Rockefeller University

[Digital Commons @ RU](#)

2020

Molecular Characterization of Mammalian De Novo DNA Methyltransferase Chromatin Recruitment

Daniel N. Weinberg

MOLECULAR CHARACTERIZATION OF MAMMALIAN DE NOVO DNA

METHYLTRANSFERASE CHROMATIN RECRUITMENT

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Daniel N. Weinberg

June 2020

MOLECULAR CHARACTERIZATION OF MAMMALIAN DE NOVO DNA METHYLTRANSFERASE CHROMATIN RECRUITMENT

Daniel N. Weinberg, Ph.D.

The Rockefeller University 2020

Precise deposition of CpG methylation is critical for mammalian development and tissue homeostasis and is often dysregulated in human diseases. The genomic localization of *de novo* DNA methyltransferases 3A (DNMT3A) and 3B (DNMT3B) is facilitated in part by chromatin "reader" domains to establish DNA methylation patterns genome-wide. Recent work has highlighted a role for the PWWP reader domain in directing recruitment of DNMT3B to actively transcribed gene bodies. However, our understanding of *de novo* DNA methyltransferase chromatin recruitment remains incomplete. Here I demonstrate using a variety of biochemical and cellular techniques that DNMT3A genomic localization is regulated by competing chromatin recruitment pathways. Under steady-state conditions, DNMT3A is targeted to non-coding intergenic regions through PWWP reader domain recognition of H3K36me2, an abundant histone post-translational modification (PTM) in cells. Following depletion of cellular H3K36me2 levels, DNMT3A is re-targeted to actively transcribed gene bodies through PWWP domain interaction with H3K36me3, thereby resembling the genomic localization of DNMT3B. Lastly, loss of PWWP domain reader functionality reveals the presence of a

previously uncharacterized chromatin reader domain within DNMT3A that serves as an alternative targeting mechanism to facilitate *de novo* methylation of Polycomb-regulated regions, including a subset of CpG islands. Perturbations to these recruitment mechanisms may underlie the pathology of developmental overgrowth and undergrowth syndromes associated with distinct germline mutations in *DNMT3A* and may also contribute to the altered DNA methylation landscapes observed in diverse cancers.

ACKNOWLEDGMENTS

I would like to thank my advisor and mentor Dr. David Allis. He has built a wonderful, friendly, and collaborative environment in which to do science that has been key to my growth and success in the lab. I also thank Dr. Elaine Fuchs for being the chair of my faculty advisory committee, as well as committee members Dr. Kivanç Birsoy and Dr. Ping Chi. My committee's feedback has been invaluable for my thesis work. I thank Dr. Yariv Houvras for acting as my external examiner.

I have benefited from interactions with a multitude of mentors and collaborators as a graduate student. Dr. Chao Lu was my rotation mentor, "lab coach", and then close collaborator upon starting his own lab at Columbia University. The bioinformatic expertise of Dr. Jacek Majewski at McGill University along with his graduate students Simon Papillon-Cavanagh and Phillip Rosenbaum and post-doctoral fellow Dr. Haifen Chen was instrumental for both projects described in this thesis. Additional collaborations with Dr. Haitao Li and his student Yuan Yue, Dr. Ben Garcia and his student Dylan Marchione, and Dr. Michael-Christopher Keogh and his technician Matthew Marunde provided valuable complementary expertise for this work. I am very grateful for all of their help; I cannot imagine my time in the lab without them.

I would like to thank all members of the Allis lab who I have come to know and learn from over the past few years: Drs. Bolaji Akala, Anja Armache, Douglas Barrows, Bryce Carey, Lijuan Feng, Leah Gates, Steven Josefowicz, Erica Korb, Benjamin Nacev, Tanya Panchenko, Richard Phillips, Alexey Soshnev, Yadira Soto-Feliciano, Liling Wan; Agata Lemiesz and Benjamin Sabari; Marisa Cerio, David Cheon, Julia

TABLE OF CONTENTS

CHAPTER 1: GENERAL INTRODUCTION

CHAPTER 2: THE HISTONE MARK H3K36ME2 RECRUITS DNMT3A AND SHAPES

THE INTERGENIC DNA METHYLATION LANDSCAPE

CHAPTER 3: DNMT3A BINDS H2AK119UB TO HYPERMETHYATE CpG ISLANDS

UPON LOSS OF PWWP READER FUNCTION

CHAPTER 4: GENERAL DISCUSSION

LIST OF FIGURES

CHAPTER 1: GENERAL INTRODUCTION

CHROMATIN LANDSCAPES

In eukaryotes, genetic information is encoded by DNA and complexed with histone

proteins to comprise the physiologic form of the genome known as chromatin. The basic

repeating unit of this biomolecular assembly is the nucleosome, comprised of two

copies of the core histone proteins H2A, H2B, H3, and H4 that together form an

octamer wrapped by 1.65 super-helical turns of DNA (Luger et al. 1997). Such an

organization accomplishes numerous structural feats to allow for the efficient packaging

and integrity of nearly two meters of linear DNA into nuclei only six microns in diameter

(Alberts et al. 2002). Chromatin influences virtually all DNA-based processes, including

replication, transcription and damage repair through regulation of nucleosome

composition, modification, and positioning (Strahl & Allis 2000, Jenuwein and Allis 2001;

Ernst and Kellis 2010).

As the substrate for myriad processes essential for cellular function, chromatin is

acted upon by an array of cellular machinery to shape its character and composition at

specific sites, or sequences, along the linear chromosomal fiber. These alterations in

turn are maintained, modified, and reset throughout organismal development to

generate chromatin landscapes that reflect cellular identity and state. Post-translational

modifications (PTMs) of histones and DNA itself are critically involved in chromatin-

mediated regulation of gene expression and genome integrity (Strahl and Allis 2000,

Jenuwein and Allis 2001). DNA and histone PTMs exert their effects through direct

physical modulation of nucleosome-DNA contacts and/or recruitment of downstream

"reader" protein complexes. Early work by Vincent Allfrey and colleagues while at the

Rockefeller Institute led them to propose a role for histone acetylation in gene regulation

(Allfrey et al. 1964). To date over 100 histone PTMs have been identified (Huang et al.

2014), many of which are dynamically controlled by enzymes catalyzing their addition

("writers") or removal ("erasers") (**Fig. 1.1**). Chemical moieties that are covalently

attached along the globular and unstructured tail domains of histone proteins run the

gamut from small adducts such as acetyl and methyl groups to larger folded protein

groups such as ubiquitin. Among them, methylation of several lysine residues located at

the amino-terminal tail of histone H3 have been extensively characterized.

Rather than being randomly distributed, histone PTMs exhibit distinct and

characteristic patterns across the genome that correlate with other structural features. In

many cases these associations were first recognized through imaging studies of intact

chromosomes and were later confirmed and extensively characterized using chromatin

immunoprecipitation followed by high-throughput sequencing (ChIP-seq) technology

(Zhou et al. 2010). For example, histone acetylation and methylation of lysine residues

at positions 4 and 36 on histone H3 are enriched at regions associated with active

transcription known as euchromatin, whereas others such as methylation of lysine 9 and

27 on histone H3 are found at sites of transcriptional repression known as

heterochromatin. Moreover, methylation of histone lysine residues exists in multiple

different valences (mono-, di-, or tri-methylation), with each form often regulated in an
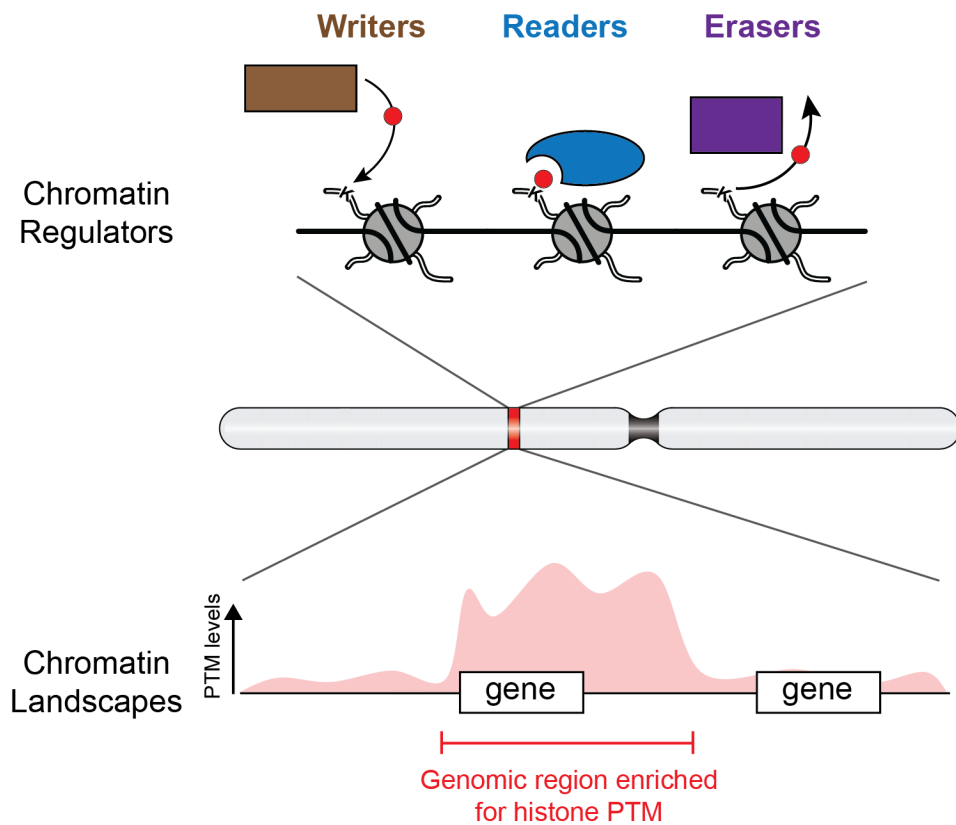
Figure 1.1: Chromatin writers, readers, and erasers establish and interpret the chromatin landscape. Chromatin regulators deposit ("writers"), remove ("erasers"), and bind ("readers") post-translational modifications on DNA and histone proteins. Their collective action creates genomic regions that are enriched or depleted for specific PTMs, creating chromatin landscapes which can be visualized through methodological approaches such as ChIP-seq.

independent fashion (i.e. catalyzed by distinct enzymatic complexes) and existing in unique configurations across the genome.

At a more granular level, histone PTMs co-localize with specific coding and non-coding elements of the genome as a result of diverse recruitment mechanisms that concentrate chromatin regulatory machinery to distinct genomic regions. Tri-methylation of lysine 36 on histone H3 (H3K36me3) has long been recognized as enriched along the gene bodies of actively transcribed genes (Kolasinska-Zwierz et al. 2009). This is in part a consequence of the sole mammalian methyltransferase for H3K36me3, SETD2, being targeted to sites of active transcription through association with subunits of the elongating RNA Polymerase II complex (Xiao et al. 2003, Edmunds et al. 2008). These histone PTM localization patterns in turn direct the further downstream engagement of chromatin effector complexes containing specialized structural reader domains that recognize specific modifications. In such a manner H3K36me3 serves as a scaffold to direct *trans*-acting factors to gene bodies, as has been demonstrated for recruitment of the gene splicing regulator ZYMND11 through its PWWP (Pro-Trp-Trp-Pro) domain (Wen et al. 2014, Guo et al. 2014).

The extent of deposition of individual histone PTMs across the chromatin landscape also varies considerably based on the relative size of their associated genomic elements. For instance, certain modifications such as acetylation of lysine 27 (H3K27ac) and tri-methylation of lysine 4 (H3K4me3) of histone H3 are enriched at relatively narrow regulatory genomic regions such as promoters and enhancers that span only a few kilobases in length. As a result, in most cell types these modifications

are found on only a small fraction of total cellular histone H3 protein as determined by quantitative mass-spectrometry measurements (Janssen et al. 2019). This stands in contrast to other more abundant modifications such as di-methylation of lysine 36 on histone H3 (H3K36me2) which is enriched across broad megabase-spanning intergenic regions and is thus present on a large fraction of the total histone H3 pool.

DNA METHYLATION IN MAMMALS: WRITERS & ERASERS

The discovery of DNA methylation in mammals by Rollin Hotchkiss while at The Rockefeller Institute of Medical Research in 1948 came shortly after the original identification of DNA as the carrier of genetic information a few years prior (Avery et al. 1944, Hotchkiss 1948). Over the decades since, insight into how DNA methylation is established, maintained, and removed has been aided by the discovery of a cohort of writers, readers and erasers that shape and interpret the DNA methylation landscape analogous to the regulation of histone PTMs. Unlike histone PTMs, modifications to the DNA backbone itself are less numerous in type as DNA methylation in mammals is limited to cytoside residues and in most cell types occurs predominately at CpG dinucleotide sequences, of which 70-80% in a cell are methylated (Ehrlich et al. 1982) (**Fig. 1.2A**).

A unique feature of DNA methylation is the separation of the establishment and maintenance writer activities into distinct enzymatic complexes. *De novo* DNA methyltransferases DNMT3A and DNMT3B (Okano et al. 1998) catalyze methylation of previously unmethylated cytosine nucleotides and thereby initially establish DNA
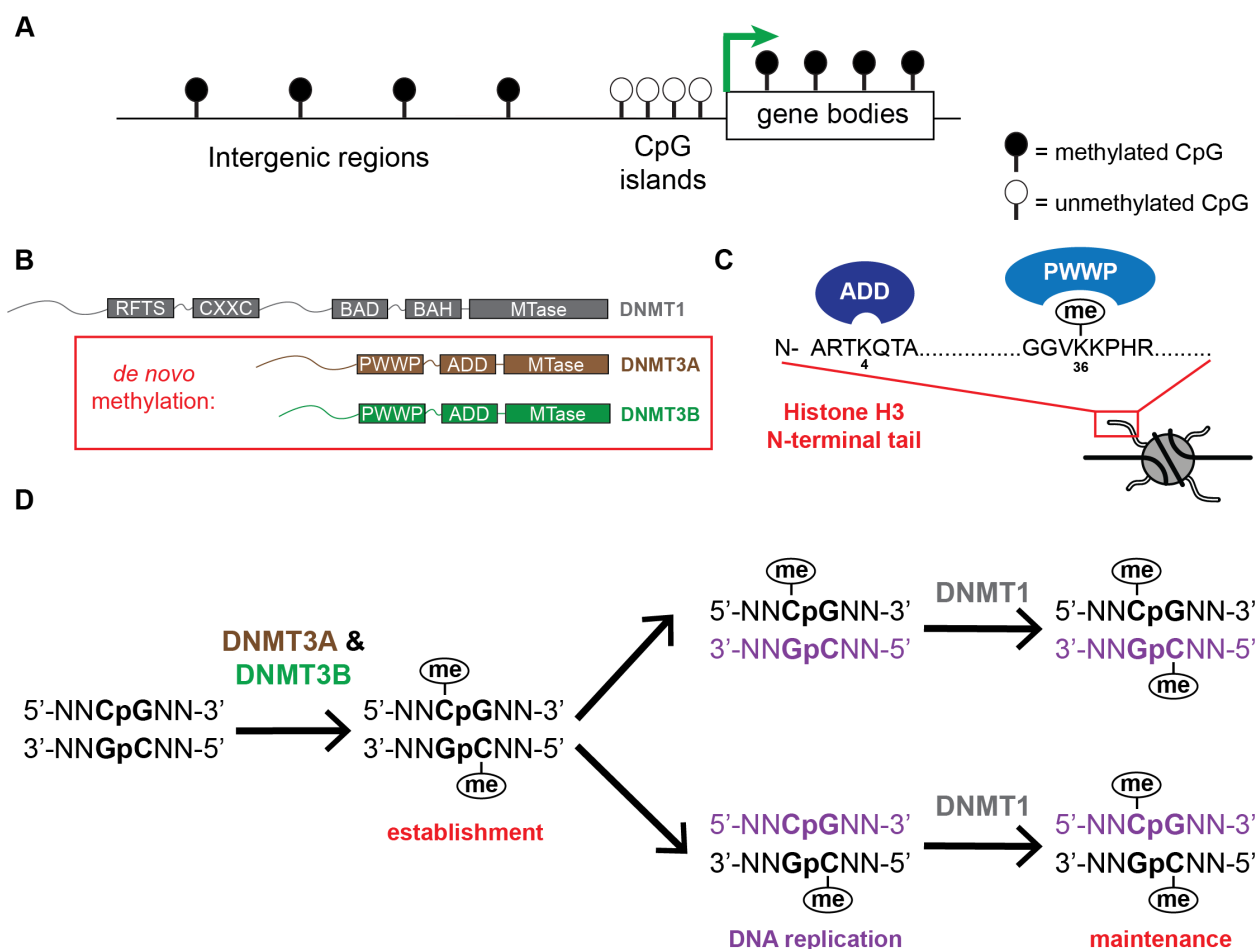
Figure 1.2: DNA methylation and DNA methyltransferases in mammals. A) DNA methylation landscape in mammals. Intergenic regions and actively transcribed gene bodies are enriched for CpG methylation. CpG-dense regions known as CpG islands are largely unmethylated in most cell types and are present at ~70% of human promoters. **B)** Schematic of the annotated domain structure of the maintenance DNA methyltransferase DNMT1 and the *de novo* DNA methyltransferases DNMT3A and DNMT3B. **C)** Schematic of binding selectivity for *de novo* DNMTs' ADD and PWWP reader domains. Both domains bind portions of the amino-terminal tail of histone H3; the ADD domains bind unmethylated H3K4, the PWWP domains bind different valences of H3K36 methylation. **D)** The *de novo* DNMTs establish the DNA methylation landscape in cells, which is then faithfully deposited on newly synthesized DNA (purple) during replication through the maintenance activity of DNMT1 on hemi-methylated substrates.

methylation in cells. They achieve this in part through the recognition of the amino-terminal tail of histone H3 using two separate chromatin reader domains: a PWWP domain and an ADD (ATRX, DNMT3, DNMT3L) domain. These reader domains serve both to target DNMT3A/B to specific genomic loci as well as to regulate their enzymatic activity and are essential for concentrating catalysis selectively across the genome (**Fig. 1.2B**).

The better characterized of the two domains, the ADD domain binds to H3K4 unmethylated histone H3 and is inhibited by the presence of H3K4 methylation, in particular by tri-methylation (H3K4me3) that is classically found at the promoters of actively transcribed genes (Ooi et al. 2007, Otani et al. 2009) (**Fig. 1.2C**). Such a mechanism is thought to explain the inverse correlation between promoter CpG methylation and gene expression. Structural analysis indicates the ADD domain of DNMT3A interacts with the catalytic domain to reduce its DNA-binding affinity, thereby placing the enzyme in an auto-inhibited state that is released upon the ADD domain binding the H3 tail (Guo et al. 2015). Furthermore, this interaction occurs independent of DNMT3L, a non-catalytic paralogue expressed in certain developmentally immature cell types such as mouse embryonic stem cells (mESCs) that can complex with either DNMT3A or DNMT3B to stimulate their catalytic activity (Suetake et al. 2004).

Many chromatin-modifying complexes are capable of recognizing and binding their catalyzed product as a mechanism to propagate existing PTMs onto newly synthesized nucleosomes, thereby transmitting epigenetic information through cell division to daughter cells. Certain histone-modifying enzymes, in addition to their

catalytic domain, contain reader domains that recognize their catalyzed product for this potential purpose (e.g. PWWP domain of the histone methyltransferase NSD1 binds H3K36me2; EED subunit of histone methyltransferase PRC2 binds H3K27me3). DNA methylation is unique in that this ability to maintain the modification across cell division ("maintenance activity") is not inherent to DNMT3A/B but rather is the responsibility of DNMT1, which recognizes hemi-methylated DNA duplex substrates and faithfully methylates the unmodified, newly synthesized DNA strand at symmetric CpG dinucleotides in a pattern identical to the template strand (**Fig. 1.2D**). Like its *de novo* counterparts, DNMT1 also is a multi-modular protein and relies on a complex interplay between its functional domains to coordinate its enzymatic activity during replication. This includes the use of an auto-inhibitory mechanism in which binding of the CXXC domain of DNMT1 to unmethylated CpG dinucleotides prevents access of the DNA to the active site, thereby preventing *de novo* methylation and allowing only for methylation of hemimethylated CpG dinucleotides (Song et al. 2011).

DNMT1 maintenance activity is necessary to preserve DNA methylation landscapes and its absence leads to the gradual loss of CpG methylation genome-wide over successive rounds of cell division, in a process known as passive demethylation. Active demethylation can also occur and is initiated through oxidation of 5-methylcytosine by ten-eleven translocation (TET) family enzymes. The generation of oxidative derivatives such as 5-hydroxymethylcytosine and further oxidative states such as 5-formylcytosine and 5-carboxylcytosine contribute to demethylation both through passive dilution over replication as well as being subject to active removal by base

8

excision repair machinery (Kohli & Zhang 2013). In contrast to CpG methylation, which

is abundant and broadly distributed across the genome, TET enzymes and their

oxidized products in most cell types are enriched at gene regulatory regions, including

enhancers and promoters (Hon et al. 2014, Rasmussen et al. 2019). Often these

regulatory regions are unmethylated CpG-dense regions of the genome which are

known as "CpG islands" (Williams et al. 2011). TET enzymes effect on gene expression

likely reflect both the erasure of 5-methylcytosine, as well as the presence of the

oxidative derivates which themselves can serve as scaffolds for chromatin reader

proteins (Zhou et al. 2014).

DNA METHYLATION & GENE SILENCING

Appreciation for the role of DNA sequence in directing DNA-templated processes such

as transcription led to interest in how modifications to the DNA backbone itself might

alter the interpretation of genetic information by the cell. Multiple labs hypothesized that

DNA methylation may act as a regulatory signal to influence gene expression in a

heritable fashion, thereby providing stability to the identity and function of a

differentiated cell (Scarano et al. 1967, Holliday & Pugh 1975, Sager & Kitchin 1975,

Riggs 1975). Such a notion was further bolstered by reports of a correlation between

DNA methylation and transcriptional silencing as part of mammalian cellular

differentiation and viral life cycle gene expression programs (Christman et al. 1977,

Desrosiers et al. 1979).

The development of a cytosine analog, 5-azacytidine, that could inhibit the

methylation of newly synthesized DNA (Jones & Taylor 1980) led to a flurry of studies

characterizing its effect on gene regulation in different contexts. Collectively, these

reports provided further experimental evidence for a role of DNA methylation in

transcriptional silencing, such as by linking 5-azacytidine treatment to the derepression

of endogenous retroviruses (ERVs) embedded in vertebrate genomes and to the

reactivation of genes important for the activity of integrated herpes simplex virus

(Groudine et al. 1981, Clough et al. 1982). A direct link was subsequently provided by

the demonstration that *in vitro* methylation of DNA inhibited expression of the encoded

gene upon introduction into vertebrate nuclei and that these exogenously introduced

changes to CpG methylation could be stably inherited over multiple cell divisions

(Vardimon et al. 1982, Stein et al. 1982). Later studies went on to show that depletion of

DNA methylation *in vivo* through the generation of mouse embryos deficient for DNMT1

led to derepression of IAP retroelements that normally have high CpG methylation

levels in somatic cells (Walsh et al. 1998). Thus, multiple lines of evidence point to a

role for DNA methylation in sustaining transcriptional repression of certain genetic

elements such as repetitive retroviruses.

Beyond CpG methylation maintaining genome integrity by preventing the

expression and potentially deleterious re-integration of ERVs, it also influences

numerous aspects of gene regulation. Interestingly, many gene promoters in

mammalian cells reside in CpG islands and remain unmethylated in most somatic cells

(Saxonov et al. 2006). However, promoter CpG islands can become methylated during

normal development and this is associated with transcriptional silencing (Shen et al.

2007, Weber et al. 2007, Meissner et al. 2008). In particular, often these cases

represent instances of monoallelic silencing in which only one of two copies of a gene in

diploid cells is expressed. Examples of monoallelic silencing include X chromosome

inactivation (XCI) in females and regulation of imprinted gene expression. XCI in female

cells serves as a form of dosage compensation by repressing gene expression to

counterbalance the presence of an additional X chromosome (Avner & Heard 2001).

During XCI, CpG islands become methylated across the inactivated X chromosome and

occurs most prominently at genes which are silenced (Sharp et al. 2011). DNA

methylation plays an essential role in this silencing, as 5-azacytidine treatment induces

re-expression of alleles on the inactive X chromosome (Xi) (Graves 1982).

Genomic imprinting is the process by which a small group of genes are selected

during gametogenesis for expression of a single parental allele in the developing

offspring. Imprinted gene regulation is controlled, at least in part, through differential

CpG methylation of imprinting control regions (ICRs) established during

spermatogenesis and oogenesis (Bartolomei & Ferguson-Smith 2011). These imprinted

regions are resistant to the genome-wide loss of DNA methylation that occurs shortly

after fertilization and maintain parent-of-origin expression patterns throughout the life of

the organism (Li et al. 1993, Morgan et al. 2005). While maternal ICRs are mostly

promoter-associated CpG islands that are methylated on the silenced allele, paternal

ICRs are often found at intergenic regions and likely influence gene regulation through

alternative mechanisms. For example, a paternal ICR at the *Igf2-H19* locus regulates

enhancer-promoter interactions by serving as a binding site for the insulator protein CTCF (Bell & Felsenfeld 2000, Hark et al. 2000). Methylation of the ICR on the paternal chromosome blocks binding of the methylation-sensitive DNA-binding function of CTCF, thereby allowing a nearby enhancer to interact with the Igf2 promoter to drive expression (Murrell et al. 2004).

CpG methylation has also been linked to inappropriate gene silencing in human disease. Many cancers exhibit aberrant methylation of promoter-associated CpG islands at genes encoding tumor suppressors and developmental regulators (Greger et al. 1989, Merlo et al. 1995, Herman et al. 1995, Gonzalez-Zulueta et al. 1995, Toyota et al. 1999, Weisenberger et al. 2006). CpG island hypermethylation in cancer is thought to facilitate oncogenic gene expression programs to promote tumor formation and progression (de Bustros et al. 1988, Merlo et al. 1995). However, the underlying mechanisms and whether they are related to those involved in monoallelic silencing during normal development remain unclear. Work in Chapter 3 will touch on one possible mechanism by which *de novo* methylation of CpG islands may be established in development and disease.

DNA METHLYATION IN DEVELOPMENT & DISEASE

Enzymes that catalyze the addition and removal of CpG methylation in mammals play a critical role in the specification and maintenance of cell fate. Initial characterization of the pan-DNMT small molecule inhibitor 5-azacytidine and its analogs revealed that treatment of the mouse C3H10T1/2 mesenchymal stem cell line induced the formation

of differentiated muscle and fat cells, suggesting that DNA methylation may act to regulate cellular identity (Jones & Taylor 1980). Genetic ablation of DNMT1/DNMT3A/DNMT3B (DNMT TKO) or TET1/TET2/TET3 (TET TKO) in mouse embryonic stem cells (mESCs) permits continued self-renewal, likely due to the strong pluripotency-associated transcriptional circuitry in these cells, but leads to profound defects in differentiation to all three germ layers (Tsumura et al. 2006, Dawlaty et al. 2014). Accordingly, depletion of CpG methylation due to deficiency for DNMT1 or both DNMT3A/B leads to growth arrest following gastrulation and embryonic lethality; likewise, TET TKO mESCs fail to support embryonic development or contribute to chimeric embryos (Li et al. 1992, Okano et al. 1999, Dawlaty et al. 2014). These defects in early mammalian development cannot be attributed solely to effects on imprinting, as DNMT3A/B double knockout embryos failed to develop despite exhibiting intact imprinting patterns due to the continued maintenance activity of DNMT1 (Okano et al. 1999).

DNA methylation regulatory pathways are similarly relevant to human physiology, as demonstrated by the existence of somatic mutations within genes encoding DNMT and TET family enzymes in tumors, as well as the discovery of germline variants in patients with developmental disorders. Recurrent hotspot mutations within the catalytic domain of DNMT3A occur in a substantial proportion of patients with acute myeloid leukemia (AML) (Ley et al. 2010), where they are proposed to act in a dominant-negative fashion to globally reduce DNMT3A catalytic activity in cells (Kim et al. 2013, Russler-Germain et al. 2014). Conditional deletion of *Dnmt3a* in mouse hematopoietic

stem cells (HSCs) both promotes self-renewal and impairs differentiation in serial transplantation experiments in mice, ultimately leading to an expanded HSC pool in the bone marrow (Challen et al. 2011). This is consistent with the phenomenon of age-related "clonal hematopoiesis" identified in asymptomatic human populations whereby single gene mutations within a hematopoietic stem or progenitor cell provide a selective growth advantage to allow for expansion of a pre-malignant clone (Jan et al. 2017). Indeed, mutations in *DNMT3A* are now recognized as some of the most prevalent drivers of clonal hematopoiesis and are associated with increased risk of hematologic cancer (Jaiswal et al. 2014).

Germline *DNMT3A* mutations have also been implicated in the pathogenesis of a developmental overgrowth and intellectual disability syndrome known as Tatton-Brown-Rahman syndrome (TBRS) (Tatton-Brown et al. 2014). TBRS-associated mutations are loss-of-function and do not permit stable DNMT3A expression, in contrast to distinct gain-of-function mutations recently identified in a separate developmental syndrome (Heyn et al. 2019). In a similar fashion to the hematopoietic system, conditional deletion of *Dnmt3a* in mouse neural progenitor cells allows for continued expansion of undifferentiated cells but results in differentiation defects (Wu et al. 2010)- whether this impact on cellular function may underpin the enlarged brain size and intellectual disabilities characteristic of TBRS patients remains unclear. Regardless, the role of DNMT3A in development appears unique from other DNA methylation machinery, as germline mutations in *DNMT3B* or *DNMT1* result in distinct clinical presentations without substantially effecting developmental growth (Klein et al. 2011, Xu et al. 1999).

Besides human developmental abnormalities caused by mutations in *DNMT1*, *DNMT3A*, and *DNMT3B* it has long been recognized that locus-specific alterations in DNA methylation status can underlie imprinting disorders in humans, over a dozen of which have been described to date (Buiting et al. 1998, Gicquel et al. 2005, Monk et al. 2019). In many instances, these disorders are defined by genetic alterations such as uniparental disomy of imprinted genomic regions or mutation of imprinted genes. However, for a subset of patients the predominant molecular feature observed is the presence of an 'epimutation': the inappropriate gain or loss of a CpG methylation at an imprinting control center either due to stochastic errors in the establishment and/or maintenance of DNA methylation or due to secondary effects from other genetic alterations (Monk et al. 2019). These genetic and epigenetic alterations in turn coalesce to dysregulate expression of imprinted genes, many of which are implicated in controlling organismal growth. Accordingly, many imprinting disorders are characterized by growth disturbances both *in utero* and postnatally (Plasschaert & Bartolomei 2014).

While no developmental syndromes are currently attributed to mutation of TET family enzymes, both genetic and metabolic perturbations to TET protein activity have been observed in multiple cancer types. *TET2* loss-of-function mutations occur in myeloid malignancies such as AML, lymphoid malignancies such as diffuse large B-cell lymphomas, and are also a frequent contributor to clonal hematopoiesis (Delhommeau et al. 2009, Busque et al. 2012, Asmar et al. 2013, Dominguez et al. 2018). In a similar fashion to loss of DNMT3A, depletion of TET2 in hematopoietic stem cells promotes self-renewal and impairs differentiation potential (Moran-Crusio et al. 2011).

Furthermore, TET2 function can be disrupted secondary to metabolic perturbations in the cell. For example, Vitamin C is a necessary co-factor for TET enzyme activity and its depletion promotes leukemogenesis in part through reduced TET2 activity (Agathocleous et al. 2017, Cimmino et al. 2017). Alternatively, in multiple cancer types, including AML, mutations in the metabolic enzymes isocitrate dehydrogenase 1 (IDH1) or 2 (IDH2) lead to the production of the oncometabolite R-2-hydroxyglutarate which acts as a competitive inhibitor to both histone demethylases and TET family hydroxylases (Xu et al. 2011). In general, tumors with reduced TET function often exhibit DNA hypermethylation of CpG islands presumably due to an inability to actively demethylate 5-methylcytosine established by DNMT enzymes at these sites (Asmar et al. 2013).

CHROMATIN 'CROSS-TALK'

Chromatin landscapes for both DNA methylation and histone PTMs are shaped through the combined activity of chromatin writers and erasers, which are targeted to specific genomic sites through diverse recruitment mechanisms. However, chromatin landscapes for individual modifications are not shaped in isolation but rather are influenced by the presence or absence of other PTMs. The mechanistic basis for this "cross-talk" between different histone and DNA modifications can take many forms. Often it involves modulation of chromatin enzyme activity and/or binding to a nucleosome by pre-existing modifications that may be adjacent or distant to the substrate for catalysis (Soshnev et al. 2016). Depending on the mechanism of action

16

this modulation can act to either enhance, or hinder, the activity of the enzyme to ultimately promote co-occurrence or mutual exclusion of the two modifications across the genome (**Fig. 1.3A,B**).

A well-known example of positive "cross-talk" is the process of gene silencing established by Polycomb Repressive Complex 1 (PRC1) and 2 (PRC2). Together, these multi-subunit complexes work in tandem to repress transcription of developmentally regulated genes in a reversible manner, forming what is referred to as facultative heterochromatin (as opposed to constitutive heterochromatin, which is more strongly transcriptionally silenced and often marked by H3K9 tri-methylation) (Peters et al. 2003). PRC2 acts as the sole H3K27 methyltransferase in mammalian cells through the action of its interchangeable EZH1/2 catalytic subunits, whereas PRC1 deposits ubiquitin on lysine 119 of histone H2A (H2AK119Ub) through the enzymatic activity of its interchangeable RING1A/B subunits (de Napoles et al. 2004, Wang et al. 2004). The presence of both modifications is reinforced at Polycomb-regulated genomic regions in cells through the ability of each complex to be recruited by the reaction product of the other. PRC1 is capable of binding the terminal PRC2-catalyzed product associated with transcriptional silencing, tri-methylation of H3K27 (H3K27me3), using a chromodomain-containing CBX family subunit (Min et al. 2003). Meanwhile, PRC2 can localize to regions enriched for PRC1-catalyzed H2AK119Ub through recognition by a ubiquitin-interacting motif (UIM) on the subunit JARID2 (Blackledge et al. 2014, Cooper et al. 2014, Cooper et al. 2016). This motif, like UIMs identified in other proteins, interacts with the hydrophobic surface patch of ubiquitin (Beal et al. 1996, Hofmann & Falquet

17
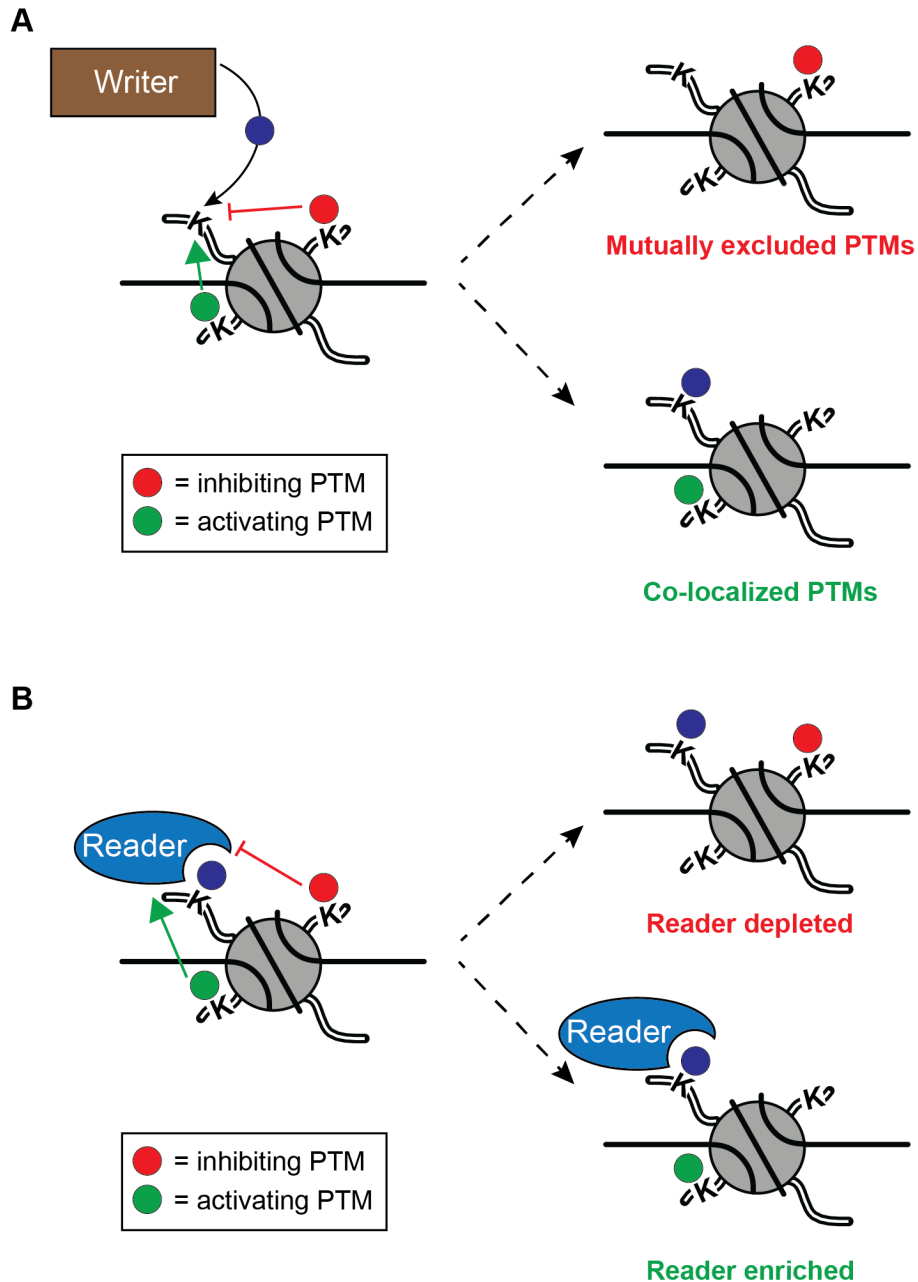
Figure 1.3: Chromatin molecular cross-talk. A) Chromatin writer (or eraser) enzymatic activity can be modulated by pre-established histone PTMs on the same or adjacent nucleosomes. Inhibitory PTMs exclude the writer-mediated deposition of new modifications, whereas activating PTMs promote writer-mediated deposition to support co-localization of the two marks across the genome. **B)** Likewise, chromatin reader binding can be modulated by adjacent PTMs to its direct chromatin binding substrate to enhance or impair reader genomic targeting.

2001, Swanson et al. 2003), in contrast to ubiquitin-dependent recruitment (UDR) domains such as that identified in 53BP1 which engage portions of the nucleosome core particle in addition to ubiquitin (Fradet-Turcotte et al. 2013, Wilson et al. 2016). However, it should be noted that while a strong positive correlation exists between H2AK119Ub and H3K27me3 genome-wide in mammalian cells the modifications do not always co-localize; an observation that may be explained in part by multiple factors that simultaneously influence PRC1/2 binding and activity as well as the presence of non-canonical forms of each protein complex (Gao et al. 2012, Zhen et al. 2016, Laugesen et al. 2019).

An additional modification that plays a key role in regulating PRC2 deposition of H3K27me3 is methylation at the adjacent lysine 36 residue on the same amino-terminal tail of histone H3. In mammals, several methyltransferases targeting H3K36 have been reported (Wagner and Carpenter 2012). While SETD2 is the only enzyme that can generate tri-methylation of H3K36 (H3K36me3), multiple methyltransferases catalyze mono- and di-methylation of H3K36 (H3K36me1/2) including NSD1/2/3 and ASH1L. Both quantitative mass-spectrometry measurements as well as ChIP-seq analysis indicate that H3K27me3 is largely excluded from nucleosomes modified with either H3K36me2 or H3K36me3 (Yuan et al. 2011, Lu et al. 2016). Furthermore, inhibition of H3K36 methyltransferase function leads to aberrantly increased PRC2 activity at sites depleted of H3K36me2/3 (Lu et al. 2016, Fang et al. 2016). The molecular basis for this negative "cross-talk" in which H3K27me3 propagation is antagonized by the presence of H3K36 methylation relies on the recognition of the lysine 36 residue of the H3 tail by the

catalytic SET domain of EZH2 (Jani et al. 2019). This region of EZH2 is narrow enough to only accommodate an interaction with unmethylated lysine 36, which in turn stimulates the catalytic activity of PRC2 towards lysine 27. Nucleosomes that are di- or tri-methylated at lysine 36 are too bulky to be accommodated within this sensing pocket and thereby prevent deposition of H3K27me3.

As previously touched upon, another example of how the chromatin landscape is shaped by antagonistic molecular "cross-talk" between chromatin modifications is the inhibition of DNMT3A interaction with histone H3 by H3K4me3. However, the lack of co-localization between CpG methylation and H3K4me3 genome-wide only helps to explain a fraction of the DNA methylation landscape in cells. Other features characteristic of mammalian genomes include the enrichment of CpG methylation within gene bodies and intergenic regions, as well as within CpG islands in certain pathologic settings (Schübeler 2015). Until recently it has remained unclear whether such enrichment is a passive process that occurs as a result of exclusion from other genomic regions or is in fact due to an active recruitment mechanism. Initial seminal work has linked DNMT3B, but not DNMT3A, to gene body DNA methylation through binding of H3K36me3 by its PWWP chromatin reader domain (Baubec et al. 2015, Morselli et al. 2015). This work will elucidate additional chromatin recruitment mechanisms for DNMT3A, both PWWP-dependent (Chapter 2) and PWWP-independent (Chapter 3), that provide further insight into how the DNA methylation landscape is regulated in mammals and how perturbations to these processes contribute to human pathophysiology.

Portions of Chapter 2 have been published as:

Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281-286 (2019).

Portions of Chapter 3 have been submitted as:

Weinberg, D. N. *et al.* DNMT3A binds H2AK119Ub to hypermethylate CpG islands upon loss of PWWP reader function. *Submitted*.

CHAPTER 2: The histone mark H3K36me2 recruits DNMT3A and shapes the

intergenic DNA methylation landscape

INTRODUCTION

Enzymes catalyzing CpG methylation play an essential role in mammalian tissue

development and homeostasis (Li et al. 1992, Okano et al. 1999, Wu et al. 2010,

Challen et al. 2011). They are also implicated in human developmental disorders and

cancers (Ley et al. 2010, Klein et al. 2011, Xu et al. 1999, Tatton-Brown et al. 2014),

supporting a critical role for DNA methylation during cell fate specification and

maintenance. Recent studies suggest that histone post-translational modifications

(PTMs) are involved in specifying patterns of DNMT localization and DNA methylation.

The presence of H3K4me3 at promoters of actively transcribed genes impairs

nucleosomal binding by the ADD chromatin reader domains of DNMT enzymes (Ooi et

al. 2007, Otani et al. 2009). More recently DNMT3B, but not DNMT3A, was

demonstrated to methylate actively transcribed gene bodies as a result of PWWP

domain-dependent recognition of H3K36me3 (Baubec et al. 2015, Morselli et al. 2015).

Despite such progress, cellular regulation of the DNA methylome remains incompletely

characterized as CpG methylation in most cell types is widespread and is found at

~80% of all CpG dinucleotides across the genome, including at non-coding regions

(Ehrlich et al. 1982). The mechanisms governing the establishment and maintenance of

intergenic DNA methylation remain poorly understood, and will be the subject of this

chapter.

We hypothesized that regulatory mechanisms for intergenic DNA methylation may be perturbed in patients with dysfunctional DNMT enzyme chromatin recruitment and/or function. Germline mutations in *DNMT3A* define Tatton-Brown-Rahman syndrome (TBRS), a childhood overgrowth disorder characterized by tall stature, facial abnormalities, and intellectual disability (Tatton-Brown et al. 2014). TBRS-associated missense mutations alter residues throughout structured regions of DNMT3A, including within the PWWP domain that has been associated with recognition of H3K36 methylated nucleosomes (Dhayalan et al. 2010). Interestingly, many clinical features of TBRS are shared with Sotos syndrome caused by haploinsufficiency of *NSD1*, a histone methyltransferase catalyzing di-methylation on H3K36 (H3K36me2) (Tatton-Brown et al. 2014, Kurotaki et al. 2002, Rayasam et al. 2003), pointing to a potential mechanistic link between the two diseases. As H3K36me2 is enriched at intergenic regions of the genome (Lu et al. 2016, Streubel et al. 2018), we wondered whether it may play a role in shaping the DNA methylation landscape and becomes dysregulated in developmental disorders and cancer. In this chapter we utilize a variety of cell-free and cell-based techniques to profile PWWP domain-dependent recruitment mechanisms of DNMT3A.

RESULTS

To better understand the role of histone PTMs in regulating CpG methylation genome-wide we first wished to characterize DNA methyltransferase (DNMT) targeting, with a particular focus on intergenic regions. We rationalized that as the only *de novo* DNA methyltransferases conserved across mammals, either DNMT3A and/or DNMT3B must be responsible for the deposition of intergenic DNA methylation and may therefore

co-localize with a specific histone PTM at intergenic loci. As no ChIP-grade antibodies currently exist for either enzyme, I generated transgenic cell lines of C3H10T1/2 mouse mesenchymal stem cells (mMSCs) expressing either DNMT3A or DNMT3B containing a hemagglutinin (HA)-epitope tag. I then profiled their genome-wide distribution patterns, along with H3K36me2 and a variety of well-characterized histone PTMs (H3K36me3, H3K27me3, H3K27ac, H3K9me3, H3K4me1) using ChIP-seq. Histone PTMs were chosen based on their association with either euchromatic (H3K36me3, H3K27ac, H3K4me1) and heterochromatic (H3K27me3, H3K9me3) portions of the genome as well as their previously well-characterized localization to specific genomic regions (i.e. H3K36me3 with gene bodies; H3K27ac and H3K4me1 with active regulatory regions). We also measured levels of CpG methylation at base-pair resolution with whole-genome bisulfite sequencing (WGBS) at high coverage (840 million reads; 45x coverage).

To rigorously assess genome-wide associations using our high-throughput sequencing data we collaborated with bioinformaticians in Dr. Jacek Majewski's group at McGill University (graduate student Simon Papillion-Cavanagh and post-doctoral fellow Dr. Haifen Chen) and Dr. Chao Lu's group at Columbia University (post-doctoral fellow Dr. Xiao Chen). Assessing the chromatin landscape of histone PTMs using a genome browser revealed enriched regions of H3K36me2 and H3K36me3 that were in close proximity and together defined Mb-sized domains which were largely exclusive from domains demarcated by H3K9me3 and H3K27me3 (**Fig. 2.1A**). Consistent with previous studies that connect H3K36me2 with transcriptional activation (Rao et al. 2005, Xiao et al. 2003), additional features associated with H3K36me2/3 domains included

Figure 2.1: H3K36me2 and H3K36me3 mark transcriptionally active euchromatin.
A) Genome browser representation of ChIP-seq normalized reads for H3K36me3,
H3K36me2, H3K27me3, H3K9me3, DNMT3A1, and DNMT3B in mMSCs at Chr2:
110.7-116.2 Mb. Levels of CpG methylation are depicted as a heat map (blue/low;
white/intermediate; red/high). Refseq genes are annotated at the bottom. **B)** Heat map
showing pairwise Pearson correlation of 10kb bins (n = 246,285) between H3K36me2/3,
H3K27me3, H3K27ac, H3K9me3, H3K4me1 and CpG methylation. **C)** Quantification of
ChIP-seq normalized reads for gene expression, H3K4me1 and H3K27ac within
H3K36me2/3 (n = 591), H3K27me3 (n = 283), or H3K9me3 (n = 545) domains. P-values
were $< 2.2 \times 10^{-16}$ for all pair-wise comparisons as determined by Wilcoxon's rank sum
test (two-sided), except for gene expression ($P = 1.72 \times 10^{-6}$) between H3K27me3 and
H3K9me3 domains. Boxes represent median and 25th-75th percentiles, whiskers are
minimum to maximum with discrete points representing outliers. **D)** 3D genome browser
representation of Hi-C chromatin conformation data from mouse myoblasts compared to
ChIP-seq normalized reads for histone PTMs in mMSCs at Chr4: 52.2-99.1 Mb. Levels
of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high).
Refseq genes are annotated at the bottom.

A Chr2: 110.7 - 116.2 Mb

H3K36me3 [0 - 97]
H3K36me2 [0 - 48]
H3K27me3 [0 - 5.98]
H3K9me3 [0 - 16]
DNMT3A [3 - 30]
DNMT3B [3 - 20]
CpGme

Ano3 *Olfr1287 Olfr1303* *Chrm5 Ryr3* *Mir1954 Fmn1* *Aqr* *Dph6* *Meis2*

0 50 100%

B

Correlation
1
0.5
0
−0.5
−1

C

Gene expression

H3K4me1

H3K27ac

D

>30

<0

H3K36me3
H3K36me2
H3K27me3
H3K9me3
CpGme

0 50 100%

Chr4: 52.2 - 99.1 Mb
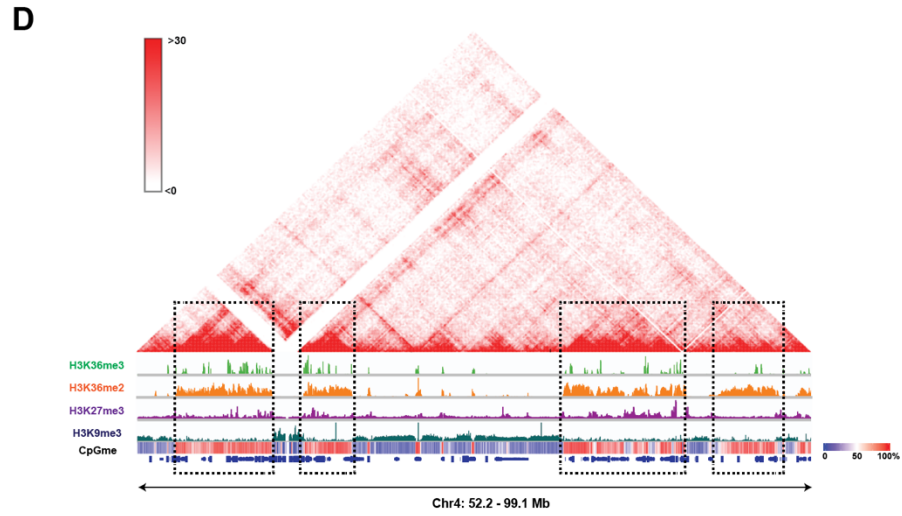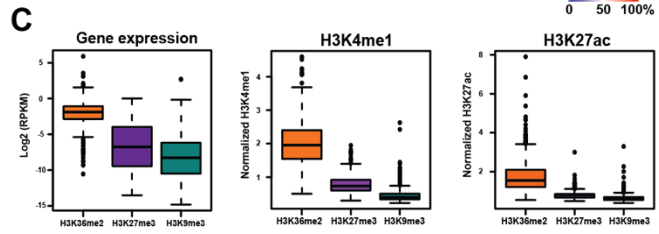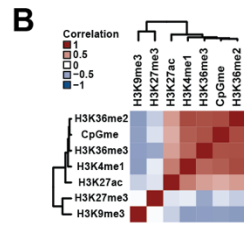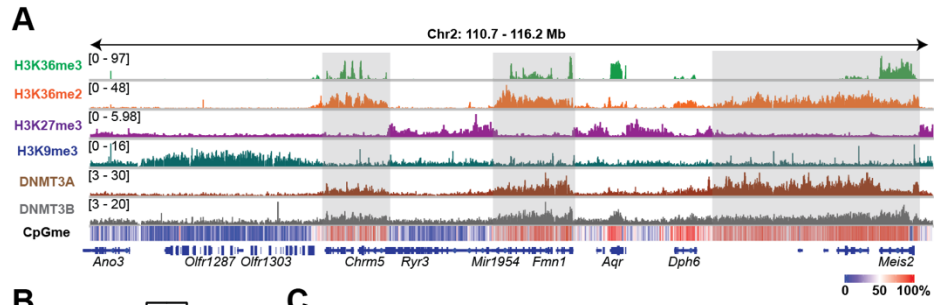
higher levels of H3K27ac, H3K4me1 and increased gene expression (**Fig. 2.1B,C**).

Moreover, H3K36me2/3 domains delineated compartments and topologically associated

domains (TADs) (Cavalli & Misteli 2013) derived from Hi-C studies of mouse myoblasts,

representing genomic regions which engage in frequent physical interactions with one

another (Doynova et al. 2017) (**Fig. 2.1D**). This suggested that H3K36me2/3 together

define transcriptionally active euchromatin that is spatially segregated from constitutive

and facultative heterochromatin.

Consistent with our hypothesis that H3K36me2 may play a role in shaping the

DNA methylation landscape, levels of the histone PTM positively correlated with CpG

methylation genome-wide (**Fig. 2.2A**). Overall we observed that the broad domains of

H3K36me2/3, H3K27me3 and H3K9me3 partitioned the genome into regions of high

(75%), intermediate (50%) and low (30%) levels of CpG methylation, respectively (**Fig.

2.1A, 2.2B**). Furthermore, DNMT3A/B ChIP-seq reads were predominantly observed in

H3K36me2/3 domains (**Fig. 2.1A, 2.2B**), suggesting that high levels of CpG methylation

in these domains are, at least in part, a result of favored DNMT3A/B recruitment. These

observations are consistent with recent reports of H3K36me3-mediated targeting of

DNMT3B activity (Baubec et al. 2015, Morselli et al. 2015). However, as this interaction

is confined to gene bodies, we wondered whether an additional chromatin *trans*-

regulatory pathway involving the deposition of H3K36me2 at intergenic regions may act

in parallel to facilitate CpG methylation at euchromatic regions.

To test this hypothesis, we contrasted the genomic distributions of H3K36me2

and H3K36me3 to better understand their relation to the localization patterns of

DNMT3A and DNMT3B. We noted that while H3K36me3 was enriched within gene
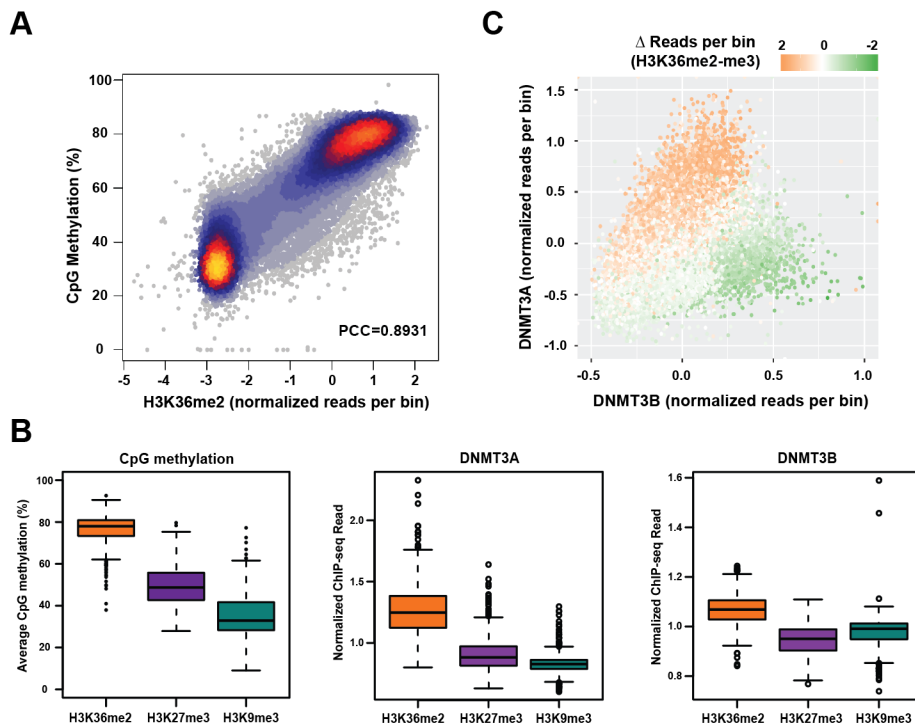
Figure 2.2: Genome-wide colocalization of DNMT3A, CpG methylation, and H3K36me2. **A)** ChIP-seq normalized reads of H3K36me2 were plotted relative to average percentage of CpG methylation in mMSCs for 100kb non-overlapping bins (n = 25,624). Pearson's correlation coefficient is indicated. **B)** Quantification of ChIP-seq normalized reads for DNMT3A1, DNMT3B and averaged CpG methylation within H3K36me2/3 (n = 591), H3K27me3 (n = 283), or H3K9me3 (n = 545) domains. P-values were < 2.2 x 10^{-16} for all pair-wise comparisons as determined by Wilcoxon's rank sum test (two-sided), except for DNMT3B (*P* = 3.73 x 10^{-14}) between H3K27me3 and H3K9me3 domains. Boxes represent median and 25th-75th percentiles, whiskers are minimum to maximum with discrete points representing outliers. **C)** ChIP-seq normalized reads per 10kb bin for DNMT3A1 (y-axis) and DNMT3B (x-axis) in mMSCs were plotted (n = 246,285). Each bin/dot was color-coded based on differences between H3K36me3 and H3K36me2 ChIP-seq reads to show selective enrichment for H3K36me2 (orange) or H3K36me3 (green).

bodies, H3K36me2 showed a more diffuse distribution that encompassed both genic and intergenic regions (**Fig. 2.3A,B**). Within actively transcribed genes, H3K36me2 covered areas downstream of the transcriptional start site through the first intron, followed by a marked switch to H3K36me3 after the first splice junction (**Fig. 2.3C**). Notably, while DNMT3B was enriched within gene bodies containing H3K36me3 as expected, the localization of DNMT3A mimicked the distribution of H3K36me2 and spanned broad intergenic regions without appreciable levels of H3K36me3 (**Fig. 2.3D**). This suggested that the divergent targeting profiles of *de novo* DNMTs might be explained by their ability to preferentially recognize different valences of H3K36 methylation. To address this possibility, we contrasted the presence of H3K36me2 and H3K36me3 genome-wide and found that DNMT3A was selectively enriched over DNMT3B at genomic regions with high levels of H3K36me2 compared to H3K36me3 (**Fig. 2.2C**). In contrast, DNMT3B was selectively enriched over DNMT3A at genomic regions with high levels of H3K36me3 compared to H3K36me2 (**Fig. 2.2C**).

As mMSCs are a somatic cell type that appears during later stages of development and persists into adulthood, we wished to confirm our findings in an earlier developmental context during which the DNA methylation landscape gets established. To that end, we turned to mouse embryonic stem cells (mESCs) grown in media containing serum with leukemia inhibitory factor (LIF), which are pluripotent cells derived from the inner cell mass of the developing blastocyst that are capable of generating all three germ layers. mESCs grown under serum/LIF conditions also correspond with the developmental period following fertilization in which DNA methylation is reestablished across the genome following its previous erasure in both
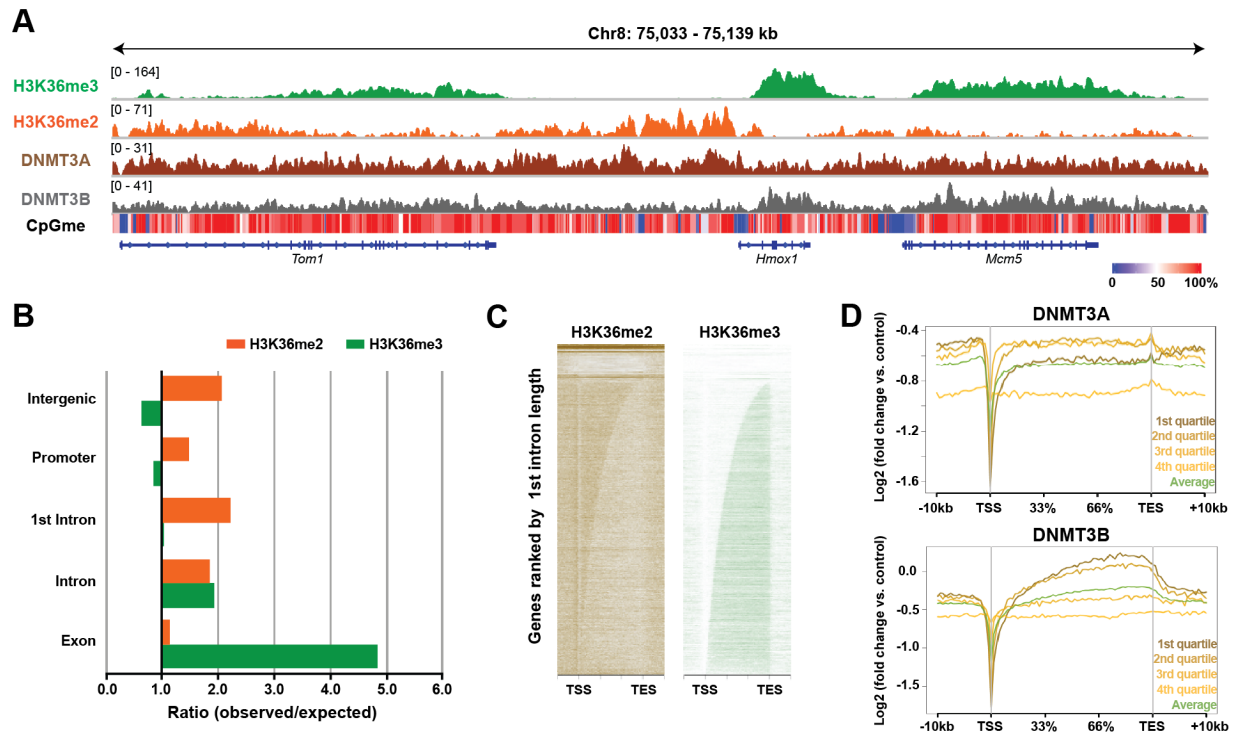
Figure 2.3: Distinct enrichment patterns of DNMT3A/B and H3K36me2/3 at euchromatin. **A)** Genome browser representation of ChIP-seq normalized reads for H3K36me3, H3K36me2, DNMT3A1, and DNMT3B in mMSCs at Chr8: 75.0-75.1 Mb. Levels of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high). Refseq genes are annotated at the bottom. **B)** Ratio of observed to expected ChIP-seq reads for H3K36me2 and H3K36me3 in annotated genomic regions. Numbers of expected reads were generated assuming equivalent genomic distribution to input. **C)** Heat maps representing ChIP-seq signal density for H3K36me2 and H3K36me3 in mMSCs across all gene bodies. Genes are ranked by first intron length. Each gene is displayed as a row. **D)** Averaged ChIP-seq normalized signal across gene bodies stratified by expression quartile, represented as log2 fold-change over input for DNMT3A1 (above) and DNMT3B (below) in parental mMSCs. Sample sizes (same for DNMT3A1 and DNMT3B) are: n = 7,524, 7,435, 6,923, 8,550 for 1st, 2nd, 3rd, and 4th quartile respectively; n = 35,777 for "Average".

male and female germ cell lineages (Messerschmidt et al. 2014). To validate this approach, I first stably expressed hemagglutinin(HA)-tagged DNMT3A2, the predominant DNMT3A isoform found during early development, in parental and DNMT3A-deficient (sgDnmt3a) mESCs. When expressed close to physiologic levels, ChIP-seq of HA-tagged DNMT3A revealed similar genome-wide localization patterns between parental and sgDnmt3a cells thereby confirming that the presence of endogenous untagged DNMT3A was not influencing the localization of the exogenous tagged form (**Fig. 2.4A,B**).

I next performed ChIP-seq for DNMT3B in mESCs and again worked with the Majewski group to compare the targeting profiles of DNMT3A and DNMT3B with the genome-wide distributions of H3K36me2/3. We observed that the selective co-enrichment between H3K36me2-DNMT3A and H3K36me3-DNMT3B was also evident in this context, although in general the distribution patterns of DNMT3A and DNMT3B were more positively correlated in mESCs than we had observed in mMSCs (**Fig. 2.4C,D**). While the reasons for this discrepancy between cell types remain unclear, it is important to note that mESCs are unique in that they also express DNMT3L, which could complex with both DNMT3A and DNMT3B to promote similar localization patterns for both enzymes. Importantly, *de novo* methylation activity of DNMT3A, measured by levels of CpG methylation following the reintroduction of DNMT3A2 into *Dnmt1* and *Dnmt3a/b* triple-knockout mESCs that lack virtually all CpG methylation (Baubec et al. 2015), tracked with levels of H3K36me2 (**Fig. 2.5A**). However, *de novo* activity of DNMT3B did not closely follow levels of H3K36me2 (**Fig. 2.5A**), but rather correlated

Figure 2.4: Genome-wide colocalization between H3K36me2 and DNMT3A2 in mouse ES cells. **A)** Immunoblots of lysates generated from parental and sgDnmt3a mESCs ectopically expressing HA-tagged DNMT3A2. Vinculin was used as a loading control. Endogenous expression of the long isoform (DNMT3A1) and short isoform (DNMT3A2) are indicated. **B)** ChIP-seq normalized reads of HA-tagged DNMT3A2 in sgDnmt3a mESCs were plotted relative to in parental mESCs for 100kb non-overlapping bins (n = 26,181). Pearson's correlation coefficient is indicated. **C)** Genome browser representation of ChIP-seq normalized reads for H3K36me2, H3K36me3, DNMT3A2, and DNMT3B in mESCs at Chr12: 86.6-87.5 Mb. Refseq genes are annotated at the bottom. Shaded areas indicate H3K36me2-enriched intergenic regions (orange) and H3K36me3-enriched genic regions (green) in parental cells. **D)** ChIP-seq normalized reads per 10kb bin for DNMT3A2 (y-axis) and DNMT3B (x-axis) in mESCs (n = 246,285). Each bin/dot was color-coded based on differences between H3K36me3 and H3K36me2 ChIP-seq reads to show selective enrichment for H3K36me2 (orange) or H3K36me3 (green).

A

B

mESC

HA-DNMT3A (normalized reads per bin) in sgDnmt3a cells

HA-DNMT3A (normalized reads per bin) in Parental cells

PCC = 0.848

C

mESC

Chr12: 86.6 - 87.5 Mb

H3K36me3 [0 - 110]
H3K36me2 [0 - 16]
DNMT3A [2 - 14]
DNMT3B [2 - 14]

Vash1 Lrrc74a Irf2bpl Cipc Tmem63c Pomt2 Tmed8 Noxred1 Ism2 Sptlc2 Snw1
Angel1   Zdhhc22 Ngb Gstz1 Ahsa1 Alkbh1

d

mESC

Dnmt3a (normalized reads per bin)

Dnmt3b (normalized reads per bin)

Δ Reads per bin
(H3K36me2-me3)

-5 -2.5 0 2.5

Figure 2.5: *De novo* DNA methylation by DNMT3A and DNMT3B in mouse ES cells. **A)** *De novo* methylation per bin by DNMT3A2 (brown) or DNMT3B (grey) upon reintroduction into DNMT triple knockout mESCs relative to H3K36me2 (n = 1,000). To generate bins, 1kb genomic tiles (n = 2,462,755) were ranked by H3K36me2 enrichment in mESCs and grouped into 1000 rank-ordered bins (2,463 tiles per group). Dashed line indicates H3K36me2 enrichment per bin. Goodness of fit was computed on a quadratic model (red line). **B)** *De novo* methylation per bin by DNMT3A2 (brown) or DNMT3B (grey) upon reintroduction into DNMT triple knockout mESCs relative to H3K36me3. To generate bins, 1kb genomic tiles (n = 2,462,755) were ranked by H3K36me3 enrichment in mESCs and grouped into 1000 rank-ordered bins (2,463 tiles per group). Dashed line indicates H3K36me3 enrichment per bin. Goodness of fit was computed on a quadratic model (red line).

with H3K36me3 as previously shown by Baubec et al. (**replicated in Fig. 2.5B**). We therefore conclude from this analysis that intergenic targeting of DNMT3A and gene-body targeting of DNMT3B together facilitate the establishment of CpG methylation in euchromatin.

To assess the requirement of H3K36me2 in targeting DNMT3A to intergenic regions, I first worked together with Dr. Kartik Rajagopalan and Dr. Chao Lu during their tenure in the Allis lab to generate and validate a panel of clonal mMSC H3K36 methyltransferase knockout lines. NSD family enzymes have been shown to catalyze H3K36me2 at intergenic regions (Kuo et al. 2011) although their expression and relative contribution to cellular H3K36me2 levels vary between cell types. In mESCs, *Nsd1* is the major NSD enzyme expressed, while both *Nsd1* and *Nsd2* are expressed in mMSCs. Therefore, we genetically ablated both *Nsd1* and *Nsd2* (sgNsd1/2) in mMSCs or *Nsd1* alone (sgNsd1) in mESCs using CRISPR/Cas9 (**Fig. 2.6A,B**). *Setd2* was disrupted separately as a control (sgSetd2) (**Fig. 2.6A**). sgNsd1/2 and sgSetd2 mMSCs showed marked and specific depletion of H3K36me2 and H3K36me3 by immunoblotting, respectively (**Fig. 2.6C**). This finding was independently confirmed by quantitative mass-spectrometry measurement of the global abundance of histone PTMs in these lines by Dylan Marchione, a student in the laboratory of Dr. Benjamin Garcia at the University of Pennsylvania (**Fig. 2.6D**). To quantitatively analyze genome-wide changes between parental and H3K36me2/3-depleted cells, we then performed ChIP-seq with *Drosophila* chromatin spike-in as an exogenous reference control (Orlando et al. 2014). The Majewski group identified a similar global reduction of H3K36me2 in sgNsd1/2 mMSCs and sgNsd1 mESCs by this analysis (**Fig. 2.7A,B**). Furthermore,
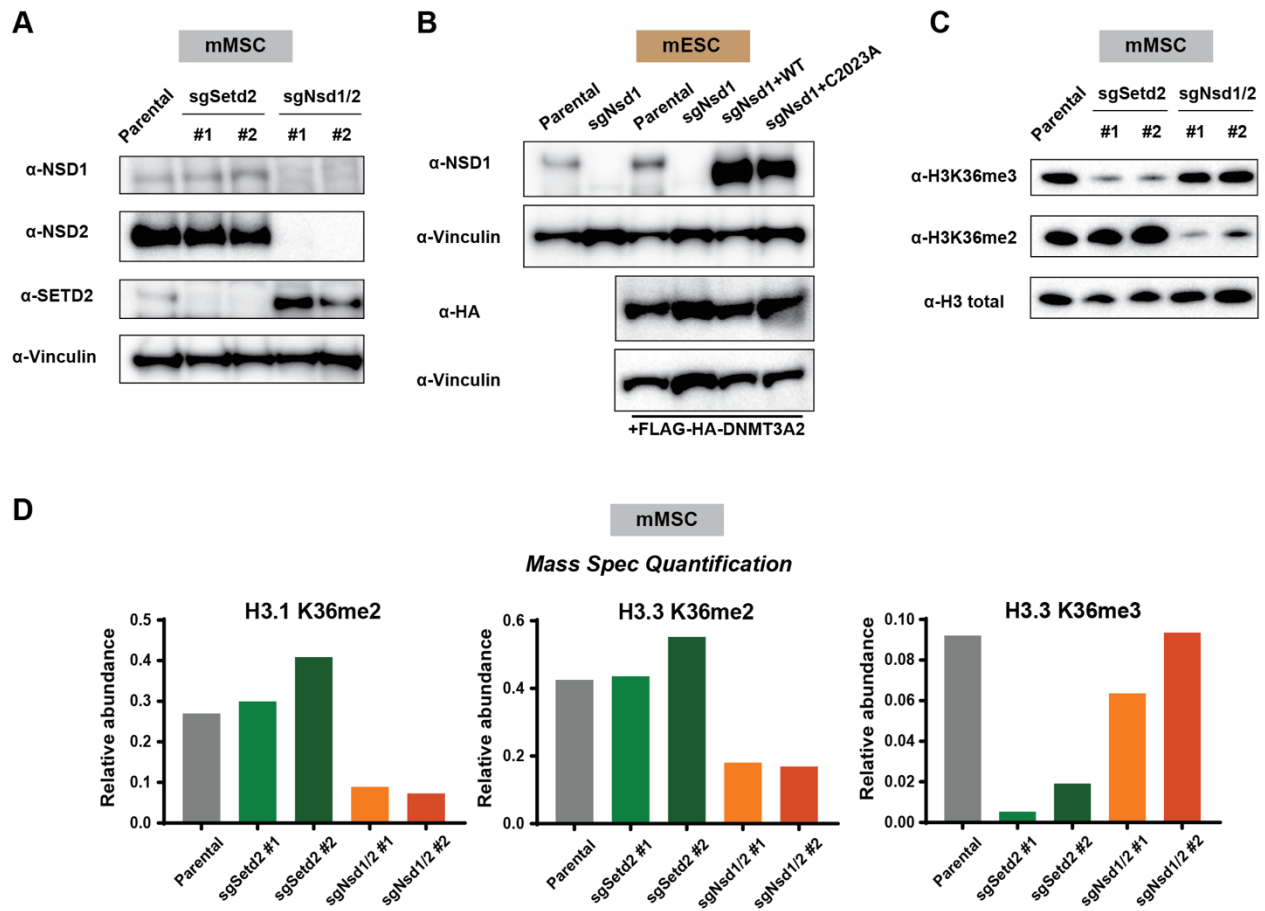
Figure 2.6: Genetic ablation of *Nsd1* and *Nsd2* in mouse MSCs and *Nsd1* in mouse ES cells. **A)** Immunoblots of lysates from parental and H3K36 methyltransferase knockout mMSC clonal lines for NSD1, NSD2, and SETD2. Vinculin was used as a loading control. **B)** Immunoblots of lysates from parental and sgNsd1 mESCs expressing HA-tagged DNMT3A. sgNsd1 cells were rescued with ectopic expression of wildtype (WT) or catalytic mutant (C2023A) NSD1. Vinculin was used as a loading control. **C)** Immunoblots of lysates generated from parental, sgSetd2, and sgNsd1/2 mMSCs for H3K36me3 and H3K36me2, with total H3 as a loading control. **D)** Quantitative mass spectrometry measurement of the abundance of histone PTMs in acid-extracted histones derived from indicated mMSC lines.
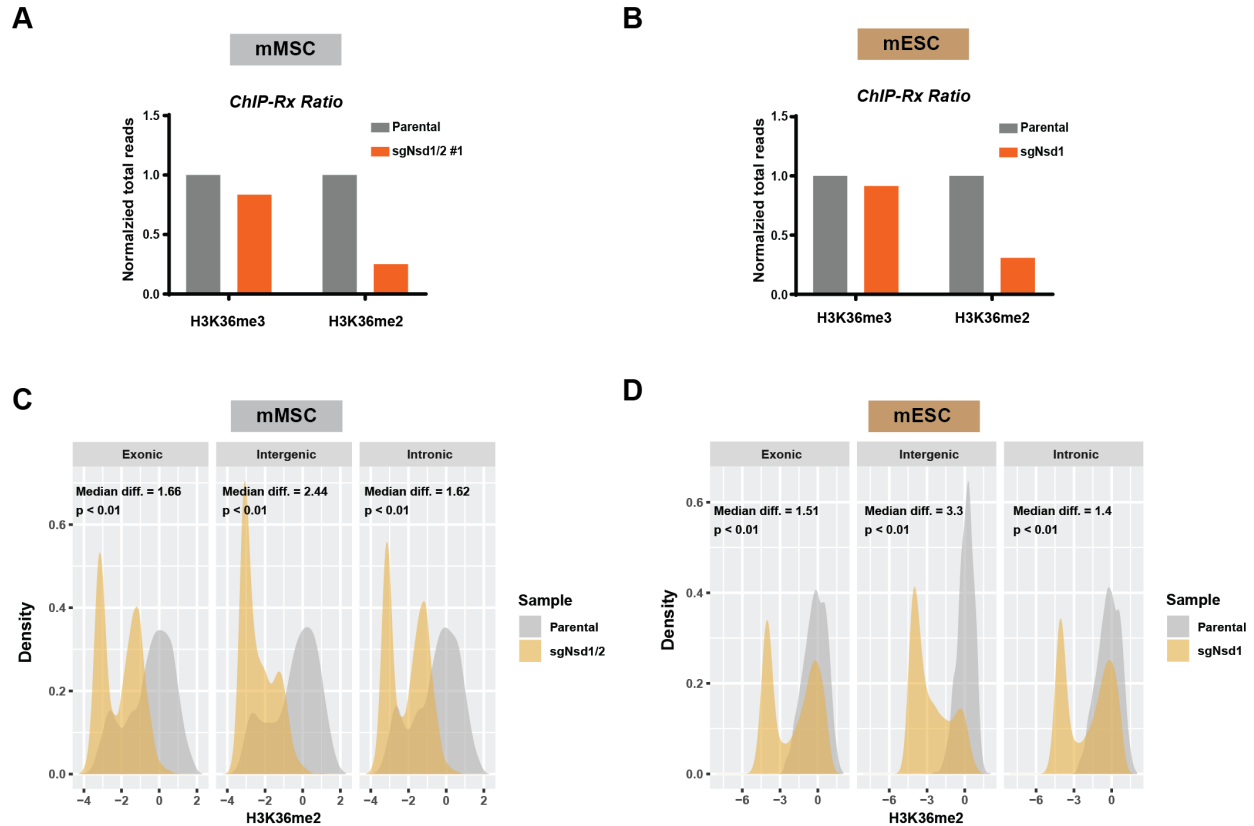
Figure 2.7: Depletion of intergenic H3K36me2 upon loss of NSD family enzymes.
A) Ratios of ChIP-seq reads for H3K36me2/3 in mMSCs between target chromatin (Mouse) and reference spike-in chromatin (*Drosophila*). **B)** Ratios of ChIP-seq reads for H3K36me2/3 in mESCs between target chromatin (Mouse) and reference spike-in chromatin (*Drosophila*). **C)** Density plots of H3K36me2 levels at intergenic (n = 1,165), exonic (n = 13,601), and intronic (n = 12,364) regions for parental (grey) and sgNsd1/2 (orange) mMSCs. Indicated p-values determined by Wilcoxon's rank sum test (two-sided). **D)** Density plots of H3K36me2 levels at intergenic (n = 1,165), exonic (n = 13,601), and intronic (n = 12,364) regions for parental (grey) and sgNsd1 (orange) mESCs. Indicated p-values determined by Wilcoxon's rank sum test (two-sided).

depletion of H3K36me2 was most prominent at intergenic regions in both lines (**Fig. 2.7C,D**). In contrast, levels of H3K36me3 were unperturbed and reductions in genic H3K36me2 were relatively modest whereas in sgSetd2 cells H3K36me3 levels were profoundly depleted as expected (**Fig. 2.6D**, **2.7A-D**). Taken together, the collective results from immunoblotting, mass-spectrometry, and ChIP-seq analysis suggested to us that H3K36 methyltransferase enzymes function distinctly at the chromatin level: NSD family enzymes deposit H3K36me2 at intergenic regions while SETD2 independently establishes H3K36me3 within actively transcribed gene bodies.

We next wished to profile DNMT3A binding patterns and DNA methylation in H3K36me2-depleted cells. To that end, I expressed HA-tagged DNMT3A1 in the mMSC clonal lines and DNMT3A2 in the mESCs clonal lines and performed ChIP-seq (**Fig. 2.8A,B**). We observed striking changes to the intergenic chromatin landscape as well as to DNMT3A localization patterns upon deletion of NSD family enzymes in both mESCs and mMSCs (**Fig. 2.8C,D**). Genome-wide analysis by the Majewski lab revealed that CpG hypomethylation in sgNsd1/2 mMSCs and sgNsd1 mESCs predominately affected intergenic regions (**Fig. 2.9A,B**). This was consistent with the loss of CpG methylation coinciding with decreases in H3K36me2 (**Fig. 2.9C,D**), which predominately is observed at intergenic regions in parental cells. Furthermore, loss of DNMT3A targeting at H3K36me2-depleted regions also coincided with reduced CpG methylation suggesting DNMT3A may require H3K36me2 for intergenic recruitment (**Fig. 2.10A,B**). To directly assess if NSD-catalyzed H3K36me2 is responsible for DNMT3A intergenic recruitment, I carried out rescue experiments in which I re-expressed either wild-type or catalytically-deficient (C2023A) NSD1 in sgNsd1 mESCs using a dual promoter Piggybac
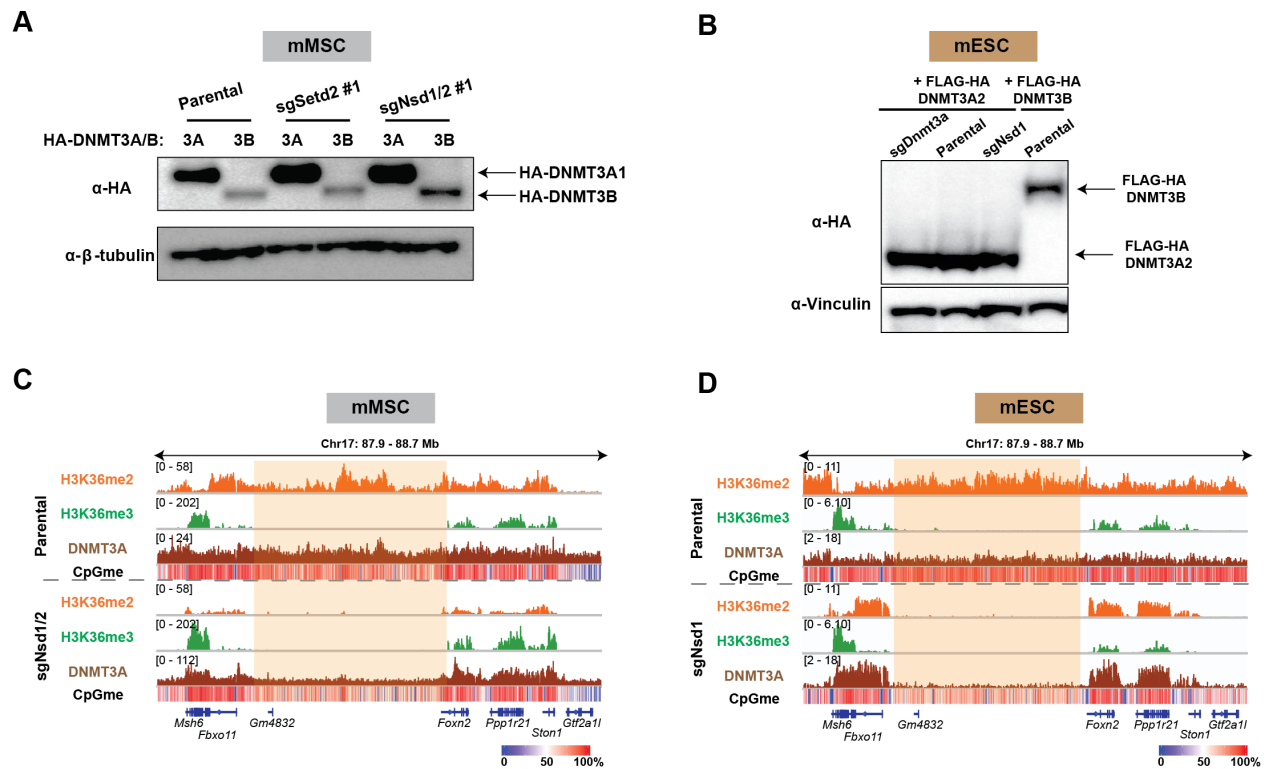
Figure 2.8: Profiling DNMT3A localization and CpG methylation in mouse MSCs and mouse ES cells depleted of intergenic H3K36me2. **A)** Immunoblots of lysates from parental, sgSetd2 and sgNsd1/2 mMSCs ectopically expressing HA-DNMT3A1 or -DNMT3B. β-tubulin was used as a loading control. **B)** Immunoblots of lysates from sgDnmt3a, parental, and sgNsd1 mESCs ectopically expressing HA-DNMT3A2 or -DNMT3B. Vinculin was used as a loading control. **C)** Genome browser representation of ChIP-seq normalized reads for H3K36me3, H3K36me2, and DNMT3A1 in parental and sgNsd1/2 mMSCs at Chr17: 87.9-88.7 Mb, as indicated. Levels of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high). Refseq genes are annotated at the bottom. Shaded area indicates H3K36me2-enriched intergenic region in parental cells. **D)** Genome browser representation of ChIP-seq normalized reads for H3K36me3, H3K36me2, and DNMT3A2 in parental and sgNsd1 mESCs at Chr17: 87.9-88.7 Mb, as indicated. Levels of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high). Refseq genes are annotated at the bottom. Shaded area indicates H3K36me2-enriched intergenic region in parental cells.
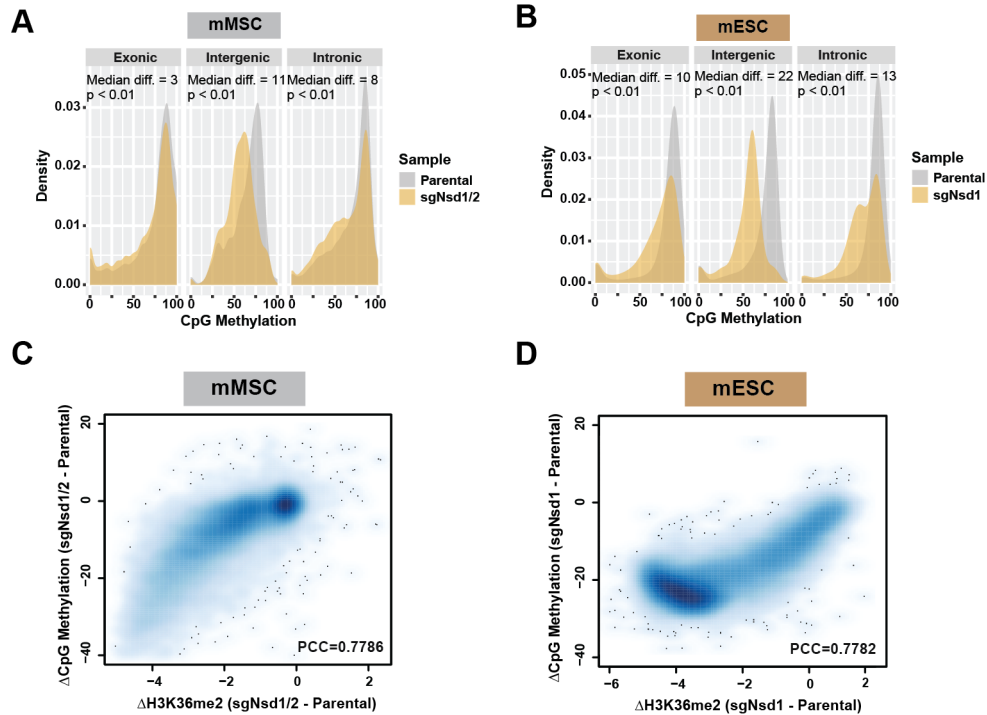
Figure 2.9: H3K36me2 depletion coincides with reductions in intergenic CpG methylation. A) Histograms for CpG methylation at intergenic (n = 1,165), exonic (n = 13,601), and intronic (n = 12,364) regions for parental (grey) and sgNsd1/2 (orange) mMSCs. Indicated p-values determined by Wilcoxon's rank sum test (two-sided). **B)** Histograms for CpG methylation at intergenic (n = 1,165), exonic (n = 13,601), and intronic (n = 12,364) regions for parental (grey) and sgNsd1 (orange) mESCs. Indicated p-values determined by Wilcoxon's rank sum test (two-sided). **C)** Percent change of averaged CpG methylation between parental and sgNsd1/2 mMSCs was plotted relative to changes in ChIP-seq normalized reads of H3K36me2 for 100kb non-overlapping bins (n = 25,611). Pearson's correlation coefficient is indicated. **D)** Percent change of averaged CpG methylation between parental and sgNsd1 mESCs was plotted relative to changes in ChIP-seq normalized reads of H3K36me2 for 100kb non-overlapping bins (n= 26,044). Pearson's correlation coefficient is indicated.

expression vector. Reintroduction of wild-type NSD1, but not the C2023A catalytic mutant form, was able to rescue global and intergenic H3K36me2 levels (**Fig. 2.6B, 2.10C,D**). Importantly, intergenic localization of DNMT3A could only be recovered through add-back of wild-type NSD1 as assessed by ChIP-qPCR of select genomic loci (**Fig. 2.10E**), demonstrating that NSD1-catalyzed H3K36me2 is required for recruitment of DNMT3A to non-coding euchromatic regions. Thus, we conclude that NSD1 plays an essential and specific role in directing DNMT3A to intergenic regions and maintaining CpG methylation at these sites.

We next sought to elucidate the mechanisms underlying the specificity of DNMT3A recruitment by H3K36me2. Both DNMT3A and DNMT3B contain PWWP 'reader' domains that can interact with H3K36 methylation *in vitro*, although their relative specificity for different valences of H3K36 methylation remains unclear (Baubec et al. 2015, Dhayalan et al. 2010). To quantitatively assess the specificity of DNMT3A$_{PWWP}$ towards nucleosomal substrates, we collaborated with EpiCypher Inc. (technician Matthew Marunde and Chief Scientific Officer Dr. Michael-Christopher Keogh). They examined interactions between the purified DNMT3A$_{PWWP}$ domain and a panel of semi-synthetic designer modified nucleosomes using an AlphaScreen proximity-based interaction assay (**Fig. 2.11A**). DNMT3A$_{PWWP}$ bound with highest affinity to nucleosomes modified with H3K36me2, followed by H3K36me3, but not H3K36me1 or any valence at H3K4, H3K9, H3K27, or H4K20 (**Fig. 2.11B,C**). To independently confirm these findings, we collaborated with Dr. Haitao Li's group at Tsinghua University (graduate student Yuan Yue) to perform quantitative isothermal titration calorimetry (ITC) assays with either H3.1 or H3.3 K36-modified peptides and DNMT3A$_{PWWP}$. These

Figure 2.10: NSD1-mediated H3K36me2 is required for intergenic DNMT3A localization. **A)** Difference in ChIP-seq normalized reads of DNMT3A1 between parental and sgNsd1/2 mMSCs was plotted relative to that of H3K36me2 for 10kb non-overlapping bins (n = 246,285). Each bin/dot was color-coded based on change of averaged CpG methylation to show lost (blue) or gained (red) CpG methylation in sgNsd1/2 cells. Pearson's correlation coefficient is indicated. **B)** Difference in ChIP-seq normalized reads of DNMT3A2 between parental and sgNsd1 mESCs was plotted relative to that of H3K36me2 for 10kb non-overlapping bins (n = 246,285). Each bin/dot was color-coded based on change of averaged CpG methylation to show lost (blue) or gained (red) CpG methylation in sgNsd1 cells. Pearson's correlation coefficient is indicated. **C)** Immunoblots of lysates generated from parental and sgNsd1 mESCs for H3K36me2, with total H3 as a loading control. sgNsd1 cells were rescued with ectopic expression of wild-type (WT) or C2023A catalytic mutant (Mut) NSD1. **D)** The enrichment (% input) of H3K36me2 at various intergenic regions in parental (black) and sgNsd1 (orange) mESCs rescued with ectopic expression of wild-type (WT) or C2023A catalytic mutant (Mut) NSD1 was measured with ChIP-qPCR. Each data point represents a genomic locus (n = 10 for H3K36me2-enriched regions, n = 4 for H3K36me2-depleted regions). Data are mean ± SD. Indicated p-values were determined by one-way ANOVA. **E)** Fold enrichment of DNMT3A at various H3K36me2-enriched versus H3K36me2-depleted intergenic regions in parental (black) and sgNsd1 (orange) mESCs rescued with ectopic expression of wild-type (WT) or C2023A catalytic mutant (Mut) NSD1 was measured with ChIP-qPCR. Each data point represents a genomic locus (n = 10). Data are mean ± SD. Indicated p-values determined by one-way ANOVA.

A mMSC

ΔDNMT3A (sgNsd1/2 - Parental)

PCC = 0.43

ΔCpG methylation
(sgNsd1/2-Parental)
-60% -30% 0% 30%

ΔH3K36me2 (sgNsd1/2 - Parental)

B mESC

ΔDNMT3A (sgNsd1 - Parental)

PCC = 0.84

ΔCpG methylation
(sgNsd1-Parental)
-60% -30% 0% 30%

ΔH3K36me2 (sgNsd1 - Parental)

C mESC

NSD1 rescue: Parental, sgNsd1 - , +WT, +Mut

α-H3K36me2

α-H3 total

D mESC

H3K36me2 ChIP (% input)

P < 0.001
P = 0.568
P < 0.001

P < 0.001
P = 0.986
P < 0.001

NSD1 rescue: - , - , +WT, +Mut

H3K36me2-enriched loci   H3K36me2-depleted loci

E mESC

DNMT3A ChIP Fold Enrichment

P < 0.001
P = 0.456
P < 0.001

NSD1 rescue: - , - , +WT, +Mut

Parental   sgNsd1

43
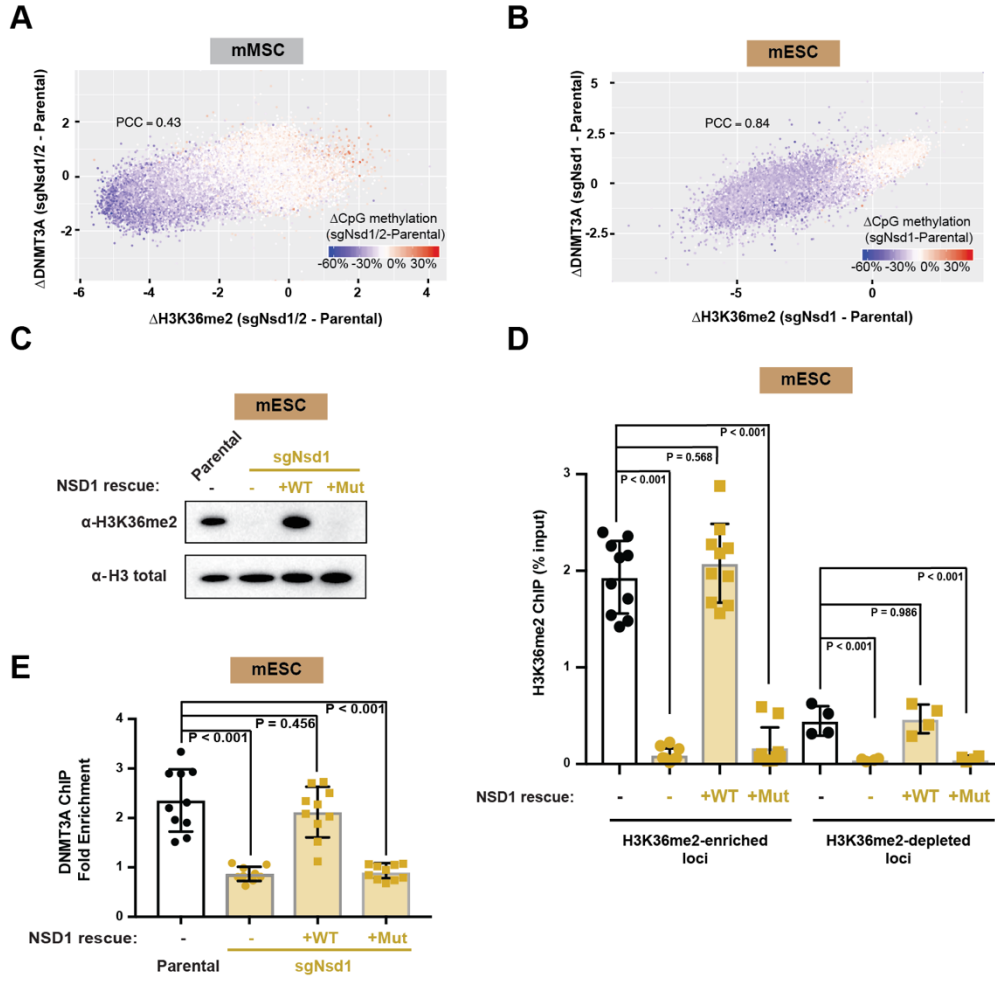
results further supported the preferential recognition of H3K36me2/3 by DNMT3A$_{PWWP}$ and indicated that binding is not substantially different between the canonical (H3.1) or variant (H3.3) forms of histone H3 (**Fig. 2.11D,E**). Taken together, these findings indicate that DNMT3A$_{PWWP}$ recognizes both methylation states but exhibits greater affinity for H3K36me2.

We speculated that such a binding preference, combined with the relative abundance of H3K36me2 compared to H3K36me3 (**Fig. 2.6D**), contributes to the favored enrichment of DNMT3A at intergenic regions. A prediction of this hypothesis is that DNMT3A will be recruited to H3K36me3-enriched gene bodies through its PWWP domain upon global depletion of H3K36me2. To test this, we examined the localization patterns of DNMT3A in H3K36me2-depleted cells in which intergenic recruitment has been abrogated. Promisingly, we noted a marked redistribution of DNMT3A to gene bodies in sgNsd1/2 mMSCs (**Fig. 2.12A**). This could not be explained through its interaction with residual genic H3K36me2, as DNMT3A now exhibited a binding profile reminiscent of DNMT3B in which signal is enriched toward 3' end of the gene body, mirroring the distribution of H3K36me3 (**Fig. 2.12B**). Moreover, while DNMT3A was depleted at regions of high H3K36me3 in parental mMSCs, in sgNsd1/2 mMSCs it was enriched at these same regions (**Fig. 2.12C**). We wished to further assess whether DNMT3A$_{PWWP}$ recognition of H3K36me3 was required for genomic re-targeting of DNMT3A in H3K36me2-depleted cells. To that end, we first disrupted SETD2 catalytic activity in sgNsd1/2 cells to create triple knockout cells (TKO) that are depleted of both H3K36me2 and H3K36me3 and observed that these cells had impaired DNMT3A localization to gene bodies (**Fig. 2.13A,B**). To independently confirm these results, we

Figure 2.11: Preferential recognition of H3K36me2 and H3K36me3 by the PWWP domain of DNMT3A. **A)** The *dCypher* approach to interrogate chromatin readers (see Methods). Biotinylated nucleosomes are immobilized on Streptavidin donor beads and GST-tagged DNMT3A$_{PWWP}$ on glutathione AlphaLISA acceptor beads. Laser excitation of the donor generates singlet oxygen that diffuses to activate emission from the acceptor: fluorescence counts are directly proportional to the amount of donor-acceptor bridged by the nucleosome-reader interaction. **B)** AlphaLISA counts for GST-DNMT3A$_{PWWP}$ interaction with semi-synthetic modified nucleosomes from two replicates. **C)** Quantification of AlphaLISA counts for isolated GST-DNMT3A$_{PWWP}$ interaction titrated against H3K36-modified nucleosomes (me0/me1/me2/me3) from mean of two replicates. **D)** ITC titration and fitting curves of human DNMT3A PWWP domain with H3.3(1-42) K36-modified peptides (me0/me1/me2/me3). **E)** ITC titration and fitting curves of human DNMT3A PWWP domain with H3.1(1-42) K36-modified peptides (me0/me1/me2/me3).
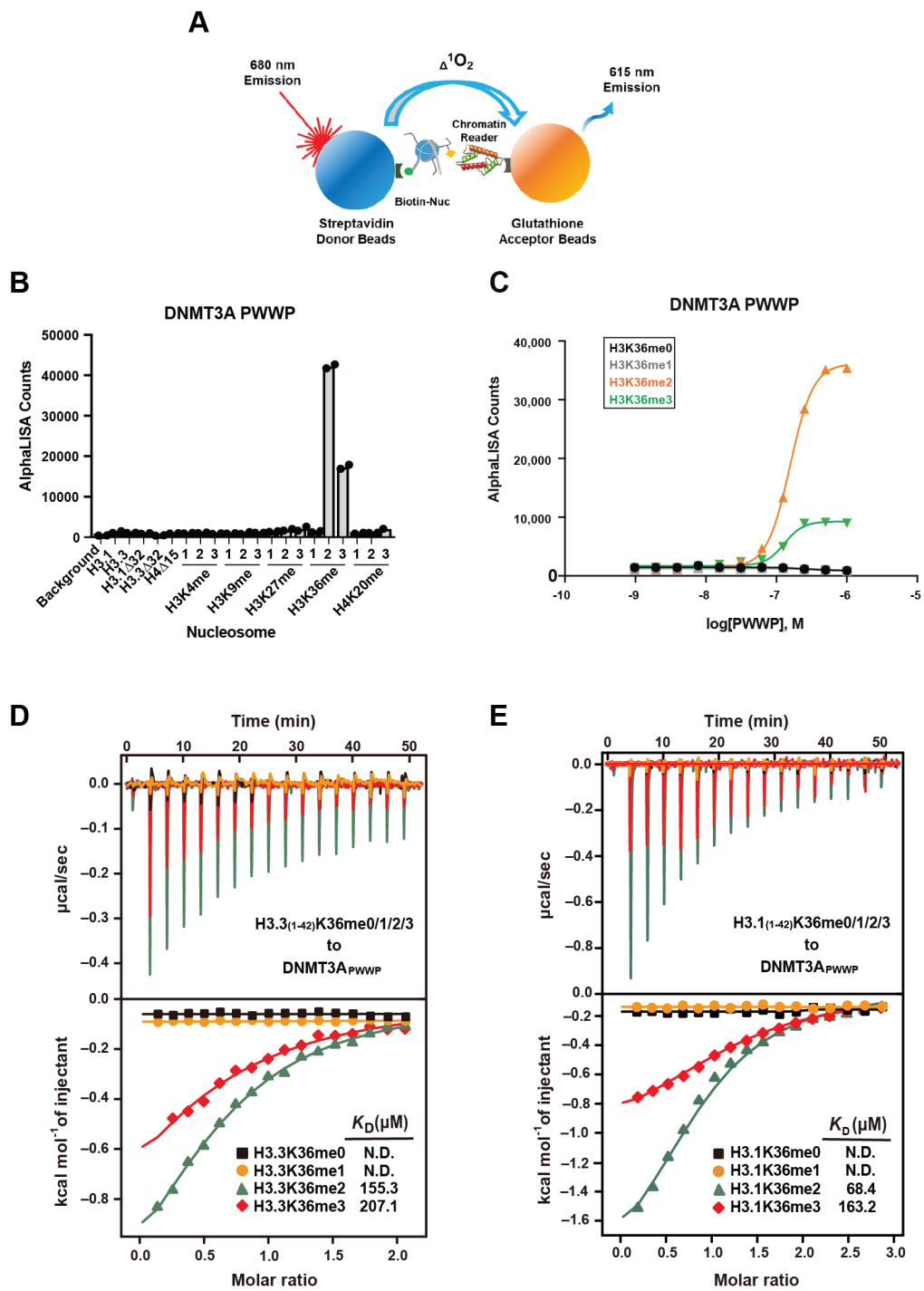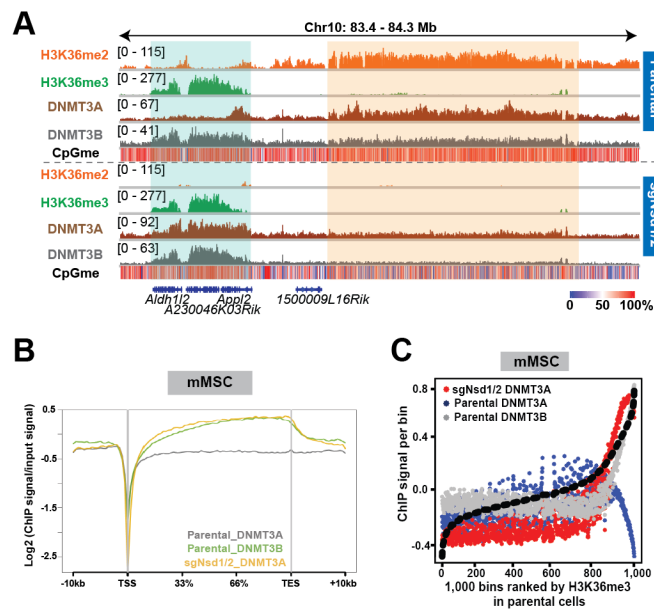
Figure 2.12: Redistribution of DNMT3A to H3K36me3-marked regions after loss of H3K36me2. **A)** Genome browser representation of ChIP-seq normalized reads for H3K36me2, H3K36me3, DNMT3A1, and DNMT3B in parental and sgNsd1/2 mMSCs at Chr10: 83.4-84.3 Mb. Levels of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high). Refseq genes are annotated at the bottom. Shaded areas indicate H3K36me2-enriched intergenic regions (orange) and H3K36me3-enriched genic regions (green) in parental cells. Upon H3K36me2 depletion, DNMT3A is re-targeted to H3K36me3-enriched gene bodies (eg. *Aldh1l2* and *Appl2*). **B)** Averaged ChIP-seq normalized signal across all gene bodies represented as log2 fold change over input for DNMT3A1 (grey, n = 14,959) and DNMT3B (green, n = 14,959) in parental mMSCs, and for DNMT3A1 (orange, n = 14,311) in sgNsd1/2 mMSCs. **C)** ChIP-seq normalized reads per bin for DNMT3A1 (blue) and DNMT3B (grey) in parental mMSCs and DNMT3A1 (red) in sgNsd1/2 mMSCs relative to H3K36me3. To generate bins, 1kb genomic tiles (n = 2,462,755) were ranked by H3K36me3 enrichment in mMSCs and grouped into 1000 rank-ordered bins (2,463 tiles per group). Dashed line indicates H3K36me3 enrichment per bin.

introduced a point mutation (D333A) in the PWWP domain that impairs DNMT3A binding to both H3K36me2 and H3K36me3 (Dhayalan et al. 2010) and found that it abrogated DNMT3A re-localization to H3K36me3-enriched regions in sgNsd1/2 cells (**Fig. 2.13C,D**). Taken together, these results indicate that DNMT3A redistribution to gene bodies is PWWP-dependent. We therefore conclude that preferential recognition of H3K36me2/3 by the PWWP domain guides DNMT3A localization across the cellular chromatin landscape.

We next asked whether a similar redistribution mechanism may occur for DNMT3B. As expected from the localization pattern of DNMT3B in cells, DNMT3B$_{PWWP}$ preferentially bound to H3K36me3 recombinant nucleosomes *in vitro* as assessed by Dr. Li's group (**Fig. 2.14A,B**). Furthermore, loss of gene-body localization for DNMT3B in H3K36me3-depleted sgSetd2 mMSCs was accompanied by a relatively modest increase in retargeting to H3K36me2-enriched intergenic regions (**Fig. 2.14C,D**). This suggests that the recognition of H3K36me2 by the PWWP domain of DNMT3B may not be of high enough affinity to promote redistribution of the enzyme. In aggregate, these data support H3K36me2/3 ratio as a key determinant for specifying targeting patterns of DNMT3A, and of DNMT3B to a lesser extent.

Having established the mechanistic basis for molecular "cross-talk" between intergenic DNA methylation and H3K36me2, we sought to explore the disease relevance of our findings. Previous work led by Dr. Simon Papillon-Cavanagh and Dr. Chao Lu reported that a subset of head and neck squamous cell carcinomas (HNSCCs) depleted of H3K36me2 due to genetic or biochemical inactivation of NSD1 shows a DNA hypomethylation signature (Papillon-Cavanagh et al. 2017). Similarly, another

Figure 2.13: DNMT3A PWWP binding to H3K36me3 is required for redistribution to gene bodies after loss of H3K36me2. **A)** Genome browser representation of ChIP-seq normalized reads for H3K36me2, H3K36me3, and DNMT3A1 in parental, sgNsd1/2, and TKO mMSCs at Chr10: 74.8-77.8 Mb, as indicated. Refseq genes are annotated at the bottom. Shaded areas indicate H3K36me3-enriched genic regions in parental cells. **B)** Quantification of ChIP-seq normalized reads of DNMT3A1 in parental, sgNsd1/2, and TKO mMSCs at 10kb non-overlapping bins enriched for H3K36me3 in parental cells (top 20% of bins, n = 54,624). Reads were normalized by read depth. P-values were $< 2.2 \times 10^{-16}$ for all pair-wise comparisons as determined by Wilcoxon's rank sum test (two-sided). Boxes represent median and 25th-75th percentiles, whiskers are minimum to maximum with discrete points representing outliers. **C)** Immunoblots of lysates from parental, sgNsd1/2, and TKO mMSCs ectopically expressing HA-tagged wildtype (WT) or D333A mutant DNMT3A1. β-actin was used as a loading control. **D)** Fold enrichment of wildtype (WT) or PWWP-mutated (D333A) DNMT3A1 at gene body versus intergenic regions in parental (black) and sgNsd1/2 (orange) mMSCs was measured with ChIP-qPCR. Each data point represents a genomic locus (n = 10). Data are mean ± SD. Indicated p-values determined by one-way ANOVA.

A Chr10: 74.8 - 77.8 Mb

B

mMSC

Normalized DNMT3A ChIP-seq by read depth

Pa | sgNsd1/2 | TKO

C

mMSC

DNMT3A: - | WT | D333A | WT

α-HA

α-β-actin

D

mMSC

DNMT3A ChIP Fold Enrichment

P < 0.001
P = 0.699
P = 0.224

HA-DNMT3A: Parental | sgNsd1/2 | Parental | sgNsd1/2
WT | PWWP Mutant
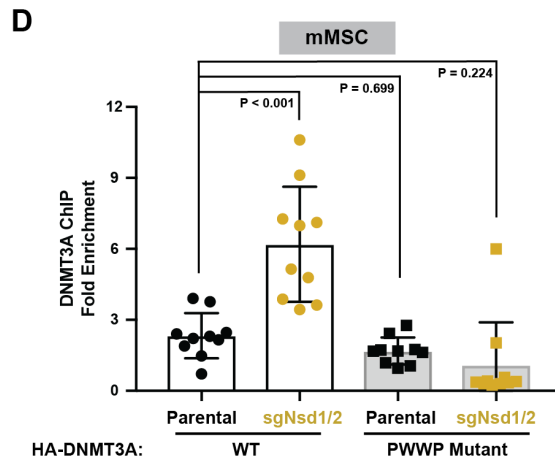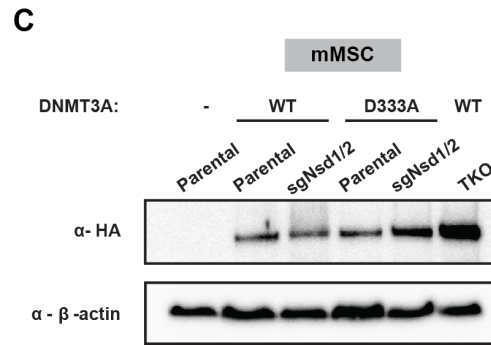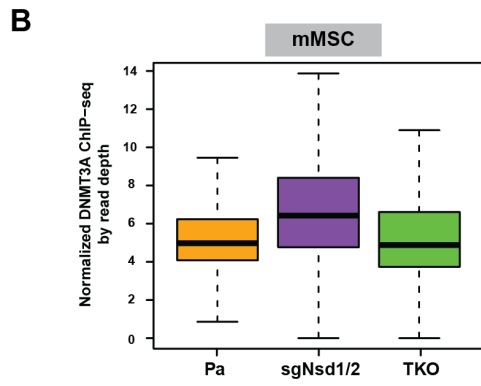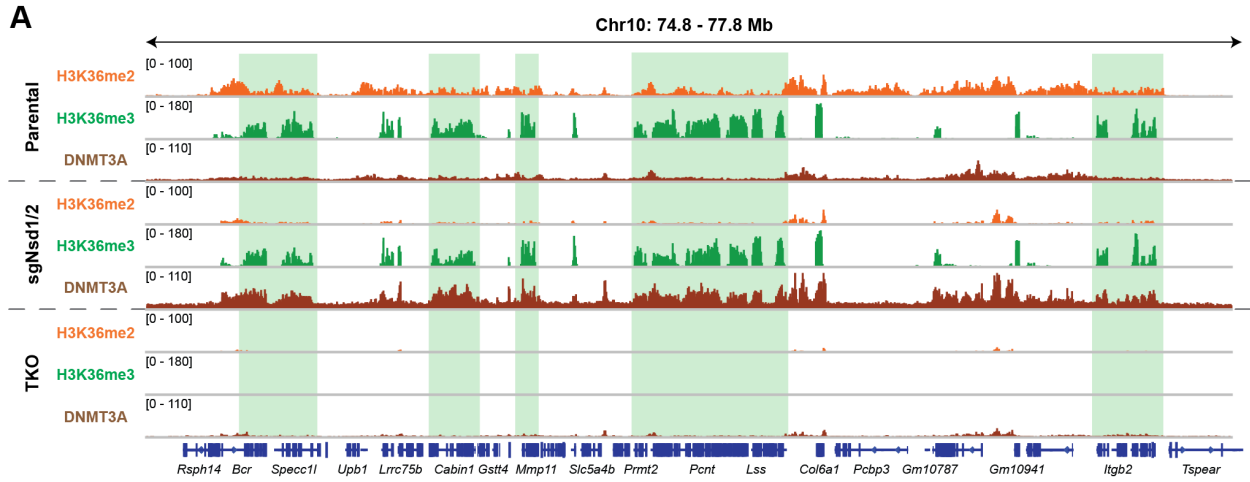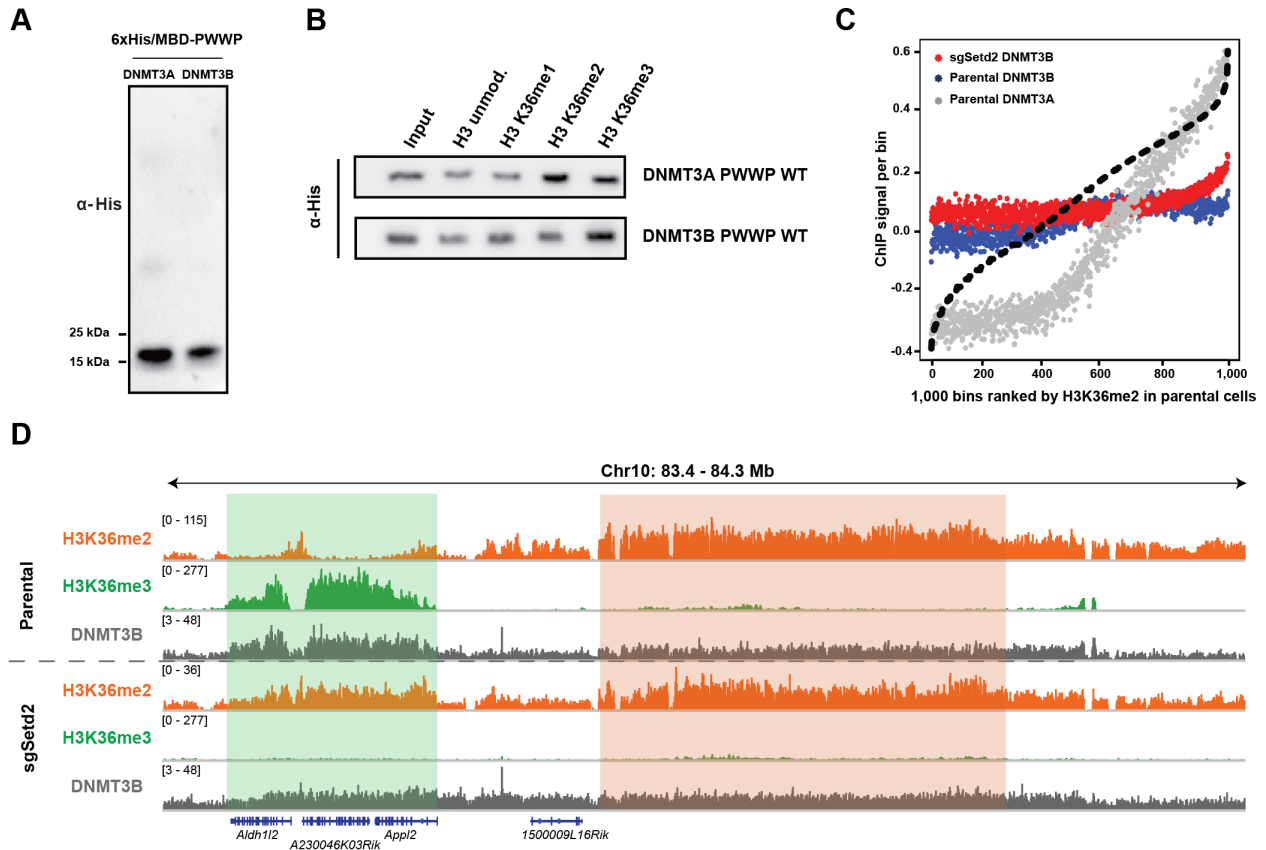
Figure 2.14: Assessment of interaction between DNMT3B and H3K36me2 *in vitro*
and in cells. **A)** Immunoblots of His/MBP-tagged recombinant PWWP domains of
DNMT3A and DNMT3B. **B)** Immunoblots of recombinant His/MBP-tagged wild-type
DNMT3A and DNMT3B PWWP domains bound to H3K36-modified recombinant
nucleosomes following the *in vitro* pull-down assay. **C)** ChIP-seq normalized reads per
bin for DNMT3A1 (grey) and DNMT3B (blue) in parental mMSCs and DNMT3B (red) in
sgSetd2 mMSCs relative to H3K36me2. To generate bins, 1kb genomic tiles were
ranked by H3K36me2 enrichment in parental mMSCs and grouped into 1000 rank-
ordered bins. Dashed line indicates H3K36me2 enrichment per bin. **D)** Genome
browser representation of ChIP-seq normalized reads for H3K36me2, H3K36me3, and
DNMT3B in parental and sgSetd2 mMSCs at Chr10: 83.4-84.3 Mb, as indicated. Refseq
genes are annotated at the bottom. Shaded areas indicate H3K36me2-enriched
intergenic regions (orange) and H3K36me3-enriched genic regions (green) in parental
cells.

51

group previously reported a significant decrease in DNA methylation in blood samples from Sotos patients with germline *NSD1* haploinsufficiency relative to controls (Choufani et al. 2015). However, it remained unclear where in the genome the DNA hypomethylation was occurring in either of these contexts. Hypothesizing that it may be at intergenic regions depleted of H3K36me2, we first profiled H3K36me2 and CpG methylation in patient-derived HNSCC cell lines carrying wild-type (Cal27 and Fadu) or mutant *NSD1* (SCC-4 and SKN-3). Despite their heterogeneous genetic background, DNA hypomethylation in *NSD1*-mutant cell lines tracked closely with genome-wide reductions in H3K36me2 as compared to *NSD1*-wild-type cell lines and occurred primarily at intergenic regions (**Fig. 2.15A,B**). With the help of Anissa Djedid and others in the Majewski lab, we next further analyzed publicly available DNA methylation arrays of HNSCC or Sotos patient samples. Probes that were hypomethylated in NSD1-inactivated compared to NSD1 wild-type HNSCC tumors were significantly enriched for intergenic regions (**Fig. 2.15C**). Similarly, Sotos syndrome patients also exhibited an enrichment of hypomethylated probes at intergenic regions compared to controls (**Fig. 2.15C**). We propose that the DNA hypomethylation observed in these publicly available datasets is specifically occurring at intergenic regions depleted of H3K36me2 due to loss of NSD1 function.

As noted previously, missense mutations in *DNMT3A* cause Tatton-Brown-Rahman syndrome (TBRS), a developmental disorder that shares many clinical features with Sotos syndrome including skeletal overgrowth, facial dysmorphism and intellectual disability (Tatton-Brown et al. 2014, Kurotaki et al. 2002, Tatton-Brown et al. 2017). We wondered whether TBRS patients also have disrupted recruitment of DNMT3A to
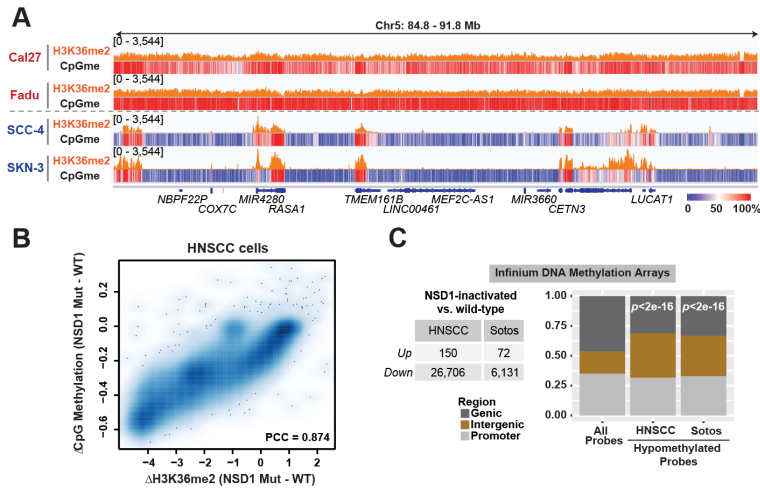
Figure 2.15: Impaired intergenic DNMT3A localization and CpG methylation in neoplastic and developmental overgrowth. **A)** Genome browser representation of ChIP-seq normalized reads for H3K36me2 in *NSD1* wild-type (Cal27, Fadu) and *NSD1* mutant (SCC-4, SKN-3) HNSCC cell lines at Chr5: 84.8-91.8 Mb. Levels of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high). Refseq genes are annotated at the bottom. **B)** Percent changes of averaged CpG methylation between *NSD1* wildtype and mutant HNSCC cell lines were plotted relative to changes in ChIP-seq normalized reads of H3K36me2 for 100kb non-overlapping bins (n = 28,395). Pearson's correlation coefficient is indicated. **C)** Left, table summarizing number of up- and down-regulated Infinium 450K DNA methylation array probes between NSD1 inactivated vs. NSD1 wild-type HNSCC tumor samples, and Sotos vs. control patient samples. Right, bar graph showing intergenic enrichment of DNA hypomethylated probes relative to all probes (n = 370,898). Indicated p-values determined by chi-squared test.

H3K36me2-enriched intergenic regions, leading to similar intergenic depletion of DNA

methylation as observed in Sotos syndrome patients. To that end, we characterized

TBRS-associated point mutations within the DNMT3A PWWP domain (W297del, I310N,

Y365C) (**Fig. 2.16A**) in collaboration with Dr. Haitao Li's group using recombinantly

expressed DNMT3A PWWP domains carrying these mutations (**Fig. 2.16B**). All

mutations tested showed reduced nucleosome binding *in vitro* (**Fig. 2.16C**), suggesting

that they may impair chromatin recruitment of DNMT3A.

Consistent with such an impairment, we found that W297del and I310N

mutations reduced the association of DNMT3A with bulk chromatin in cells (**Fig. 2.17A**),

which was accompanied by reduced protein levels as recently reported (Heyn et al.

2019) (**Fig. 2.17B**). Whether the reduced protein stability of TBRS-associated mutants

is a cause or consequence of their impaired "reader" function remains unclear. Further

examination of chromatin-bound wild-type, W297del, I310N or Y365C mutant DNMT3A

revealed a substantial decrease in the levels of H3K36me2 on nucleosomes in complex

with mutant DNMT3A (**Fig. 2.17C**). Accordingly, ChIP-seq analysis showed that the

I310N mutation abrogated targeting of DNMT3A to H3K36me2 genome-wide (**Fig.

2.17D,E**). These results suggest that even mutant DNMT3A that manages to associate

with bulk chromatin fails to be targeted to H3K36me2-enriched genomic regions. Thus,

impaired recruitment of DNMT3A and reduced CpG methylation at H3K36me2-enriched

intergenic euchromatic regions appear to be common features associated with TRBS

and Sotos syndrome. In support of this notion, unsupervised hierarchical clustering

performed by Dr. Chao Lu's group of published DNA methylation array profiles

(Choufani et al. 2015, Jeffries et al. 2019) indicated that patients with Sotos syndrome,
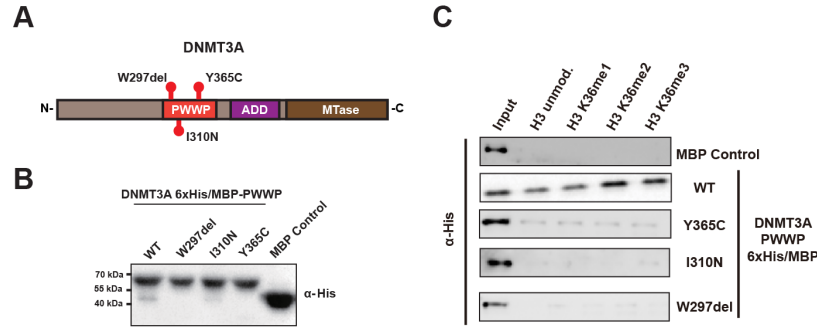
Figure 2.16: TBRS-associated mutations abrogate binding to H3K36me2 *in vitro*.
A) Schematic of DNMT3A conserved structural domains with indicated TBRS-associated mutations. **B)** Immunoblots of recombinant His/MBP-tagged DNMT3A PWWP domain containing TBRS-associated mutations. MBP alone was used as a control. **C)** Immunoblots of recombinant His/MBP-tagged DNMT3A wild-type and mutant PWWP domains (Y365C, I310N, and W297del) bound to H3K36-modified recombinant nucleosomes following the *in vitro* pull-down assay.

Figure 2.17: TBRS-associated mutations in DNMT3A are loss-of-function. **A)** Immunoblots of soluble or chromatin-associated lysates generated from cells ectopically expressing HA-tagged wildtype (WT), or mutant (W297del or I310N) DNMT3A. β-tubulin and LaminB1 were used as loading controls for the soluble and chromatin-associated fractions, respectively. **B)** Immunoblots of lysates from parental mMSCs ectopically expressing HA-tagged wildtype (WT) or TBRS-mutant DNMT3A from one experiment. β-tubulin was used as a loading control. **C)** Immunoblots of nucleosomes bound to HA-tagged wildtype (WT) or mutant (W297del, I310N, or Y365C) DNMT3A after ChIP with anti-HA tag antibodies. Total H3 was used as a loading control and to normalize for differences in protein expression and nucleosome pull-down efficiency between samples. **D)** Genome browser representation of ChIP-seq normalized reads for H3K36me2, DNMT3A1 wildtype (WT), and DNMT3A1 I310N mutant in mMSCs at Chr13: 62.2-66.8 Mb. Levels of CpG methylation are depicted as a heat map (blue/low; white/intermediate; red/high). Refseq genes are annotated at the bottom. **E)** ChIP-seq normalized reads of DNMT3A1 wild-type or PWWP mutant (I310N) were plotted relative to that of H3K36me2 for 100kb non-overlapping bins (WT: n = 25,694; I310N: n = 25,757). Pearson's correlation coefficient is indicated.

A

HA-DNMT3A: WT   W297del   I310N

Fraction: S C   S C   S C

α-HA

α-LaminB

α-β-tubulin

B

Whole Cell Lysate

WT W297del I310N Y365C

α-HA

α-β-tubulin

C

ChIP: α-HA

HA-DNMT3A: WT W297del I310N Y365C

α-H3K36me2

α-H3 total

Immunoblotting

D

Chr13: 62.2 - 66.8 Mb

H3K36me2 [0 - 47]

DNMT3A WT [0 - 47]

DNMT3A I310N [0 - 47]

CpGme

Gm3604 Mir713   Ptch1 Ercc6l2   Ctsl   Cntnap3 Mfsd14b   Gm10324

0   50   100%

E

DNMT3A WT

DNMT3A normalized reads

H3K36me2 normalized reads

PCC = 0.72

DNMT3A I310N

DNMT3A normalized reads

H3K36me2 normalized reads

PCC = -0.45

57

but not EZH2 mutation-associated Weaver syndrome (Tatton-Brown et al. 2017), shared

a similar DNA methylome with TBRS patients (**Fig. 2.18A**). This suggests intergenic

DNA hypomethylation is not a common feature of developmental overgrowth disorders

but rather reflects the common molecule convergence caused by disruption of a NSD1-

DNMT3A *trans*-chromatin pathway in both TBRS and Sotos syndrome.

DISCUSSION

Our findings provide a more comprehensive framework for understanding how DNA

methylation landscapes are established and maintained in euchromatin. Selective

targeting of DNMT3A and DNMT3B, guided by their PWWP chromatin reader domains,

directs CpG methylation to H3K36me2-enriched intergenic non-coding regions and

H3K36me3-enriched gene bodies, respectively (**Fig. 2.18B**). Simultaneously, the

presence of H3K4me3 at active promoters disrupts the interaction between the ADD

domain of DNMT3A/B and histone H3, thereby preventing CpG methylation at these

sites (Ooi et al. 2007, Otani et al. 2009). We favor a model in which the PWWP and

ADD domains work in tandem to direct the chromatin recruitment and licensing of

enzymatic activity of *de novo* DNMTs. In addition, promoters containing CpG islands are

further protected from DNA methylation through the recruitment of several CXXC

domain-containing proteins that preferentially recognize CpG-dense regions, including

TET family enzymes involved in active DNA demethylation and the H3K36 demethylase

KDM2B (Williams et al. 2011, Boulard et al. 2015). In such a manner, we propose that

CpG methylation is established across broad euchromatic regions through the

combined actions of DNMT3A/B while focal sites of transcriptional initiation are spared.
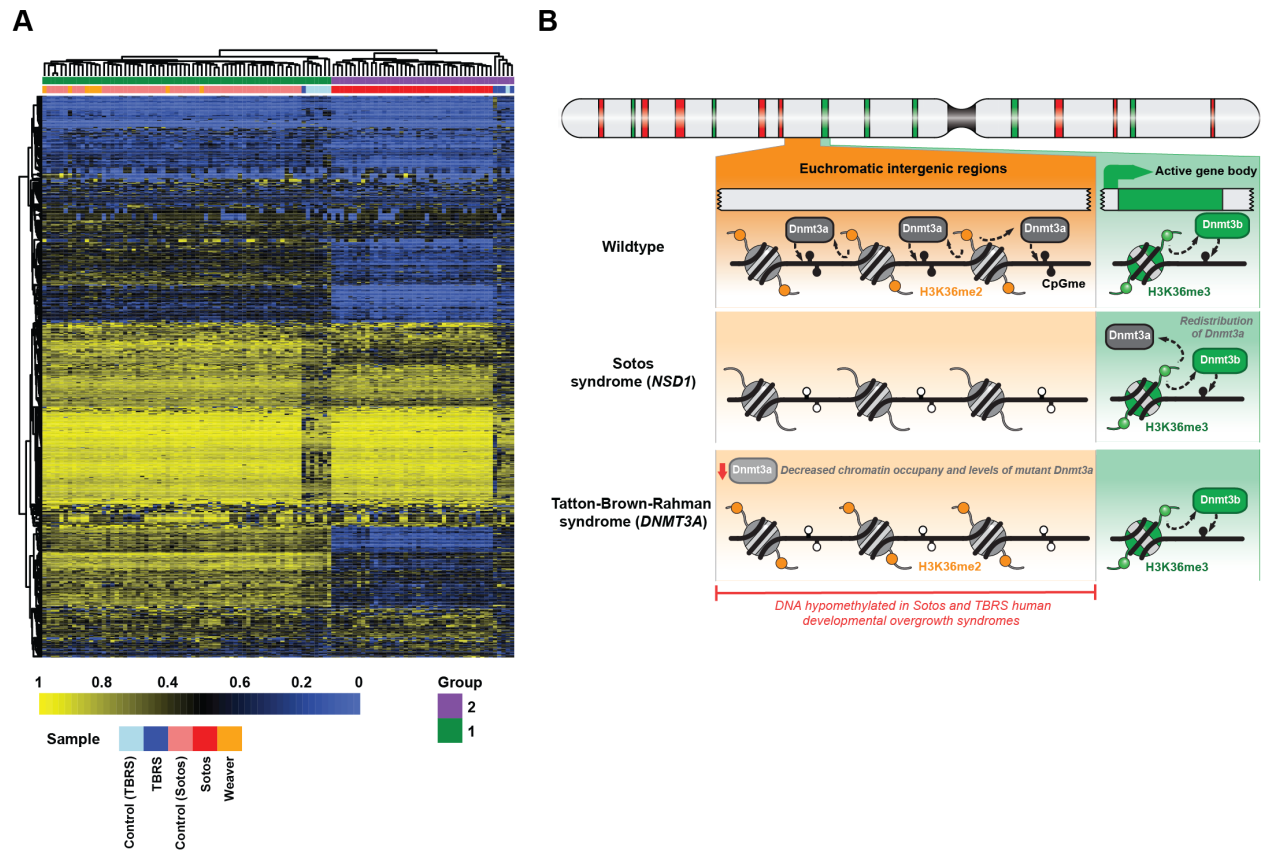
Figure 2.18: TBRS and Sotos syndrome patients share a similar DNA hypomethylation signature. **A)** Unsupervised hierarchical clustering of publically available Infinium Human Methylation 450K array profiles of blood samples from TBRS, Sotos and Weaver syndrome patients and controls, based on top 1000 most variable probes. **B)** Model depicting chromatin landscape changes in Sotos and Tatton-Brown-Rahman (TBRS) human developmental overgrowth syndromes. In normal development, DNMT3A and DNMT3B act in parallel to methylate CpG dinucleotides at H3K36me2-enriched intergenic and H3K36me3-enriched genic regions, respectively. Haploinsufficiency of *NSD1* and depletion of intergenic H3K36me2 levels in Sotos syndrome abrogates PWWP-mediated intergenic recruitment of DNMT3A, leading to intergenic DNA hypomethylation and DNMT3A redistribution to H3K36me3-enriched gene bodies. Mutations within the PWWP domain of DNMT3A in TBRS impair chromatin occupancy and reduce cellular levels of the protein, thereby also resulting in intergenic DNA hypomethylation.

However, the occurrence of ectopic DNA methylation at CpG islands in tumor samples (Toyota et al. 1999, Weisenberger et al. 2006) suggests mechanisms may exist to target *de novo* DNMT activity to sites of transcriptional regulation, a possibility that will be elaborated upon in the next chapter.

Our work also highlights the need for further structural characterization of DNMT3A/B bound to H3K36me2/3 to determine the basis of selective recognition by their PWWP domains. Such understanding may prove critical for further insight into the chromatin recruitment mechanisms underlying the more than two dozen PWWP domain-containing proteins in humans, which are implicated in a variety of cellular processes including chromatin remodeling, mismatch repair of DNA, and histone acetylation (Qin & Min 2014). The ability of these proteins to discriminate between H3K36me2 and H3K36me3 using their PWWP domains has largely not been tested using nucleosomal substrates. Future characterization of the biochemical interactions of PWWP domains should ideally rely upon modified nucleosome (as opposed to peptide) substrates, as PWWP domains are one of the few classes of chromatin reader domains to bind both modified histones and DNA simultaneously (Weaver et al. 2018).

Multivalent binding to both modified histones and DNA enables PWWP domains to achieve high binding affinity for their substrate but may also spatially constrain potential interactions with histone PTMs. This feature was recently highlighted by the first reported structure of a PWWP domain in complex with a nucleosome, which was solved for the human transcriptional coregulator LEDGF with H3K36me3 using cryo-electron microscopy (Wang et al. 2020). Beyond capturing the well-characterized

binding of an aromatic cage in the PWWP domain to the methyl-lysine residue of the

histone H3 tail, the structure also revealed how two positively-charged patches on the

surface of the domain to either side of the H3K36me3-binding pocket coordinate and

interact with both DNA gyres that wrap around the histone octamer. Such a mode of

binding is likely conserved among PWWP domain-containing proteins (Wang et al.

2020) and the positively charged patches may serve as interfaces for regulation, such

as through post-translational modifications.

Looking beyond regulation of DNMT3A/B, our findings also do not exclude the

possibility that DNMT1 activity is also regulated by H3K36 methylation. This could help

to explain reductions in intergenic DNA methylation levels in H3K36me2-depleted cells.

Such an observation would not be predicted to occur solely on the basis of impaired

recruitment of DNMT3A and/or DNMT3B, as they together play a minimal role in DNA

methylation maintenance compared to DNMT1. Indeed, UHRF1 is an obligate co-factor

for DNMT1 methylation maintenance activity and has been observed to be enriched at

non-genic euchromatic regions (Rajakumara et al. 2011). Furthermore, UHRF1 contains

a Tudor chromatin reader domain that is structurally related to PWWP family domains

and similarly recognizes both modified histones and DNA in tandem (Weaver et al.

2018). While some Tudor domains recognize H3K36 methylation, others do not (Lu &

Wang 2013), and that of UHRF1 remains incompletely characterized with its binding

preference to date only being assessed for histone peptide substrates, as opposed to

more physiologic nucleosomal substrates (Arita et al. 2012).

Alternative explanations for impaired DNMT1 maintenance function could also be at work. For example, a well-characterized consequence of intergenic H3K36me2 depletion is aberrant spreading of PRC2-catalyzed H3K27me3 (Lu et al. 2016, Jani et al. 2019). Antagonism between DNA methylation and H3K27me3 has been observed in a variety of contexts (Bartke et al. 2010, Brinkman et al. 2012, Reddington et al. 2013), although the molecular underpinnings remain elusive. It remains possible that increased deposition of H3K27me3 at intergenic regions by PRC2 in H3K36me2-depleted cells leads to reduced DNMT1 maintenance activity, thereby lowering intergenic CpG methylation levels. Addressing such a scenario will likely involve further characterization of the DNA methylation landscape in cells which are depleted of both H3K27me3 and H3K36me2.

From a pathophysiologic perspective, our studies demonstrate that intergenic DNA hypomethylation resulting from a dysfunctional NSD1-DNMT3A *trans*-regulatory pathway represents a mechanistic link between two phenotypically overlapping human overgrowth syndromes (**Fig. 2.18B**). This pathway is also likely to facilitate tissue neoplastic overgrowth as loss of *Nsd1* or *Dnmt3a* promotes tumor development in mouse models of squamous cell carcinomas (Quintana et al. 2013, Rinaldi et al. 2017). How depletion of DNA methylation at non-coding genomic regions alters cellular function remains unclear and will be speculated upon in the general discussion of this work. Notably, germline mutations in additional epigenetic regulators, including EZH2 and histone H1, cause Sotos- and TRBS-like developmental disorders (Tatton-Brown et al. 2017). Our current belief is that a complex cross-talk between H3K36me2,

H3K27me3, DNA methylation, and possibly other chromatin factors such as histone H1 works to compartmentalize the genome into discrete compartments to allow for precise gene regulatory control during development and throughout life.

It should also be noted that alterations to the NSD1-DNMT3A axis occur in other developmental disorders and cancers, pointing to additional areas for further investigation. Somatic mutations in *DNMT3A* are common in acute myeloid leukemia, a malignancy where recurrent translocations that fuse *NSD1* to *NUP98* leading to aberrant localization of NSD1 are also observed (Ley et al. 2010, Jaju et al. 2001). As NSD1 appears to act upstream of DNMT3A, more work is warranted to better understand how NSD family enzymes are recruited to chromatin and how these mechanisms are disrupted in disease. Likewise, the work in this chapter does not exclude the potential for additional recruitment mechanisms for DNMT3A, possibly involving chromatin interactions that are independent of PWWP domain function. Indeed, gain-of-function mutations in DNMT3A that result in ectopic DNA hypermethylation at polycomb repressive domains were recently found to cause delayed growth and microcephaly in mice and human (Heyn et al. 2019, Sendžikaitė et al. 2019) and will be the topic of the next chapter.

CHAPTER 3: DNMT3A binds H2AK119Ub to hypermethylate CpG islands upon loss of PWWP reader function

INTRODUCTION

As demonstrated in Chapter 2, the localization of DNMT3A and DNMT3B is facilitated by PWWP domain recognition of H3K36 methylation to target intergenic regions and gene bodies for CpG methylation, respectively. The *de novo* DNA methyltransferases are also normally excluded from binding CpG islands (CGIs) (Wu et al. 2010), which likely contributes to the unmethylated status of many of these regions. However, CpG methylation at a subset of CGIs that are regulated by Polycomb repressive complexes has been observed during embryogenesis (Chen et al. 2019), cellular differentiation (Mohn et al. 2008), and cancer progression (Ohm et al. 2007, Schlesinger et al. 2007, Widschwendter et al. 2007). Furthermore, disruption of binding between the PWWP domain DNMT3A and H3K36 methylation due to hotspot missense mutations observed in paraganglioma tumors and patients with microcephalic dwarfism also promotes CpG hypermethylation of Polycomb-regulated regions (Remacha et al. 2018, Heyn et al. 2019, Dukatz et al. 2019).

This chapter explores the hypothesis that these observations may be explained by the existence of an uncharacterized recruitment mechanism that competes with the PWWP domain for DNMT3A targeting. To explore a role for PWWP-independent recruitment in pathophysiologic contexts, we sought to compare the effects of PWWP domain deletion with dwarfism- and paraganglioma-associated missense mutations on

DNMT3A localization. We rationalized such an approach could also indicate whether the patient-derived DNMT3A PWWP mutations are neomorphic or if their effects on the DNA methylation landscape could be explained solely through loss of PWWP domain functionality. Based on the previous association of these mutations with hypermethylation of Polycomb-regulated regions we focused our initial attention on the relationship of DNMT3A PWWP mutants to PRC2-catalyzed H3K27me3 and PRC1-catalyzed H2AK119Ub. Analogous to the last chapter, here we use both cell-free and cell-based techniques to profile PWWP domain-independent recruitment mechanisms of DNMT3A.

RESULTS

Disease-associated hotspot missense mutations in DNMT3A impair binding of its PWWP domain to di- and tri-methylated histone H3 at lysine 36 (H3K36me2, H3K36me3) (Dukatz et al. 2019, Heyn et al. 2019) and also promote DNA hypermethylation of Polycomb-regulated regions through an uncharacterized mechanism (Remacha et al. 2018, Heyn et al. 2019). To better understand regulation of DNMT3A mutant targeting, we stably expressed wild-type, PWWP domain deletion mutant (ΔPWWP), and patient-derived PWWP domain mutant (K299I, R318W, W330R, and D333N) forms of haemagglutinin(HA)-tagged DNMT3A1—the predominant isoform in somatic cells (Chen et al. 2002, Manzo et al. 2017)—in C3H10T1/2 mouse mesenchymal stem cells. We next obtained genome-wide binding profiles for each mutant and teamed up with members of Dr. Majewski's group (graduate student Phil

Figure 3.1: Disease-associated DNMT3A PWWP domain mutations abrogate recruitment to H3K36me2-enriched regions. **A)** Heat map showing genome-wide, pairwise Pearson correlations across 10-kb bins (n = 245,842) between H3K36me2, H3K36me3, H3K27me3, H2AK119Ub, DNMT3A wild-type, DNMT3A PWWP mutants, and Ring1b. **B)** Scatterplots showing genome-wide, pairwise Pearson correlations across 10-kb bins (n = 245,842) for ChIP-seq normalized reads between H3K36me2, H3K36me3, H3K27me3, H2AK119Ub, DNMT3A wild-type, DNMT3A PWWP mutants, and RING1B. Pearson's correlation coefficients are indicated. **C)** Genome browser representation of ChIP-seq normalized reads for H3K36me2, H3K36me3, H3K27me3, H2AK119Ub, DNMT3A wild-type and DNMT3A PWWP mutants in mouse MSCs at chromosome 6: 50.9-51.2 Mb. The shaded areas indicate H3K36me2-enriched (orange) and H2AK119Ub-enriched (red) genomic regions. **D)** Ratio of observed-to-expected ChIP-seq reads for DNMT3A1 wild-type, ΔPWWP, and H2AK119Ub in annotated genomic regions. Numbers of expected reads were generated assuming equivalent genomic distribution to input. **E)** ChIP-seq normalized reads for DNMT3A1 wild-type (grey) in mouse MSCs relative to H3K36me2 at intergenic regions. Intergenic regions greater than 10-kb (n = 13,990) were ranked and sorted by mean H3K36me2 enrichment in MSCs. The black line indicates mean H3K36me2 enrichment per bin. **F)** ChIP-seq normalized reads for DNMT3A1 ΔPWWP (orange) in mouse MSCs relative to H3K36me2 at intergenic regions. Intergenic regions greater than 10-kb (n = 13,990) were ranked and sorted by mean H3K36me2 enrichment in MSCs. The black line indicates mean H3K36me2 enrichment per bin.

A

B
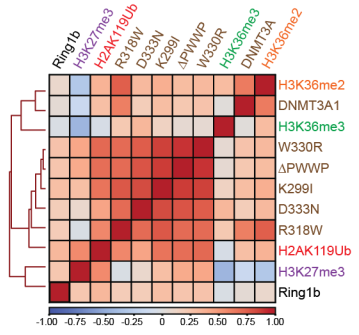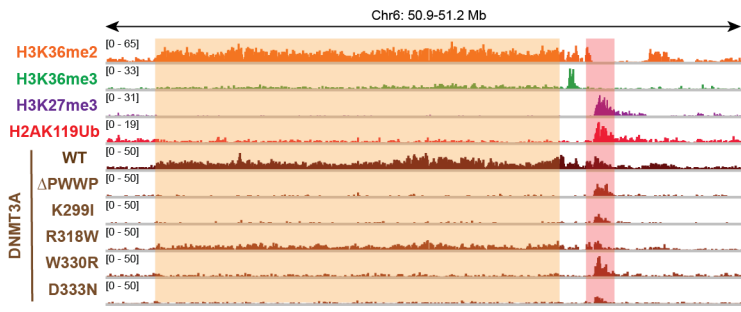
C Chr6: 50.9-51.2 Mb
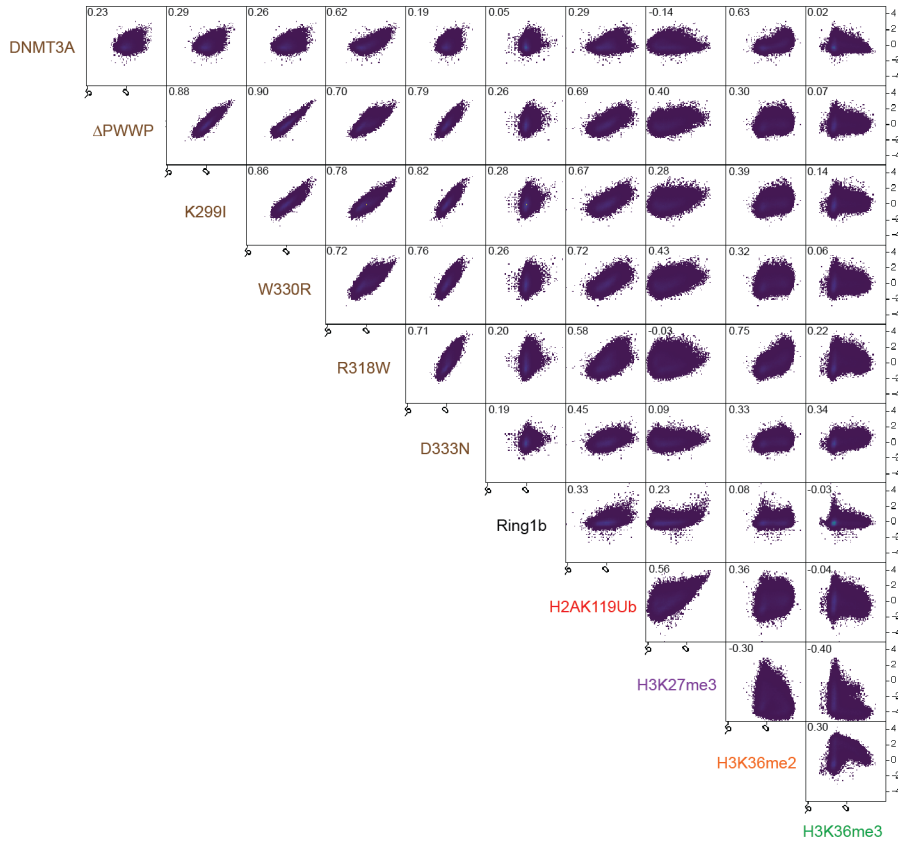
D
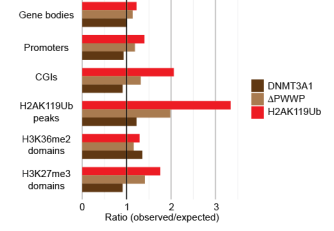
E
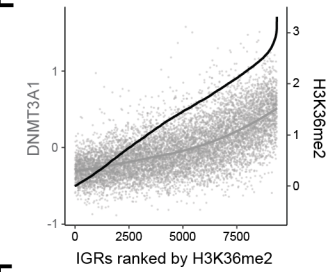
F
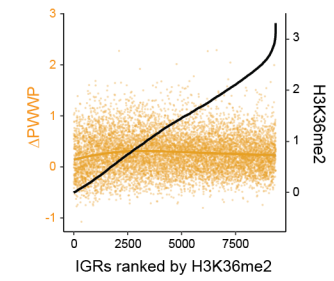
Rosenbaum) to compare them to the distribution patterns of histone post-translational modifications (H3K36me2, H3K36me3, H3K27me3, H2AK119Ub) and the PRC1 subunit RING1B (Lu et al. 2016) using chromatin immunoprecipitation followed by sequencing (ChIP-seq). Unlike wild-type DNMT3A, ΔPWWP and patient-derived DNMT3A mutants failed to co-localize with H3K36me2 genome-wide (**Fig. 3.1A,B**). Consistent with this, DNMT3A mutant binding was reduced at H3K36me2-enriched intergenic regions relative to wild-type DNMT3A (**Fig. 3.1C-E**). Rather, binding of both DNMT3A ΔPWWP and patient-derived mutants were highly correlated and observed primarily at Polycomb-regulated regions as defined by the presence of H3K27me3 or H2AK119Ub (**Fig. 3.1A-D**). Thus, disease-associated mutations in DNMT3A abrogate PWWP-mediated targeting to intergenic regions and promote mis-targeting to Polycomb-regulated regions through loss of PWWP reader domain functionality.

Upon closer examination, we noted that DNMT3A mutant binding more positively correlated genome-wide with H2AK119Ub compared to H3K27me3 (**Fig. 3.1A,B**). Moreover, DNMT3A mutants were preferentially recruited to H2AK119Ub-enriched genomic regions which lacked H3K27me3, but not to regions enriched solely for H3K27me3 suggesting that their localization may be regulated in a PRC1-dependent and PRC2-independent manner (**Fig. 3.2A**). In support of this, DNMT3A mutant binding genome-wide was associated more strongly with the presence of H2AK119Ub than H3K27me3 even after controlling for a potential confounding effect from the other histone PTM (**Fig. 3.2B**). A lack of a role for PRC2-catalyzed H3K27me3 in targeting was confirmed by an inability of the mutants to localize to H3K27me3 peak regions (**Fig.**

68

Figure 3.2: DNMT3A mutants colocalize with H2AK119Ub independently of H3K27me3. A) Genome browser representation of ChIP-seq normalized reads for H2AK119Ub, H3K27me3, and DNMT3A PWWP mutants in mouse MSCs at chromosome 14: 122.6-123.1 Mb and chromosome 11: 117.7-117.8 Mb. Genes from the RefSeq database are annotated at the bottom. The shaded areas indicate H3K27me3-enriched (purple) and H2AK119Ub-enriched (red) genomic regions. **B)** Genome-wide partial correlations of ChIP-seq normalized reads across 10-kb bins (n = 245,842) in parental MSCs. Left: relationships between DNMT3A PWWP mutants and H2AK119Ub after controlling for H3K27me3. Right: relationships between DNMT3A PWWP mutants and H3K27me3 after controlling for H2AK119Ub. **C)** Averaged ChIP-seq normalized read signal at H3K27me3 peaks ± 5-kb (n = 137,523), represented as log2(ChIP signal/input signal) for DNMT3A1 wild-type, DNMT3A PWWP mutants, H2AK119Ub, and H3K27me3 in parental mouse MSCs. **D)** Enrichment heat map depicting ChIP-seq normalized reads centered at H2AK119Ub peaks ± 5-kb within broad domains of H3K27me3 (n = 13,620) for DNMT3A wild-type, DNMT3A PWWP mutants, RING1B, H2AK119Ub, H3K27me3, H3K36me2, and H3K36me3. Regions are sorted by H2AK119Ub enrichment. **E)** Enrichment heat map depicting ChIP-seq normalized reads centered at H2AK119Ub peaks ± 5 Kb outside of broad domains of H3K27me3 (n = 17,438) for DNMT3A wild-type, DNMT3A PWWP mutants, RING1B, H2AK119Ub, H3K27me3, H3K36me2, and H3K36me3. Regions are sorted by H2AK119Ub enrichment.

A

Chr14: 122.6-123.1 Mb

H2AK119Ub [0 - 12]
H3K27me3 [0 - 4.16]

DNMT3A
ΔPWWP [0 - 32]
K299I [0 - 27]
R318W [0 - 75]
W330R [0 - 56]
D333N [0 - 29]

Pcca   Ggact   Tmtc4

Chr11: 117.7-117.8 Mb

[0 - 22]
[0 - 4.32]
[0 - 61]
[0 - 41]
[0 - 142]
[0 - 84]
[0 - 72]

Tnrc6c   Tmc6   Tmc8   Tk1   Afmid
6030468B19Rik
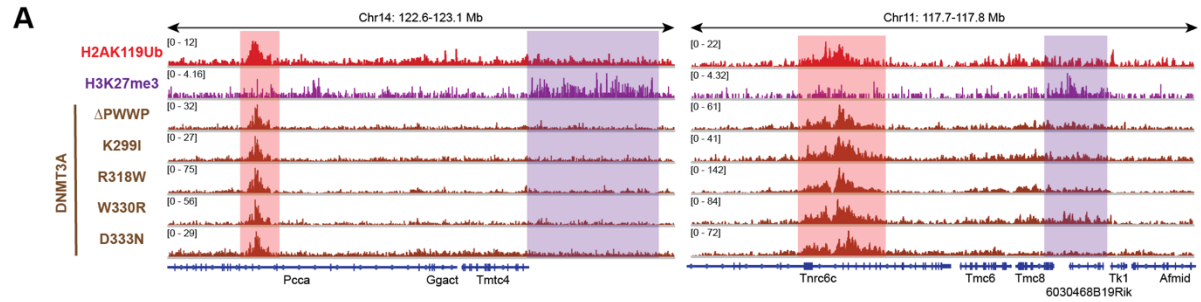
B

Partial Correlation Analysis

H2AK119Ub vs. DNMT3A1

DNMT3A1	r	p
ΔPWWP	0.593	2×10^{-16}
K299I	0.625	2×10^{-16}
R318W	0.708	2×10^{-16}
W330R	0.623	2×10^{-16}
D333N	0.471	2×10^{-16}

H3K27me3 vs. DNMT3A1

DNMT3A1	r	p
ΔPWWP	0.035	2×10^{-16}
K299I	-0.132	2×10^{-16}
R318W	-0.497	2×10^{-16}
W330R	0.050	2×10^{-16}
D333N	-0.201	2×10^{-16}

C

H3K27me3 peaks

DNMT3A
ΔPWWP
K299I
W330R
R318W
D333N
H2AK119Ub
H3K27me3

-5.0   peak center   5.0Kb

D

DNMT3A1 ΔPWWP K299I W330R R318W D333N Ring1b H2AK119Ub H3K27me3 H3K36me2 H3K36me3

H2AK119Ub peaks within H3K37me3 peaks, sorted by H2AK119Ub

Kb: -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0

0.0 0.4 0.8

E

DNMT3A1 ΔPWWP K299I W330R R318W D333N Ring1b H2AK119Ub H3K27me3 H3K36me2 H3K36me3

H2AK119Ub peaks outside H3K37me3 peaks, sorted by H2AK119Ub

Kb: -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0 -5.0 5.0

0.0 0.4 0.8

70

3.2C). Indeed, DNMT3A mutant localization to H2AK119Ub peaks was observed regardless of the accompanying presence (**Fig. 3.2D**) or absence (**Fig. 3.2E**) of H3K27me3, further suggesting that recruitment to Polycomb-regulated regions occurs separately from PRC2 activity.

While canonical PRC1 recognizes H3K27me3 through chromodomain-containing CBX subunits, a variant form of PRC1 is targeted to CGIs in a PRC2-independent fashion (Tavares et al. 2012, Wu et al. 2013, Blackledge et al. 2014). We thus examined whether DNMT3A mutant may be targeted to CGIs due to the action of PRC1. In aggregate, DNMT3A ΔPWWP localization was significantly enriched at CGIs, whereas wild-type DNMT3A binding was significantly depleted (**Fig. 3.3A**). We wondered whether DNMT3A mutants were recruited to all CGIs or only a specific subset. Genome-wide analysis revealed that CGIs were more strongly associated with the presence of H2AK119Ub (6,386 of 16,023 CGIs; 39.8%) than H3K27me3 (3,650 of 16,023 CGIs; 22.8%), despite the broader genomic distribution of H3K27me3 (**Fig. 3.3B**). We observed enhanced localization of DNMT3A mutants specifically at H2AK119Ub-enriched CGIs, in contrast to wild-type DNMT3A which failed to be targeted to CGIs consistent with the relative depletion of H3K36me2 and H3K36me3 at these sites (**Fig. 3.3C**). Furthermore, targeting of DNMT3A mutants to H2AK119Ub-enriched CGIs could not be explained by the presence of PRC1 itself, as the RING1B subunit was relatively uniformly enriched across all CGIs (**Fig. 3.3C**). Together, these results suggest that DNMT3A mutant localization to CGIs is determined by the presence of H2AK119Ub.
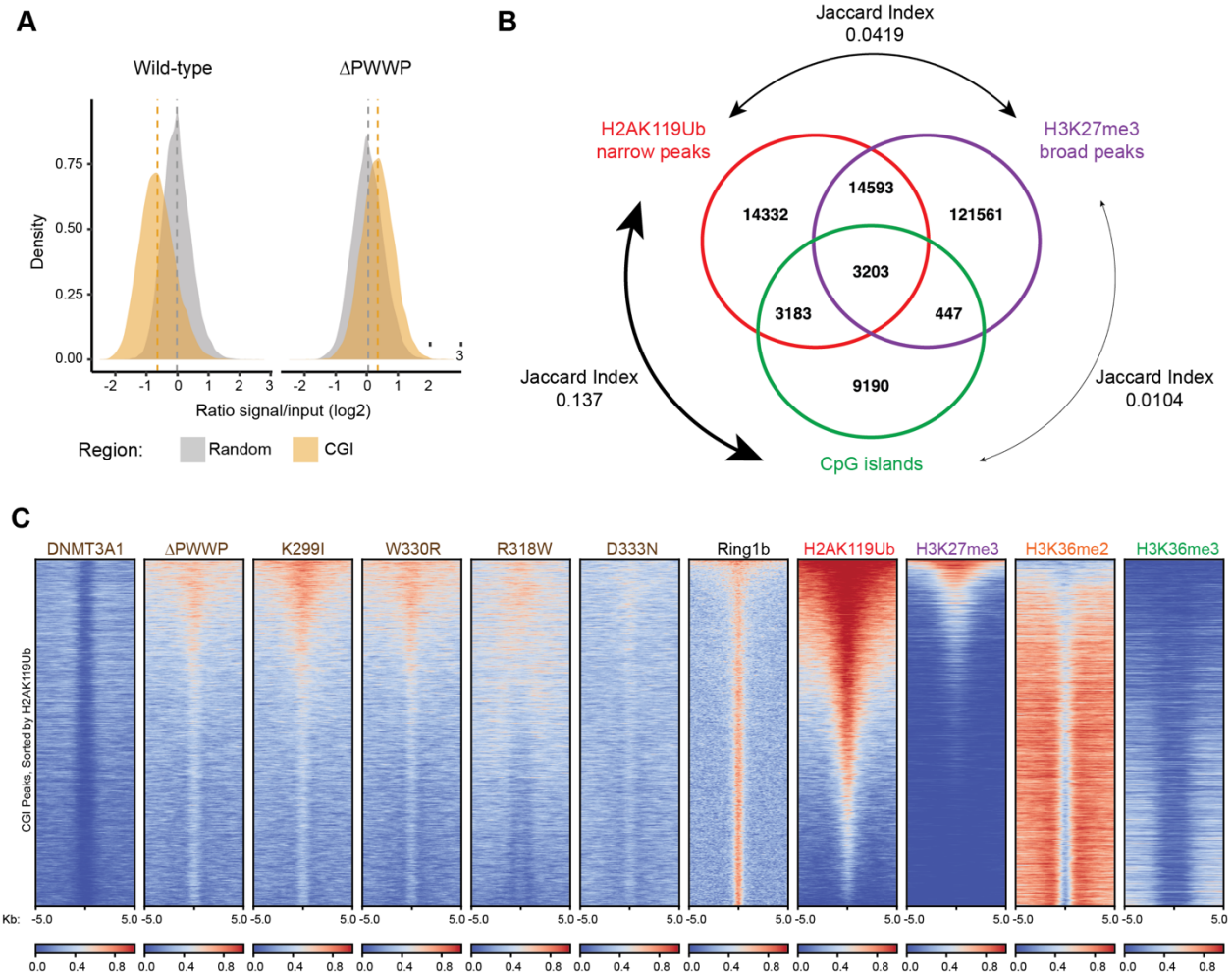
Figure 3.3: Loss of PWWP reader domain function redistributes DNMT3A to CpG islands. A) Observed DNMT3A1 wild-type and ΔPWWP enrichment at CpG islands (*n* = 15,492) in MSCs compared to expected signal represented by enrichment at randomly-shuffled regions (*n* = 15,492). **B)** Overlap analysis of H2AK119Ub narrow peaks, H3K27me3 broad peaks, and CpG islands in mouse MSCs. Jaccard index for pairwise comparisons are indicated. **C)** Enrichment heat map depicting ChIP-seq normalized reads centered at CpG islands ± 5-kb for DNMT3A wild-type, DNMT3A PWWP mutants, Ring1b, H2AK119Ub, H3K27me3, H3K36me2, and H3K36me3. Regions are sorted by H2AK119Ub enrichment.

PRC1 contains two interchangeable ubiquitin ligase catalytic subunits, RING1A and RING1B, which are redundant for depositing H2AK119Ub genome-wide (de Napoles et al. 2004, Wang et al. 2004). To directly test if PRC1 is required for DNMT3A mutant targeting, we used sgRNA to genetically ablate both *Ring1a* and *Ring1b* in MSCs (sgRing1a/b) using CRISPR-Cas9 and stably expressed DNMT3A ∆PWWP and patient-derived mutants at equivalent levels to parental MSCs (**Fig. 3.4A**). ChIP-seq in sgRing1a/b MSCs revealed profound decreases in H2AK119Ub levels at both CGI- and non-CGI-containing regions, which was accompanied by impaired localization of DNMT3A mutants (**Fig. 3.4B**). Genome-wide, depletion of H2AK119Ub in sgRing1a/b cells was accompanied by reduced recruitment of both DNMT3A mutants (**Fig. 3.4C**). Similar trends were also observed at CGIs for DNMT3A ∆PWWP and gain-of-function mutants (**Fig. 3.4D**). As PRC1 in certain contexts can support downstream recruitment of PRC2 (Blackledge et al. 2014, Cooper et al. 2014, Cooper et al. 2016), we sought to confirm that loss of DNMT3A mutant recruitment could not be explained by changes to H3K27me3. Partial correlation analysis indicated that reductions in DNMT3A mutant recruitment tracked more closely with changes to H2AK119Ub than H3K27me3 in sgRing1a/b cells (**Fig. 3.5A**). Moreover, loss of DNMT3A mutant localization at H2AK119Ub-depleted regions was observed despite the persistence of H3K27me3 at some sites (**Fig. 3.5B**).

We asked whether DNMT3A mutant recruitment is mediated through direct interaction with H2AK119Ub-modified nucleosomes. We reasoned that wild-type DNMT3A may possess a latent ability to bind H2AK119Ub, which becomes activated in

Figure 3.4: PRC1-catalyzed H2AK119Ub is required for localization of DNMT3A to CpG islands. A) Immunoblots of lysates generated from parental and sgRing1a/b mouse MSCs that ectopically express HA-tagged DNMT3A1 PWWP mutants. Vinculin was used as a loading control. **B)** Genome browser representation of ChIP-seq normalized reads for H2AK119Ub and DNMT3A1 PWWP mutants in parental and sgRing1a/b mouse MSCs at chromosome 12: 110.3-110.7 Mb. CGIs (green) and genes from the RefSeq database are annotated at the bottom. The shaded areas indicate H2AK119Ub-enriched (red) genomic regions. **C)** Difference in ChIP-seq normalized reads of DNMT3A1 ΔPWWP, K299I, R318W, W330R, and D333N between parental and sgRing1a/b mouse MSCs relative to that of H2AK119Ub for 10-kb non-overlapping bins genome-wide ($n = 245,842$). Pearson's correlation coefficient is indicated. **D)** Difference in ChIP-seq normalized reads of DNMT3A1 ΔPWWP, K299I, R318W, W330R, and D333N between parental and sgRing1a/b mouse MSCs relative to that of H3K27me3 for CGIs ($n = 15,492$). Pearson's correlation coefficient is indicated.

A

HA-DNMT3A1:

Parental: ΔPWWP, K299I, R318W, W330R, D333N
sgRing1a/b: ΔPWWP, K299I, R318W, W330R, D333N

α-HA

α-Vinculin

B

Chr12: 110.3-110.7 Mb

Parental
H2AK119Ub [0 - 22]
DNMT3A:
ΔPWWP [0 - 29]
K299I [0 - 36]
R318W [0 - 105]
W330R [0 - 55]
D333N [0 - 33]

sgRing1a/b
H2AK119Ub [0 - 22]
DNMT3A:
ΔPWWP [0 - 29]
K299I [0 - 36]
R318W [0 - 105]
W330R [0 - 55]
D333N [0 - 33]

CGIs

Ppp2r5c   Dync1h1   1700001K19Rik Wdr20

C

Genome-wide

Δ[DNMT3A1 ΔPWWP] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.47

Δ[DNMT3A1 K299I] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.42

Δ[DNMT3A1 R318W] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.38

Δ[DNMT3A1 W330R] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.51

Δ[DNMT3A1 D333N] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.20

D

CGIs

Δ[DNMT3A1 ΔPWWP] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.48

Δ[DNMT3A1 K299I] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.52

Δ[DNMT3A1 R318W] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.47

Δ[DNMT3A1 W330R] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.52

Δ[DNMT3A1 D333N] (sgRing1a/b vs. parental) — ΔH2AK119Ub (sgRing1a/b vs. parental) — PCC = 0.30

A

Partial Correlation Analysis
(Parental vs. sgRing1a/b)

ΔH2AK119Ub vs. ΔDNMT3A1

DNMT3A1	r	p
ΔPWWP	0.398	2×10^{-16}
K299I	0.366	2×10^{-16}
R318W	0.408	2×10^{-16}
W330R	0.437	2×10^{-16}
D333N	0.167	2×10^{-16}

ΔH3K27me3 vs. ΔDNMT3A1

DNMT3A1	r	p
ΔPWWP	0.052	2×10^{-16}
K299I	0.025	2×10^{-16}
R318W	-0.177	2×10^{-16}
W330R	0.065	2×10^{-16}
D333N	0.014	2.8×10^{-8}

Figure 3.5: Genetic ablation of PRC1 abrogates recruitment of DNMT3A mutants despite persistence of H3K27me3. A) Genome-wide partial correlations of ChIP-seq normalized reads across 10-kb bins (n = 245,842) between parental and sgRing1a/b MSCs. Left: relationships between changes in DNMT3A PWWP mutants and H2AK119Ub after controlling for H3K27me3. Right: relationships between changes in DNMT3A PWWP mutants and H3K27me3 after controlling for H2AK119Ub. **B)** Genome browser representation of ChIP-seq normalized reads for H2AK119Ub, H3K27me3, and DNMT3A PWWP mutants in parental and sgRing1a/b mouse MSCs at chromosome 11: 21.96-22.03 Mb and chromosome 16: 57.0-57.2 Mb. Genes from the RefSeq database are annotated at the bottom. The shaded areas indicate H2AK119Ub-enriched (red) genomic regions.

76

cells upon loss of PWWP domain reader function. We therefore examined the interactions between purified full-length DNMT3A1 and semi-synthetic nucleosomes in vitro in collaboration with EpiCypher Inc. (technician Matthew Marunde and Chief Scientific Officer Dr. Michael-Christopher Keogh). Indeed, we observed DNMT3A1 bound to nucleosomes modified with H2AK119Ub, in addition to those with H3K36me2 or H3K36me3, but did not recognize unmodified nucleosomes or those containing H2BK120Ub or H3K36me1 (**Fig. 3.6A**). We speculated that full-length DNMT3A1 may contain a ubiquitylation-dependent recruitment region (UDR), similar to a structured region identified within 53BP1 that mediates interaction with H2AK15Ub-modified nucleosomes (Fradet-Turcotte et al. 2013, Wilson et al. 2016). Consistent with this, Predictor of Natural Disordered Regions (PONDR) analysis revealed the presence of an unannotated ordered domain in the amino-terminal region of DNMT3A1 that could serve as a putative UDR (**Fig. 3.6B**).

The amino-terminal region of DNMT3A1 is absent from DNMT3A2—a shorter isoform expressed primarily in early development—and has minimal amino acid sequence conservation with that of DNMT3B, the other mammalian *de novo* DNA methyltransferase (Manzo et al. 2017). Purified full-length DNMT3B engaged nucleosomes modified with H3K36me3 and H3K36me2, but unlike DNMT3A1 failed to interact with H2AK119Ub-modified nucleosomes in vitro (**Fig. 3.6C**). The amino-terminal portion of DNMT3B also lacked a well-demarcated ordered domain (**Fig. 3.6D**), suggesting UDR-mediated recognition of H2AK119Ub may be specific to DNMT3A1. Furthermore, interaction with H2AK119Ub-modified nucleosomes was not observed for
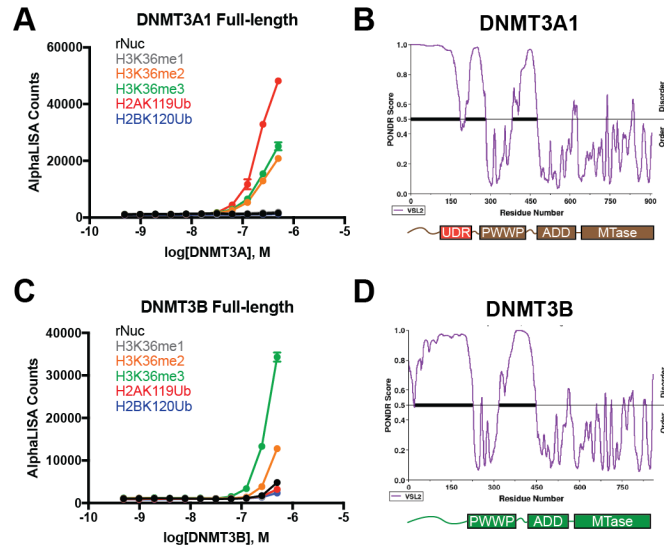
Figure 3.6: DNMT3A interacts directly with H2AK119Ub-modified nucleosomes. A)
Quantification of AlphaLISA counts for isolated GST-tagged DNMT3A full-length
interaction, titrated against H2AK119Ub, H2BK120Ub, and H3K36-modified
nucleosomes. **B)** Graph of intrinsic disorder for DNMT3A1. PONDR (Predictor of
Natural Disordered Regions) VSL2 scores are indicated on the y axis and the amino
acid positions are indicated on the x axis, with the domain structure of DNMT3A1 shown
below. **C)** Quantification of AlphaLISA counts for isolated GST-tagged DNMT3B full-
length interaction, titrated against H2AK119Ub, H2BK120Ub, and H3K36-modified
nucleosomes. **D)** Graph of intrinsic disorder for DNMT3B. PONDR (Predictor of Natural
Disordered Regions) VSL2 scores are indicated on the y axis and the amino acid
positions are indicated on the x axis, with the domain structure of DNMT3B shown
below.

the purified DNMT3A PWWP domain (DNMT3A$_{PWWP}$) (**Fig. 3.7A**). To assess whether the amino-terminal UDR region of DNMT3A1 is necessary and sufficient for targeting to H2AK119Ub-enriched regions we stably expressed wild-type, PWWP deletion, and domain swap mutants of DNMT3A1, DNMT3A2, and DNMT3B in parental MSCs (**Fig. 3.7B,C**). Deletion of the PWWP domain redistributed DNMT3A1 to H2AK119Ub-enriched regions but not DNMT3A2 (3A2 ∆PWWP) or DNMT3B (3B ∆PWWP), however swapping of the amino-terminal region of DNMT3A1 into DNMT3B (N-3A1/3B ∆PWWP) was sufficient to confer co-localization with H2AK119Ub (**Fig. 3.7D**). Furthermore, we found that DNMT3A1 ∆PWWP required the UDR (residues 160-219) for targeting to H2AK119Ub-enriched regions whereas the amino-terminal disordered region (residues 1-159) was dispensable (**Fig. 3.7E**).

Recent work has implicated PRC1 in X chromosome inactivation through deposition of H2AK119Ub independent of PRC2 activity (de Napoles et al. 2004, Almeida et al. 2017, Zylicz et al. 2019), leading us to speculate that DNMT3A mutants may be targeted to the inactive X chromosome (Xi). We confirmed in parental MSCs that DNMT3A1 ∆PWWP, but not DNMT3A1 wild-type, co-localized with the long non-coding RNA *Xist* responsible for establishing and maintaining X chromosome inactivation (**Fig. 3.8A**). All deletion and domain swap mutants which localized to H2AK119Ub-enriched regions on autosomes also localized to Xi in parental MSCs, suggesting that targeting to Xi may be PRC1-dependent (**Fig. 3.8B**). Accordingly, ablation of PRC1 abrogated recruitment of DNMT3A1 ∆PWWP to Xi (**Fig. 3.8C**). We noted that all DNMT constructs containing the amino-terminal UDR domain of DNMT3A

Figure 3.7: H2AK119Ub interaction potential is specific for DNMT3A and resides within a putative N-terminal ubiquitin-dependent recruitment region. A) Quantification of AlphaLISA counts for isolated GST-tagged DNMT3A$_{PWWP}$ interaction, titrated against H2AK119Ub, H2BK120Ub, and H3K36-modified nucleosomes. **B)** Immunoblots of lysates generated from parental mouse MSCs that ectopically express HA-tagged wild-type, deletion, and domain swap mutants of DNMT3A1, DNMT3A2, and DNMT3B. Vinculin was used as a loading control. **C)** Schematic of wild-type, deletion, and domain swap mutants of DNMT3A1, DNMT3A2, and DNMT3B. **D)** Fold enrichment of DNMT3A1, DNMT3A2, DNMT3B, and their corresponding deletion mutants at various H2AK119Ub-enriched versus H2AK119Ub-depleted regions in mouse MSCs, measured by ChIP-qPCR. **E)** Fold enrichment of DNMT3A1 ΔPWWP and co-deletion mutants of N-terminal disordered (Δ1-159) and ordered (Δ160-219) domains at various H2AK119Ub-enriched versus H2AK119Ub-depleted regions in mouse MSCs, measured by ChIP-qPCR. P values were determined by one-way analysis of variance (ANOVA).

A DNMT3A PWWP

rNuc
H3K36me1
H3K36me2
H3K36me3
H2AK119Ub
H2BK120Ub

B

HA-DNMT:

α-HA

α-Vinculin

C

3A1 WT
3A1 ΔPWWP
3A2 WT
3A2 ΔPWWP
3B WT
3B ΔPWWP
N-3A1/3B ΔPWWP

D

E
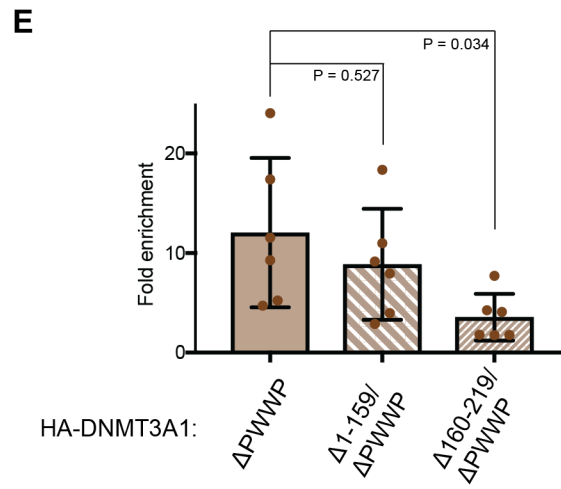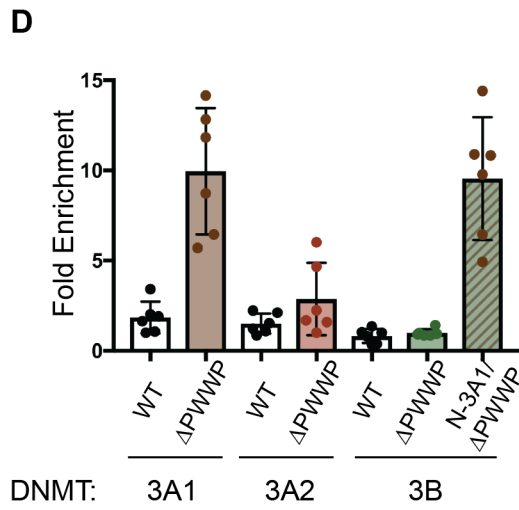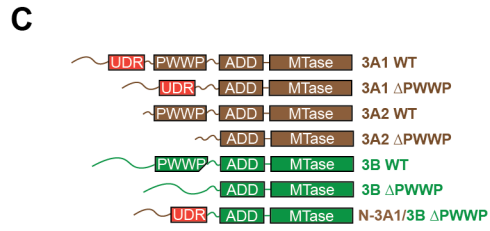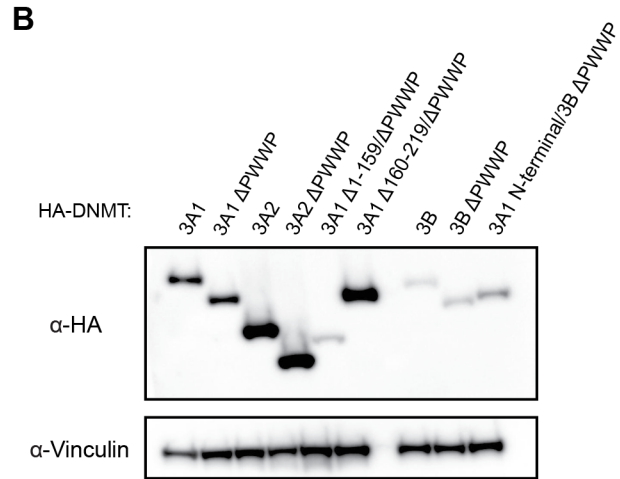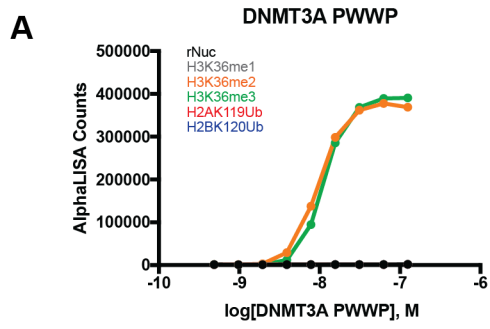
Figure 3.8: Localization of DNMT3A PWWP mutants to inactive X chromosome. A)
ImmunoFISH for DNMT3A wild-type and ΔPWWP (green) and Xist RNA (red) in
parental MSCs. Nuclei are stained with DAPI (blue). **B)** Summary of localization to
H2AK119Ub-enriched regions (by ChIP) and the inactive X chromosome (by
ImmunoFISH) in parental MSCs for wild-type, deletion, and domain swap mutants of
DNMT3A1, DNMT3A2, and DNMT3B. DNMT3A/B constructs containing the DNMT3A1
UDR region are indicated in bold. **C)** Summary of localization to H2AK119Ub-enriched
regions (by ChIP) and the inactive X chromosome (by ImmunoFISH) for DNMT3A1
ΔPWWP in parental and sgRing1a/b MSCs. **D)** Summary of localization to H2AK119Ub-
enriched regions (by ChIP) and the inactive X chromosome (by ImmunoFISH) for
DNMT3A1 wild-type in parental, sgNsd1/2, and H3K36M-expressing MSCs. N.D. = not
determined.

A

	α-HA	Xist	DAPI	Merged
DNMT3A1 wild-type				
DNMT3A1 ΔPWWP				
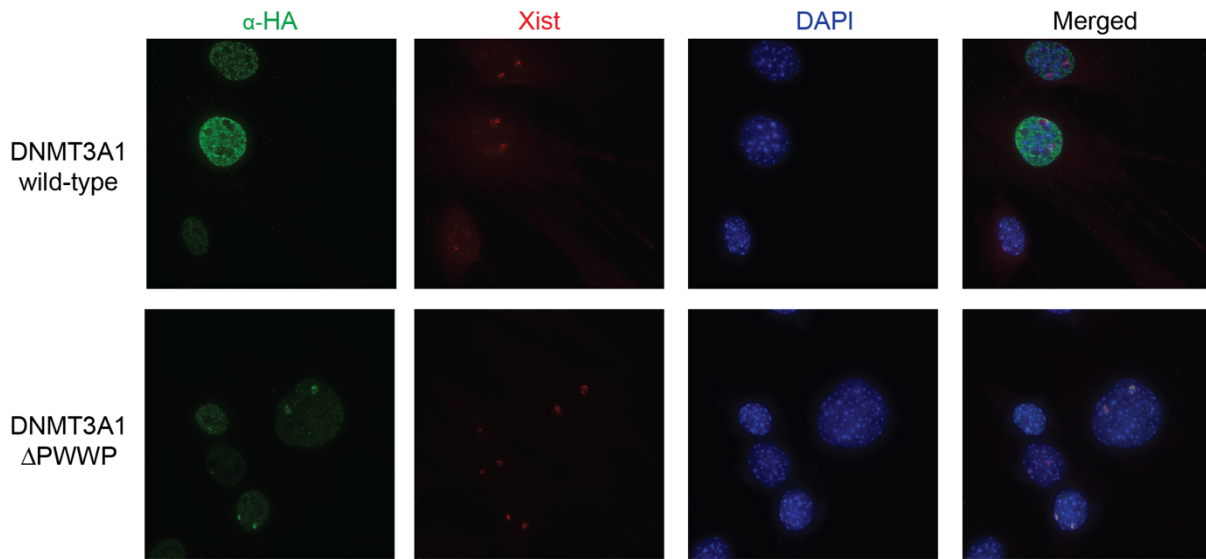
B

In parental MSCs:

DNMT expressed	Co-localizes with H2AK119Ub?	Co-localizes with Xist by IF?
3A1 WT	No	No
3A1 ΔPWWP	Yes	Yes
3A2 WT	No	No
3A2 ΔPWWP	No	No
3B WT	No	No
3B ΔPWWP	No	No
N-3A1/ 3B ΔPWWP	Yes	Yes
Δ1-159/ 3A1 ΔPWWP	Yes	Yes
Δ160-219/ 3A1 ΔPWWP	No	No

(DNMT constructs containing putative UDR domain)

C

~UDR~~ADD~MTase **3A1 ΔPWWP**

MSCs	Co-localizes with H2AK119Ub?	Co-localizes with Xist by IF?
Parental	Yes	Yes
sgRing1a/b	No	No

D

~UDR~PWWP~ADD~MTase **3A1 WT**

MSCs	Co-localizes with H2AK119Ub?	Co-localizes with Xist by IF?
Parental	No	No
sgNsd1/2	No	No
H3K36M	N.D.	Yes

(**boldfaced in Fig. 3.8B**) localized to H2AK119Ub-enriched regions and Xi except for wild-type DNMT3A1, which had intact PWWP domain functionality. We wondered whether PWWP-mediated recruitment may outcompete UDR-mediated recruitment in parental MSCs to determine DNMT3A1 localization. To test this, we expressed wild-type DNMT3A1 in MSCs that were globally depleted of both H3K36me2 and H3K36me3 due to expression of the dominant-acting oncohistone H3K36M (Fang et al. 2016, Lu et al. 2016). Cellular depletion of the PWWP binding substrates H3K36me2 and H3K36me3 resulted in a marked shift of wild-type DNMT3A1 to Xi, but this did not occur in H3K36me2-depleted sgNsd1/2 MSCs that had preserved levels of H3K36me3 (**Fig. 3.8D**). We conclude that DNMT3A1 targeting in parental MSCs is organized in a hierarchical manner, with targeting to H3K36me2-enriched regions being the most favored, followed by regions enriched for H3K36me3 and then H2AK119Ub.

Patient samples harboring hotspot missense mutations in DNMT3A are characterized by DNA hypermethylation of Polycomb-regulated regions (Remacha et al. 2018, Heyn et al. 2019). To determine if hypermethylation may be a direct consequence of DNMT3A mutant recruitment by H2AK119Ub, we performed Reduced Representation Bisulfite Sequencing (RRBS) in parental and sgRing1a/b MSCs expressing either wild-type or W330R-mutant DNMT3A. DNA hypermethylation was observed at H2AK119Ub-enriched CGIs in W330R-expressing parental cells that was abrogated by deletion of PRC1 and concomitant loss of DNMT3A1 W330R recruitment (**Fig. 3.9A**). Genome-wide, expression of DNMT3A1 W330R promoted hypermethylation of H2AK119Ub-enriched regions in a PRC1-dependent manner (**Fig.**

Figure 3.9: DNMT3A-mediated CpG island hypermethylation is dependent on PRC1. A) Genome browser representation of ChIP-seq normalized reads for H2AK119Ub and DNMT3A1 W330R and RRBS data for CpG methylation (black) in parental and sgRing1a/b mouse MSCs expressing either wild-type or W330R DNMT3A1 at chromosome 7: 130.2-130.3 Mb and chromosome 10: 13.9-14.0Mb. CGIs (green) and genes from the RefSeq database are annotated at the bottom. The shaded areas indicate H2AK119Ub-enriched (red) genomic regions. **B)** Histograms for CpG methylation within parental H2AK119Ub peak regions (n = 320,708 CpGs) in parental and sgRing1a/b mouse MSCs expressing DNMT3A1 wild-type (grey) or W330R (orange). **C)** Histograms for CpG methylation outside parental H2AK119Ub peak regions (n = 1,801,839 CpGs) in parental and sgRing1a/b mouse MSCs expressing DNMT3A wild-type (grey) or W330R (orange). **D)** Histograms for CpG methylation at CpG islands within parental H2AK119Ub peak regions (n = 250,121 CpGs) in parental and sgRing1a/b mouse MSCs expressing DNMT3A1 wild-type (grey) or W330R (orange). **E)** Histograms for CpG methylation at CpG islands outside parental H2AK119Ub peak regions (n = 356,051 CpGs) in parental and sgRing1a/b mouse MSCs expressing DNMT3A1 wild-type (grey) or W330R (orange).
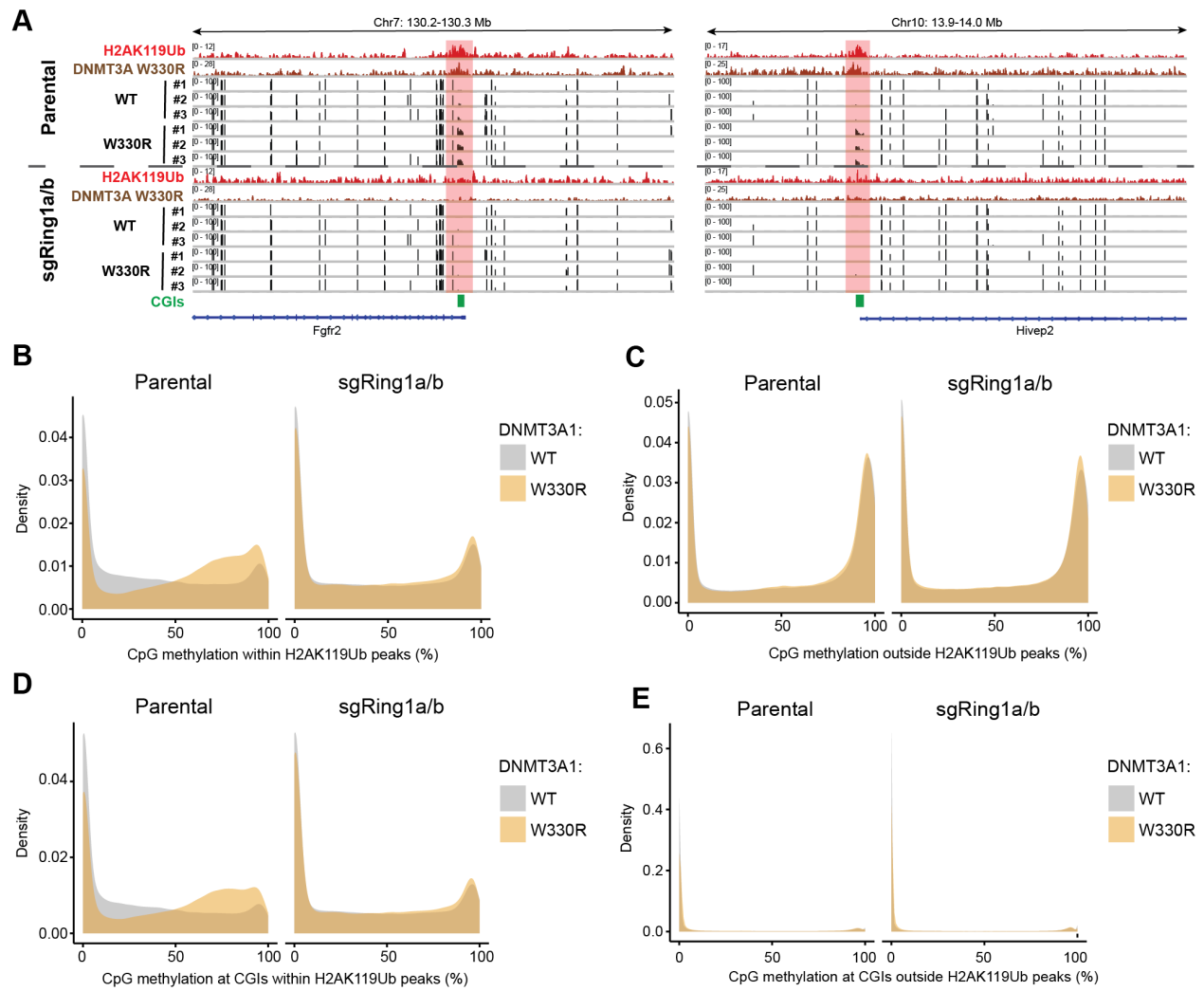
3.9B) and elicited minimal changes at H2AK119Ub-depleted regions (**Fig. 3.9C**).

Similarly, CGI hypermethylation in W330R-mutant cells also required PRC1 and

coincided with the presence (**Fig. 3.9D**) but not the absence (**Fig. 3.9E**) of H2AK119Ub

in parental MSCs. We conclude that PRC1 is required for aberrant methylation of

H2AK119Ub-enriched regions including CGIs by DNMT3A mutants.

DISCUSSION

Our findings provide evidence for molecular cross-talk between PRC1-catalyzed

H2AK119Ub and DNMT3A that drives CGI hypermethylation upon mutational

inactivation of PWWP reader function (**Fig. 3.10**). We propose under steady-state

conditions PWWP-mediated recruitment of DNMT3A1 in cells predominates due to

greater abundance of H3K36me2 and H3K36me3 compared to H2AK119Ub (Weinberg

et al. 2019, Goldknopf & Busch 1977, West & Bonner 1980). Thus, loss of PWWP-

mediated recruitment either due to loss of substrate availability or impaired recognition

of H3K36 methylated nucleosomes causes DNMT3A1 to rely instead on UDR-mediated

recruitment. Alternative promoter usage that regulates expression of the short isoform

DNMT3A2 may serve as another mechanism for preventing DNA methylation of PRC1-

regulated regions in early developmental contexts (Chen et al. 2002). Consistent with

this, in mouse embryonic stem cells DNMT3A1 exhibits a stronger propensity to localize

to Polycomb-regulated regions than DNMT3A2 (Manzo et al. 2017).

Further structural characterization of the DNMT3A1 UDR domain will be

necessary to understand its selectivity for H2AK119Ub-modified nucleosomes and
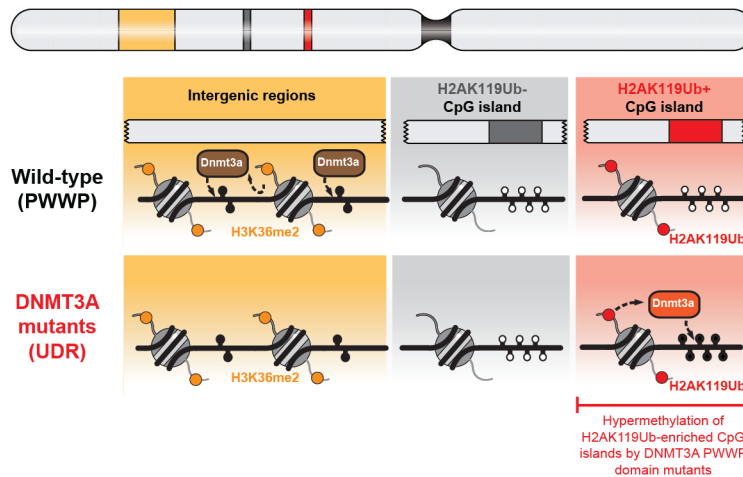
Figure 3.10: Model for CpG island hypermethylation by DNMT3A mutants. Model depicting changes in DNMT3A genomic localization patterns due to missense mutations associated with paragangliomas and microcephalic dwarfism. Wild-type DNMT3A is recruited to intergenic regions through PWWP-mediated recognition of H3K36me2. Gain-of-function mutations within the PWWP domain abrogate binding to H3K36-methylated nucleosomes, leading to a redistribution of DNMT3A to H2AK119Ub-enriched CpG islands through UDR-mediated recognition of H2AK119Ub. As a result, H2AK119Ub-enriched CpG islands become hypermethylated.

reveal similarities to other histone ubiquitylation reader modules, including the UDR

domain of 53BP1 which engages H2AK15Ub-modified nucleosomes through interaction

with the acidic patch of the nucleosome core particle (Wilson et al. 2016). Our findings

suggest that additional unannotated histone ubiquitylation reader modules may exist in

other chromatin regulators. However, the lack of amino acid sequence similarity

between the UDR domains of 53BP1 and DNMT3A1 may indicate a lack of a conserved

motif for identifying UDRs. This stands in contrast to previously identified ubiquitin-

interacting motifs (UIMs) which facilitate protein interactions with a hydrophobic surface

patch on ubiquitin regardless of whether it is conjugated to a histone protein (Beal et al.

1996, Hofmann & Falquet 2001, Swanson et al. 2003).

We propose DNMT3A missense mutations observed in paragangliomas and

microcephalic dwarfism share a common mechanism of action, although additional

study is required to assess their functional impact on cellular and organismal growth.

However, we note that interaction with H2AK119Ub-modified nucleosomes is an

inherent property of wild-type DNMT3A and is not a neomorphic capability conferred by

the disease-associated mutations. This raises the possibility that DNMT3A recruitment

to, and *de novo* methylation of, H2AK119Ub-enriched regions may occur in other

physiologic and pathophysiologic contexts through changes in the balance between

PWWP-mediated and UDR-mediated targeting. Indeed, *de novo* methylation of CGIs by

wild-type DNMT3A has been reported in hematopoietic stem cells upon cytokine

stimulation (Spencer et al. 2017) and many Polycomb-regulated gene promoters

become DNA methylated during neuronal differentiation (Mohn et al. 2008). Moreover,

phosphorylation of DNMT3A can direct its recruitment to heterochromatic regions and block its binding to promoters (Deplus et al. 2014, Kumar & Lassar 2014). It is interesting to note that DNA methylation changes during X chromosome inactivation largely involve hypermethylation of CGIs (Sharp et al. 2011), potentially implicating a switch to UDR-mediated recruitment of wild-type DNMT3A1 in this developmental process as well.

Further work is warranted to assess how post-translational modifications to DNMT3A affect targeting in development and disease. While CpG hypermethylation is observed at tumor suppressors and developmental regulators in many cancer types, certain tumors exhibit a CpG island methylator phenotype (CIMP) in which remodeling of the DNA methylation landscape is more extensive (Toyota et al. 1999, Issa 2004). The underlying mechanisms driving hypermethylation in CIMP+ tumors remain obscure, limiting our understanding of how different promoter-associated CGIs are selected for *de novo* methylation in tumors and their role in cancer progression (Hughes et al. 2013). Curiously, in colorectal cancer the presence of CIMP is highly correlated with activating mutations in BRAF and KRAS (Weisenberger et al. 2006). We speculate UDR-mediated recruitment of DNMT3A may be promoted by oncogenic signaling pathways to create the patterns of CGI hypermethylation observed in CIMP+ colorectal cancers and other diverse cancer types.

CHAPTER 4: GENERAL DISCUSSION

UNRAVELING THE DNA METHYLOME

Changes to the DNA methylation landscape have been implicated in the pathogenesis

of numerous diseases, including developmental growth syndromes (Tatton-Brown et al.

2014), imprinting disorders (Buiting et al. 1998, Gicquel et al. 2005, Monk et al. 2019),

and malignancies (Ley et al. 2010), among others. In many cases, these perturbations

have been linked directly to mutational alteration of DNA methylation machinery. The

conceptual framework for this thesis was to use insights from human genetics to better

understand how the DNA methylome is established, maintained, and altered in disease.

While extensive characterization of the DNA methylation landscape in a variety of

cellular contexts has given a robust snapshot of where CpG methylation is typically

enriched and depleted across the mammalian genome, our collective understanding of

the molecular mechanisms underpinning these landscapes has lagged. The *de novo*

DNA methyltransferases responsible for establishing DNA methylation have chromatin

reader domains that recognize specific histone PTMs. These reader domains impair *de*

novo DNMT recruitment to active promoters marked by H3K4me3 (Ooi et al. 2007,

Otani et al. 2009) and recruit DNMT3B to methylate within the body of actively

transcribed genes (Baubec et al. 2015, Morselli et al. 2015). However, these findings do

not provide a comprehensive understanding of DNA methylation dynamics as it has

long been appreciated that CpG methylation is widespread across the genome and not

limited solely to gene bodies (Schübeler 2015).

Profiling DNMT3A chromatin recruitment during my graduate studies revealed the existence of independent mechanisms for the establishment of DNA methylation at intergenic regions (Chapter 2) and Polycomb-regulated regions including CpG islands (Chapter 3), thereby providing an enhanced framework for understanding establishment of the DNA methylome. Insights for each of these processes was informed by molecular and clinical characterization of human developmental syndromes involving overgrowth (Tatton-Brown-Rahman and Sotos syndromes) and undergrowth (microcephalic dwarfism), respectively. The implications of these findings will be discussed in the subsequent sections, but key questions regarding how DNA methylation is established genome-wide remain. First, the catalytic domains of DNMT3A and DNMT3B exhibit distinct substrate preferences for naked DNA based on the sequence composition flanking CpG dinucleotides (Handa & Jeltsch 2005). Integrating this knowledge with histone-mediated recruitment mechanisms may be required for more complete understanding of DNA methylation patterns, especially in certain disease contexts. For example, recurrent missense mutations within the catalytic domain of DNMT3A are commonly observed in acute myeloid leukemia (Ley et al. 2010) and alter substrate sequence preference of the enzyme (Emperle et al. 2018). Second, previous work suggests DNMTs may interact with *trans*-acting factors such as the H3K9 di-methyltransferases G9A/GLP (Epsztejn-Litman et al. 2008, Dong et al. 2008, Chang et al. 2011) and the imprinting regulator ZFP57 (Quenneville et al. 2011) to maintain methylation patterns in certain contexts. However, whether these interactions also help to establish the DNA methylation landscape remains less clear. Finally, the amino-

terminal intrinsically disordered regions (IDR) of DNMT3A and DNMT3B await further study. While the IDR of DNMT3A is clearly dispensable for both PWWP-dependent and UDR-dependent targeting mechanisms, it could still influence recruitment by facilitating weak affinity interactions with other proteins to form molecular condensates in a similar manner to other previously studied IDRs (Hnisz et al. 2017).

DNA MAINTENACE METHYLATION: AN UNSOLVED PUZZLE

Mechanisms relating to *de novo* DNMT targeting provide insight into the establishment, but not the maintenance, of the DNA methylation landscape. This is due to the high fidelity of maintenance methylation performed by DNMT1 during S phase (Bostick et al. 2007). As a consequence, loss of DNA methylation cannot be explained by impairment of DNMT3A and DNMTB function alone (Chen et al. 2003). Indeed, we have confirmed in mouse MSCs that combined genetic ablation of *Dnmt3a* and *Dnmt3b* fails to induce DNA hypomethylation of H3K36me2-enriched intergenic regions or H3K36me3-enriched gene bodies. This indicates that the DNA hypomethylation observed at intergenic regions in sgNsd1/2 MSCs, sgNsd1 mESCs, and *NSD1*-mutant HNSCC lines must be due, at least in part, to an effect on regulators of DNA methylation other than *de novo* DNMTs.

Based on our current understanding of how DNA methylation can be erased, two non-mutually exclusive possibilities exist: 1) increased active demethylation by TET family enzymes or 2) increased passive demethylation due to impaired DNMT1 maintenance methylation. Evidence for the existence of both mechanisms is available in

other developmental contexts, such as the active and passive erasure of DNA methylation genome-wide in primordial germ cells and in preimplantation embryos (Messerschmidt et al. 2014). Increased levels of 5-hydroxymethylcytosine (5-hmC) upon knockdown of NSD family enzymes would provide support for enhanced TET enzyme activity, as 5-hmC is exclusively generated through active demethylation. However, how such a mechanism would operate remains unclear.

A more likely scenario based on prior work examining chromatin modification crosstalk is a defect in DNA methylation maintenance upon loss of NSD family enzyme function. Biochemical or genetic inactivation of NSD1/2 leads to the aberrant spreading of H3K27me3 by PRC2 at intergenic regions that are depleted of H3K36me2 (Lu et al. 2016, Streubel et al. 2018). H3K36 methylation appears to antagonize PRC2 activity through a sensing mechanism within its EZH2 catalytic subunit, which acts to prevent deposition of H3K27me3 outside of Polycomb-regulated regions (Jani et al. 2019). Notably, DNA methylation and H3K27me3 have also been observed to antagonize the presence of the other, although the underlying mechanism remains unclear (Bartke et al. 2010, Brinkman et al. 2012, Reddington et al. 2013). This suggests that increased intergenic H3K27me3 levels in NSD knockout cells may impair DNMT1 function to lead to DNA hypomethylation at these sites. A prediction of this model would be that inhibition of PRC2 function before knockdown of NSD family enzymes would rescue intergenic DNA methylation levels.

COMPETING, HIERARCHICAL RECRUITMENT OF A CHROMATIN READER

Our findings suggest that PWWP-mediated and UDR-mediated recruitment compete to regulate DNMT3A targeting. In parental MSCs, DNMT3A recruitment is primarily determined by PWWP domain binding to intergenic H3K36me2. Loss of PWWP reader function causes a shift in DNMT3A localization to H2AK119Ub-enriched regions through an alternative recruitment pathway involving its UDR domain. Furthermore, while DNMT3A$_{PWWP}$ has the potential to recognize both H3K36me2 and H3K36me3, DNMT3A only localizes to H3K36me3-enriched gene bodies upon global depletion of H3K36me2 levels. Global depletion of both H3K36me2 and H3K36me3 is required to redistribute DNMT3A to H2AK119Ub-enriched regions, thereby mimicking the effects of PWWP domain deletion. Targeting of DNMT3A therefore proceeds in a hierarchical manner, in which localization to H3K36me2-enriched intergenic regions is the most favored, followed by H3K36me3-enriched gene bodies, and finally H2AK119Ub-enriched regions (**Fig. 4.1A,B,C**).

We propose that such hierarchical localization preferences can be explained by competition between the mechanisms that regulate DNMT3A targeting. Based on our findings, at least two factors set the hierarchy: the binding affinity of DNMT3A for each substrate, and the molar ratio of the substrates in the cell. H3K36me2 is approximately 4-5 fold more abundant than H3K36me3 or H2AK119Ub based on quantitative mass-spectrometry measurements and exhibits a much broader distribution genome-wide, thereby making it an ideal sink for chromatin-bound DNMT3A. This may prevent DNMT3A from localizing to Polycomb-regulated regions despite its relatively high affinity

Figure 4.1: Competing, hierarchical recruitment of DNMT3A regulates establishment of the DNA methylation landscape. A) Under steady-state conditions, DNMT3A is recruited to intergenic regions due to PWWP-mediated recognition of H3K36me2 serving as the predominant recruitment mechanism. **B)** Upon global depletion of cellular H3K36me2 due to inactivation of NSD family enzymes in Sotos syndrome or head and neck squamous cell carcinomas, DNMT3A is recruited to actively transcribed gene bodies due to PWWP-mediated recognition of H3K36me3 serving as the predominant recruitment mechanism. Intergenic regions also become CpG hypomethylated as a result of reduced H3K36me2. **C)** Loss of PWWP reader domain functionality due to paraganglioma- or microcephalic dwarfism-associated mutations causes DNMT3A to rely on UDR-mediated recognition of H2AK119Ub for recruitment. As a result, H2AK119Ub-enriched regions including CpG islands and the inactive X chromosome have enhanced DNMT3A recruitment and become CpG hypermethylated. We speculate post-translational regulation of DNMT3A may also promote UDR-mediated recruitment through reducing PWWP reader functionality.
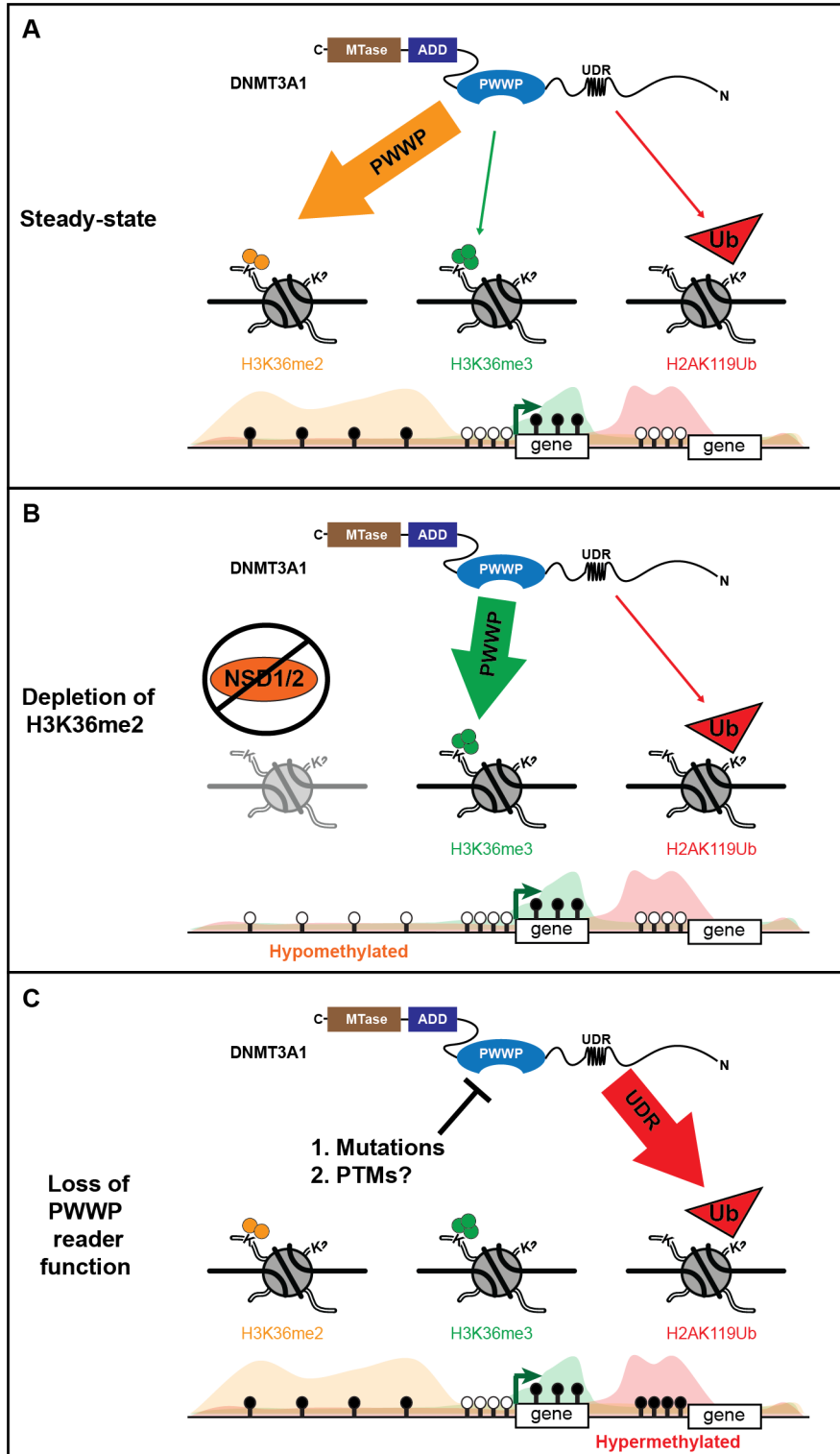
binding to H2AK119Ub-modified nucleosomes. However, other factors could also be at work and the contributions of the ADD domain in facilitating PWWP-mediated versus UDR-mediated binding awaits further characterization.

Similar to DNMT3A, many chromatin regulators contain multiple chromatin reader modules and previous work has largely focused on how adjacent reader domains can act in a cooperative fashion. Our findings highlight that in some cases reader domains of a chromatin regulator can actually compete for binding, with one recruitment mechanism ultimately predominating and determining localization. While this study has looked at mutational effects on DNMT3A recruitment, our findings suggest transcriptional and post-translational regulation of DNMT3A could also be used to toggle between PWWP- and UDR-mediated recruitment. Such a regulatory strategy could be a more widespread phenomenon that is employed to change the localization patterns of chromatin regulators depending on the developmental context. Alterations to the chromatin landscape that alter the molar ratio of a chromatin reader's substrates could also be utilized in physiologic and pathophysiologic contexts to induce a switch in recruitment mechanism. For example, changes to the cellular abundance of histone PTMs during development and aging is well-document (Booth & Brunet 2016). Further work is necessary to identify other chromatin regulators that may exhibit hierarchical localization and to determine the regulatory logic underlying the competing recruitment mechanisms.

CELLULAR FUNCTIONS OF DNMT3A RECRUITMENT

Key questions remain regarding the cellular role of DNMT3A-mediated DNA methylation at H3K36me2-enriched intergenic regions and H2AK119Ub-enriched Polycomb-regulated regions. For example, DNA methylation is classically associated with transcriptional repression and yet we find it is predominately enriched at TADs associated with active transcription. How to reconcile these findings? While H3K36me2-enriched TADs are broadly characterized by high levels of DNA methylation, focal sites of transcriptional regulation including active enhancers and promoters within these TADs are DNA hypomethylated. We propose that CpG methylation may serve as an activation barrier that prevents inappropriate transcription at euchromatic regions of the genome. Thus, DNA methylation would not be needed at H3K27me3-enriched facultative heterochromatin or at H3K9me3-enriched constitutive heterochromatin in which nucleosomes are tightly packed and not accessible to transcriptional regulators. In this way, *trans*-acting factors such as transcription factors would be required to deplete CpG methylation at their regulatory binding sites within accessible euchromatin regions to effect transcription, thereby providing an additional layer of transcriptional control.

Multiple areas of future investigation are warranted to better understand how DNA methylation might act as an activation barrier within euchromatic regions. It may be particularly important for genomic integrity, as endogenous retroviruses (ERVs) embedded within mammalian genomes are derepressed following DNA hypomethylation and could potentially reintegrate at new genomic sites (Groudine et al.

1981). Further work will be necessary to determine if ERVs specifically located within H3K36me2-enriched regions are transcriptionally activated upon ablation of NSD family enzymes. Other regulatory elements within intergenic regions may also be affected by DNA hypomethylation resulting from global depletion of NSD-catalyzed H3K36me2. As previously mentioned, the role of DNA methylation at intergenic enhancers may provide new insights into transcriptional regulation. For example, localization of some transcription factors is inhibited by methylation of their binding sites (Domcke et al. 2015). Additionally, binding of the zinc finger protein CTCF to insulator elements to regulate physical contacts between nearby enhancers and promoters is blocked by DNA methylation (Bell & Felsenfeld 2000, Wang et al. 2012). Many of these insulator elements are located at intergenic regions and control chromatin looping to facilitate long-range interactions of cis-acting regulatory elements (Phillips & Corces 2009). This suggests that intergenic DNA methylation may also play a role in shaping three-dimensional chromatin conformation. Thus, further profiling of the transcriptome and higher-order chromatin architecture in NSD knockout lines may provide valuable new insights into the cellular function of intergenic DNMT3A recruitment.

Questions also remain concerning the function of DNMT3A UDR-mediated recruitment to Polycomb-regulated regions. Our findings highlight that while some H2AK119Ub-enriched regions also have high levels of H3K27me3, others do not and are instead located within broad domains of H3K36me2. While the former likely represent targets of canonical Polycomb-mediated gene repression, the latter are poorly characterized. Could DNMT3A recruitment to H2AK119Ub-enriched, promoter-

associated CGIs provide an alternative mechanism for transcriptional repression independent of PRC2 function? And what is the consequence of hypermethylating regions that have high levels of both H2AK119Ub and H3K27me3? Interestingly, initial characterization of patient samples harboring *DNMT3A* PWWP domain mutations suggests that DNA hypermethylation is accompanied by reduced H3K27me3 (Heyn et al. 2019), consistent with previous reports of antagonism between these two modifications (Bartke et al. 2010, Brinkman et al. 2012, Reddington et al. 2013). This may reflect a shift of repression at these sites from being PRC2-dependent to being DNMT3A-dependent, thereby altering the regulatory factors needed to promote gene expression.

DEVELOPMENTAL FUNCTIONS OF DNMT3A RECRUITMENT

Mutations in *DNMT3A* result in developmental overgrowth or undergrowth depending on whether they are loss-of-function or gain-of-function, respectively. We propose that these developmental abnormalities can be traced back to mutational effects on DNMT3A recruitment. Our findings demonstrate that TBRS-associated mutations decrease protein stability and PWWP-mediated intergenic recruitment of DNMT3A, whereas dwarfism-associated mutations redistribute DNMT3A to H2AK119Ub-enriched regions and result in hypermethylation of a subset of CpG islands. While this study does not address the developmental functions of DNMT3A chromatin recruitment pathways, the clinical phenotypes reported in patients may provide some clues to inform future studies.

Defects in imprinting could provide a possible unifying explanation for the stark contrast in growth phenotypes caused by different germline mutations in DNMT3A. Many of the imprinted genes identified in humans regulate organismal growth and perturbations to their parent-of-origin expression patterns can result in developmental defects (Bartolomei & Ferguson-Smith 2011, Monk et al. 2019). DNMT3A is the *de novo* DNA methyltransferase responsible for establishing parental imprints during spermatogenesis as well as maternal imprints during oogenesis (Kaneda et al. 2004) and both DNMT3A1 and DNMT3A2 are expressed in both germ cell lineages (Lucifero et al. 2007, La Salle & Trasler 2006). Curiously, the most well-characterized paternally imprinted regions are intergenic and promote fetal growth whereas maternally imprinted regions are typically promoters and restrict fetal growth (Bartolomei & Ferguson-Smith 2011). Based on the conclusions of this work, it may be worthwhile to assess if DNMT3A establishes sex-specific imprinting patterns due to the utilization of PWWP-mediated recruitment in spermatogenesis, in contrast to oogenesis where UDR-mediated recruitment may predominate. Through a similar mechanism to what was proposed in Chapter 3 for X chromosome inactivation, these differences in recruitment preferences between cellular lineages could be through post-translational regulation of DNMT3A1. For example, signaling pathways unique to oogenesis could promote phosphorylation of DNMT3A1 that would impair PWWP domain reader functionality and thereby lead to alternative recruitment by the UDR domain. Germline mutations could disrupt the delicate balance between these two recruitment mechanisms, thereby perturbing imprinting and dysregulating developmental growth.

We propose that post-translational regulation of DNMT3A may be hijacked by oncogenic signaling pathways to inappropriately activate UDR-mediated recruitment, leading to hypermethylation of CpG islands and transcriptional repression of tumor suppressors. New lines of inquiry should be pursued to test this model. For instance, does hypermethylation in cancers lacking *DNMT3A* mutations occur selectively at H2AK119Ub-enriched CGIs as well? If so, this would provide compelling preliminary evidence for such a phenomenon. Additional work will be necessary to characterize the upstream signals that may bias DNMT3A to rely on its UDR domain for targeting, including whether DNMT3A acquires additional post-translational modifications during cellular transformation. If UDR-mediated DNMT3A recruitment is indeed broadly implicated in cancer-associated CGI hypermethylation, it may represent an attractive new therapeutic target for disrupting oncogenic gene expression programs.

Methods

<u>Plasmid construction and lentivirus production for cell culture</u>

sgRNAs directed against mouse *Dnmt3a*, *Nsd1*, *Nsd2*, *Setd2, Ring1,* and *Rnf2* were cloned into px458 (Addgene #48138, a gift from Feng Zhang). Mouse Dnmt3a and Dnmt3b cDNA sequences from Horizon Dharmacon were cloned into pCDH-EF1-MCS-Neo and piggybac (pCAGGS-IRES-Neo, a gift from H. Niwa, Institute of Molecular Embryology and Genetics, Kumamoto, Japan) with an N-terminal FLAG-HA epitope tag. Mouse Nsd1 cDNA from Horizon Dharmacon was cloned into pPB-CAG-3xFLAG-empty-pgk-hph (Addgene #48754, a gift from Austin Smith). All cloning including generation of deletion, domain-swap, and patient-associated mutations was performed using Gibson assembly (NEB). The TRBS missense mutations W297del, I310N and Y365C correspond to mouse Dnmt3a residues W293, I306 and Y361 respectively. The paraganglioma and dwarfism missense mutations K299I, R318W, W330R, and D333N correspond to mouse DNMT3A residues K295, R314, W326, and D329, respectively. To produce lentivirus, 293T cells were transfected with the lentiviral vector and helper plasmids (psPAX2, pVSVG). Supernatant containing lentivirus was collected and filtered 48 hours later for transduction.

<u>Cell culture, CRISPR/Cas9 gene editing, and generation of stable cell lines</u>

293T (ATCC) and C3H10T1/2 (ATCC) cells were cultured in Dulbecco's modified Eagles' medium (DMEM, Invitrogen) with 10% fetal bovine serum (FBS, Sigma). Fadu (ATCC), SKN-3 (JCRB cell bank), Cal27 (ATCC), SCC-4 (ATCC) cells were cultured in

DMEM:F12 medium (Invitrogen) with 10% FBS. V6.5 mouse embryonic stem cells (C57BL/6 x 129S4/SvJae F1) were maintained on gelatin-coated plates in Knockout DMEM (Gibco) supplemented with 15% ES-cell-qualified FBS (Gemini), 0.1 mM 2-mercapoethanol, 2mM L-glutamine (Life technologies) and LIF. *Drosophila* S2 cells were cultured in Schneider's *Drosophila* Medium (Invitrogen) containing 10% heat-inactivated FBS. All cell lines tested negative for mycoplasma contamination. To generate knockout lines, mESCs and C3H10T1/2 cells were transfected with sgRNA-containing px458 using Xfect mESC transfection reagent (Takara) or Lipofectamine 2000 (Invitrogen), incubated for forty-eight hours, and single GFP+ cells sorted into 96 well plates. Clones were expanded, screened for global reduction of H3K36me2/3 or H2AK119Ub by immunoblot, and individually verified by Sanger sequencing of the target loci. To generate transgenic C3H10T1/2 lines expressing epitope-tagged DNMT3A/B, cells were transduced with concentrated lentivirus as previously described (Papillon-Cavanagh et al. 2017). Transduced cells were grown under G418 selection (1000 µg/ml) 48hrs after transduction and selected for one week before being harvested for immunoblot or ChIP-seq. To generate transgenic mESC lines, ~5x10[6] cells were electroporated with piggybac expression vectors plus transposase (pBase) in a 3/1 ratio using Amaxa ESC Nucleofector Kit (VPH-1001, program A-023, Lonza). Cells were plated on gelatin-coated plates and grown under G418 selection (500 µg/ml) or Hygromycin selection (100 µg/ml) 48hrs after electroporation for at least 2 passages before being harvested for immunoblot or ChIP-seq analysis.

Immunoblotting

Fractionated or whole cell lysates were resolved by SDS-PAGE, transferred to a

nitrocellulose or PVDF membrane, blocked in 5% non-fat milk in PBS plus 0.5% Tween-

20, probed with primary antibodies, and detected with horseradish peroxidase-

conjugated anti-rabbit or anti-mouse secondary antibodies (GE Healthcare). Primary

antibodies were: anti-H3K36me2 (Cell Signaling Tech, 2901), anti-H3K36me3 (Active

Motif, 61101), anti-NSD1 (Abbexa, abx135901), anti-NSD2 (Millipore Sigma,

MABE191), anti-SETD2 (Abcam, ab31358), anti-DNMT3A (Abcam, ab188470), anti-

Vinculin (Cell Signaling Tech, 13901), anti-His (ZSGB-Bio,TA-02), anti-Lamin B1 (Cell

Signaling Tech, 12586), anti-β-Tubulin (Cell Signaling Tech, 2128), anti-β-actin (Abcam,

ab8224), anti-H3 (Abcam, ab1791) and anti-HA (Biolegend, 901501). The specificities

of anti-H3K36me3 and anti-H3K36me2 antibodies were validated using *Setd2* knockout

and *Nsd1/2* double knockout cell lines.

Histone acid extraction, histone derivatization and PTM analysis by nano-liquid

chromatography-mass-spectrometry (nano-LC-MS)

Cells were lysed in nuclear isolation buffer (15 mM Tris pH 7.5, 60 mM KCl, 15 mM

NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose, 10 mM sodium butyrate, 1 mM DTT,

500 μM AEBSF, 5 nM microcystin) containing 0.3% NP-40 alternative on ice for 5 min.

Nuclei were pelleted and resuspended in 0.4 N H_2SO_4 followed by 1.5 hr rotation at 4°C.

After centrifugation, supernatants were collected, proteins were precipitated in 33%

TCA overnight on ice, washed with acetone, and resuspended in deionized water. Acid-

extracted histones (5-10 µg) were resuspended in 100 mM ammonium bicarbonate (pH 8), derivatized using propionic anhydride and digested with trypsin as previously described (Sidoli et al. 2016). After a second round of propionylation the resulting histone peptides were desalted using C18 Stage Tips, dried using a centrifugal evaporator, and reconstituted using 0.1% formic acid in preparation for LC-MS analysis. Nanoflow liquid chromatography was performed using a Thermo Scientific™ Easy nLC™ 1000 equipped with a 75 μm x 20 cm column packed in-house using Reprosil-Pur C18-AQ (3 μm; Dr. Maisch GmbH, Germany). Buffer A was 0.1% formic acid and Buffer B was 0.1% formic acid in 80% acetonitrile. Peptides were resolved using a two-step linear gradient from 5% to 33% B over 45 min, then from 33% B to 90% B over 10 min at a flow rate of 300 nL/min. The HPLC was coupled online to an Orbitrap Elite mass spectrometer operating in the positive mode using a Nanospray Flex™ Ion Source (Thermo Scientific) at 2.3 kV. Two full MS scans (m/z 300-1100) were acquired in the orbitrap mass analyzer with a resolution of 120,000 (at 200 m/z) every 8 DIA MS/MS events using isolation windows of 50 m/z each (e.g. 300-350, 350-400…650-700). MS/MS spectra were acquired in the ion trap operating in normal mode. Fragmentation was performed using collision-induced dissociation (CID) in the ion trap mass analyzer with a normalized collision energy of 35. AGC target and maximum injection time were 5e5 and 50 ms for the full MS scan, and 3e4 and 50 ms for the MS/MS can, respectively. Raw files were analyzed using EpiProfile 2.0 (Yuan et al. 2018). The area for each modification state of a peptide was normalized against the total signal for that peptide to give the relative abundance of the histone modification.

dCypher nucleosome binding assays

5 µL of 250 nM GST-DNMT3A$_{PWWP}$ (Active Motif, 31541), GST-DNMT3A1 (Reaction Biology, DMT-21-125), or GST-DNMT3B (Reaction Biology, DMT-21-126) was incubated with 5 µL of 10nM biotinylated nucleosomes (EpiCypher, 16-9001) for 30 minutes at room temperature in Binding buffer (20 mM Hepes pH 7.5, 250 mM NaCl, 0.01% BSA, 0.01% NP-40, 1 mM DTT) in a 384-well plate. A mix of 10 uL of 2.5 µg/mL glutathione acceptor beads (PerkinElmer, AL109M) and 5 µg/mL streptavidin donor beads (PerkinElmer, 6760002) was prepared in Bead Buffer (20 mM Hepes pH 7.5, 250 mM NaCl, 0.01% BSA, 0.01% NP-40) and added to each well. The plate was incubated at room temperature in subdued lighting for 60 minutes and AlphaLISA signal measured on a PerkinElmer 2104 EnVision (680 nm laser excitation, 570 nm emission filter +/- 50 nm bandwidth). Each binding interaction was performed in duplicate.

Nucleosome Pull-Down Assays

1 µg biotinylated nucleosomes (Epicypher, 16-0006; 16-0322; 16-0319; 16-0320) were immobilized on 10 µL of BSA blocked Dynabeads MyOne Streptavidin C1 (Invitrogen, 65001) in 20 mM Tris pH 7.5, 100 mM NaCl, 0.1% NP40 and 1 mM DTT buffer plus protease inhibitors (Selleck, K4000) for 30 min on ice. The beads were then incubated with 1 µM or 10 µM His-tagged DNMT3A$_{PWWP}$ or DNMT3B$_{PWWP}$ proteins at 4°C for 4h. After five washes with 1 mL of binding buffer, the protein-bound beads were pooled and

the pull-down products were analyzed by Western blotting with α-His antibody (ZSGB-Bio, TA-02).

The recombinant DNMT3A$_{PWWP}$ and DNMT3B$_{PWWP}$ proteins for nucleosome pull-down assays were prepared by the following procedures: the DNMT3A$_{PWWP}$ (residues 278-427) domains of wild type and mutant human DNMT3A were cloned into a modified pRSFDuet vector with an N-terminal 6xHis-MBP tag. Recombinant proteins were overexpressed in *E.coli* strain BL21 (DE3) induced by 0.2 mM isopropyl-1-thio-D-galactopyranoside (IPTG) at 16°C overnight. Cells were harvested and resuspended in buffer containing 20 mM Tris, pH 8.0, 300 mM NaCl, and 5% glycerol. After cell lysis and centrifugation, proteins were purified by Dextrin Beads 6FF (Smart-Lifesciences, SA026100) and eluted with 20 mM maltose elution buffer. The eluted proteins were further purified by gel filtration Superdex 200 10/300 GL (GE Healthcare). Human DNMT3B$_{PWWP}$ (residues 206-355) was cloned into pET28b vector and then expressed at 16 °C overnight. The protein was purified by HisTrap column and Superdex 75 10/300 (GE Healthcare). All gel filtration buffer contained 20mM Tris–HCl pH 7.5 and 100mM NaCl for the nucleosome pull-down assay. Protein purity was analyzed by SDS-PAGE and western blot. Proteins concentrations were determined by UV spectroscopic measurement at 280 nm.

Recombinant DNMT3A$_{PWWP}$ expression and ITC titration assay

The PWWP domain (residues 278-427) of human DNMT3A (DNMT3A$_{PWWP}$) was sub-cloned into a modified pET28b vector with an N-terminal 6xHis-GB1 tag. Recombinant protein was overexpressed in E. coli BL21 (DE3) induced by 0.2 mM isopropyl-1-thio-D-galactopyranoside (IPTG) at 16°C overnight. Cells were harvested and resuspended in buffer containing 20mM Tris, pH 7.5, 300 mM NaCl, 10 mM imidazole and 5% glycerol. After cell lysis and centrifugation, the His-GB1-DNMT3A$_{PWWP}$ supernatant was purified by a HisTrap column (GE Healthcare). After PreScission protease digestion to cleave the 6xHis-GB1 tag, DNMT3A$_{PWWP}$ was further purified on a HiTrap Heparin HP Column (GE Healthcare). The protein sample was finally polished over a Superdex75 10/300 GL (GE Healthcare) column in buffer containing 20mM Tris, pH 7.5 and 150mM NaCl. Isothermal titration calorimetry (ITC) was performed at 25°C with a MicroCal PEAQ-ITC (Malvern Panalytical). DNMT3A$_{PWWP}$ was exchanged to ITC buffer containing 150 mM NaCl and 20 mM Hepes-Na, pH 7.4 by gel filtration. Lyophilized peptide was directly dissolved in ITC buffer for titration. Concentrations of DNMT3A$_{PWWP}$ and H3.3$_{1-42}$ or H3.1$_{1-42}$ peptides were 0.2 mM and 2 mM, respectively. Protein concentration was determined by the UV280nm absorbance. Peptides were quantified by weighing on a large scale. Acquired ITC data were analyzed by *Origin 8.0* (GE Healthcare) using the "One Set of Binding Sites" fitting model.

Chromatin Fractionation

C3H10T1/2 cells were washed with PBS and lysed in Buffer A (10 mM HEPES, 10 mM KCl, 1.5 mM MgCl$_2$, 0.34 M Sucrose, 10% Glycerol, 0.5 mM PMSF, 0.1% TritonX-100)

on ice for 8 min. Centrifugation was carried out at 1,300 x g at 4°C for 5 min and supernatant was collected (cytosolic fraction). The nuclei pellet was further lysed in Buffer B (3 mM EDTA, 0.2 mM EGTA, 0.2 mM PMSF) on ice for 30 min and centrifugation performed (1,700 x g at 4°C for 5 min) to obtain the supernatant (nuclear soluble fraction). Cytosolic and nuclear soluble fractions were combined to make the soluble fraction. The insoluble pellet was lysed in SDS sample loading buffer, boiled, and sonicated to yield the chromatin fraction. The protein concentration of each fraction was measured and equal amounts analyzed by immunoblot.

Immunolabeling with fluorescent in situ hybridization (ImmunoFISH)

Cells were grown on glass #1.5 coverslips coated with Poly-D-lysine in 24 well plates. For imaging, cells were washed with PBS and fixed for 10 min at room temperature with 4% formaldehyde solution. After washing twice with PBS, cells were permeabilized with 0.1% Triton X-100 in PBS for 5 min at room temperature. Immunolabeling was performed with a 1:500 dilution of anti-HA antibody (Biolegend, 901501) in PBS for 1 hr at room temp, followed by three washes with PBS, incubation with secondary antibody for 1 hr at room temperature, and then again washed three times with PBS. Cells were re-fixed with 4% formaldehyde for 10 min and then hybridization of the Stellaris RNA FISH probe for mouse Xist with Quasar 570 dye (BioSearch Tech, SMF-3011-1) was performed. First, cells were washed with PBS and incubated with Stellaris Wash Buffer A (BioSearch Tech, SMF-WA1-60) for 5 min. Coverslips were then transferred to a dark, humidified container and cells were incubated at 37 ºC for 4 hr with Stellaris FISH

hybridization buffer (SMF-HB1-10) containing a 1:100 dilution of the probe. Next, cells were washed with Wash buffer A for 30 min at 37 °C, counterstained with DAPI in Wash Buffer A for 30 min at 37 °C, and then incubated with Stellaris Wash Buffer B (BioSearch Tech, SMF-WB1-20) for 5 min at room temperature before being mounted onto a slide. Images were taken and deconvoluted using a DeltaVision Restoration Microscopy system (GE Healthcare) with an Inverted Olympus IX-70 microscope using a 40x objective.

Chromatin Immunoprecipitation

Cross-linking ChIP in mMSCs and mESCs was performed using ~2×10^7 cells per immunoprecipitation. Prior to fixation, media was aspirated and cells washed once with PBS. Cells were cross-linked directly on the plate using 1% paraformaldehyde for 5 min at room temperature with gentle shaking and glycine added to quench (final concentration 125 mM, incubated for 5 min at room temperature). Cells were washed once with cold PBS, scraped off the plates, and pelleted. To obtain a soluble chromatin extract, cells were resuspended in 1 mL LB1 (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1x Complete protease inhibitor) and incubated rotating at 4°C for 10 min. Samples were centrifuged, resuspended in 1 mL LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1x Compete protease inhibitor), and incubated rotating at 4°C for 10 min. Finally, samples were centrifuged, resuspended in 1 mL LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na deoxycholate, 0.5% N-lauroylsarcosine, 1% Triton

X-100, 1x Complete protease inhibitor), and homogenized by passing two times through

a 27-gauge needle. Chromatin extracts were sonicated for 8 min (anti-HA ChIP) or 12

min (anti-histone PTM ChIP) using a Covaris E220 focused ultra-sonicator at peak

power 140, duty factor 5, and cycles/burst 200. For histone PTM ChIP-Rx, after

centrifugation soluble chromatin was spiked-in with soluble chromatin from *Drosophila*

S2 cells that was similarly prepared and equivalent to 5-10% of the mouse/human cell

chromatin. The lysates were incubated with 100 µl Pierce anti-HA beads (Themo

Scientific, 88836) or with anti-H3K4me1 (Abcam, ab8895), anti-H3K9me3 (Abcam,

ab8898), anti-H3K27ac (Active Motif, 39133), anti-H3K27me3 (Cell Signaling Tech,

9733), anti-H2AK119Ub (Cell Signaling, 8240), anti-H3K36me2 (Cell Signaling, 2901) or

anti-H3K36me3 (Active Motif, 61101) antibodies bound to 75 µl protein A or protein G

Dynabeads (Invitrogen) and incubated overnight at 4°C with 5% kept as input DNA.

Magnetic beads were sequentially washed with low salt buffer (150 mM NaCl; 0.1%

SDS; 1% Triton X-100; 1 mM EDTA; 50 mM Tris-HCl), high salt buffer (500 mM NaCl;

0.1% SDS; 1% Triton X-100; 1 mM EDTA; 50 mM Tris-HCl), LiCl buffer (150 mM LiCl;

0.5% Na deoxycholate; 0.1% SDS; 1% Nonidet P-40; 1 mM EDTA; 50 mM Tris-HCl)

and TE buffer (1 mM EDTA; 10 mM Tris-HCl). For ChIP-seq, beads were resuspended

in elution buffer (1% SDS, 50 mM Tris-HCl pH 8.0, 10mM EDTA, 200 mM NaCl) and

incubated for 30 min at 65°C. After centrifugation the eluate was reverse cross-linked

overnight at 65°C. The eluate was then treated with RNaseA for 1 hr at 37°C and with

Proteinase K (Roche) for 1 hr at 55°C and DNA was recovered using Qiagen PCR

purification kit. For ChIP-Immunoblot, TE washes were replaced with TBS 0.1% Tween-

20, and the beads were resuspended in Laemmli buffer and boiled to de-crosslink

proteins before separation by SDS-PAGE. For HNSCC lines, sonication was performed

on a BioRuptor UCD-300 for 60 cycles at 4°C, 10s on followed by 20s off, centrifuging

every 15 cycles, and ChIP-Rx for H3K36me2 (Cell Signaling, 2901) using 2% S2

chromatin spike-in was performed on a Diagenode SX-8G IP-Star Compact using

Diagenode Automated Ideal ChIP-seq kit according to manufacturer's instructions.

ChIP-qPCR & ChIP-sequencing (ChIP-seq)

ChIP-qPCR was performed using the Applied Biosystems StepOnePlus system and

SYBR green dye. Fold enrichment was calculated by dividing the enrichment (% input)

at individual genomic loci by the averaged signal from multiple negative control regions

(e.g. H3K36me2-depleted or H2AK119Ub-depleted loci). For ChIP-seq, library

preparation was carried out using KAPA HTP Illumina or KAPA Hyper Prep library

preparation reagents following the manufacturer's protocol. ChIP libraries were

sequenced using Illumina HiSeq 2000, 2500 or 4000 at 50bp single reads.

RNA-sequencing

Total RNA was extracted from cell pellets using the AllPrep DNA/RNA/miRNA Universal

Kit (Qiagen) according to instructions from the manufacturer. Library preparation was

performed with ribosomal RNA (rRNA) depletion according to instructions from the

manufacturer (Epicentre) to achieve greater coverage of mRNA and other long non-

coding transcripts. Paired-end sequencing (125 bp) was performed on the Illumina HiSeq 2500 or 4000 platform.

Reduced Representation Bisulfite Sequencing (RRBS)

100 ng of genomic DNA were digested with 100U of MspI (NEB) and end-repaired/A-tailed using Kapa Hyper Prep kit (Kapa Biosystems). After ligation of Illumina-sequencing compatible indexes, DNA was purified using a 1X Agencourt AMPure XP bead clean up (Beckman Coulter). Bisulfite conversion was carried out using the Zymo EZ DNA kit (Zymo Research) using the following program: 55 cycles: 95°C 30s, 50°C 15 min, 4°C hold. Libraries were amplified 17 cycles using Uracil+ Ready mix (Kapa Biosystems, KK2801). The resulting libraries, mean size 388bp, were normalized to 3nM, pooled and clustered on a pair end read flow cell and sequenced for 100 cycles on a NovaSeq6000 (Illumina. Primary processing of sequencing images was done using Illumina's Real Time Analysis software (RTA). CASAVA 1.8.2 software was then used to demultiplex samples and generate raw reads and respective quality scores. Analysis of bisulfite treated sequence reads was carried as previously described (Garrett-Bakelman et al. 2015), using CUTADAPT instead of FLEXBAR to remove adapter sequences.

Whole Genome Bisulfite Sequencing (WGBS)

Whole-genome sequencing libraries were generated from 1000 ng of genomic DNA spiked with 0.1% (w/w) unmethylated λ DNA (Roche Diagnostics) previously

fragmented to 300–400 bp peak sizes using the Covaris focused-ultrasonicator E210. Fragment size was controlled on a Bioanalyzer High Sensitivity DNA Chip (Agilent), and the KAPA High Throughput Library Preparation Kit (KAPA Biosystems) or NxSeq AmpFREE Low DNA Library Kit (Lucigen) applied. End repair of the generated dsDNA with 3′ or 5′ overhangs, adenylation of 3′ ends, adaptor ligation, and clean-up steps were carried out as per manufacturer's recommendations. The cleaned-up ligation product was then analyzed on a Bioanalyzer High Sensitivity DNA Chip (Agilent). Samples were then bisulfite converted using the EZ-DNA Methylation Gold Kit (Zymo Research) according to the manufacturer's protocol. DNA was amplified by 9 cycles of PCR using the Kapa Hifi Uracil + DNA polymerase (KAPA Biosystems) according to the manufacturer's protocol. The amplified libraries were purified using Ampure XP Beads (Beckman Coulter), validated on Bioanalyzer High Sensitivity DNA Chips, and quantified by PicoGreen. Sequencing of the WGBS libraries was performed on the Illumina HiSeqX system using 150-bp paired-end sequencing.

Analysis of ChIP-seq data (Chapter 2)

Raw sequencing reads were aligned using BWA (Li & Durbin 2009) version 0.7.17 with default parameters to hg19, mm10 and dm6 for human, mouse and drosophila data, respectively. Aligned reads were summed into bins according to different annotations (CpG islands (CGIs), promoters, genic regions, 1kb, 10kb and 100kb bins) using bedtools (Quinlan & Hall 2010) version 2.22.1. Annotations of CGIs and RefSeq transcripts for hg19 and mm10 genomes were downloaded from UCSC Table Browser.

Promoters were defined as 5kb regions centered on RefSeq TSS. For genic regions, we used the longest isoform. ChIP-Seq coverage tracks were generated and visualized using IGV 2.3 (Robinson et al. 2011, Thorvaldsdottir et al. 2013). The ChIP enrichment was computed using ChIP-Rx (ChIP with reference exogenous genome), which uses spike-in Drosophila chromatin as internal control (Orlando et al. 2014). For each ChIP-Seq profile, the ChIP-Rx ratio (denoted as Rx) was calculated as follows:

$$Rx = \frac{SP/SP\_dmel}{IP/IP\_dmel},$$

Where *SP* and *SP_dmel* are the percentages of reads in the target sample mapped to mouse (or human) and spike-in Drosophila genome respectively (analogously, *IP* and *IP_dmel* for input). Tracks were normalized by dividing by the total sequencing depth and multiplying by the ChIP-Rx ratio. The signal was multiplied by 10^10 in all samples. For ChIP-Seq data analysis, binned reads were divided by total sequencing depth and normalized over input (using log2). *De novo* methylation plots were generated as in Baubec et al.

To call broad domains of H3K36me2, H3K27me3 and H3K9me3, ChIP signals (after input normalization) were segmented to identify changes in signal average and variance using PELT (Killick et al. 2012, Wambui et al. 2015). Segments were merged within a distance of 10bp, and then segments >500kb outputted as the large domains. Bedtools was used to aggregate the signals of CpG methylation, gene expression, and ChIP within these broad domains.

Gene body plots were generated using ngs.plot.r (Shen et al. 2014) with fragment length of 300bp and a flanking length of 10kb. For gene coordinates, we used a BED file of the top 50% most expressed genes, sorted by distance between the TSS and first internal exon, relative to total gene length, when appropriate.

Annotated genomic regions within H3K36me2 domains were identified by a custom PERL script. ChIP-seq reads for H3K36me2 and H3K36me3 in these regions were observed by featureCounts (Liao et al. 2014). Expected ChIP-seq reads with equivalent total reads number for H3K36me2 and H3K36me3 in the same regions were simulated by the "shuffle" function of Bedtools.

Analysis of ChIP-seq data (Chapter 3)

Raw reads were aligned to mouse (UCSC mm10) and *Drosophila* (UCSC dm6) genome builds using BWA version 0.7.17 (Li & Durbin 2009) with default parameters. All ChIP-seq samples were normalized relative to input DNA (see supplement for signal-input pairs) and normalized for read depth as described previously (Harutyunyan et al. 2019). We define S_i and N_i as the read counts in the i-th genomic compartment for the ChIP sample and input sample respectively. T_S and T_N are the total reads for the ChIP and input samples respectively. We use c to denote a pseudocount (1 for all the analyses presented here) which we use to avoid zeroes in the denominator when normalizing, and in logarithms downstream in the analysis. Finally, we calculate normalized signal for the i-th bin (S_i^{norm}) as follows:

$$S_i^{norm} = \log_2 \frac{(S_i + c)/T_S}{(N_i + c)/T_N}$$

When comparing between samples (parental vs. sgRing1a/b cells), ChIP-seq for broad histone marks (H3K27me3, H3K36me2, H3K36me3) were also normalized using exogeneous (*Drosophila*) reference chromatin (ChIP-Rx), which enables us to quantitatively compare ChIP-seq enrichment across experiments (Orlando et al. 2014). For ChIP-Rx, we calculate binned normalized signal as before, but we additionally multiply by the ChIP-Rx ratio (denoted Rx) for standardization across samples. We define s as the percentage of reads mapped to the mouse genome in the ChIP sample and s_{dmel} as the percentage of reads mapped to the spiked-in *Drosophila* genome in the ChIP sample. Similarly, i and i_{dmel} are defined for the corresponding input sample. Rx and Rx-normalized signal are then calculated as follows:

$$Rx = \frac{s/s_{dmel}}{i/i_{dmel}} \qquad\qquad S_i^{Rx-norm} = \log_2 \frac{Rx\,(S_i + c)/T_S}{(N_i + c)/T_N}$$

Depending on the analysis, scores were normalized genome-wide at either 200-bp or 10-kb bin resolution. For aggregate plots and heat maps (e.g. Fig. 1c,d), ChIP-seq signal was normalized for 200-bp bins, and then smoothed by taking the mean across five bins. The score in each 200-bp bin therefore describes the mean score over 1 kb centered at that bin. Normalization was performed using the *bamCompare* tool from *deepTools* version 3.3.1 (Ramírez et al. 2016) with appropriate arguments for adding pseudocounts, smoothing, scaling by Rx, T_N and T_S, and taking the $\log_2$ of the ratio.

Read count statistics (T_N, T_S, s, i) were obtained using the *flagstat* tool from *samtools* version 1.9 (Li et al. 2009). For analyses spanning specific regions such as CGIs, smoothed 200-bp scores were mapped onto bed intervals describing the regions of interest using the *bedmap --wmean* function from *BEDOPS* version 2.4.37 (Neph et al. 2012). Genome-wide analyses, 10-kb normalized ChIP-seq signal was used.

The annotations of RefSeq genes and CGIs for the mm10 genome were downloaded from the UCSC Table Browser and UCSC genome annotation database. For genic regions, we kept only the longest region (TSS to TES) if multiple isoforms exist. We considered promoters to be 1 kb region upstream from the TSS. For analysis of intergenic regions, we took the complement of all genes using the *complement* tool from *bedtools* version 2.27.1 (Quinlan et al. 2010), then retained only regions at least 10 kb long.

Genome-wide comparisons were performed as described below: Genome-wide pairwise correlations were generated using the *multiBigwigSummary* and *plotCorrelation* tools from *deepTools* version 3.3.1 (Ramírez et al. 2016). Enrichment heat maps and aggregate plots were also generated with *deepTools* (*computeMatrix*, *plotProfile* and *plotHeatmap* tools). ChIP-seq coverage tracks were visualized using IGV version 2.4.8 (Robinson et al. 2011). Peaks for H2AK119Ub were called using the *callpeak* function from *MACS2* version 2.1.2 (Zhang et al. 2008) with default parameters. Broad peaks/domains for H3K27me3 and H3K36me2 were called identically, except with the addition of the *--broad* flag. "Delta" values represent the

difference in log2-normalized score for each 10-kb bin between conditions (sgRing1a/b – parental). Subtracting scores between bedGraph files, plotting and correlation analysis were done using AWK and R scripts. Enrichment by genomic annotation was computed in R using the *annotatr* package (Cavalcante & Sartor 2017) and then plotted using *geom_bar* from *ggplot2*. To test enrichment of DNMT3A1 wild-type vs. ΔPWWP at CGIs, smoothed histograms of enrichment were generated for normalized ChIP-seq signal at CGIs and an equivalent number of shuffled regions of equal size to CGIs. These regions were randomly selected genome-wide, excluding blacklist regions, using the *shuffle* tool from *bedtools* (Quinlan et al. 2010). Number of overlaps between peak sets and CGIs was calculated with the *overlap* tool from *bedtools* (Quinlan et al. 2010). Jaccard indices were calculated using the *jaccard* tool from *bedtools* (Quinlan et al. 2010). Partial correlation analyses were conducted in R using the *ppcor* package (Kim 2015). All ChIP-seq analyses were restricted to regions outside of ENCODE blacklisted regions (Amemiya et al. 2019). For region-specific analyses, regions overlapping with blacklisted regions were excluded.

Analysis of RNA-seq data

Raw sequencing reads were aligned using STAR (Dobin et al. 2013) version 2.5.3a to the reference genomes mentioned above. Gene-level read counts were obtained from featureCounts v1.5.3 using the USCS gene annotation and normalized to total sequencing depth.

Analysis of RRBS data

Raw reads were aligned to the mouse genome (UCSC mm10) using Bismark (Krueger

& Andrews 2011). Alignments were merged for each sample's three replicates. The

resulting alignment files were processed using the *extract* and *mergeContext* functions

from *MethylDackel* version 0.4.0 (https://github.com/dpryan79/MethylDackel) with

options set to calculate methylation for each CpG, keeping only CpGs with a coverage

of at least 10X. To compare between samples (all combinations of parental or

sgRing1a/b and W330R or wild-type DNMT3A1), we restricted further analysis to CpGs

which meet the 10X coverage threshold across all samples. We therefore retained

2,122,547 CpGs genome-wide, of the 2,365,218 CpGs represented in at least one of

the four RRBS samples (89.74%). Smoothed histograms of per-CpG methylation scores

were generated in R using *geom_density* from *ggplot2*. All RRBS analyses were

restricted to regions outside of ENCODE blacklisted regions. For region-specific

analyses, regions overlapping with blacklisted regions were excluded.

Analysis of WGBS data

Raw reads were aligned to mouse or human genome builds (UCSC mm10 or hg19)

using BWA (version 0.6.1) after converting the reference genome to bisulfite mode.

Low-quality sequence at the 3′ ends were trimmed. For the overlapping paired-end

reads, we clipped the 3′ end of one of them to avoid double counting on both forward

and reverse strand. After alignment, we filtered duplicated or poorly mapping reads

(>2% mismatches or aberrant insert size). To call methylation of individual CpGs, we

used Samtools (Li et al. 2009) (version 0.1.18) mpileup. We filtered CpGs with less than

5x coverage, overlapping with SNPs from dbSNPs (build 137), or located within the

ENCODE DAC blacklisted regions or Duke excluded regions (Consortium 2012). For

visualization in IGV, the coordinates of individual CpGs were artificially extended to the

midpoint between their neighbors, as previously described (Hovestadt et al. 2014).

Partially methylated domains (PMDs), were called in a similar fashion (Berman et al.

2012). Average methylation levels were computed in 10 kb bins and neighboring bins

with methylation levels <70% merged into putative PMDs. Only regions >1mb were

considered true PMDs.

DNA methylation clustering

Illumina HumanMethylation450 BeadChip methylation beta values of Tatton-Brown-

Rahman Syndrome (TBRS), Sotos Syndrome and Weaver Syndrome patient and

control samples were obtained from the GEO database (accession numbers:

GSE74432, GSE128801). After removing probes that are on the sex chromosomes,

unsupervised hierarchical clustering was performed using the 1,000 most variable sites

between TBRS patient and control samples. Clustering was performed using the

pheatmap R function (Pretty Heatmaps v1.0.10) with parameters: clustering_method =

'complete', clustering_distance_cols = 'euclidean', and cutree_cols = 2.

References

Agathocleous, M. *et al.* Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature* **549**, 476-481 (2017).

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. *Molecular Biology of the Cell*. (Garland Science, 2002).

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci.* **51**, 786-794 (1964).

Almeida, M. *et al.* PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081-1084 (2017).

Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).

Arita, K. *et al.* Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc. Natl. Acad. Sci.* **109**, 12950-12955 (2012).

Asmar, F. *et al.* Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. *Haematologica* **98**, 1912-1920 (2013).

Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **79**, 137-158 (1944).

Avner, P., & Heard, E. X-chromosome inactivation: counting, choice, and initiation. *Nat. Rev. Genet.* **2**, 59-67 (2001).

Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-484 (2010).

Bartolomei. M. S. & Ferguson-Smith, A. C. Mammalian genomic imprinting. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520,** 243–247 (2015).

Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M., & Pickart, C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl. Acad. Sci.* **93**, 861-866 (1996).

Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-485 (2000).

Berman, B. P. *et al*. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44**, 40-46 (2012).

Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and Polycomb domain formation. *Cell* **157**, 1445-1459 (2014).

Booth, L. N., & Brunet, A. The aging epigenome. *Mol. Cell.* **62**, 728-744 (2016).

Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764 (2007).

Boulard, M., Edwards, J. R. & Bestor, T. H. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* **47,** 479–485 (2015).

Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128-1138 (2012).

Buiting, K. *et al.* Sporadic imprinting defects in Prader-Willi syndrome and Angelman syndrome: implications for imprint-switch models, genetic counseling, and prenatal diagnosis. *Am. J. Hum. Genet.* **63**, 170-180 (1998).

Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44**, 1179-1181 (2012).

Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics* **33**, 2381-2383 (2017).

Cavalli, G., & Misteli, T. Functional implications of genome topology. *Nat. Struct. Mol. Biol.* **20**, 290-299 (2013).

Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44,** 23–31 (2011).

Chang, Y. *et al.* MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferases GLP/G9a. *Nat. Commun*. **2**, 533 (2011).

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.* **277**, 38746-38754 (2002).

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

Chen, Z., Yin, Q., Inoue, A., Zhang, C., & Zhang, Y. Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. *Sci. Adv.* **5**, eaay7246 (2019).

Choufani, S. *et al.* NSD1 mutations generate a genome-wide DNA methylation signature. *Nat. Commu.* **6,** 10207 (2015).

Christman, J. K., Price, P., Pedrinan, L. & Acs, G. Correlation between hypomethylation of DNA and expression of globin genes in Friend erytholeukemia cells. *Eur. J. Biochem.* **81**, 53-61 (1977).

Cimmino, L. *et al.* Restoration of TET2 function blocks aberrant self-renewal and leukemia progression. *Cell* **170**, 1079-1095 (2017).

Clough, D. W., Kunkel, L. M. & Davidson, R. L. 5-azacytidine-induced reactivation of a herpes simplex thymidine kinase gene. *Science* **216**, 70-73 (1982).

Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

Cooper, S. *et al.* Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* **7**, 1456-1470 (2014).

Cooper, S. *et al.* Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nat. Commun.* **7**, 13661 (2016).

Dawlaty, M. M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell.* **29**, 102-111 (2014).

de Bustros, A. *et al.* The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia. *Proc. Natl. Acad. Sci.* **85**, 5693-5697 (1988).

de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* **7**, 663-676 (2004).

Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360**, 2289-2301 (2009).

Deplus, R. *et al.* Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. *Cell Rep.* **8**, 743-753 (2014).

Desrosiers, R. C., Mulder, C., & Fleckenstein, B. Methylation of herpesvirus saimiri DNA in lymphoid tumor cell lines. *Proc. Natl. Acad. Sci.* **76**, 3839-3843 (1979).

Dhayalan, A. *et al.* The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. *J. Biol. Chem.* **285,** 26114–26120 (2010).

Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575-579 (2015).

Dominguez, P. M. *et al.* TET2 deficiency causes germinal center hyperplasia, impairs plasma cell differentiation, and promotes B-cell lymphomagenesis. *Cancer Discov.* **8**, 1632-1653 (2018).

Dong, K. B. *et al.* DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* **27**, 2691-2701 (2008).

Doynova, M. D., Markworth, J. F., Cameron-Smith, D., Vickers, M. H. & O'Sullivan, J. M. Linkages between changes in the 3D organization of the genome and transcription during myotube differentiation in vitro. *Skelet Muscle* **7,** 5 (2017).

Dukatz, M. *et al.* H3K36me2/3 binding and DNA binding of the DNA methyltransferase DNMT3A PWWP domain both contribute to its chromatin interaction. *J. Mol. Biol.* **431**, 5063-5074 (2019).

Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYBP/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **23**, 406-420 (2008).

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

Emperle, M. *et al.* The DNMT3A R882H mutant displays altered flanking sequence preferences. *Nucleic Acids Res.* **46**, 3130-3139 (2018).

Epsztejn-Litman, S. *et al.* De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat. Struct. Mol. Biol.* **15**, 1176-1183 (2008).

Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817-825 (2010).

Fang, D. *et al.* The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science* **352**, 1344-1348 (2016).

Fradet-Turcotte, A. *et al.* 53BP1 is a reader of the DNA-damage-induced histone H2A Lys 15 ubiquitin mark. *Nature* **499**, 50-54 (2013).

Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **10**, 344-356 (2012).

Garrett-Bakelman, F. E. *et al.* Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base pair resolution. *J. Vis. Exp.* **96**, e52246 (2015).

Gicquel, C. *et al.* Epimutation of the telomeric imprinting center region on chromosome 11p15 in Silver-Russell syndrome. *Nat. Genet.* **37**, 1003-1007 (2005).

Goldknopf, I. L. & Busch, H. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc. Natl. Acad. Sci.* **74**, 864-868 (1977).

Gonzalez-Zulueta, M. *et al.* Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Res.* **55**, 4531-4535 (1995).

Graves, J. A. 5-azacytidine-induced re-expression of alleles on the inactive X chromosome in a hybrid mouse cell line. *Exp. Cell Res.* **141**, 99-105 (1982).

Greger, V., Passarge, E., Höpping, W., Messmer, E., & Horsthemke, B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum Genet.* **83**, 155-158 (1989).

Groudine, M., Eisenman, R. & Weintraub, H. Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. *Nature* **292**, 311-317 (1981).

Guo, R. *et al.* BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol. Cell* **56**, 298-310 (2014).

Guo, X. *et al.* Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* **517**, 640-644 (2015).

Handa, V., & Jeltsch, A. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *J. Mol. Biol.* **348**, 1103-1112 (2005).

Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486-489 (2000).

Harutyunyan, A. S. *et al.* H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. *Nat. Commun.* **10**, 1262 (2019).

Herman, J. G. *et al.* Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res.* **55**, 4525-4530 (1995).

Heyn, P. *et al.* Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.* **51**, 96 (2019).

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13-23 (2017).

Hofmann, K, & Falquet, L. A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* **26**, 347-350 (2001).

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

Hon, G. C. *et al.* 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* **56**, 286-297 (2014).

Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* **175**, 315-332 (1948).

Hovestadt, V. *et al*. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537-541 (2014).

Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D., & Zhao, Y. SnapShot: histone modification. *Cell* **159**, 458 (2014).

Hughes, L. A. *et al.* The CpG island methylator phenotype: what's in a name? *Cancer Res.* **73**, 5858-5868 (2013).

Issa, J. P. CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **4**, 988-993 (2004).

Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488-2498 (2014).

Jaju, R. J. *et al.* A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood* **98,** 1264–1267 (2001).

Jan, M., Ebert, B. L., & Jaiswal, S. Clonal hematopoiesis. *Semin. Hematol.* **54**, 43-50 (2017).

Jani, K. S. *et al.* Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase. *Proc. Natl. Acad. Sci.* **116**, 8295-8300 (2019).

Janssen, K. A., Coradin, M., Lu, C., Sidolia, S., & Garcia, B. A. Quantitation of single and combinatorial histone modifications by integrated chromatography of bottom-up and middle-down polypeptide tails. *J. Am. Soc. Mass. Spectrom.* **30**, 2449-2459 (2019).

Jeffries, A. R. *et al.* Growth disrupting mutations in epigenetic regulatory molecules are associated with abnormalities of epigenetic aging. *Genome Res.* **29**, 1057-1066 (2019).

Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074-1080 (2001).

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

Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903 (2004).

Killick, R. *et al*. Optimal detection of changepoints with a linear computation cost. *J Am Stat Assoc* **107**, 1590-1598 (2012).

Kim, S. ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665-674 (2015).

Kim, S. J. *et al*. A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood* **122**, 4086-4089 (2013).

Klein, C. J. *et al.* Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* **43,** 595–600 (2011).

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

Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**, 376-381 (2009).

Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. *Bioinformatics* **27**, 1571-1572 (2011).

Kumar, D. & Lassar A. B. Fibroblast growth factor maintains chondrogenic potential of limb bud mesenchymal cells by modulating DNMT3A recruitment. *Cell Rep.* **8**, 1419-1431 (2014).

Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44,** 609–620 (2011).

Kurotaki, N. *et al.* Haploinsufficiency of NSD1 causes Sotos syndrome. *Nat. Genet.* **30,** 365–366 (2002).

La Salle, S. & Trasler, J. M. Dynamic expression of DNMT3a and DNMT3b isoforms during male germ cell development in the mouse. *Dev. Biol.* **296**, 71-82 (2006).

Laugesen, A., Højfeldt, J. W. & Helin, K. Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Mol. Cell* **74**, 8-18 (2019).

Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363,** 2424–2433 (2010).

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

Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-365 (1993).

Li, H. & Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

Liao, Y., Smyth, G. K., & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

Lu, R., & Wang, G. G. Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem. Sci.* **38**, 546-555 (2013).

Lu, C. *et al.* Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* **352**, 844-849 (2016).

Lucifero, D. *et al.* Coordinate regulation of DNA methyltransferase expression during oogenesis. *BMC Dev. Biol.* **7**, 36 (2007).

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260 (1997).

Manzo, M. *et al.* Isoform-specific localization of DNMT3A regulates DNA methylation fidelity at bivalent CpG islands. *EMBO J.* **36**, 3421-3434 (2017).

Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770 (2008).

Merlo, A. *et al.* 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat. Med.* **1**, 686-692 (1995).

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

Min, J., Zhang, Y., & Xu, R. M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823-1828 (2003).

Mohn, F. *et al.* Lineage-specific Polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30**, 755-766 (2008).

Monk, D., Mackay, D. J. G., Eggermann, T., Maher, E. R., & Riccio, A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat. Rev. Genet.* **20**, 235-248 (2019).

Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11-24 (2011).

Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14 Spec No1:** R47-R58 (2005).

Morselli, M. *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* **4,** (2015).

Murrell, A., Heeson, S., & Reik, W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat. Genet.* **36**, 889-893 (2004).

Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919-1920 (2012).

Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* **39**, 237-242 (2007).

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

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

Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448,** 714–717 (2007).

Orlando, D. A. *et al.* Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. *Cell Reports* **9,** 1163–1170 (2014).

Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.* **10**, 1235-1241 (2009).

Papillon-Cavanagh, S. *et al.* Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat. Genet.* **49,** 180–185 (2017).

Peters, A. H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577-1589 (2003).

Phillips, J. E., & Corces, V. G. CTCF: master weaver of the genome. *Cell*, **137**, 1194-1211 (2009).

Plasschaert, R. N. & Bartolomei, M. S. Genomic imprinting in development, growth, behavior and stem cells. *Development* **141**, 1805-1813 (2014).

Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

Qin, S., & Min, J. Structure and function of the nucleosome-binding PWWP domain. *Trends Biochem. Sci.* **39**, 536-547 (2014).

Quenneville, 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* **44**, 361-372 (2011).

Quintana, R. M. *et al.* A transposon-based analysis of gene mutations related to skin cancer development. *J. Invest. Dermatol.* **133,** 239–248 (2013).

Rajakumara, E. *et al.* PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol. Cell* **22**, 275-284 (2011).

Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-W165 (2016).

Rao, B., Shibata, Y., Strahl, B. D. & Lieb, J. D. Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide. *Molecular and Cellular Biology* **25,** 9447–9459 (2005).

Rasmussen, K. D. *et al.* TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. *Genome Res.* **29**, 564-575 (2019).

Rayasam, G. V. *et al.* NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22,** 3153–3163 (2003).

Reddington, J. P. *et al.* Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).

Remacha, L. *et al.* Gain-of-function mutations in DNMT3A in patients with paraganglioma. *Genet. Med.* **20**, 1644-1651 (2018).

Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9-25 (1975).

Rinaldi, L. *et al.* Loss of Dnmt3a and Dnmt3b does not affect epidermal homeostasis but promotes squamous transformation through PPAR-γ. *Elife* **6,** (2017).

Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26 (2011).

Russler-Germain, D. A. *et al.* The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* **25**, 442-454 (2014).

Sager, R. & Kitchin, R. Selective silencing of eukaryotic DNA. *Science* **189**, 426-433 (1975).

Saxonov, S., Berg, P., & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci.* **103**, 1412-1417 (2006).

Scarano, E., Iaccarino, M., Grippo, P. & Parisi, E. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. *Proc. Natl. Acad. Sci.* **57**, 1394-1400 (1967).

Schlesinger, Y. *et al.* Polycomb-mediated methylation of Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* **39**, 232-236 (2007).

Schübeler D. Function and information content of DNA methylation. *Nature* **517**, 321-326 (2015).

Sendžikaitė, G., Hanna, C. W., Stewart-Morgan, K. R., Ivanova, E. & Kelsey, G. A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. *Nature Communications* **10**, 1884 (2019).

Sharp, A. J. *et al.* DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.* **21**, 1592-1600 (2011).

Shen, L. *et al.* Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* **3**, 2023-2036 (2007).

Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).

Sidoli, S., Bhanu, N. V., Karch, K. R., Wang, X. & Garcia, B. A. Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. *J Vis Exp.* **111**, e54112 (2016).

Song, J., Rechkoblit, O., Bestor, T. H., & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036-1040 (2011).

Soshnev, A. A., Josefowicz, S. Z., & Allis, C. D. Greater than the sum of parts: complexity of the dynamic epigenome. *Mol. Cell* **62**, 681-694 (2016).

Spencer, D. H. *et al.* CpG Island hypermethylation mediated by DNMT3A is a consequence of AML progression. *Cell* **168**, 801-816 (2017).

Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A., & Cedar, H. Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc. Natl. Acad. Sci.* **79**, 61-65 (1982).

Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).

Streubel, G. *et al.* The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in embryonic stem cells. *Mol. Cell* **70**, 371-379 (2018).

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* **279**, 27816-27823 (2004).

Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L, & Radhakrishnan, I. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *EMBO J.* **22**, 4597-4606 (2003).

Tatton-Brown, K. *et al.* Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* **46,** 385–388 (2014).

Tatton-Brown, K. *et al.* Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. *The American Journal of Human Genetics* **100,** 725–736 (2017).

Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678 (2012).

Thorvaldsdottir, H., Robinson J. T. & Mesirov J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14**, 178-192 (2013).

Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci.* **96**, 8681-8686 (1999).

Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a, and Dnmt3b. *Genes Cells* **11**, 805-814 (2006).

Vardimon, L., Kressman, A., Cedar, H., Maechler, M. & Doerfler, W. Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proc. Natl. Acad. Sci.* **79**, 1073-1077 (1982).

Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol Cell Biol.* **13**, 115-126 (2012).

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

Wambui, G. D. et al. The Power of the Pruned Exact Linear Time (PELT) Test in Multiple Changepoint Detection. *American Journal of Theoretical and Applied Statistics* **4**, 581-586 (2015).

Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-878 (2004).

Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680-1688 (2012).

Wang, H., Farnung, L., Dienemann, C. & Cramer, P. Structure of H3K36-methylated nucleosome-PWWP complex reveals multivalent cross-gyre binding. *Nat. Struct. Mol Biol.* **27**, 8-13 (2020).

Weaver, T. M., Morrison, E. A., & Musselman, C. A. Reading more than histones: the prevalence of nucleic acid binding among reader domains. *Molecules* **23**, E2614 (2018).

Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457-466 (2007).

Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* **38**, 787-793 (2006).

Wen, H. *et al.* ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* **508**, 263-268 (2014).

West, M. H. & Bonner, W. M. Histone 2B can be modified by the attachment of ubiquitin. *Nucleic Acids Res.* **8**, 4671-4680 (1980).

Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **38**, 1571-158 (2007).

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

Wilson, M. D., *et al.* The structural basis of modified nucleosome recognition by 53BP1. *Nature* **536**, 100-103 (2016).

Wu, H. *et al.* Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes. *Science* **329,** 444–448 (2010).

Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell* **49**, 1134-1146 (2013).

Xiao, T. *et al.* Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* **17,** 654–663 (2003).

Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402,** 187-191 (1999).

Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17-30 (2011).

Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983-7989 (2011).

Yuan, Z. F. *et al.* EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *J Proteome Res.* **17**, 2533-2541 (2018).

Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).

Zhen, C. Y. *et al.* Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. *Elife* **5**, (2016).

Zhou, V.W., Goren, A. & Bernstein, B.E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7-18 (2011).

Zhou, T. *et al.* Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol. Cell* **54**, 879-886 (2014).

Zylicz, J. J. *et al.* The implication of early chromatin changes in X chromosome inactivation. *Cell* **176**, 182-197 (2019).