

Interdependency Between Two Epigenetic Modifications At Select Loci In Mouse
Chromatin

Honors Thesis
Presented to the College of Agriculture and Life Sciences
Of Cornell University
in Partial Fulfillment of the Requirements for the
Biology Honors Program

by
Michael Peter Motley
May 2012

Supervisors James A. Hagarman, Paul D. Soloway

ABSTRACT

Chromatin contains numerous different epigenetic marks, including histone-tail modifications and DNA methylation. Many of these marks have varying effects on the expression state of a gene at different points in development and life. However, there is accumulating evidence that these modifications do not act independently, and that interaction between epigenetic marks is necessary for proper cell function. Our previous study showed mutual antagonism between two epigenetic marks, histone 3 lysine 27 trimethylation (H3K27me3) and DNA methylation (DNAm), upstream of the promoter of the *Rasgrf1* gene in mouse embryonic stem (ES) cells. H3K27me3 is deposited on chromatin by Polycomb Repressive Complex 2 (PRC2), and was shown to both restrict and be restricted by the placement of DNA methylation at this locus. Here, I look at the methylation profiles of DNA from wild type mouse ES cells and cells with a mutation in *Eed* of the PRC2 complex that results in a global loss of H3K27me3 to determine if this mutual antagonism occurs genome-wide. Using a combination of a Methyl DNA Immunoprecipitation microarray (MeDIP-chip) and sodium bisulfite sequencing, I show that H3K27me3 does indeed influence DNA methylation at numerous promoters in the embryonic mouse genome. Instead of seeing a consistent increase in DNA methylation after loss of H3K27me3 however, I found that only some of these promoters showed DNAm enrichment, while others showed depletion in DNAm. Additionally, several genes were shown to have both DNAm enrichment and depletion in the same promoter. My findings suggest that mutual antagonism is not observed consistently genome wide, and that more complexity exists regarding the interaction between these two marks.

INTRODUCTION

In eukaryotes, the expression of genetic material is regulated not only by transient transcription factors, but additionally by more permanent changes in its accessibility due to the rearrangement of chromatin(1). Chromatin, stored in the nucleus, is the association of DNA and several protein factors which include the nucleosome, an octomer of histone proteins that the DNA is wrapped around. The rearrangement of chromatin, due to factors such as the modification to these histones and to the DNA itself, has an impact on the packing of genes and their ability to be expressed. Two classes of modifications that are given great attention are post-translational modifications to histones and DNA methylation, which have been found to be key factors in pluripotency and the differentiation of the cell in early development.

Types of Epigenetic Marks

The N-terminal tails of histones are subject to several different covalent modifications, such as acetylation, ubiquitination, and mono-, di-, and trimethylation, and these alterations affect how chromatin is packed. For some marks, such as acetylation, the presence or absence of the mark is correlated with activation or repression of transcription, regardless of which residue it is placed upon (2-5). Methylation of histone tail residues however, is much more specific; while methylation of lysine 4 on the tail of histone 3 (H3K4me) correlates with active chromatin transcription(6-8), methylation at histone 3 lysine 9 (H3K9me) facilitates the formation of heterochromatin and gene silencing(8, 9). Trimethylation at Histone 3 Lysine 27 (H3K27me3) is a repressive mark that is placed on chromatin by the Polycomb Repressive Complex 2 (PRC2) and is associated with repression of genes during cellular development (10-14).

The methylation of cytosine nucleotides, or DNA methylation (DNAm) is also associated with repression of transcription. It primarily exists in the context of CpG

dinucleotides, and is usually found at most CpGs in the genome except at CpG-rich sites called CpG islands. These islands, usually found in the promoters of genes, are characteristically unmethylated, except in the case of transcriptional repression, present in both healthy and pathogenic cell types(15, 16). DNA methylation is laid down in early development after a global demethylation event prior to implantation (17-19). This *de novo* methylation of CpGs is performed by DNA Methyltransferases (DNMTs), namely DNMT3a and DNMT3b (20), and later maintained by DNMT1 (21-23). CpG islands become methylated when cells look to repress transcription of that gene. This is true both in healthy cells, where active methylation of CpG islands is associated with imprinting, X chromosome inactivation and the repression of transposable elements (15), and in cells that exhibit pathogenic phenotypes, where accumulation of methylation at these sites due to errors can cause genetic instability, pathogenic phenotypes, and cancer (16).

These modifications distinguish different functional regions of chromatin, or chromatin states, defined as the culmination of all of the epigenetic modifications at that region. Additionally, these epigenetic modifications are reversible. Methylation for example, can be passively demethylated by inactivity of DNMT1, or actively demethylated by proteins such as MBD2 and possibly the DNA repair pathway (24). As a result, epigenetic states are able to change in certain cell types at several points and developmental stages in an organism's life, and these states ultimately have an effect on overall gene expression at these times.

Interaction Between Marks

Recent research has shown that epigenetic marks do not act independently. Instead, it has been found that many of these marks collaborate with one another, require one another, or antagonize each other to function properly (25). For example, in embryonic stem cells (ES cells),

repressive H3K27me and activating H3K4me3 are found to exist simultaneously at sites that code for developmental transcription factors binding keeping genes expressed at low levels. However, this coincidence of marks, known commonly as a bivalent state, is thought to keep these genes “poised” for activation or silencing at a later time, and plays a role in the general developmental processes of an organism (14, 26).

Placement of chromatin marks may additionally be dependent on other marks. DNA methylation has indeed been shown to regulate deacetylation of histone residues (27, 28) and either prevent placement of methylation of H3K4 or actively demethylate H3K4me3 (27, 29). In contrast, an increasing amount of evidence has suggested DNA methylation requires histone methyltransferases for its placement. For example, EZH2, a component of the PRC2 complex, has been shown to directly interact with DNMT1, DNMT3A, and DNMT3B and be required for their recruitment (30). This finding is consistent with results of other studies linking H3K9 methyltransferase activity with DNA methylation (31-33). Additionally, unmethylated H3K4 has been demonstrated to recruit DNMT3L and facilitate *de novo* DNA methylation (34, 35), in seemingly direct opposition to the thought that it is the DNA methylation that influences H3K4 methylation (27, 29). However, both of these findings may indeed show some truth, and if so provide an example of a third kind of interaction, in which both marks affect each other simultaneously. This kind is best demonstrated in the observations of mutual antagonism.

In mutual antagonism, two epigenetic marks impede each other’s placement on chromatin. This interaction has been found in several studies, one of which showed that in *Arabidopsis thaliana*, histone 2 variant H2A.Z and DNA methylation were shown to limit each other (36). Loss of H2A.Z on chromatin due to a mutation in the Swr1 complex led to hypermethylation across the genome. Meanwhile enrichment and depletion of DNA methylation

on chromatin due to a mutation in the MET1 DNA methyltransferase of *A. thaliana* were shown to cause respective depletion and enrichments in H2A.Z at those same regions (36).

Our own laboratory has demonstrated an example of mutual antagonism in mammals in the imprinting control region upstream of the *Rasgrf* gene in mouse stem cells (37). In one of our experiments, treatment of mouse embryonic fibroblasts (MEFs) with DNMT inhibitor 5-azacytidine showed increased concentration of H3K27me3 placement at the imprinting control region of the paternal allele, a region that is typically highly methylated. In another, artificially inducing methylation at the normally unmethylated maternal locus decreased levels of H3K27me3 usual to that region. Additionally, we found that deactivation of a functional PRC2 complex in mouse embryoid bodies and trophoblast outgrowths led to a drop in H3K27me3 at the maternal locus and subsequent presence of methylation at the same site. We therefore concluded that DNAm actively impedes the placement of H3K27me3, and that reciprocally, H3K27me3 blocks placement of DNAm at the *Rasgrf* locus (37). This finding is supported by evidence from proteomic analysis of HeLa S3 cells that Eed and SUZ12, components of PRC2, are depleted on methylated DNA (38).

Mutual antagonism and other epigenetic interactions help check each other and help establish appropriate epigenomic states for proper cell function. As a result, errors in these interactions have been associated with pathogenic phenotypes (25). It has recently been shown for instance that ChIP analysis of colon cancer cells show colocalization of DNAm and H3K27me3 (39, 40), in contrast with our recent study (37). Others support this claim, believing that in cancer this colocalization “locks in” stem cells temporarily repressed via H3K27me3 by reinforcing the mark with DNAm (41) with assistance from H3K9me (42). Learning the mechanisms behind these chromatin mark interactions is crucial to understanding how they break

down in these periods of cell crisis, and whether they can be reconstituted to save a cell from cancer or developmental problems.

Investigation of H3K27me and DNAm Mutual Antagonism

Current work in the laboratory is looking at the interaction between DNAm and H3K27me3 genome-wide. Based on previous work (37), we are looking to determine if these two repressive marks display mutual antagonism in promoters outside of *Rasgrf1*. If such antagonism is present across the genome, we expect that the global depletion of one mark from chromatin will lead to the increase in the placement of the other. In this study, I look to answer half of this question, determining whether global depletion of H3K27me3 leads to an increase in DNAm at promoters in the mouse genome.

My study utilizes two lines of mouse embryonic stem cells, the first being a wild type V6.5 strain, and the second being an *eed*^{17Rn5-3554SB} (*eed*^{-/-}) mutant, in which the Eed protein of the PRC2 complex is rendered inoperable. Previous studies have shown that Eed is required in conjunction with EZH2 and SUZ12 for PRC2's successful operation, with a mutation in the protein leading to loss of repression (43), and undetectable levels of H3K27me3 on chromatin (44). We additionally confirmed this loss of H3K27me3 by a western blot (J.A. Hagarman, in preparation). As a result, the mutation should allow us to test whether or not DNAm has encroached on those spaces usually occupied by the histone mark in its absence

To test the presence of DNA methylation, we first performed a genome-wide Methyl DNA Immunoprecipitation and subsequent cohybridization of DNA from both strains to a promoter microarray (MeDIP-chip) on DNA from three replicate cultures of embryonic cells from each line (40). This array probed 50nt fragments within +1kb and -3kb of all promoters in the genome. Relative counts of methylated DNA from the V6.5 and *eed*^{-/-} cells were made at

each 50nt probe, and these counts were grouped into bins of 100 nucleotides. Aligning the relative methylated DNA counts across the genome among all three replicates showed peaks of enrichment and depletion of DNAm within promoters of mutant. This method allowed us to identify which promoters showed changes in DNAm as a result of H3K27me3 depletion. We validated these results by confirming a number of peaks via bisulfite sequencing. Bisulfite sequencing uses sodium bisulfite to convert those unmethylated cytosine bases in DNA into uracils, which are later converted to thymidine bases after PCR. However, the sodium bisulfite will not react with methyl-cytosine, and will thus leave the identity of the nucleotide intact (45). As a result, short stretches of DNA can be sequenced for the presence or absence of DNAm at single nucleotide resolution (46). We utilized this method to validate a number of amplified fragments of DNA from the two cell types, each of which corresponded to a region flagged by the MeDIP-chip analysis as a peak.

In summary, we used two independent methods to test our hypothesis that knockdown in levels of H3K27me leads to a global increase in DNA methylation at promoters across the mouse genome. Overall our data shows that H3K27me3 depletion does lead to significant changes in DNAm levels at ES cell promoters throughout the genome, and in many cases antagonized DNAm placement. In numerous other cases however, we found the converse phenomenon to be true, where H3K27me3 facilitated placement of DNAm. This evidence suggests more complexity than a genome-wide trend of mutual antagonism between these two marks.

METHODS

Extraction of DNA from ES cells

The DNA used in the experiment was extracted by J.A. Hagarman from ES cells generated from a V6.5 wild type and *Eed*^{17Rn5-3554SB} (*eed*^{-/-}) mouse cell line and eluted in QIAGEN EB Buffer (J.A. Hagarman, in preparation).

Methylated DNA Immunoprecipitation Microarray

The MeDIP was performed by J.A. Hagarman using a protocol outlined by Mohn, 2009 (47), using 10ug of anti-5-MeC antibody (Eurogentec, BI-MECY-0100) in IP buffer (10mM Na-Phosphate pH7, 140mM NaCl, 0.05% Triton), and anti-mouse M-280 Dynabeads (Invitrogen, 112.01D) (J.A. Hagarman, in preparation). The immunoprecipitate was then amplified using the GenomePlex Complete Whole Genome Amplification kit (Sigma, WGA2), and submitted to Cornell University Core Laboratories Facility, which ran the DNA on a NimbleGen Mouse DNA Methylation 3x720,000 CpG Island Plus RefSeq Promoter Array (NimbleGen, 05924537001). Three replicates of the cohybridization microarray were performed. Data analysis was performed by the Core Facility using NimbleGen NimbleScan© software under default settings.

Bisulfite Treatment and PCR

DNA extracted from ES cells was treated using the MethylEasy™ Xceed Bisulfite Conversion Kit (Human Genetic Signatures, ME002). The manufacturers protocol was followed as instructed, using 5µg of DNA in 20µL of DEPC H₂O, and eluting with 20µL of 70°C Reagent 5. I then amplified 2µL of the treated DNA from each sample in a 25µL PCR reaction using Ex Taq™ HS DNA Polymerase (TaKaRa). The conditions and mixture used mirrored that of the protocol described by the MethylEasy™ kit (*Mixture: DEPC H₂O: 16.875µL; 10x Ex Taq Buffer: 2.5µL; 2.5mM ExTaq dNTP mixture: 2µL; 20µM Forward primer: .75µL; 20µM Reverse*

primer: .75 μ L; 5U/ μ L Ex Taq DNA Polymerase: .125 μ L. Conditions: 95°C 3 minutes; (95°C 1 minute; 50°C 2 minutes; 72°C 2 minutes)x32 cycles; 72°C 10 minutes; 4°C 30 seconds). Primers used are shown in Table 1. I then ran each PCR product on a 1.5% agarose gel, and excised bands of the appropriate size (see Table 1). Bands were then purified using the MinElute® Gel Extraction Kit (QIAGEN) using the manufacturer's instructions for microcentrifuge use with the following alterations: elution buffer EB was preheated to 55°C prior to elution, and the eluate was used to elute the column a second time to increase DNA yield.

Ligation and Transformation of Treated DNA

I then ligated the eluted DNA into pCR®2.1-TOPO® using the TOPO TA Cloning Kit (Invitrogen, 45-0641) and transformed these vectors into TOP10 competent cells using the manufacturer's instructions. Cells were plated on ampicillin selective media, as well as X-gal for a blue-white screen, and incubated overnight, also according to manufacturer specifications (50 μ L and 100 μ L of inoculated SOC medium was used). Plates were stored at 4°C following incubation. I analyzed 48 white transformants for inclusion of each fragment using a PCR reaction outlined by the manufacturer, with the following alterations: (*15 μ L Mixture: DEPC H₂O: 9.7 μ L; 10mM dNTP: 30 μ L; Forward primer: .75 μ L; Reverse primer: .75 μ L, .2% BSA: .15 μ L, 5U/ μ L GoTaq® DNA polymerase: .15 μ L. Conditions: 95°C 2 minutes; (95°C 30 seconds, 55°C 30 seconds, 72°C 1 minute)x40 cycles; 72°C 5 minutes, 4°C 1 minute).* Only those transformants that clearly included the fragment of interest were used for sequencing.

Transformant DNA Collection and Sequencing

I grew 24 transformants of each amplicon and cell type (V6.5, eed) in 2mL LB cultures with ampicillin overnight. DNA from these cultures was then prepared using the QIAprep® Miniprep Kit and a QIAvac 24 Plus vacuum manifold (QIAGEN). Manufacturer's procedures were

followed, with the exception of the use of 40µL of elution buffer EB preheated to 55°C prior to elution, and the use of EZ-10 Spin Columns (Bio Basic Inc) instead of the spin columns provided in the kit. Each set of 24 samples were then sequenced by a Illumina/Solexa Genome Analyzer through the Cornell University Core Laboratories Facility, using a M13 Reverse primer compatible with the pCR®2.1-TOPO® vector.

Data Analysis of Bisulfite Sequencing

Sequence reads were compiled and analyzed using the online QUMA methylation statistical software (<http://quma.cdb.riken.jp/>). At least 15 of the 24 DNA reads from both the treated V6.5 line and treated eed^{-/-} line were compared with the reference sequence for differences in DNA methylation. Clones were discounted if they possessed less than 95.0% percent cytosine conversion or were less than 90.0% aligned with their genomic reference sequence. Results were evaluated using a Fishers Exact Test for methylation at each CpG.

Table 1: Primers Used To Validate DNA Methylation Peaks

Each primer set from P1-P7 was used in pair to amplify the corresponding regions on later figures. Primer sets P1-P5 were designed by J.A. Hagarman. Primer Sets are named as they appear in the figures below. Those primer sets that are unlabeled were unable to be amplified for various reasons later mentioned, and were not used in later analysis. The laboratory code for each primer is also listed, as well as the direction of the primer and its melting temperature (T_m).

Primer Set	Gene Name	Product Size	Lab Code	F/R	T_m	Sequence
P1	Gas5	421	PDS1980	Forward	59	TGAATGTATTGTTTTTTGAGAGTAATTG
			PDS1981	Reverse	60	AAAAAACCCACCATCAAATAAACT
P2	Pdgrl	425	PDS1758	Forward	60	TTGTTATTGGATGGTTTTTGTGTA
			PDS1959	Reverse	59	ACCCCTACTATCCTTAAACACACCA
P3	Pdgrl	482	PDS1760	Forward	59	TGGTGTGTTTAAGGATAGTAAGGGT
			PDS1761	Reverse	59	TAAAAAACCTTTTAAAAACCCCTC
P4	Miip/Fv1	250	PDS1835	Forward	58	GGTAAGGAATTTAATGATTTTGTGATT
			PDS1836	Reverse	60	AAACTTCAATCTATTTTCCCAAC
P5	Miip/Fv1	186	PDS1839	Forward	58	ATAGGTAGTTAGTTGTAGGGGGTTG
			PDS1840	Reverse	58	CCAATAATTAATTATCACCAATTTTATTTA
P6	Fbxo/ Ankrd16	240	PDS1918	Forward	56	TTATAGTGGAGATAAGTTAGGTTTGTAG
			PDS1919	Reverse	57	AAAAAAGTTCTATTATAAAATCTCCC
P7	Fbxo/ Ankrd16	276	PDS1926	Forward	58	AAGGATTAGAAATTTAAGAGAGATAAGGT
			PDS1927	Reverse	59	TTCCATAAACAACCCATAAATAACA
--	Tnp1	257	PDS1819	Forward	60	GGGGTTTAGTTGTGGAAGTATAAGG
			PDS1820	Reverse	56	AACAAAAGTTACTAACTCATATACACCAT
--	Tnp1	125	PDS1821	Forward	60	AATTGAATAAATATGGGGGAGGTT
			PDS1822	Reverse	60	AAAAAATCTAACACCCAAAACATCA
--	Prss21	289	PDS1823	Forward	59	TTGTTTTTTGTAGTAATATGATGGATGA
			PDS1824	Reverse	62	CTACCCTAAACCCCATCACTCAAC
--	Prss21	234	PDS1825	Forward	60	GAGAGAGGTAGTTAGAGTTAGGGGTGT
			PDS1826	Reverse	56	ACATACAAACATAATACAACAATCAAA
--	Prss21	260	PDS1827	Forward	56	TTTAGTTTGATTGTTGTATTATGTTTG
			PDS1828	Reverse	59	ATCACATTAATCTTCTACCTCTACCTC
--	Miip/Fv1	288	PDS1833	Forward	59	GGAAATAAAAGTTTGAAGATGAATTTTT
			PDS1834	Reverse	57	TTTATCCAAAAATACCAAAAAACAAC
--	Miip/Fv1	272	PDS1837	Forward	56	GAGTTTAAATAGGAGGAGTTTTTTT
			PDS1838	Reverse	60	AAAAACCTTCATTAACACAATCCA
--	Herc3	283	PDS1841	Forward	60	TTTTTTTGTGTTTGTGTTTGGTTGT
			PDS1842	Reverse	56	CCTACATTAAGTTAAAAAATCTCAAAT
--	Herc3	265	PDS1843	Forward	60	AAAGGGGTAGGGGATTTATTATTTG
			PDS1844	Reverse	59	AACCAAATTTAACAAAAAATTCCAA
--	Herc3	106	PDS1845	Forward	59	TTGGAATTTTTGTAAATTTGGTT
			PDS1846	Reverse	56	CTTTATAAAAAAGTTACCCAATTCC
--	Herc3	281	PDS1847	Forward	60	GGTAAGTTTTTTATAAAGTTTTTTTTGGG
			PDS1848	Reverse	56	CAATCACAACACTACAAAACCTCTCT
--	Herc3	160	PDS1849	Forward	56	TTTGTAAATAGAGTTGGTTTAGAGAGG
			PDS1850	Reverse	59	CAACCATCAACTAATAATAAACAACCA

--	Herc3	181	PDS1851	Forward	58	GAATGGTTGTTTATTATTAGTTGATGG
			PDS1852	Reverse	58	CAAATATAAAAAAATACAATCTCAACCTC
--	Fbxo/ Ankrd16	207	PDS1920	Forward	59	AGAGGATTTGTTTTTTTAGGTGTTG
			PDS1921	Reverse	57	AAAAAACTTCCTATTATAAAATCTCCC
--	Fbxo/ Ankrd16	246	PDS1922	Forward	58	GAAGGAAATTTTTTTAGAGTTTTTTAGA
			PDS1923	Reverse	60	CTAAAACCTAACTAACCCACCCCTA
--	Fbxo/ Ankrd16	230	PDS1924	Forward	59	GAGTTTTTTAGAGTTTTTTGGGTTTT
			PDS1925	Reverse	60	CTAAAACCTAACTAACCCACCCCTA
--	Fbxo/ Ankrd16	286	PDS1928	Forward	56	TTTGTAAGGAAGGAGTAGGTAGTTT
			PDS1929	Reverse	56	CAATTTCTTTACATTTAACAATACATTC

RESULTS

ES cell deficiency in H3K27me3 causes DNAm enrichment in 861 genes

The three independent MeDIP-chip experiments surveyed 20,404 promoters and 15,980 CpG islands. Of these regions, the NimbleScan© software found evidence of significant DNAm changes in 2,933 promoters and 1,413 genes according to the Ensembl annotation of the NCBIM37 assembly of the mouse genome. The three arrays were shown to have good correlation of peak intensities (0.74484, 0.70824, 0.64928, pairwise Pearson correlation). Of these identified genes, 861 displayed increased DNAm in the *eed*^{-/-} cell line. Validation of one of the enriched peaks which correlated to a promoter of the *Gas5* gene [Figure 1A] showed high significance at two of the five dinucleotides ($p < .01$, Fisher's exact test) in the amplicon, confirming that DNA methylation increased in the *eed*^{-/-} line [Figure 1B-D]. This gene was also shown to be enriched for H3K27me3 in the absence of DNAm, and is believed to exhibit mutual antagonism similar to that of *Rasgrfl* (J.A. Hagarman, in preparation). Gene ontology analysis of those genes with increased DNA methylation in *eed*^{-/-} cells showed that many of these genes help operate either sensory perception or development (J.A. Hagarman, in preparation).

ES cell deficiency in H3K27me3 causes DNAm depletion in 552 genes

Despite those 861 genes showing increases in DNA methylation in the *eed*^{-/-} line, 552 genes of the 1,413 showed decreased methylation. These genes tended to have higher CpG content in their promoters and corresponded with chromatin organization. (J.A. Hagarman, in preparation) Analysis of the two amplicons used to validate depletion of the promoter at the *Pdgrfl* growth factor gene [Figure 2A] showed only one peak of those analyzed showed high significance ($p < .001$, Fisher's exact test) [Figure 2B-D]. Of the CpGs in the amplicon, site 36 was found to possess a single nucleotide polymorphism at site creating an artificial CpG site that was read as

significant. However despite this result, the significance of the one true CpG is still adequate in confirming a drop in DNAm at that site.

ES cell deficiency in H3K27me3 causes both enrichment and depletions at 18 sites.

In addition to identifying promoters which had only depleted or enriched levels, the MeDIP-chip also identified 18 genes that showed significant levels of both enrichment and depletion in DNAm in the *eed*^{-/-} cells [Table 2]. Validation was performed for two of these promoters, one that encompasses the bidirectional promoters of *Miip* and *Fv1* on chromosome 2 [Figure 3A], and one that includes the *Fbxo18* and *Ankrd16* genes on chromosome 4 [Figure 4A]. The bisulfite sequence data shows that three of the four CpGs found in the enriched amplicon of the *Miip* and *Fv1* genes had significant increases in DNAm in *eed*^{-/-} cells ($p < .035$, Fisher's exact test) [Figure 3B,C,F], and both CpGs found in the depleted amplicon of the gene pair had significant decreases in DNAm ($p < .05$, Fisher's exact test) [Figure 3D, E,G]. Contrastingly, while one of the CpGs in the depleted amplicon of the *Fbxo18-Ankrd 16* pair displayed a highly significant decrease in DNAm in the *eed*^{-/-} line ($p < .0003$, Fisher's exact test) [Figure 4D,E,G], none of the 13 CpG sites tested in the enriched region of the pair showed any significant change in DNAm between the wild type and mutant lines [Figure 4B,C,F]. It should be noted however that the PCR product used to validate this amplicon existed at one end of the peak, and that differences in DNAm may still be found elsewhere within the peak, although due to bisulfite conversion constraints we were unable to design primers to span its entirety. I should also indicate that these seven amplicons were not the only regions surveyed, but that these were the only products of 24 [See Table 1] that could be successfully sequenced; many of the PCR products either failed to amplify, failed to be gel purified or ligated, or failed to be sequenced

[See Table 3]. Overall, three of the four peaks used to validate those genes with both enriched and depleted genes were confirmed, while one peak remains to be validated.

Table 2: Genes with promoters found to have enrichment and depletion in DNAm in H3K27me3 deficient ES cells

Nucleotide position of the promoter regions found by the microarray to have both enrichment and depletion are shown, in addition to the names of the genes those promoters correspond to. Enrichment and Depletion scores are represented as \log_2 increases or decreases, based on sign. For those three gene pairs, both nucleotide start and end sequences for respective promoters are shown.

Chr #	Promoter Start	Promoter End	Gene Name	Depletion (\log_2)	Enrichment (\log_2)
1	196443022	196447022	Plxna2	-1.60	0.98
2	11696379	11700379	Fbxo18/Ankrd16	-1.22	1.38
2	11698154	11702154			
4	113832988	113836988	Skint11	-1.55	1.24
4	147240087	147244087	Fv1/Miip	-1.95	1.48
4	147241828	147245828			
5	147641465	147645465	Rpl21	-1.13	1.14
6	52153586	52157586	Hoxa3	-1.21	1.29
6	58856120	58860120	Herc3	-1.59	1.67
7	92061290	92065290	Olfr290	-1.52	1.11
7	110862109	110866109	Olfr243	-1.21	1.03
8	36647851	36651851	Mfhas1	-1.23	2.2
8	83014943	83018943	Gypa	-1.49	1.52
8	109414539	109418539	Chtf8/Cirhla	-1.13	2.53
8	109416493	109420493			
9	110847662	110851662	Tdgf1/Crrc2	-1.00	1.58
10	129310811	129314811	Olfr814	-1.39	1.67
11	121688531	121692531	Ptchd3	-1.85	1.43
14	80170119	80174119	Pcdh8	-1.37	1.50
15	61813895	61817895	Myc	-1.28	1.09
16	59215437	59219437	Olfr199	-1.44	1.43

Table 3: Summary of Procedures Performed and Success of Procedures

The total number of laboratory procedures I performed in this experiment is listed with how many of these procedures succeeded in producing a relevant product for further experiments/data analysis. These procedures are outlined in Methods. A success of a Bisulfite PCR reaction was recorded as the presence of a definable band of DNA in the purification gel. As the amount of DNA recovered from the gel purification was not measured, success of this procedure was recorded as whether the product purified was successfully transformed into cells in sufficient numbers to perform 48 transformant verification PCR reactions. A success of a transformant verification PCR reaction was defined as one where at least 24 of the transformants contained the desired fragment. The success of a sequencing reaction was considered as a PCR fragment whose parallel sequencing produced at least 15 of 24 sequences that contained less than 10% sequence mismatches and less than 5% unconverted CpH dinucleotides as analyzed by QUMA.

<u>Procedure Type</u>	<u>Attempts</u>	<u>Successes</u>	<u>% Success</u>
Bisulfite Treatment of DNA	14		
Bisulfite PCR	152	82	53.9
Gel Purification, Ligation, and Transformation of Products	100	56	56.0
Transformant Verification PCR (x48)	42	32	76.2
Sequencing Reaction (x24)	31	16	51.6

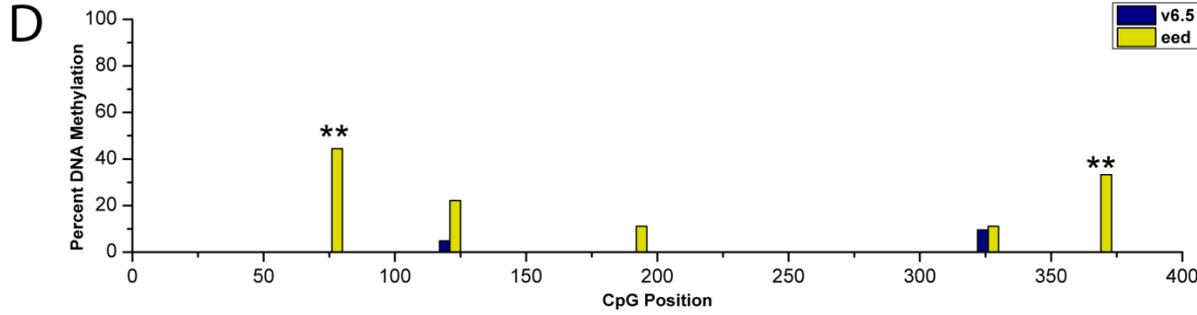
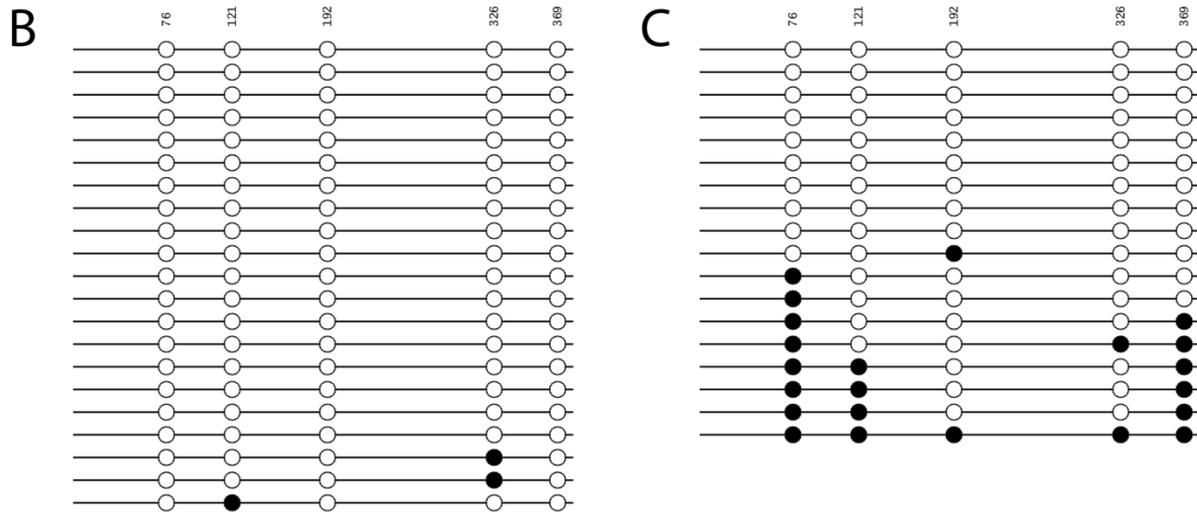
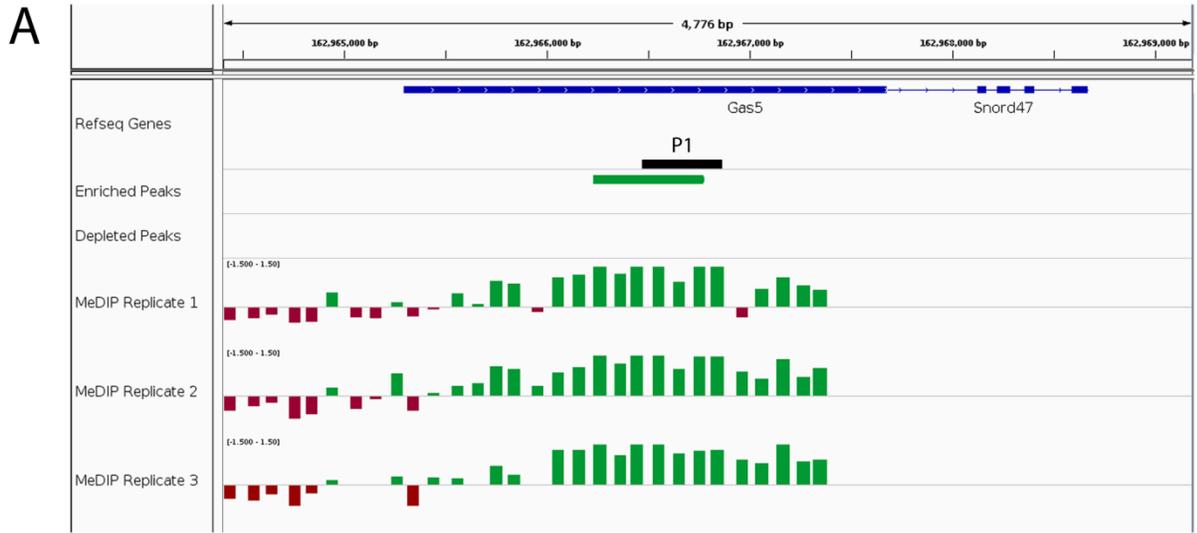


Figure 1. Validation of Enriched Methylation Peak at Gene *Gas5*. (A) The location of *Gas5* in the mouse genome is shown above and data from the 3 MeDIP replicates is shown at bottom. Green vertical bars represent the \log_2 degree of enrichment of DNAm at the corresponding bin, while red vertical bars represent the \log_2 degree of depletion. The calculated average of the three replicates generated an enriched peak, shown as a green bar spanning the length of the peak. Also shown is a black bar indicating where the validation PCR product spans in relation to the peak (see Table 1). (B,C) The methylation profile of DNA at the PCR product P1 in V6.5 and *eed*^{-/-} ES cells, respectively. Each circle represents one CpG as found in the product, and each row represents an independent clone used in the analysis. White circles indicate that the CpG was not methylated at that site, while black circles indicate methylation at that CpG. (D) A graphical comparison of methylation between V6.5 and *eed*^{-/-} ES cells at each CpG site in PCR product P1. Blue bars indicate percent methylation among clones at a given CpG site in V6.5. Right adjacent yellow bars indicate percent methylation among clones in *eed*^{-/-}. The horizontal axis shows relative position of each CpG in the PCR product. ** indicates high significance at $p < .01$ (Fisher's exact test).

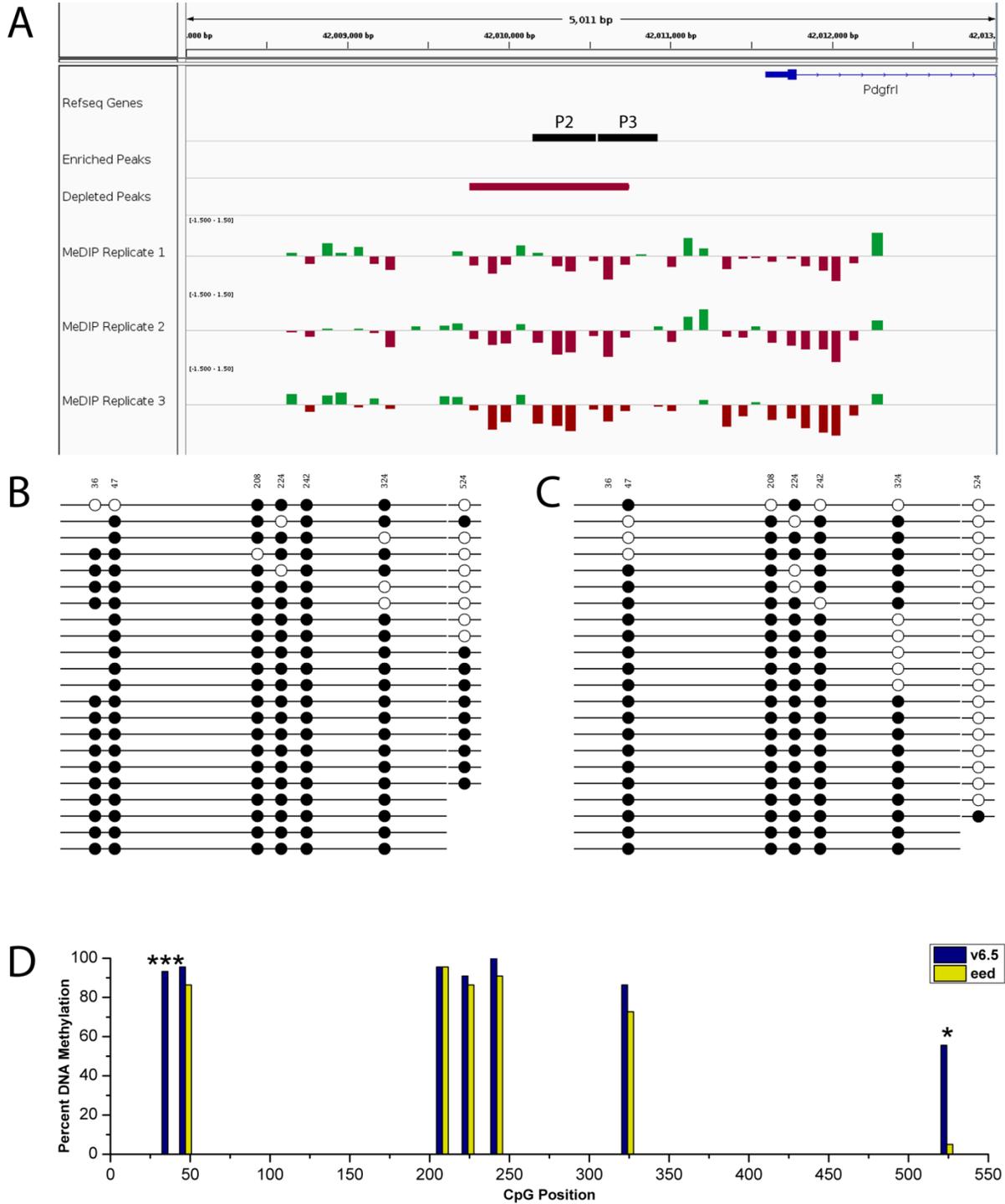


Figure 2. Validation of Depleted Methylation Peak at Gene *Pdgfrl*. (A) The location of *Pdgfrl* is shown with data from the 3 MeDIP replicates. The calculated average peak of depletion is shown as a red bar spanning the length of the peak, and black bars indicating where the validation PCR products P2 and P3 spans in relation to the peak (see Table 1). (B,C) The methylation profile of DNA at products P2 and P3 in V6.5 and *eed*^{-/-} ES cells, respectively. (D) A graphical comparison of methylation between V6.5 and *eed*^{-/-} ES cells at each CpG site in PCR products P2 and P3. * indicates significance at $p < .05$ (Fisher's exact test). *** represents a SNP at site 36 in which no CpG is found in the *eed*^{-/-} cell line.

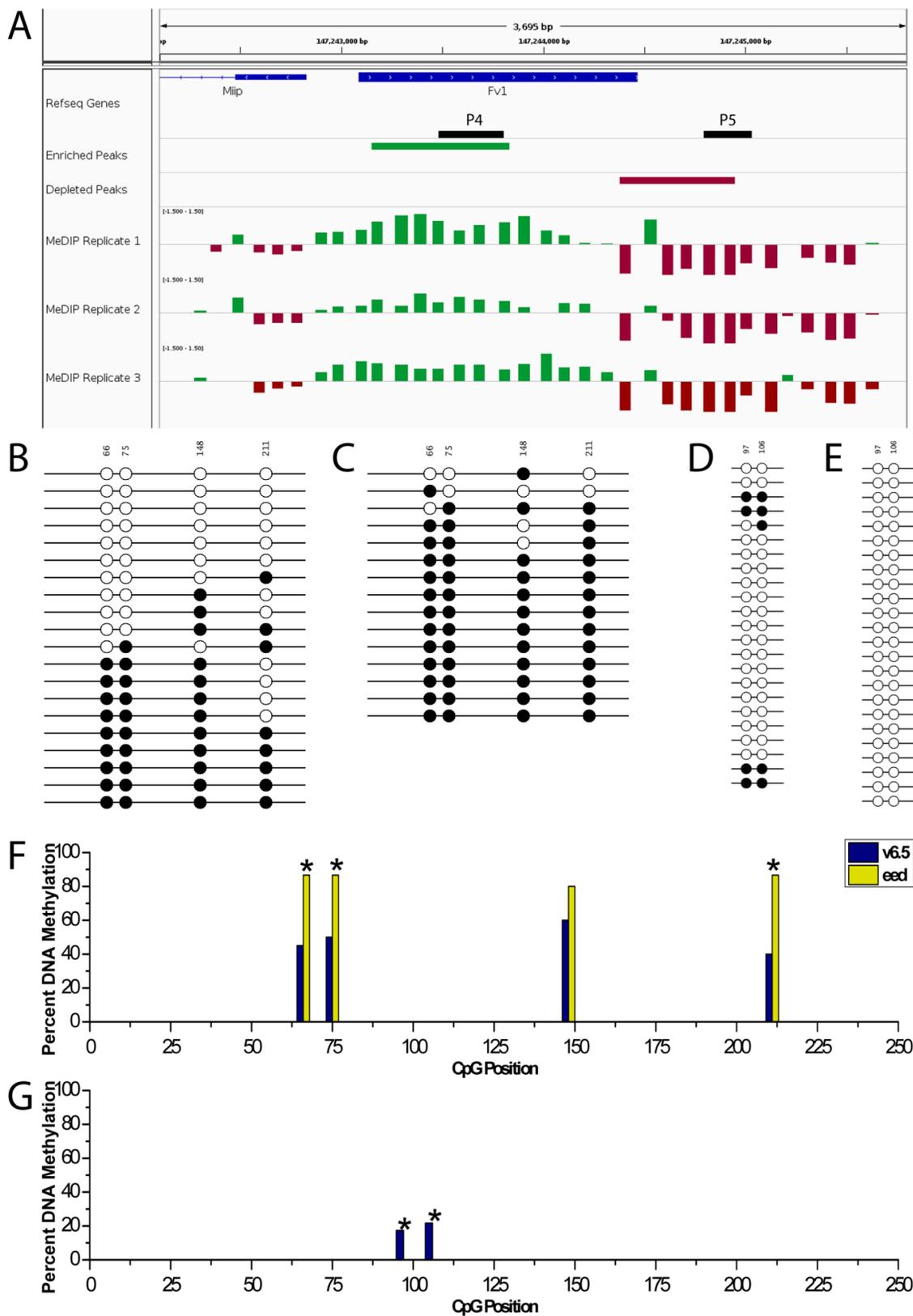


Figure 3. Validation of Depleted Methylation Peak at Gene Pair *Fv1/Miip*. (A) The location of the pair is shown with data from the 3 MeDIP replicates. The calculated average peaks of enrichment and depletion are shown as green and red bars spanning the length of the peak, respectively. Black bars indicate where the validation PCR products P4 and P5 span in relation to the peaks (see Table 1). (B,C) The methylation profile of DNA at product P4 in V6.5 and *eed*^{-/-} ES cells, respectively. (D,E) The respective methylation profile of DNA in V6.5 and *eed*^{-/-} ES cells. (F, G) A graphical comparison of methylation between V6.5 and *eed*^{-/-} ES cells at each CpG site in PCR products P4 and P5, respectively. * indicates significance at $p < .05$ (Fisher's exact test).

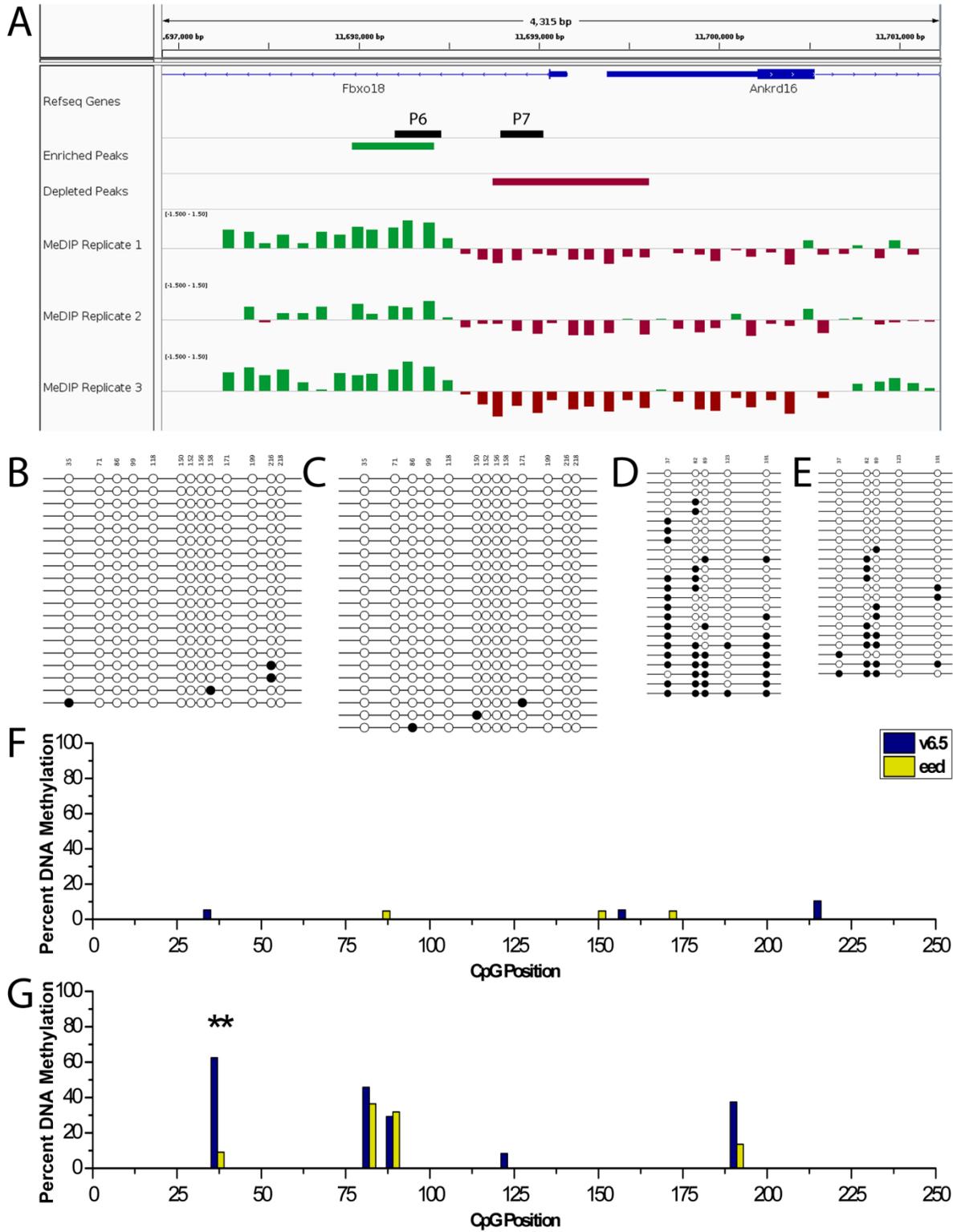


Figure 4. Validation of Depleted Methylation Peak at Gene Pair *Fbxo18/Ankrd16*. (A) The location of the pair is shown with data from the 3 MeDIP replicates. The calculated average peaks of enrichment and depletion are shown as green and red bars spanning the length of the peak, respectively. Black bars indicate where the validation PCR products P6 and P7 span in relation to the peaks (see Table 1). (B,C) The methylation profile of DNA at product P6 in V6.5 and *eed*^{-/-} ES cells, respectively. (D,E) The respective methylation profile of DNA at product P7 in V6.5 and *eed*^{-/-} ES cells. (F, G) A graphical comparison of methylation between V6.5 and *eed*^{-/-} ES cells at each CpG site in PCR products P6 and P7, respectively. ** indicates high significance at $p < .01$ (Fisher's exact test).

DISCUSSION

Using a genome-wide approach, we have confirmed that H3K27 trimethylation does indeed influence DNA methylation at numerous promoters in the genome of mouse ES cells. As discussed, mechanisms behind this interaction are not well understood. However, we found that DNAm is not widely antagonized by H3K27me₃. Instead, some promoters showed enrichment in the absence of an operational PRC2 complex, while others showed a notable depletion in its absence, and a small few showed both enrichment and depletion. While our recent discovery of hypomethylation leading to increased H3K27 trimethylation provides evidence that both marks mutually regulate each other (Hagarman, in preparation), a global trend of mutual antagonism at promoters of genes must be ruled out in mouse embryonic stem cell lines.

An alternative hypothesis postulated by another lab is that the PRC2 complex may cripple itself by binding to DNAm-poor sites and recruiting factors that induce *de-novo* methylation of nearby chromatin, reducing its ability to methylate H3K27. This model is based on findings that DNA methylation by itself can inhibit PRC2 complex binding to chromatin and that PRC2 may recruit DNA methyltransferases in postnatal neural stem cells (48), and can be supported by our findings that some H3K27me₃ depleted regions showed decreased DNA methylation. However, it is unlikely based on our lab's previous studies that this alternative hypothesis is a global trend either (37). This demonstrates that the interaction between DNA methylation and H3K27me₃ is likely to involve more than DNMT and PRC2 accessibility, and that a number of additional factors may dictate whether these marks coincide or are excluded from each other.

One possible factor leading to the variation in H3K27 trimethylation's effects on DNA methylation could be PRC2's interaction with *cis* DNA elements such as promoters and

enhancers. It is possible that only certain sequences allow PRC2 to recruit certain DNMTs. Thus those sequences that do allow placement of DNA methylation would show drops in DNA methylation in the mutant, and those sequences that prevent PRC2 from recruiting certain DNMTs might show increases in DNA methylation when the chromatin may be more accessible for other DNMTs. This sequence specificity may be assisted by one of several cofactors that the core proteins of PRC2 associate with (49). However, while our study looked at gene function of those affected loci, we did not characterize *cis* DNA elements at these regions, and so further investigation into are data is required.

Additionally, it should be noted that these findings only indicate interactions exhibited in mouse embryonic stem cells. Multiple studies have shown that after differentiation, chromatin marks influence each other differently. For example, the bivalent chromatin structure of coinciding activating and repressive marks described by Bernstein *et. al.* is lost as a cell develops, and those cells formerly held in the “poised” state lose either H3K27me or H3K4me after differentiating (14). These differences between stem cells and differentiated cells is further supported by our finding that those genes that we identified as having this interaction mostly concern development; other genes involved in later processes may yet to be actively regulated. Thus, our study only captures a snapshot of epigenetic regulation when cells may exhibit a cycle of chromatin states as it passes through development. Unfortunately, we cannot not identify changes amongst these mutants at later time points using similar methods; global inactivation of PRC2 allows cells to survive at the embryonic stage, but it is ultimately lethal to the fetus (50).

Despite this limited look at chromatin state at this time point, some reports have indicated that some cancer cells may display epigenetic interactions similar to stem cells, supporting the stem cell hypothesis of tumorigenesis (41). Indeed, a similar inverse correlation between

H3K27me3 and DNAm3 was found in colon cancer cell lines (40). If our findings are indeed stage-specific, then these interactions between H3K27me3 and DNA methylation could then serve as a model for how these marks are regulated in cancer cells as well as stem cells. This could then lead to finding means by which to inhibit these specific interactions, and the ultimate development of new anti-cancer drugs.

Data collection for validation of methylation peaks came with extreme difficulty. One pitfall of using bisulfite treatment is that the conversion of unmethylated cytosines to thymines creates numerous problems in designing primers used for PCR. After bisulfite treatment, all non-CpGs cytosines are converted to thymines, causing a loss in the complexity in the DNA. For a 20bp primer for example, the number of unique sequences available drops from numbers in the low trillions to numbers in the billions. Additionally, primers must also be absent of any CpG dinucleotides in their sequence, as differences in methylation would create differences in primer annealing and amplification (51). Additionally, the conversion of cytosines to thymines can also permit hybridization of previously dissimilar genes, creating complex annealing and secondary structures that can confound sequencing efforts. Indeed, in our attempt to validate a peak in several unreported genes, sequence data received back from clones had consistent drop-offs after less than a hundred base pairs. Finally, the bisulfite process can be incredibly destructive to DNA, with the harsh conditions capable of degrading DNA during the denaturing and conversion steps. At many times during data collection, DNA was unable to be recovered from gel purification because no DNA viable for a PCR reaction remained after conversion. While the time over which these caustic steps are carried out can be limited to increase DNA yield, for larger PCR fragments this results in a large number of unconverted cytosines that render the sequence useless to analyze (51). As a result, neither I nor anyone else in the laboratory could

design primers that generated a product larger than 500bp. Consequently, combined with the reduced number of site-specific PCR primers available, we were unable to survey all sections of any given peak, since only few regions had sequences that matched these specifications. Despite these limitations however, I and my advisor were able to generate enough data to adequately validate the MeDIP-chip data.

In summary, I found evidence that DNA methylation increases, decreases, or both at a number of gene promoters in mouse embryonic stem cell in the absence of an active Polycomb Repressive Complex 2, exhibiting a genome-wide interaction between DNA methylation and H3K27 trimethylation. Further work is being performed in our laboratory both to examine the levels of H3K27 trimethylation between wild type and DNMT triple knockout stem cells, and to identify transcript abundance of the affected genes amongst all three mutants – wild type, *eed*^{-/-}, and the DNMT mutant. We hope that these findings further elucidate the interaction between these two marks, and that further studies will discover the mechanism behind this interaction, whether direct or indirect.

ACKNOWLEDGEMENTS

I would like to thank Dr. Paul Soloway for the opportunity to work in his laboratory throughout this past year, and for all of his guidance and support in carrying out this project. I would also like to thank Dr. James Hagarman for training me in the procedures necessary for this project and helping me structure the experiments. I should also thank James Putnam, without whom I would be tremendously lost trying to find reagents within the laboratory for my experiments. Finally, I would like to thank Dr. Jonathan Flax, Patrick Murphy, David Taylor, Leah Hellerstein, and the rest of the Soloway group for keeping me sane during the countless hours I spent in the lab trying to complete this project.

REFERENCES

1. A. Bird, "DNA methylation patterns and epigenetic memory," *Genes Dev.* **16**, 6 (2002).
2. T. HEBBES, A. THORNE, C. CRANEROBINSON, "A Direct Link between Core Histone Acetylation and Transcriptionally Active Chromatin," *EMBO J.* **7**, 1395 (1988).
3. Y. Zhang, D. Reinberg, "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails," *Genes Dev.* **15**, 2343 (2001).
4. M. BRAUNSTEIN, A. ROSE, S. HOLMES, C. ALLIS, J. BROACH, "Transcriptional Silencing in Yeast is Associated with Reduced Nucleosome Acetylation," *Genes Dev.* **7**, 592 (1993).
5. M. Grunstein, "Histone acetylation in chromatin structure and transcription," *Nature* **389**, 349 (1997).
6. B. D. Strahl, R. Ohba, R. G. Cook, C. D. Allis, "Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena," *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14967 (1999).
7. H. Santos-Rosa *et al.*, "Active genes are tri-methylated at K4 of histone H3," *Nature* **419**, 407 (2002).
8. M. Lachner, N. O'Carroll, S. Rea, K. Mechtler, T. Jenuwein, "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins," *Nature* **410**, 116 (2001).
9. S. Rea *et al.*, "Regulation of chromatin structure by site-specific histone H3 methyltransferases," *Nature* **406**, 593 (2000).
10. R. Cao *et al.*, "Role of histone H3 lysine 27 methylation in polycomb-group silencing," *Science* **298**, 1039 (2002).
11. B. Czermin *et al.*, "Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal polycomb sites," *Cell* **111**, 185 (2002).
12. A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, D. Reinberg, "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein," *Genes Dev.* **16**, 2893 (2002).
13. J. Muller *et al.*, "Histone methyltransferase activity of a Drosophila polycomb group repressor complex," *Cell* **111**, 197 (2002).
14. B. Bernstein *et al.*, "A bivalent chromatin structure marks key developmental genes in embryonic stem cells RID A-1306-2007," *Cell* **125**, 315 (2006).

15. A. BIRD, "Cpg-Rich Islands and the Function of Dna Methylation," *Nature* **321**, 209 (1986).
16. P. A. Jones, D. Takai, "The role of DNA methylation in mammalian epigenetics," *Science* **293**, 1068 (2001).
17. M. MONK, M. BOUBELIK, S. LEHNERT, "Temporal and Regional Changes in Dna Methylation in the Embryonic, Extraembryonic and Germ-Cell Lineages during Mouse Embryo Development," *Development* **99**, 371 (1987).
18. T. Kafri *et al.*, "Developmental Pattern of Gene-Specific Dna Methylation in the Mouse Embryo and Germ Line," *Genes Dev.* **6**, 705 (1992).
19. M. BRANDEIS *et al.*, "Sp1 Elements Protect a Cpg Island from De-Novo Methylation," *Nature* **371**, 435 (1994).
20. M. Okano, D. W. Bell, D. A. Haber, E. Li, "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development," *Cell* **99**, 247 (1999).
21. T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, "Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases," *J. Mol. Biol.* **203**, 971 (1988).
22. T. H. Bestor, "Activation of Mammalian Dna Methyltransferase by Cleavage of a Zn Binding Regulatory Domain," *EMBO J.* **11**, 2611 (1992).
23. H. LEONHARDT, A. PAGE, H. WEIER, T. BESTOR, "A Targeting Sequence Directs Dna Methyltransferase to Sites of Dna-Replication in Mammalian Nuclei," *Cell* **71**, 865 (1992).
24. M. Szyf, "DNA methylation and demethylation as targets for anticancer therapy," *Biochemistry-Moscow* **70**, 533 (2005).
25. H. Cedar, Y. Bergman, "Linking DNA methylation and histone modification: patterns and paradigms," *Nat. Rev. Genet.* **10**, 295 (2009).
26. V. Azuara *et al.*, "Chromatin signatures of pluripotent cell lines RID A-4491-2010," *Nat. Cell Biol.* **8**, 532 (2006).
27. T. Hashimshony, J. Zhang, I. Keshet, M. Bustin, H. Cedar, "The role of DNA methylation in setting up chromatin structure during development," *Nat. Genet.* **34**, 187 (2003).
28. X. Nan *et al.*, "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex," *Nature* **393**, 386 (1998).
29. L. Lande-Diner *et al.*, "Role of DNA methylation in stable gene repression," *J. Biol. Chem.* **282**, 12194 (2007).

30. E. Vire *et al.*, "The Polycomb group protein EZH2 directly controls DNA methylation," *Nature* **439**, 871 (2006).
31. J. Jackson, A. Lindroth, X. Cao, S. Jacobsen, "Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase," *Nature* **416**, 556 (2002).
32. H. Tamaru, E. Selker, "A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*," *Nature* **414**, 277 (2001).
33. B. Lehnertz *et al.*, "Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin," *Curr. Biol.* **13**, 1192 (2003).
34. S. K. T. Ooi *et al.*, "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA," *Nature* **448**, 714 (2007).
35. D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, "Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation," *Nature* **449**, 248 (2007).
36. D. Zilberman, D. Coleman-Derr, T. Ballinger, S. Henikoff, "Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks," *Nature* **456**, 125 (2008).
37. A. M. Lindroth *et al.*, "Antagonism between DNA and H3K27 Methylation at the Imprinted Rasgrf1 Locus," *PLoS Genet.* **4**, e1000145 (2008).
38. T. Bartke *et al.*, "Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation," *Cell* **143**, 470 (2010).
39. Y. Schlesinger *et al.*, "Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer," *Nat. Genet.* **39**, 232 (2007).
40. E. N. Gal-Yam *et al.*, "Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line," *Proc. Natl. Acad. Sci. U. S. A.* **105**, 12979 (2008).
41. M. Widschwendter *et al.*, "Epigenetic stem cell signature in cancer," *Nat. Genet.* **39**, 157 (2007).
42. J. E. Ohm *et al.*, "A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing," *Nat. Genet.* **39**, 237 (2007).
43. L. Boyer *et al.*, "Polycomb complexes repress developmental regulators in murine embryonic stem cells," *Nature* **441**, 349 (2006).
44. N. D. Montgomery *et al.*, "The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation," *Current Biology* **15**, 942 (2005).

45. M. FROMMER *et al.*, "A Genomic Sequencing Protocol that Yields a Positive Display of 5-Methylcytosine Residues in Individual Dna Strands," *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1827 (1992).
46. A. Meissner *et al.*, "Genome-scale DNA methylation maps of pluripotent and differentiated cells," *Nature* **454**, 766 (2008).
47. F. Mohn, M. Weber, D. Schuebeler, T. Roloff, "Methylated DNA Immunoprecipitation (MeDIP)," *Methods in Molecular Biology* **507**, 55 (2009).
48. H. Wu *et al.*, "Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes," *Science* **329**, 444 (2010).
49. R. Margueron, D. Reinberg, "The Polycomb complex PRC2 and its mark in life," *Nature* **469**, 343 (2011).
50. A. He *et al.*, "Polycomb Repressive Complex 2 Regulates Normal Development of the Mouse Heart," *Circ. Res.* **110**, 406 (2012).
51. L. C. Li, R. Dahiya, "MethPrimer: designing primers for methylation PCRs," *Bioinformatics* **18**, 1427 (2002).