in sperm, while the paternal assay returned high methylation levels. **d** Methylation is lost at all known and putative gDMR, though the decrease is very small at *Dlk1-Gtl2 IG*. *Error bars* represent s.e.m.; $^{**}p < 0.01$; $^{***}p < 0.001$; *n.s.* not significant

Methylation can be restored following loss of DNMT3A/B at the majority of imprinted gDMR in DNMT3A2-rescued ES cells

Our initial work reported above established a clear ability for DNMT3A2 to restore methylation marks at key imprinted gDMRs such as *Igf2r* in 3abKO cells. We now extended this analysis to the other known and putative gDMRs indicated above. For convenience, all of the assays from our work are presented together in Fig. 6a. Eight of 10 known gDMRs assayed gained methylation when compared to 3abKO (Fig. 6a), with only *Grb10*, in addition to *KvDMR*, showing a failure to regain methylation. Of the four putative gDMR assayed, only one (*6330408a02Rik*) did not show any increase in methylation. The maternal

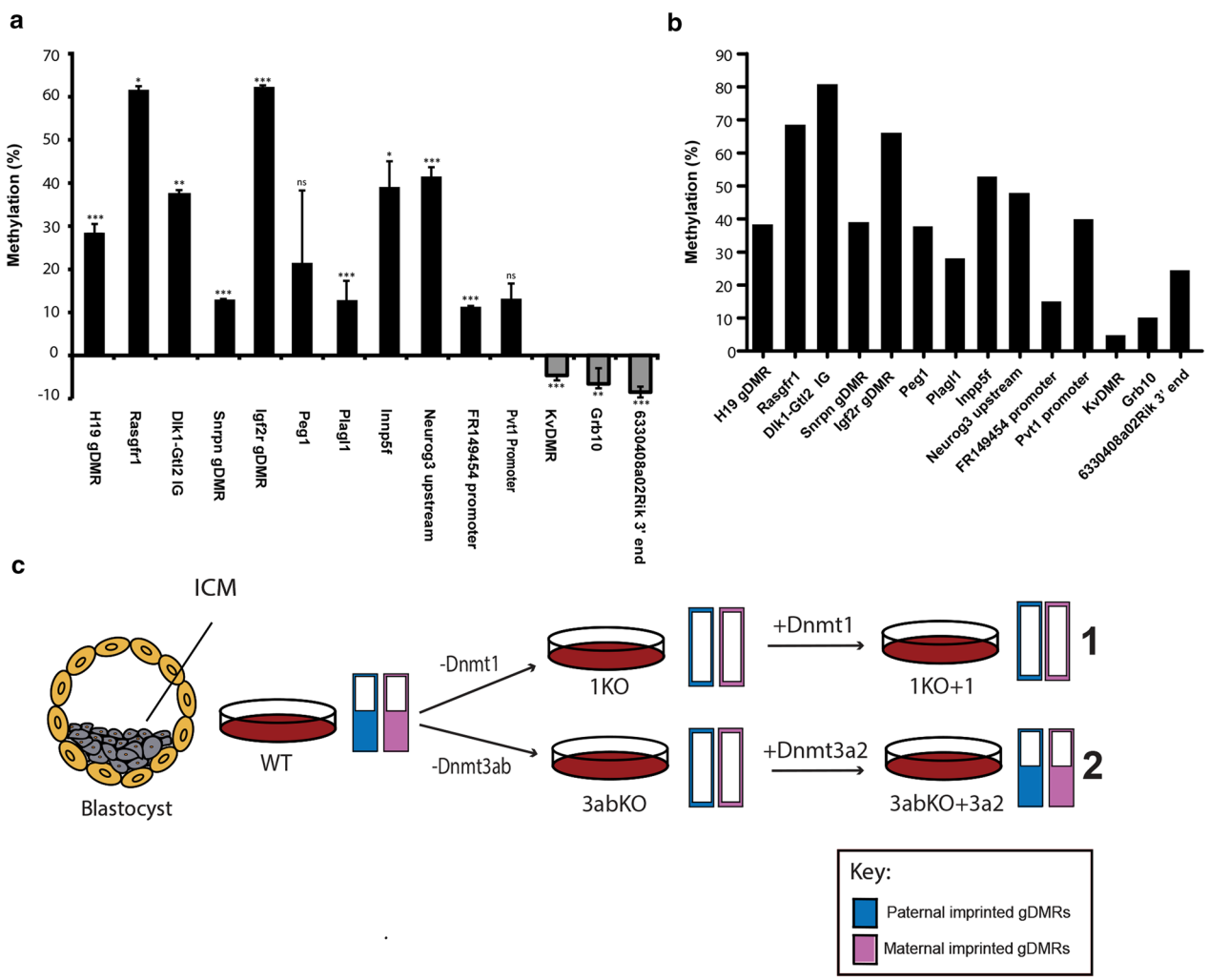


Fig. 6 Summary of methylation responses at known and putative imprinted gDMR. **a** Changes in methylation seen in 3abKO cells rescued with DNMT3A2 for all of the known and putative gDMRs. Eleven gDMRs of fourteen which could be assayed showed gains in methylation, with nine of these reaching significance. *KvDMR, Grb10* and *6330408a02Rik* failed to recover methylation levels, instead showing significant additional reductions in methylation when compared to the 3abKO ES cells. **b** Absolute methylation levels in the 3abKO + 3a2 cells at the various gDMR. **c** Schematic summarising the changes in methylation seen in the two types of knockout and rescue. WT ESC cells grown in petri dishes were derived originally from inner cell mass (ICM) of early embryo and retained 50% methylation at most imprinted gDMR (*half-filled bars*: paternal at *left*, maternal at *right*). While loss of DNMT1 (pathway 1, *top*) gave comparable hypomethylation to loss of DNMT3A/B (pathway 2, *bottom*), no recovery of methylation at either paternally or maternally methylated imprinted gDMRs was seen in DNMT1-rescued cells, whereas rescue with DNMT3A2 in 3abKO cells restored methylation non-discriminately at both paternally and maternally imprinted gDMRs (pathway 2)

gDMR *Igf2r* showed the largest recovery of methylation when compared to the 3abKO ES cells at 62.30%. Gain in methylation was seen at all the known paternally methylated gDMRs assayed-*H19*, *Rasgrf1* and *Dlk1-Gtl2 IG* (Fig. 6a). Absolute methylation values for the 3abKO + 3a2 cells are shown in Fig. 6b for comparison. Notably, *KvDMR, Grb10* and *6330408a02Rik 3′* end not only fail to regain methylation (Fig. 6a), but instead continue to lose it in the 3abKO + 3a2 cell. This suggests that in the absence of DNMT3A/B the loci do not remain stable but rather continue to lose methylation (Fig. 6a). Our findings for the individual loci are summarised in Fig. 6c and in Table 1.

Discussion

Maintenance methylation is a vital process as it is responsible for the stable inheritance of this epigenetic signature from mother to daughter cells during the process of mitosis. At one time, DNMT1 was thought to be the only enzyme associated with maintenance of methylation

Table 1 Summary of findings with regard to methylation at gametic differentially methylated regions at imprinted loci

Locus	Chr	Chromosomal region[a] delineated by pyro assay	Origin[b]	gDMR status[c]	Sperm meth %	Somatic meth %[d]	Location within gene	CpG island	Gain in 3abKO + 3a2 ESCs
H19	7	142,580,262–142,580,434	P	Known	84.19	51.06	Intergenic	No	Yes
Rasgrf1	9	89,872,365–89,872,512	P	Known	89.00	52.67	Intergenic	No	Yes
Dlk1-Gtl2 IG	12	109,528,521–109,528,661	P	Known	96.33	57.92	Intergenic	Yes	Yes
Snrpn	7	60,004,993–60,005,163	M	Known	4.41	43.42	Promoter/Exon 1	Yes	Yes
Igf2r	17	12,960,690–12,962,806	M	Known	3.00	45.71	Intronic	Yes	Yes
Peg1	6	30,687,444–30,688,524	M	Known	15.50	51.75	Intronic	Yes	Yes
Plagl1	10	13,091,014–13,091,154	M	Known	10.80	42.83	Promoter/Exon 1	Yes	Yes
Inpp5f	7	128,688,173–128,688,290	M	Known	13.10	57.19	Intronic	Yes	Yes
Grb10	11	12,025,894–12,026,044	M	Known	7.38	45.80	Intronic	Yes	No
KvDMR	7	143,295,771–143,295,910	M	Known	6.33	50.54	Intronic	Yes	No
6330408a02Rik 3′	7	13,260,963–13,261,135	M	Putative	19.43	47.71	Exon 13	Yes	No
Neurog3 upstream	10	62,127,922–62,128,093	M	Putative	15.44	43.84	Intragenic	No	Yes
FR149454 promoter	11	119,258,958–119,259,182	M	Putative	15.33	58.67	Intronic	No	Yes
Pvt1 promoter	15	62,037,136–62,037,311	M	Putative	19.42	43.30	Intragenic	No	Yes

Data are presented for each gDMR for which a validated pyrosequencing assay (pyro assay) could be established. Known gDMR are listed first and then putative, with paternal imprints preceding maternal (none of the putative gDMRs were paternal)

Chr chromosome, *Origin* parent of origin of methylation mark, *meth* methylation

[a] mm10 release

[b] Parental origin of methylation: *p* paternal chromosome, *m* maternal

[c] Whether gDMR is well characterised (known) or recently discovered (putative)

[d] Somatic methylation = average methylation value across three adult tissues

due to its preferential binding to hemi-methylated DNA and its presence at the replication foci [4, 5]. Chen et al. [39] showed, however, that DNMT3A and DNMT3B were also important for maintenance methylation at some repeats and, using a qualitative technique, at certain imprinted loci. In a previous study, we confirmed that DNMT3A and 3B were needed at a few selected imprinted loci using a more quantitative approach and extended this observation to transiently imprinted genes, which also require DNMT3A/B for maintenance in ESCs [13]. Here we looked in greater depth at all the known gametic DMR as well as some newly identified imprinted gDMR and confirm their reliance (with 1–2 exceptions such as *Dlk1-Gtl2* IG) on the DNMT3A/B enzymes for maintenance of methylation. Interestingly, the decrease in methylation at these gDMRs was found to be approximately the same in 1KO and 3abKO cells, suggesting an equal contribution by DNMT3A/B and DNMT1 in maintenance of methylation at imprinted gDMRs in ESCs.

Overexpression of DNMT1 in 1KO cells resulted in a global increase in methylation as reported previously; similar global increases in methylation were observed here in DNMT3A/B rescued cells using LUMA, although this increase does not bring the methylation level to the normal WT level globally. This could be due to a number of reasons: (1) it may indicate the presence of some sequences which are refractory to remethylation in ESCs; (2) the expression of the cDNA in the rescued cells may

not be as high as the endogenous levels of DNMT3A2; and/or (3) some sequences may require both DNMT3B and DNMT3A2 to fully recover methylation to WT levels [11].

We showed here that in the 1KO + 1 cells there was no gain in methylation seen at any of the gDMR examined, confirming earlier results from a number of groups. In contrast, 3abKO cells rescued with DNMT3A2 showed clear and reproducible gains in methylation at the majority of imprinted gDMR. These results were confirmed using up to three techniques per locus-pyrosequencing, clonal analysis and COBRA. Additionally, the transcriptional status of *H19* and *Igf2* responded appropriately to the loss and regain of methylation, confirming that functional imprinting was being affected, at least at these loci (other loci showed transcription levels which were too low to reliably quantitate in these cells). While some previous studies have found that none [38] or only one [39] imprinted locus showed any gains in methylation on rescue, these were based on more qualitative techniques and in many cases could not examine the locus except at a low level of resolution using techniques such as Southern blotting. Here we show gain in methylation of greater than 10% at 11/14 gDMR, with substantially greater gains at most. Average gain was 28%, which in the context of an incomplete overall rescue as indicated from the global methylation levels (above) represents a corrected gain of close to 50%.

There were a few loci (3/14-*KvDMR*, *Grb10* and *6330408a02Rik* 3′ end), which showed no gain in methylation, and in fact displayed evidence for further hypomethylation relative to the 3abKO cells. This latter is not unexpected since we have shown that ESC rely on DNMT3A/B for maintenance methylation as well as de novo activity, so if these three loci are refractory to the action of DNMT3A2 in the rescued line, they would be expected to continue to lose methylation. Our examination of ENCODE data and of the current literature has found so far no common denominator for these three loci. Nevertheless, these results show that for the majority of imprinted loci, methylation at the gDMR, and in some cases functional imprinting, can be restored in a somatic cell type without germline passage.

What mechanism is associated with imprint recovery in 3abKO + 3a2 and not 1KO + 1 cells? This is particularly puzzling since the two rescued cell types both have all three enzymes present. Two possibilities are that (1) loss of DNMT3A/B proteins could alter histone marks on chromatin, which then act to attract de novo methylation by DNMT3A2 on rescue or (2) loss of DNMT1 protein causes a change in histone marks, which mean that even after rescue, the DNA cannot be remethylated. Notably, triple KO cell lines lacking all three enzymes also fail to show imprint restoration when rescued [39], suggesting that it is the loss of DNMT1 which leads to an irreversible change in epigenetic potential, precluding rescue with DNMT3A2. It has been reported that the loss of DNMT1 results in loss of H3K9me3 in ESC [50]. One possibility is that loss of H3K9me3 occurs in 1KO cells, but not 3abKO cells, and that the presence of this mark facilitates remethylation by DNMT3A2 in the latter. It has also been reported that the PWWP domain of DNMT3A is linked with targeting of chromatin carrying H3K36me3 [51]. Loss and gain of methylation marks on imprinted gDMRs could be due to the presence and absence of such interactions between methyltransferases and histone marks associated with chromatin, which require further experimental exploration in this system.

We clearly identified three gDMR, including *KvDMR*, where methylation once lost cannot be recovered. This supports other evidence, suggesting that mechanism of imprinting and response to methylation loss and recovery can vary among imprinted genes [52]. In future, it will be interesting to compare the histone marks associated with *KvDMR* and with those associated with gDMRs that recover methylation in rescued cells. The *Dlk1-Gtl2* IG was also interesting in that it showed overmethylation in our experiments, gaining almost 40% methylation in 3abKO + 3a2. The tendency of this locus to become hypermethylated in human ES and iPS cells has been noted before [53, 54] and may reflect some fundamental mechanistic feature of imprinting at this locus, which in practise could act as a barrier to somatic reprogramming efforts.

During the course of writing, a paper from the Wong group investigating the behaviour of UHRF1 rescue cells found that a number of imprinted genes showed gains in methylation in that system too [55]. Methylation gain was only seen at some of the imprinted loci, and there was no clear link to the location of the gDMR, the presence of antisense transcripts or the type of imprint. Furthermore, they investigated common histone marks and found no relationship between any specific mark and the ability of the locus to gain methylation in the rescues. They did not, however, investigate transcriptional changes at the loci in their cells. Their data, taken together with the findings we present here, show that gametically acquired methylation at imprinted loci can be reset somatically in certain circumstances.

Conclusions

We have shown that (1) both DNMT1 and DNMT3A/B loss generate similar methylation changes at imprinted gDMRs in ESCs; (2) recovery of imprints in 1KO + 1 cell lines is not possible but imprints can be recovered in DNMT3A2-rescued 3abKO cells; and (3) there are some

exceptional gDMRs where imprints, once lost, cannot be re-established. Our findings highlight important differences between the two cell systems and indicate that it may be possible to restore imprints somatically under certain circumstances, an observation of clear relevance for imprinting disorders. This may provide a useful model system in which to further explore reprogramming.

Additional files

Additional file 1: Table S1. Details of the primers used for RT-qPCR during the study.

Additional file 2: Table S2. Details of the primers used for pyrosequencing during the study. The gametic differentially methylated region (gDMR) is given at left. Some primer sets are commercially available from Qiagen, as indicated. The sequence of the unconverted DNA, as well as the bisulfite-converted sequence, is given for ease of identification. Which primer carried the biotin modification is also indicated.

Additional file 3: Figure S1. Controls to confirm the methyltransferase activities in the ESC used during the study. (A) Westerns showing the presence or absence of DNMT1 (top) or of DNMT3A2 (bottom) in the various cell lines indicated. ACTB and GAPDH are loading controls. The size of the expected protein is shown at left. A number of DNMT3B antibodies tried proved unreliable. (B) RT-PCR showing the presence or absence of transcripts for the various enzymes in the ESC used. The expected sizes are shown at left. *Hprt* was a loading control.

Authors' contributions

AT made the initial observations, carried out wet laboratory and statistical analyses of the first five loci, assembled figures and wrote the first draft. SJM carried out bioinformatics and wet laboratory analyses of all other loci, did statistical analysis, assembled figures and contributed to writing. REI carried out all westerns. From a paper ran RT-qPCR and assembled figures. GP carried out RT-PCR and RT-qPCR. KON carried out some cell culture, supervised laboratory work and contributed to design and writing. CPW conceived and designed the study, interpreted results and wrote the MS. All authors read and approved the final manuscript.

Author details

[1] Genomic Medicine Research Group, Biomedical Sciences Research Institute, Centre for Molecular Biosciences, University of Ulster, Coleraine BT52 1SA, UK. [2] Present Address: Terry Fox Laboratory, BC Cancer Agency, 675 W 10th Ave, Vancouver, BC V5Z 1G1, Canada. [3] Present Address: Centre for Experimental Medicine, The Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7AE, UK.

Acknowledgements

We are very grateful to Masaki Okano for the gift of cells. We thank Philip Logue, Rhonda Black, Bernie McKay and Keith Thomas for technical support and members of the Genomic Medicine Research Group for comments.

Competing interests

The authors declare that they have no competing interests.

Availability of data and material

All data generated or analysed during this study are included in this published article (and its supplementary information files).

Ethics approval and consent

Animals were housed and kept in accordance with UK Home Office regulations and relevant University Institutional guidance. Ethical approval was obtained from the University of Ulster Ethics Committee.

Funding

Work in the C.P.W. laboratory was supported by a Grant from the Medical Research Council (MR/J007773/1). AT was the recipient of a Vice Chancellor's Research Studentship from Ulster University.

Received: 9 August 2016 Accepted: 8 November 2016
Published online: 22 November 2016

References

1. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev. 2013;14:204–20.
2. Yoder JA, Bestor TH. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet. 1998;7:279–84.
3. Bestor TH. The DNA methyltransferases of mammals. Hum Mol Genet. 2000;9:2395–402.
4. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 1992;71:865–73.
5. Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 1996;122:3195–205.
6. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004;279:48350–9.
7. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19:219–20.
8. Chen T, Ueda Y, Xie S, Li E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem. 2002;277:38746–54.
9. Robert M-F, Morin S, Beaulieu N, Gauthier F, Chute IC, Barsalou A, et al. DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. Nat Genet. 2003;33:61–5.
10. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429:900–3.
11. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. 2010;42:1093–100. http://www.nature.com/doifinder/10.1038/ng.708.
12. Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. 2003.
13. Rutledge CE, Thakur A, O'Neill KM, Irwin RE, Sato S, Hata K, et al. Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes. Development. 2014;141:1313–23.
14. Olek A, Walter J. The pre-implantation ontogeny of the H19 methylation imprint. Nat Genet. 1997;17:275–6.
15. Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, et al. Programming and inheritance of parental DNA methylomes in mammals. Cell. 2014;157:979–91.
16. Shemer R, Birger Y, Riggs AD, Razin A. Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. Proc Natl Acad Sci USA. 1997;94:10267–72.
17. Guo F, Li X, Liang D, Li T, Zhu P, Guo H, et al. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell. 2014;15:447–58. http://www.sciencedirect.com/science/article/pii/S1934590914003415.
18. Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature. 2012;484:339–44.
19. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99:247–57.

20. Stewart CL, Stuhlmann H, Jähner D, Jaenisch R. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc Natl Acad Sci USA. 1982;79:4098–102.
21. Santos F, Hendrich B, Reik W, Dean W. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol. 2002;241:172–82.
22. Bartolomei MS, Ferguson-Smith AC. Mammalian genomic imprinting. Cold Spring Harb Perspect Biol. 2011;3:1–17.
23. John RM, Lefebvre L. Developmental regulation of somatic imprints. Differentiation. 2011;81:270–80.
24. Davis TL, Yang GJ, McCarrey JR, Bartolomei MS. The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. Hum Mol Genet. 2000;9:2885–94.
25. Adalsteinsson B, Ferguson-Smith A. Epigenetic control of the genome—lessons from genomic imprinting. Genes. 2014;5:635–55.
26. Smallwood SA, Kelsey G. De novo DNA methylation: a germ cell perspective. Trends Genet. 2012;28:33–42.
27. Sutcliffe JS, Nakao M, Christian S, Orstavik KH, Tommerup N, Ledbetter DH, et al. Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. Nat Genet. 1994;8:52–8.
28. Buiting K, Saitoh S, Gross S, Dittrich B, Schwartz S, Nicholls RD, et al. Inherited microdeletions in the Angelman and Prader–Willi syndromes define an imprinting centre on human chromosome 15. Nat Genet. 1995;9:395–400.
29. Zwart R, Sleutels F, Wutz A, Schinkel AH, Barlow DP. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev. 2001;15:2361–6.
30. Bartolomei MS. Genomic imprinting: employing and avoiding epigenetic processes. Genes Dev. 2009;23:2124–33.
31. Forne T, Oswald J, Dean W, Saam JR, Bailleul B, Dandolo L, et al. Loss of the maternal H19 gene induces changes in Igf2 methylation in both cis and trans. Proc Natl Acad Sci USA. 1997;94:10243–8.
32. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001;294:2536–9.
33. Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002;129:1983–93.
34. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429:900–3.
35. Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, et al. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet. 2007;16:2272–80.
36. Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011;43:811–4.
37. Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, et al. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature. 2007;448:318–24.
38. Tucker KL, Beard C, Dausmann J, Jackson-Grusby L, Laird PW, Lei H, et al. Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev. 1996;10:1008–20.
39. Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 2003;23:5594–605.
40. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol. 2000;25:169–93.
41. Walsh CP, Bestor TH. Cytosine methylation and mammalian development. Genes Dev. 1999;13:26–34. http://www.genesdev.org/cgi/content/full/13/1/26.
42. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005;15:1451–5.
43. Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature. 2002;415:810–3.
44. Mancini-Dinardo D, Steele SJS, Levorse JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev. 2006;20:1268–82.
45. Shin J-Y, Fitzpatrick GV, Higgins MJ. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J. 2008;27:168–78.
46. Woodfine K, Huddleston JE, Murrell A. Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. Epigenetics Chromatin. 2011;4:1.
47. Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells. 2006;11:805–14.
48. Oda M, Yamagiwa A, Yamamoto S, Nakayama T, Tsumura A, Sasaki H, et al. DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. Genes Dev. 2006;20:3382–94.
49. Karimi M, Luttropp K, Ekström TJ. Global DNA methylation analysis using the luminometric methylation assay. Methods Mol Biol. 2011;791:135–44.
50. Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell. 2011;44:361–72.
51. Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, et al. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem. 2004;279:25447–54.
52. Bartolomei MS, Ferguson-Smith AC. Mammalian genomic imprinting. Cold Spring Harb Perspect Biol. 2011;3. doi:10.1101/cshperspect.a002592.
53. Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature. 2010;465:175–81.
54. Mo C-F, Wu F-C, Tai K-Y, Chang W-C, Chang K-W, Kuo H-C, et al. Loss of non-coding RNA expression from the DLK1-DIO3 imprinted locus correlates with reduced neural differentiation potential in human embryonic stem cell lines. Stem Cell Res Ther. 2015;6:1. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4417332&tool=pmcentrez&rendertype=abstract\nhttp://www.ncbi.nlm.nih.gov/pubmed/25559585\nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4417332.
55. Qi S, Wang Z, Li P, Wu Q, Shi T, Li J, et al. Non-germ line restoration of genomic imprinting for a small subset of imprinted genes in ubiquitin-like PHD and RING finger domain-containing 1 (Uhrf1) null mouse embryonic stem cells. J Biol Chem. 2015;290:14181–91.

PAPER-II

Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression.

Karla M. O'Neill, Rachelle Irwin, Sarah-Jayne Mackin, Sara-Jayne Thursby, Avinash Thakur Ciske Bertens, Laura Masala, Jayne Loughery, Darragh McArt and Colum P. Walsh.

Epigenetics & Chromatin201811:12

https://doi.org/10.1186/s13072-018-0182-4

The main aims of this paper were to:

1. To develop the first non-cancerous human differentiated cells stably depleted in the main maintenance methyltransferase DNMT1.
2. To look for classes of genes sensitive to the genome-wide changes in methylation consequent to DNMT1 depletion.
3. To correlate transcriptional alterations and methylation changes in the cells.
4. To determine if possible, why specific regions become differentially methylated and not others.

CONTRIBUTION

I performed an initial bioinformatic analysis of the methylation levels of the DNMT1-depleted clonally derived cell lines to help generate lists of genes that were differentially methylated when compared to WT hTERT-1604 cells. This formed the basis of two previous versions of this manuscruipt and the submission to two different journals, with respective rounds of revision. I carried out gene ontology analyses on the enriched genes that displayed significant differential methylation when compared to WThTERT-1604 cells such as that presented in Fig. 1F. Upon the lists of enriched genes, I performed downstream statistical analyses on the gene classes that were identified to be enriched in both the methylation and transcriptional array data and modelled this data using the Galaxy Interface to create user defined tracks on UCSC Genome Browser. I helped to design a panel of pyrosequencing assays for the olfactory receptors, grew all the cell lines in culture, extracted and bisulfite converted the DNA for

quantitative methylation analyses. To comply with submission guidelines, I created and deposited all the IDAT file data for both the 450k methylation and HT12 transcription array data for the manuscript onto the NCBI GEO archive along with supplemental information to promote reproducibility and data sharing under the terms of Open Access. I did not have any input on the polycomb data presented in the paper.

Summary of the major findings

1. The KD of DNMT1 in the cell lines examined highlighted enriched groups of genes sensitive to this loss of methylation- the protocaderin genes (PCDH), fat homeostasis/body mass genes, olfactory receptors (OR), cancer testis antigen (CTA) genes and those of the UGT1A complex.

2. The methylation levels of both the fat/body mass genes and the UGT1A genes were affected by both DNA methylation and the presence of polycomb marks.

3. Widespread changes in methylation in each of the KD cell lines was not accompanied by large-scale transcriptional derepression, with only a few hundred genes showing dysregulation, and the fold change in transcription being small.

O'Neill *et al. Epigenetics & Chromatin (2018) 11:12*
https://doi.org/10.1186/s13072-018-0182-4

Epigenetics & Chromatin

RESEARCH

Open Access

Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression

Karla M. O'Neill[1,5†], Rachelle E. Irwin[1†], Sarah-Jayne Mackin[1], Sara-Jayne Thursby[1], Avinash Thakur[1,6], Ciske Bertens[1,2], Laura Masala[1,3], Jayne E. P. Loughery[1], Darragh G. McArt[4] and Colum P. Walsh[1*]

Abstract

Background: DNA methylation plays a vital role in the cell, but loss-of-function mutations of the maintenance methyltransferase *DNMT1* in normal human cells are lethal, precluding target identification, and existing hypomorphic lines are tumour cells. We generated instead a hypomorphic series in normal hTERT-immortalised fibroblasts using stably integrated short hairpin RNA.

Results: Approximately two-thirds of sites showed demethylation as expected, with one-third showing hypermethylation, and targets were shared between the three independently derived lines. Enrichment analysis indicated significant losses at promoters and gene bodies with four gene classes most affected: (1) protocadherins, which are key to neural cell identity; (2) genes involved in fat homoeostasis/body mass determination; (3) olfactory receptors and (4) cancer/testis antigen (CTA) genes. Overall effects on transcription were relatively small in these fibroblasts, but CTA genes showed robust derepression. Comparison with siRNA-treated cells indicated that shRNA lines show substantial remethylation over time. Regions showing persistent hypomethylation in the shRNA lines were associated with polycomb repression and were derepressed on addition of an EZH2 inhibitor. Persistent hypermethylation in shRNA lines was, in contrast, associated with poised promoters.

Conclusions: We have assessed for the first time the effects of chronic depletion of DNMT1 in an untransformed, differentiated human cell type. Our results suggest polycomb marking blocks remethylation and indicate the sensitivity of key neural, adipose and cancer-associated genes to loss of maintenance methylation activity.

Keywords: DNMT1, EZH2, Protocadherin, Body mass, Cancer/testis antigen

Background

DNA methylation is an important mechanism for epigenetic regulation of genes in both mouse and human [1]. It occurs mainly at the CpG dinucleotide, and methylation at this symmetrical site is efficiently maintained during replication by the action of the DNA methyltransferase 1 (DNMT1) enzyme [2]. Methylation is known to play an important role in regulating imprinted loci [3], genes on the inactive X chromosome [4] and germline-specific genes [5] in mouse.

Where methylation occurs at the promoter of a gene, it is strongly associated with the silencing of transcription, particularly if there is a high density of CpGs, a so-called CpG island (CGI). However, studies have shown that most CGI are intrinsically protected from methylation [6, 7] and only a small number shows dynamic changes during development, mostly in the three classes mentioned above [5, 8], though there may be others which have not

*Correspondence: cp.walsh@ulster.ac.uk
†Karla M. O'Neill and Rachelle E. Irwin contributed equally to this work
[1] Genomic Medicine Research Group, Centre for Molecular Biosciences, School of Biomedical Sciences, Ulster University, Cromore Road, Coleraine BT52 1SA, UK
Full list of author information is available at the end of the article

yet been clearly defined. As you move outward from an island, the shores and shelves show higher levels of methylation and greater dynamic response [9], though here the link to changes in gene activity is less clear [10]. Methylation is also associated with larger regions of inert chromatin, such as the inactive X, pericentromeric repeats and regions rich in transposable elements [1], generally consistent with a repressive role. Recent genome-wide surveys have also indicated that high levels of methylation are found in the bodies of active genes, where they may facilitate transcription [11, 12]. In keeping with this, we and others recently showed that artificially decreasing intragenic methylation levels reduced steady-state transcript levels, consistent with a positive role for methylation in the gene body [11–13].

Another major system for epigenetic repression is via histone modification, particularly by the polycomb group of proteins, with EZH2 being one of the main enzymes involved [14]. A number of studies suggest an interplay between polycomb- and DNMT-mediated repression, with a generally negative correlation between DNA methylation and the H3K27me3 mark deposited by EZH2 [15, 16]. Supporting this, a loss of DNA methylation caused a reshaping of the histone landscape and derepression of some polycomb targets in mouse ES cells [17], suggesting that DNA methylation helps to determine where polycomb marks are deposited.

While DNMT1 is the main maintenance methyltransferase, there also appears to be an important role for the de novo enzymes DNMT3A and DNMT3B in complementing that activity at some loci [18, 19]. In order to clarify which genes are most sensitive to DNMT1 loss in human, a number of studies have been carried out using mutations within the gene to assess the effects of loss of methylation [19–22]. While this has been a fruitful approach in mouse embryonic stem (ES) cells, where null mutants are tolerated, differentiation of the mouse cells leads to cell death [20, 22, 23], whereas DNMT1 disruption in human ES cells is not tolerated even in undifferentiated cells [24]. Genetic ablation in adult differentiated cells also leads to cell death within a few cell cycles, before passive demethylation of the genome can occur [23, 25]. One of the best-studied systems in humans consists of HCT116 colon cancer cells carrying a hypomorphic allele in the DNMT1 gene together with a DNMT3B knockout (HCT116 DKO cells) [26–28]. Blattler et al. [29] found that there was widespread and relatively uniform demethylation across the genome in the DKO cells, with small effects at CGI (most of which are normally unmethylated anyway) and relatively few genes showing derepression. There was no enrichment by gene ontology (GO) analysis, but some effect at enhancers: however, this is complicated by the presence of the DNMT3B knockout alleles. Acute depletion of DNMT1 using an siRNA-mediated approach in embryonal carcinoma cells also found regions of low CpG density (open sea, shelf) to be the most affected by loss of methylation [70]. Among the small number of dysregulated genes, there was some enrichment for cell morphogenesis and phosphorylation pathways.

Neither of these cancer cell lines, however, are a good model for the normal differentiated cell as they are transformed, aneuploid, hypermethylated, and contain a number of different mutations in key regulatory genes. Additionally, acute depletion of DNMT1 results in cell cycle delay, triggering of the DNA damage response and increased rates of cell death [24, 25, 30], making it difficult to separate acute and chronic effects.

To circumvent some of the difficulties outlined above, we generated a series of isogenic human cell lines derived from the hTERT-immortalised normal fibroblast line hTERT1604 as previously described [30]. These are normosomic and non-transformed, and by using a stably incorporated plasmid with an shRNA targeting *DNMT1* we were able to isolate a number of clonally derived lines to allow identification of any cell line-specific effects. While these showed initially the range of shared features indicative of a global response to the loss of this critical regulator, including cell cycle delay, demethylation of imprinted genes and others, they could be cultured for longer under selection [30], allowing identification of loci with particular sensitivity for decreased maintenance methyltransferase activity. Here we set out to completely characterise the methylation changes seen in the cell lines using the Illumina Infinium HumanMethylation450 BeadChip (450k) array platform [31] and subsequent analysis using the RnBeads pipeline [32]. These approaches were chosen due to their high reproducibility and low inter-operator variability, ensuring the reliable and sensitive detection of alterations in methylation. A sample of the observations was then further verified using locus-specific assays. In addition and for the same reasons, we used the HT-12 Expression v4 BeadChip array, to assay changes in transcription in our cell lines.

Methods

Cell culture

The parental or wild-type (WT) adherent hTERT1604 lung fibroblast cell line [33] was cultured in 4.5 g/l glucose DMEM (Thermofisher, Loughborough, UK) supplemented with 10% FBS and 2× NEAA (Gibco/Thermofisher). Generation of the hTERT1604 cell lines stably depleted of DNMT1 using a pSilencer construct (Thermofisher) has been previously described [30]. Knockdown (KD) cells were maintained as for WT, but medium was supplemented with 150 μg/ml hygromycin

B (Invitrogen/Thermofisher, Paisley, UK), which was removed at least 48 h before any experimental procedure. Treatment of cells with siRNA for 24 h was as previously described [34]: for the pulse-chase experiment cells were afterwards allowed to recover in normal media and passaged as required for up to 36 days. The siRNA (Dharmacon ON-TARGETplus SMARTpool) for *DNMT1* and *DNMT3B*, as well as scrambled control, was obtained from Invitrogen/Thermofisher. HCT116 and double knockout (DKO) cells [27] were cultured in 1 g/l glucose DMEM (Gibco) supplemented with 10% FBS and 1× NEAA (Gibco). DZNeP (Sigma-Aldrich, Dorset, UK) was used at a final concentration of 1 μM.

DNA extraction and bisulphite conversion

Genomic DNA was harvested from cells in log phase of growth. Samples were incubated overnight at 55 °C in lysis buffer [50 mM Tris pH 8, 0.1 M EDTA (both Sigma-Aldrich), 0.5% SDS, 0.2 mg/ml proteinase K (Roche, West Sussex, UK)], with rotation, and DNA was subsequently isolated using the standard phenol/chloroform/isoamyl alcohol (25:24:1 pH8, Sigma-Aldrich) extraction method. DNA quality was verified using gel electrophoresis and UV absorbance measurements at 260/280 and 260/230 nm using a Nanodrop UV spectrophotometer (Labtech International, Ringmer, UK). Bisulphite conversion of 500 ng of DNA was carried out using the EpiTect bisulphite kit (Qiagen, Crawley, UK) according to the manufacturer's instructions.

Hybridisation to 450K array and bioinformatic analyses

Three samples from each cell line were used to prepare DNA, with at least one biological repeat in each set. DNA was assessed for purity and integrity as above prior to quantification using the Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific) as per manufacturer's instructions. In total, 500 ng of high-quality bisulphite-converted (Zymo Research) DNA was checked for purity and fragmentation on a bioanalyser and then loaded on the Infinium HumanMethylation450 BeadChip [31] and imaged using an Illumina iScan (Cambridge Genomic Services). Output files in IDAT format were processed using the RnBeads [32] methylation analysis package (v1.0.0) which carries out all the analysis from import to differential methylation within the R platform (3.2.0). Briefly, quality control used the built-in probes on the array and included filtering out of probes containing SNPs, and checking for hybridisation performance. Normalisation was then carried out using the SWAN method in minfi [35] after background subtraction with methylumi.noob. The exploratory analysis module was used to generate probe density distributions and scatter graphs. The differential methylation analyses was based on a combined ranking score, which combined absolute effect size, relative effect sizes and p-values from statistical modelling into one score where rank is computed as the most conservative value among mean difference in means, mean in quotients and combined p value across sites in the region: the enrichment analysis used the combined rank among the 1000 best-ranking regions and a hypergeometric test to identify GO terms in the AmiGO 2 database [36]. Pairwise comparison of triplicate samples from each cell line against WT hTERT was also made to determine change in beta value and associated combined p-value, adjusted for multiple comparison using false discovery rate (FDR). Some tailored analyses were also carried out using custom scripts in R. Additional GO studies were performed using DAVID (v6.7) [37].

We used the GALAXY platform [38] to map sites showing highly reproducible changes (FDR < 0.05) against the locations of RefSeq genes or ChromHMM regions on the UCSC genome browser [39] for each cell line. GO category genes which showed changes in methylation at multiple sites in more than one KD cell line were scored as true hits (Yes in the FDR column), while GO categories with few or no sites reproducibly altered across replicates (FDR > 0.05) or where methylation changes were small (< 0.1 β), inconsistent in direction, or not found in more than one KD cell line, were scored as false positives. Absolute β levels were used to measure median methylation across genes of interest using custom workflows in GALAXY, with further statistical analyses in Statistical Package for the Social Sciences software (SPSS) version 22.0 (SPSS UK Ltd).

Locus-specific methylation analysis

Amplification was carried out using the PyroMark PCR kit (Qiagen) with 2 μl bisulphite-converted DNA, 12.5 μl MasterMix, 2.5 μl CoralLoad Concentrate, 1.25 μl each primer (10 μM) and 5.5 μl nuclease-free H_2o using the following conditions: 15 min at 95 °C followed by 45 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 30 s and a final elongation step of 72 °C for 10 min. Pyrosequencing was carried out on the PyroMark Q24 System, according to the manufacturer's instructions (Qiagen). Most assays were designed in-house using the PyroMark Assay Design software 2.0 (*LEP*, *MAGEA12*, *OR10J5*, *OR51E2*, *OR2AG1*, *PCDHA2*, *PCDHC4*, *UGT1A1*, *UGT1A4*) prior to synthesis (Metabion, Germany): see Additional file 1: Table S1 for details: *DAZL*, *SYCP3*, *D4Z4* and *NBL2* were as described [34, 40]. In some cases, pre-designed pyrosequencing primers were obtained from Qiagen (*GABRQ* PM00133483, *GHSR* PM00014350, *SNRPN* PM00168252). Clonal analysis was carried out as previously described [30].

Hybridisation to HT-12 microarray and bioinformatic analyses

Total RNA was extracted using the RNeasy minikit (Qiagen) as per manufacturer's instructions, including a DNase step. RNA integrity was verified via gel electrophoresis, and quality and quantity were verified using a SpectroStar (BMG Labtech, Aylesbury, UK) and a bioanalyser (Agilent Technologies, Cheadle, UK). Two hundred nanograms of total RNA underwent linear amplification using the Illumina TotalPrep RNA Amplification Kit (Life Technologies/Thermofisher, Paisley, UK) following the manufacturer's instructions. Microarray experiments were performed at Cambridge Genomic Services, University of Cambridge, using the HumanHT-12 v4 Expression BeadChip (Illumina, Chesterford, UK). After scanning the data were loaded in GenomeStudio (Illumina) and then processed in R (version 3.2.2). The data were filtered to remove any non-expressed probes using the detection p-value from Illumina, transformed using the variance stabilization transformation (VST) from lumi and normalised using the quantile method. Comparisons were made using the limma package with results corrected for multiple testing using false discovery rate (FDR) testing.

RNA and protein analysis

Transcriptional assays at individual loci using RT- and RT-qPCR were carried out essentially as in [34]: primer sequences are listed in Additional file 1: Table S1. Protein was extracted from cells growing in log phase using protein extraction buffer (50 mM Tris–HCl, 150 mM NaCl, 1% Triton-X, 10% glycerol, 5 mM EDTA; all Sigma-Aldrich) and 0.5 µl protease inhibitor mix (Sigma-Aldrich). For Western blotting, 30 µg protein was denatured in the presence of 5 µl 4× LDS sample buffer (Invitrogen) and 2 µl 10× reducing agent (Invitrogen) in a total volume of 20 µl nuclease-free water (Qiagen) via incubation at 70 °C. Proteins were separated by SDS-PAGE and then electroblotted onto a nitrocellulose membrane (Invitrogen) and blocked in 5% non-fat milk for 1 h at room temperature (RT). Membranes were incubated with anti-DNMT1 (a kind gift from Guoliang Xu) and anti-β-actin (Abcam ab8226) overnight at 4 °C, followed by HRP-conjugated secondary antibody incubation at RT using ECL (Invitrogen).

Statistical analysis

Statistical analysis was performed by the RnBeads package, or separately in Excel (Microsoft Office Professional Plus 2013), Prism (Graphpad) or SPSS (v22.0). Experiments were carried out in triplicate and included at least one biological replicate. PCR results were analysed using Student's paired *t*-test. Pyrosequencing results were analysed by ANOVA within representative runs and using Student's *t*-test on the average of multiple runs. Error bars on all graphs show standard error of the mean (SEM) or in the case of HT12 array data, 95% confidence interval (CI), unless otherwise stated. Asterisks are used to represent probability scores as follows: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.005$ or n.s. not significant.

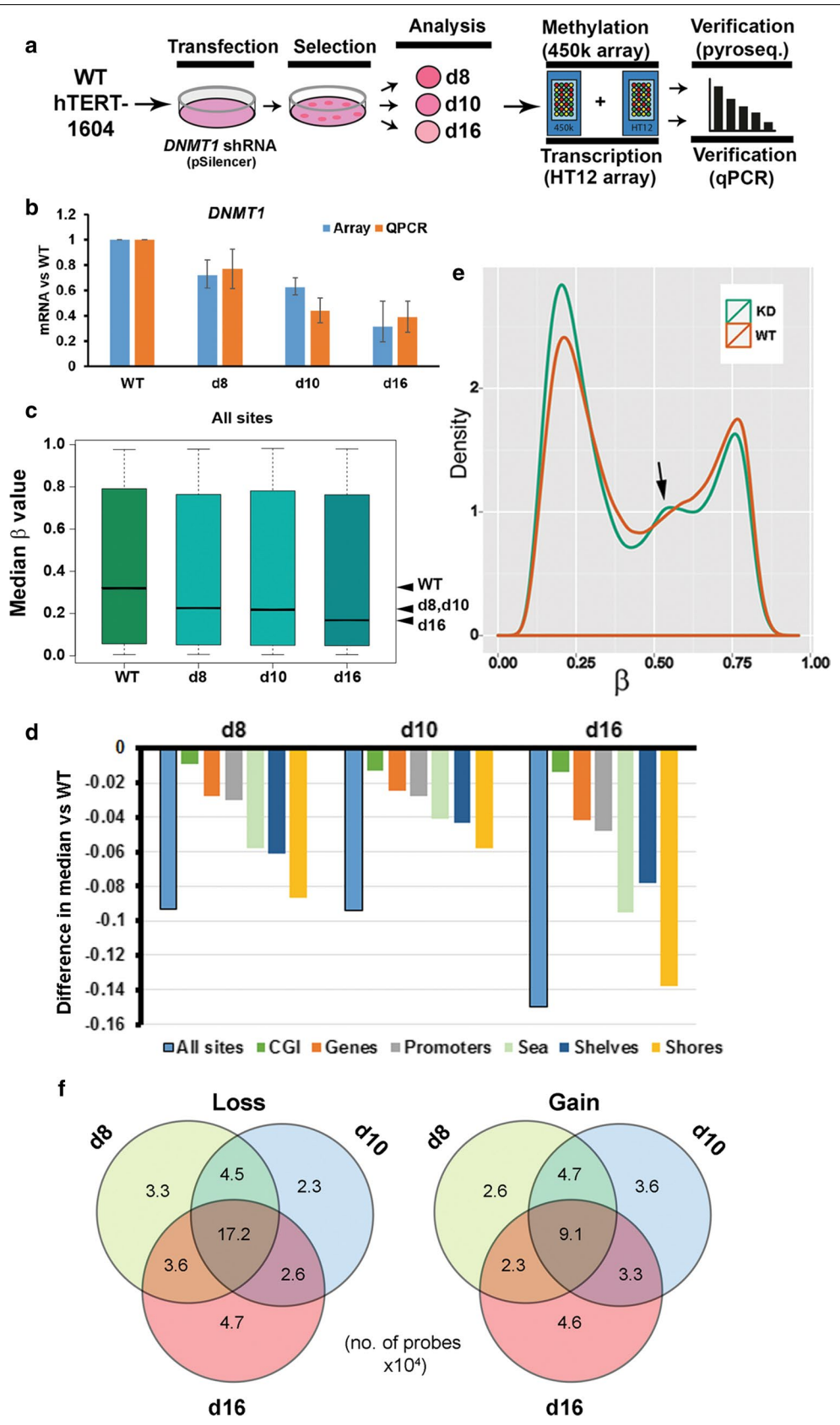
Results

Generation of isogenic hTERT1604 fibroblast cell lines

Isogenic lines carrying an shRNA construct targeting *DNMT1* were generated by transfecting the hTERT-immortalised human lung fibroblast cell line hTERT-1604 with pSilencer plasmid containing an shRNA (Fig. 1a). The generation and initial characterisation of these isogenic cell lines have been previously described [30]. Here we took two sublines typical of the intermediate levels of knockdown (KD) seen (d8 and d10) as well as one line (d16) with relatively low levels of mRNA, with good agreement between reverse transcription quantitative PCR (qPCR) and array results (Fig. 1b; all $p < 0.05$ except d8 array). We also confirmed knockdown at the protein level using Western blotting, with HCT116 cells mutated in DNMT1 and DNMT3B [27] as controls (Additional file 3: Fig. S2A).

(See figure on next page.)

Fig. 1 Cell line generation and overall changes seen in methylation levels. **a** Experimental approach: WT hTERT1604 fibroblasts were transfected with shRNA-containing plasmid and grown in selective medium; colonies of resistant cells were expanded, and three (d8, d10, d16) showing reduced DNMT1 levels were then analysed using genome-wide methylation and transcription arrays on the Illumina platform. **b** Levels of *DNMT1* mRNA in cell lines from array and qPCR: error bars represent 95% confidence intervals around median, and standard error of the mean (SEM), respectively. All three knockdown (KD) lines were significantly depleted at $p < 0.05$ for both assays (except d8 array). **c** Overall methylation levels in WT and KD cells as measured by 450K: a β value of 1 equates to 100% methylation. Median values are indicated by the line, and whiskers represent interquartile range. The positions of the medians are also indicated at right (arrowheads). **d** The difference in median β value between each KD cell line and WT is shown first for all sites assayed (see **c** above) and then for each type of genomic element. CGI, CpG island; shore, region adjacent to CGI; shelf, adjacent to shore; sea, all other. See also Additional file 3: Fig. S2B. **e** Probe density distributions; in KD there is a decrease in the number of fully methylated sites (β closer to 1) and an increase in the number of unmethylated sites (β closer to 0), as well as in probes showing intermediate levels of methylation (arrow). **f** Numbers of sites ($\times 10^4$) showing significant changes in methylation (FDR < 0.05) compared to WT: the set of common sites is largest in each case, with close to twice as many sites commonly losing methylation in comparison with those gaining


a
Transfection
Selection
Analysis
Methylation (450k array)
Verification (pyroseq.)
WT hTERT-1604
DNMT1 shRNA (pSilencer)
d8
d10
d16
450k
HT12
Transcription (HT12 array)
Verification (qPCR)
b
DNMT1
Array
QPCR
mRNA vs WT
WT
d8
d10
d16
c
All sites
Median β value
WT
d8,d10
d16
e
KD
WT
Density
β
d
Difference in median vs WT
All sites
CGI
Genes
Promoters
Sea
Shelves
Shores
f
Loss
Gain
3.3
4.5
2.3
17.2
3.6
2.6
4.7
2.6
4.7
3.6
9.1
2.3
3.3
4.6
(no. of probes x10⁴)

Characterisation of overall changes in absolute methylation levels in depleted lines

Using the 450K array [31] and processing in RnBeads [32] to assess methylation levels across the genome (Fig. 1c), there was still a wide range of methylation values (given for the array as a value β ranging from 0 to 1) in KD lines as compared to WT, but the median values were decreased as expected in all three with d8 being comparable to d10, while d16 was lower (arrowheads at right). Principle components analysis and examination of the sites showing greatest differences in methylation between the stable lines confirmed that d8 and d10 were most similar (Additional file 2: Fig. S1). Probes on the array were annotated by location relative to genomic features, and while all regions showed a decrease in methylation, the difference in median values was smallest for CGI, which were unmethylated anyway in parental cells ($\beta < 0.1$ in WT), while the separation in medians was greatest at shelves and shores, where methylation levels were higher (Additional file 3: Fig. S2B). This can most clearly be seen by plotting the difference in medians (Fig. 1d). Both WT and the KD cell lines showed the typical bimodal probe density distribution pattern reported in most cell types [31] (Fig. 1e). Overall, there was an increase in the numbers of less methylated probes ($\beta < 0.25$) in the KD cell lines and a decrease in the numbers of highly methylated probes ($\beta > 0.65$). For individual regions CGI again showed the smallest change, while gene bodies (genes) appeared most altered (Additional file 3: Fig. S2C).

To determine whether methylation was lost stochastically in each KD cell line given the variation seen (Additional file 2: Fig. S1), or was more targeted, we determined the degree to which affected sites were shared between the three cell lines (Fig. 1f). The largest set of sites losing methylation (17.2×10^4) was that shared between all three KD lines, supporting a non-random loss. A spike in numbers of probes showing intermediate levels of methylation ($\beta \sim 0.50$) in KD cell lines in the density profile plot (Fig. 1e, arrow) had indicated that a possible gain in methylation might also be occurring at some sites. Analysis showed that a substantial number (9.1×10^4) of sites gaining methylation are shared between all three KD lines, indicating reproducible gains in methylation at particular CpGs.

Overall pattern of sites showing significant differential methylation on DNMT1 depletion

We compared WT cells to all three KD lines using the RnBeads package in R and combined rank scoring (see methods). This confirmed that d16 has the greatest number of demethylated sites using a false discovery rate (FDR) cut-off of $p < 0.05$, but at $p < 0.001$ all three lines have comparable numbers of hypo- and hypermethylated sites (Additional file 4: Fig. S3A), with more sites losing than gaining. An analysis of the 1000 best-ranking sites highlights sites common to all three KD lines (Additional file 4: Fig. S3B), confirming that there are large numbers of sites which respond in the same way in each KD, with an excess of probes showing loss over gain.

We then looked to see whether shared probes were enriched in any particular gene region. As we were interested in changes which might cause altered transcription, we focussed on CGI, promoters and gene bodies (hereafter genes) rather than shores, shelves or open sea, where correlations with transcriptional output are harder to assess. Using a hypergeometric test in RnBeads, both promoters and genes, but not CGI, showed significant enrichment in demethylated probes for particular gene ontology (GO) terms. Table 1 indicates the top 3 ontology classes under biological process (BP) and molecular function (MF). For loss of methylation, examining common genes and processes suggested that three classes of genes were common to the enriched GO terms, which we grouped as follows: (1) genes involved in neuroepithelial differentiation; (2) genes involved in fat homoeostasis/body mass (FBM); and (3) olfactory receptor genes (groups 1–3 in Table 1), all of which will be dealt with below. The only orphan GO term whose members had multiple high-confidence demethylated sites was GO:0007506 gonadal mesoderm formation, which largely consists of members of the *TSPY* gene family on the Y chromosome. For gain of methylation, the same was true in that a relatively small number of histone modifier genes (group 4), represented under several GO terms, were responsible for many of the hits. In addition, the GO terms for glucuronosyltransferase activity (GO:0015020) and for regulation of megakaryocyte differentiation were also represented (Table 1). These were then curated by looking for sites showing reproducible changes (FDR < 0.05) in all KD lines (described more fully in "Methods" section), which indicated strong support [Yes (Y) in confirm column, Table 1] for all GO categories showing loss, but only in two showing gain (GO:0015020 and GO:0004984). We then set about verifying these targets.

Loss of methylation at the protocadherin gamma gene cluster particularly affects the A and B class variable genes

A main contributor to the enrichment of neuroepithelial genes are the protocadherin genes. Protocadherin α, β and γ (*PCDHA*, *PCDHB* and *PCDHG*) genes are located in three linked clusters on chromosome 5 and give rise to neural cell–cell adhesion proteins, with significant loss of methylation across the whole region in all three cell lines (Additional file 4: Fig. S3C). The α and γ proteins have a variable extra-cellular recognition domain, either A, B

Table 1 Gene ontology analysis for differentially methylated sites

Type	GO FID	*P*	OR	Ex	Obs	Total	GO Term	Grp	confirm
Loss									
Promoter									
BP	0098609	0.0011	3.0454	4.2148	12	189	Cell–cell adhesion	1	Y
	0007156	0.0011	3.4722	3.0998	10	139	Homophilic cell adhesion via plasma membrane	1	Y
	0010982	0.0015	88.2036	0.0669	2	3	Regulation of high-density lipoprotein particle clearance	2	Y
MF	0004888	0.0001	1.9709	24.2681	44	1055	Transmembrane signalling receptor activity	3	Y
	0005509	0.0001	2.2488	14.3768	30	625	Calcium ion binding	1	Y
	0004871	0.0003	1.7441	33.6302	54	1462	Signal transducer activity	3	Y
Gene									
BP	0007506	0	130.3775	0.1339	5	7	*Gonadal mesoderm development*		Y
	0032375	0.0001	25.9783	0.2295	4	12	Negative regulation of cholesterol transport	2	Y
	0045409	0.0001	77.705	0.0956	3	5	Negative regulation of interleukin-6 biosynthetic process	2	Y
MF	0008083	0.0009	3.5742	3.0015	10	158	Growth factor activity	3	Y
	0004984	0.0014	2.5939	6.136	15	323	Olfactory receptor activity	3	Y
	0038023	0.0014	1.7776	22.9102	38	1206	Signalling receptor activity	3	Y
Gain									
Promoter									
BP	0035574	0	443.1106	0.4729	14	15	Histone H4-K20 demethylation	4	N
	0045653	0	147.6833	0.5359	14	17	Negative regulation of megakaryocyte differentiation		N
	0016577	0	26.4022	1.0404	15	33	Histone demethylation	4	N
MF	0035575	0	452.3692	0.4637	14	15	Histone demethylase activity (H4-K20 specific)	4	N
	0032451	0	21.0879	1.1747	15	38	Demethylase activity	4	N
	0015020	0	10.1109	0.8965	7	29	*Glucuronosyltransferase activity*		Y
Gene									
BP	0035574	0	280.0725	0.4039	14	16	Histone H4-K20 demethylation	4	N
	0045653	0	140.0181	0.4544	14	18	Negative regulation of megakaryocyte differentiation		N
	0006335	0	31.0869	0.8078	14	32	DNA replication-dependent nucleosome assembly	4	N
MF	0035575	0	287.2955	0.3942	14	16	Histone demethylase activity (H4-K20 specific)	4	N
	0032451	0	24.654	0.9856	15	40	Demethylase activity	4	N
	0004984	0	4.4768	7.9586	31	323	Olfactory receptor activity	3	Y

BP biological process, *MF* molecular function, *GO FID* gene ontology family identification code, *P* probability value, *OR* odds ratio, *Ex* expected number of hits, *Obs* observed number, *Total* total number of genes in that family, *Grp*-see below; confirm Y/N, confirmation given by FDR tracks Yes/No

Groups (Grp): 1 = neuroepithelium; 2 = Fat homoeostasis/body mass (FBM); 3 = olfactory receptor; 4 = histone modifier

or C-type, attached to a constant transmembrane and intracellular domain. This is achieved at the gene level by alternative 5′ exons encoding the variable region being spliced to the constant region exons. Figure 2a shows the tracks containing sites with significant (FDR < 0.05) methylation differences between KD and WT cells for the *PCDHG* cluster. These reveal loss of methylation (in red in Fig. 2a) at most A and B class variable exons in all three KD cell lines, but not at the C class variable or the constant exons. Array probes were present in this region, and examination of the absolute rather than relative methylation (amber, top track in Fig. 2a) confirmed high levels of methylation in WT, where median β values were high for all variable exons (Fig. 2b). Methylation decreased in all three KD lines, with d10 showing the least effect (Fig. 2b). Methylation was substantially altered at all A and B class variable exons, but not at the C class (Fig. 2c). We could experimentally verify the loss of methylation at A2 (Fig. 2d), and no change at C4 (Fig. 2e), using pyrosequencing assays (pyroassay).

Some demethylation of other neuroepithelial genes in this GO category was also seen from the array, such as *S100P*, *ROBO1* and *PAX6*, with significant ($p < 0.05$) demethylation of *S100P* in two-thirds of KD cell lines confirmed by pyrosequencing (not shown).

Loss of methylation at other targets including fat homoeostasis/body mass (FBM) genes

Another class of genes showing enrichment all appear to be involved in some aspect of triglyceride processing,

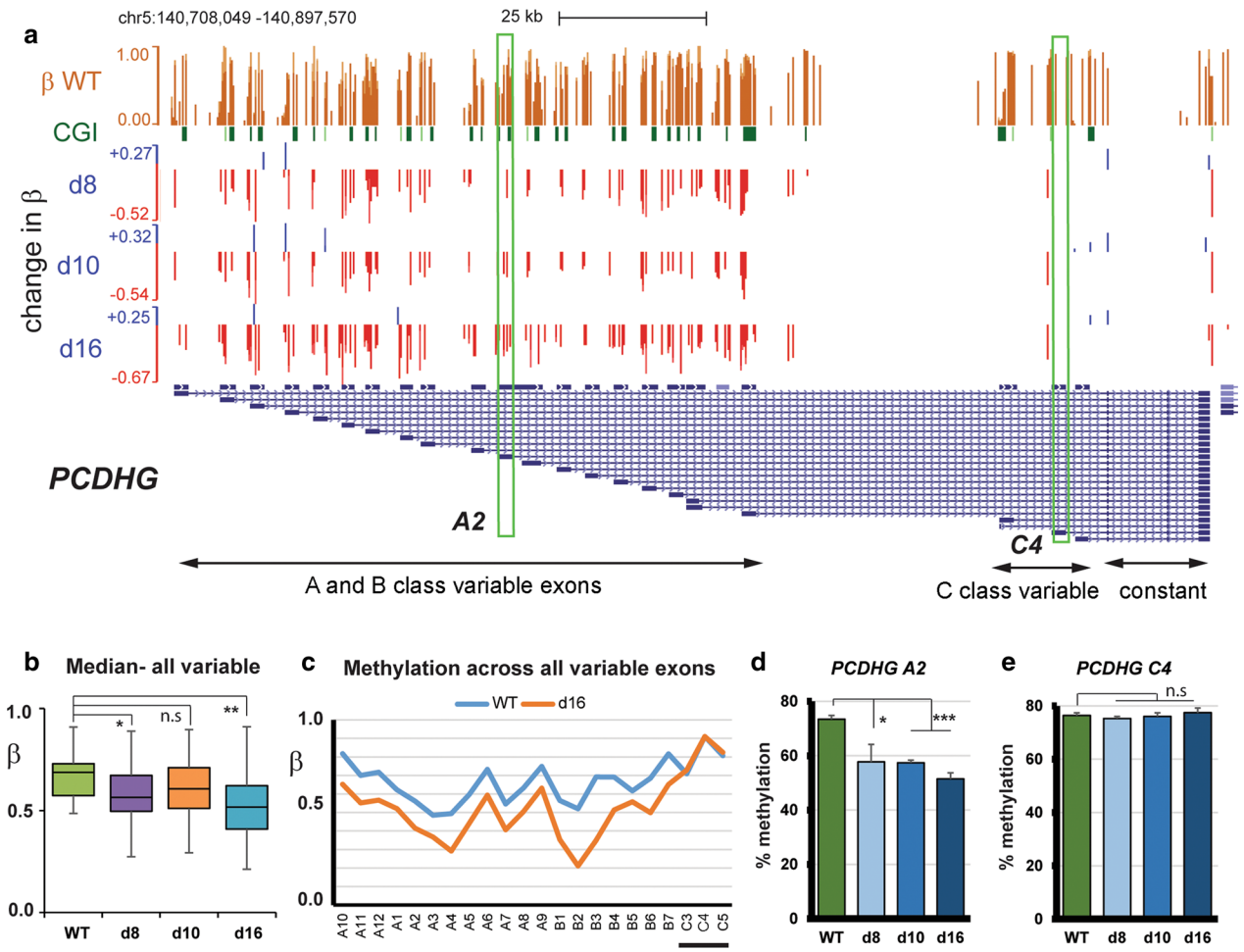


Fig. 2 Loss of methylation at the protocadherin γ (*PCDHG*) cluster of neuroepithelial identity genes. **a** Structure of the *PCDHG* cluster showing the 5′ variable exons (A, B and C classes) which are spliced to the 3′ constant exons (right). The top track (amber) shows absolute *β* values in the WT fibroblast cells from the 450K array, which range from 1(fully methylated) to zero (unmethylated). Only the sites showing significant differences from WT (FDR < 0.05) in each cell line are shown in the three tracks below, with decreases in red representing loss of methylation, and gains in blue. The size of the bar is proportional to the magnitude of change: maxima and minima are indicated on the scales at left. The locations of CpG islands (CGI) are also shown. Pyroassay locations are boxed. **b** Median *β* values for all variable exons. Significant differences (Mann–Whitney *U*) are indicated: **p* < 0.05; ***p* < 0.1; n.s., not significant. **c** Methylation at each exon in WT and d16 cells obtained by taking the median of the absolute *β* value for all probes at that exon. The variable class C exons are underlined. **d** Average methylation values in WT and KD cells obtained from a pyrosequencing assay (pyroassay) designed to cover CpGs in the A2 exon. Bars represent SEM; ****p* < 0.001, *t*-test. **e** Methylation at the C4 variable exon by pyroassay, shown as a control

energy homoeostasis and body weight regulation (Table 1), including leptin (*LEP*), ghrelin/growth hormone secretagogue receptor *(GHSR)* and genes encoding the very low density lipoproteins *APOC1*, *APOC2* and *APOC3*. Median levels of methylation in the gene bodies were approximately 45% in WT ($\beta = 0.45$) and showed significant ($p < 0.05$) decreases in the KD lines (Fig. 3a). Most individual genes also showed substantial loss, with the exception of the *ANXA* genes (Fig. 3b). Loss of methylation at the *LEP* and *GHSR* promoters was confirmed using pyroassay (Fig. 3c).

Olfactory receptor (OR) genes appeared in a number of GO categories as having lost methylation, though some gains in the gene body were also indicated (Table 1). ORs encode G protein-coupled receptor proteins and are members of a large gene family, many of which are grouped into major clusters, particularly on chromosome 11 [41]. To buffer against stochastic effects due to the large gene family involved, we carried out a second analysis starting instead with sites in promoters showing reliable methylation loss compared to WT (FDR < 0.05) in the triplicates of each KD line

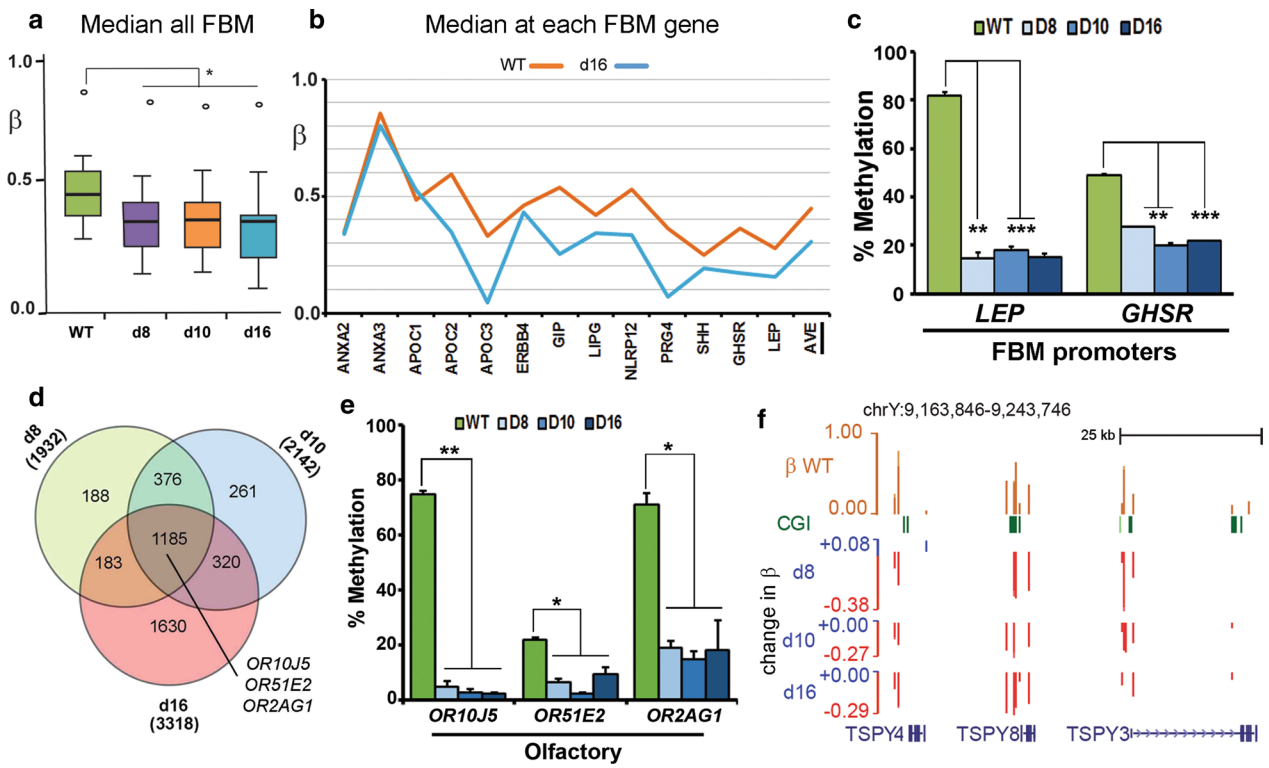


Fig. 3 Loss of methylation at fat homoeostasis/body mass (FBM) genes, olfactory receptors and the TSPY genes. **a** Median β values for all FBM genes (following curation) in WT and KD cell lines; significant differences (Mann–Whitney *U*) are indicated. **b** Median methylation values at each FBM gene in WT and d16 cells. **c** Average methylation levels obtained from pyroassays at the Leptin (*LEP*) and Ghrelin/growth factor receptor secretagogue (*GSHR*) promoters. **d** CpG located in promoters which showed highly reproducible loss of methylation (FDR < 0.05) were identified. The set of sites common to all three KD cell lines ($n = 1185$) was found to be enriched for olfactory receptors (such as the three indicated) using the DAVID tool. **e** Pyroassays designed for the three olfactory receptor genes from (**d**) confirmed methylation was consistently reduced across all KD cell lines. **f** Browser view showing loss of methylation (red) at the CGI-containing promoters for members of the *TSPY* family on the Y chromosome

and then overlapping these (Fig. 3d) to see which sites were common to all three KD cell lines (Additional file 5: Table S2). Ontology analysis of these common sites using DAVID independently highlighted signalling receptor genes and more particularly olfactory receptors ($n = 21$). This group of OR genes also showed significant demethylation compared to WT (Kruskal–Wallis, $p < 0.05$) across the genes when median methylation at all available probes was analysed (Additional file 4: Fig. S3D). We chose three of these genes—*OR10J5*, *OR51E2* and *OR2AG1*—located on different chromosomes and could verify loss of methylation in all KD lines (Fig. 3e).

The final GO category of genes (GOFMID:0007506) showing loss of methylation (Table 1) consists largely of the *TSPY* gene family (*TSPY1-4*, *8* and *10*) located on the Y chromosome and thought to be implicated in both normal gonadal development and in gonadoblastoma [42]. These also showed clear evidence of demethylation (Fig. 3f).

Gains in methylation affect the UGT1A locus

As indicated above, with respect to gains in methylation only two of the GO classes identified in the genome-wide screen (Table 1) contained multiple sites showing significant gains in methylation (FDR < 0.05, > 0.1 gain in β). One of these was the olfactory genes, discussed above: the other GO term GO:0015020 was largely comprised of members of the *UGT1A* family. This gene family has a similar structure to the *PCDHG* cluster, where unique alternate 5′ exons splice to common 3′ exons, but in this case codes for a series of nine UDP-glucuronosyltransferase enzymes (UGTs). Substantial gains in methylation can be seen at the upstream promoters controlling the 5′ exons (Fig. 4a), most of which lack CGI. Median methylation levels also showed clear increases overall in the KD lines (Fig. 4b), though these did not reach significance. Most individual exons also showed a sharp increase (Fig. 4c), with A1 being a clear exception in all lines. We confirmed a significant gain in methylation in each cell line at A4 (Fig. 4d) but no alteration at A1 (Fig. 4e). In contrast to the clear gains in

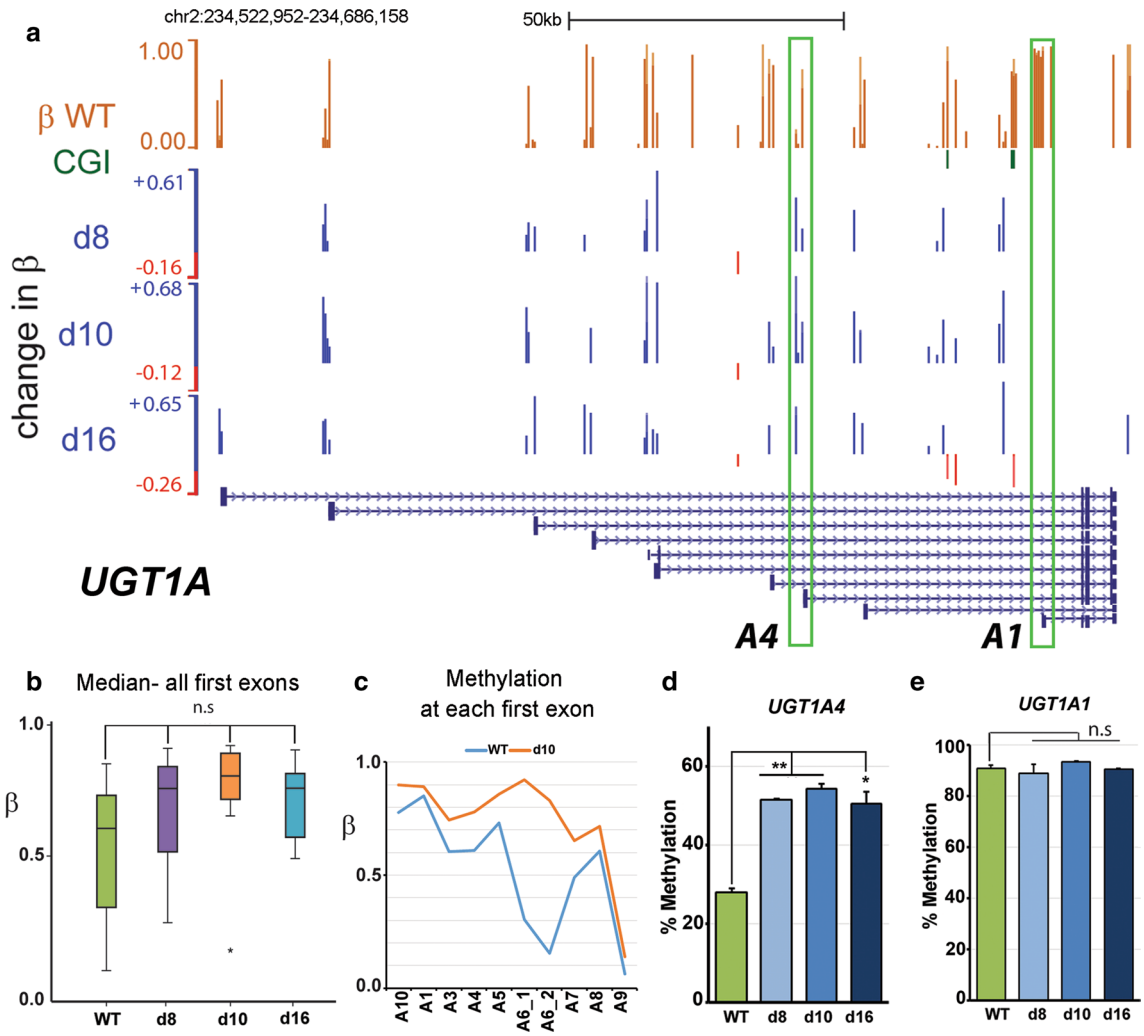


Fig. 4 Gains in methylation at the clustered *UGT1A* locus. **a** Structure of the *UGT1A* cluster showing the 5′ variable exons (*UGT1A1–UGT1A10*) which are spliced to the 3′ exons (right). Key to tracks as before; pyroassay locations (*UGT1A1* and *UGT1A4*) are boxed. **b** Median β values for all first exons: though medians are higher in KD lines these failed to reach statistical significance. **c** Median absolute β values at individual first exons in WT and d16 cells. **d** Average methylation values in WT and KD cells obtained from a pyrosequencing assay (pyroassay) designed to cover CpGs in the A4 exon. **e** Methylation by pyroassay at the A1 exon, shown as a control

all three lines for *UGT1A*, the histone modifier group also identified as gaining methylation (Table 1, group 4) contained few FDR-supported sites and these often did not overlap between cell lines, with median β levels also not differing significantly (Additional file 4: Fig. S3E).

A cluster of loci showing gain of methylation on the X chromosome

Given that there were considerable numbers of probes showing gain in methylation, but few of the GO classes from the RnBeads analysis contained testable targets by our criteria, we tried an alternative analysis as for the OR above. Sites associated with promoters and which showed reliable (FDR < 0.05) gains were identified in each KD line, and then the lists of cognate genes were compared to find those which were common to all three cell lines (Fig. 5a). Examination of the 201 promoters from this analysis (Additional file 5: Table S2) failed to show any significantly enriched terms in DAVID. However, several of the genes showing the greatest gain in methylation were located on the X chromosome, including *GABRQ* and members of the *MAGE* family of cancer/testis antigens such as *MAGEA12*. Mapping of FDR sites to the X chromosome showed that adjacent domains could vary in methylation level by more than 80% in either direction (Fig. 5b). Pyroassays for *GABRQ* and the neighbouring

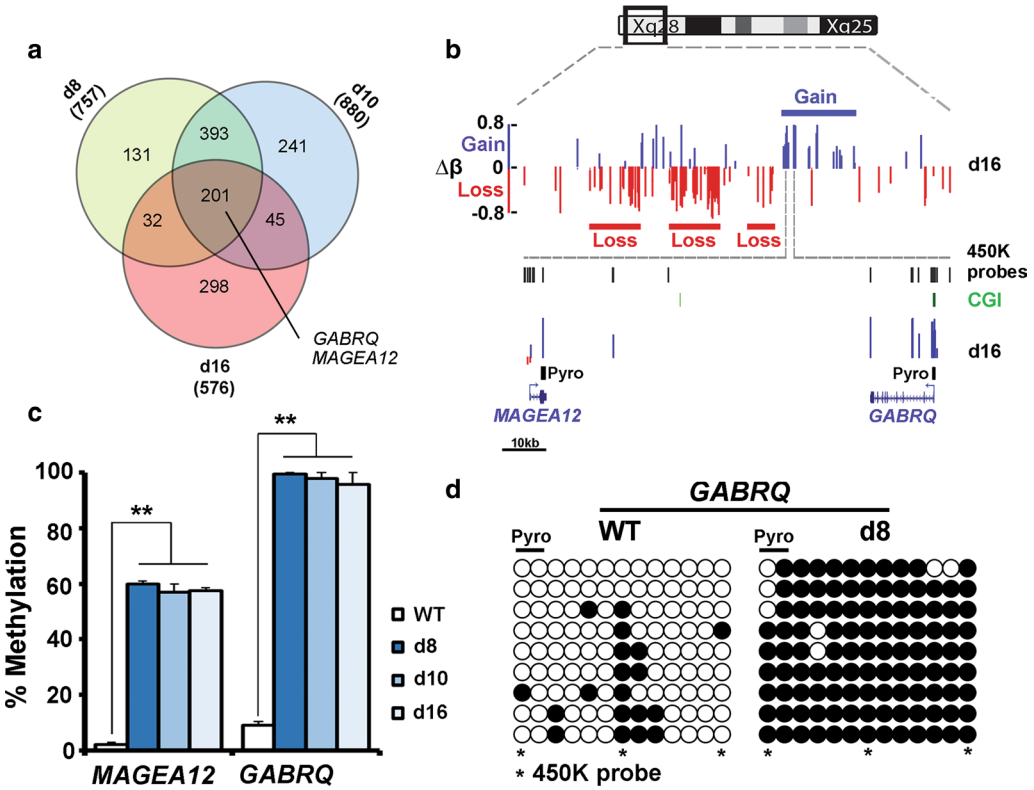


Fig. 5 Gains in methylation on the X chromosome. **a** Sites reliably showing gain in methylation and located in promoters were analysed to identify those common to all three KD lines ($n=201$). Some of these sites showing the greatest change in methylation were located on the X chromosome including *MAGEA12* and *GABRQ*. **b** Schematic showing the locations of the two genes adjacent to each other on X in a region showing gain in methylation. Tracks indicate the locations of all 450K probes and CGI; the positions of the pyroassays are also indicated; the scale bar pertains to the bottom part of the schematic; Δ*β*, change in beta value. **c** Methylation as determined by pyroassay at the two genes indicated in **a**, **b**. **d** Clonal analysis of *GABRQ* in WT and d8. Filled circles represent methylated sites, open circles unmethylated. The CpG which were also analysed by the pyroassay (pyro) and the 450K array (asterisk) are indicated

MAGEA12 gene confirmed significant gains in methylation at the *GABRQ* promoter and in the *MAGEA12* gene body (Fig. 5c). Clonal analysis for *GABRQ* indicated a uniform increase in methylation (78 vs. 16%) across all adjacent CpG at this locus (Fig. 5d). Both direction and degree of change in methylation were highly correlated between pyrosequencing and the 450K array across all sites which were covered by both types of assay ($r=0.916$ for loss of methylation $r=0.818$ for gain in methylation).

Transcriptional changes are enriched at cancer/testis antigen genes on X and Y

To see whether methylation changes were accompanied by large-scale changes in transcription, we carried out a genome-wide screen using the HT12 array which assays most RefSeq genes. Figure 7a shows the distribution of changes comparing d8 and WT: genes which showed >2 fold change (FC) and with scores of $p<0.05$ are highlighted, with the greater spread to the right indicating a greater tendency to derepression. Relatively small numbers of genes were affected (Fig. 6b), particularly at higher stringency (FDR < 0.01), and d16 showed fewest dysregulated genes. To determine common targets, we looked for shared genes (Fig. 6c). DAVID analysis on the genes common to all three ($n=70$; Additional file 6: Table S3) indicated significant enrichment for genes coding for MAGE domains (Fig. 6d). MAGE genes on the X chromosome were previously identified as showing large changes in methylation (Fig. 5): also appearing here was a *TSPY* family member (Table 1, Fig. 3f). Upregulation of members of these gene classes could be verified by RT-PCR (Fig. 6e) and showed similar direction of change to the array, and greater magnitude, by RT-qPCR (Fig. 6f). Consistent with the transcriptional upregulation, median methylation levels at the promoters of these genes were lower than WT (Fig. 6g). Interestingly, there was an overall increase in intragenic (as opposed to promoter) methylation in the larger group of transcriptionally dysregulated genes common to d8 and d10 ($n=764$, see Fig. 6h and Additional file 6: Table S3), which may

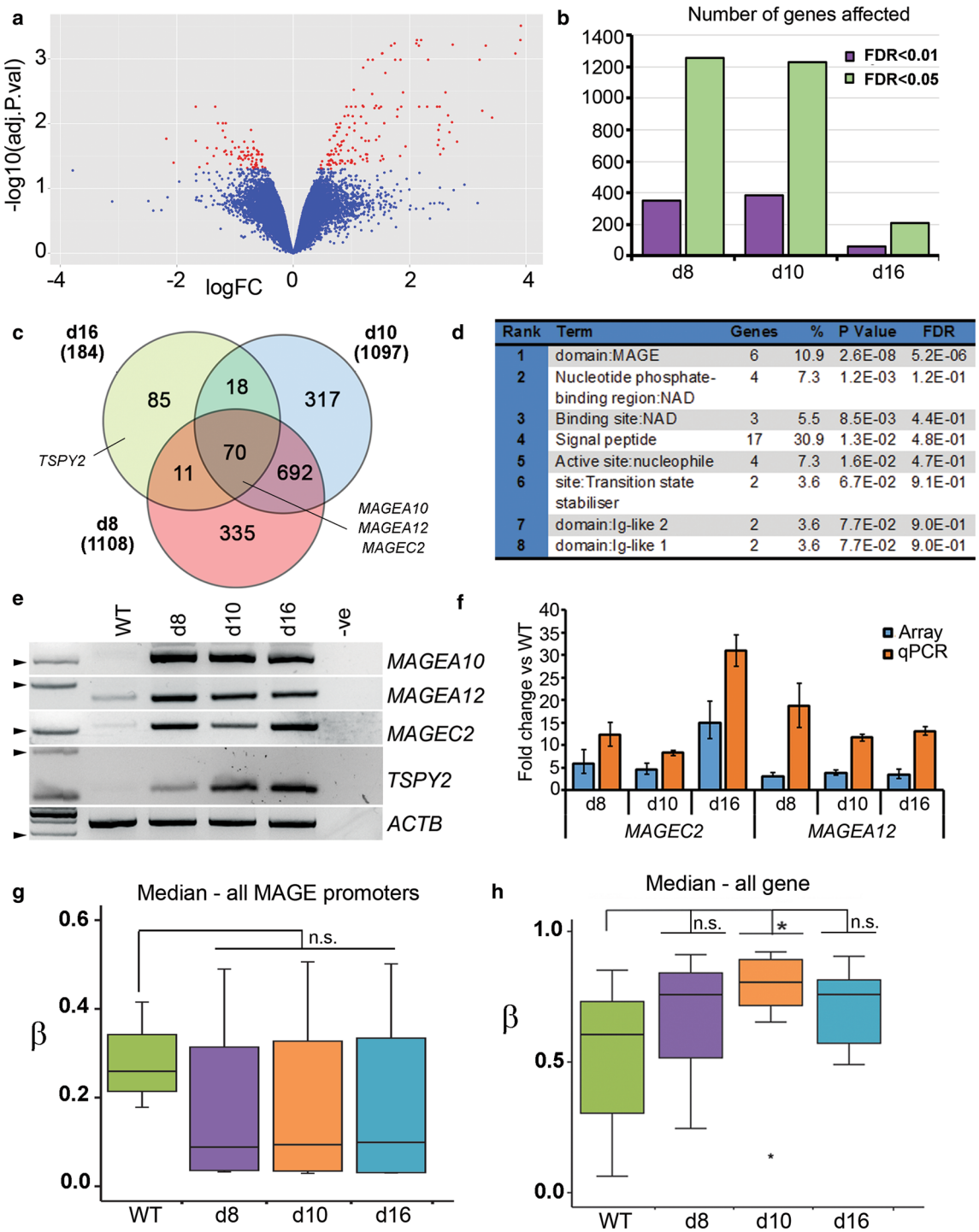
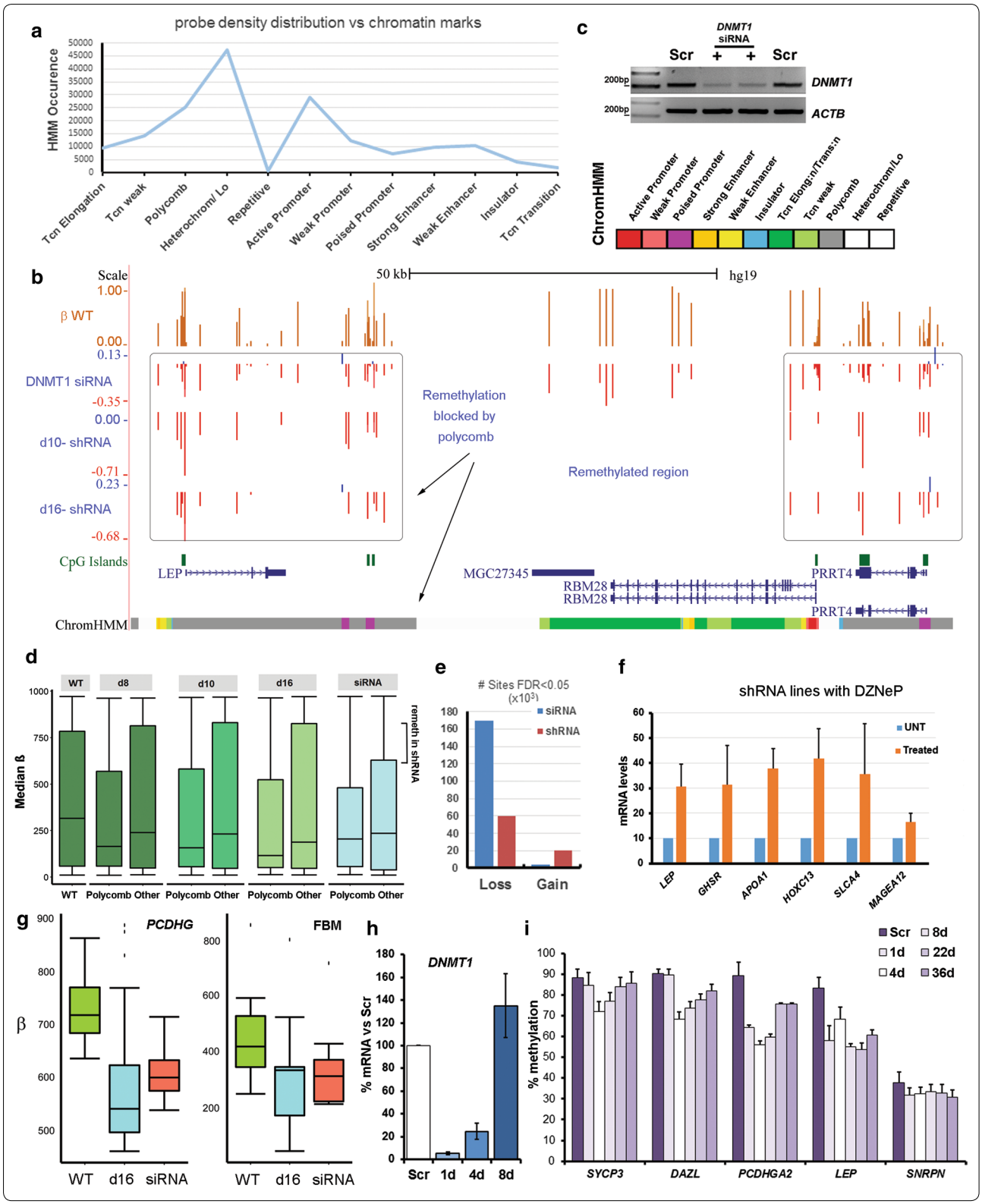


Fig. 6 Transcriptional dysregulation of genes on the X and Y chromosomes correlates with methylation changes. **a** Volcano plot showing log fold change (FC) in transcription as measured by HT12 array versus FDR-corrected significance values: genes with > 2FC and FDR < 0.05 are highlighted in red. **b** Numbers of dysregulated genes at different FDR thresholds for the different KD lines. **c** Genes common to more than one KD line at FDR < 0.05; total numbers in each cell line are indicated in brackets. **d** Ontology enrichment output from DAVID for the genes common to all KD lines. **e** RT-PCR analysis of the three *MAGE* genes on X and a member of the *TSPY* gene family on Y highlighted in DAVID analysis (C). *ACTB* is a loading control; −ve, negative control lacking cDNA. A 100-bp ladder is shown at left with the 200-bp band indicated by an arrowhead. **f** Transcription levels of indicated MAGE genes from the HT12 array or by qPCR. Error bars are 95% CI for the array, SEM for qPCR; fold change was significant ($p < 0.05$) in all cases. **g** Median β values on 450K array for probes at MAGE promoters were decreased, though failed to reach significance. **h** Gene body methylation was increased in transcriptionally upregulated genes


a
probe density distribution vs chromatin marks
HMM Occurence
Tcn Elongation
Tcn weak
Polycomb
Heterochrom/ Lo
Repetitive
Active Promoter
Weak Promoter
Poised Promoter
Strong Enhancer
Weak Enhancer
Insulator
Tcn Transition
c
DNMT1 siRNA
Scr + + Scr
200bp
DNMT1
ACTB
ChromHMM
b
Scale
50 kb
hg19
β WT
DNMT1 siRNA
d10- shRNA
d16- shRNA
Remethylation blocked by polycomb
Remethylated region
CpG Islands
LEP
MGC27345
RBM28
PRRT4
ChromHMM
d
WT
d8
d10
d16
siRNA
Median ß
remeth in shRNA
WT Polycomb Other
e
Sites FDR<0.05 (x10³)
siRNA
shRNA
Loss
Gain
f
shRNA lines with DZNeP
UNT
Treated
mRNA levels
LEP
GHSR
APOA1
HOXC13
SLCA4
MAGEA12
g
PCDHG
FBM
β
WT d16 siRNA
h
DNMT1
% mRNA vs Scr
Scr 1d 4d 8d
i
Scr 8d 1d 22d 4d 36d
% methylation
SYCP3
DAZL
PCDHGA2
LEP
SNRPN

(See figure on previous page.)
Fig. 7 Methylation loss is concentrated at regions normally repressed by polycomb. **a** Distribution of probes showing significant loss per chromatin state—numbers of probes are shown at left, chromatin states below: tcn, transcription; heterochrom/Lo, heterochromatin or low signal; repetitive, repeat DNA. **b** Region around the *LEP* gene: tracks as before, with the addition of data from cells treated with siRNA for 72 h (top). A track showing ChromHMM chromatin states from NHLF foetal lung fibroblasts is shown at bottom: grey, polycomb-repressed; green, transcriptionally active (full colour key at top right). **c** DNMT1 mRNA levels by qPCR following treatment with siRNA (+) for 72 h compared with scrambled control (Scr). ACTB is shown as a control; ladder as above. **d** Median β values for all regions (WT) compared to medians for polycomb-repressed regions (Polycomb), or all other regions (Other) in the cell lines indicated at top; remeth, remethylated. **e** Numbers of probes showing loss and gain in methylation in hTERT cells following treatment with siRNA for 72 h compared with the shRNA lines (averaged); #, number. **f** mRNA levels for the indicated genes in shRNA lines treated with the EzH2 inhibitor DZNeP; UNT, untreated; bars represent SEM, experiment carried out in duplicate. **g** Median β values for all variable exons at the *PCDHG* locus (left) and for fat/body mass genes (FBM, right): compare d16 shRNA lines with cells treated with siRNA. **h** *DNMT1* mRNA levels in WT cells exposed to siRNA for 48 h, then allowed to recover in normal medium; comparisons were made to a scrambled siRNA negative control (Scr). **i** Methylation levels by pyroassay at the loci indicated during the transient KD and recovery shown in (**h**); timepoints are in days. All loci showed significant loss of methylation: *LEP* and *SNRPN* showed no significant gain versus lowest methylation level, while *PCDHGA2* showed no significant gain between d22 and d36

reflect increasing gene body methylation accompanying transcription.

Regions hypomethylated in shRNA lines correlate with polycomb repression

To investigate why losses in methylation occurred at the same positions in all KD lines, we used ENCODE data to look at chromosomal distribution, replication timing and chromatin features which might be important, since the DNMTs have no DNA sequence specificity themselves. Of these, the chromatin marks were most informative, in particular the ChromHMM dataset on lung fibroblasts which partitioned the genome into different types of chromatin based on a set of distinguishing histone marks and other features [43]. This indicated that probes significantly losing methylation in our shRNA lines are most densely distributed across regions which are normally polycomb-repressed or are heterochromatic/low-signal regions in lung fibroblasts (Fig. 7a). Specifically, many regions show a striking correlation between polycomb marking and methylation loss, such as the *LEP* and neighbouring *PRRT4* genes (Fig. 7b): in contrast, the intervening *MGC27345* and *RBM28* genes at that locus, which are highly methylated in WT cells (top track), show little or no loss of methylation and have chromatin marks associated with transcription.

These data suggested that polycomb-repressed regions might be more susceptible to demethylation than others. To test whether these regions lost methylation more readily than others, we treated hTERT1604 with siRNA for 72 h, which led to acute depletion of the *DNMT1* mRNA (Fig. 7c). We found, however, that there was little difference between polycomb-repressed and other regions in terms of demethylation in the siRNA-treated lines (Fig. 7d), in contrast to the shRNA lines where losses were concentrated at the former (Fig. 7d). This could also be seen at the *LEP* locus, where *MGC27345* and *RBM28* now showed loss of methylation following siRNA treatment (Fig. 7b, siRNA track). Also of note, almost no probes showed gains in methylation relative to WT in the siRNA cells (Fig. 7e), indicating that this effect is associated exclusively with chronic treatment. These results suggested that gains of methylation had occurred only in shRNA lines and had effectively restored methylation to near WT levels at most regions outside of those marked as polycomb-repressed.

Since transcriptional analysis did not highlight dysregulation of polycomb regions in shRNA cells (Fig. 6d), we tested to see whether polycomb-mediated repression was being maintained there in the absence of DNA methylation. To do this, we treated with DZNep, an inhibitor of EZH2, and confirmed the upregulation of a positive control gene *SLCA4* (Fig. 7f) as previously reported [44]. Likewise, *HOXC13*—a known polycomb target—showed derepression (Fig. 7f). The FBM genes marked by polycomb including *LEP* showed reactivation to a comparable degree to *SLCA4*, whereas the *MAGEA12* gene which is in a heterochromatic region not marked by polycomb showed little effect (Fig. 7f).

To further investigate the difference between acute and chronic DNMT1 depletion in these cells, we first examined the effects of acute depletion by siRNA on the loci identified in the stable lines: this confirmed that loci such as the clustered protocadherins and the fat/body mass genes also lose methylation on short-term depletion by siRNA (Fig. 7g). Following treatment, cells were then allowed to recover in the absence of siRNA for an extended period (36 days). DNMT1 levels returned to normal rapidly (Fig. 7h). Examination of the methylation response at various gene classes was very instructive. Germline genes (*SYCP3*, *DAZL*), which are known to become de novo methylated to high levels during somatic differentiation [5, 34], showed initial loss versus a scrambled control (Scr), followed by remethylation over time to near WT levels (Fig. 7i), confirming that the hTERT cells possess sufficient de novo activity to remethylate

the genome, as already suggested (Fig. 7b–e). Imprinted genes are normally unable to regain methylation somatically [45], and we could confirm that the *SNRPN* imprint control region failed to remethylate (Fig. 7i). The polycomb-marked genes *LEP* and *PCDHGA2* were also refractory to de novo methylation, either showing no gain (*LEP*) or reaching a plateau at an intermediate level of recovery only (*PCDHGA2*) (Fig. 7i).

Gain in methylation is associated with poised promoters in shRNA lines

Having established that loss of methylation in shRNA lines is linked to polycomb repression, we wished to determine what features are associated with gains in methylation in these chronically depleted cell lines. As indicated, gains were not seen genome-wide following acute depletion using siRNA (Fig. 7e) and specific loci such as *UGT1A* showed instead loss of methylation on acute treatment (Fig. 8a, siRNA track), suggesting that hypermethylation is associated with longer-term culture of the shRNA-containing cell lines. To investigate what features might be associated with such loci, we looked to see which chromatin states in shRNA lines showed the highest median β for probes which gained methylation and the largest difference in methylation (Fig. 8b). This identified weak and poised promoter categories, and comparing shRNA lines to WT (Fig. 8c), the median values were more different for poised than for weak promoters (0.4 vs. 0.2, Cohen's *D* test). These results suggested that poised promoters attract de novo methylation particularly strongly. Consistent with this, hypermethylation in the shRNA lines is centred around the *UGT1A* promoters and not the common 3′ exons (Fig. 8a). A heterochromatic location may contribute to over-methylation, since genes in adjacent active chromatin show restoration of normal methylation (Fig. 8a, compare siRNA to d10, d16 for *DGKD*), but not hypermethylation. While *UGT1A* transcription levels were very low compared to expressing cells by RT-qPCR (not shown), available HT12 array data showed a consistent decrease in transcription in all three shRNA lines (Fig. 8d, left), correlated with gains in methylation at the cognate promoters (Fig. 8d, right).

Further analysis confirmed that while gains in methylation were seen across all the UGT1A exons in all shRNA lines (Fig. 8e), all of these exons showed a loss of methylation following acute depletion with siRNA. We took advantage of our transient depletion and recovery experiment (Fig. 7h, i) to examine levels of methylation at *UGT1A4* using pyrosequencing: this showed that while the region indeed loses methylation on acute depletion, it undergoes steady de novo methylation following recovery and at day 36 was the only gene examined whose methylation exceeded that seen in the scrambled control (32.4 vs. 31.3%), suggesting that these genes are indeed susceptible to hypermethylation.

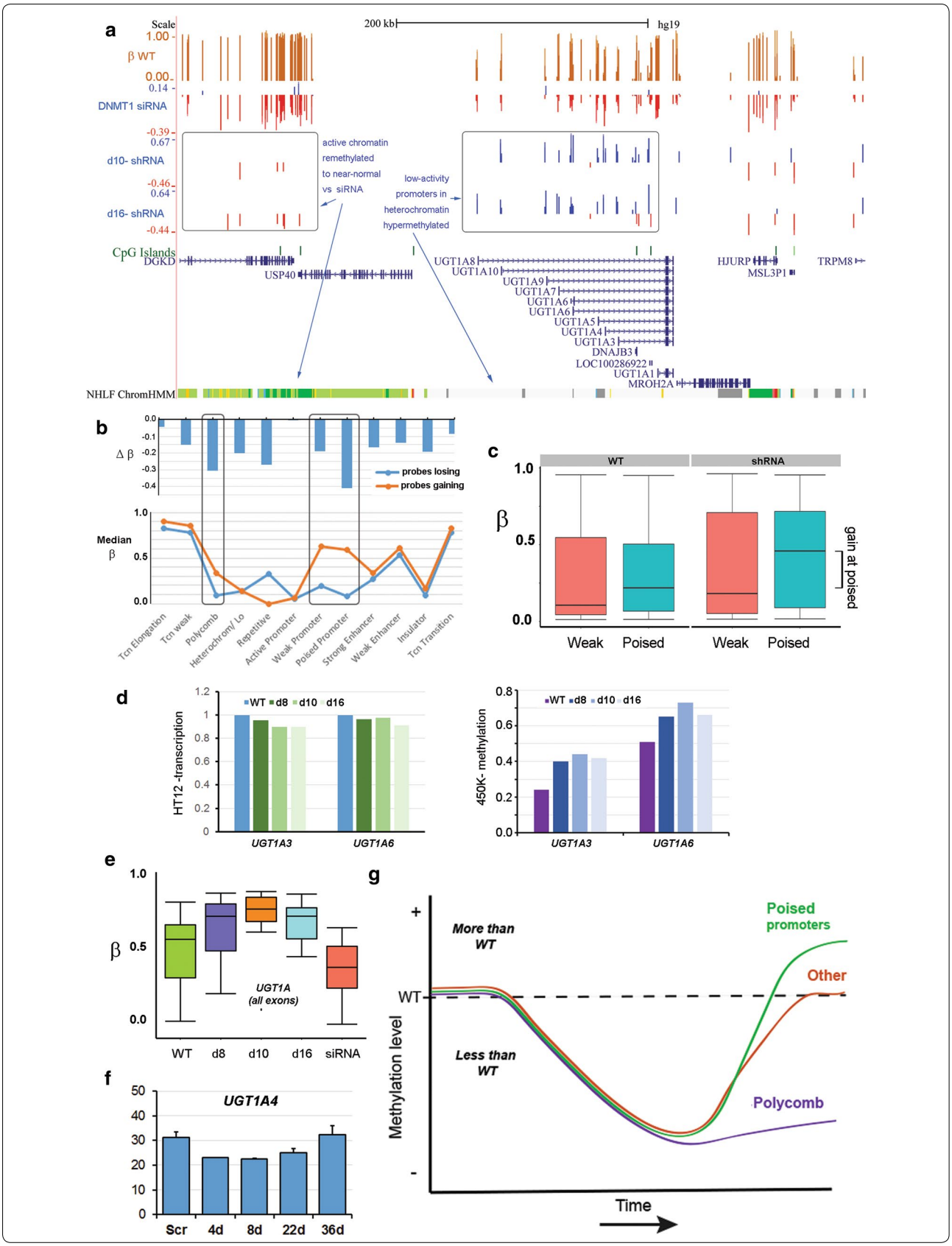
One possible reason for the gains in methylation seen in the shRNA lines could be over-expression of a de novo enzyme. Previous reports have indicated that between them, DNMT3B and DNMT1 account for the majority of methylation in cultured adult human cells and that there may be a role for DNMT3B in maintenance as well as de novo methylation [27]. We saw little change in *DNMT3B* levels in the *DNMT1* KD lines from the HT12 transcriptional array (Additional file 7: Fig. S4A) or RT-PCR (not shown), indicating that gains in methylation are not due to *DNMT3B* over-expression. To investigate a possible role in maintenance methylation, we carried out a transient siRNA treatment and could achieve robust knockdown of *DNMT3B* in the cells (Additional file 7: Fig. S4B). While some germline genes showed little effect, loci previously shown to require DNMT3B including *D4Z4* and *NBL2* did show loss of methylation (Additional file 7: Fig. S4C), confirming that we had achieved a functional depletion. Examination of the loci identified in our *DNMT1* shRNA clones showed that these loci also showed loss of methylation in *DNMT3B* KD cells (Additional file 7: Fig. S4C), suggesting that loci which remain hypomethylated in the shRNA clones also require input from DNMT3B to retain WT methylation levels.

Discussion

Summary and model

We and others have previously shown that acute depletion of DNMT1 using siRNA triggered the DNA damage response and cell cycle perturbations in human cell lines, making it difficult to identify genes which are directly controlled by methylation. Here we used isogenic shRNA-containing derivatives of a normosomic lung fibroblast cell line to look at the effects of chronic depletion of the protein. We characterised the alterations in methylation and transcription using microarrays in three different cell lines, processing them using a highly reproducible pipeline, and verified changes using locus-specific pyrosequencing or RT-qPCR assays. Additionally, we compared the effects on methylation of this chronic depletion to the effects of acute depletion using siRNA, as well as investigating possible contributions by DNMT3B. Finally, we investigated the correlations between chromatin state and DNA methylation and showed a role for polycomb-mediated repression at some of the loci.

Our results show that while both siRNA and shRNA-treated cells lose methylation overall as would be expected, only the latter show gains in methylation, most likely reflecting selection against the deleterious effects


a
Scale
1.00
β WT
0.00
0.14
DNMT1 siRNA
-0.39
0.67
d10- shRNA
-0.46
0.64
d16- shRNA
-0.44
200 kb
hg19
active chromatin remethylated to near-normal vs siRNA
low-activity promoters in heterochromatin hypermethylated
CpG Islands
DGKD
USP40
UGT1A8
UGT1A10
UGT1A9
UGT1A7
UGT1A6
UGT1A6
UGT1A5
UGT1A4
UGT1A3
DNAJB3
LOC100286922
UGT1A1
MROH2A
HJURP
MSL3P1
TRPM8
NHLF ChromHMM
b
Δ β
probes losing
probes gaining
Median β
Tcn Elongation
Tcn weak
Polycomb
Heterochromy/ Lo
Repetitive
Active Promoter
Weak Promoter
Poised Promoter
Strong Enhancer
Weak Enhancer
Insulator
Tcn Transition
c
WT
shRNA
β
gain at poised
Weak
Poised
Weak
Poised
d
HT12 -transcription
WT d8 d10 d16
UGT1A3
UGT1A6
450K- methylation
WT d8 d10 d16
UGT1A3
UGT1A6
e
β
UGT1A (all exons)
WT
d8
d10
d16
siRNA
f
UGT1A4
Scr
4d
8d
22d
36d
g
More than WT
WT
Less than WT
Poised promoters
Other
Polycomb
Methylation level
Time

(See figure on previous page.)
Fig. 8 Methylation gain is concentrated at poised promoters. **a** *UGT1A* locus showing siRNA treatment data (top), shRNA lines (middle) and chromatin states (bottom); grey, heterochromatin/low signal; green, transcriptionally active (for full key see previous fig). **b** Median *β* levels for probes gaining and losing in shRNA lines (bottom) and median changes in methylation (Δ*β*) versus WT for different chromatin states. **c** Boxplots of methylation values for probes falling within weak and poised promoter chromatin regions in WT or shRNA lines (averaged). **d** Transcription at the *UGT1A3* and *UGT1A6* genes decreases (relative to WT, set to 1) in all three shRNA lines as methylation (*β* value) increases, as indicated by HT12 and 450K arrays, respectively. **e** Median methylation (*β*) across all *UGT1A* exons decreases in siRNA-treated cells, but shows gains in all shRNA lines. **f** Methylation at *UGT1A2* during the transient KD and recovery experiment shown in Fig. 7h, i; differences are significant between control (Scr) and d4, but not Scr versus d36. **g** Model for methylation changes which occurred over time following chronic (shRNA) depletion of *DNMT1*: while polycomb-marked regions (purple) resisted remethylation, most regions ("other", red) regained normal or near-normal levels, while poised promoters (green) tended to become hypermethylated

of hypomethylation during clonal expansion and culture. Figure 8e shows what we propose to have occurred: shRNA treatment gave initial widespread demethylation in all three clonal lines, since each line shows the presence of some highly demethylated sites distributed across the genome, but methylation seems to have recovered at most CpGs (Fig. 8e red line). Comparison to normal chromatin patterns in human lung fibroblasts indicated that remaining hypomethylation in the expanded cells was concentrated at regions normally marked for repression by polycomb (Fig. 8e purple line), while the smaller number of regions becoming hypermethylated relative to the parental cell line are associated with poised promoters (green line). TET expression was not detected, and the cells had little or no 5-hydroxymethylation (5hmC; data not shown), in keeping with other reports [46], suggesting that the hypermethylation does not represent 5hmC. Likewise, no over-expression of DNMT3B was detected.

In terms of what type of gene was particularly affected by chronic DNMT1 KD, the enrichment analyses and laboratory verification consistently pointed at the same small group of gene categories, namely (1)neuroepithelial genes, and in particular the protocadherins; (2) fat homoeostasis/body mass genes; (3) olfactory receptors; (4) the cancer/testis antigens; and (5) the *UGT1A* complex.

Protocadherins are major targets of DNA methylation in human cells

Emerging evidence suggests that the clustered protocadherin genes may be central to specifying individual neural cell identity [47, 48] and they have been shown to become heavily methylated during embryonic development in mouse [49], suggesting that stable repression of non-transcribing copies is a programmed event during development. Recent work has shown that DNMT3B is important for de novo methylation at these loci and suggested that dysregulated expression may contribute to the phenotype in immunodeficiency, chromosome abnormalities and facial anomalies (ICF) syndrome [50], where *DNMT3B* is frequently mutated [51], and we found that depletion of DNMT3B was accompanied by loss of methylation at *PCDHGA2*. The *PCDHA* and *PCDHB* loci are heterochromatic and show persistent loss of methylation, as does the 5′ end of the *PCDHG* locus which is polycomb-repressed, but not the 3′ end which shows little loss of methylation and has instead chromatin marks associated with weak transcription (Additional file 3: Fig. S2B). Meehan and co-workers recently showed that long-term loss of DNA methylation in mouse *Dnmt1* −/− ES cells led to spreading of polycomb marks (in particular H3K27me3): their analyses singled out the *Pcdh* genes, which were heavily methylated in WT but not mutant ESC, as also shown by others [52]. Reddington et al. [17] also showed an increase in H3K27me3. A similar sequence of events in our human cells would cause an increase in H3K27me3 on *PCDH* genes and potentially help block remethylation. The sensitivity of the protocadherin cluster to methylation changes may explain why these genes are frequently identified in screens for differentially methylated loci in cancer [53]. The lack of derepression in our stable fibroblast cells is unsurprising here since expression of these genes is restricted to neurons [54]: they are also, with the exception of part of the *PCDHG* complex, heterochromatic rather than polycomb-repressed and may as such be harder to reactivate.

Fat/body mass genes can be repressed by DNA methylation and polycomb

Currently, there is much interest in the possibility that altered diet, folate status or exposure to environmental toxins may lead to stable changes in the human methylome which particularly affect metabolic processes, as this offers an attractive mechanism by which it may be possible to partly explain the foetal origins of adult disease [55, 56]. Enrichment analysis in our cells identified the FBM genes involved in the common processes of lipid storage and body mass homoeostasis, including *LEP*, *GHSR* and the *APOC* cluster. These loci are readily demethylated on acute DNMT1 depletion and remain demethylated in chronically depleted cells where many other loci have

recovered methylation. These loci are heavily marked by polycomb in normal fibroblasts, rather than being heterochromatic, which can potentially explain both their resistance to remethylation and their lack of transcriptional depression in the stable lines. In keeping with this, inhibition of the polycomb repressor EZH2 which generates H3K27me3 marks could reactivate these genes, as well as the canonical polycomb targets the *HOX* genes. These results suggest that in cells which have both DNA methylation and polycomb-mediated repression, both layers of repression must be removed to achieve gene activation. Interestingly a recent report by Hajkova and colleagues showed that reprogramming of germ cells in mouse also required both removal of DNA methylation and alteration of polycomb marks [57].

Olfactory genes are methylated and largely inert

Olfactory receptors are also involved in specification of neural cell identity, where individual receptors are expressed in only a small group of cells in the olfactory epithelium [58]. They are largely monoallelically expressed, and methylation has been implicated as playing a role in their control [59, 60]. The OR gene family is the largest in the genome, with approx. 380 active members, many organised into "gene factories" where they are flanked by many more pseudogenes and repeats, such as the large cluster on chr11 [41]. These regions are often transcriptionally inert and heterochromatic, which together with the requirement for tissue-specific factors may explain their lack of derepression.

Cancer/testis antigen genes are particular targets for demethylation and activation

The *TSPY* and *MAGE* genes fall into a functionally defined group known as the cancer/testis antigen (CTA) genes ([61, 62]; http://www.cta.lncc.br/) which are expressed during testis development normally, but which are aberrantly expressed in some tumours, such as melanoma and gonadoblastoma (e.g. *TSPY2*). This latter property makes them of particular interest for cancer immunotherapy, and monoclonal antibodies against some CTA members have already gained clinical approval [63]. CTA genes have been shown previously to lose methylation and become derepressed in several cancer cell types after treatment with the methyltransferase inhibitor 5′aza-2-deoxycytidine (Aza) [64–66] and in the HCT116 DNMT1 mutant line [66, 67] using locus-specific approaches. Our study (1) shows in an unbiased genomic screen that CTA genes are the genes most affected by loss of maintenance activity, (2) shows this for the first time in a normal, differentiated cell line and (3) highlights the subset of CTA genes which are particularly dependent on maintenance activity to keep them repressed. It is noteworthy that the majority of these genes are on the X chromosome, which shows major fluxes in methylation in our stable lines. The genes are largely associated with heterochromatin, rather than polycomb repression, and do not respond to EZH2 inhibition, but rather directly to loss of methylation, which may reflect some difference in heterochromatin marking on the X. Strategies to demethylate and turn on these genes in tumour cells (e.g. with Aza) to facilitate cancer vaccine development may be worthwhile to pursue, given that these genes are the most responsive to loss of methylation in our cell lines.

UGT1A genes and other poised promoters are susceptible to hypermethylation

From the enrichment analysis, the *UGT1A* gene cluster was highlighted in terms of genes gaining methylation. These genes are known to be highly expressed in skin fibroblasts postnatally, and to be repressed in non-expressing tissues by methylation [68, 69]. The WT cells already had substantial levels of methylation but the increased methylation in the stable cell lines led to small but consistent decreases in transcription on the HT12 array, though levels were so low these could not be confirmed by Taqman qPCR (data not shown). It may be that the particular marks associated with a recent inactivation of the *UGT1A* cluster in the fibroblasts during adaptation to cell culture led to an increased de novo activity here, and in our transient KD experiment we saw the greatest gains in methylation at *UGT1A4*. Consistent with this, hypermethylation relative to the WT cells was associated with weak and poised promoters genome-wide, and the latter showed the greatest tendency to gain methylation above normal WT levels in the shRNA-containing lines.

Lack of transcriptional changes in part due to polycomb

It is notable that while there was widespread changes in methylation in the KD cell lines, this was not accompanied by large-scale transcriptional derepression, with only a few hundred genes showing dysregulation, and the fold change in transcription being small. Of the four gene classes identified as most affected in terms of methylation, only one—that containing the *TSPY* and *MAGE*

genes—showed robust transcriptional derepression. A lack of global changes in transcription, also reported by others [29, 70], is likely due to in part to the absence of transcription factors in fibroblasts needed to transcribe neural or adipocyte genes at high levels. However, many of the regions showing most persistent hypomethylation are polycomb-marked and this is likely to be sufficient in itself, as it is for example in *Drosophila*, to maintain repression of these genes. However, we could show that in the presence of an EZH2 inhibitor, polycomb-marked loci which lacked DNA methylation, such as those involved in fat homoeostasis/body mass regulation, became upregulated, along with canonical polycomb targets such as the HOX genes. Our results therefore indicate both that the polycomb system is sufficient in itself to repress and also that polycomb-repressed regions appear to be refractive to remethylation, which may be due to the action of FBXL10 [71]. It has previously been proposed that the two systems work in parallel, with their own sets of targets and a degree of mutual exclusivity [15–17]: our results would support such a conclusion.

Comparison to other recent work

Two recent studies have also examined the effects of DNMT1 mutation on DNA methylation and gene transcription in human, albeit in cancer cells [29, 70]. Acute depletion of DNMT1 using an siRNA-mediated approach found, as we did, regions of low CpG density (open sea, etc.) to be most affected, but differed in finding more evidence for cell morphogenesis and phosphorylation pathways being affected [70]. This might reflect differences between acute and chronic depletion and the high levels of cell death during acute depletion. Blattler and colleagues [29] also found that relatively few genes were dysregulated in *DNMT1/3B* double KO HCT116 cells, but some cancer/testis genes (the related GAGE genes) were upregulated, along with Krüppel-associated box genes, while chaperonins figured prominently among down-regulated genes. The latter two gene classes may therefore be more dependent on DNMT3B, or the combination of DNMT1 and 3B, for their maintenance; alternatively the differences may be due to the experiment being carried out in colon cancer cells rather than, as here, in non-transformed fibroblasts.

Conclusions

In conclusion, our study sheds new light on the loci which are most sensitive to sustained loss of maintenance activity in humans and shows an interplay between polycomb and DNA methylation-mediated repression in these differentiated cells.

Additional files

Additional file 1: Table S1. Details of the primers used in this study.

Additional file 2: Figure S1. Variation between shRNA clonal lines. **(A)** Relative similarities between cell lines based on principal component analysis (PCA) of the 450K data; three independent cultures of each line were analysed. Note the clustering of lines d8R and d10R. The fraction of total variance explained by each component is indicated in brackets. (**B**) The 1000 sites most variably methylated between cell lines were used for hierarchical clustering. The location of sites with respect to CpG island is indicated at left. Beta values are depicted as shades from red (low) to blue (high).

Additional file 3: Figure S2. Changes in methylation levels by genomic element. **(A)** Protein levels in knockdown lines by western blotting. As a control HCT116 colon cancer cells which are WT or have a homozygous mutation in *DNMT1* (KO) are shown: the DNMT1-specific top band is indicated by the arrowhead at right. **(B)** Median levels of methylation are shown for each genomic element (listed at top). The positions of medians are also indicated at right (arrowheads). The differences between WT and KD medians were used to plot Fig. 1d. **(C)** Density distribution of methylation at the three main elements involved in gene regulation, shown by cell line. Demethylation seems most marked at gene bodies (Genes), indicated by increased density of probes at low methylation (β) values.

Additional file 4: Figure S3. Further analysis of enriched genes. **(A)**Total numbers of sites showing significant changes in methylation at different false discovery rates (FDR). Some sites showing gain were found in each KD cell line alongside the more numerous sites showing loss. **(B)** Differential methylation between WT and all KD lines using the 1000 best-ranking sites as identified by RnBeads (red). The majority of high-scoring sites common to all three lines lost methylation, but approx. one-third showed gain. **(C)** Methylation changes at neural identity genes on chromosome 5. Protocadherins in the α and γ families (*PCDHA* and *PCDHG* genes) have a clustered arrangement, while genes for the β family members are arranged individually. Tracks are as in Fig. 3. The position of the C class variable exons in the *PCDHA* and *PCDHG* clusters are also shown: gain in methylation relative to the siRNA-treated cells can be seen in the boxed regions, which includes the *PCDHG* constant exons, corresponding to transcriptionally active chromatin (green). **(D)** Median β values for gene bodies for olfactory receptors identified by DAVID: differences were significant by Mann-Whitney U (MWU). **(E)** Median β values for the promoters of genes in the histone modifier group identified by enrichment analysis in Table 1. No significant differences between WT and KD were found by MWU.

Additional file 5: Table S2. Details of the hypomethylated and hypermethylated genes from Figs. 3d and 5a, respectively.

Additional file 6: Table S3. Details of the genes showing transcriptional changes in KD cell lines from Fig. 6c.

Additional file 7: Figure S4. Role of DNMT3B in hTERT1604. **(A)** DNMT3B mRNA levels from the HT12 transcription array (3 probes) did not differ substantially in *DNMT1* shRNA cell lines from WT cells. **(B)** Successful depletion of *DNMT3B* mRNA using siRNA for 48hr, versus a scrambled control (Scr). **(C)** Methylation levels by pyroassay at the indicated loci: KD, knockdown. Methylation levels at 72hr were similar (not shown).

Authors' contributions

KON and REI supervised and carried out the majority of wet laboratory work and assembled figures; SJM and SJT carried out the majority of the bioinformatics analyses; AT carried out the DNMT3B work; CB and LM contributed results on specific loci; JL carried out initial KD experiments; DGM supervised and carried out bioinformatics analyses; CPW designed the experiments, carried out bioinformatics analyses, interpreted results and wrote the MS. All authors read and approved the final manuscript.

Author details
[1] Genomic Medicine Research Group, Centre for Molecular Biosciences, School of Biomedical Sciences, Ulster University, Cromore Road, Coleraine BT52 1SA, UK. [2] Academie Life Science, Engineering & Design, Saxion University, M.H. Tromplaan 28, 7500 Enschede, Netherlands. [3] Department of Obstetrics and Gynecology, University of Sassari, Via Vienne 2, 7100 Sassari, Italy. [4] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast BT9 7AE, UK. [5] Present Address: The Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast BT9 7AE, UK. [6] Present Address: Terry Fox Laboratory, BC Cancer Research Centre, 675 West 10th Avenue, Room 13-112, Vancouver, BC V5Z 1L3, Canada.

Acknowledgements
We thank Julien Bauer, Daniel Harkin and Steven McLoughlin for help with bioinformatics analysis, Andrew Irwin, Lee McCahon, Anna McLaughlin and Bob Goodman for technical help, Paul Thompson for advice and other members of the laboratory for critical comments. We are obliged to Bert Vogelstein for the HCT116 WT and DKO cells.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
Data from the 450K and HT-12 arrays have been deposited with the Gene Expression Omnibus database at the National Centre for Biotechnology Information, USA, under the Series number GSE90012. Supplementary Figures and Tables are available in the online version. Cell lines or other materials are available from the corresponding author on request.

Ethics approval and consent to participate
Not applicable.

Funding
Work in the Walsh laboratory was funded by grants from the Medical Research Council (MR/J007773/1) and the ESRC/BBSRC (ES/N000323/1).

Publisher's Note

Received: 4 January 2018 Accepted: 21 March 2018
Published online: 29 March 2018

References

1. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev. 2013;14:204–20.
2. Edwards JR, Yarychkivska O, Boulard M, Bestor TH. DNA methylation and DNA methyltransferases. Epigenetics Chromatin. 2017;10:23. https://doi.org/10.1186/s13072-017-0130-8
3. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993;366:362–5.
4. Beard C, Li E, Jaenisch R. Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev. 1995;9:2325–34.
5. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39:457–66.
6. Krebs AR, Dessus-Babus S, Burger L, Schübeler D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. Elife [Internet]. 2014;3:e04094.
7. Wachter E, Quante T, Merusi C, Arczewska A, Stewart F, Webb S, et al. Synthetic CpG islands reveal DNA sequence determinants of chromatin structure. Elife [Internet]. 2014;3:1–16.
8. Walsh CP, Bestor TH. Cytosine methylation and mammalian development. Genes Dev [Internet]. 1999;13:26–34.
9. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet [Internet]. 2009;41:178–86.
10. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet [Internet]. 2012;13:484–92.
11. Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F, et al. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell. 2013;155:121–34.
12. Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science. 2010;329:444–8.
13. Irwin RE, Thakur A, O'Neill KM, Walsh CP. 5-Hydroxymethylation marks a class of neuronal gene regulated by intragenic methylcytosine levels. Genomics [Internet]. 2014;104:383–92.
14. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature [Internet]. 2011;469:343–9.
15. Lynch MD, Smith AJH, De Gobbi M, Flenley M, Hughes JR, Vernimmen D, et al. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J [Internet]. 2012;31:317–29.
16. Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet. 2010;6:1–14.
17. Reddington JP, Perricone SM, Nestor CE, Reichmann J, Youngson NA, Suzuki M, et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol. 2013;14:R25.
18. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem [Internet]. 2005;74:481–514.
19. Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 1996;122:3195–205.
20. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69:915–26.
21. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19:219–20.
22. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99:247–57.
23. Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet. 2001;27:31–9.
24. Liao J, Karnik R, Gu H, Ziller MJ, Clement K, Tsankov AM, et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet. 2015;47:469–78.
25. Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, et al. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet. 2007;39:391–6.
26. Rhee I, Jair KW, Yen RW, Lengauer C, Herman JG, Kinzler KW, et al. CpG methylation is maintained in human cancer cells lacking DNMT1. Nature. 2000;404:1003–7.
27. Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, et al. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature. 2002;416:552–6.
28. Egger G, Jeong S, Escobar SG, Cortez CC, Li TW, Saito Y, et al. Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc Natl Acad Sci USA. 2006;103:14080–5.
29. Blattler A, Yao L, Witt H, Guo Y, Nicolet CM, Berman BP, et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. Genome Biol [Internet]. 2014;15:469.
30. Loughery JE, Dunne PD, O'Neill KM, Meehan RR, McDaid JR, Walsh CP. DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response. Hum Mol Genet. 2011;20:3241–55.
31. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98:288–95.
32. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods. 2014;11:1138–40.

33. Ouellette MM, McDaniel LD, Wright WE, Shay JW, Schultz RA. The establishment of telomerase-immortalized cell lines representing human chromosome instability syndromes. Hum Mol Genet. 2000;9:403–11.
34. Rutledge CE, Thakur A, O'Neill KM, Irwin RE, Sato S, Hata K, et al. Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes. Development. 2014;141:1313–23.
35. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.
36. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25:288–9.
37. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocol. 2009;4:44–57.
38. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 2005;15:1451–5.
39. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, et al. The UCSC genome browser database. Nucleic Acids Res. 2003;31:51–4.
40. Kondo T, Bobek MP, Kuick R, Lamb B, Zhu X, Narayan A, et al. Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Hum Mol Genet. 2000;9:597–604.
41. Glusman G, Yanai I, Rubin I, Lancet D. The complete human olfactory subgenome. Genome Res. 2001;11:685–702.
42. Lau YFC, Li Y, Kido T. Gonadoblastoma locus and the TSPY gene on the human y chromosome. Birth Defects Res Part C Embryo Today Rev. 2009;87:114–22.
43. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature [Internet]. 2011;473:43–9.
44. Fujiwara T, Saitoh H, Inoue A, Kobayashi M, Okitsu Y, Katsuoka Y, et al. 3-Deazaneplanocin A (DZNep), an inhibitor of S-adenosylmethionine-dependent methyltransferase, promotes erythroid differentiation. J Biol Chem. 2014;289:8121–34.
45. Tucker KL, Beard C, Dausmann J, Jackson-Grusby L, Laird PW, Lei H, et al. Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev. 1996;10:1008–20.
46. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. 2012;22:467–77.
47. Yagi T. Molecular codes for neuronal individuality and cell assembly in the brain. Front Mol Neurosci [Internet]. 2012;5:45.
48. Rubinstein R, Thu CA, Goodman KM, Wolcott HN, Bahna F, Mannepalli S, et al. Molecular logic of neuronal self-recognition through protocadherin domain interactions. Cell. 2015;163:629–42.
49. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet [Internet]. 2010;42:1093–100.
50. Toyoda S, Kawaguchi M, Kobayashi T, Tarusawa E, Toyama T, Okano M, et al. Developmental epigenetic modification regulates stochastic expression of clustered Protocadherin genes, generating single neuron diversity. Neuron. 2014;82:94–108.
51. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature. 1999;402:187–91.
52. Otani J, Kimura H, Sharif J, Endo TA, Mishima Y, Kawakami T, et al. Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. PLoS ONE. 2013;8:e82961.
53. van Roy F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat Rev Cancer [Internet]. 2014;14:121–34.
54. Chen WV, Maniatis T. Clustered protocadherins. Development [Internet]. 2013;140:3297–302.
55. Barker DJP. The developmental origins of chronic adult disease. Acta Paediatr Suppl. 2004;93:26–33.
56. Irwin RE, Pentieva K, Cassidy T, Lees-Murdock DJ, McLaughlin M, Prasad G, et al. The interplay between DNA methylation, folate and neurocognitive development. Epigenomics [Internet]. 2016;8:863.
57. Hill PWS, Leitch HG, Requena CE, Sun Z, Amouroux R, Roman-Trufero M, et al. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature [Internet]. 2018;555:392–6.
58. McClintock TS. Achieving singularity in mammalian odorant receptor gene choice. Chem Senses. 2010;35:447–57.
59. MacDonald JL, Gin CSY, Roskams AJ. Stage-specific induction of DNA methyltransferases in olfactory receptor neuron development. Dev Biol. 2005;288:461–73.
60. Colquitt BM, Markenscoff-Papadimitriou E, Duffie R, Lomvardas S. Dnmt3a regulates global gene expression in olfactory sensory neurons and enables odorant-induced transcription. Neuron. 2014;83:823–38.
61. Simpson AJG, Caballero OL, Jungbluth A, Chen Y-T, Old LJ. Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer. 2005;5:615–25.
62. Almeida LG, Sakabe NJ, de Oliveira AR, Silva MCC, Mundstein AS, Cohen T, et al. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res. 2009;37:D816.
63. Gjerstorff MF, Andersen MH, Ditzel HJ. Oncogenic cancer/testis antigens: prime candidates for immunotherapy. Oncotarget. 2015;6:15772–87.
64. Samlowski WE, Leachman SA, Wade M, Cassidy P, Porter-Gill P, Busby L, et al. Evaluation of a 7-day continuous intravenous infusion of decitabine: inhibition of promoter-specific and global genomic DNA methylation. J Clin Oncol. 2005;23:3897–905.
65. Karpf AR. A potential role for epigenetic modulatory drugs in the enhancement of cancer/germ-line antigen vaccine efficacy. Epigenetics. 2006;1:116–20.
66. Koslowski M, Bell C, Seitz G, Lehr HA, Roemer K, Müntefering H, et al. Frequent nonrandom activation of germ-line genes in human cancer. Cancer Res. 2004;64:5988–93.
67. James SR, Link PA, Karpf AR. Epigenetic regulation of X-linked cancer/germline antigen genes by DNMT1 and DNMT3b. Oncogene. 2006;25:6975–85.
68. Belanger AS, Tojcic J, Harvey M, Guillemette C. Regulation of UGT1A1 and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells. BMC Mol Biol. 2010;11:9.
69. Sumida K, Kawana M, Kouno E, Itoh T, Takano S, Narawa T, et al. Importance of UDP-glucuronosyltransferase 1A1 expression in skin and its induction by UVB in neonatal hyperbilirubinemia. Mol Pharmacol. 2013;84:679–86.
70. Tiedemann RL, Putiri EL, Lee J-H, Hlady RA, Kashiwagi K, Ordog T, et al. Acute depletion redefines the division of labor among DNA methyltransferases in methylating the human genome. Cell Rep [Internet]. 2014;9:1554–66.
71. Boulard M, Edwards JR, Bestor TH (2015) FBXL10 protects polycomb-bound genes from hypermethylation. Nat Genet 47(5):479–85.

PAPER-III

Comparison of DNMT1 inhibitors by methylome profiling identifies unique signature of 5-aza-2'deoxycytidine.

Sarah-Jayne Mackin, Karla M. O'Neill, and Colum P. Walsh.

Epigenomics. 2018 Aug; 10(8):1085-1101.

doi: 10.2217/epi-2017-0171

The main aims of this paper were to:

1. Identify gene targets specifically sensitive to DNMT1 inhibitors in normal human cell lines by pharmacological depletion of DNMT1 using Aza.
2. Compare the depletion of DNMT1 using Aza to that of a transient knockdown of DNMT1 using DNMT1 targeting siRNA.
3. Identify any novel effects of DNMT1 depletion on global methylation and on specific genomic regions.

CONTRIBUTION

I performed all of the bioinformatic analysis on the transiently and pharmacologically DNMT1 depleted human fibroblastic cell lines (created by co-author Karla O'Neill) using a modular methylation analysis package called RnBeads with some tailored script in the R interface. Using the differential methylation results from the initial bioinformatic analysis, I performed all the downstream gene enrichment, and subsequent statistical analysis of differentially methylated gene classes, in addition to screening the same gene classes in complementary transcription array data of the same cell lines. I created all the figures for the manuscript using either RnBeads or Adobe Illustrator and I wrote the majority of the manuscript and created the GEO submission to deposit all of the 450k methylation data and associated metadata sheets

Summary of the major findings;

1. Treatment of non transformed hTERT cells with Aza, and DNMT1-targeting siRNA both caused an overall hypomethylation event.

2. The hypomethylation caused by Aza was particulary enriched at the promoters of histone genes.

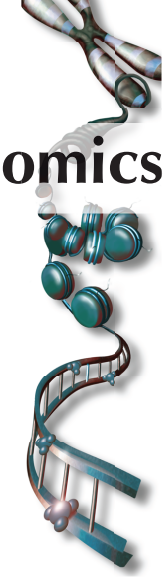
3. A common set of genes enriched for proteins involved in phosphorylation & acetylation were affected by both treatments.

4. Gains of methylation were observed in Aza-treated hTERTs particulary at GPCRs.

Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics



Comparison of DNMT1 inhibitors by methylome profiling identifies unique signature of 5-aza-2′deoxycytidine

Sarah-Jayne Mackin[1], Karla M O'Neill[1,2] & Colum P Walsh*[,1]

[1]Genomic Medicine Research Group, Centre for Molecular Biosciences, Biomedical Sciences Research Institute, Ulster University, Coleraine, BT52 1SA, UK

[2]Current address: The Wellcome–Wolfson Institute for Experimental Medicine, Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7AE, UK

*Author for correspondence: cp.walsh@ulster.ac.uk

Aim: 5-aza-2′deoxycytidine (Aza) is used to treat myelodysplastic syndrome and is in trials for other cancers. It acts chiefly as a hypomethylating agent inhibiting DNMT1. A lack of understanding of off-target effects in normal cells hinders wider usage. **Materials & methods:** We compared treatment of the same normosomic, nontransformed fibroblast cell line with Aza and SMARTpool siRNA against *DNMT1*. Methylation and transcription were assayed using Illumina 450k and HT12 arrays. **Results:** Both Aza and DNMT1 siRNA caused overall hypomethylation, with siRNA more efficient at demethylating gene bodies. Hypomethylation at the promoters of many histones, and hypermethylation at multiple sites genome wide, were unique to Aza treatment. **Conclusion:** Aza had important unique effects and targets compared with DNMT1 inhibition via siRNA.

First draft submitted: 19 December 2017; Accepted for publication: 23 February 2018; Published online: 31 July 2018

Keywords: cancer epigenetics • DNA methylation • DNA methyltransferase

Aberrations in DNA methylation are a major characteristic of cancer progression [1,2] and in particular of some hematological malignancies such as myeloid dysplastic syndromes (MDS) [3]. DNA methylation is a post-translational modification which serves as an epigenetic method to regulate transcription, modulate chromatin structure and repress transposable elements [4]. DNA methylation is established through the addition of a methyl group from a donor molecule of *S*-adenosyl-methionine to the 5′ position on a cytosine within a cytosine-guanine (CG) dinucleotide [5]. Current epigenetic treatments are thought to mainly affect DNMT1, one of the three functional methyltransferases in human and chiefly responsible for the maintenance of DNA methylation and ensuring the integrity of methylation post replication [6,7].

There are currently a range of known DNMT1 inhibitors such as the nucleoside inhibitors 5-aza-2′ deoxycytidine (Aza), 5-azacytidine and zebularine [7,8], all of which are modified at the 5′ position of their pyrimidine ring. Aza in particular has been found to be the best nucleoside inhibitor to promote DNA demethylation [9] and in this role as a hypomethylating agent it has become US FDA- and UK NICE -approved (as decitabine or dacogen) for the treatment of MDS [3,10], and is currently under clinical trial for use in acute myeloid leukemia (AML) patients who are ineligible for intensive chemotherapy [11]. Around 15% of confirmed patients with MDS respond to epigenetic therapy through an improved blood cell count, particularly a smaller count of malignant cells [12,13]. This epigenetic treatment with Aza has been shown to prolong the overall survival while lowering the risk of progression to the leukemic stages in patients with higher risk MDS, in comparison to MDS that is treated conventionally [7,14]. However, despite clinical success with use of this pharmacological agent the mode of action behind Aza and its effect on myeloid malignancy remains unclear.

Aza functions so well as a hypomethylating agent because it acts as a cytosine analog and can be easily incorporated into the replicating DNA in proliferating cells and acts as a suicide substrate. This phenomenon is due to the irreversible covalent bonding of the cytosine analog, chiefly with the DNA methyltransferase DNMT1 [15,16], which is the main activity present in almost all cells, although DNMT3A and DNMT3B, which are present at low levels in

Future Medicine

most adult tissues [17] may also be affected. With DNMT1 now unable to function as maintenance methyltransferase, normal methylation is prevented at CG dinucleotides of newly synthesized DNA and the enzyme becomes targeted for premature degradation [18]. This ultimately leads to a hypomethylation, most likely through excision of the enzyme: analog: DNA conjugate by DNA repair enzymes and synthesis of a new, unmethylated replacement [16]. The widespread demethylating effect of Aza on multiple genes across many cell lines and tissues, as shown both by ourselves [19–21] and others [7,9,15,22] is therefore well established. The nucleoside inhibitor 5-azacytidine on the other hand becomes preferentially incorporated into the RNA [23,24]. Genome-wide demethylation in cancerous cells at promoter regions in particular is one of the primary aims of oncological epigenetic therapy. In cell lines, methylation levels at transcription start sites can be pharmacologically altered to cause hypomethylation and consequent transcriptional re-activation of epigenetically silenced tumor suppressor genes (TSGs) [7,22,25].

However, despite clinical success in the treatment of MDS by prolonging the time to leukemic transformation [24] specific DNMT inhibitors that are both efficient and have clearly elucidated gene targets have not yet been identified: this is in part due to different genetic backgrounds in distinct cancers [26]. There is a real need for baseline data in a normal cell which can be compared with direct DNMT1 inhibition. Here we aimed to investigate the exact gene targets that were particularly sensitive to changes in methylation post-treatment with Aza in the normosomic, nontransformed human fibroblast cell line hTERT-1604, comparing directly to a transient knockdown of DNMT1 in the same cell line. To allow quantitation of smaller effects, we used the Infinium HumanMethylation 450k BeadChip Array, which can very reliably measure small effects [27].

Material & methods

Cell culture

hTERT-1604 fibroblasts were cultured as previously described [21] in 4.5 g/l glucose Dulbecco's Modified Eagle's Medium (D-MEM) supplemented with 10% Fetal Bovine Serum (FBS) and 2× Non-Essential Amino Acids (NEAA). Pharmacological inhibition of DNMT1 with Aza (sold as 5-aza-2′ deoxycytidine [Sigma-Aldrich, Dorset, UK] and resuspended in medium) and transient DNMT1 depletion using siRNA (SMARTpool, Thermo Fisher Scientific, Loughborough, UK) in hTERT-1604 fibroblasts were also carried out as described therein, along with subsequent validation. Briefly, for Aza treatment 1×10^6 hTERT-1604 cells were seeded onto a 90 mm plate in complete medium and allowed to attach overnight, before replacing with medium containing 1 uM Aza, or medium alone (untreated [UT]). Medium +/- Aza was replaced at 24 and 48 h, and cells harvested at 72 h. For siRNA, 1×10^6 cells were resuspended in 2 ml of medium containing 0.2 uM final concentration of targeting siRNA or a scrambled control (Thermo Fisher Scientific), then allowed to attach in the well for 24 h before replacing with fresh medium without siRNA; cells were harvested at 72 h.

DNA extraction & pyrosequencing

DNA was extracted from each of the cell lines in log phase using the Genomic DNA Purification Kit (Fermentas, now Thermo Fisher Scientific) as per manufacturer's instructions. The concentration and purity of the DNA extracted from each cell line were measured using a Nanodrop ND-1000 spectrophotometer (Labtech, Ringmer, UK). Integrity of the DNA samples was determined by running 3 µl of each sample on a 1% agarose gel with 1 kb ladder (Invitrogen, Paisley, Scotland) by gel electrophoresis and imaged using UV light in the Kodak Gel Logic 200 imaging system (Carestream Health UK Ltd, Hemel Hempsted, UK) . An initial aliquot (500 ng) of DNA was used to test the effectiveness of the treatments by bisulfite conversion using the EpiTect Bisulfite Kit (Qiagen, Crawley, UK) according to the manufacturer's instructions followed by pyroassay, as was described previously [21]: briefly, an initial PCR was carried out using the PyroMark kit (Qiagen) before pyrosequencing on the PyroMark Q24 according to the manufacturer's instructions. Details of assays are given in Rutledge *et al.* [21]: average methylation across all CpG (DPEP-5CpG, SYCP3- 4CpG, DAZL- 4CpG) is shown; experiments were carried out in triplicate with at least one biological repeat.

Illumina Human Methylation 450k BeadChip Array

Prior to array analysis the DNA underwent additional quantification using PicoGreen® dsDNA quantification assay (Thermo Fisher Scientific). For this, 2.5 µl PicoGreen and 497.5 µl 1 × TE buffer (10mM Tris (pH 8), 1mM EDTA) were added together and combined to a solution of 3 µl DNA and 497 µl 1 × TE buffer. This combined 1 ml solution was left to incubate for 5 min. Following incubation, 150 µl of the solution was aliquoted into six wells of a 96-well plate (Orange Scientific, Braine-l'Alleud, Belgium) per DNA sample and fluorescence

read on the plate reader at 485–520 nm. The subsequent fluorescence values were incorporated into a calculation that included information gained via the generation of a standard curve using λ DNA (effectively the R^2 value from the slope of the line). The DNA at a concentration of 50 ng/μl was sent to Cambridge Genomic Services (CGS; Cambridge, UK) who bisulfite converted the DNA in-house (Zymo Research kit, Cambridge Bioscience, Cambridge, UK) prior to assessing methylation levels on the Infinium Human Methylation 450k BeadChip Array (Illumina, Chesterford, UK). The outputs from the DNA methylation array were subsequently scanned using an Illumina iScan (Illumina) by CGS.

RNA extraction & reverse transcription quantitative PCR

RNA was extracted from each sample using the RNeasy Mini Kit (Qiagen) as per the manufacturer's instructions: concentration and purity was checked using the Nanodrop ND-1000 spectrophotometer (Labtech International, Ringmer, UK) and integrity by running on a 1% agarose gel and imaged as before. Prior to sending the samples the RNA was quantified using the RiboGreen® quantification assay (Thermo Fisher Scientific). The reverse transcription quantitative PCR protocol used for the test loci (*DAZL*, *DPEP3*, *SYCP3*) and *DNMT1* was described in Rutledge *et al.* [21], primers as detailed therein except *DNMT1* F: GTGGGGGACTGTGTCTCTGT and R: TGAAAGCTGCATGTCCTCAC (204 bp).

HT12 transcription array

The RNA extracted from each sample was concentrated using a SpectroStar (BMG, Labtech, Aylesbury, UK) and a Bioanalyzer (Agilent Technologies, Cheadle, UK) by CGS before using the HumanHT-12 v4 Expression BeadChip (Illumina) to examine gene expression. Two hundred nanogram of RNA was required as an input per reaction and was subjected to linear amplification with the Illumina TotalPrep RNA Amplification Kit (Life Technologies, Paisley, UK). The resultant cDNA was screened for concentration, integrity and purity using the above mentioned SpectroStar and Bioanalyzer. Screened cDNA was then hybridized to the BeadChip overnight, followed by washing, staining and scanning using the BeadArray Reader (Illumina) as per manufacturer's instructions.

Bioinformatic analysis of methylation & transcription array data

The methylation array was analyzed by calculating the β-value. The β is calculated as the signal of the methylation-specific probes over the sum of the signals from both the methylation- and unmethylated-specific probes [28]. Total methylation of a nominated CG dinucleotide is given as a β of one, whereas total absence of methylation is represented by a β of zero. Generally probes that have signals where the p-value >0.05 are removed before downstream analysis is conducted. GenomeStudio (Illumina v3.2) was used by CGS after scanning the methylation and transcriptional array data for initial data processing. Subsequently the RnBeads (v1.2.2) package [28] was used to further process the methylation array data using R (version 3.2.2 [2015–08–14]) in R Studio (version 0.99.903). The package offers a tailored pipeline approach for preprocessing, normalization, quality control, exploratory analysis, differential methylation and enrichment analysis. Each analysis module populates an interactive html to facilitate analysis. An initial principal component analysis using three siRNA- and three Aza-treated samples was used to obtain the most tightly clustered samples and data was then renormalized with using two siRNA and two Aza samples.

The sample, and control probes belonging to the HT12 transcription array data were also analyzed in R in RStudio using the lumi (2.22.0) and the limma package (3.26.3). Probes that failed to have intensity values in this case that were not significantly different ($p > 0.01$) from the allocated negative controls were removed. Lumi Variance Stabilization Transformation was used to transform the data, followed by quantile normalization. Limma was used to perform false discovery rate (FDR) testing.

Gene ontology

Functional classification of enriched gene lists identified through either methylation or transcription array analysis was performed using the enrichment module in RnBeads, and in some cases also using the online tool DAVID (Database for Annotation, Visualization and Integrated Discovery) v6.8 [29] and the PANTHER v12.0 classification system [30].

10.2217/epi-2017-0171

Data visualization in Galaxy

To allow comparison of gene features aligned with their associated probes browser extensible data (BED) and bedGraph files were generated. Methylation values were visualized by converting the decimal β-value (0.0–1.0) into integers (1–1000) by multiplying by 1000, then intervals were compared using custom scripts in Galaxy, an open-source web-based platform [31]. The interface allowed the mapping of probes to the Refseq human reference genome (hg38) in the University of California at Santa Cruz (UCSC) genome browser [32] and quantification of average β-levels across genomic regions, etc.: tables of data were extracted and further statistical analysis and boxplots generated using SPSS (v22.0, IBM UK Ltd, Portsmouth, UK). In addition to plotting actual β-scores, Δβ was plotted to show the changing methylation levels across the assembly.

Results

Overall hypomethylation was greater in cells treated with siRNA than Aza

In the present investigation, methylation changes after treatment with 5′-aza-2-deoxycytosine (Aza) or with a more specific inhibitor of DNMT1 were compared in hTERT-1604 cells. We availed of cells which we had previously treated with Aza and siRNA, where we had shown that both treatments cause demethylation and reactivation of key indicator genes known to be transcriptionally repressed by DNA methylation as described [21]. This cell line is normosomic, nontransformed and differentiated, making it a better model for normal tissue than the majority of cell lines. Briefly, cells were treated with Aza (1 μM) in fresh medium each day for 3 days, or with siRNA against *DNMT1* once on day 1, then harvested at 72 h for DNA and RNA (Figure 1A). We found both demethylation and reactivation for specific germline genes including *DAZL*, *SYCP3* and *DPEP3* with both Aza and *DNMT1* siRNA using locus-specific assays (Figure 1B) as described previously [21]. Here, we extended our study to examine genome-wide methylation levels using the Illumina 450k array, with data processing using the RnBeads pipeline in R. Principal component analysis of probes passing the quality control steps confirmed the tight clustering of samples according to treatment, with PC1 and PC2 explaining 55.33 and 25.47% of the variance, respectively (total 80.80%) (Figure 2A). It can be seen that siRNA samples are more different from UT than the Aza samples are to the UT samples. To further investigate the pattern of methylation change across the distinctly segregated sample groups, hierarchical clustering was first performed across the 10,000 most variable sites (Figure 2B) followed by 3000 of the most variable promoters that experienced a change in methylation (Figure 2C). The quantitative output from the BeadChip array is given as a decimal value β ranging from 0.0 to 1.0, representative of methylation at each CpG assessed, with high values indicating high levels of methylation and vice versa. Both clustering maps employ Euclidean distance as their dissimilarity metric and in each case the *DNMT1* siRNA-treated samples show a greater density of probes with lower methylation in comparison to either the UT or Aza-treated samples. This pattern was echoed in the regional methylation profiles in Figure 2D where the siRNA samples showed increased numbers of probes with lower β-values across gene bodies, CpG islands and promoters, and reduced numbers of probes with high β-values. There was a small plateau effect visible in gene body probes around 25% methylation in the siRNA samples where many more probes exhibited an intermediate level of methylation in comparison to that of the UT or Aza-treated samples. As expected Aza-treated samples also exhibited hypomethylation across all three of the regions (Figure 2D) but the overall profile shape remained similar to that of the UT, with the demethylating effect less pronounced than that of the *DNMT1*-targeting siRNA.

Distinct gene classes are most affected by the different treatment types

Differential methylation analysis was carried out between UT and both the siRNA- and the Aza-treated samples. Scatter plots showing probes with FDR-adjusted p-values <0.05 were generated as part of this analysis (Figure 3A) and indicated that treatment with siRNA caused more individual probes to significantly lose methylation than treatment with Aza (Figure 3A), and that there were a large number of probes that also gained methylation in the Aza-treated cells. This was also evident when examining methylation at the promoters, in particular when volcano plots were generated using the change in β (Δβ) values (equivalent to difference in methylation vs UT) between treatments (Figure 3B). For samples treated with siRNA, the promoters were seen to mostly lose methylation on the whole (from β 0.00 to -0.4) with only some gaining methylation (from β 0.00 to +0.3). We noticed, however, that with Aza there is nearly an equal spread of promoters that are both gaining and losing methylation when compared with UT β-values (0.0). In addition to the differences in direction of methylation change the magnitude (y-axis) also differed with siRNA more potent at inducing hypomethylation at the promoters than Aza (adj. p-values up to

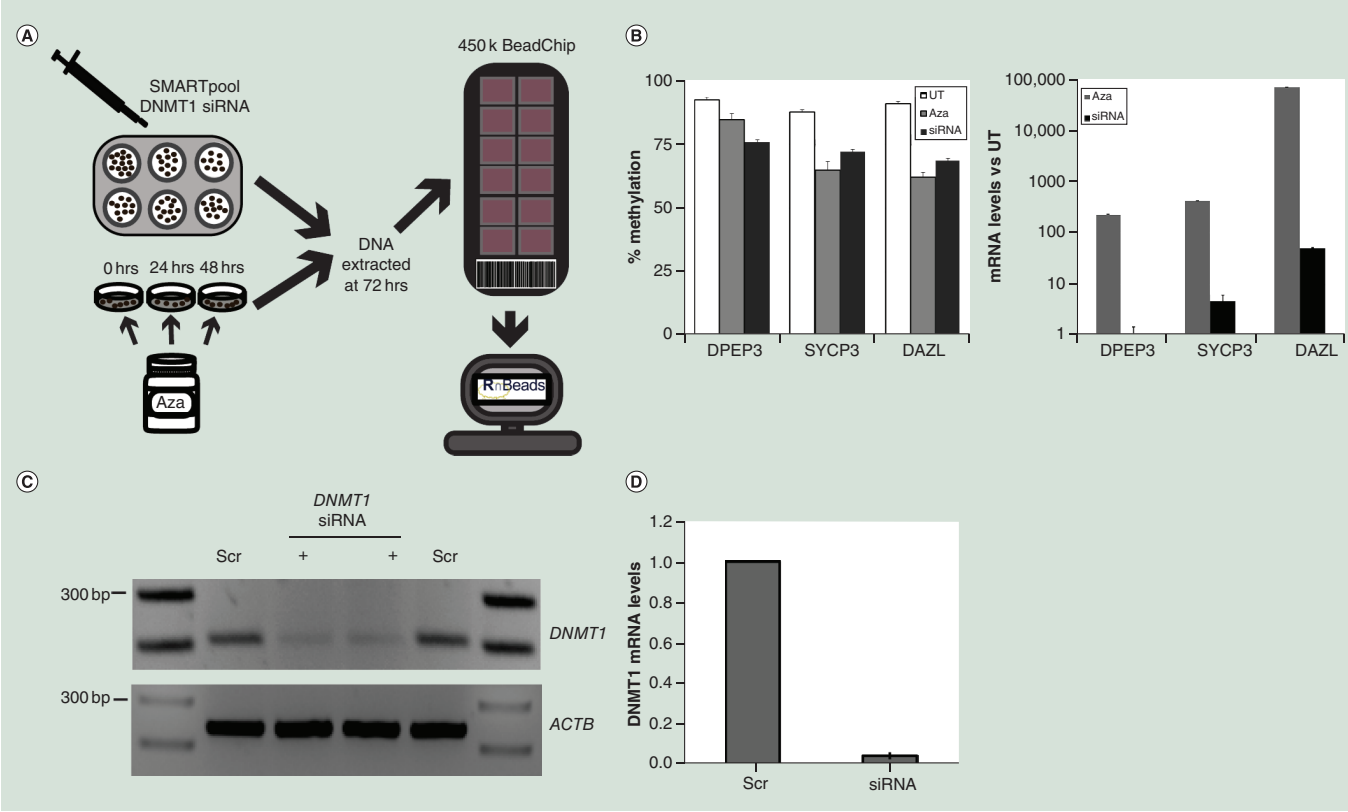


Figure 1. Schematic of experimental approach for comparative treatment of hTERT-1604 with 5-aza-2′deoxycytidine and *DNMT1* siRNA, and confirmation of effects. (A) The normosomic, nontransformed fibroblast cell line hTERT-1604 was treated over 3 days with 5-aza-2′deoxycytidine, or once on day 1 with SMARTpool siRNA targeting *DNMT1* respectively, and untreated hTERT-1604 used as a control. DNA was collected for analysis of methylation levels using Illumina 450k BeadChip methylation arrays. Output IDAT files were analyzed in R using the RnBeads pipeline and in-house scripts. **(B)** Methylation at three indicator genes showed decreases after both treatments compared with untreated cells as measured by pyrosequencing assay (left) and mRNA levels increased in reverse transcription quantitative PCR assays consistent with derepression (right). Methylation is the average across all CpG assayed; error bars are standard error of the mean; all differences were significant at $p < 0.05$ except *DPEP3* mRNA levels for siRNA **(C)** Reverse transcription-PCR showing effectiveness of siRNA targeting *DNMT1*. Scr, scrambled control siRNA; *ACTB* is a control gene; a 100 bp ladder was used **(D)** Quantification of *DNMT1* decrease using reverse transcription quantitative PCR: Scr, scrambled siRNA control. Difference was significant by t-test ($p < 0.01$).
(B) Reproduced with permission from [21].

10^{-4} vs 10^{-2}) (Figure 3B). We then wanted to explore if any of these probes experiencing hypomethylation were enriched in any particular gene class.

A hypergeometric enrichment analysis was carried out in RnBeads and the top three enriched gene ontology (GO) terms for both biological process and molecular function identified for promoters and genes experiencing hypomethylation in each treatment (Figure 3C). The three enriched GO terms for promoters following siRNA treatment showed considerable overlap, that is, detection/sensory perception of chemical stimulus. Many of the targets in these categories are olfactory receptors, a large class of G protein-coupled receptors. When looking at which gene bodies are enriched in probes losing methylation, related terms also surface (odorant binding, GDP metabolic activity), but other terms are also identified including tyrosine phosphorylation of STAT protein and cytokine receptor activity. Interestingly there was no overlap between the top enriched categories affected by the Aza treatment among both promoters and genes. Instead two themes appeared evident – terms associated with DNA replication and histone interactions.

To help visualize these changes in methylation across the genome, we created user-defined tracks viewable on the UCSC genome browser with the use of the Galaxy interface. Figure 3D shows an example locus with significant methylation changes between siRNA-treated cells and UT at *TLR2*, which belongs to the enriched group 'Detection of Chemical Stimulus' and exhibits demethylation (minimum -11%) at multiple probes spanning the CpG islands

10.2217/epi-2017-0171

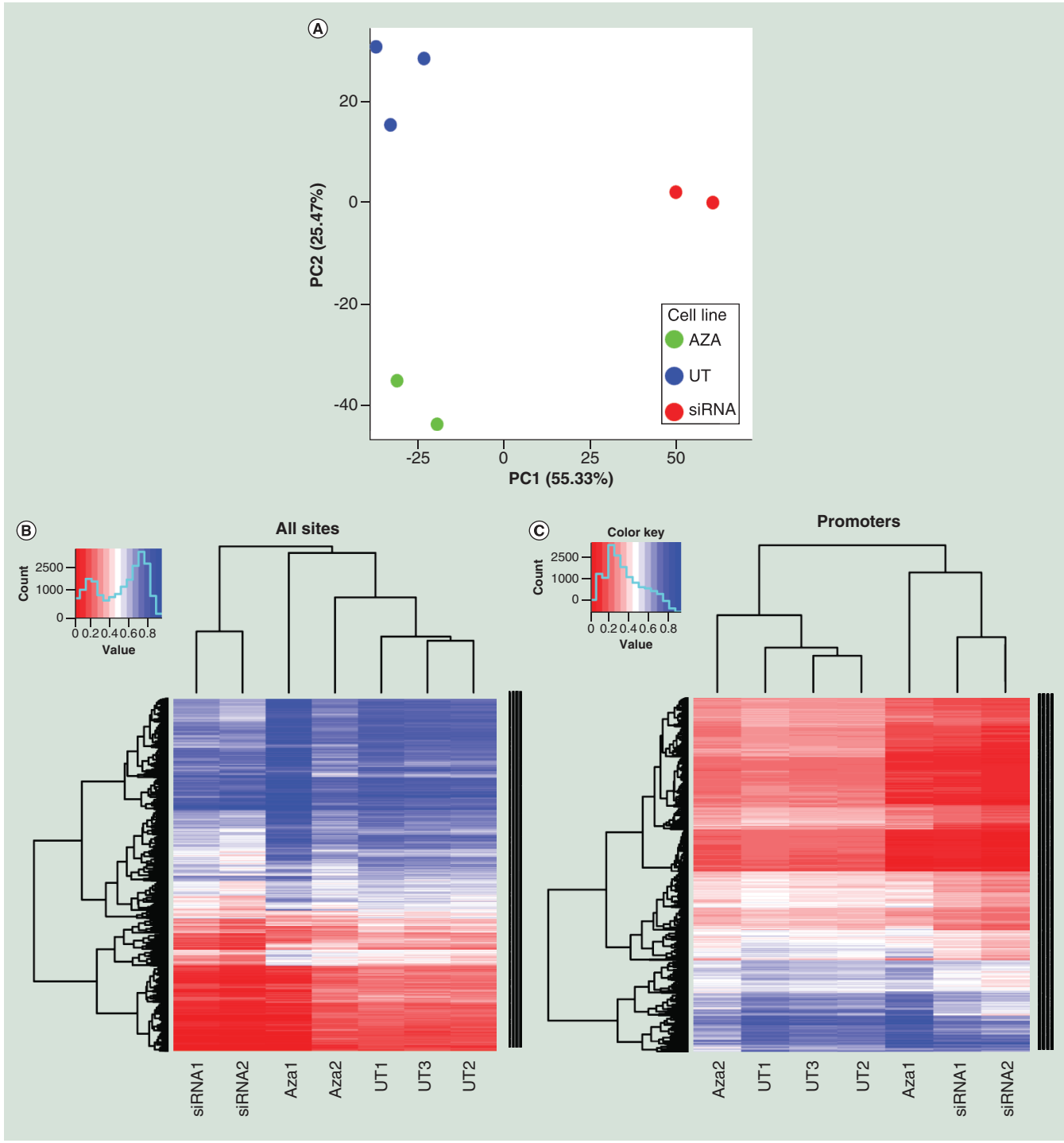


Figure 2. 5-aza-2′deoxycytidine and *DNMT1* siRNA both caused overall hypomethylation, with the latter proving more efficient. (A) Principal component analysis presented in a scatter plot showing how the cells clustered by treatment. This principal component analysis is exclusive of one anomalous Aza sample, and one siRNA sample. Variance for principle component one and two are as labeled on the x and y axis, respectively. **(B)** Hierarchical clustering was performed using 10,000 most variable sites using Euclidean distance as a dissimilarity metric. Blue represents high methylation values, white intermediate and red low values, showing how the methylation patterns differ between treatments, with siRNA showing more low and intermediate values. **(C)** Hierarchical clustering was used to map the 3000 most variable promoters, with Euclidean distance employed as above. Again siRNA appeared to more probes showing low and intermediate methylation as indicated by the dense band of red respective to the siRNA samples, with more variability seen between the two Aza samples. **(D)** Methylation value densities of each sample group across gene bodies, CpG islands and promoters were graphed to demonstrate the different profiles of methylation change across the respective probes, with all the treated and untreated cell lines graphed for comparison. The siRNA samples showed a distribution markedly shifted toward the left (less methylated), particularly for gene bodies. The most noticeable change in Aza-treated samples was a decrease in numbers of highly methylated probes at CpG islands. Aza: 5-aza-2′deoxycytidine.

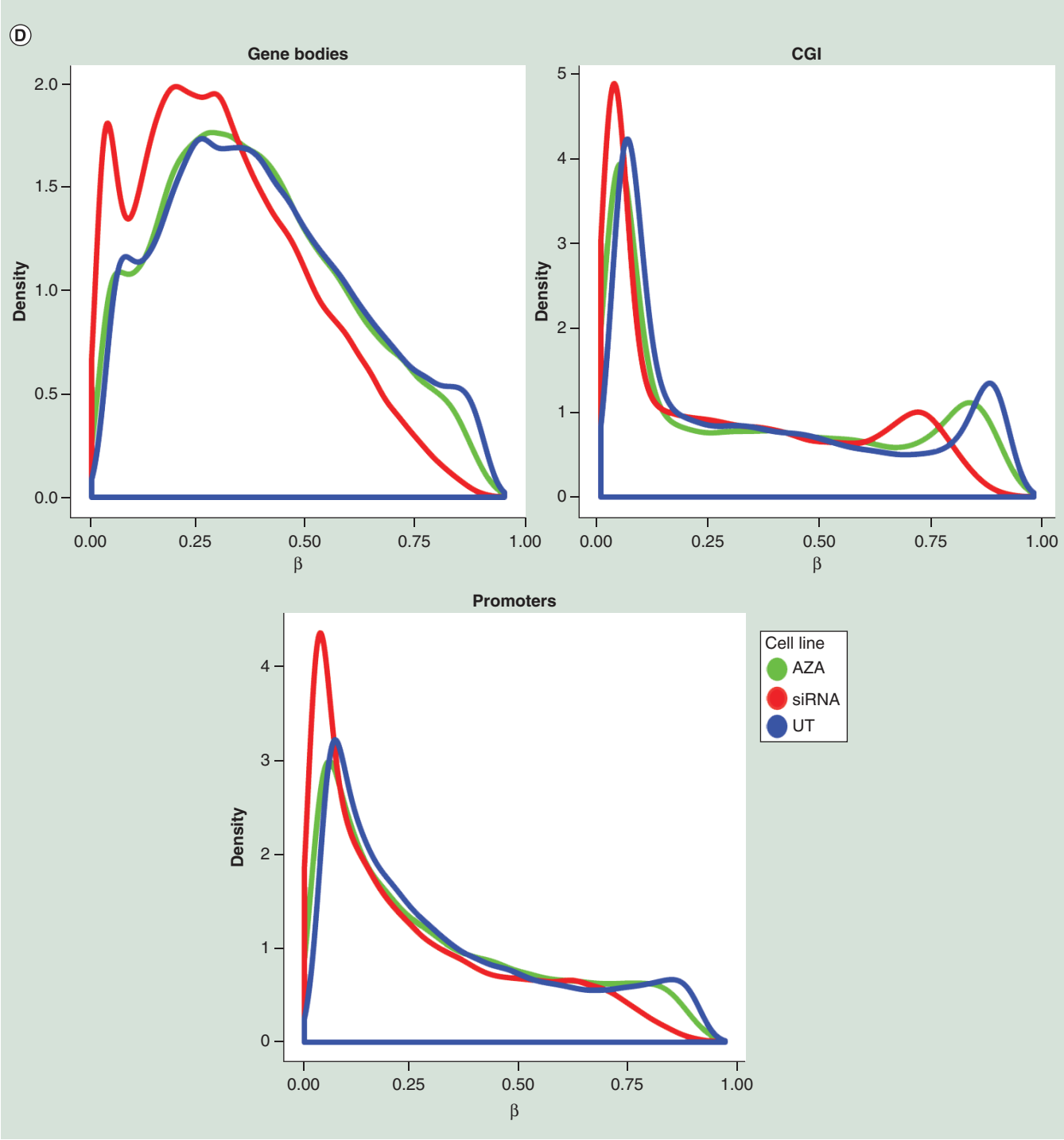


Figure 2. 5-aza-2′deoxycytidine and *DNMT1* siRNA both caused overall hypomethylation, with the latter proving more efficient (cont.). **(A)** Principal component analysis presented in a scatter plot showing how the cells clustered by treatment. This principal component analysis is exclusive of one anomalous Aza sample, and one siRNA sample. Variance for principle component one and two are as labeled on the x and y axis, respectively. **(B)** Hierarchical clustering was performed using 10,000 most variable sites using Euclidean distance as a dissimilarity metric. Blue represents high methylation values, white intermediate and red low values, showing how the methylation patterns differ between treatments, with siRNA showing more low and intermediate values. **(C)** Hierarchical clustering was used to map the 3000 most variable promoters, with Euclidean distance employed as above. Again siRNA appeared to more probes showing low and intermediate methylation as indicated by the dense band of red respective to the siRNA samples, with more variability seen between the two Aza samples. **(D)** Methylation value densities of each sample group across gene bodies, CpG islands and promoters were graphed to demonstrate the different profiles of methylation change across the respective probes, with all the treated and untreated cell lines graphed for comparison. The siRNA samples showed a distribution markedly shifted toward the left (less methylated), particularly for gene bodies. The most noticeable change in Aza-treated samples was a decrease in numbers of highly methylated probes at CpG islands. Aza: 5-aza-2′deoxycytidine.

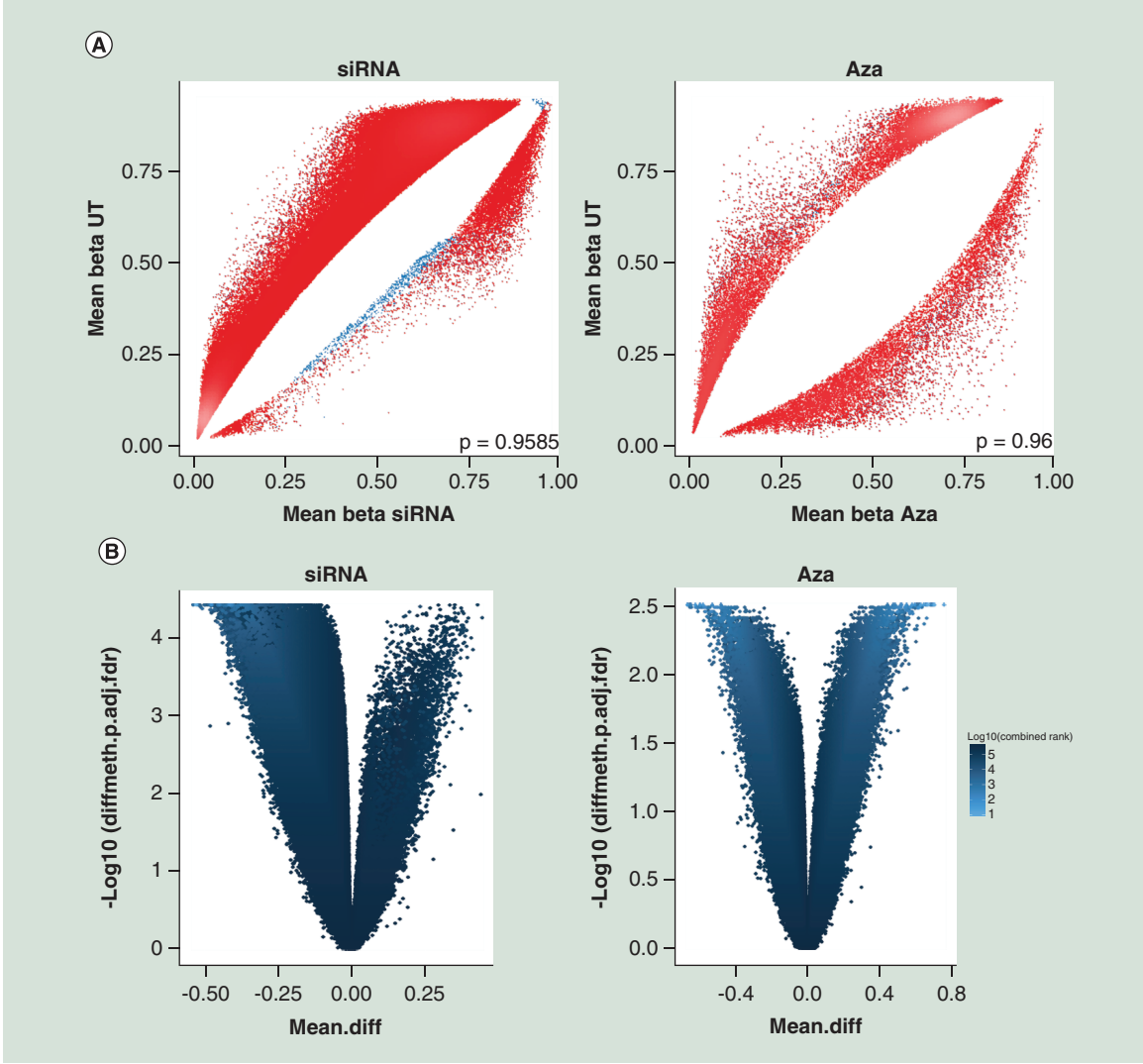


Figure 3. Differential methylation analysis highlights distinct gene classes which are affected by each treatment. **(A)** Scatterplots comparing treated with untreated at the site level, looking only at probes with FDR-adjusted p-value <0.05. Samples treated with siRNA have more hypomethylated probes than Aza-treated samples. **(B)** Volcano plots for promoter differential methylation comparing treatment with siRNA to that of Aza (using FDR-adjusted p-value). A negative value for the difference between means (mean.diff) signifies a loss of methylation, for which siRNA-treated samples have a greater population of probes compared with Aza. Aza-treated samples show a more nearly similar spread of probes which have both negative and positive mean.diff. **(C)** Top gene ontology categories enriched for hypomethylated probes using 'combined rank among the 1000 best ranking regions' for both promoters and gene bodies for both treatments and color coded to show similar functional annotation. The gene classes' enriched intertreatment are distinct from each other, but within treatment groups show substantial commonality. **(D)** Map of the *TLR2* locus from the accession group 'Detection of Chemical Stimulus' enriched in demethylated probes following siRNA treatment, showing the first exon and intron at bottom, and location of CpG islands (green). Above are shown the locations of probes on the array and whether these gained (blue) or lost (red) methylation; a scale bar is shown at left, with maximum and minimum methylation changes as labeled. The chromosomal location and span are shown at top. **(E)** *CHAF1B* belongs to the accession group 'DNA replication-dependent nucleosome assembly' enriched in demethylated probes following Aza treatment: legend as in D.
Aza: 5-aza-2′deoxycytidine; BP: Biological process; ExpCount: Expected count; MF: Molecular function.

located at the promoter. Aza-treated samples also experienced demethylation here (min -22%) but across fewer probes in the region, and also showed probes gaining methylation here. *CHAF1B* is an example from the accession group 'DNA replication-dependent nucleosome assembly' enriched in the Aza-treated cells (Figure 3E), which showed the opposite pattern.

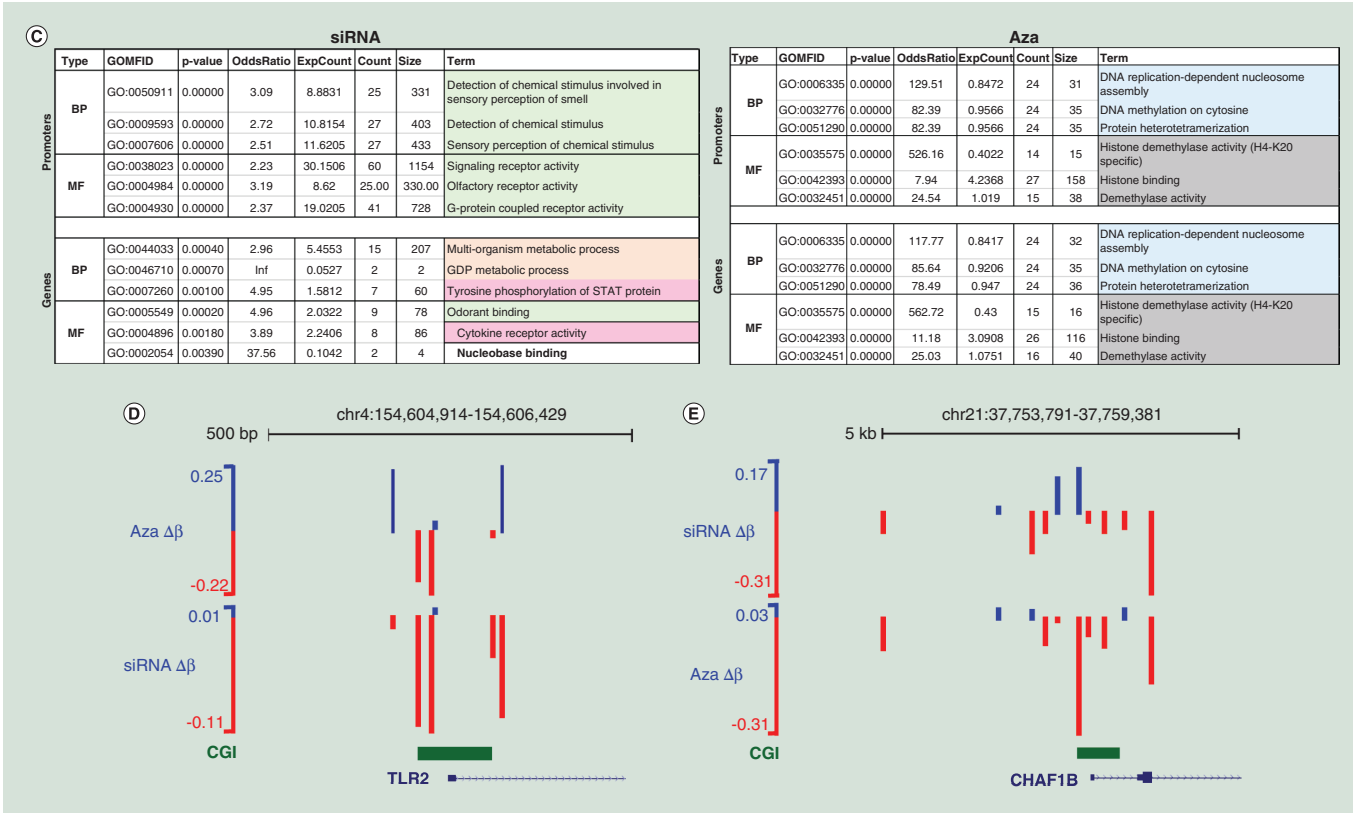
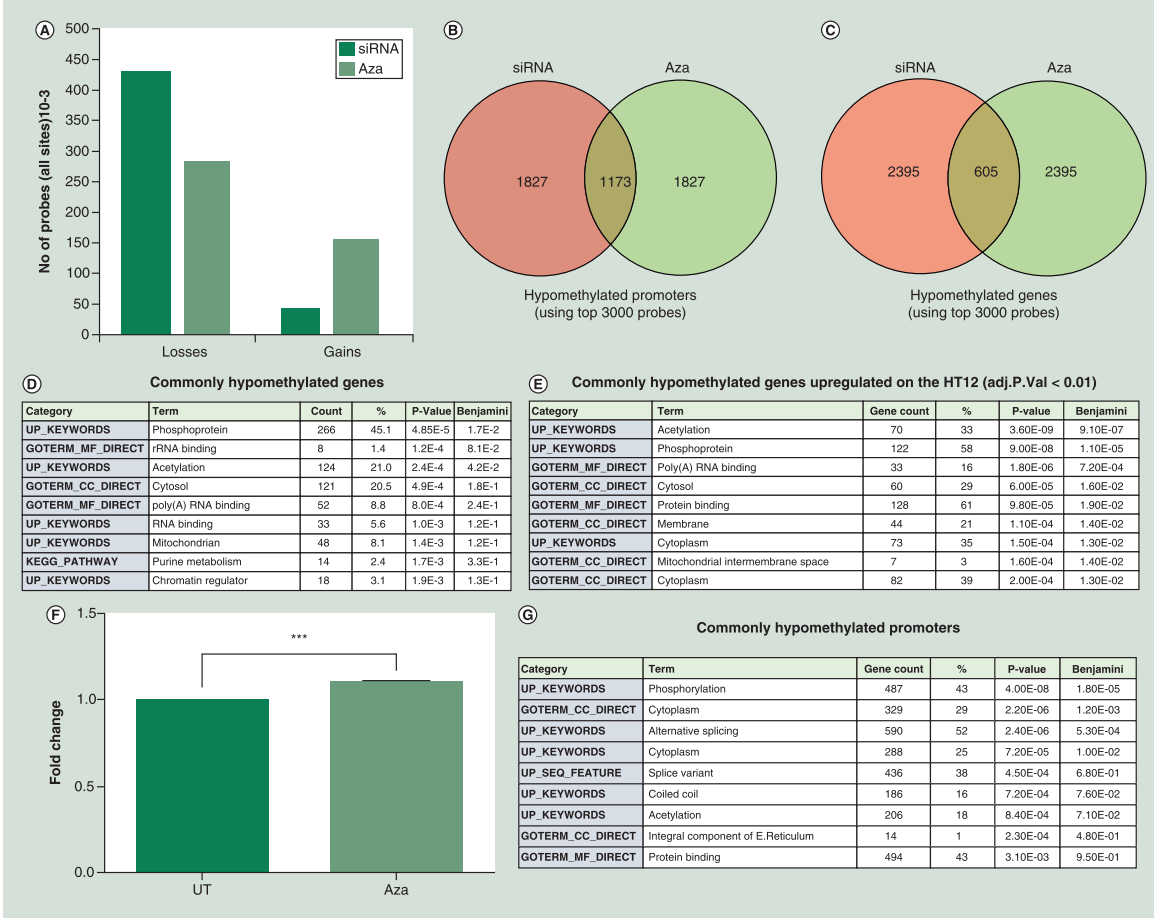


Figure 3. Differential methylation analysis highlights distinct gene classes which are affected by each treatment (cont.). (A) Scatterplots comparing treated with untreated at the site level, looking only at probes with FDR-adjusted p-value <0.05. Samples treated with siRNA have more hypomethylated probes than Aza-treated samples. **(B)** Volcano plots for promoter differential methylation comparing treatment with siRNA to that of Aza (using FDR-adjusted p-value). A negative value for the difference between means (mean.diff) signifies a loss of methylation, for which siRNA-treated samples have a greater population of probes compared with Aza. Aza-treated samples show a more nearly similar spread of probes which have both negative and positive mean.diff. **(C)** Top gene ontology categories enriched for hypomethylated probes using 'combined rank among the 1000 best ranking regions' for both promoters and gene bodies for both treatments and color coded to show similar functional annotation. The gene classes' enriched intertreatment are distinct from each other, but within treatment groups show substantial commonality. **(D)** Map of the *TLR2* locus from the accession group 'Detection of Chemical Stimulus' enriched in demethylated probes following siRNA treatment, showing the first exon and intron at bottom, and location of CpG islands (green). Above are shown the locations of probes on the array and whether these gained (blue) or lost (red) methylation; a scale bar is shown at left, with maximum and minimum methylation changes as labeled. The chromosomal location and span are shown at top. **(E)** *CHAF1B* belongs to the accession group 'DNA replication-dependent nucleosome assembly' enriched in demethylated probes following Aza treatment: legend as in D.
Aza: 5-aza-2′deoxycytidine; BP: Biological process; ExpCount: Expected count; MF: Molecular function.

A common set of genes coding for proteins involved in phosphorylation & acetylation are affected by both treatments

To quantitatively address the numbers of probes losing and gaining methylation we looked at all of the probes with either a positive or negative Δβ per treatment (Figure 4A). Cells treated with *DNMT1*-targeting siRNA had considerably more probes losing methylation than those treated with Aza, but there were more losses than gains with each treatment, as expected. Notably, while some probes belonging to the siRNA-treated samples experienced gains in methylation there were close to three-times as many probes gaining methylation in Aza-treated cells, which was unexpected (Figure 4A).

Due to the comparable nature of both of the treatments with respect to hypomethylation we wanted to examine if Aza treatment did demethylate the same targets as siRNA and so examined the overlap in genes (and promoters) that were demethylated through each treatment. Differentially methylated regions were ranked in RnBeads according to a combination of both statistical significance and effect size, where the latter was composed of the absolute difference in DNA methylation (between UT and treatment) in addition to the relative ratio of mean DNA methylation levels (β) between sample groups, that is, Aza and siRNA. There were 1173 overlapping



D

Category	Term	Count	%	P-Value	Benjamini
UP_KEYWORDS	Phosphoprotein	266	45.1	4.85E-5	1.7E-2
GOTERM_MF_DIRECT	rRNA binding	8	1.4	1.2E-4	8.1E-2
UP_KEYWORDS	Acetylation	124	21.0	2.4E-4	4.2E-2
GOTERM_CC_DIRECT	Cytosol	121	20.5	4.9E-4	1.8E-1
GOTERM_MF_DIRECT	poly(A) RNA binding	52	8.8	8.0E-4	2.4E-1
UP_KEYWORDS	RNA binding	33	5.6	1.0E-3	1.2E-1
UP_KEYWORDS	Mitochondrian	48	8.1	1.4E-3	1.2E-1
KEGG_PATHWAY	Purine metabolism	14	2.4	1.7E-3	3.3E-1
UP_KEYWORDS	Chromatin regulator	18	3.1	1.9E-3	1.3E-1

E

Category	Term	Gene count	%	P-value	Benjamini
UP_KEYWORDS	Acetylation	70	33	3.60E-09	9.10E-07
UP_KEYWORDS	Phosphoprotein	122	58	9.00E-08	1.10E-05
GOTERM_MF_DIRECT	Poly(A) RNA binding	33	16	1.80E-06	7.20E-04
GOTERM_CC_DIRECT	Cytosol	60	29	6.00E-05	1.60E-02
GOTERM_MF_DIRECT	Protein binding	128	61	9.80E-05	1.90E-02
GOTERM_CC_DIRECT	Membrane	44	21	1.10E-04	1.40E-02
UP_KEYWORDS	Cytoplasm	73	35	1.50E-04	1.30E-02
GOTERM_CC_DIRECT	Mitochondrial intermembrane space	7	3	1.60E-04	1.40E-02
GOTERM_CC_DIRECT	Cytoplasm	82	39	2.00E-04	1.30E-02

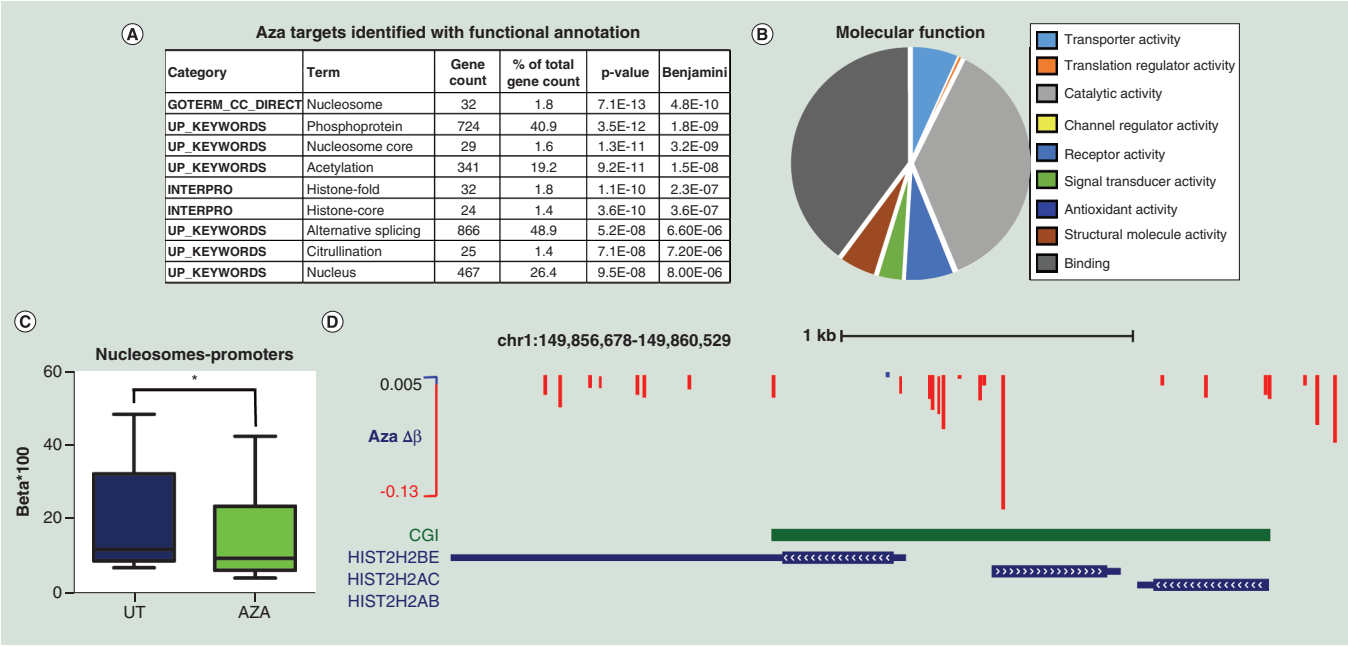
G

Category	Term	Gene count	%	P-value	Benjamini
UP_KEYWORDS	Phosphorylation	487	43	4.00E-08	1.80E-05
GOTERM_CC_DIRECT	Cytoplasm	329	29	2.20E-06	1.20E-03
UP_KEYWORDS	Alternative splicing	590	52	2.40E-06	5.30E-04
UP_KEYWORDS	Cytoplasm	288	25	7.20E-05	1.00E-02
UP_SEQ_FEATURE	Splice variant	436	38	4.50E-04	6.80E-01
UP_KEYWORDS	Coiled coil	186	16	7.20E-04	7.60E-02
UP_KEYWORDS	Acetylation	206	18	8.40E-04	7.10E-02
GOTERM_CC_DIRECT	Integral component of E.Reticulum	14	1	2.30E-04	4.80E-01
GOTERM_MF_DIRECT	Protein binding	494	43	3.10E-03	9.50E-01

Figure 4. Common targets experience demethylation in both treatments. (A) Column graph showing the number of total probes that show an absolute loss and gain in (methylation) β scores overall for each treatment when compared with the untreated hTERT-1604 cells. Probes belonging to the siRNA-treated samples predominantly show demethylation, while there were a substantial number of probes in Aza-treated cells which showed gains in methylation too. **(B)** The overlap between the 3000 top-ranked hypomethylated promoters between sample groups is shown in a Venn diagram. Of this group, 1173 promoters were hypomethylated by both treatments, and an exactly equal number of unique probes (1827) were seen for each treatment. **(C)** Analysis as for B but with genes (gene bodies) showed 605 hypomethylated genes in common between the respective treatments, but greater numbers of uniquely targeted genes. **(D)** DAVID was used to perform functional annotation for the 605 commonly hypomethylated genes between siRNA and Aza treatments. Benjamini, p-value corrected by Benjamini–Hochberg method. **(E)** Commonly hypomethylated genes from D, which also showed upregulation on the HT12 array (a fold change of >1.0 on the HT12 and an adjusted p-value of < 0.01) were analyzed by DAVID, which further highlighted phosphoprotein and acetylation terms. **(F)** Transcriptional fold change for the enriched 'acetylation' targets seen in Figure 4E was calculated using data generated from the HT12 transcription array. Wilcoxon signed rank test indicated a significant upregulation of these genes overall in the Aza samples when compared with untreated ($p < 0.001$). **(G)** Functional annotation was also carried on the 605 commonly hypomethylated promoters seen in (B) using DAVID. Acetylation appears to be the only enriched gene class that appears to show up as hypomethylated for both promoters and gene bodies (D above).
Aza: 5-aza-2′deoxycytidine; DAVID: Database for Annotation, Visualization and Integrated Discovery.

hypomethylated promoters (Figure 4B) while there were only 605 common hypomethylated genes (Figure 4C). To further investigate the enriched common genes (605) in particular, their IDs were extracted and added to the GO pipeline in DAVID, which functionally categorized the enriched genes into classes such as phosphorylation, rRNA binding and acetylation (Figure 4D). Very similar results were obtained when analyzing GO terms in PANTHER (data not shown).

To explore if methylation changes at these enriched hypomethylated genes had any downstream transcriptional effects, we assessed transcription using the HT12 array. With the strongest Benjamini score of 9E10–07, acetylation



Category	Term	Gene count	% of total gene count	p-value	Benjamini
GOTERM_CC_DIRECT	Nucleosome	32	1.8	7.1E-13	4.8E-10
UP_KEYWORDS	Phosphoprotein	724	40.9	3.5E-12	1.8E-09
UP_KEYWORDS	Nucleosome core	29	1.6	1.3E-11	3.2E-09
UP_KEYWORDS	Acetylation	341	19.2	9.2E-11	1.5E-08
INTERPRO	Histone-fold	32	1.8	1.1E-10	2.3E-07
INTERPRO	Histone-core	24	1.4	3.6E-10	3.6E-07
UP_KEYWORDS	Alternative splicing	866	48.9	5.2E-08	6.60E-06
UP_KEYWORDS	Citrullination	25	1.4	7.1E-08	7.20E-06
UP_KEYWORDS	Nucleus	467	26.4	9.5E-08	8.00E-06

Figure 5. Promoters targeted by 5-aza-2′deoxycytidine but not DNMT1 siRNA. (A) DAVID was used to examine the 1827 uniquely hypomethylated Aza promoters from Figure 4B, with results ranked according to Benjamini score. **(B)** The annotation of the uniquely hypomethylated Aza promoters was further explored using PANTHER to determine functional categorization specifically for molecular function where 'binding' and 'catalytic activity' were the most populated gene classes. **(C)** A boxplot was generated for the uniquely hypomethylated Aza-treated promoters and gene bodies for the 'nucleosome' category highlighted in (B). Thirty-two histones were identified as enriched under the 'nucleosome' accession and are significantly hypomethylated when compared with untreated hTERT-1604. **(D)** A schematic of three of the enriched histones belonging to the nucleosome accession from UCSC with Aza Δβ tracks showing methylation loss (red probes) across the region, with the maximally demethylated probe at -13% occurring inside a CpG island which spans parts of all three histones.
Aza: 5-aza-2′deoxycytidine.

was the first common enriched term to appear with nearly 70 genes hypomethylated and upregulated (Figure 4E). The Wilcoxon signed rank test was performed on the transcriptional fold change of the genes treated with Aza that fell under the acetylation category and these were also significantly upregulated compared with the UT ($p < 0.001$) (Figure 4F).

Using the same approach, DAVID functional categorization was performed on the commonly affected 1173 promoters. While some similar enriched gene terms appeared, there were distinct categories enriched in this analysis of common hypomethylated promoters such as alternative splicing/splice variant (Figure 4G). Overall we show here that a relatively small number of common gene classes are affected by both treatments and many of these showed transcriptional effects after being treated with Aza.

Aza uniquely targets histone genes through promoter hypomethylation

We decided to examine the 1827 hypomethylated promoters initially that are unique to Aza treatment (Figure 5A). Through functional categorization of the uniquely hypomethylated promoters in DAVID, gene terms such as nucleosome, phosphorylation and nucleosome core were significantly enriched (Figure 5B). Further exploration of the same 1827 hypomethylated promoters in PANTHER for molecular function highlighted two key categories affected, catalytic activity and binding (Figure 5C). Consistent with the functional annotation analysis, binding is one of the main terms of enrichment that overlaps with phosphoprotein in Figure 5B. Using an analysis workflow in Galaxy we could confirm that methylation levels of the gene targets belonging to the nucleosome family in particular are significantly lower than that of the UT hTERTs (Figure 5D). A cluster containing three histone genes, which belong to the nucleosome category, is shown in Figure 5E with the user defined track indicating Δβ against UT hTERT. *HIST2HAC* showed the maximum decrease in methylation of 12% at its promoter here, but with consistent hypomethylation across this region (Figure 5E).

10.2217/epi-2017-0171

Observed gains of methylation is a phenomenon almost exclusively caused by Aza

To investigate the gains in methylation in treated cells noted in Figure 4A we compared all the genes and promoters which had a positive Δβ for each treatment (Figure 6A & B, respectively) using Venn diagrams. These showed that the majority of gains are observed as an effect of the Aza treatment at both genes and promoters, while only a few are observed in the siRNA-treated cells. Enrichment analysis in RnBeads for Aza was able to assign some functional categorization of the gains in both genes and promoters for biological process and molecular function (Figure 6C) with overlap present at, for example, sensory perception of chemical stimulus and signaling receptor activity. Through matching the 2471 genes that were uniquely hypermethylated upon Aza treatment to information provided on the HT12 transcription array for the same clones, 26 genes were shown to also be downregulated (Figure 6D & E). RnBeads generated a word cloud made up of the 100 best ranking probes that are gaining methylation according to their molecular function (Figure 6F) showing that G-protein-coupled receptors (GPCRs) are most affected, with many being olfactory receptors. One of the other GPCRs is *DRD5*, a dopamine receptor that shows maximum gains of 24% when treated pharmacologically (Figure 6G).

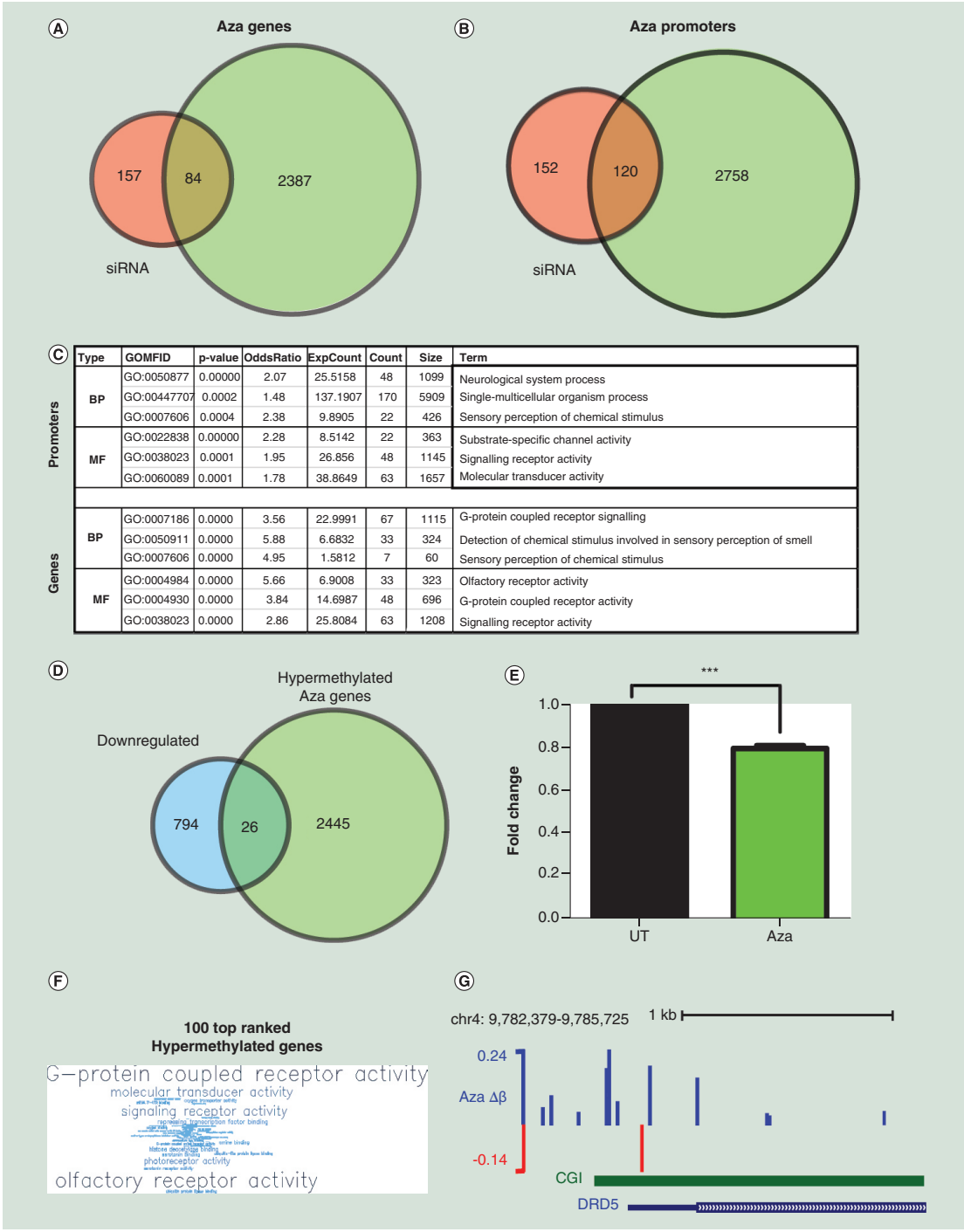
Discussion

Epigenetic therapies are being taken up in the clinic, with the use of pharmacological DNMT inhibitors such as Aza showing some success in the treatment of MDS and the deferment of subsequent progression to AML [33]. However what genes are specifically affected by such epigenetic therapies remains unclear. In the present investigation, target genes susceptible to a change in methylation levels after Aza treatment as compared with siRNA were identified. We found that siRNA was more effective at causing loss of methylation; both treatments targeted genes coding for proteins which become phosphorylated or acetylated, and the genes show transcriptional dysregulation in Aza samples; Aza uniquely targeted histone genes for promoter hypomethylation; and gains of methylation were also seen only in Aza-treated samples.

While both treatments caused overall hypomethylation in the treated cells in both gene bodies and promoters, siRNA proved more efficient at demethylation within the transcription units. These results are in agreement with those of others [34] who observed that pharmacological treatment with Aza caused hypomethylation in the gene bodies of HCT116 cells and subsequently altered the expression of cancer-associated genes. Together this confirms that a drop in methylation in the gene body could be a previously unrecognized but important therapeutic target for DNMT inhibitors. An important role for gene body methylation in contributing to transcription levels has also been indicated by a number of recent studies in other systems [20,35]. This type of methylation has an important function in regulating alternative splicing [36] and this is particularly noteworthy in the context of MDS given the latter's abundance of spliceosome gene mutations [37].

The enriched hypomethylated 'acetylation' genes were significantly upregulated compared with the UT hTERT-1604 cells. This suggests a negative correlation between methylation and expression levels in gene bodies. This novel finding shows that pharmacological treatment with Aza also operates under a different system than traditionally thought by causing transcriptional dysregulation. It has been reported that the methylation present in gene bodies of some leukemic subtypes may have both a functional and clinical impact on disease progression [38] and this may suggest further research is needed in this area.

Methylation at promoter regions is known to silence genes, and often aberrant hypermethylation at these regions is associated with cancer and inappropriate gene silencing [9,39]. Here, however, we saw that Aza uniquely targeted histone transcription units for promoter hypomethylation including some 32 histones including *HIST1H2AJ*, *HIST1H1E* and *HIST2H2AC*, which all fell under the 'acetylation' accession. We observed that the histones were hypomethylated by three separate approaches: using enrichment analysis in RnBeads, DAVID and PANTHER; comparing methylation levels versus UT using Galaxy workflows and statistical analysis in SPSS; by manual curation on UCSC with the use of user-defined interactive tracks against the human genome (hg38). Therefore, it is concluded that Aza can specifically target histones. Interestingly histone deacetylase (HDAC) inhibitors, for example, valproic acid [40] have been used clinically for some time and these can also revert closed chromatin to a more open conformation and in doing so provide an additional mechanism for gene silencing. Some reports have indicated an element of cross-talk between DNA and histone methylation to reduce gene expression, due potentially to the recruitment of HDAC after a 5-methylcytosine-binding protein has attached to the gene promoter [41]. HDAC inhibitors are now used clinically with DNMT1 inhibitors as a combination therapy. This form of combination therapy has shown success with respect to the antineoplastic response in AML both clinically [42,43] and in the laboratory [44].



	Type	GOMFID	p-value	OddsRatio	ExpCount	Count	Size	Term
Promoters	BP	GO:0050877	0.00000	2.07	25.5158	48	1099	Neurological system process
		GO:00447707	0.0002	1.48	137.1907	170	5909	Single-multicellular organism process
		GO:0007606	0.0004	2.38	9.8905	22	426	Sensory perception of chemical stimulus
	MF	GO:0022838	0.00000	2.28	8.5142	22	363	Substrate-specific channel activity
		GO:0038023	0.0001	1.95	26.856	48	1145	Signalling receptor activity
		GO:0060089	0.0001	1.78	38.8649	63	1657	Molecular transducer activity
Genes	BP	GO:0007186	0.0000	3.56	22.9991	67	1115	G-protein coupled receptor signalling
		GO:0050911	0.0000	5.88	6.6832	33	324	Detection of chemical stimulus involved in sensory perception of smell
		GO:0007606	0.0000	4.95	1.5812	7	60	Sensory perception of chemical stimulus
	MF	GO:0004984	0.0000	5.66	6.9008	33	323	Olfactory receptor activity
		GO:0004930	0.0000	3.84	14.6987	48	696	G-protein coupled receptor activity
		GO:0038023	0.0000	2.86	25.8084	63	1208	Signalling receptor activity

Figure 6. Observed gains of methylation is a phenomenon almost exclusively caused by 5-aza-2′deoxycytidine. (A) Common gains in methylation across the probes annotated as in gene bodies (genes) are presented in a Venn diagram. Gains were taken to be anything with a positive difference (not FDR corrected). Aza-treated samples experienced considerably more hypermethylation than the siRNA-treated samples, with only 84 genes commonly hypermethylated between the two treatment groups. **(B)** As for (A) but promoter-associated probes; 120 promoters were commonly gaining methylation between the treatment groups. **(C)** Enrichment table from the RnBeads differential methylation output showing the top enriched hits for Aza (biological process/molecular function) for all of the genes, and promoters that were identified as hypermethylated when compared with the UT. Enriched gene classes 'sensory perception of chemical stimulus' and 'signaling receptor activity', however, were enriched at both probes associated with genes and promoters, and contains many intron-less olfactory receptor genes. **(D)** Venn identifying the hypermethylated Aza probes (p-value < 0.05) which are also downregulated on the HT12 transcription array (fold change <1.0 when compared with UT hTERT-1604). **(E)** For the 26 genes that were both hypermethylated and downregulated as shown in Figure 6D their fold-change collectively was significantly lower in the Aza-treated samples than the UT hTERT-1604 cells. **(F)** Word cloud taken from the enrichment analysis in RnBeads showing the enriched gene classes that were hypermethylated in the Aza samples using the 100 best-ranking selected for molecular function. **(G)** The enriched hypermethylated gene target DRD5 mapped against the human reference assembly and an Aza δβ track showing maximum gains of 24% 5′ of the gene.
Aza: 5-aza-2′deoxycytidine; UT: Untreated.

The hypomethylation seen in the gene bodies here has been argued elsewhere to increase the accessibility of chromatin [44]. The enrichment for hypomethylated histones with Aza treatment may partly explain known clinical interactions [45] that take advantage of combining the DNMT inhibitors and HDAC inhibitors [46] in the treatment of myeloid neoplasms. This form of dual combination treatment approach reduced DNMT activity to create a net demethylation while the HDAC inhibitors are responsible for an overall gain in histone acetylation – this synergistic approach maintains an active and open chromatin state to promote transcriptional activation of, for example, TSGs. The absence of TSGs themselves in the enrichment analysis may reflect the absence of GO terms specifically matched with TSGs, which have a range of different molecular functions and process involvement.

We could not rule out that perhaps the gains in methylation observed in the enriched GPCRs contribute to the relapse in AML in sufferers as key TSGs could be silenced. GPCRs are known to act as oncogenes and tumor suppressors through their regulation of oncogenic signaling networks (reviewed in [47]). To support this point there have been several largescale genome-wide association studies (GWAS) in multiple human tumor samples that have uncovered novel GPCRs altered in cancer [47]. GPCRs may be a suitable therapeutic target as personalized medicine begins to diminish the clinical dependency on nonspecific pharmacological or chemotherapeutic agents. In our study, we were able to identify GPCRs that were hypermethylated including *MAS1* (on oncogene involved in leukemia [48], *USP33* and *PARD3* – candidate and identified TSGs, respectively in lung cancer and in lung squamous cell carcinoma [49,50].

Among the other targets experiencing demethylation, some hypomethylated promoters were unique to Aza treatment and therefore potentially off-target with respect to the reduction in DNMT1 such as phosphoprotein, histone-fold and nucleosome core. These demethylated targets, and the gains in methylation seen at other loci, may be a result of effects of Aza on DNMT3A or DNMT3B, or through a more indirect mechanism such as the reported effects of Aza on the histone modifier G9a [51].

Conclusion

In conclusion, our use of 450k to study the comparative effects of a pharmacological knockdown using Aza, and DNMT1 targeting siRNA in the immortalized hTERT-1604 cell line has confirmed many observations from previous literature with respect to demethylating effects. However, there was also a notable hypermethylating effect, particularly at some GPCRs. Our results suggest Aza is also having an important effect on methylation unrelated to the direct inhibition of DNMT1, thus suggesting further avenues for therapeutic improvements.

Summary points

- 5-aza-2'deoxyctidine (Aza) is a DNA methyltransferase (DNMT) inhibitor used to treat some cancers
- We know relatively little about targets and off-target effects, particularly in normal cells.
- Here we compared drug treatment to siRNA knockdown of the main methyltransferase DNMT1 in non-transformed cells.
- Both caused hypomethylation, particularly at genes for proteins involved in phosphorylation and acetylation, but siRNA was more efficient.
- There were distinct gene classes affected by each, histones being particularly affected by Aza.
- Gains of methylation were uniquely seen in Aza-treated cells, with G-protein coupled receptors common targets.

Authors' contributions

S-J Mackin carried out all of the bioinformatic analysis and assembled figures. KM O'Neill performed the knockdown experiments, extracted the DNA and RNA, and supervised assaying on the Illumina 450k BeadChip Array and the HT12 Transcription Array. CP Walsh designed and supervised the study. S-J Mackin and CP Walsh wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank R Irwin for suggestions and informative discussions as well as the other members of the Walsh Lab and Genomic Medicine Research Group for constructive comments. The authors thank J Bauer at Cambridge Genomic Services for help and advice.

Financial & competing interests disclosure

This work was funded by a grant from the Medical Research Council (MR/J007773/1). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA methylation in cancer and aging. *Cancer Res.* 76(12), 3446–3450 (2016).
2. Weisenberger DJ. Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J. Clin. Invest.* 124(1), 17–23 (2014).
3. National Institute for Health and Clinical Excellence. Azacitidine for the treatment of myelodysplastic syndromes, chronic myelomonocytic leukaemia and acute myeloid leukaemia (TA218). (2011)
4. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat. Rev.Genet.* 14, 204–220 (2013).
5. Klimasauskas S, Kumar S, Roberts RJ, Cheng X. *Hhal* methyltransferase flips its target base out of the DNA helix. *Cell* 76(2), 357–369 (1994).

• **Shows how methyltransferases physically move the cytosine out of the helix, and by extension how 5-aza-2′deoxycytidine (Aza) can lead to 'jamming' of the enzyme.**

6. Bestor TH. Cloning of a mammalian DNA methyltransferase. *Gene* 74(1), 9–12 (1988).
7. Fahy J, Jeltsch A, Arimondo PB. DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies. *Expert Opin. Ther. Pat.* 22(12), 1427–1442 (2012).
8. Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J. Mol. Biol.* 321(4), 591–599 (2002).
9. Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20(1), 85–93 (1980).
10. Kaminskas E, Farrell AT, Wang YC, Sridhara R, Pazdur R. FDA drug approval summary: azacitidine (5-azacytidine, Vidaza) for injectable suspension. *Oncologist* 10(3), 176–182 (2005).
11. ClinicalTrials.gov: NCT02472145 (2017). https://clinicaltrials.gov/ct2/show/NCT02472145
12. Fenaux P, Mufti GJ, Hellstrom-Lindberg E *et al.* Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, Phase III study. *Lancet Oncol.* 10(3), 223–232 (2009).
13. Lubbert M, Suciu S, Hagemeijer A *et al.* Decitabine improves progression-free survival in older high-risk MDS patients with multiple autosomal monosomies: results of a subgroup analysis of the randomized Phase III study 06011 of the EORTC Leukemia Cooperative Group and German MDS Study Group. *Ann. Hematol.* 95(2), 191–199 (2016).
14. Yun S, Vincelette ND, Abraham I, Robertson KD, Fernandez-Zapico ME, Patnaik MM. Targeting epigenetic pathways in acute myeloid leukemia and myelodysplastic syndrome: a systematic review of hypomethylating agents trials. *Clin. Epigenet.* 8, 68–77 (2016).
15. Cheishvili D, Boureau L, Szyf M. DNA demethylation and invasive cancer: implications for therapeutics. *Br. J. Pharmacol.* 172(11), 2705–2715 (2015).
16. Juttermann R, Li E, Jaenisch R. Toxicity of 5-aza-2′-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc. Natl Acad. Sci. USA* 91(25), 11797–11801 (1994).

• **First paper to show how Aza inhibits DNMT1.**

17. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 99(3), 247–257 (1999).
18. Ghoshal K, Datta J, Majumder S *et al.* 5-aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. *Mol. Cell. Biol.* 25(11), 4727–4727 (2005).
19. Lynch CA, Tycko B, Bestor TH, Walsh CP. Reactivation of a silenced H19 gene in human rhabdomyosarcoma by demethylation of DNA but not by histone hyperacetylation. *Mol. Cancer* 1, 2 (2002).
20. Irwin R, Thakur A, O'Neill KM, Walsh CP. 5-Hydroxymethylation marks a class of neuronal gene regulated by intragenic methylcytosine levels. *Genomics* 104(5), 383–392 (2014).
21. Rutledge CE, Thakur A, O'Neill KM *et al.* Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes. *Development* 141(6), 1313–1323 (2014).
22. Santini V, Kantarjian HM, Issa JP. Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. *Ann. Intern. Med.* 134, 573–586 (2011).
23. Herman JG, Umar A, Polyak K *et al.* Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA* 95(12), 6870–6875 (1998).

24. Raj K, Mufti GJ. Azacytidine (Vidaza(R)) in the treatment of myelodysplastic syndromes. *Ther. Clin. Risk Manag.* 2(4), 377–388 (2006).
25. Pandiyan K, You JS, Yang X *et al.* Functional DNA demethylation is accompanied by chromatin accessibility. *Nucleic Acids Res.* 41(7), 3973–3985 (2013).
26. Traina F, Visconte V, Elson P *et al.* Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. *Leukemia* 28(1), 78–87 (2014).

• **Study looking at the effects of mutations in components of the methylation machinery in predicting response to Aza treatment.**

27. Bibikova M, Barnes B, Tsan C *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* 98(4), 288–295 (2011).
28. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* 11(11), 1138–1140 (2014).
29. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4(1), 44–57 (2009).
30. Thomas PD, Campbell MJ, Kejariwal A *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13(9), 2129–2141 (2003).
31. Afgan E, Baker D, van den Beek M *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44(W1), W3–W10 (2016).
32. Kent WJ, Sugnet CW, Furey TS *et al.* The human genome browser at UCSC. *Genome Res.* 12(6), 996–1006 (2002).
33. Diesch J, Zwick A, Garz A, Palau A, Buschbeck M, Götz KS. A clinical-molecular update on azanucleoside-based therapy for the treatment of hematologic cancers. *Clin. Epigenetics* 8, 71 (2016).
34. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26(4), 577–590 (2014).

•• **Study showing Aza decreases gene body methylation concomitant with a decrease in gene transcription.**

35. Smallwood S, Kelsey G. *De novo* DNA methylation: a germ cell perspective. *Trends Genet.* 28(1), 33–42 (2012).
36. Shukla S, Kavak E, Gregory M *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479(7371), 74–79 (2011).
37. Yoshida K, Sanada M, Shiraishi Y *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478(7367), 64–69 (2011).
38. Kulis M, Heath S, Bibikova M *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* 44(11), 1236–1242 (2012).

•• **Shows that the two subtypes of chronic lymphocytic leukemia are distinguished by differences in gene body methylation.**

39. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 3(6), 415–428 (2002).
40. Kuendgen A, Gattermann N. Valproic acid for the treatment of myeloid malignancies. *Cancer* 110(5), 943–954 (2007).
41. Jones PL, Veenstra GJ, Wade PA *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* 19(2), 187–191 (1998).
42. Garcia-Manero G, Kantarjian HM, Sanchez-Gonzalez B *et al.* Phase I/II study of the combination of 5-aza-2′-deoxycytidine with valproic acid in patients with leukemia. *Blood* 108(10), 3271–3279 (2006).
43. Fredly H, Gjertsen BT, Bruserud O. Histone deacetylase inhibition in the treatment of acute myeloid leukemia: the effects of valproic acid on leukemic cells, and the clinical and experimental evidence for combining valproic acid with other antileukemic agents. *Clin. Epigenetics* 5(1), 12 (2013).
44. Lemaire M, Momparler LF, Farinha NJ, Bernstein M, Momparler RL. Enhancement of antineoplastic action of 5-aza-2′-deoxycytidine by phenylbutyrate on L1210 leukemic cells. *Leuk. Lymphoma* 45(1), 147–154 (2004).
45. Gore SD, Baylin S, Sugar E *et al.* Combined DNA methyltransferase and histone deacetylase inhibition in the treatment of myeloid neoplasms. *Cancer Res.* 66(12), 6361–6369 (2006).
46. Plass C, Oakes C, Blum W, Marcucci G. Epigenetics in acute myeloid leukemia. *Semin. Oncol.* 35(4), 378–387 (2008).
47. Feigin ME. Harnessing the genome for characterization of G-protein coupled receptors in cancer pathogenesis. *FEBS J.* 280(19), 4729–4738 (2013).
48. Rabin M, Birnbaum D, Young D, Birchmeier C, Wigler M, Ruddle FH. Human ros1 and mas1 oncogenes located in regions of chromosome 6 associated with tumor-specific rearrangements. *Oncogene Res.* 1(2), 169–178 (1987).
49. Wen P, Kong R, Liu J *et al.* USP33, a new player in lung cancer, mediates Slit–Robo signaling. *Protein Cell* 5(9), 704–713 (2014).
50. Wang L, Zhang H, Hasim A *et al.* Partition-defective 3 (PARD3) regulates proliferation, apoptosis, migration, and invasion in esophageal squamous cell carcinoma cells. *Med. Sci. Monit.* 23, 2382–2390 (2017).

51. Wozniak RJ, Klimecki WT, Lau SS, Feinstein Y, Futscher BW. 5-Aza-2′-deoxycytidine-mediated reductions in G9A histone methyltransferase and histone H3 K9 di-methylation levels are linked to tumor suppressor gene reactivation. *Oncogene* 26(1), 77–90 (2007).

•• **Paper describing how treatment with Aza leads to indirect decreases in the H3K9 methyltransferase G9A levels and subsequent chromatin changes.**

PAPER-IV

Imprint Stability and Plasticity during Development

Sarah-Jayne Mackin, Avinash Thakur, Colum P. Walsh.

Reproduction. 2018 Aug; 156 (2):R43-R55.

doi: 10.1530/REP-18-0051

The main aims of this paper were to:

1. Summarise what is known about the importance of imprints for normal mammalian growth and body size.
2. Recap upon the main types of control that acts upon an imprinted locus.
3. Discuss the ability of imprints to be reprogrammed somatically, and therefore after initial establishment.
4. Examine the conservation of mammalian imprints in human, mouse and livestock species.
5. Cast an eye forward to the future of research into the field of genomic imprinting.

CONTRIBUTION

I wrote most of this manuscript and performed an intensive literatue search for the technologies presented in Table 1. I also completed the referencing. Avinash Thakur created Figure 1 and Figure 2.

Summary of the major findings;

This review endevours to cover the recent advancements in genomic imprinting from the main technologies use to identify imprints in livestock and human species, to a summary of the locus control at imprints and reprogramming abilities at imprinted loci.

REPRODUCTION
REVIEW

Imprint stability and plasticity during development

Sarah-Jayne Mackin, Avinash Thakur*, and Colum P Walsh

Genomic Medicine Research Group, School of Biomedical Sciences, Ulster University, Northern Ireland, UK

Correspondence should be addressed to C P Walsh; Email: cp.walsh@ulster.ac.uk

*(A Thakur is currently at Terry Fox Laboratory, BC Cancer Research Centre, 675 West 10th Avenue, Room 13-112, Vancouver, BC V5Z 1L3, Canada)

Abstract

There have been a number of recent insights in the area of genomic imprinting, the phenomenon whereby one of two autosomal alleles is selected for expression based on the parent of origin. This is due in part to a proliferation of new techniques for interrogating the genome that are leading researchers working on organisms other than mouse and human, where imprinting has been most studied, to become interested in looking for potential imprinting effects. Here, we recap what is known about the importance of imprints for growth and body size, as well as the main types of locus control. Interestingly, work from a number of labs has now shown that maintenance of the imprint post implantation appears to be a more crucial step than previously appreciated. We ask whether imprints can be reprogrammed somatically, how many loci there are and how conserved imprinted regions are in other species. Finally, we survey some of the methods available for examining DNA methylation genome-wide and look to the future of this burgeoning field.

Reproduction (2018) **156** R43–R55

Introduction

Genomic imprinting is a classic epigenetic phenomenon, and after X-inactivation, one of the best understood. In its simplest form, imprinting refers to genes showing mono-allelic expression depending on their parent of origin (Bartolomei & Ferguson-Smith 2011, Barlow & Bartolomei 2014). There is however considerable variation seen in where in the body and when the imprinted expression is seen. Imprints need to be set up when the alleles are separated during germ line development and then the imprints need to be interpreted by the transcriptional machinery in the target tissues. Recent work has advanced our understanding of (1) the processes behind establishment of imprints as well as (2) how the imprinting mark is maintained and interpreted. For the former, there is an excellent review by Kelsey (Hanna & Kelsey 2014). Here, we wish to focus on the latter process of maintenance, as well as ask what we have learned from species other than mouse, and touch on the technical difficulties, which arise when working in organisms other than mouse and human, where the most resources are available.

Importance of known imprints for body mass and feeding

Effects on growth and body size

One of the most enduring explanations for why imprinting may have arisen is the parental conflict hypothesis (Haig & Graham 1991, Moore & Haig 1991). Many imprinting phenomena involve effects on growth, nutrition or metabolic balance, and this theory was partly based on observations from uniparental conceptuses where the paternal contribution seemed particularly important for the embryo, while the maternal genome was important for placental development (Barton *et al.* 1984). This turned out to be partly due to the paternal expression of the major growth factor insulin-like growth factor-2 (IGF2), and the maternal expression of an antagonistic binding partner insulin-like growth factor-2 receptor (IGF2R), one of the first observations leading to the idea of parental conflict in imprinting. Several imprinted genes were also shown to be dysregulated in the placentas of these mouse conceptuses (Walsh *et al.* 1994), including genes specifically imprinted only in placenta such as *Mash2* (Guillemot *et al.* 1995). Indeed, the placenta is generally considered to have a higher number of imprinted genes than in other tissues (Court *et al.* 2014); though in the mouse and human, many of these are maternally expressed. An alternative to the parental conflict hypothesis, called coadaptation theory, also stresses the importance of interactions between offspring and parents for nutrient provision and acquisition (Keverne & Curley 2008).

Much of the support for these theories has been gleaned in mouse systems. Interestingly however, a single nucleotide substitution in a non-coding region of *IGF2* underlies a major QTL in pigs that affects muscle growth, heart size and fat deposition (Jeon *et al.* 1999,

ISSN 1470–1626 (paper) 1741–7899 (online)
https://doi.org/10.1530/REP-18-0051
Online version via www.reproduction-online.org

Van Laere *et al.* 2003). The mutation disrupts the binding of a recently evolved nuclear protein ZBED6 that appears to be acting as a repressor of *IGF2* transcription, not only in pigs but potentially in all placental mammals, where it is highly conserved (Markljung *et al.* 2009, Younis *et al.* 2018). A more general defect called large offspring syndrome (LOS) in livestock (namely sheep and cattle) also involves perturbed imprinted gene expression, thought to be due to preimplantation disturbances. LOS is associated with an increase in gestational length, 50–80% increase in birth weight, hypoglycaemia and an enlarged tongue (Chen *et al.* 2013). The phenotypic characteristics of LOS are reminiscent of those exhibited by Beckwith–Wiedemann syndrome in humans, which can be driven by defects at the *IGF2* imprint control region (ICR) or the nearby *KvDMR* (Weksberg *et al.* 2010). Rivera and colleagues found however that LOS was not characterised by loss of imprinting at any one locus, but seemed instead to be accompanied by misregulation of multiple imprinted genes, with greater dysregulation seen in the largest offspring (Chen *et al.* 2015). The requirement for delivery of large offspring by caesarean section, the frequent post-natal mortality, and other complications, mean that LOS restricts the full use of modern reproductive strategies in ruminants.

The callipyge muscle hypertrophy phenotype observed in sheep is also caused by a SNP, in this case located intergenically between known imprinted genes, the paternally expressed *DLK1* and maternally expressed *MEG3* long non-coding RNA (lncRNA) within the telomeric region of chromosome 18. This region is known to share a high degree of homology with other mammals including cattle and humans (Freking *et al.* 2002*a*,*b*). Callipyge syndrome can cause adverse changes to normal muscle development, body shape and even changes to meat quality (Freking *et al.* 2002*a*) and the composition of their carcass (Bidwell *et al.* 2014). Interestingly, the expression of this unique phenotype is the only distinctive example of polar overdominance in mammals (Freking *et al.* 2002*a*). Heterozygous callipyge sheep have a normal maternal allele while carrying the callipyge SNP on the paternal allele (Bidwell *et al.* 2014).

Neurological effects involving feeding and weight gain

Another general feature of imprinted genes is that many are transcribed primarily in the neurons and in some cases only show imprinted expression there. While initially this posed difficulties for the parental conflict hypothesis, which was originally proposed on the basis of imprinting affecting placentation and embryonic growth, with the discovery that some of the genes controlled feeding behaviour or weight gain later in life, these results could be reconciled with the hypothesis. Thus, Prader–Willi syndrome in humans (OMIM176270), where transcription of the neuronal genes at the *SNRPN* cluster are affected, involves excessive eating (hyperphagia), while the *Peg3* gene in mouse is crucial for feeding behaviour (Curley *et al.* 2005). Likewise, loss of expression of the *Gnas* and *Gnasxl* genes in mice can give hypo- and hyperphagia respectively, while a number of other imprinted genes also have effects on fat/body mass ratios (Peters 2014).

Furthermore, the aforementioned LOS or large calf syndrome can often cause a reluctance to suckle in young cattle. This may be due to the dysregulation of a number of imprinted genes reported in these animals (Zhang *et al.* 2004, Chen *et al.* 2015). However apart from these few studies, little work appears to have been done to date investigating imprinted genes and feeding behaviours outside of the mouse and human systems.

Imprinting disorders in hybrid species: the case of the hinny

A classic example of parent-of-origin effects in livestock is the horse–donkey cross: when the horse is the sire, the resultant offspring is a mule; if the donkey is the sire, then a hinny results, and these are quite distinguishable animals even though their genomic complements are the same. Initially, it was argued that this may be due to differences in uterine environment and the placental–uterine exchange. Early investigation identified that serum concentrations of the placental hormone equine chorionic gonadotrophin were considerably higher in mule than in hinny pregnancies, suggestive of paternal genome-specific expression (Allen 1969). Recent use of RNA sequencing of trophoblast tissue from embryos derived from each reciprocal cross of horse and donkey (Wang *et al.* 2013) identified 15 'core' imprinted genes that are conserved in equids, mouse and human, including genes such as *SNRPN, H19* and *PEG3*. Interestingly, paternal expression seems favoured, as 10/15 core imprints – and all of the larger group of 78 genes identified in that study – showed paternal bias, but with many only imprinted in placenta. Additionally, some of the genes in the second, larger group showed incomplete or variable silencing. The authors speculate that this variability may reflect the flexibility of the structure and function of the placenta, which varies widely between mammals. Interestingly, XIST does not seem to be one of the imprinted genes, and X-inactivation in the placenta is random in both the horse parent and the hybrids (Wang *et al.* 2012).

Establishment of parental-specific methylation at imprints

It was recognised already from early studies that imprints would most likely be established in the germ line, where each allele of the diploid pair is separate and therefore can be separately marked. While studies on germ cell development were long hampered by the

difficulty of access to the early stages of development of this migratory cell population, and the small numbers of cells involved, recent advances in cell labelling and separation, coupled with the refinement of sequencing techniques and single-cell approaches (see below) have meant that we now have a much deeper understanding of the process of imprint establishment. This area has been dealt with very expertly by others in a recent review (Hanna & Kelsey 2014). We will concentrate here on insights gained in some studies published since then, confining ourselves first to a brief recap of the varied imprinting mechanisms.

The main types of imprinted loci and their structures

Current studies suggest that there are three main types of imprinted locus, controlled by different mechanisms. This has implications for where the imprint controllers are located and what they look like, as well as how the locus might respond to different perturbations.

Insulator

The archetype here is the cluster of genes around the *H19* locus (Bartolomei & Ferguson-Smith 2011): these are controlled by an intergenically located ICR that is not a CpG island (CGI) (Bird 1987) but is relatively CG rich, making it sensitive to DNA methylation. The ICR contains several repeats of a sequence that binds the insulator protein CTCF. Methylation blocks CTCF binding on the paternal allele at the locus (Bell & Felsenfeld 2000, Hark *et al.* 2000), preventing *H19* transcription but facilitating expression of *Igf2* and, in mouse, the *Insulin 2* gene expressed in the yolk sac. The mark on the ICR that is set up in the germ cells is called a primary differentially methylated region (1° DMR) or gametic DMR (gDMR). Methylation that occurs post implantation, for example, on the *H19* promoter after it has already been silenced, is known as a secondary (2°) DMR or somatic DMR (sDMR). There are a number of other imprinted loci with CTCF-binding sites, in some cases lacking CGI, where this protein is thought to play an important role (Prickett *et al.* 2013).

Long non-coding RNA promoter

Here the index locus is *Igf2r* that, unlike *H19*, is an orphan imprinted gene. The DMR at the *Igf2r* promoter was, against expectation, the secondary DMR, and the primary DMR and functional ICR were found to be located at an intragenic island. This turned out to control transcription of an antisense lncRNA, whose transcription blocked the sense transcript in cis (Sleutels *et al.* 2002). This type of arrangement is also seen at the *Kcnq1*, *Grb10* and *Dlk1* loci among others (Barlow & Bartolomei 2014) and as more lncRNA are uncovered, this is becoming the largest group.

Promoter

For some imprinted loci such as *Snrpn*, *Plagl1* and *Grb10*, transcription appears to be directly regulated by methylation of the promoter. Here, a promoter CGI is methylated to block transcription rather than any indirect mechanism. Although there is no antisense transcript, the *Snrpn* promoter does drive the production of a long RNA called *Snurf/Snrpn* that extends far downstream and contains multiple small RNA species that are processed from it (Buiting 2010), so it could be argued that it falls into the lncRNA category. However, for other genes in this group such as *Grb10*, no associated lncRNA have so far been found.

Recent insights into imprint establishment

Seminal work from the Bestor lab (Ooi *et al.* 2007) showed that the methyltransferase cofactor DNMT3L, which is only required in germline, bound to CGI but was blocked by methylation of the fourth lysine on the histone 3 tail (H3K4me3), and further that imprinting was disrupted in the knockout (KO) mice. This link seemed strengthened by the finding that KDM1B, a histone demethylase, was required to establish DNA methylation at imprints (Ciccone *et al.* 2009). However, the idea that DNMT3L might be specific to imprints was weakened by studies charting the establishment of methylation in oocytes, which showed that not only did many thousands of non-imprinted CGI became hypermethylated during oocyte maturation in a DNMT3L-dependent fashion, but DNMT3L was also required for methylation of many other non-CGI sequences (Kato *et al.* 2007, Smallwood *et al.* 2011, Kobayashi *et al.* 2012), suggesting that DNMT3L was in fact more of a general cofactor for the *de novo* methyltransferases, required primarily in the germline. What appeared instead to distinguish imprinted loci from other genes was that they could *maintain* their DMR in the face of the two main waves of remodelling in the early embryo, namely active and passive demethylation in the first few cell divisions and then the widespread *de novo* methylation seen post implantation. While few imprints are established in the male germ line, indications are that here too it is the maintenance of these DMR in the face of remodelling that separates them from other loci (Wang *et al.* 2014).

In effect, this marks a real shift in thinking with regards to imprinting, moving the emphasis away from the mechanisms that might target regions to become imprinted and towards identifying factors that are responsible for maintaining methylation differences at specific loci in the face of extensive epigenomic remodelling. Since this is a crucial new insight, we wish to expand on these findings below and begin by looking in more detail at the protein factors that may be involved prior to implantation, and then look at events post implantation.

Maintenance of imprints pre implantation

Methylation marks are applied in a sex-specific manner during gametogenesis at the gDMRs and several proteins play important roles at various stages in this process. In particular, *Pgc7* (also referred to as *Stella*) is a critical component required in early maternal development and encodes a protein that has a SAP-like domain and a splicing factor-like domain (Aravind & Koonin 2000). It has an important role in protecting the methylation status of imprinted genes by limiting demethylation in early embryogenesis (Nakamura *et al.* 2007) and as such is crucial for normal development. PGC7 has been experimentally shown to be actively expressed in primordial germ cells (PGCs) in mouse from E7.25 to E15.5 (Sato *et al.* 2002) but more importantly in the immature ovaries of the neonate females (Payer *et al.* 2003). Through depletion of the protein in the oocytes, Payer *et al.* (2003) found there was a reduction in the number of blastocysts, successful implantations and a reduction in the number of viable offspring. Nakamura *et al.* (2007) initially showed that PGC7 was required to maintain methylation on most imprinted loci in the preimplantation embryo, and later went on to show that PGC7 bound to H3K9me2 at both paternally and maternally marked loci protects against active demethylation by TET3, a finding confirmed by others (Nakamura *et al.* 2012, Szabó & Pfeifer 2012).

Another factor necessary for DNA methylation maintenance at ICRs in both mouse and human is ZFP57 (Kruppel-like zinc finger protein). Removal of the protein from the mouse zygote by use of both maternal and zygotic deletions caused embryonic death mid-gestation with a loss of all maternally methylated imprints (Li *et al.* 2008). In humans too, mutations in *ZFP57* causes hypomethylation at multiple imprinted loci (Mackay *et al.* 2008). ZFP57 binds to a target hexanucleotide found at imprinted gDMRs, but only when methylated, and recruits DNMT1 and its cofactor UHRF1 to maintain methylation on the marked alleles (Quenneville *et al.* 2011). Additionally, through KAP1, it recruits SETDB1 that modifies the chromatin by adding H3K9me3. Indeed, loss of KAP1 also causes a failure to maintain imprints in the preimplantation embryo (Messerschmidt *et al.* 2012). Following this protective step in early development, DNA methylation is thought to be maintained along with this chromatin mark throughout the offspring's life at the respective ICRs (Proudhon *et al.* 2012).

Post-implantation changes at transient and stable imprints

The small number of known imprinted loci characteristically show (1) methylation differences between gametes (gDMRs), (2) dependence on DNMT3L (Kato *et al.* 2007) and (3) an ability to maintain differential methylation during preimplantation development (Hanna & Kelsey 2014). However, the genome-wide studies of the methylation landscape in germ cells mentioned earlier found that a relatively large number of loci (>1000) were methylated in a DNMT3L-dependent fashion in oocytes, but not sperm, and that many of them maintained high levels of methylation on the maternal allele until at least the blastocyst stage (Borgel *et al.* 2010, Smallwood *et al.* 2011, Shirane *et al.* 2013). Initial work by Michael Weber and colleagues (Borgel *et al.* 2010), suggested that not only imprinted loci, but also genes that were specifically expressed in the germline showed these features. Work from our own lab confirmed this (Rutledge *et al.* 2014) and further uncovered a class of genes expressed in brain showing these characteristics, suggesting three classes of genes with DNMT3L-programmed gDMRs: imprinted, germline and brain-specific genes (Fig. 1A). However, in stark contrast to the canonical imprinted loci, most of the brain and germline genes examined lost their differential methylation in adult tissues tested due to *gains* in methylation on the paternal, unmethylated allele. However, it was possible that among these latter two classes there were some true imprinted genes, which would show allele-specific expression in a tissue- or stage-specific fashion.

To identify such genes, Ferguson-Smith and colleagues used ZFP57 binding as an extra criterion in addition to the three indicated above (Strogantsev *et al.* 2015). Using this approach, they could confirm recently identified imprints such as *Cdh15* and *Gpr1*, and also uncovered evidence for imprinting at the *Fkbp6* gene. *Fkbp6* showed methylation in oocytes but not sperm in their study (Strogantsev *et al.* 2015), and maternal methylation was maintained in placenta at e16.5. In keeping with this, there was also predominant expression from the paternal allele as assayed using a transcribed SNP in interspecific mouse crosses, and this allele also carried H3K4me3 marks. Notably, in brain, both transcription and H3K4me3 continued to be associated with the paternal allele, even though the DNA methylation was assayed at 80%, rather than the expected 50%. This suggests that (1) either a subset of cells in brain continues to maintain the DNA methylation difference (in keeping with the low levels of transcription in this tissue); (2) some of the methylation being assayed may in fact be 5-hydroxymethylation (5hmC- see below), which is more prominent in brain and may cloud the picture, or most intriguingly (3) the DNA methylation difference is not as important as the histone marks. In addition to this complexity, *Fkbp6* may show differences in imprinting due to genetic variation, since sperm methylation varies between mouse strains (Rutledge *et al.* 2014), and there have been reports of polymorphic imprinting in humans (Hanna *et al.* 2016). ZFP57 is also capable of mediating such strain-specific effects (Strogantsev *et al.* 2015).

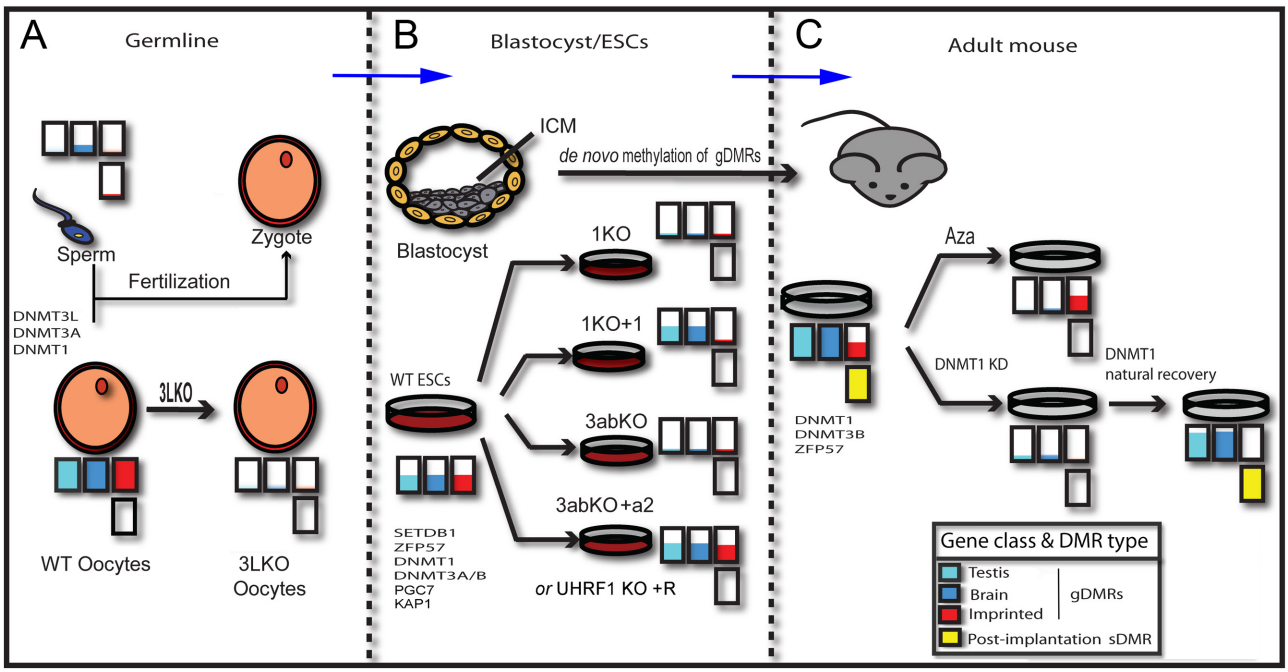


Figure 1 Some experimentally observed dependencies for imprint and imprint-like genes in mouse. Three classes of genes have gametic Differentially Methylated Regions (gDMR): testis, brain and imprinted genes (see key at bottom right: filled bar, 100% methylation). Genes that acquire methylation post implantation (somatic or sDMR) are also represented. (A) Germline: All three classes of gDMRs lose methylation in oocytes lacking DNMT3L. (B) Blastocyst/ESC: The gDMRs retain their methylation aided by DNMT1 and by DNMT3A/B, with other possible contributors indicated. Loss of methylation on imprints in *Dnmt1* knockout (1KO) cells cannot be rescued with *Dnmt1* cDNA (1Ko+1). In contrast, imprints in *Dnmt3a/b KO* (3abKO) can recover following rescue with *Dnmt3a2* (3abKO+a2). (C) Adult: Unlike true imprints, the testis and brain gDMRs cannot maintain the unmethylated allele in the face of the wave of *de novo* methylation after implantation. Likewise, they are more sensitive to inhibitors such as 5′-aza-2′deoxycytidine (Aza).

Using one of the other three criteria above, namely dependence on DNMT3L, the Bourc'his lab also identified novel imprinted gDMRs including *Cdh15, AK008011, Zfp777* and *Zfp787* (Proudhon *et al.* 2012). The *Cdh15* gDMR was found to control transcription of an allele-specific RNA in the hypothalamus, but there was no evidence for imprinted expression at the other three loci. All four gDMRs again lost their parental mark through *gains* in methylation in most adult tissues, highlighting the crucial role of protection from *de novo* methylation for imprint stability. They proposed that such genes, which would include *Fkbp6*, *Cdh15*, *Gpr1* and others, should be referred to as *transient imprints*. The subtlety of regulation of these loci led to the question of how functionally important this was. In a subsequent paper, they showed however that for the *Gpr1* locus, DNA methylation did indeed play an important role in fine-tuning transcription levels (Duffie *et al.* 2014). At this locus, interestingly, the histone marks seem to play a secondary role.

In contrast to these transient imprints, the canonical imprinted loci maintain methylation differences at the gDMR even in tissues that do not express the associated gene (Woodfine *et al.* 2011, Court *et al.* 2014, Wang *et al.* 2014), highlighting the existence of mechanisms to (1) prevent demethylation of the methylated (usually maternal) allele and (2) protect the unmethylated (usually paternal) allele from gaining methylation. Studies comparing the sizes of DMR concluded that a certain degree of shrinkage occurs from the gametic state as the embryo matures (Court *et al.* 2014) but that the boundaries do not markedly shift upstream or downstream, consistent with a mechanism centred on the sequence determinants in the gDMRs.

Possible role of other cytosine modifications at imprinted loci

The dynamic changes in DNA methylation seen during early development are now known to be driven both by passive dilution through replication in the absence of maintenance activity, as well as active demethylation via the action of the three ten-eleven translocation (Tet) enzymes (Guo *et al.* 2014, Wang *et al.* 2014). The Tets have been shown to catalyse the oxidation of normal 5-methylcytosine (5mC), first to 5-hydroxymethylation (5hmC), then to 5-formlycystosine (5fC) and 5-carboxycytosine (5caC), at which stage the modified nucleotide is excised from DNA by the base excision repair enzyme thymidine DNA glycosylase (TDG) or possibly removed through an as-yet uncharacterised

decarboxylase activity (reviewed in Hahn *et al.* 2014). Active demethylation explains the rapid loss of 5mC in the pronuclei of the fertilised egg in the absence of replication, particularly notable for the male pronucleus, and lies behind much of the swift demethylation seen at some other time points such as germ cell specification and preimplantation development (reviewed in Hill *et al.* 2014). Blocking the action of the Tets thus prevents the demethylation at ICR in the germ cells required to reset the imprints, and it is only through the protective action of PGC7 (Nakamura *et al.* 2012, Szabó & Pfeifer 2012) that imprints can be maintained pre implantation when Tet activity drives rapid demethylation at other methylated regions (Guo *et al.* 2014, Wang *et al.* 2014).

In addition to being an intermediate in the active demethylation pathway, however, there is some evidence that 5hmC may also act as a stable epigenetic mark in itself (Hahn *et al.* 2014). Support for this derives from the observations that 5hmC levels in brain are too high to be consistent with being an intermediate in active demethylation in the non-replicating neurons and that proteins with high affinities for 5hmC, which may represent specific readers of this modification, appear to exist (Iurlaro *et al.* 2013, Spruijt *et al.* 2013). Thus, 5hmC is frequently detected at high levels in neural tissues in the absence of 5fC and 5caC and may play an instructive role. Since it cannot be distinguished from 5mC by standard bisulfite modification and sequencing, it may contribute a substantial proportion of the DNA methylation reported at some loci. Examination of imprinted genes using methods that detect 5hmC have indicated however that the majority of methylation seen at ICR is indeed 5mC in adult tissues, though 5hmC is relatively enriched at imprinted loci compared to the levels elsewhere in the genome (Hernandez Mora *et al.* 2005). The 5hmC was present on the same allele as the 5mC, suggesting to the authors that it might be a stable mark here and not a sign of active demethylation. However, it is possible that this might reflect a dynamic turnover of methyl groups here, since Zhang and colleagues recently showed that G9a/GLP was required to recruit *de novo* methyltransferases to counteract TET activity at imprinted ICR (Zhang *et al.* 2016). However, levels of oxidised methyl cytosine products at ICR were not measured directly in the latter experiments. Much of the 5hmC that was detected by Hernandez Mora and colleagues was in transcribed regions of imprinted genes, not ICR, in keeping with an earlier report for *H19* (Nestor *et al.* 2012). Intriguingly, an association has been reported between 5hmC levels in *H19* and small-for-gestational-age babies (Piyasena *et al.* 2015).

Thus, while there is good evidence for high levels of 5hmC and other modified cytosines being seen at imprints during reprogramming, levels appear to be much lower in adult tissues except brain, and may be more related to transcription in the latter, since found in the gene bodies.

Can somatic reprogramming of imprints occur?

One of the key features of imprinted regions is their programming in the germ line, when the parental alleles are separated. Early work showed that in *Dnmt1*−/− ES cells (1KO – Fig. 1B), imprints lost DNA methylation and could not recover it even when DNMT1 was restored using a cDNA (1KO+1), unlike the rest of the genome. However, when the rescued ES cells were used to make mice, the imprints were restored following passage through the germline (Tucker *et al.* 1996). This is thought to be due to the imperviousness of imprints to the normal reprogramming events seen in the soma. Indeed, the inability to reprogram in somatic tissues is a key difference between the imprinted gDMRs and the other two classes of gDMR, namely those found in germline and brain genes (Borgel *et al.* 2010, Rutledge *et al.* 2014), whose methylation can be restored in 1KO+1 cells (Fig. 1B). We and others have confirmed these results both in ES cells (Chen *et al.* 2003, Thakur *et al.* 2016) and in adult cell lines (Chen *et al.* 2003, Wernig *et al.* 2007) (Fig. 1B and C). Interestingly, imprints appear to be resistant to demethylation as well as remethylation: treatment of cells with 5′Aza-2′deoxytidine (Aza) results in robust demethylation of most targets, but not imprints (O'Neill *et al.* 2018).

However, recently two ES cell systems have been described in which imprints can apparently be restored, breaking one of the main rules of imprinting. Wong and colleagues, working with UHRF1-deficient ES cells (Qi *et al.* 2015), found that most imprinted gDMRs lost methylation, as expected, since UHRF1 appears to be a vital cofactor for DNMT1 in somatic cells (Bostick *et al.* 2007, Sharif *et al.* 2007). Surprisingly however, they found that when they rescued these cells using a UHRF1 cDNA that not only was methylation restored on bulk DNA, but also is a subset of the imprinted loci showed recovery. In particular, the *H19*, *Nnat* and *Dlk1* gDMRs showed some recovery of methylation, though not to WT levels, and ZFP57 binding was restored at 2/3 loci (Qi *et al.* 2015). While it would be reasonable to assume that an underlying chromatin mark might be retained, allowing restoration of methylation, there was no clear correlation between recovery and chromatin marks at the loci (Qi *et al.* 2015).

Using a different ES cell system, we too found somatic recovery of imprints could occur under some conditions. In cells lacking DNMT3A and DNMT3B (3abKO), imprinted gDMRs lost their parental allele-specific methylation to almost the same extent as in DNMT1-deficient cells. However, in contrast to rescue experiments in 1KO cells, when DNMT3A2 was added back to the 3abKO cells, we found that DNA methylation was restored at most gDMRs examined (11/14), with many achieving levels similar to the WT cells (Fig. 1B). Results were confirmed using up to three different techniques (Thakur *et al.* 2016).

Taken together, these two sets of results independently confirm that in some ES cell types, methylation can be restored on imprinted gDMRs outside of the germline opening the way for further exploration of the factors and signals that may be involved.

Implications for reproductive biology

How many imprinted loci are there?

In the index species mouse, where most data are available and there have been some dedicated searches, the number of verified imprinted loci remains relatively steady at around 150. Despite a few high-profile studies suggesting more may exist, the number of genes with confirmed uniparental expression has not increased greatly. Many of the known imprinted genes in mouse (http://www.mousebook.org/imprinting-gene-list (accessed 24/1/2018)) are also imprinted in human (http://igc.otago.ac.nz/1601summarytable.pdf (accessed 24/1/2018)). A recent study by Wang *et al.* (2014) identified a small number of new germline DMR, for which uniparental methylation has been verified by ourselves or others (Thakur *et al.* 2016), but the effects on transcription are yet to be confirmed. Likewise a study in human by Court *et al.* (2014) using a combination of arrays and whole-genome bisulfite sequencing (WGBS – see below) added some new DMR. These were largely confined to known imprinted loci, although some were novel placenta-specific DMR. One caveat with the latter is that the differential methylation appears to have been established postnatally, which would require a radical rethink of mechanisms. However, some evidence exists to suggest that this may also occur in mouse (Wagschal & Feil 2006).

In addition to the 'true' imprints, there are a number of 'transient' imprinted genes that show widespread imprinting in the early embryo that later becomes confined to one tissue (see 'Post-implantation changes' section above). This larger group of neuronal and germline genes have gDMRs in the early embryo that become erased post-implantation, largely through the *de novo* methylation of the unmethylated paternal allele (Borgel *et al.* 2010, Proudhon *et al.* 2012, Rutledge *et al.* 2014). These loci have not all been tested for uniparental expression in early development, but may contribute to parent-of-origin effects in early embryogenesis.

How conserved are imprints in other species?

Initial reports of imprinting in mouse were quickly followed by the investigation of corresponding human loci, and the subsequent discovery that at least one of the index loci *IGF2R* was not imprinted in all humans, but may instead show polymorphic imprinting (Xu *et al.* 1993). As the number of well-characterised imprinted loci in mouse grew, the general trend was for the homologous locus in human to be imprinted, but with a number of notable exceptions (Hanna & Kelsey 2014, Peters 2014). In general, though, the emerging complexity of imprint locus control has tended to discourage attempts to verify in detail in the second species, with an assumption of similarity being adopted on the whole unless forced to be re-evaluated due to clinical or experimental evidence.

Some early studies also tried to map the index imprinted loci in marsupials in order to test parental conflict theories as well as to investigate evolutionary mechanisms. This met with some success, despite the difficulty of the exercise. There is now support for an origin for imprinting at the time of divergence of marsupials (metatheria) and egg-laying mammals (montremes), which lack imprinting (Renfree *et al.* 2013). The oldest imprinted loci with a conserved DMR are *H19* and *PEG10*, though a number of other imprinted genes such as *IGF2R* and *PEG1* are also imprinted, but lack a conserved DMR. Many other loci appear to have acquired imprinting in eutheria (Renfree *et al.* 2013). *PEG10* is a retroposon-derived gene, and it has long been suggested that there is a link between methylation at newly acquired retrotransposons and imprinting (Yoder *et al.* 1997). It is interesting in this context that rodentia are notable for having the highest number of imprinted loci and also a unique methyltransferase locus *Dnmt3c*, initially thought to be a pseudogene based on an early draft of the rat genome (Lees-Murdock *et al.* 2004) but now shown to produce a functional enzyme confined to the male rodent germline that appears dedicated to retroposon methylation (Barau *et al.* 2016).

ART and imprints

In livestock species, there has also been a lively interest in imprinted loci (O'Doherty *et al.* 2012), partly due to the widespread use of artificial reproduction techniques (ART), which are thought to be particularly associated with perturbed methylation and development including syndromes such as the aforementioned LOS (see above), enlarged placenta and perinatal death. These have been tied to changes in methylation at imprinted loci in a number of studies in cattle (O'Doherty *et al.* 2012, 2014). While studies in mouse have found substantial support for ART-induced alterations in imprints (Rivera *et al.* 2008, Denomme & Mann 2012), in humans, any link between methylation changes at imprinted loci and standard IVF has been controversial, with most reviews of the area not able to fully exclude an effect on imprints or other epigenetically regulated loci (Grafodatskaya *et al.* 2013). However, the addition of large enough concentrations of inhibitors such as aphidicolin have conclusively been shown to alter the programmed demethylation of the genome (Guo *et al.* 2014) at the pronuclear stage, so

Table 1 Technologies for genome-wide methylation analysis in animals.

Type[a]	Acronym	Technology	Spp[b]	Cost[c]	Coverage[d]	meC[e]	Input[f]	Res[g]	Notes[h]	Reference
Array	450K/EPIC	450K/850K probe array (Illumina)	Hu	L	Exome, CGI, regulatory	L	M	L	No TEs	Bibikova *et al*. (2011)
	EDMA	EmbryoGENE DNA methylation analysis	Bo	L	Exome, CGI, promoter	L	M	L	Qualitative, no TEs	Shojaei Saadi *et al*. (2014)
Reseq	WGBS	Whole-genome bisulfite sequencing	Any	H	Total	H	H	H	Bioinformatic analysis-heavy	Cokus *et al*. (2008), Lister *et al*. (2008), Meissner *et al*. (2008)
	RRBS	Reduced representation bisulfite sequencing	Any	M	Fraction	M	M	H	Partial CGI cover	Meissner *et al*. (2005)
	PBAT	Post-bisulfite adaptor tagging	Any	M	Total	H	L	H	Amplification-free	Miura *et al*. (2012)
IP	MeDIP-seq	Methylated DNA immunoprecipitation	Any	M	Fraction	M	M	H	Favours sparse CpG	Weber *et al*. (2005)
	MBD-CAP	Methyl binding domain CAPture	Any	M	Fraction	M	M	H	Favours CGI	Brinkman *et al*. (2010), Serre *et al*. (2010)
uCGI	MRE-seq	Methylation-sensitive restriction enzyme	Any	M	Fraction	M	M	L	uCGI-specific	Maunakea *et al*. (2010)
	CFP1-seq	CFP1 protein IP	Any	M	Fraction	M	M	H	uCGI-specific	Thomson *et al*. (2010)
SC	scWGBS	Single-cell WGBS	Any	H	Total	H	L	H	Bioinfo-heavy	Farlik *et al*. (2015)
	scRRBS	Single-cell RRBS	Any	M	Fraction	M	L	H	Bioinfo-heavy	Guo *et al*. (2013)
	scPBAT	Single-cell PBAT	Any	H	Total	H	L	H	Bioinfo-heavy	Smallwood *et al*. (2014)
modC	SMRT	Single molecule real time (PacBio)	Any	H	Total	H	H	H	All modifications	Flusberg *et al*. (2010)
	hMeDIP-Seq	5hmC antibody IP	Any	M	Fraction	M	M	H	5hmC-specific	Jin *et al*. (2011)
	oxBS-seq	Oxidative bisulfite sequencing (CEGX)	Any	M	Total	H	H	H	5hmC and 5mC	Booth *et al*. (2012)
	TAB-Seq	TET-assisted bisulfite sequencing	Any	M	Total	H	H	H	5hmC and 5mC	Yu *et al*. (2012, 2018)
	fCAB-seq		Any	M	Total	H	H	H	5fC-specific	Song *et al*. (2013)
	JBP1-seq	JBP1 5hmC-binding protein IP	Any	M	Fraction	M	M	H	5hmC-specific	Cui *et al*. (2014)

[a]Type: Array, microarray; Reseq, resequencing; IP, immunoprecipitation; uCGI, targets unmethylated CGI; SC, single-cell; modC, targets modified 5mC. [b]Spp: species: Hu, human-specific; Bo, bovine-specific; any, can in theory be applied to any. [c]Cost: L, Low; M, Medium; H, High. [d]Coverage: genome-wide coverage: Exome, exons; CGI, CpG islands; Regulatory, promoters and enhancers; Total, sequences whole genome; Fraction, subsets the genome but includes all types of sequence. [e]meC: fraction of all meC covered: Low, Medium, High. [f]Input: amount and purity of DNA: Low, lower quality and amount needed; Medium, better quality and more needed; High, highly pure DNA at higher amounts. [g]Res: resolution at target, i.e. are all 5mC covered in target areas (e.g. RRBS) or only a fraction at each location (arrays): Low, Medium, High. [h]Notes: TE, transposable elements; 5hmC, 5-hydroxymethylcytosine detection only; 5fC, 5-formylcytosine only.

environmental perturbations of imprinting remain at least theoretically plausible.

Use of sequencing technologies to investigate known and potential imprinted regions

There have been a number of approaches taken to identify potentially imprinted genes and to more closely define the size of the imprint, partly dictated by the tools and approaches available in the differing species. In humans for example, microarrays have made assessing methylation easier, but low heterozygosity levels and incomplete data make assignment of parental allele and gametic imprints – even with resequencing approaches – problematic, whereas in mice, no methylation arrays are available but interspecific crosses can be generated to maximise heterozygosity. Ruminants on the other hand have a long reproductive cycle and inter-strain hybrids are not commonly used, and with some exceptions, there are no arrays here either. For these reasons, resequencing approaches offer the most widely applicable tool (Table 1). While sequencing can be restrictive in cost, it represents a powerful discovery platform.

Wang and colleagues for example used WGBS to investigate the methylation status of the 54 imprinted DMR in the mouse genome (Wang *et al.* 2014), and in particular to sort unclassified DMR into gametic or somatic categories. They compared oocyte, sperm and early embryo, and began by confirming they could correctly sort all the DMR whose status was known (29) into gametic vs somatic types. They then used the same criteria to classify 25 imprinted DMRs whose status was unknown into gametic or somatic DMR, which is essential for determining where primary control of each locus is being exerted. At least four of these newly classified DMRs could subsequently be independently confirmed using a pyrosequencing-based assay (Thakur *et al.* 2016).

In humans, microarray technologies and in particular the Illumina EPIC array, have become so reliable in terms of assessing methylation levels quantitatively that confirmation by a second technology has become almost redundant. The low relative cost of the array makes it an attractive approach for initial screening and a comparison of uniparental disomies could correctly identify all 30 DMRs for which the array had probes (one DMR at *MEG3/DLK1* is not covered). They additionally identified 21 novel putative imprints, 15 specific to placenta, albeit a number of them were at known imprinted loci already (Court *et al.* 2014). However, the low relative resolution of the array meant that they needed to use WGBS to accurately delineate the DMR.

Reliable standardised methylation microarrays are largely confined to humans, though the success of the 450K and EPIC platforms has meant that Illumina appears to be considering a similar BeadChip array for mouse. An exception is cattle, where the EDMA array has been developed for assessing genome-wide methylation (Shojaei Saadi *et al* 2014), though this is based on cutting with methylation-sensitive restriction enzymes rather than oligos. While the array gives good internal consistency and technical reproducibility and could correctly identify methylation status at a number of imprinted regions in this species, it gives relative rather than absolute levels of methylation and is therefore difficult to correlate with outputs from many other assays (Desmet *et al.* 2016), though not all (Ispada *et al.* 2018). Whole-genome sequencing represents the gold standard for exploratory work and can definitively assess methylation across all sites. Confirming parental origin of the methylation mark requires heterozygous SNPs

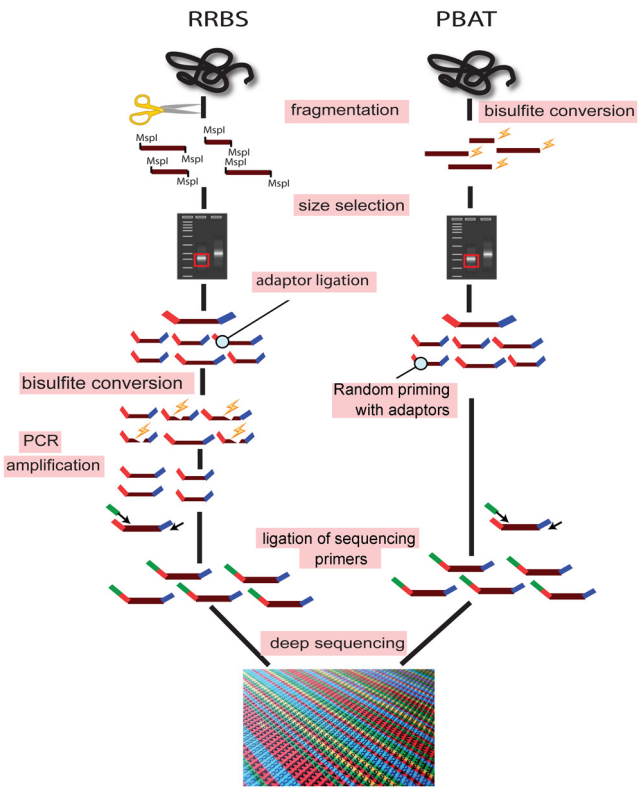


Figure 2 Schematic comparison of reduced representation bisulfite sequencing (RRBS) vs post-bisulfite adaptor tagging (PBAT). In RRBS (left), the DNA is initially fragmented by a methylation-insensitive restriction endonuclease such as *Msp*I to an optimal fragment size of between 400 and 600 bp, and this fraction isolated following gel electrophoresis or similar. Adaptors are ligated to these DNA fragments that are subsequently treated through bisulfite conversion to convert unmethylated cytosines to thymines for the differential readout. Amplification is carried out using primers specific to the adaptor sequences, but there is substantial loss of sample at this point due to strand breakage (flashes) between adaptors caused by bisulfite. PBAT (right) skips the initial enzyme digestion and instead takes advantage of the bisulfite-induced breakage to fragment the DNA here. Instead of PCR, the adaptors are used for two rounds of random primer extension, which serves to produce numerous reads from a small amount of input genomic DNA.

however. While Liu and colleagues could find sufficient variation in an inter-strain cross for mice (Wang *et al.* 2014), generating cross-bred cattle is not trivial given cost implications and slow reproductive times and most reported work in the area does not feature such animals.

An alternative, more widely applicable and cost-effective approach for assessing methylation is to subset the genome and do targeted resequencing. Reduced representation bisulfite sequencing has been the most extensively used method in this category so far (Meissner *et al.* 2005), but is being superseded by post-bisulfite adaptor tagging (PBAT) (Miura *et al.* 2012), which appears to offer improvements in yield and efficiency, particularly when starting with small amounts of material such as oocytes (Smallwood *et al.* 2014) (Fig. 2). Apart from cost considerations, a significant bottleneck for resequencing-based analyses is the extensive bioinformatic analysis required, with concomitant need for higher-specification computing hardware, local network clusters and data warehousing, all adding substantial capital and human resource costs to the equation (Lewitter *et al.* 2009). While this has resulted in the majority of such work being undertaken at large central institutes that have developed the required infrastructure, often with national support, there remains a niche for smaller players out-sourcing the sequencing to larger centres with surplus capacity and who may in future also be able to avail of cloud-based storage and analysis capabilities (Liu *et al.* 2014).

Another point to consider here is that some species that may be of interest for biologists wishing to understand the origins or phylogenetic reach of imprinting may not be sequenced yet. In cases where a reference genome framework is unavailable, assembling sequence reads and defining variation become considerably more challenging (Cahais *et al.* 2012).

Finally, the impact of third-generation sequencing technologies is yet to be fully felt, given that this market is currently in flux and no clear front-runner appears yet to have emerged. Current technologies require a separate sequence library for each type of cytosine modification for example (meC, 5hmC, 5fC, 5caC) (Yu *et al.* 2012), whereas certain new approaches (Table 1) such as PacBio's technology have the capability to also call methylated bases while sequencing, obviating the need for multiple sequence sets (Lister & Ecker 2009).

Conclusions and future prospects

Imprinting clearly plays an important role in body size determination and nutrient transfer to the young, making assessment of methylation at key loci a useful tool for reproductive biologists and breeders alike. Despite the improvement in technologies for assessing and delineating imprinting, the number of verified imprinted loci has not increased very significantly. However, our understanding of the mechanisms by which these regions are established and maintained in early life, as well as the contexts in which some plasticity in imprinting may occur, have greatly increased. We can look forward to further insights from comparative genomic approaches based on the increasing application of sequencing technologies to an ever-expanding range of organisms, including the so far rather surprisingly neglected livestock species, which play such an important role in human food production and welfare.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

The authors' work was supported in part by the Medical Research Council (grant number MR/J007773/1) and a grant jointly funded by the ESRC/BBSRC (grant number ES/N000323/1).

Acknowledgements

The authors thank members of the Walsh lab for helpful comments and apologize to colleagues whose work could not be cited owing to length limitations.

References

Allen WR 1969 Factors influencing pregnant mare serum gonadotrophin production. *Nature* **223** 64–65. (https://doi.org/10.1038/223064a0)

Aravind L & Koonin EV 2000 SAP – a putative DNA-binding motif involved in chromosomal organization. *Trends in Biochemical Sciences* **25** 112–114. (https://doi.org/10.1016/S0968-0004(99)01537-6)

Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, Guillou F & Bourc'his D 2016 The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354** 909–912. (https://doi.org/10.1126/science.aah5143)

Barlow DP & Bartolomei MS 2014 Genomic imprinting in mammals. *Cold Spring Harbor Perspectives in Biology* **6** (https://doi.org/10.1101/cshperspect.a018382)

Bartolomei MS & Ferguson-Smith AC 2011 Mammalian genomic imprinting. *Cold Spring Harbor Perspectives in Biology* **3** a018382. (https://doi.org/10.1101/cshperspect.a002592)

Barton SC, Surani MA & Norris ML 1984 Role of paternal and maternal genomes in mouse development. *Nature* **311** 374–376. (https://doi.org/10.1038/311374a0)

Bell AC & Felsenfeld G 2000 Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405** 482–485. (https://doi.org/10.1038/35013100)

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL *et al.* 2011 High density DNA methylation array with single CpG site resolution. *Genomics* **98** 288–295. (https://doi.org/10.1016/j.ygeno.2011.07.007)

Bidwell CA, Waddell JN, Taxis TM, Yu H, Tellam RL, Neary MK & Cockett NE 2014 New insights into polar overdominance in callipyge sheep. *Animal Genetics* **45** (Supplement 1) 51–61. (https://doi.org/10.1111/age.12132)

Bird AP 1987 CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics* **3** 342–347. (https://doi.org/10.1016/0168-9525(87)90294-0)

Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W & Balasubramanian S 2012 Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336** 934–937. (https://doi.org/10.1126/science.1220671)

Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T & Weber M 2010 Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genetics* **42** 1093–1100. (https://doi.org/10.1038/ng.708)

Bostick M, Kyong Kim J, Esteve P, Clark A & Pradhan S 2007 UHRFI plays a role in maintaining DNA methylation in mammalian cells. *Science* **317** 1760–1764. (https://doi.org/10.1126/science.1147939)

Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J & Stunnenberg HG 2010 Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* **52** 232–236. (https://doi.org/10.1016/j.ymeth.2010.06.012)

Buiting K 2010 Prader-Willi syndrome and Angelman syndrome. *American Journal of Medical Genetics: Part C, Seminars in Medical Genetics* **154C** 365–376. (https://doi.org/10.1002/ajmg.c.30273)

Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, Chiari Y, Belkhir K, Ranwez V & Galtier N 2012 Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources* **12** 834–845. (https://doi.org/10.1111/j.1755-0998.2012.03148.x)

Chen T, Ueda Y, Dodge JE, Wang Z & Li E 2003 Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology* **23** 5594–5605. (https://doi.org/10.1128/MCB.23.16.5594-5605.2003)

Chen Z, Robbins KM, Wells KD & Rivera RM 2013 Large offspring syndrome: a bovine model for the human loss-of-imprinting overgrowth syndrome Beckwith-Wiedemann. *Epigenetics* **8** 591–601. (https://doi.org/10.4161/epi.24655)

Chen Z, Hagen DE, Elsik CG, Ji T, Morris CJ, Moon LE & Rivera RM 2015 Characterization of global loss of imprinting in fetal overgrowth syndrome induced by assisted reproduction. *PNAS* **112** 4618–4623. (https://doi.org/10.1073/pnas.1422088112)

Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E & Chen T 2009 KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **461** 415–418. (https://doi.org/10.1038/nature08315)

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M & Jacobsen SE 2008 Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452** 215–219. (https://doi.org/10.1038/nature06745)

Court F, Tayama C, Romanelli V, Martin-Trujillo A, Iglesias-Platas I, Okamura K, Sugahara N, Simon C, Moore H, Harness JV *et al.* 2014 Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Research* **24** 554–569. (https://doi.org/10.1101/gr.164913.113)

Cui L, Chung TH, Tan D, Sun X & Jia XY 2014 JBP1-seq: a fast and efficient method for genome-wide profiling of 5hmC. *Genomics* **104** 368–375. (https://doi.org/10.1016/j.ygeno.2014.08.023)

Curley JP, Pinnock SB, Dickson SL, Thresher R, Miyoshi N, Surani MA & Keverne EB 2005 Increased body fat in mice with a targeted mutation of the paternally expressed imprinted gene Peg3. *FASEB Journal* **19** 1302–1304. (https://doi.org/10.1096/fj.04-3216fje)

Denomme MM & Mann MR 2012 Genomic imprints as a model for the analysis of epigenetic stability during assisted reproductive technologies. *Reproduction* **144** 393–409. (https://doi.org/10.1530/REP-12-0237)

Desmet KL, Van Hoeck V, Gagne D, Fournier E, Thakur A, O'Doherty AM, Walsh CP, Sirard MA, Bols PE & Leroy JL 2016 Exposure of bovine oocytes and embryos to elevated non-esterified fatty acid concentrations: integration of epigenetic and transcriptomic signatures in resultant blastocysts. *BMC Genomics* **17** 1004–1022. (https://doi.org/10.1186/s12864-016-3366-y)

Duffie R, Ajjan S, Greenberg MV, Zamudio N, Escamilla del Arenal M, Iranzo J, Okamoto I, Barbaux S, Fauque P & Bourc'his D 2014 The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. *Genes and Development* **28** 463–478.

Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schonegger A, Klughammer J & Bock C 2015 Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Reports* **10** 1386–1397. (https://doi.org/10.1016/j.celrep.2015.02.001)

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J & Turner SW 2010 Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7** 461–465. (https://doi.org/10.1038/nmeth.1459)

Freking BA, Keele JW, Nielsen MK, Leymaster KA 2002*a* Evaluation of the ovine callipyge locus: II. Genotypic effects on growth, slaughter, and carcass traits. *Journal of Animal Science* **76** 2549–2559 (https://doi.org/10.2527/1998.76102549x)

Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, Jirtle RL & Smith TP 2002*b* Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Research* **12** 1496–1506. (https://doi.org/10.1101/gr.571002)

Grafodatskaya D, Cytrynbaum C & Weksberg R 2013 The health risks of ART. *EMBO Reports* **14** 129–135. (https://doi.org/10.1038/embor.2012.222)

Guillemot F, Caspary T, Tilghman SM, Copeland NG, Gilbert DJ, Jenkins NA, Anderson DJ, Joyner AL, Rossant J & Nagy A 1995 Genomic imprinting of Mash2, a mouse gene required for trophoblast development. *Nature Genetics* **9** 235–242. (https://doi.org/10.1038/ng0395-235)

Guo H, Zhu P, Wu X, Li X, Wen L & Tang F 2013 Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Research* **23** 2126–2135. (https://doi.org/10.1101/gr.161679.113)

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu TP, Hu B *et al.* 2014 Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15** 447–459. (https://doi.org/10.1016/j.stem.2014.08.003)

Hahn MA, Szabo PE & Pfeifer GP 2014 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics* **104** 314–323. (https://doi.org/10.1016/j.ygeno.2014.08.015)

Haig D & Graham C 1991 Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell* **64** 1045–1046. (https://doi.org/10.1016/0092-8674(91)90256-X)

Hanna CW & Kelsey G 2014 The specification of imprints in mammals. *Heredity* **113** 176–183. (https://doi.org/10.1038/hdy.2014.54)

Hanna CW, Penaherrera MS, Saadeh H, Andrews S, McFadden DE, Kelsey G & Robinson WP 2016 Pervasive polymorphic imprinted methylation in the human placenta. *Genome Research* **26** 756–767. (https://doi.org/10.1101/gr.196139.115)

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM & Tilghman SM 2000 CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405** 486–489 (https://doi.org/10.1038/35013106)

Hernandez Mora JR, Sanchez-Delgado M, Petazzi P, Moran S, Esteller M, Iglesias-Platas I & Monk D 2005 Profiling of oxBS-450K 5-hydroxymethylcytosine in human placenta and brain reveals enrichment at imprinted loci. *Epigenetics* **13** 182–191. (https://doi.org/10.1080/15592294.2017.1344803)

Hill PW, Amouroux R & Hajkova P 2014 DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104** 324–333. (https://doi.org/10.1016/j.ygeno.2014.08.012)

Ispada J, de Lima CB, Sirard MA, Fontes PK, Nogueira MFG, Annes K & Milazzotto MP 2018 Genome-wide screening of DNA methylation in bovine blastocysts with different kinetics of development. *Epigenetics and Chromatin* **11** 1. (https://doi.org/10.1186/s13072-017-0171-z)

Iurlaro M, Ficz G, Oxley D, Raiber E, Bachman M, Booth MJ, Andrews S, Balasubramanian S & Reik W 2013 A screen for hydroxymethylcytosine and formyl cytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology* **14** R119. (https://doi.org/10.1186/gb-2013-14-10-r119)

Jeon JT, Carlborg O, Tornsten A, Giuffra E, Amarger V, Chardon P, Andersson-Eklund L, Andersson K, Hansson I, Lundstrom K *et al.* 1999 A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nature Genetics* **21** 157–158. (https://doi.org/10.1038/5938)

Jin SG, Wu X, Li AX & Pfeifer GP 2011 Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Research* **39** 5015–5024. (https://doi.org/10.1093/nar/gkr120)

Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M & Sasaki H 2007 Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell

development in the mouse. *Human Molecular Genetics* **16** 2272–2280. (https://doi.org/10.1093/hmg/ddm179)

Keverne EB & Curley JP 2008 Epigenetics, brain evolution and behaviour. *Frontiers in Neuroendocrinology* **29** 398–412. (https://doi.org/10.1016/j.yfrne.2008.03.001)

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y *et al.* 2012 Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genetics* **8** e1002440. (https://doi.org/10.1371/journal.pgen.1002440)

Lees-Murdock DJ, McLoughlin GA, McDaid JR, Quinn LM, O'Doherty A, Hiripi L, Hack CJ & Walsh CP 2004 Identification of 11 pseudogenes in the DNA methyltransferase gene family in rodents and humans and implications for the functional loci. *Genomics* **84** 193–204. (https://doi.org/10.1016/j.ygeno.2004.02.004)

Lewitter F, Rebhan M, Richter B & Sexton D 2009 The need for centralization of computational biology resources. *PLoS Computational Biology* **5** e1000372. (https://doi.org/10.1371/journal.pcbi.1000372)

Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P & Ferguson-Smith AC 2008 A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Developmental Cell* **15** 547–557. (https://doi.org/10.1016/j.devcel.2008.08.014)

Lister R & Ecker JR 2009 Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Research* **19** 959–966. (https://doi.org/10.1101/gr.083451.108)

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH & Ecker JR 2008 Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133** 523–536. (https://doi.org/10.1016/j.cell.2008.03.029)

Liu B, Madduri RK, Sotomayor B, Chard K, Lacinski L, Dave UJ, Li J, Liu C & Foster IT 2014 Cloud-based bioinformatics workflow platform for large-scale next-generation sequencing analyses. *Journal of Biomedical Informatics* **49** 119–133 (https://doi.org/10.1016/j.jbi.2014.01.005)

Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP *et al.* 2008 Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. *Nature Genetics* **40** 949–951. (https://doi.org/10.1038/ng.187)

Markljung E, Jiang L, Jaffe JD, Mikkelsen TS, Wallerman O, Larhammar M, Zhang X, Wang L, Saenz-Vash V, Gnirke A *et al.* 2009 ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth. *PLoS Biology* **7** e1000256. (https://doi.org/10.1371/journal.pbio.1000256)

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y *et al.* 2010 Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466** 253–257. (https://doi.org/10.1038/nature09165)

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES & Jaenisch R 2005 Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research* **33** 5868–5877. (https://doi.org/10.1093/nar/gki901)

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB *et al.* 2008 Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454** 766–770. (https://doi.org/10.1038/nature07107)

Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A & Knowles BB 2012 Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* **335** 1499–1502. (https://doi.org/10.1126/science.1216154)

Miura F, Enomoto Y, Dairiki R & Ito T 2012 Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Research* **40** e136. (https://doi.org/10.1093/nar/gks454)

Moore T & Haig D 1991 Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in Genetics* **7** 45–49. (https://doi.org/10.1016/0168-9525(91)90230-N)

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M *et al.* 2007 PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology* **9** 64–71. (https://doi.org/10.1038/ncb1519)

Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y & Nakano T 2012 PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486** 415–419 (https://doi.org/10.1038/nature11093)

Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ & Meehan RD 2012 Tissue type is a major modifier of the 5′-hydroxymethylcytosine content of human genes. *Genome Research* **22** 467–477. (https://doi.org/10.1101/gr.126417.111)

O'Doherty A, O'Shea LC & Fair T 2012 Bovine DNA methylation imprints are established in an oocyte size-specific manner, which are coordinated with the expression of the DNMT3 family proteins. *Biology of Reproduction* **86** 1–10. (https://doi.org/10.1095/biolreprod.111.094946)

O'Doherty A, O'Gorman A, al Naib A, Brennan L, Daly E, Duffy P & Fair T 2014 Negative energy balance affects imprint stability in oocytes recovered from postpartum dairy cows. *Genomics* **104** 177–185. (https://doi.org/10.1016/j.ygeno.2014.07.006)

O'Neill KM, Irwin RE, Mackin SJ, Thursby SJ, Thakur A, Bertens C, Masala L, Loughery JEP, McArt DG & Walsh CP 2018 Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression. *Epigenetics and Chromatin* **11** 12. (https://doi.org/10.1186/s13072-018-0182-4)

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD *et al.* 2007 DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448** 714–717. (https://doi.org/10.1038/nature05987)

Payer B, Saitou M, Barton SC, Thresher R, Dixon JPC, Zahn WH, Cartlon MBL, Nakano T & Surani MA 2003 Stella is a maternal effect gene required for normal early development in mice. *Current Biology* **13** 2110–2117. (https://doi.org/10.1016/j.cub.2003.11.026)

Peters J 2014 The role of genomic imprinting in biology and disease: an expanding view. *Nature Reviews Genetics* **15** 517–530. (https://doi.org/10.1038/nrg3766)

Piyasena C, Reynolds RM, Khulan B, Seckl JR, Menon G & Drake AJ 2015 Placental 5-methylcytosine and 5-hydroxymethylcytosine patterns associate with size at birth. *Epigenetics* **10** 692–697. (https://doi.org/10.1080/15592294.2015.1062963)

Prickett AR, Barkas N, McCole RB, Hughes S, Amante SM, Schulz R & Oakey RJ 2013 Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome Research* **23** 1624–1635. (https://doi.org/10.1101/gr.150136.112)

Proudhon C, Duffie R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyan VK *et al.* 2012 Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Molecular Cell* **47** 909–920 (https://doi.org/10.1016/j.molcel.2012.07.010)

Qi S, Wang Z, Li P, Wu Q, Shi T, Li J & Wong J 2015 Non-germ line restoration of genomic imprinting for a small subset of imprinted genes in ubiquitin-like PHD and RING finger domain-containing 1 (Uhrf1) null mouse embryonic stem cells. *Journal of Biological Chemistry* **290** 14181–14191. (https://doi.org/10.1074/jbc.M114.626697)

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A *et al.* 2011 In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular Cell* **44** 361–372. (https://doi.org/10.1016/j.molcel.2011.08.032)

Renfree MB, Suzuki S & Kaneko-Ishino T 2013 The origin and evolution of genomic imprinting and viviparity in mammals. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences* **368** 20120151. (https://doi.org/10.1098/rstb.2012.0151)

Rivera RM, Stein P, Weaver JR, Mager J, Schultz RM & Bartolomei MS 2008 Manipulations of mouse embryos prior to implantation result in aberrant expression of imprinted genes on day 9.5 of development. *Human Molecular Genetics* **17** 1–14. (https://doi.org/10.1093/hmg/ddm280)

Rutledge CE, Thakur A, O'Neill KM, Irwin RE, Sato S, Hata K & Walsh CP 2014 Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes. *Development* **141** 1313–1323. (https://doi.org/10.1242/dev.104646)

Sato M, Kimura T, Kurokawa K, Fujita Y, Abe K, Masuhara M, Yasunaga T, Ryo A, Yamamoto M & Nakano T 2002 Identification of PGC7, a new gene expressed specifically in preimplantation embryos and germ cells.

Mechanisms of Development **113** 91–94. (https://doi.org/10.1016/S0925-4773(02)00002-3)

Serre D, Lee BH & Ting AH 2010 MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Research* **38** 391–399. (https://doi.org/10.1093/nar/gkp992)

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K *et al.* 2007 The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450** 908–912. (https://doi.org/10.1038/nature06397)

Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T & Sasaki H 2013 Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genetics* **9** e1003439. (https://doi.org/10.1371/journal.pgen.1003439)

Shojaei Saadi HA, O'Doherty AM, Gagne D, Fournier E, Grant JR, Sirard MA & Robert C 2014 An integrated platform for bovine DNA methylome analysis suitable for small samples. *BMC Genomics* **15** 451–468. (https://doi.org/10.1186/1471-2164-15-451)

Sleutels F, Zwart R & Barlow DP 2002 The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415** 810–813. (https://doi.org/10.1038/415810a)

Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR & Kelsey G 2011 Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genetics* **43** 811–814. (https://doi.org/10.1038/ng.864)

Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W & Kelsey G 2014 Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11** 817–820. (https://doi.org/10.1038/nmeth.3035)

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu HC 2013 Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153** 678–691. (https://doi.org/10.1016/j.cell.2013.04.001)

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F *et al.* 2013 Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152** 1146–1159. (https://doi.org/10.1016/j.cell.2013.02.004)

Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M, McEwen K, Sun B, Pedersen R & Ferguson-Smith AC 2015 Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biology* **16** 112–119. (https://doi.org/10.1186/s13059-015-0672-7)

Szabó PE & Pfeifer GP 2012 H3K9me2 attracts PGC7 in the zygote to prevent Tet3-mediated oxidation of 5-methylcytosine. *Journal of Molecular Cell Biology* **4** 427–429. (https://doi.org/10.1093/jmcb/mjs038)

Thakur A, Mackin SJ, Irwin RE, O'Neill KM, Pollin G & Walsh C 2016 Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2. *Epigenetics and Chromatin* **9** 53–67. (https://doi.org/10.1186/s13072-016-0104-2)

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD *et al.* 2010 CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464** 1082–1086. (https://doi.org/10.1038/nature08924)

Tucker KL, Beard C, Dausmann J, Jackson-Grusby L, Laird PW, Lei H, Li E & Jaenisch R 1996 Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes and Development* **10** 1008–1020. (https://doi.org/10.1101/gad.10.8.10080)

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M *et al.* 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425** 832–836. (https://doi.org/10.1038/nature02064)

Wagschal A & Feil R 2006 Genomic imprinting in the placenta. *Cytogenetic and Genome Research* **113** 90–98. (https://doi.org/10.1159/000090819)

Walsh C, Glaser A, Fundele R, Ferguson-Smith A, Barton S, Surani MA & Ohlsson R 1994 The non-viability of uniparental mouse conceptuses correlates with the loss of the products of imprinted genes. *Mechanisms of Development* **46** 55–62. (https://doi.org/10.1016/0925-4773(94)90037-X)

Wang X, Miller DC, Clark AG & Antczak DF 2012 Random X inactivation in the mule and horse placenta. *Genome Research* **22** 1855–1863. (https://doi.org/10.1101/gr.138487.112)

Wang X, Miller DC, Harman R, Antczak DF & Clark AG 2013 Paternally expressed genes predominate in the placenta. *PNAS* **110** 10705–10710. (https://doi.org/10.1073/pnas.1308998110)

Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, Yang L, Zhang J, Li G, Ci W *et al.* 2014 Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157** 979–991. (https://doi.org/10.1016/j.cell.2014.04.017)

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL & Schubeler D 2005 Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genetics* **37** 853–862. (https://doi.org/10.1038/ng1598)

Weksberg R, Shuman C & Beckwith JB 2010 Beckwith-Wiedemann syndrome. *European Journal of Human Genetics* **18** 8–14. (https://doi.org/10.1038/ejhg.2009.106)

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE & Jaenisch R 2007 In-vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448** 318–324. (https://doi.org/10.1038/nature05944)

Woodfine K, Huddleston JE & Murrell A 2011 Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenetics and Chromatin* **4** 1. (https://doi.org/10.1186/1756-8935-4-1)

Xu Y, Goodyer CG, Deal C & Polychronakos C 1993 Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochemical and Biophysical Research Communications* **197** 747–754. (https://doi.org/10.1006/bbrc.1993.2542)

Yoder JA, Walsh CP & Bestor TH 1997 Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* **13** 335–340. (https://doi.org/10.1016/S0168-9525(97)01181-5)

Younis S, Schonke M, Massart J, Hjorteberg R, Sundstrom E, Gustafson U, Bjornholm M, Krook A, Frystyk J, Zierath JR *et al.* 2018 The ZBED6-IGF2 axis has a major axis has a major effect on growth of skeletal muscle and internal organs in placental mammals. *PNAS* **115** 2048–2057. (https://doi.org/10.1073/pnas.1719278115)

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B *et al.* 2012 Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149** 1368–1380. (https://doi.org/10.1016/j.cell.2012.04.027)

Yu M, Han D, Hon GC & He C 2018 Tet-assisted bisulfite sequencing (TAB-seq). *Methods in Molecular Biology* **1708** 645–663. (https://doi.org/10.1007/978-1-4939-7481-8_33)

Zhang S, Kubota C, Yang L, Zhang Y, Page R, O'Neill M, Yang X & Tian XC 2004 Genomic imprinting of H19 in naturally reproduced and cloned cattle. *Biology of Reproduction* **71** 1540–1544. (https://doi.org/10.1095/biolreprod.104.031807)

Zhang T, Termanis A, Ozkan B, Bao XX, Culley J, de Lima Alves F, Rappsilber J, Ramsahoye B & Stancheva I 2016 G9a/GLP complex maintains imprinted DNA methylation in embryonic stem cells. *Cell Reports* **15** 77–85. (https://doi.org/10.1016/j.celrep.2016.03.007)

Received 26January2018
First decision 5March2018
Revised manuscript received 5April2018
Accepted 8May2018

General Discussion

DNA methylation involves the transfer of a methyl group from 5-adenosylmethionine to 5'carbon of a cytosine residue, converting it to 5- methylcytosine (Gibney and Nolan, 2010) and it is essential for differentiation and normal mammalian development, due to its effects on long term stable gene repression (Cedar & Bergman, 2009). The work presented in this thesis, focuses on the effects of loss of key DNA methyltransferases in human model and mouse model systems to identify and validate genes classes sensitive to a transient and stable loss of methylation in an attempt to shed some light on their reprogramming capabilities and point towards the mechanisms, which allow this to happen.

Identification of enriched human gene classes that are particulary sensitive to a stable loss of maintenance methylation and those responsive to s transient loss of methylation

A chronic depletion of DNMT1 in the hTERT cell lines caused by either pharmacological treatment with Aza, transiently with siRNA or through a stable knockdown with shRNA caused hypomethylation at defined enriched gene classes, and interestingly not all at the same gene classes. The pharmacological DNMT1 inhibitor Aza in the fibroblast cell line was shown to cause hypomethylation of the histone genes, and genes encoding proteins, which become phosphorylated or acetylated (Paper III), whilst siRNA treatment resulted in hypomethylation of gene classes involved in the sensory perception of smell, metabolic processes and phosphorylation (Paper III). Interestingly, long-term KD by shRNA instead led to cell lines with hypomethylation concentrated atthe PCDH locus, fat/body mass genes, the CTA, and the Olfactory genes (Paper II). These enriched hypomethylated gene classes are described further below;

PCDH

The enriched PCDH locus (A & B) identified in Paper II for example is heterochromatic by nature and exhibited a persistent loss of methylation. We identified this region to be polycomb repressed in fibroblasts, where a transcriptionally repressive state has been established through interactions with H3K27me3 (van Kruijsbergen et al., 2015). A number of genome-wide studies in mouse ESCs established that polycomb marking and DNA methylation appeared in large part to cover different areas of the genome, such that genes were either marked by

H3K27me3 *or* by DNA methylation, and not usually by both, suggesting that the two systems of repression were mutually exclusive (Otani et al., 2013).

There are 70 members of the PCDH family currently identified, and segregated into two broad groups- clustered, and non-clustered. PCDHα, -β and -γ are clustered sequentially on Chr 5q31 in human (Hulpiau and van Roy, 2011) as shown in Figure 1. *PCDH* forms a gene cluster similar to that of the immunoglobulins and T-cell receptors. *PCDH* genes are highly expressed during neural development, as such as the brain matures the expression level of PCDH falls (Hulpiau and van Roy, 2011). Their role is of a transmembrane protein involved in calcium ion binding, synaptic formation, axonal projection and neuronal survival. These three clusters have large variable (V) exons, each of which encodes an extracellular domain, as well as a transmembrane domain and a short cytoplasmic region. Both α and γ cluster variable exons are spliced to a set of constant (C) exons (red shading below) which encode cluster-specific additional intracytoplasmic domains (ICD). The PCDHγ cluster genes however do not share this property with the α and β clusters, and so lack the additional ICD. Somewhat confusingly, there are also C-type variable exons, which have a slightly different function than the standard variable exons (yellow shading below). Functionally the PCDHs are involved in neural circuit formation and have been found to be expressed in the neurons and glia as cell-cell adhesion proteins (Takeichi, 2007).

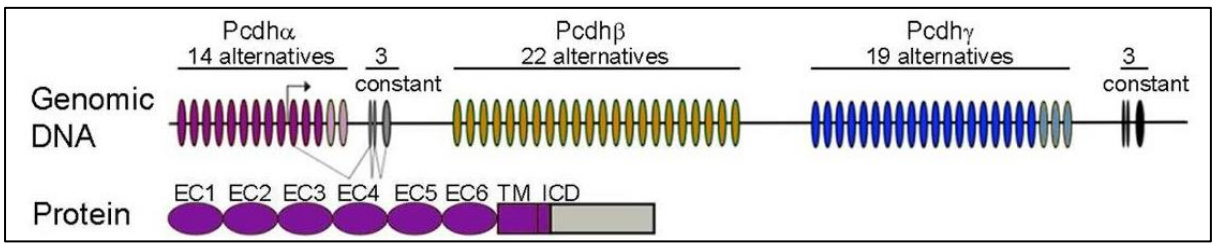


Figure 1- Schematic of the clustered PCDH family. The protocadherin locus has three different clusters of possible variable exons adjacent to each other**.** The two PCDHα and PCDHγ clusters have three constant exons, responsible for encoding a cluster specific intracellular domain which is combined through mRNA splicing of the two clusters *(*Image reused from Rubinstein et al., 2015).

Interest in protocadherin structure and function has grown greatly during recent years, as it has emerged that they play a key role in specifying neural cell shape and identity. Knockout studies conducted in the laboratories of Tom Maniatis and Takeshi Yagi have shown that mice lacking the protocadherins either die in utero (C-class variable exon deletion) or survive to birth but

have severe neurological problems including gait and memory (Chen et al., 2012). The presence of a different "bar-code" of PCDH proteins on adjacent neural cells allows the neurites to distinguish self-self from self - non-self-contacts and so form functional interactions with its neighbours to build the neural net (Rubinstein et al., 2015).Only 1-3 PCDH promoters in any cluster are being expressed, while the other 65+ are inactive. Crucially with respect to our studies, the inactive promoters are methylated, and Yagi and others have shown that demethylation using Aza or cells with mutations in the methyltransferase genes can cause re-activation of the silenced alleles (Morishita and Yagi, 2007).

The shRNA treated cell lines showed significant DNA hypomethylation across the extended locus, in particular at the upstream variable exons with little difference seen in the constant region. Maintaining greater diversity of specific mRNAs and PCDH proteins would require repression of all other non-transcribing PCDHs than those being actively expressed, as such this is an important event in early mammalian development. The Pcdh-γ proteins have been found to be expressed embryonically, (mice null for Pcdh-γ die neonatally (Wang et al., 2002)) when neural development is very active.Pcdh-γ proteins have also been found to be expressed postnatally in the viable embryo until the brain has matured (Frank et al., 2005).

Repression of Pcdh genes can be caused epigenetically through silencing by DNA methylation. This was shown in two mouse cells lines, where Kawaguchi et al (2008) observed that methylation of *Pcdh*-α promoters and the 5'region of each variable exon was correlated with a lower expression level (Kawaguchi et al., 2008). The diversity of the clustered PCDH alleles therefore relies upon the epigenetic regulation of promoter choice and the variety of alternative transcripts (Toyoda et al., 2014). The *de novo* methyltransferase 3B is responsible for the de novo methylation of some of these promoters in early development ultimately regulating the expression of the variety of potential *Pcdh* isoforms in individual cells (Toyoda et al., 2014).

Fat Body Mass Genes

In Paper II we were able to show that upon comparison to normal chromatin patterns in hTERT cell lines, the clonal stable knockdown cell lines exhibited hypomethylation at regions known to be marked by polycomb; for example, the hypomethylated *LEP*, *GHSR* and *APOC* gene classes are heavily marked by polycomb in normal fibroblasts. We showed that the shRNA lines undergo gradual remethylation at most loci throughout the genome, except for the

polycomb-marked regions. One reason these could be protected from a remethylation event might be the multidomain chromosomal protein FBXL10, and its localisation at polycomb repressive marks.

Furthermore, in our study we could show that inhibition of EZH2, the main enzyme responsible for depositing H3K27me3 marks, was sufficient to boost transcription from some of the polycomb-repressed and demethylated genes, suggesting that the histone marking is sufficient in itself to keep these genes repressed in the absence of DNA methylation. This both supports the idea that polycomb and DNA methylation can act independently to repress and explains in part why we see so little gene reactivation in the stable cell lines.

Cancer/Testis Antigen Genes

In addition to the *PCDH* and fat/bodymass genes discussed above we also identified the cancer/testis antigen (CTA) genes to be targets for demethylation and transcriptional depression. Demethylation occurred to a greater extent in the gene bodies in the DNMT1-depleted hTERT (Paper II), and in both the promoters and gene bodies in cells treated with siRNA (Paper III). In particular the *TSPY* and *MAGE* genes were shown to be enriched, some of over 200 *CTA* genes, which are characterised by being expressed at the protein level in the human germ line and also in tumours, but nowhere else, and by having poorly understood or poorly-conserved function (Simpson et al., 2005; Almeida et al., 2009). The *CTA* genes can be classified into two groups- X chromosome *CTA* genes and autosomal *CTA*. Single copy *CTA* genes such as *BAGE* and *HAGE* (Gjerstorff et al., 2015) are expressed meiotically and post-meiotically in male germ cells, with a characteristically varied expression level in cancer cells, even in the same tumour (Gjerstorff and Ditzel, 2012).

Historically, the identification of tumour antigens with cytolytic capacities was conceived to be an avenue for developing immunotherapy approaches for cancer. Ideally the antigen would be consistently and stably expressed by a tumour, not expressed by somatic tissue whilst being essential for the cancer cell (Simpson et al., 2005). MAGE1 (now known as melanoma-associated antigen A1) was the first such antigen to be identified in a patient with melanoma, followed by MAGE2 and MAGE3. MAGE1 was also identified in further cancer cell types, but interestingly not in somatic tissue except testis. The chromosomal location of the *MAGE* genes showed 12 further closely related genes, with additional *MAGE* clusters MAGEB, and C downstream. MAGEA, -B and -C clusters were all identified to be expressed in cancers and

testis, whilst additional clusters MAGED-MAGEL were expressed in non-cancerous normal tissue (Chomez et al., 2001). Many *CTA* genes, namely those localised to the X chromosome are co-expressed in tumours, with multiple *CTA* genes being expressed.

We showed using both small molecule and RNA inhibitor approaches that the *CTA* genes were affected by the loss of maintenance methyltransferase activity in our normal non-cancerous human cell line. Promoter demethylation is considered to be the main driver in the expression of these genes, particularly those located on the X chromosome (Simpson et al., 2005). Epigenetic reprogramming is central to activation of the *CTA* genes, previously shown using the same DNMT inhibitor Aza (there termed 5-aza-dC) in cancerous cell lines (Sigalotti et al., 2002; Karpf et al., 2004). Such reprogramming occurs in two phases where both DNA methylation and chromatin undergo alterations- initially in gametogenesis, and then early embryogenesis. The *CTA* genes, particularly those found on the X chromosome have methylated CGI in normal somatic tissue that are subsequently demethylated during development in spermatogenesis. Following this trend, the densely methylated CGI of *CTA* promoters can be demethylated experimentally causing cells that do not typically express *CTA* genes to express them. In tumour cells, global hypomethylation events are thought to echo this same effect. One important line of research is to identify whether the expression of *CTA* genes is contributory to tumorigenesis or is a functionally null by-product of the chromatin reconfiguration of this gene class. However, quite apart from their normal function (if any), the *CTA* have proven clinical utility as targets for anti-tumour immunotherapy.

Ultimately, treatment with e.g. 5-aza dC can increase the expression of *CTA* genes, making them more visible to the immune system. There is a continuing effort to identify novel targets for immunotherapy and in particular the *CTA* genes as they characterised in the literature to have to restricted expression in somatic tissue, in addition to their depression in solid epithelial cancers (Thomas et al., 2018).

The MAGE proteins have a common central MAGE homology domain (MHD) identified to be important for protein-protein interaction (Taniura et al., 2005). Of such proteins, the SKI-interacting protein (SKIP) binds to MAGEA1 when there is an appropriate carboxyl terminus available. SKIP works by connecting DNA binding proteins to additional proteins that have a role in the activation or conversely the repression of transcription ultimately implicating a range of signalling pathways such as Vitamin D and NOTCH1 (Simpson et al., 2005). MAGEA1 has been found to disrupt SKIP-mediated NOTCH1 through binding to SKIP and subsequently

recruiting HDAC. In this functionality, MAGEA1 acts as a transcriptional repressor (Laduron et al., 2004). However, the role of MAGEA1 in the germline has not been fully elucidated, it has been postulated that MAGEA1 could repress the expression of genes that are crucially important to normal development and differentiation. A similar functionality of the MAGE genes in cancer cells would be a contributory factor to tumorigenesis and a malignant phenotype.

It is important to identify the *CTA* genes sensitive to depression in cancer cells. Koslowski et al (2004) identified that most of *CTA* genes were activated in gametogenesis (namely premeiotic stages) whilst the germline genes are expressed in gametogenesis at the late meiotic and postmeiotic stages (Koslowski et al., 2004). Koslowski tested the induction of *CTA* genes under the influence of DNMT1 inhibitor 5-aza dC in somatic cells, where similar to the results in Paper II using shRNA targeted to DNMT1 that showed that 80% of *CTA* genes were expressed, unlike those of the germline genes at only 18%. Interestingly we did not investigate the impact on *CTA* genes with depleted *de novo* methyltransferase 3B, but Koslowski did not find any induced expression of the *CTA* genes in human colorectal cancer cell lines collected from individuals. Simpson et al (2005) was able to confirm that genomic hypomethylation was essential and adequate for the expression of the *CTA* genes; in particular, *DNMT1* combined with the activity of *DNMT3B* is implicated in the transcriptional repression of this gene class in somatic cells.

Olfactory genes are also enriched targets of hypomethylation

Another gene class identified as having a significant loss of DNA methylation at promoters and gene bodies in a DNMT1 deficient cell system as represented in Paper II, and in Paper III are the olfactory receptor genes (*OR*). Whilst being largely inert as a group both shRNA and siRNA targeting DNMT1 showed similar effects in this gene class in the hTERT model system. Treatment with Aza (5-aza dC) however conversely also caused hypermethylation in some *OR* genes in this system when compared to the untreated hTERTs.

The *OR* genes are the largest gene family in the mammalian genome, part of the G-protein coupled receptor (GPCR) hyperfamily (Glusman et al., 2001). There are two different types of *OR*- Class I and Class II- in mammals; both classes are part of the GPCR gene family. Over 300 *OR* genes have been identified and are frequently found to be located in clusters making

up 0.1% of the human genome. This gene family are expressed, like the imprinted genes (Paper I and Paper IV) in a monoallelic fashion, established early in development following the blastula stage (Takagi and Oshimura, 1974).

The glomeruli that form part of the neural circuit is responsible for the olfactory input to the brain, and this is in turn innervated by olfactory sensory neurons. The neurons in the same glomerulus express the same *OR*. Odour discrimination is more effective when each olfactory sensory neuron expresses one allele of an individual *OR* gene. To promote the expression of a single *OR* gene the mechanism of action leading to their transcription operates through homeobox transcription factors e.g. Lhx2, Emx2 and Ebf (McClintock, 2010). A popular hypothesis suggests that epigenetic control of *OR* gene transcription is important for the repressive events necessary to ensure that only a single type of *OR* is produced.

The *OR* genes that show substantial losses of methylation as we have shown in Paper II and Paper III and are made up of a single exon which follow a few noncoding exons, and the TSS of which can be 100s-1000s bps upstream. Methylation plays a functional role in the control and expression of the *OR* genes, this and their heterochromatic state supports the inertness of this gene class. The silencing of *OR* gene expression also complements the preferential single *OR* gene expression required to silence all of the other *OR* gene classes. The polycomb complex and chromatin remodelling elements are abundantly found in immature olfactory sensory neurons where the ultimate *OR* gene choice is confirmed. The *OR* gene cluster has additional repressive histone marks e.g. H3K9me2, in addition to low levels of histone acetylation, both of which are associated with poor levels of transcription.

In summary we showed that the siRNA treatment led to widespread loss of methylation across the genome, as did Aza, but the unique targets of Aza likely are a result of 1) the drug affecting DNMT3A and DNMT3B as well as DNMT1, the only target for the siRNA/shRNA; 2) other side-effects such as the reported effect on G9A/GLP histone methyltransferase. On the other hand, the more limited pattern of hypomethylation seen in the shRNA lines is almost certainly due to remethylation of the genome over time as the cell lines are grown out and expanded. Here the profile of hypomethylation reflects instead regions which are resistant to remethylation.

Methyltransferase deficient human model systems provide evidence for gains in methylation upon treatment with 5-aza-2'dC offering a different response than treatment with DNMT1 targeting siRNA

As previously shown by this lab group, a chronic depletion of DNMT1 in particular through transient siRNA mediated knockdown, revealed global losses of methylation across multiple candidate gene targets, in particular those with a fundamental role in early development (Rutledge et al., 2014). However, classical approaches were not able to reveal and identify enriched gene classes that experienced perturbations of methylation out with the candidate genes. With the introduction of microarrays such as the Illumina BeadChip 450k array and subsequent gene ontology, analyses of alternative DNA methylation profiles were revealed on a genome wide basis (Figure 2).

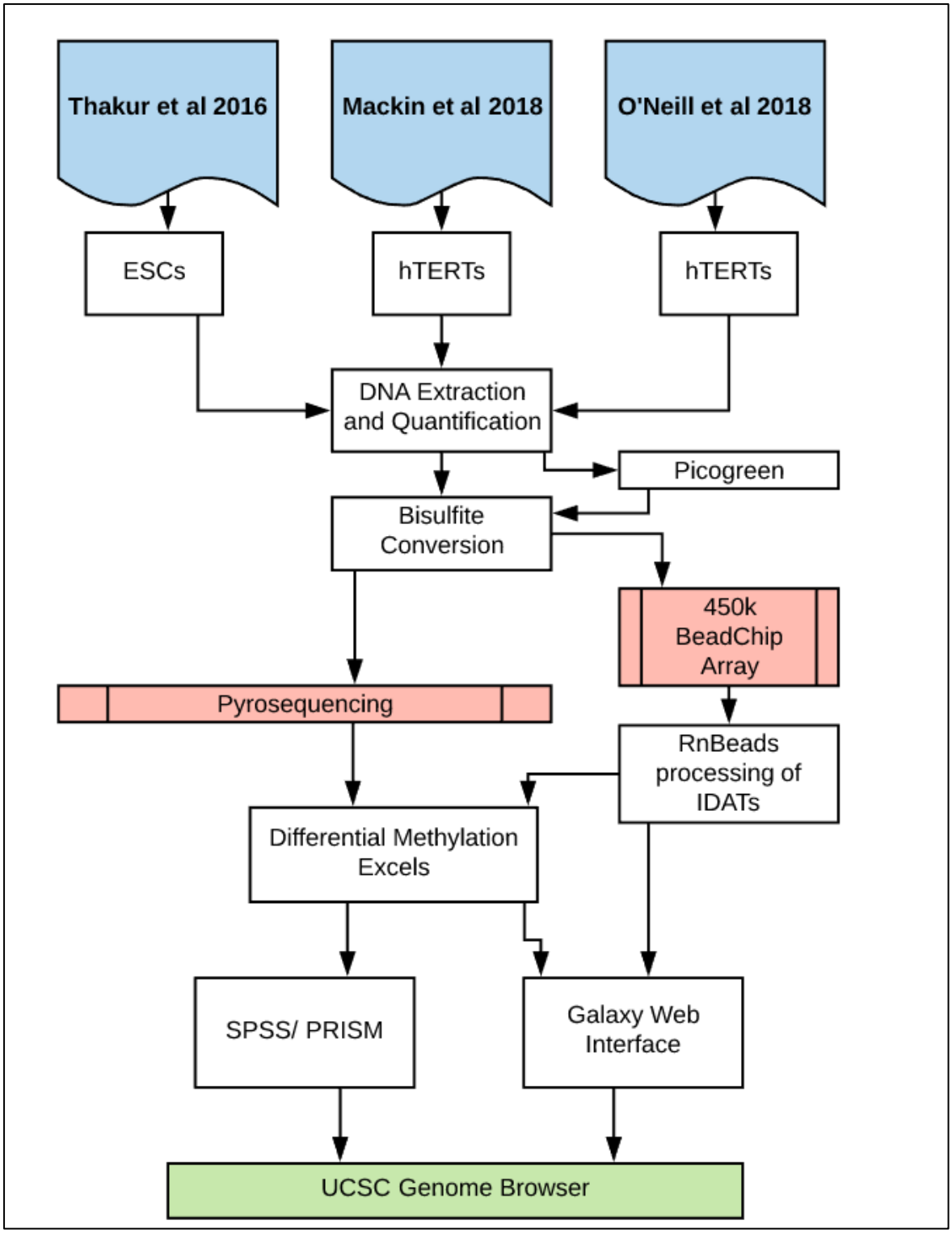


Figure 2- Pipeline Methylation Analysis using both wet-lab and bioinformatic techniques. The methodology presented in Papers I, II and III in this thesis all utilise part or all of this methylation pipeline. The mESCs in Thakur et al. (2016) are unsuitable for use in 450k BeadChip arrays but in-house selective CG methylation profiling was carried out on a smaller scale with the use of pyrosequencing. Despite the human and mouse samples segregating with respect to their sequencing platform, the common outputs could be presented in multiple common bioinformatic interfaces for interpretation.

The application of such next-generation approaches revealed that DNMT1-targeting siRNA and shRNA-treated hTERTS cells do lose methylation globally as predicted, and for the siRNA-treated cells particularly at gene bodies; but uniquely only a stably mediated knockdown of DNMT1 using shRNA in hTERTS showed gains in methylation (Paper II). This pattern of hypermethylation was echoed in the pharmacologically depleted hTERT cell lines treated with Aza (Paper III) despite previous evidence concentrating on the widespread demethylating effects of Aza by us and others. The phenomenon of hypermethylation, namely of CG dense promoter regions has been frequently associated with the silencing of tumour suppressor genes in tumorigenesis or the aberrant methylation of genes fundamental to the regulation of DNA repair (Ehrlich, 2002; Timp and Feinberg, 2013). However, we observed hypermethylation of multiple gene classes in non-tumorigenic *de novo* methyltransferase depleted cells lines- that is in human normosomic lung fibroblast cell lines and mouse ESCs. It could be assumed that such hypermethylation is a consequence of survival in a cultured micro-environment against hypomethylation.

To help solve this riddle, and to establish if short-term depletion of DNMT1 could also lead to gains in methylation, we examined the effects of transient or acute DNMT1 depletion in our hTERT cells by exposing the cells to siRNA against DNMT1 for 48-72hrs, then analysing the methylation state using the 450k array as before (Paper II and III). In these cells we saw little or no gains in methylation (Figure 3), indicating that gains are not a short-term response to depletion.

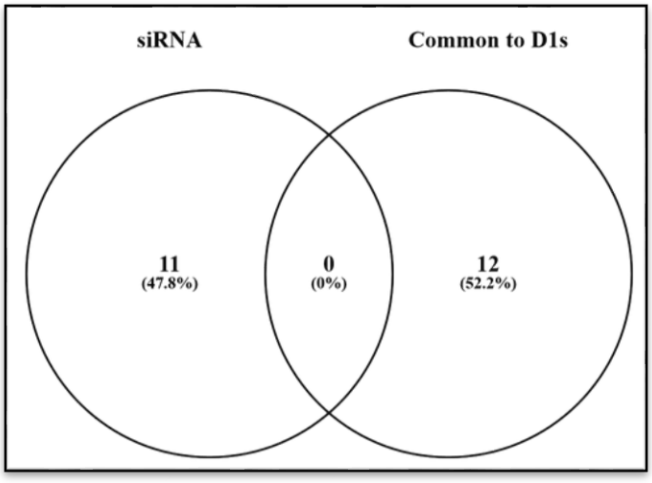


Figure 3: Commonly hypermethylated genes. Genes gaining methylation that are common to all of the DNMT shRNA lines (d8, d10 and d16) from Paper II compared to the common hypermethylated genes of all the DNMT1 siRNA treated hTERTs from Paper III.

Instead, the cells showed more widespread demethylation across the genome, including 1) in regions where little demethylation is seen in the stable (shRNA) lines and 2) regions where the shRNA lines show gains. These results were very telling, since this indicates that the most likely sequence of events in the derivation of the stable shRNA lines is that the cells initially lose methylation everywhere, but during selection for the stable integrants and outgrowth, many areas become remethylated back to a normal level of methylation, and so show no significant difference from WT in the shRNA lines when analysed. The "hit-and-run" experiment described in Paper II provides solid support for this idea, since we saw initial loss of methylation followed by slow but steady remethylation at many genes. Regions which show gains in the shRNA lines are therefore likely to be the result of "overshoot" where de novo methylation during the outgrowth phase does not stop when the normal level of methylation is reached, and instead accumulates to a higher level, a good example being the *UGT1* gene cluster. Again, we could show experimentally that a gene at the *UGT1* locus initially lost methylation, then regained normal levels, before overshooting to a hypermethylated state in the "hit-and-run" experiment. We went further and looked to see what properties such regions have and found that the regions becoming hypermethylated are more likely to be poised promoters (Figure 4). We summarised our understanding of these processes in the model at the end of Paper II. Given these insights, it is tempting to speculate that the cells treated with the small molecule inhibitor may also be undergoing such a process, with an initial demethylation followed by a rapid overshoot at some genes. However, in order to prove this we would need to set up a repeat experiment where we expose cells to the agent then take samples at timed intervals following exposure. Examination of the methylation levels at genes which were identified as hypermethylated in the initial study could then establish, using this timecourse, if they are initially demethylated before subsequent recovery and then overshoot.

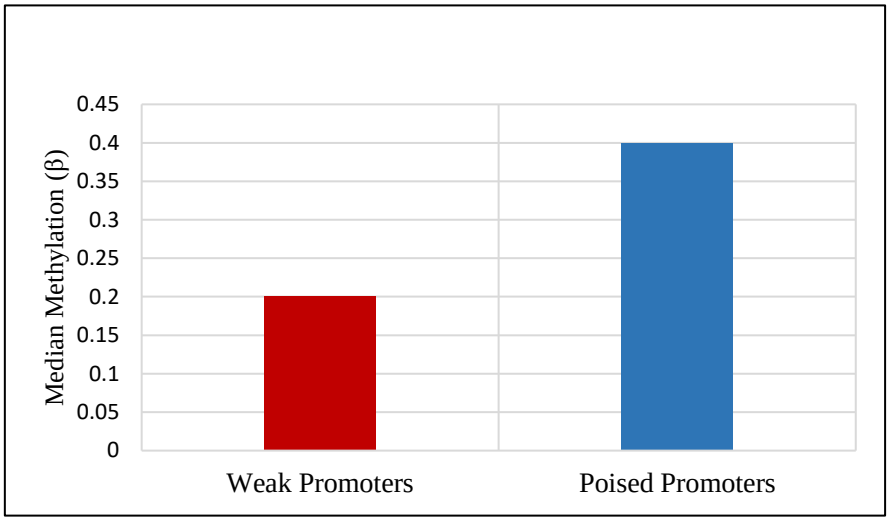


Figure 4- Median methylation values for weak and poised promoter categories, and comparing shRNA lines to WT. Cohen's D test revealed that the median methylation value (β) value for poised promoters is 0.2β higher than that of the weak promoters when shRNA lines were compared to WT.

Gains in methylation at the gene classes identified in Paper III in Aza treated hTERTs e.g. GPCRs could be significant in the treatment avenues of AML sufferers. GPCR methylation gains could contribute to the relapse in AML sufferers as key TSGs could be silenced. GPCRs are known to act as oncogenes and tumour suppressors through their regulation of oncogenic signalling networks. GPCRs may also be a suitable therapeutic target- *MAS1* was hypermethylated, on oncogene involved in leukaemia, and *USP33*/*PARD3* are both TSGs in lung cancer, and lung squamous cell carcinoma. However recent literature re-addresses the role DNA methylation plays in sensitising or priming e.g. AML sufferers to subsequent therapies following Aza treatment. As an alternative to hTERTs, Vispe (2015) used leukaemic cell line KG1 treated with Aza to ascertain the global methylation changes and concluded that DNA damage mechanisms are distinct from DNA methylation in the treated cells (Vispe et al., 2015). Furthermore, 450k analysis of the treated KG1 cells also exhibited distinct hypermethylation as we did in Paper II and III supporting the role of Aza in DNA hypermethylation.

Despite the observation of hypermethylation in DNMT1 depleted cell lines- both normosomic and cancerous -through treatment with Aza, the success of such treatment for sensitising patients to downstream therapeutic treatment might not lie in the effects of methylation of tumour suppressor or other such tumorigenic genes. Recently, two research groups have been able to refine the response of Aza and its anti-tumour activity (Roulois et al., 2015; Chiappinelli et al., 2015). The research from these two groups successfully elucidated that a low dose

treatment of a demethylating agent such as Aza (almost 10-fold lower than we used) primes the immune system by stimulating the activation of ERVs and launching an anti-viral response.

This viral response might be implicated in the induction of tumour antigen expression and the alteration of cell signalling pathways and/or apoptotic thresholds. However, the exact downstream pathway of how Aza treatment impacts myeloid malignancy is yet to be fully elucidated. In our study we used one higher dose Aza limited to only one cell line which may have masked such effects and offered a different enriched gene ontology landscape. Merit still exists however in the approach we have undertaken as previous work from our lab (Rutledge et al., 2014) using this dosage highlighted similar gene classes. Further studies in our hTERT system would be warranted, using different doses and timings for Aza treatment as indicated above and a more comprehensive sequencing approach such as WGBS, to see if this viral immune response occurs in normal cells as well as cancer cell lines.

Alternatively, treatment with Aza in combination with a HDAC inhibitor such as valproic acid has also shown success. Histone involvement is not to be overlooked in the phenomenon of DNA hypermethylation presented in Paper II and III. Histones were identified as an enriched hypermethylated gene class in both papers.

UGT Genes were also observed to experience gains in methylation in hTERTs

The issue remains as to why there are gene classes more susceptible to hypermethylation than others. One theory is, as indicated above, that such genes exhibiting regions of hypermethylation e.g. *UGT1A* (Paper II) have poised promoters which is rendering them susceptible. Such promoters are frequently found in developmentally important genes (e.g. those on the X chromosome and in the germ line) and frequently have chromatin configured in a bivalent manner with high levels of repressive H3K4me3 and/or H3K27me3.

The *UGT1A* genes, though not normally considered bivalent and despite their already high methylation levels, had further increased methylation in the stable cell lines, and showed small but consistent decreases on the HT12 array. Perhaps marks associated with recent inactivation of the *UGT1A* cluster in fibroblasts during adaptation to cell culture led to increased *de novo* activity, and are therefore showing the highest gains in methylation.

The *UGTs* (UDP-glucuronosyltransferase) as a gene class belong to a superfamily of many isoforms who all share a role in metabolic activity, responsible for 35% of all phase II metabolic

reactions (Yang et al., 2017) and are as such a crucial gene class in pharmacogenetics (Barbarino et al., 2014). The *UGT1A* locus encodes nine isoforms through an exon sharing mechanism (Figure 5). This involves the transcripts of the first exon cassette being spliced to exons 2-5 consequentially expressing individual *UGT* isoforms in a tissue and age specific manner (Habano et al., 2015). The variety of *UGT* isoforms also has a pivotal role in the pathogenesis of neonatal jaundice, Gilbert's syndrome and Criglerr-Najjar syndrome which are all closely related to a deficiency of *UGT1A1*, typically expressed in the liver (Bartlett and Gourley, 2011).

These properties make the transcription of the *UGT* isoforms important for drug efficacy and determine adverse effects to drug treatment. It has been postulated that epigenetic mechanisms e.g. DNA methylation have an effect upon the regulation of this locus and as such the response to drug therapy as has been shown previously in colorectal cancer cells treated with alternative DMEs (drug metabolising enzymes) (Habano et al., 2009; Belanger et al., 2010). As a gene class the UGTs have been documented in cancerous cell lines to be hypermethylated.

Irinotecan is used clinically to treat colorectal cancer, particularly in metastatic cases (Cunningham et al., 2001; Gagnon et al., 2006), the active metabolite of this treatment is SNF2 ultimately mediated through *UGT* (Gagne et al., 2002). The *UGT1A* gene encodes an enzyme which is responsible for catalysing conjugation reactions with glucuronic acid, and changes to this gene arising through polymorphisms or exon sharing are important for affecting toxicity and clinical outcome in patients treated with Irinotecan. In this instance Gagne et al. (2002) were able to show that DNA methylation is responsible for the repression of UGT1A1 expression in colon cancer cells. The downstream effect of hypermethylation of this gene has been associated with levels of the anticancer agent SN-38 within the tumorigenic cells, inactivating it which can ultimately influence clinical response to Irinotecan treatment (Gagne et al., 2002).

It may well be that the hypermethylation in the Aza-treated samples reflects a similar "bounce" or recovery following demethylation, in this case due to Aza, but in Paper II from the outgrowth of the shRNA-containing clones. We know that the hTERT cells (used in both studies) contain at least the DNMT3B *de novo* methyltransferase so can add methyl groups to naked DNA. It is unclear whether DNMT3A plays a role too: some DNMT3A transcript was detected by array in the cells, but we have not tested for functionality by DNMT3A-specific knockdown.

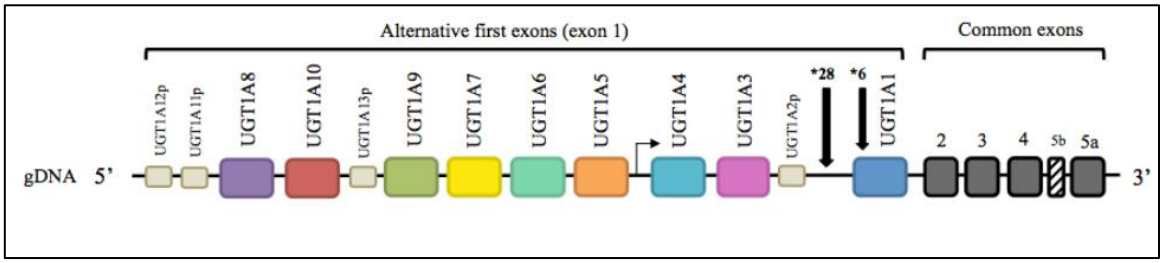


Figure 5: Schematic of the human UGT1A locus (not to scale). This gene locus is responsible for the transcription of nine UGT1 enzymes. The locus spanning a region of 200kb contains numerous alternative first exons with their own promoters to make up exon 1. The individual exons for the respective isoforms are combined with the common exons (2-5 shown in grey) by splicing out the intervening sequences. The alternatively spliced isoforms of UGT1A are known as UGT1As_i2 and are created when exon 5b is used, either with or instead of exon 5a. Image reused from Barbarino et al., 2014.

A number of known and putative imprinted gene loci mESCs are capable of recovery where methylation has been lost

We have identified a model system depleted in the *de novo* methyltransferases that experiences methylation levels that are higher than WT values. A recovery in methylation to WT levels is observed upon rescue with a common isoform of the de novo methyltransferase DNMT3A.

In the same mESCs models the *de novo* methyltransferases Dnmt3a/3b have a role in the maintenance of methylation in the mESCs

Paper 1 showed that a loss of DNMT1, and a loss of DNMT3A/B had equivalent effects on methylation levels at delineated imprinted regions in mouse ESCs with both methyltransferases exhibiting some activity of methylation maintenance at the imprints assayed, but that only rescue with DNMT3A2 provided restoration of methylation at selected imprints. Whilst we hoped to identify the correct isoform of DNMT3B responsible for the recovery of methylation in imprints such as *Rasgfr1* and *Igf2r* we were not able to delineate which domain of the e.g. DNMT3A2 was most active in this process. We postulated that perhaps the PWWP domain was most actively involved due to its known association with H3K36me3 (Ge et al., 2004). Since the completion of this thesis further work in the lab has attempted to investigate this relationship by rescuing the 3ab KO cells with five different point mutations (as shown in Table 5) to determine which of the DNMT3A domains are crucial for the rescue of DNA methylation.

Plasmids with Point Mutations
Glutamine 304 to Alanine
Aspartic Acid 308 to Alanine
Methionine 325 to Tryptophan
Glutamic Acid 533 to Alanine
Arginine 580 to Alanine

Table 1: Mutated plasmids received from a collaborating lab with point mutations in DNMT3A. The point mutations in *de novo* methyltransferase 3A will allow us to determine which of its multiple domains are important for the recovery of DNA methylation if in the methylation depleted system 3ab KO cells (Paper 1).

The imprints investigated in Paper I were limited to a mouse ESCs model system, however in the hTERT-1604 treated with DNMT1 targeting shRNA cell lines used in Paper II it has since been identified in the lab that there is a loss of imprinting at loci including *H19*, *SNRPN*, *KvDMR*, and *GRB10*. These have been identified through mining of the user defined tracks for these human cell lines at the co-ordinates delineated for the gDMRS of these imprinted genes (Court et al., 2014; Woodfine et al., 2011) using both UCSC genome browser and the Galaxy Bioinformatics user interface. This approach could allow for the identification of clinically relevant imprinted gDMRs that are sensitive to a transient loss of DNA methylation in human differentiated cells. Preliminary data from the lab has already shown that there are novel gains in methylation at specific regulatory regions of imprinted loci in the d10 and d16 cell lines (Paper 1 when mapped against published co-ordinates for imprinted loci in Court et al., 2014) e.g. a gain of methylation has been observed at a secondary DMR at the upstream promoter of the locus involved in Silver-Russell syndrome which was particularly sensitive to the loss of DNA methylation. Developing new pyrosequencing assays for this novel secondary DMR and others like it at additional imprint DMRs could further inform diagnosis for this and other imprinting disorders, as existing tests may not be entirely accurate. The identification of such epigenomic signatures could be used to develop early intervention strategies for affected children.

It would be fitting to investigate why the imprints in Paper I do recover by determining which chromatin marks are associated with imprinted regions and whether or not this correlates with the ability of these imprints to recover DNA methylation in a methylation depleted system. Through alteration of these chromatin modifications and/or histone marks through e.g. Epi-CRISPR (Vojta et al., 2016) we could identify if these would prevent recovery in these cell lines, or in fact have no effect at all on recovery. Recovery could be instigated in the Dnmt3a2-rescued cells by their ability to 'remember' that the maternal copy of e.g. *Snrpn* must be methylated. Conversely when Dnmt1 was stably depleted in the mouse ESCs, this may have resulted in the indirect loss of the H3K9me3 repressive histone mark from the known germline imprints such as *Snrpn* and it may be this loss that could be preventing the ESCs from remembering that methylation must be placed on the maternal allele. If this hypothesis about the importance of the presence of histone marks at imprinted loci proves to be correct this mouse model system may help us to understand the conditions which are required for imprint establishment in other mammalian species such as human.

A paper published by the Stancheva group in 2016 (Zhang et al., 2016) aimed to identify a mechanism responsible for the imprints and discover why some imprints are capable of regaining methylation whilst others are unable to; and investigate the histone marks associated with imprinted regions to examine whether or not they were essential for the stable maintenance of genomic imprint in a ESC model system. Zhang et al. (2016) concentrated much of their efforts on the G9a/GLP complex, a knockdown of which reduced DNA methylation at maternally and paternally methylated ICRs in mouse ESCs. This complex was required to stabilise imprinted DNA methylation in the ESCs through DNMT1 recruitment. In turn, this has the effect of counteracting TET dioxygenase-dependent demethylation pathways. Furthermore, the G9a/GLP complex was observed to stabilise imprinting by the recruitment of *de novo* DNA methyltransferases to ICRs of the ESCs, in turn antagonising TET recruitment at these regions.

Conclusion

Through the use of both bioinformatic and wet-lab experimental approaches, gene classes sensitive to a transient loss of methylation have been identified in methyltransferase depleted human hTERTs and mouse ESCs. We have provided evidence for the i) recovery of methylation in putative imprinted regions; ii) maintenance activity of Dnmt3a/b in mouse ESCs iii) identification of the loci most sensitive to a stable loss of maintenance methylation and their interaction with polycomb marks and; iv) the novel discovery of gains in methylation of genes following the 5-aza-dC treatment of hTERTS. It is evident from the data presented in this thesis that DNA methylation plays a clear role in the regulation and control of a range of gene classes in mouse and human systems.

References

Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen., *et al* (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens, *Nucleic Acids Research* 37 (Database issue) D816-9.

Barbarino, J.M., Haidar, C.E., Klein, T.E. and Altman, R.B. (2014) PharmGKB summary: very important pharmacogene information for UGT1A1, *Pharmacogenetics and Genomics* 24 (3): 177-183.

Bartlett, M.G. & Gourley, G.R. (2011) Assessment of UGT polymorphisms and neonatal jaundice, *Semin. Perinatol*. 35 (3): 127-133.

Belanger, A., Tojcic, J., Harvey, M. & Guillemette, C. (2010) Regulation of UGT1A and HNF1 transcription factor gene expression by DNA methylation in colon cancer cells, *BMC Molecular Biology* 11 (9): 1-11.

Boulard, M., Edwards, J.R. & Bestor, T.H. (2015) FBXL10 protects Polycomb-bound genes from hypermethylation, *Nature Genetics* 47 (5): 479-485.

Cedar, H. & Bergman, Y. (2009) Linking DNA methylation and histone modification: patterns and paradigms, *Nature Reviews.Genetics*, 10 (5): 295-304.

Chen, W.V., Alvarez, F.J., Lefebvre, J.L., Friedman, B., Nwakeze, C., Geiman, E., Smith, C., Thu, C.A., Tapia, J.C., Tasic, B., Sanes, J.R. & Maniatis, T. (2012) Functional significance of isoform diversification in the protocadherin gamma gene cluster, *Neuron*, 75 (3): 402-409.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., *et al* (2015) Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA including Endogenous Retroviruses, *Cell*, 162 (5): 974-986.

Chomez, P., De Backer, O., Bertrand, M., De Plaen, E., Boon, T. & Lucas, S. (2001) An overview of the MAGE gene family with the identification of all human members of the family, *Cancer Research*, 61 (14): 5544-5551.

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., *et al* (2014) Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment, *Genome Research*, 24 (4): 554-569.

Cunningham, D., Maroun, J., Vanhoefer, U. & Van Cutsem, E. (2001) Optimizing the use of irinotecan in colorectal cancer, *The Oncologist*, 6 (Suppl 4): 17-23.

Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little, *Oncogene*, 21 (35): 5400-5413.

El Hajj, N., Dittrich, M. & Haaf, T. (2017) Epigenetic dysregulation of protocadherins in human disease, *Seminars in Cell & Developmental Biology*, 69: 172-182.

Frank, M., Ebert, M., Shan, W., Phillips, G.R., Arndt, K., Colman, D.R. & Kemler, R. (2005) Differential expression of individual gamma-protocadherins during mouse brain development, *Molecular and Cellular Neurosciences*, 29 (4): 603-616.

Gagne, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G. & Guillemette, C. (2002) Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38), *Molecular Pharmacology*, 62 (3): 608-617.

Gagnon, J.F., Bernard, O., Villeneuve, L., Tetu, B. & Guillemette, C. (2006) Irinotecan inactivation is modulated by epigenetic silencing of UGT1A1 in colon cancer, *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 12 (6): 1850-1858.

Ge, Y.Z., Pu, M.T., Gowher, H., Wu, H.P., Ding, J.P., Jeltsch, A. & Xu, G.L. (2004) Chromatin targeting of de novo DNA methyltransferases by the PWWP domain, *The Journal of Biological Chemistry*, 279 (24): 25447-25454.

Gibney, E.R. & Nolan, C.M. (2010) Epigenetics and gene expression, *Heredity*, 105 (1): 4-13.

Gjerstorff, M.F., Andersen, M.H. & Ditzel, H.J. (2015) Oncogenic cancer/testis antigens: prime candidates for immunotherapy, *Oncotarget*, 6 (18): 15772-15787.

Gjerstorff, M.F. & Ditzel, H.J. (2012) Limited SP17 expression within tumors diminishes its therapeutic potential, *Tissue Antigens*, 80 (6): 523-527.

Habano, W., Gamo, T., Sugai, T., Otsuka, K., Wakabayashi, G. & Ozawa, S. (2009) CYP1B1 but not CYP1A1 is downregulated by promoter methylation in colorectal cancers, *J Oncology*, 34: 1085-1091.

Habano, W., Kawamura, K., Iizuka, N., Terashima, J., Sugai, T. & Ozawa, S. (2015) Analysis of DNA methylation landscape reveals the roles of DNA methylation in the regulation of drug metabolizing enzymes, *Clinical Epigenetics* (7) 105-015-0136-7.

Hulpiau, P. & van Roy, F. (2011) New insights into the evolution of metazoan cadherins, *Molecular Biology and Evolution*, 28 (1): 647-657.

Isbel, L., Prokopuk, L., Wu, H., Daxinger, L., Oey, H., Spurling, A., Lawther, A.J., Hale, M.W. & Whitelaw, E. (2016) Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse, eLife, (5), 10.7554/eLife.15082.

Karpf, A.R., Lasek, A.W., Ririe, T.O., Hanks, A.N., Grossman, D. & Jones, D.A. (2004) Limited gene activation in tumor and normal epithelial cells treated with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine, *Molecular Pharmacology*, 65 (1): 18-27.

Kawaguchi, M., Toyama, T., Kaneko, R., Hirayama, T., Kawamura, Y. & Yagi, T. (2008) Relationship between DNA methylation states and transcription of individual isoforms encoded by the protocadherin-alpha gene cluster, *The Journal of Biological Chemistry*, 283 (18): 12064-12075.

Koslowski, M., Bell, C., Seitz, G., Lehr, H., Roemer, K., Munteferling, H., *et al*. (2004) Frequent Nonrandom Activation of Germ-Line Genes in Human Cancer, *Cancer Research*, 64: 5988-5993.

Laduron, S., Deplus, R., Zhou, S., Kholmanskikh, O., Godelaine, D., De Smet, C., *et al.* (2004) MAGE-A1 interacts with adaptor SKIP and the deacetylase HDAC1 to repress transcription, *Nucleic Acids Research*, 32 (14): 4340-4350.

McClintock, T.S. (2010) Achieving singularity in mammalian odorant receptor gene choice, *Chem Senses*, 35, (6): 447-457.

Morishita, H. & Yagi, T. (2007) Protocadherin family: diversity, structure, and function, *Current Opinion in Cell Biology*, 19 (5): 584-592.

Otani, J., Kimura, H., Sharif, J., Endo, T.A., Mishima, Y., Kawakami, T., Koseki, H., Shirakawa, M., Suetake, I. & Tajima, S. (2013) Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells, *PLoS One*; 8(12): e82961.

Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., *et al.* (2015) DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts, *Cell*, 162 (5): 961-973.

Rubinstein, R., Thu, C.A., Goodman, K.M., Wolcott, H.N., Bahna, F., Mannepalli, S., *et al.* (2015) Molecular logic of neuronal self-recognition through protocadherin domain interactions, *Cell*, 163 (3): 629-642.

Rutledge, C.E., Thakur, A., O'Neill, K.M., Irwin, R.E., Sato, S., Hata, K. & Walsh, C.P. (2014) Ontogeny, conservation and functional significance of maternally inherited DNA methylation at two classes of non-imprinted genes, *Development* (Cambridge, England) 141 (6):1313-1323.

Satie, A.P., Rajpert-De Meyts, E., Spagnoli, G.C., Henno, S., Olivo, L., Jacobsen, G.K., *et al.* (2002) The cancer-testis gene, NY-ESO-1, is expressed in normal fetal and adult testes and in spermatocytic seminomas and testicular carcinoma in situ, *Laboratory Investigation; a Journal of Technical Methods and Pathology*, 82 (6): 775-780.

Sigalotti, L., Coral, S., Nardi, G., Spessotto, A., Cortini, E., Cattarossi, I., *et al.* (2002) Promoter methylation controls the expression of MAGE2, 3 and 4 genes in human cutaneous melanoma, *J Immunotherapy*, 25 (1): 16-26.

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T. & Old, L.J. (2005) Cancer/testis antigens, gametogenesis and cancer, *Nature Reviews Cancer*, 5 (8): 615-625.

Takeichi, M. (2007) The cadherin superfamily in neuronal connections and interactions, *Nature Reviews Neuroscience*, 8 (1): 11-20.

Taniura, H., Kobayashi, M. & Yoshikawa, K. (2005) Functional domains of necdin for protein-protein interaction, nuclear matrix targeting, and cell growth suppression, *Journal of Cellular Biochemistry*, 94 (4): 804-815.

Thomas, R., Al-Khadairi, G., Roelands, J., Hendrickx, W., Dermime, S., Bedognetti, D. & Decock, J. (2018) NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives, *Frontiers in Immunology*, (9): 947.

Timp, W. & Feinberg, A.P. (2013) Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host, *Nature Reviews Cancer*, 13 (7): 497-510.

Toyoda, S., Kawaguchi, M., Kobayashi, T., Tarusawa, E., Toyama, T., Okano, M., *et al.* (2014) Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity, *Neuron*, 82 (1): 94-108.

van Kruijsbergen, I., Hontelez, S. & Veenstra, G.J. (2015) Recruiting polycomb to chromatin, *The International Journal of Biochemistry & Cell Biology*, 67: 177-187.

Vispe, S., Deroide, A., Davoine, E., Desjobert, C., Lestienne, F., Fournier, L., *et al.* (2015) Consequences of combining siRNA-mediated DNA methyltransferase 1 depletion with 5-aza-2'-deoxycytidine in human leukemic KG1 cells, *Oncotarget*, 6 (17): 15265-15282.

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., *et al.* (2016) Repurposing the CRISPR-Cas9 system for targeted DNA methylation, *Nucleic Acids Research*, 44 (12): 5615-5628.

Wang, X., Weiner, J.A., Levi, S., Craig, A.M., Bradley, A. & Sanes, J.R. (2002) Gamma protocadherins are required for survival of spinal interneurons, *Neuron*, 36 (5): 843-854.

Weiser, T.S., Guo, Z.S., Ohnmacht, G.A., Parkhurst, M.L., Tong-On, P., Marincola, F.M., *et al*. (2001) Sequential 5-Aza-2 deoxycytidine-depsipeptide FR901228 treatment induces apoptosis preferentially in cancer cells and facilitates their recognition by cytolytic T lymphocytes specific for NY-ESO-1, *Journal of Immunotherapy (Hagerstown, Md.: 1997)*, 24 (2): 151-161.

Whitehurst, A.W. (2014) Cause and consequence of cancer/testis antigen activation in cancer, *Annual Review of Pharmacology and Toxicology*, 54: 251-272.

Woodfine, K., Huddleston, J.E. & Murrell, A. (2011) Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue, *Epigenetics & Chromatin*, 4 (1): 1.

Yang, N., Sun, R., Liao, X., Aa, J. & Wang, G. (2017) UDP-glucuronosyltransferases (UGTs) and their related metabolic cross-talk with internal homeostasis: A systematic review of UGT isoforms for precision medicine, *Pharmacological Research*, 121: 169-183.

Zhang, T., Termanis, A., Ozkan, B., Bao, X.X., Culley, J., de Lima Alves, F., *et al*. (2016) G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells, *Cell Reports*, 15 (1): 77-85.

Zhang, T., Haws, P. & Wu, Q. (2004) Multiple variable first exons: a mechanism for cell- and tissue- specific gene regulation, *Genome Research*, 14: 79-84.

Achievements

Published Abstracts

1. **Mackin, S.J.,** Walsh, C.P. (2015) Differential responses of clinically important gene classes to transient loss of DNA methylation in human differentiated cells. Ulster Medical Journal 84(3): 200 2014.

2. **Mackin, S.J.**, O'Neill, K.M., Irwin, R.E., Walsh, C.P. (2016) Identifying clinically relevant imprinted gDMRs sensitive to a transient loss of DNA methylation in human differentiated cells. Ulster Medical Journal 86(1): 55-68.

3. **Mackin, S.J.,** O'Neill, K.M., Walsh, C.P. (Pending Date) Comparison of DNMT1 inhibitors by methylome profiling identifies unique epigenetic signature of Dacogen. Ulster Medical Journal.

4. Thakur, A., **Mackin, S.J.**, R.E. Irwin., O'Neill, K.M., Pollin, G., Walsh, C.P (Pending date) Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2. Ulster Medical Journal.

5. O'Neill, K.M., R.E. Irwin, **Mackin, S.J**., Thakur, A., Thursby, SJ., Bertens, C., Masala, L., Loughery, J., McArt, D., Walsh, C.P (Pending date) Depletion of DNMT1 in differentiated human cells highlights key classes of independent genes. Ulster Medical Journal.

Additional Research Training

22/09/2014 Effective Researcher: Let's Get Started with your Doctorate.

23/09/2014 An Introduction to Strategic Career Planning Workshop.

23/09/2014 Optimising Your Professional Development.

24/09/2014 Planning your Research.

08/10/2014 Workplace Experience/Learning: Presentation of a Research Paper at Journal Club of the Genomic Medicine Research Group.

10/10/2014 Developing Information Skills for Effective Research (Life and Health Sciences).

13/10/2014 Introduction to Refworks v2 Tutorial.

16/10/2014 Workplace Experience/Learning: Introduction to the Sanger Sequencing Protocol and Analysing Sequencing Results from ABI Prism.

16/10/2014 Introduction to Learning and Teaching for Postgraduate Tutors and Demonstrators.

17/11/2014 Workplace Experience/Learning: Introduction to Pyrosequencing and Using Pyromark Q24 Assay Design Software.

12/12/2014 Workplace Experience/Learning: Galaxy 101 Tutorial.

22/12/2014 Completion of the Research Integrity Ulster University Online Course.

01/01/2015 Demonstrator for Medical Cell Biology (BMS104) (January-April 2015).

17/02/2015 Workplace Experience/Learning: Illumina Biobanking Interactive Webinar.

11/03/2015 Workplace Experience/Learning: Health & Safety Induction in Gut and Blood Lab.

23/03/2015 Effective Researcher: Let's Carry on with your Doctorate.

25/03/2015 Strategically Planning your Career.

27/03/2015 Workplace Experience/Learning: Introduction to DAVID Bioinformatic Analysis.

22/04/2015 Workplace Experience/Learning: Research and Consent under the Human Tissue Act Training.

07/05/2015 Work-place experience/learning: Journal Club presentation on the non-germline restoration of genomic imprinting.

08/05/2015 Holding Pyrosequencing Tutorials - for undergraduate students ahead of final year research projects.

20/06/2015 Soapbox Science Volunteer: Women in Science, Belfast.

23/07/2015 Effective Researcher: 2nd Year Doctoral Skills.

18/08/2015 Athena Swan Focus Group meeting on ‘Women in Media.’

04/09/2015 Poster presentation at Irish Society of Human Genetics, DCU, Dublin.

01/10/2015 Demonstrator for Introductory Chemistry (BMS105) (Oct-December 2015).

01/10/2015 Roche LC480 Training workshop.

28/10/2015 Webinar from the UK Data Service 'Data Sources for Research Students.'

18/11/2015 Adding Value to your Doctoral Training.

23/11/2015 Workshop at Newcastle University on Bioconductor/RStudio: Microarray analysis (4 days).

10/12/2015 LC480 Training workshop: Part Two held by Roche at Ulster University.

13/01/2016 Presenting a Bioconductor Workshop Report at the Genomic Medicine Research Group lab meeting.

02/02/2016 Presentation and Focus group leader at a Women in Careers evening at Our Lady's Grammar School. Newry.

05/04/2016 Workshop on R- Advanced Graphics and Programming at Newcastle University (4 days).

21/05/2016 Athena Swan Focus Group Meeting of Female PhD researchers.

27/05/2016 Oral Presentation at "Shaping Healthier Communities" Conference, Ulster University (Coleraine).

03/06/2016 Attendance at Illumina 450K Methylation Workshop, London.

22/06/2016 Attendance at a Presentation led by Exiqon on Exosomal Micro RNAs as Biomarkers for Prostate Cancer

25/06/2016 Step up to Science Demonstrator at Ulster University for local schools.

02/09/2016 Poster Presentation for the "Dolly at 20" Symposium at the Roslin Institute, University of Edinburgh.

09/09/2016 Oral Presentation at Irish Society of Human Genetics Autumn Meeting, City Hospital Belfast.

21/09/2016 Webinar on the Pipeline Analysis of Methylation Array data in a longitudinal study.

22/09/2016 Published Abstract in Ulster Medical Journal - January 2017.

22/09/2016 Published Abstract in Ulster Medical Journal - September 2015.

28/10/2016 An Oral Presentation at 7th C-TRIC Annual Translational Medicine Conference, Londonderry.

09/02/2017 What can you do with a PhD? A Workshop on Planning a Career Outside of Academia Workshop.

27/07/2017 Registered as a STEM ambassador with W5 for Northern Ireland.

Certificates

Certified Training from School of Mathematics and Statistics, University of Newcastle (UK) for Statistical Modelling, Programming with R, Efficient R Programming, Advanced Graphics and Bioconductor.

Successful certified completion of the CPPD module 'Introduction to Statistical Programming' Ulster University (2016).

Conference Presentations

Oral Presentations

- 7th Annual Translational Medicine (TMED7) Conference, L/Derry, Northern Ireland- Big Data and Data Analytics in Precision Medicine (Oct 2015).
- 19th Irish Society of Human Genetics Annual Scientific Meeting, City Hospital Belfast, Northern Ireland- Identifying clinically relevant imprinted gDMRs sensitive to a transient loss of DNA methylation in human differentiated cells (Sept 2016).
- Cancer Translational Research Group (CTRG) Annual Symposium, QUB, Northern Ireland- Does the Acute Myeloid Leukaemia treatment Dacogen act primarily through interfering with the enzyme DNA methyltransferase 1? (Oct 2016).

- Faculty of Life and Health Sciences Postgraduate Research Conference, Ulster University, Coleraine- Shaping Healthier Communities. Identifying known and putative imprints responsive to restoration of methylation in rescued mouse embryonic stem cells, (May 2016).

Poster Presentation

- 18th Irish Society of Human Genetics Annual Scientific Meeting, DCU, Dublin, Ireland. Differential responses of clinically important gene classes to transient loss of DNA methylation in human differentiated cells. **Mackin, S.J.,** O'Neill, K.M., Walsh, C.P, (Sept 2015).

- Coming of Age: The Legacy of Dolly at 20 Symposium, Roslin Institute Edinburgh. Widespread recovery of methylation at gametic imprints in hypomethylated cells following rescue with DNMT3A2. Thakur, A., **Mackin, S.J,** Irwin, R.E., O'Neill, K.M., Pollin, G., Walsh, C.P. (Sept 2016).

- 50th Anniversary European Society of Human Genetics, Copenhagen, Denmark. Widespread recovery of methylation at gametic imprints in hypomethylated mouse stem cells following rescue with DNMT3A2 Thakur, A., **Mackin, S.J,** Irwin, R.E., O'Neill, K.M., Pollin, G., Walsh, C.P. (May 2017).

- 19th Irish Society of Human Genetics Annual Scientific Meeting, Croke Park, Dublin, Ireland. Comparison of DNMT1 inhibitors by methylome profiling identifies unique epigenetic signature of Dacogen. **Mackin, S.J,** Thakur, A., Walsh, C.P (September 2017).

Grants

Training Grant from the Genetics Society (UK) £958.47

Travel Grant from the Genetics Society (UK) £723.58

Travel Grant from the Irish Society of Human Genetics €600.00

Awards

Young Investigator Award for Best Postgraduate Oral at 19th Annual Scientific Meeting of the Irish Society of Human Genetics (2016).