MECHANISMS OF EPIGENETIC REGULATION CHARACTERIZED USING

NOVEL SINGLE MOLECULE AND TRADITIONAL METHODS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Patrick James Murphy

January 2013

Mechanisms of Epigenetic Regulation Characterized Using Novel Single

Molecule and Traditional Methods

Doctoral Dissertation

Patrick James Murphy, Ph.D.

Cornell University 2013

Abstract

Epigenetic mechanisms, including histone modification and DNA methylation, are fundamental for controlling gene regulation. Not surprisingly, aberrant placement of modifications can cause defects in fundamental cellular processes such as proliferation, migration, and differentiation, and may contribute to tumorigenesis. Additionally, epigenetic state can be influenced by environmental factors that include diet, environmental toxins, and behavior. Therefore, it is critical to understand genome wide epigenetic mark placement. Today we have a deep understanding of what epigenetic modifications are, and where they are placed throughout the genome, but generally, even after 75 years of research, we only have a superficial understanding of what mechanisms determine which genes are chemically modified, and how these epigenetic modifications lead to an altered cell fate.

Here I will discuss experiments and techniques that have enabled us to characterize epigenetic mark regulation in two ways. First, I will present work characterizing the establishment of DNA methylation during mouse epigenetic

reprogramming using the imprinted *Rasgrf1* model locus. Our data indicate that a repeat region, located 3-prime the DMR (Differentially Methylated Region) functions as a promoter for a pitRNA (piRNA-targeted non-coding RNA), which is necessary for deposition of methylation, and functions only in *cis*. From these data I propose a model where deposition of methylation takes place in a co-transcriptional manner. Second, I will discuss the development of a single molecule nanofluidic based technology capable of detecting intact chromatin, assaying for epigenetic marks, purifying DNA based on methylation state, quantifying relative epigenetic mark abundance, and detecting multiple simultaneously occurring epigenetic marks. Single molecule based studies offer an attractive means for assaying chromatin molecules, because they allow for direct inspection of molecules, without ensemble averaging, and can be preformed using very small amounts of input material. This new technique called SCAN (Single Chromatin molecule Analysis in Nanochannels) has enable us to conclude that normal antagonism of H3K27me3 and DNA methylation breaks down during cellular immortalization. The findings presented here have not only helped to characterize epigenetic mark regulation, but will ultimately lead to new questions and innovative research projects.

Biographical Sketch

Patrick was born to Patrick and Karen Murphy on April 11th 1985 in Buffalo, NY. Over the next nine years his family grew from 4 to 7. In addition to his parents, Patrick's siblings Sean, Brian, Sarah, and James provided him with the love and support that would ultimately mold him into the person he has become today. From age 4 to 20 Patrick lived in an old brick house on Monroe Avenue in Angola, a suburb of Buffalo. During his early life in Angola he was an active member of numerous organizations including the Boy Scouts of troop 528.

He attended Lake Shore Central Schools where his favorite subjects included Science, Math, and Art. He takes the most pride in his accomplishments as Captain of the Varsity Track team, becoming an Eagle Scout, and as High School Student Body President. He credits an Elementary school ALPHA gifted and talent program for introducing him to abstract thought and problem solving. He credits the Boy Scouts of America for introducing him to the beauty of nature, and he credits his various art teachers for enabling him to embrace creativity. During the summer between his sophomore and junior year Patrick was invited to participate in the Howard Hughes Medical Institute for Young Scholars at Villanova University. There he was introduced to what would ultimately become his life's passion, basic biology research.

Patrick attended college at St John's University, where he continued to explore basic research, and joined the laboratory of Christopher Bazinet. There

he began as a volunteer, learning basic molecular biology and genetic techniques, and finished as a paid research technician. After completing his Bachelors of Science degree in only 3 years, Patrick moved on to the National Institutes of Health in Washington, DC. There he was a member of Elissa Lei's laboratory, where he researched chromatin nuclear structure in drosophila using biochemistry based methods. After 1 year at the National Institutes of Health, Patrick was accepted to graduate school at Cornell University in Ithaca NY. At Cornell Patrick became a member of Paul Soloway's lab, where he investigated mammalian epigenetic mechanisms. At Cornell, he met the love of his life Kristin Blauvelt, and after dating for 4 years, the two were married on July 6th of 2012. Following his time at Cornell, Patrick will be joining Brad Cairns' lab at the Huntsman Cancer Institute in Salt Lake City, where he will begin researching vertebrate early embryo epigenetic reprogramming.

Dedication

<u>To my father Patrick:</u>

Who at times worked multiple jobs in industry, sacrificed his health and his sanity, to

provide for his family, and to give them a better life than he ever had.

I thank you for showing me how working hard can lead to intense satisfaction.

I thank you for various life lessons: "Work with your head, not with your muscles."

"Measure twice, cut once." "If it's worth doing, it's worth doing right."

<u>To my mother Karen:</u>

Who went back to school for multiple degrees including a Masters in Education, all the

while raising 5 children, and working numerous part time jobs.

I thank you for teaching me to think in an abstract way, to always be creative, to have

pride in myself, and that determination will ultimately win over circumstance.

<u>To my wife Kristin:</u>

I thank you for always supporting me and loving me unconditionally.

I thank you for teaching me it's always best to think about things before reacting.

I thank you for teaching me that it's okay to be quiet sometimes.

You are my voice of reason, and I love you.

<u>To my siblings:</u>

Who are a direct reflection of my parents and always make me feel at home.

Thank you for keeping my life balanced and for being my best friends.

Acknowledgements

I would like to begin by thanking my advisor Paul Soloway. He has helped to mold me and challenge me as a scientist. When I first arrived at Cornell I would consider myself similar to a young colt. I was, enthusiastic, energetic, eager to run on the fast track, but above all I was unrefined. Paul has been the best trainer I could ask for over the past 5 years. He has taken the reins and taught me how to think critically, plan experiments carefully, and organize my thoughts clearly. More so, I thank Paul for encouraging me to be a creative and to be critical with my questions. Good mentorship is truly the backbone of science research, and Paul is a prime example. The impact of his mentorship will go far beyond his personal research achievements.

I would also like to thank my various advisors here at Cornell, including Harold Craighead, Nate Sutter, Eric Richards, John Schimenti, Eric Alani, and Scott Coonrod. Their support and guidance have helped to keep me on track as both a student, and as a young scientist.

I have been fortunate enough to have a number of superb labmates over the years. Id like to begin by thanking Chelsea Brideau and Nadia Drake, who are more friends than labmates. From the moment I joined Paul's lab, they were prime examples of what it takes to succeed, and how to balance life with science as a graduate student. I thank Ruqian Zhao for her support and mentorship. I would never have been successful with my thesis project had Ruqian not been

Table of Contents

List of Figures

List of Abbreviations and Symbols

♂   -   Male

♀   -   Female

α   -   anti (antibody against)

5azaC   -   5-aza-2'-deoxycytidine

5hmC   -   5-hydroxymethyl-cytosine

5mC   -   5-methyl-cytosine

APD   -   avalanche photodiode

B6   -   C57BL/6 strain

bp   -   base pair

BPA   -   bisphenol A

BS-Seq   -   bisulfite sequencing

cDNA   -   complimentary DNA

ChIP   -   chromatin immuno-precipitation

ChIP-chip -   chromatin immuno-precipitation combined with microarray

ChIP-Seq -   chromatin immuno-precipitation combined with sequencing

DES   -   diethylstilbestrol

DMD   -   differentially methylated domain

DMEM   -   Dulbecco's modified Eagle's medium

DMR   -   differentially methylated region

DNMT   -   DNA methyltransferases

Enh	-	enhancer
ESC	-	embryonic stem cells
FCS	-	fluorescence correlation spectrometry
FRET	-	Föster's Resonance Energy Transfer
FVB	-	FVB/n strain
GFP	-	green fluorescent protein
H2AZ	-	histone 2AZ
H2B	-	histone 2B
H3	-	histone 3
H3K27	-	lysine 27 of H3
H3K27me	-	methylation on lysine 27 of H3
H3K27me3	-	trimethylation on lysine 27 of H3
H3K36me	-	methylation on lysine 36 of H3
H3K4	-	lysine 4 of H3
H3K4me	-	methylation on lysine 4 of H3
H3K4me3	-	trimethylation on lysine 4 of H3
H3K9me	-	methylation on lysine 9 of H3
H3K9me3	-	trimethylation on lysine 9 of H3
H4	-	histone 4
H4K16ac	-	acetylation on lysine 16 of H4
H4K20me	-	methylation on lysine 20 of H4

HAT	-	histone acetyltransferase
HDAC	-	histone deacetylase
HDM	-	histone demethylase
HMT	-	histone methyltransferase
ICR	-	imprinting control region
IP	-	immuno-precipitations
LGABN	-	licking, grooming and arched-back nursing
LIF	-	leukemia inhibitory factor
MBD	-	methylated DNA binding domain containing protein
MF	-	mouse fibroblasts
Mnase	-	micrococcal nuclease
ncRNA	-	non-protein coding RNA
piRNA	-	PIWI interacting RNA
pitRNA	-	piRNA-targeted non-coding RNA
Pro	-	promoter
R2	-	*Igf2r* region 2
RD	-	repeat deletion mutant
RPMI		Roswell Park Memorial Institute medium
RT	-	room temperature
RT-PCR	-	reverse transcription PCR
SCAN	-	single chromatin molecule analysis in nanochannels

SMD	-	Single Molecule Detection
SNP	-	single nucleotide polymorphism
TBS	-	tris-buffered saline
TCA	-	tricarboxylic acid
TF	-	transcription factor
TKO	-	triple knockout
TP1	-	*TIMP1*
Wt	-	wild type
XCI	-	X-chromosome inactivation

I. Introduction

I.1. A brief history of epigenetics

I.1.1. The problem of development – part 1

In the 1930s the field of organismal development and morphogenesis, which at the time was referred to as "epigenesis", struggled with a particularly complex problem. Although basic laws of Mendelian inheritance (Mendel, 1865) were widely accepted, few understood how they applied to epigenesis. A prominent developmental biologist at the time, Conrad Waddington wondered how each cell, which contained identical genetic material, could transition into a wide variety of phenotypes during development and cellular differentiation. It seemed that genetics alone could not explain this process. Ultimately to resolve this dilemma, Waddington proposed an idea that would later unify the fields of epigenesis and genetics. Waddington used the term "epigenetics" to describe the combined effects and casual interaction of genes with their products as they contribute to a developmental phenotype (Waddington, 1939). Mechanisms controlling epigenetic phenomena remained a mystery however, and it wasn't until nearly 15 years later that scientists began using epigenetics to solve the problem of development.

I.1.2. The heritable subunit

In the mid 1940s the DNA molecule was determined to be the heritable subunit of the nucleus (Avery et al., 1944) and encode the instructions for building an organism. Subsequently, Francis Crick, who helped to discover the structure of the DNA double helix (Watson and Crick, 1953), proposed the central dogma of molecular biology (Crick, 1956, 1958), which explains how genotype leads to phenotype, and describes a two step process where DNA passes information to RNA, which then transfers that information to a peptide during protein synthesis (**Fig I.1**). Crick extended this linear model to include the idea that proteins NEVER transfer information to DNA or to RNA and histones have no genetic information encoded within them. Although on the surface, this idea was accurate, further inspection revealed exceptions, and again developmental differentiation presented a problem.

I.1.3. The problem of development – part 2

If each cell contains identical DNA sequence, and proteins can transfer no additional information to DNA, how is it that a zygote can develop along a variety of complex lineages, and potentially be converted into any cell necessary to build a higher organism? Presumably, something other than DNA must be mitotically heritable, and that something must be changing during differentiation in order to generate such a broad phenotypic array. Between 1969 and 1975, a number of scientists began to realize that an enzyme protein [later defined as DNA

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have
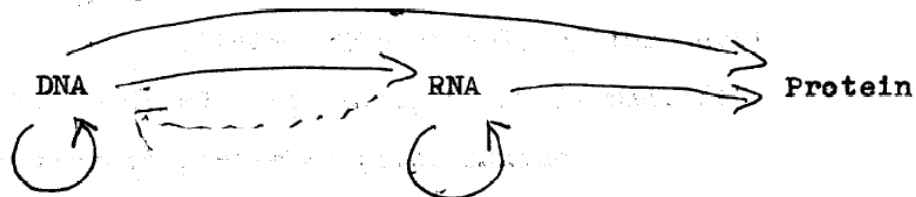
but never
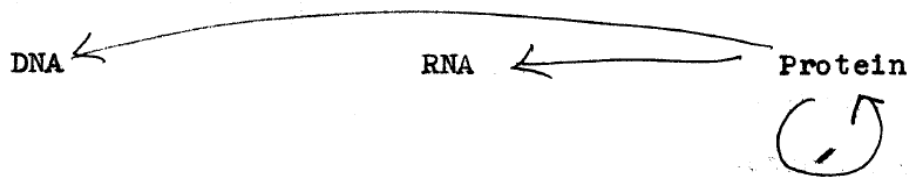
where the arrows show the transfer of information.

Requirements for protein synthesis.

Fig I.1 - James Crick's thoughts on protein synthesis
James Crick diagrams the passage of information from DNA to RNA, and eventually to a protein during peptide synthesis. This is used to describe how DNA carries the heritable genetic information for creating an organism. In his diagram he specifically illustrates how proteins can never pass information to DNA or to RNA. This passage of information was later accounted for using epigenetic phenomena, disproving Crick's model.

methyltransferase (Yen et al., 1992) could place methyl groups on specific DNA sequences, and provide for a non-genetic means of inheritance (Griffith and Mahler, 1969; Riggs, 1975). In this sense, methyl groups can carry information from proteins to DNA. The genomic location of these methyl groups is not only mitotically heritable (Scarano, 1971), but can potentially change in a cell type specific manner during developmental differentiation (Holliday and Pugh, 1975). This discovery not only directly contradicted one principle from Crick's central dogma, namely that proteins do not transfer information to DNA, but also provided a possible solution to the 30 year old problem first proposed by Waddington in the 1930s. Thus, many scientists studying the molecular mechanisms controlling non-genetic inheritance began to adopt Waddington's term use it when referring to these phenomena as "epigenetics."

I.1.4. The epigenetic landscape

With a new found appreciation for Waddington's ideas, the field of epigenetics began to blossom. One idea that gained particular momentum was Waddington's metaphorical "epigenetic landscape." (**Fig I.2**) In this metaphor Waddington uses a rolling ball to symbolize a genotype and a downward sloping hill, upon which the ball rests, to symbolize epigenetics. As the ball travels down the hill it encounters a number of paths, each one bringing the ball in a variety of different directions and leading to a variety of different outcomes. The idea is that a genotype encounters a number of different signals as it progresses through

4
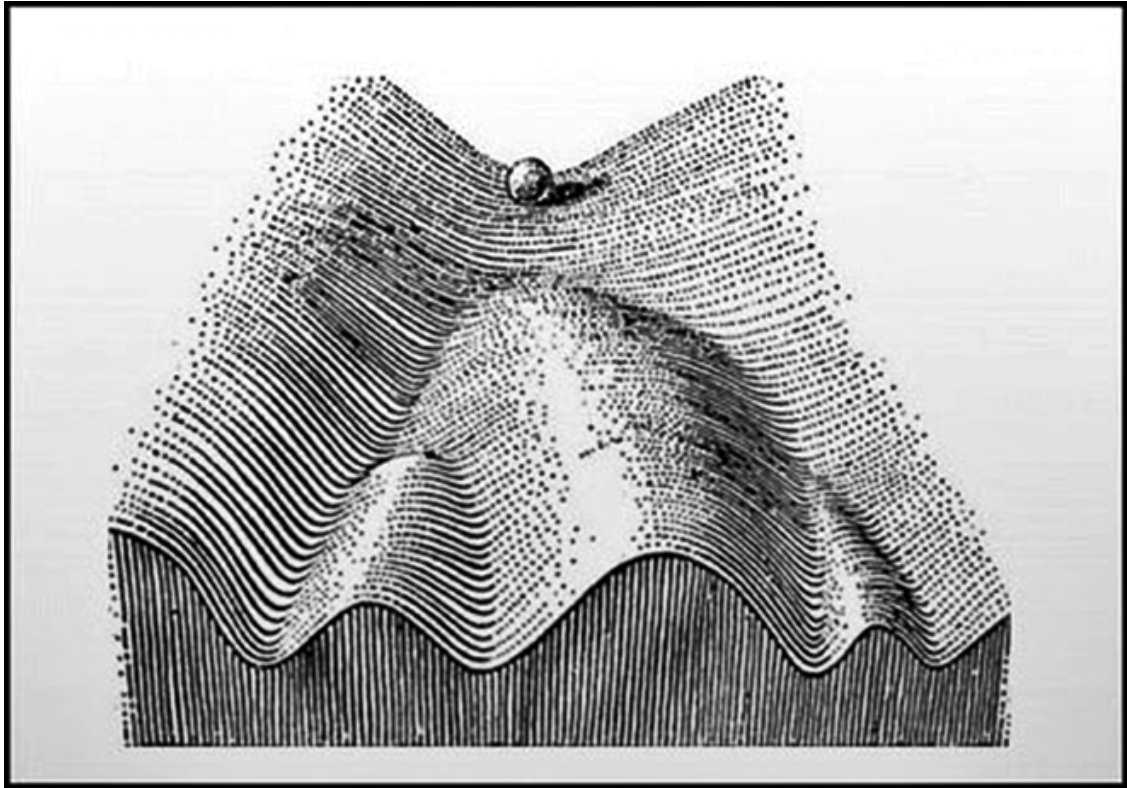
Fig I.2 – Conrad Waddington's Epigenetic Landscape
Waddington uses this illustration as a metaphor to describe epigenetics. The ball begins at the top of the epigenetic landscape and begins to roll down the hill. As it rolls, it encounters a number of hills and valleys that form trails and paths. Each path leads the ball in a different direction and will ultimate bring it one of a variety of resting places. The round ball at the top is used to represent either a single gene or a cell as it changes during differentiation. As it progresses through differentiation it moves across the epigenetic landscape and is guided to its eventual phenotypic expression pattern.

development. The path each gene takes, as it reacts to these signals, will confer ultimate phenotypic expression. In other words, those who study epigenetics are not studying genes *per se*, they are studying paths and trails leading to phenotypic outcomes. Instead, if we imagine the ball that rolls down the hill as if it were a cell, the epigenetic landscape would then represent all possible paths the cell can take leading to differentiation and lineage commitment. Epigenetics is therefore the study of mitotically and meiotically heritable mechanisms, specifically chemical modifications to DNA and chromatin structure, which control gene expression, cellular differentiation, and lineage commitment.

In general those who founded the field of epigenetics were motivated by one fundamental question: "What controls where the ball rolls as it moves down the hill?" In the same way, many modern day molecular epigeneticists ask the following question: "What are the mechanisms that determine which genes are chemically modified, and how do those modifications alter cell fate?"

I.2. Chemical modifications to the chromatin fiber

I.2.1. Epigenetic phenomena and mechanisms in humans

In mice and in humans, females possess two copies of the X-chromosome, while males possess only a single copy of each the X-chromosome and the Y-chromosome (Takagi and Sasaki, 1975). Although this

imbalance of X-chromosomes is required for proper sex determination (Sinclair et al., 1990), if all genes on the X-chromosome were transcribed as autosomes are, it would create a significant imbalance of both transcript and protein levels between males and females. To account for this, and create a balance between males and females, mammals have evolved a mechanism to silence expression of one female X-chromosome (Barakat and Gribnau, 2012). During mouse embryonic development, shortly after implantation a process referred to as X-chromosome inactivation (XCI) is initiated, and with respect to the parental origin, silencing is random (Lyon, 1961). A number of epigenetic silencing mechanisms, including non-protein coding RNAs (ncRNA) (Wutz et al., 1997), post-translational histone protein modifications (Zhao et al., 2008), and DNA methylation (Nesterova et al., 2008), are essential in this process. Failure to maintain human X-chromosome expression levels and regulate XCI has been linked to a number of disease phenotypes including hemophilia, manic depression, and bipolar disorder (Dorus, 1983; Valleix et al., 2002; Dobyns et al., 2004).

Aside from XCI, ncRNAs, histone modifications, and DNA methylation are important for other non-genetic phenomena. For example, at a number of loci throughout the genome only a single copy of a given allele is transcribed. At these mono-allelically expressed genes the parent of origin pre-determines which allele is expressed, and which is silenced (Cattanach and Kirk, 1985; Surani et

al., 1987). This parental allele specific gene expression pattern, referred to as genomic imprinting, is not only heritable, but is controlled by allele specific epigenetic marks, which will be discussed in the next section. In humans, mutations that lead to improper regulation of imprinted gene expression cause a number of developmental abnormalities which manifest as diseases in adults, including Beckwith-Weideman, Prader-Willi, and Angelman syndromes (Robertson, 2005). Because epigenetic marks control numerous important processes, including imprinting and X-chromosome inactivation, research has focused on understanding how epigenetic modifications are regulated, and what determines where epigenetic marks are placed throughout the genome. Various studies have used imprinted loci to study epigenetic mechanisms, including ncRNAs, histone modifications, and DNA methylation.

I.2.2. Modification to DNA

Within the nucleus DNA wraps itself around a protein octamer structure called the nucleosome. This higher order structure of DNA and its associated proteins is often referred to as the "chromatin fiber". A variety of different chemical modifications can be made to a host of substructures contained within the chromatin fiber. These chemical modifications may have both direct and indirect effects on the expression of genes residing within the underlying DNA sequence. The simplest of these modifications is DNA methylation, and in vertebrates occurs almost exclusively in the form of methyl groups on the fifth

carbon of the cytosine located 5' to a guanine nucleotide. The product, 5-methyl-

ctyosine (5mC) is dependent on the universal methyl donor, S-adenosyl

methionine, and on the catalytic activity of one of three DNA methyltransferases

(DNMT). DNMT3a and DNMT3b are the only known *de novo*

methyltransferases, and are responsible for placing methylation on previously

unmodified cytosines. DNMT1 is referred to as the maintenance

methyltransferase because its role is to place methyl groups on cytosines that

are the product of recent replication and have become hemi-methylated.

I.2.3. The role of methylation

The ability of 5mC to directly inhibit gene transcription is well established,

with both *in vitro* and in vivo studies having found that gene transcription and

subsequent protein synthesis is inhibited when 5mC is present in promoter

regions (Watt and Molloy, 1988; Götz et al., 1990). Additionally, DNA

methylation can function indirectly; when present at unique binding sites 5mC

has been shown to both promote and inhibit transcription factor (TF) binding. For

example, proteins that contain a methylated DNA binding domain (MBD) require

the presence of 5mC for effective binding (Boyes and Bird, 1991), while others,

like the insulator binding protein CTCF can be repelled by 5mC (Bell and

Felsenfeld, 2000). Early research often characterized the effect of 5mC either on

individual loci, or by using *in vitro* systems, making it impossible to address more

general genome wide concepts. Recently, chromatin immuno-precipitation (ChIP)

(Jackson, 1978) has been combined with genome wide analysis platforms, like

ChIP-chip, where microarray based technologies are taken advantage of (Ren et

al., 2000), and ChIP-Seq, where whole genome sequencing platforms are utilized

(Barski et al., 2007; Johnson et al., 2007). These new technologies, which allow

for genome wide mapping of protein DNA interactions, make it possible to

analyze wide ranging genomic regulatory principles. Notably, recent whole

genome studies have confirmed that 5mC regulates binding of TFs and

demonstrated that this process is fundamental for genome wide transcriptional

regulation (Hogart et al., 2012; Wang et al., 2012). Of note however, these

whole genome studies rely on correlations of 5mC with TF binding to make broad

and general conclusions. They do not directly observe the presence or absence

of methylation and TF co-occupancy. Proper placement of 5mC is critical for

proper genome wide epigenetic/chromatin organization and transcriptional

regulation.

I.2.4. Histone modifications

Modifications can also be made to other substructures contained within

the chromatin fiber. The structure and location of the nucleosome, relative to the

DNA, can be critical for determining the function of the underlying DNA

sequence. For example, the repositioning of nucleosomes over gene promoters

has been shown to actively regulate transcription (Hirschhorn et al., 1992;

Boeger et al., 2003), and altering nucleosome structure by substituting entire

core histone proteins for histone variants, like Histone 2AZ (H2AZ), has been shown to regulate establishment of genomic expression domains (Meneghini et al., 2003). Interestingly, in *Arabidopsis* these same complexes responsible for remodeling and reorganizing nucleosomes affect genome wide DNA methylation levels (Jeddeloh et al., 1999), indicating that there is coordination of various epigenetic phenomena.

The amino acids that make up these histone proteins can also be modified, ultimately leading to changes in underlying DNA function (Allis et al., 1985). Although there are numerous important post-translational histone modifications (Tan et al., 2011), the majority of research has focused on methyl and acetyl group additions to lysine residues placed on the N-terminal tails of histone 3 (H3) and histone 4 (H4). Modification to lysines on N-terminal H4 tails typically coincides with changes to the overall structure of the chromatin fiber. For example, acetylation on lysine 16 of H4 (H4K16ac) controls chromatin compaction and formation of higher order structure (Shogren-Knaak et al., 2006), while methylation on H4 lysine 20 (H4K20me) controls recruitment of DNA damage repair enzymes (Sanders et al., 2004). Modification to lysines on N-terminal H3 tails typically coincides with changes to gene expression levels. H3 lysine 9 methylation (H3K9me), for example, is associated with gene repression and the formation of dense highly stable compact structures called heterochromatin (Lachner et al., 2001). H3 lysine 36 methylation (H3K36me), on

the other hand, associates with gene expression, polymerase elongation, and a less dense structure called euchromatin (Strahl et al., 2002). At genes important for early development, which require precisely regulated temporal expression, two epigenetic marks, one repressive (H3 lysine 27 trimethylation - H3K27me3), and one activating (H3 lysine 4 trimethylation - H3K4me3), are found together (Bernstein et al., 2006). The co-occupant nature of these marks is thought to prepare a gene for dynamic changes in expression patterns. Similar to the "DNA code", the term "histone code" is often used to describe how these chemical modifications are functionally "coding" for specific molecular mechanisms (Strahl and Allis, 2000; Jenuwein and Allis, 2001).

I.2.5. Epigenetic writers, readers, and erasers

Earlier I discussed the enzymes responsible for adding methyl groups to DNA, however I have not yet addressed the enzymes responsible for modifying histones. As mentioned earlier, histones can be modified in a variety of ways; for the purposes of this dissertation I will only discuss methylation, and acetylation. Often the enzymes responsible for placing a specific epigenetic modification are referred to as the "writers" (Arrowsmith et al., 2012). Histone acetyltransferases (HATs) and histone methyltransferases (HMTs) are two forms of epigenetic writers. HAT enzymes were originally discovered in Tetrahymena and in yeast (Brownell et al., 1996; Kuo et al., 1998), and were shown mechanistically to be important for gene activation. Since then, 18 different HATs have been

discovered, with their function ranging from DNA damage repair to cell cycle progression (Arrowsmith et al., 2012). The first discovered HMTs on the other hand were shown to include both repressors, like Suv39h1 (Aagaard et al., 1999; Rea et al., 2000), and activators, like Set7 (Wang et al., 2001), depending on which lysine was targeted. To date 60 protein methyltransferase enzymes have been identified, and a SET domain is a common feature among them (Arrowsmith et al., 2012).

Factors that bind to specific histone modifications are often referred to as epigenetic "readers". Readers are a large and diverse group of proteins, where the function of a given protein is often dictated by the domain structure contained with in it. For example, HP1, which contains a Chromo domain, specifically binds to H3K9me3, and is required for spreading of repressive heterochromatin and DNA compaction (Lachner et al., 2001). Readers can be subdivided based on the domain they contain. These domains include Bromo, Tudor, MBT, Chromo, PWWP and PHD (Arrowsmith et al., 2012). Others readers, like EED, bind different histone marks, like H3K27me3, and facilitate spreading of a more flexible and reversible repressive chromatin state (Margueron et al., 2009).

In order for chromatin state to be reversed, often histone modifications need to be removed so that other different modifications can be placed. Enzymes responsible for removal of epigenetic marks are often called "erasers". Generally

proteins responsible for removal of methyl groups are called histone demethylases (HDMS) and when acting upon lysine residues, often fall within the Jumonji subfamily. For examples, Jumonji protein JMJD3 removes methylation on H3K27 (Agger et al., 2007), and relieves repression, while LSD1, not a Jumonji protein, removes methylation from H3K4 and promotes gene silencing (Shi et al., 2004). Enyzmes responsible for removing acetyl groups are called histone deacetylases (HDACs). Typically these enzymes are capable of removing acetyl groups from a number of animo acids on the same histone tail, and function in gene transcriptional repression (Rundlett et al., 1996; Laherty et al., 1997).

Earlier I discussed DNA methylation writers (DNMTs) and readers (MBD containing proteins). Similar to histone mark marks, DNA methylation can also be removed, and can occur in different ways. Removal of methylation from DNA was first discovered in zebra fish, where 5mC deaminase AID converts 5mC to thymine and MBD4 facilitates the incorporation of fully unmethylated cytosine by mismatch repair (Rai et al., 2008). Alternatively, during mouse embryonic epigenetic reprogramming the Tet proteins are responsible for removal of 5mC by converting it first to 5-hydroxymethyl-cytosine (5hmC) (Tahiliani et al., 2009). Interestingly, this newly discovered epigenetic modification (5hmC) has been shown to be fundamental for maintenance of stem cell plouripotentcy (Ito et al., 2010) and for proper chromatinization in the nuclei of purkinje neurons

14

(Kriaucionis and Heintz, 2009). This indicates that 5hmC has function significance outside of its role as a 5mC demethylation intermediate.

Although much is known about these epigenetic modifications, including where they reside, how they change during developmental differentiation, and which enzymatic factors are responsible for their placement, we in the field of epigenetics still cannot definitively answer one simple question: **"What determines where in the genome chemical modifications are placed?"**

I.3. Epigenetics and disease

I.3.1. Environmental effects

Waddington's original research focused on epigenetic inheritance as a byproduct of environmental manipulation (Waddington, 1952). Furthermore, epigenetic disruption can occur as a result of environmental stimuli even with no evidence for genetic mutation. For example, when female rats maintained a high level of care for their young, which was characterized behaviorally as increased licking, grooming and arched-back nursing (LGABN), there was a significant decrease in stress responses of offspring. Molecularly, the altered stress response was attributed to decreased levels of 5mC over the glucocorticoid receptor gene promoter in the hippocampus of mice that received increased care (Weaver et al., 2004). Often these mutations are referred to as "epimutations",

because direct sequence mutation cannot be attributed to the phenotype.

Maternal diet can also induce epigenetic changes in offspring. For example, during pregnancy, female mice who consume a diet high in methyl donors gave birth to pups with hypermethylated DNA. At the yellow agouti viable (*A(vy)*) locus, where a transposable element is inserted, the hypermethylation presents itself as an epimutation in the form of a coat color change, which persists into adulthood and is transiently heritable even when progeny are fed a control diet (low methyl donor) (Wolff et al., 1998; Waterland and Jirtle, 2003). A common component of household plastics called BisPhenol A (BPA) has been shown to have a similar effect on the *A(vy)* locus (Dolinoy et al., 2007). Mice exposed to moderate BPA levels (similar to those found in the environment) while *in utero* or as neonates experienced higher overall body weight, increased breast and prostate cancer, and decreased fertility as adults. Not surprisingly, these mice had decreased levels of 5mC at the *A(vy)* locus, suggesting that the more severe phenotypic effects are the byproduct of the environmentally altered epigenetic state. These and other studies (Weinhouse et al., 2011; Doshi et al., 2012) suggest that a chemical compound commonly found in the environment can influence the epigenetic state in humans, and potentially lead to heritable increases in cancer susceptibility, obesity, and infertility.

I.3.2. Epigenetics and cancer

Epigenetic modifications were first linked to cancer nearly 30 years ago (Feinberg and Vogelstein, 1983a) while studying DNA methylation in human cancer tissue. In these studies, global hypomethylation, combined with local hypermethylation was associated with colorectal carcinogenesis. Supporting this claim, the tumor suppressor *c-Ha-ras* gene was found to be hypomethylated in six of eight human carcinomas when compared to normal tissue (Feinberg and Vogelstein, 1983b). Emphasis on epigenetic mechanisms relating to cancer has grown recently, and a number of epigenetic biomarkers have been identified for various cancer types. Depending on genomic location, DNA can either be hyper- or hypomethylated relative to normal cells in both a tissue specific and cancer specific manner (Doi et al., 2009; Lendvai et al., 2012). Histone modification can also play a role in cancer. Notably loss of H4K16ac and H4K20me (both of which were mentioned previously to help maintain chromatin integrity) accompanied by global DNA hypomethylation, has been found to occur in leukemias (Fraga et al., 2005). Mutations in an important H3K4 methyltransferase MLL1 have been linked to a variety of acute leukemias in humans (Yokoyama et al., 2002; Dou et al., 2006). Whether modifications are at the root of oncogenesis, or are a byproduct of carcinogenic progression remains unknown (Martin et al., 2011). Regardless of origin, studying the mechanisms by which epigenetic modifications operate, both as a function of environment and as they relate to disease state, will

ultimately prove valuable for a variety of related fields, including developmental biology, epidemiology, and oncology.

I.4. Regulation of imprinted DNA methylation

I.4.1. A model system

Often to gain insight into complex epigenetic mechanisms, direct genetic mutations are made, and the epigenetic effects are monitored. Because these studies often require a number of manipulation steps, researchers often prefer to study model genes where epigenetic patterns are reliable, consistent, and highly reproducible. Imprinted loci satisfy all these requirements, and are thus used as model systems. As discussed earlier, genomic imprinting occurs when gene expression patterns are determined by the parent of origin. For example, at the imprinted mouse *Igf2* locus (**Fig I.3**) (Ferguson-Smith et al., 1991) the paternally inherited copy of *Igf2* is expressed and the *H19* gene is repressed (Bartolomei et al., 1991). Alternatively, when inherited from the mother, *H19* is expressed and *Igf2* is silenced (Kalscheuer et al., 1993). Shortly after *Igf2* imprinting was discovered, imprinted epigenetic modifications, like 5mC, were determined to be the cause of these unique expression patterns (Li et al., 1993). A number of other genomic regions, including the mouse *Rasgrf1* locus (Plass et al., 1996), have been discovered to have imprinted expression patterns, and at a number of loci epigenetic (Ferguson-Smith et al., 1993; Li et al., 1993; Wutz et al., 1997)
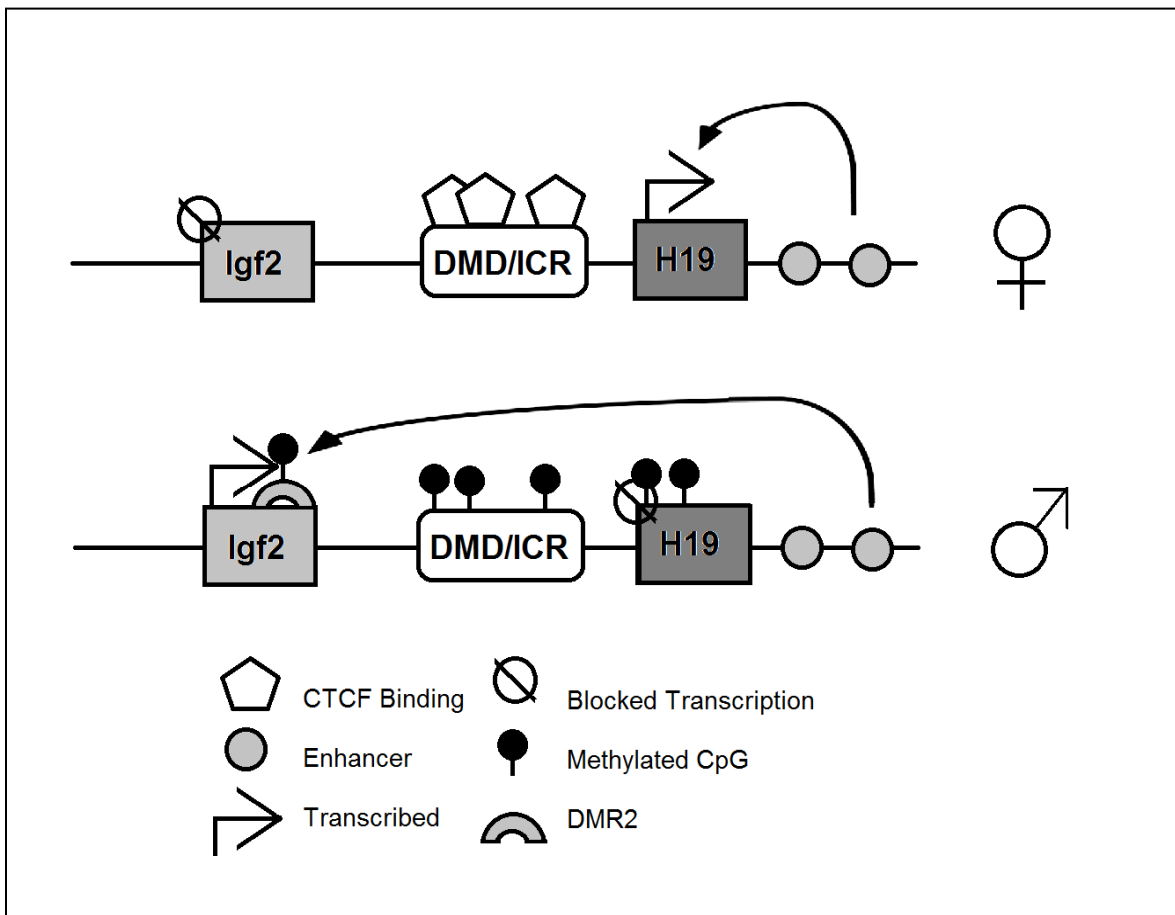
Fig I.3 – Model for regulation of imprinting at the mouse *H19/Igf2* locus.
Modified from: Wan and Bartolomei, 2008.
Shown are the maternally expressed *H19* gene, and the paternally expressed *Igf2* gene. Their shared enhancers are marked by shaded circles. Enhancer to promoter interaction is marked by arrows. On the maternal chromosome, the unmethylated differentially methylated domain (DMD) and imprinted control region (ICR) bind the CTCF protein and forms an insulator that prevents the shared enhancers from communicating with *Igf2*. When CTCF is bound, the enhancers instead facilitates activation of the nearby *H19*. On the paternal chromosome, the methylated (CH_3) DMD/ICR cannot bind CTCF and *Igf2* gene is expressed.

modifications have proven to be key regulators. Although every imprinted gene

has not been studied in an in depth manner, all well characterized imprinted loci

contain differentially methylated regions (DMRs), where 5mC levels are

determined by the parental mode of inheritance.

I.4.2. *Rasgrf1* imprinting

In mouse neonatal brain, the *Rasgrf1* gene is expressed exclusively from

the paternal allele. Allele specific expression patterns are indirectly controlled by

imprinted 5mC on a 251 base pair (bp) retro-transposon like sequence located

30,000bp 5' to the *Rasgrf1* promoter and 3' to an unidentified enhancer. Using

gel shift techniques, and the beta-globin enhancer reporter system (Bell et al.,

1999), studies demonstrated that the DMR not only binds the insulator binding

protein CTCF, but also has enhancer blocker activity (Yoon et al., 2005).

Although a number of CTCF binding regions were identified, only a single site

was identified using the recently characterized consensus CTCF binding

sequence (Essien et al., 2009; Rhee and Pugh, 2011). Because CTCF binding is

inhibited by the presence of 5mC, CTCF fails to bind the paternal allele and

enhancer to promoter interaction is maintained. On the silent maternal allele,

where the DMR is unmethylated, CTCF binds, and prevents the enhancer from

interacting with the promoter (**Fig I.4**). Hence only the paternal allele is

expressed, while the maternally inherited *Rasgrf1* allele remains silent (Yoon et

al., 2005).
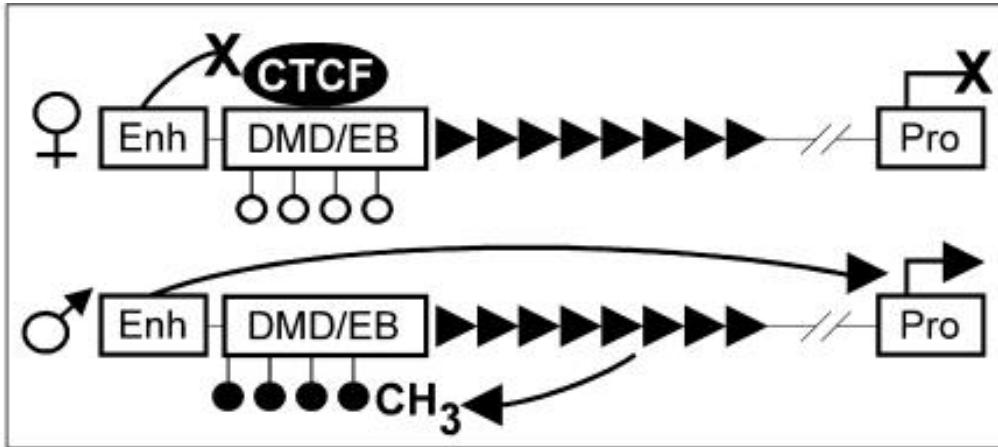
Fig I.4 - Model depicting the *Rasgrf1* DMD and repeat as a binary switch that regulates *Rasgrf1* imprinting
(modified from Yoon et al., 2005)
Shown are the putative neonatal enhancer (Enh), promoter (Pro), repeats (rightward-pointing filled triangles) and the DMD methylated (filled circles) and unmethylated (open circles) on the paternal and maternal alleles, respectively. Curved lines ending in an X indicate blocked interactions or activities, and lines ending in an arrowhead indicate those that are permitted. On the maternally transmitted allele, there is no DNA methylation and CTCF binding facilitates the enhancer-blocking activity of the DMD, which prevents a yet-to-be-identified upstream enhancer from activating *Rasgrf1* transcription. On the paternally inherited allele, the *Rasgrf1* DMD is methylated and prevents CTCF binding. Because CTCF is not bound, the enhancer blocker does not function on the paternal allele and enhancer-to-promoter communicate to permit expression of the paternal *Rasgrf1* allele.

I.4.3. The *Rasgrf1* repeat region

Imprinted methylation at *Rasgrf1* is controlled by a region of repetitive

DNA located adjacent to the DMR. (**Fig I.4**) The repetitive region consists of a

degenerate 41nt sequence repeated 40 times. By deleting this region during

gametogenesis, and during embryonic development, it was determined that the

repeats are both necessary (Yoon et al., 2002) and sufficient (Park et al., 2012)

for establishing 5mC in the male germline and for maintaining 5mC during early

somatic development (Holmes et al., 2006). The repeats combined with the DMR

constitute the first documented *cis*-regulatory sequence responsible for

deposition of 5mC, and together act as a binary switch that regulates *Rasgrf1*

imprinted expression. Prior to these studies focusing on *Rasgrf1*, no *cis*-acting

5mC regulatory sequence had been identified. One significant question remains

however: "**What features of the repeat region enable it to act in such a way?**"

Answering this question may provide insight into a mechanism to explain how

DNA methylation is localized in the genome at various times during development.

I. 5. Combinatorial relationships of epigenetic marks

I. 5.1. H3K27me3 and 5mC at *Rasgrf1*

The unmethylated maternally inherited copy of the *Rasgrf1* DMR happens

to be marked by H3K27me3, while the paternally inherited allele has none.

Interestingly, deletion of the paternal *Rasgrf1* repeat region not only caused loss

of 5mC in *cis*, but also caused acquisition of H3K27me3 to the paternal DMR (Lindroth et al., 2008). Direct inhibition of DNMT1 with 5-aza-2'-deoxycytidine (5azaC), not only caused a decrease in 5mC levels, but also caused increased levels of H3K27me3 specifically on the paternally inherited DMR (Lindroth et al., 2008). These experiments indicated that there is an antagonistic relationship between H3K27me3 and 5mC at *Rasgrf1*. Specifically, the presence of one epigenetic mark excludes the placement of the other mark. Recently published data demonstrated that this antagonistic relationship is maintained genome wide in normal tissue (Brinkman et al., 2012; Hagarman et al., 2013). **It is possible however, that the antagonism is specific to a subset of cell types.**

I.5.2. Is epigenetic coordination the norm?

This type of coordinate regulation is fairly common through out the epi-genome, where numerous other combinatorial epigenetic effects have been observed. In *Arabidopsis*, when present over the promoter there is mutual antagonism between 5mC and H2AZ (Zilberman et al., 2008), indicating that not only can 5mC affect TF binding (as is the case with the previously mentioned CTCF), but it can also directly impact chromatin structure. In yeast, H2AZ antagonizes H3K36me, a mark associated with active gene transcription within the gene body (Li et al., 2005). Combinatorial relationships have also been observed in higher organisms, and can act cooperatively rather than antagonistically. For example, in mammalian embryonic stem cells, H3K4me3

and H3K27me3 are often found together, on the same nucleosome over promoters of genes where precise temporal expression is required for developmental progression. Because these marks typically oppose one another (one is associated with active transcription while the other with silencing) they are thought to render a gene "poised" enabling it to be rapidly activated by H3K4me3 or repressed by H3K27me3 (Bernstein et al., 2006). In recent years numerous marks have been assayed, and mark coordination has been implicated in a variety of important processes (Ram et al., 2011) including developmental differentiation (Mikkelsen et al., 2007), gametogenesis (Ooi et al., 2007), and DNA replication (Eaton et al., 2011). Many of these studies that assay for coordination of epigenetic marks rely on correlations to conclude that marks are physically together on the same chromatin fragment. This post process analytical method remains controversial however. For example, the combination of H3K4me3 and H3K27me3, which render a gene "poised", were once thought to antagonistic. Prior to sequential ChIP experiments, where the product from one ChIP is used for a second ChIP, work based on the *Drosophila* Hox gene clusters concluded that the activating H3K4me mark, placed by the Trithorax complex, restricted to the repressive H3K27me mark, placed by the Ploycomb complex (Papp and Müller, 2006). This conclusion was based on correlations from ChIP and gene expression datasets. Later, sequential ChIP (reChIP) experiments (which will be discussed the next section) were used to determine

that, although these marks serve opposite functions, they are not actually

antagonistic (Bernstein et al., 2006).

I.5.3. Analysis of coordinate epigenetic marks

The ChIP technique is used almost exclusively to determine where in the

genome certain epigenetic marks reside, and is traditionally performed only on

one epigenetic mark at a time. In these experiments DNA from pull downs using

separate antibodies against distinct epigenetic marks are sequenced, and then

the location of each epigenetic mark is mapped to the genome separately. After

genomic mapping, the overlap of these marks is typically reason enough to infer

casually they occur together on the same chromatin fragment. This is a flawed

inference; only after sequential ChIP can one confirm whether various marks do

indeed reside on the same individual chromatin fragment (Bernstein et al., 2006).

In sequential ChIP experiments, two immuno-precipitations (IPs) are performed

where the purified chromatin from the first pull down is used as the input for the

second pull down. Though useful, this technique is extremely difficult to perform,

and often times not sensitive enough to detect co-occupancy of two epigenetic

marks within rare cells. This is due to the large amounts of input required for a

relatively low yield. Sequential ChIP is completely impractical when attempting to

assay for simultaneous presence of more than two marks. Alternatively, a variant

of ChIP can be used to assay for coincidence of histone modifications and mC

(Brinkman et al., 2012); here immuno precipitated chromatin is treated with

bisulfite and sequenced. Additionally mass spectrometry can be used to quantify combinations of histone marks (Johnson et al., 2004), and when combined with ChIP, can be used to assay for two epigenetic marks on different histones within the same nucleosome (Voigt et al., 2012). Each method is labor intensive, highly inefficient, and requires large input quantities for application. **In fact, no currently available technique can assay for multiple epigenetic marks from rare populations of cells.** Recently many labs, including ours, have begun to develop custom methods to solve this problem.

I.6. New techniques for biological research

I.6.1. Single molecule methods in biology

As alluded to in the previous section, the highly variable and dynamic nature of the chromatin fiber within a population of cells contributes to high background levels, increased false negatives, and overall low resolution when performing macro-scale assays like ChIP. For this reason, single molecule based studies offer an attractive alternative when assaying chromatin molecules. They allow for direct inspection of molecules, without ensemble averaging, and can be preformed using very small amounts of input material (Wang et al., 2008a; Gorman et al., 2010; Jin et al., 2010). In recent work, single molecule studies on chromatin have facilitated the visualization of DNA repair proteins as they navigate through nucleosomes (Gorman et al., 2010), and have helped to

confirm the existence of centromeric hemisome (Dalal et al., 2007b; Wang et al., 2008b). These and other key findings could not have occurred without the use of single molecule based techniques.

Typically, when assaying, single chromatin molecules, samples are affixed to a solid surface and a high resolution camera is used to image numerous highly ordered arrays in an automated manner such that numerous molecules can be assayed simultaneously. Techniques like this have been used to sequence DNA (Braslavsky et al., 2003; Levene et al., 2003), assay for the assembly of transcription machinery (Blair et al., 2012), as well as identify histone proteins on chromatin molecules (Cerf et al., 2012). Although these techniques are powerful, because the molecules are affixed to a surface, sample recovery is difficult, and throughput it limited.

As an alternative, nanofluidic channels are often used for assaying biologically relevant molecules. The small dimension of the nano-channels, which often creates a sub-femtoliter confinement volume, when combined with Fluorescence Correlation Spectroscopy (FCS), enable samples to be analyzed at relatively high concentrations (>1nM), and allow for increased overall throughput while maintaining single molecule resolution (Elson, 2011). For example, using a combined nanofluidic and FCS based platform, DNA sequences were assayed in high throughput at near physiological concentrations (Levy and Craighead,

2010). When combined with multicolor fluorescence microscopy, single molecule nanofluidic based techniques may allow for the detection and characterization of multi-protein complexes as single molecules. In this manner, if these types of techniques were adapted to assay chromatin, they could potentially be useful for simultaneously detecting multiple epigenetic marks on chromatin purified from only a few cells.

Because methods for extracting small amounts of material are limited, often traditional macro-sized extractions are performed, and only a small amount of purified material is analyzed. Essentially, the power of single molecule techniques to assay small amounts of material is negated by the fact that such small amounts are not available when assaying rare and difficult to purify cell types.

I.6.2. A contained system

In recent years numerous labs have begun to take advantage of micro-fluidic systems to extract material that is too precious to be analyzed using more traditional techniques. For these types of studies it has become common to design custom micro scale devices for extraction and "in-line" analysis of material. For example, fluidic channels have been designed that can execute all aspects of gene expression analysis on single cells, including cell capture, lysis, reverse transcription, and quantitative polymerase chain reaction (qPCR) (White

et al., 2011). Because the experimental pipe-lines are contained within fluidic devices, these assays require no physical handling of reagents, essentially reducing sample loss and increasing sensitivity. Others have applied similar techniques for assaying histone modifications. In these systems numerous buffer containing reservoirs and actuated valves are enclosed within microfluidic channels, allowing for the blocking, washing and binding steps necessary to implement small scale chromatin immuno-precipitation (Wu et al., 2009, 2012). Similarly, cell extraction devices have been fabricated to purify small amounts of DNA from only a few cells (Pasquardini et al., 2011).

Ideally, if these types of extraction devices were fused to a nanofluidic single molecule detection platform, it may allow for efficient extraction and high-throughput analysis of epigenetic marks on chromatin isolated from very few cells. This would first require the micro-scale extraction device, and buffer conditions to be altered in order to purify chromatin and be compatible with a nanoscale analysis device. It would also require the nanofluidic channels to be altered to allow chromatin to flow and epigenetic marks to be characterized. Finally, it would require those molecules containing a given epigenetic mark to be purified and recovered for downstream sequencing. Developing a system to detect epigenetic marks on single chromatin molecules in a nanofluidic system has formed the bulk of my graduate work.

I.7. Experimental objectives

The broad focus of this dissertation will present work that furthers the understanding of epigenomic mark regulation.

To address the following questions:

1. How is 5mC localized during its establishment in germ cells?

Using the highly characterized *Rasgrf1* system as a model, I will demonstrate that the repeat region functions as a promoter for a non-coding RNA, and through the PIWI pathway it is able to impart *cis* mediated 5mC during male gametogenesis.

2. Can single molecule methods be used to assay for epigenetic marks on chromatin molecules?

I will describe single molecule nanofluidic FCS based techniques that are capable of detecting intact chromatin, assaying for epigenetics marks, and purifying DNA based on 5mC.

3. Is the relationship between H3K27me3 and 5mC that is observed at *Rasgrf1* maintained genome wide, and can simultaneous co-occupancy of these two epigenetics marks be detected using single molecules techniques?

I will describe single molecule nanofluidic assays for directly detecting

multiple epigenetic marks simultaneously and use it to demonstrate that

genome wide 5mC prevents the placement of H3K27me3 in normal cells,

but breaks down during cellular immortalization.

II. How is 5mC localized during its establishment in germ cells?

II.1 Abstract - piRNA mediated *Rasgrf1* imprinting

Previous work has determined that imprinting at the *Rasgrf1* locus is directly controlled by 5mC, which is placed on a differentially methylated region (DMR) located 30kb 5' to the gene promoter, and is dependent on a *cis* acting repeat region (Yoon et al., 2002, 2005; Herman et al., 2003; Holmes et al., 2006). Although the pattern of methylation and the mechanism by which imprinted expression is controlled have been rigorously studied, the precise model to explain acquisition of methylation is not well resolved. Chiefly, it remains unknown how the *Rasgrf1* repetitive domain is able to impart 5mC at the neighboring sequence. Preliminary data and pilot experiments suggested that a ncRNA plays a role in this process. This hypothesis was confirmed during a collaboration with Hiroyuki Sasaki, through which we were able to determine that the repeat region functions through the PIWI/piRNA repressive pathway to deposit paternal allele-specific 5mC during mouse gametogenesis at embryonic day 16.5 (e16.5). Our results indicate that the repeats function as a promoter for transcription of a piRNA-targeted ncRNA (pitRNA), which is required for proper methylation of the DMR. Experiments proved that repeat-mediated deposition is dependent on transcription through the DMR in *cis* and not in *trans*. These data led to a new model for co-transcriptional acquisition of imprinted 5mC at the *Rasgrf1* locus.

II.2 Introduction

II.2.1 *Rasgrf1* imprinting

The *Rasgrf1* gene is paternally methylated and expressed in mouse
neonatal brain (Plass et al., 1996), and allele specific expression is regulated by
5mC on the DMR, which contains an RMER4B like sequence element, and is
located 30kb 5' to the *Rasgrf1* promoter. (**Fig II.1A**). Located directly 3' to the
DMR is a highly repetitive region, which is not only required for establishment of
5mC over the DMR (Yoon et al., 2002), but is also required for maintenance of
5mC in the zygote (Holmes et al., 2006) (**Fig II.1B**). When the repeat region was
removed at any point between fertilization and epiblast formation the embryonic
DMR became hypomethylated in somatic DNA of neonates. When the repeats
were removed subsequent to this critical window, there was no effect on 5mC.
These studies demonstrated that the repeats combined with the DMR constitute
the first documented *cis*-regulatory sequence responsible for deposition of 5mC,
and together act as a binary switch that regulates *Rasgrf1* imprinted expression.
In independent studies the DMR was found to directly bind the insulator binding
protein CTCF (Yoon et al., 2005). This binding occurred only in the absence of
capability, the *Rasgrf1* DMR and repeats were inserted into the chicken beta-
globin enhancer and neo reporter system (Chung et al., 1993). In this assay,
5mC, and was therefore hypothesized to occur on the unmethylated maternal
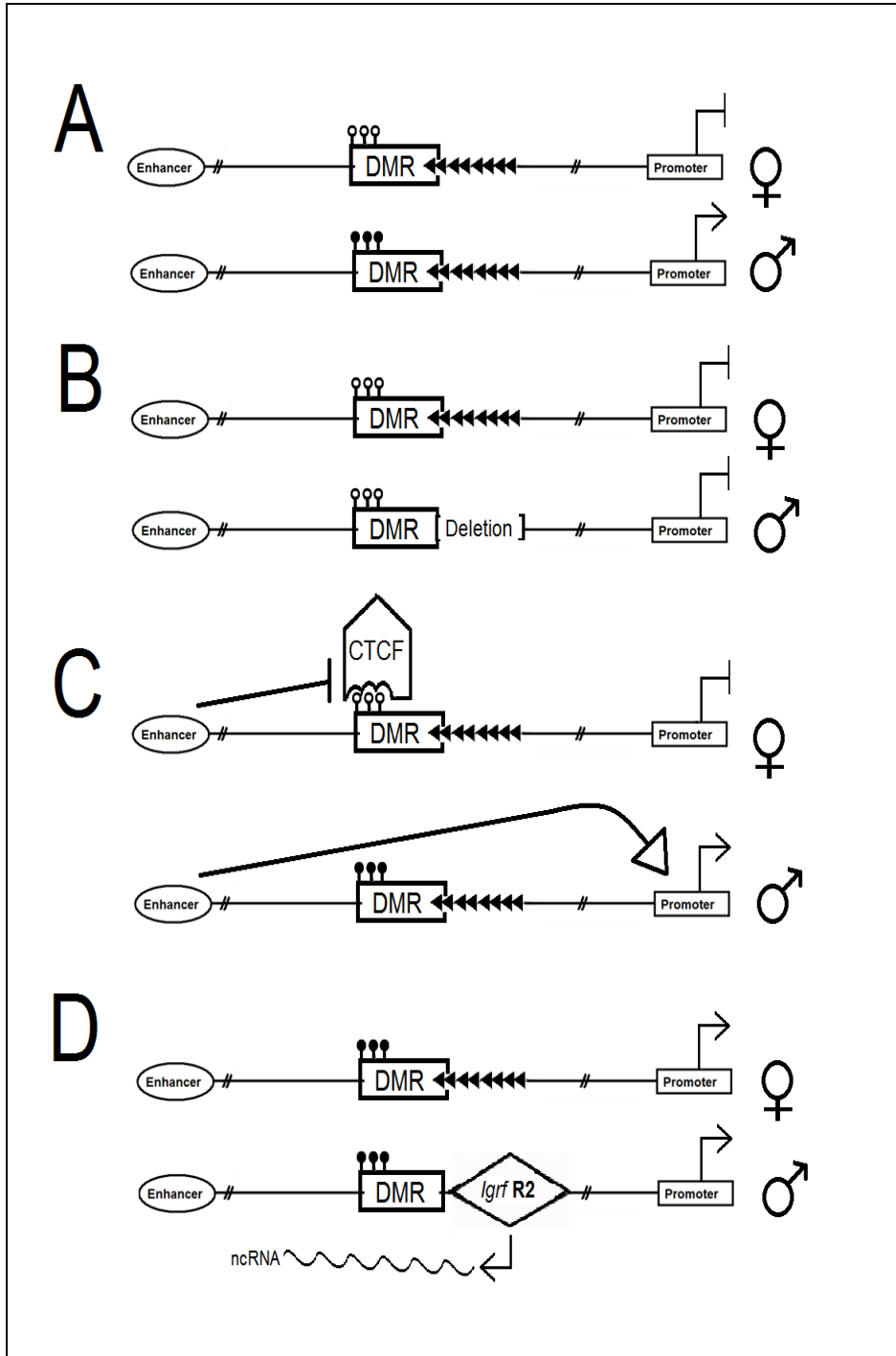
Fig II.1 - Model to explain patterns of imprinted 5mC and expression at the *Rasgrf1* locus.

Methylation is illustrated by dark lollipops and lack of methylation is illustrated by open lollipops. Black arrowheads represent repeat sequences. The expression over the promoter is represented as on with an arrow, or off with lack of an arrow. ncRNA transcript is illustrated using a long wavy line.

A) 5mC is present over the paternal differentially methylated region (DMR) and absent from the maternal DMR. This pattern correlates directly with promoter expression.

B) When the repeat region is deleted from the paternal allele there is failure to establish methylation over the DMR, and improper gene expression.

C) The unmethylated maternal DMR binds the insulator binding protein CTCF and prevents expression. On the paternal allele CTCF cannot bind, the enhancer can interact with the promoter, and facilitate monoallelic expression.

D) The R2 region from *Igf2*r, which is the promoter for an ncRNA, was inserted in place of the *Rasgrf1* repeat region on the paternal allele. It successfully recapitulated wild type Rasgf1 paternal establishment of 5mC, however it caused aberrant 5mC in trans to the normally unmethylated maternal allele.

and not the methylated paternal allele (**Fig II.1C**). To test for enhancer blocker

when placed between the enhancer and promoter, the *Rasgrf1* sequences

significantly reduced enhancer to promoter communication regardless of

orientation, indicating that the DMR sequence functions as an enhancer blocker.

These combined studies led to a model where on the silent maternal allele the

unmethylated DMR binds CTCF and prevents the enhancer from interacting with

the promoter. On the paternal allele however, where the DMR is methylated (in a

repetitive sequence dependent manner), CTCF cannot bind and the enhancer

efficiently communicates with the promoter, leading to imprinted mono-allelic

expression.

II.2.2 Evidence for ncRNAs at *Rasgrf1*

Two main questions that came from these studies were namely: "How do

the repeats function during establishment of methylation in the paternal germ

line?" and "How do they operate to maintain methylation in the developing

embryo?" Interestingly, *Igrf2r* is also imprinted in mouse, and similar to *Rasgrf1,*

imprinted expression is maintained through allele specific methylation in *cis*. In

contrast to *Rasgrf1* however, methylation is present on the maternal *Igf2r* allele

and absent on the paternal allele. Within the second intron of the *Igrf2r* locus is a

differentially methylated region known as region 2 (R2). When this region is

methylated on the maternal allele there is proper mono-allelic maternal

expression, and when R2 is deleted from the paternal allele, the gene loses

imprinted methylation and is expressed regardless of parental origin (Wutz et al.,
1997).

At the time, it was hypothesized that sequence features contained within
the R2 sequence were able to impart maternal specific 5mC (Birger et al., 1999),
and in a similar way, different sequence features contained within *Rasgrf1*
repeats were able to impart paternal specific 5mC. In a way, the mechanisms
acting on the maternal allele of *Igf2r* were the similar to those acting on the
paternal allele of *Rasgrf1*. To test this hypothesis, a mutant mouse was made
where the R2 region was inserted in place of the *Rasgrf1* repeat region (**Fig
II.1D**). When the mutation was transmitted through the paternal allele, it was able
to impart 5mC to the paternal *Rasgf1* DMR. Surprisingly however, paternal
transmission also led to improper acquisition of 5mC on the maternal DMR in
trans (Herman et al., 2003). Independent work led to the discovery that the R2
region functions as a promoter for a long ncRNA (named *Air*), which is paternally
transcribed and required for proper imprinted expression at *Igf2r* (Sleutels et al.,
2002). Subsequently, it was hypothesized that the observed *trans* affect was
mediated by improper regulation of an unidentified endogenous *Rasgrf1* ncRNA,
and that ncRNA is able to regulate 5mC at the DMR.

Regulation of 5mC by ncRNAs has several precedents. In plants, RNA
that functions during post-transcriptional gene silencing of transgenes is thought

to guide 5mC methyltransferases to specific DNA based on sequence homology (Jones et al., 1999). In human cells (HeLa), shRNAs generated *in vitro*, which were made complementary to the RASSF1A promoter, can target *de novo* DNA methylation and induce gene silencing (Castanotto et al., 2005). PIWI proteins, responsible for processing of short 26-31bp small RNAs called piRNAs, which are required for male gametogenesis, have also been shown to mediate deposition of 5mC (Kuramochi-Miyagawa et al., 2008). These piRNAs, which interact with PIWI subfamily proteins, function to suppress retrotransposon activation during primordial germ cell epigenetic reprogramming, and specifically, transposon sequences are hypomethylated when a member of the PIWI protein family (MILI) is mutated in mouse (Kuramochi-Miyagawa et al., 2004; Aravin et al., 2007b).

Because small RNAs by nature are complementary to the DNA sequence from which they are derived, often it is hypothesized that they serve as a guide, and provide protein complexes with the capacity to target specific genomic locations. Many mammalian piRNAs arise as a product of a biogenesis cycle, which is often referred to as "the ping-pong cycle", and requires transcription of two complimentary ncRNAs (**Fig II.2**) (Aravin et al., 2008). Some have even speculated that complimentary piRNAs, produced from the ping-pong cycle, can recruit both PIWI and chromatin modifying complexes to simultaneously cleave target RNAs and establish new repressive epigenetic states. This new epigenetic
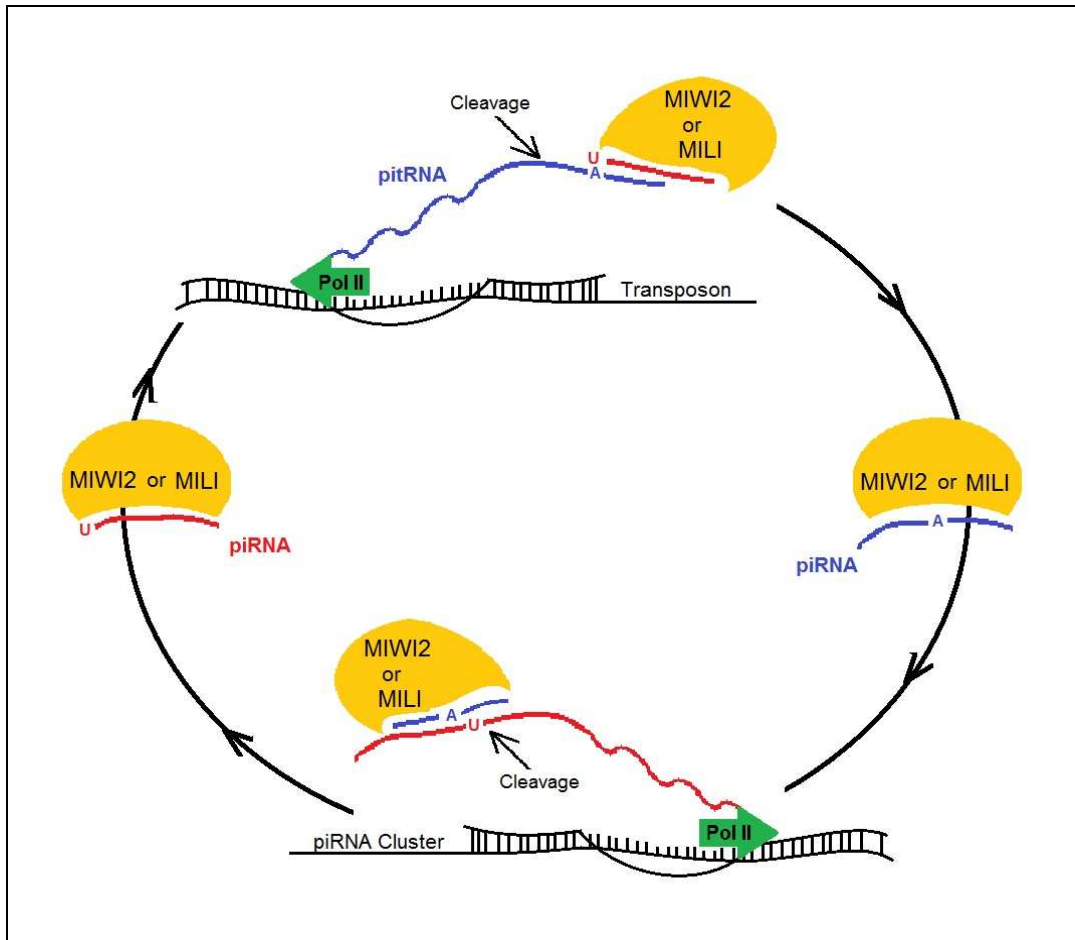
Fig II.2 - A model of mouse ping-pong cycle
In the male germline, an active transposable element, which is transcribed in an opposite orientation to a piRNA cluster, is the substrate for primary piRNA (blue) processing. Presumably, these piRNAs interact with either MILI or MIWI2 to form a complex. The complex is then guided to the piRNA cluster by base complimentarily, where cleavage of the secondary antisense cluster derived piRNA (red) occurs. The secondary piRNA then forms a second complex, which facilitates processing of the transposon derived piRNA. This process leads to cyclic amplification of both piRNAs, transcript silencing, and ultimately acquisition of m5C to the transposon.
Adapted from (Brennecke et al., 2007).

state is then proposed to be stabilized by the recruitment of DNA

methyltransferase (Aravin et al., 2008) (**Fig II.2**).

In summary, a repeat region within the *Rasgrf1* locus controls methylation

patterns at the adjacent DMR. The DMR is homologous to an RMER4B

retrotransposon sequence, and like other retrotransposon sequences, 5mC

establishment takes place during male gametogenesis. When the R2 mutation

was made to the paternal allele at the *Rasgrf1* locus, there was improper

acquisition of 5mC in *trans* presumably through ncRNA mechanisms. These

combined data led to the hypothesis that establishment of 5mC at the *Rasgrf1*

DMR was dependent on ncRNAs. Because *Rasgrf1* RMER4B epigenetic

reprogramming takes place in the germ line, at the same time and in the same

tissue as PIWI mediated transposon silencing, piRNAs were the most likely

candidate for regulating *Rasgrf1* methylation. So, we embarked on a

collaboration with an expert in the field of piRNA mediated epigenetic

reprogramming, Hiroyuki Sasaki.

II.3 Materials and methods

II.3.1 Collaborator contributions

Spermatogonia isolation, preliminary bisulfite sequencing (BS-Seq), small

RNA library preparation, sequencing and mapping analysis, were performed in

the Sasaki lab and are described in our collaborative publication (Watanabe et al., 2011b). The experiments, which I performed, are described below.

II.3.2 Mouse breeding

Mice harboring the BJR3 targeted *Rasgrf1* mutation, in which the repeat region is deleted, were generated previously and maintained on the C57BL/6 (B6) background (Yoon et al., 2002). Creation of mice harboring the RC1 transgene (which contains the entire *Rasgrf1* locus), was on the FVB/n (FVB) background, took place in the Soloway lab, and is described in an additional publication, in which I hold co-authorship (Park et al., 2012). These two mutant mouse lines (BJR3 and RC1) were crossed over a series of generations to create mice that were both hemizygous for RC1 and homozygous for BJR3. Reciprocal wild type outbreed FVB to B6 mouse crosses were also made, and used for the purpose of generating offspring with a high level of sequence polymorphism.

II.3.3 Mouse embryonic day 16.5 tissue collection

To determine the precise date of conception, females from various breeding crosses, were monitored twice daily (in the morning and in the evening) for the presence of vaginal plugs. Sixteen days after plugs were observed females were dissected and embryos were harvested. From each individual male embryo, two testicles and a single limb were collected. DNA was extracted from the limb and used for genotyping or for methylation analysis. RNA was

extracted from the testicles and used for expression analysis. In cases where genotypes were known prior to dissection (for example: homozygous BJR3 embryos) no direct genotyping was performed and up to 12 embryonic testes were pooled prior to RNA extraction.

II.3.4 pitRNA and piRNA cluster expression analysis

RNA was extracted from either e16.5 embryonic or adult testes and complimentary DNA (cDNA) was made for Reverse Transcription PCR (RT-PCR) using random oligonucleotides. To assess pitRNA expression, cDNA was amplified using primers specific to the *Rasgrf1* pitRNA DMR sequence. To amplify the *Rasgrf1* pitRNA in either Wt mice or in BJR3 homozygotes, and to amplify the RC1 pitRNA in a BJR3 homozygous mouse background, primers were: F-ATACGGGCAACCTTGGGATCATAGGCA, and R-CAAAATTCTTACTACACATGGCACA, to give a 251bp product. To amplify RC1 in a wild type (Wt) *Rasgrf1* background, primers were: F-CTGCACCGCTGCCGCTAAGC and R-ATCACTAGTGCGGCCGGCCGCCTGCA, which generate a 198bp product. To amplify the endogenous *Rasgrf1* pitRNA in RC1 mutants, primers were: F-CTGCACCGCTGCCGCTAAGC, and R-GCAGCAGTAGCAGTCGTGGT, which generate an 85bp product. To assess imprinted expression of the piRNA cluster cDNA was amplified using primers specific to the cluster region located on chromosome 7. Primers were: F-TGTTAACAGTTGAGGTATTTATTTTTG, and

R- TAGTCACCTTAATGGGAGCAAAATC to produce a 139bp product, and spanned a single nucleotide polymorphism (SNP) between FVB (SNP=A) and B6 (SNP=T). To control for total cDNA loading, *Rpl32* (F- CATGCACACAAGCCATCTACTCA, R- TGCTCACAATGTGTCCTCTAAGAAC), which generates a 128bp product, and *Timp1* (F- ACTCTTCACTGCGGTTCTGGGAC, R- GTCATAAGGGCTAAATTCATGGG) primers were used, which produced a 389bp fragment from DNA and a 194bp fragment from cDNA.

II.3.5 Mouse genotyping

To genotype the RC1 and BJR3 mutant mice, DNA was extracted from either adult tails for embryonic soma, and the following primers were used for multiplex PCR amplification: F- GCACTTCGCTACCGTTTCGC, R- TGTCCTCCACCCCTCCACC, and W- TTTCTGCCATCATCCCAGCC, where F and R amplify the BJR3 allele and F and W amplify the RC1 allele. Using these primers in multiplex PCR, the Wt allele should be amplified from W and R to produce a 190bp fragment, the RC1 allele should be amplified using W and R to produce a 280bp product, and the BJR3 allele should be amplified with F and R to produce a 351bp product.

II.3.6 Methylation analysis

Methylation analyses of the RC1 and BJR3 alleles were performed using the Sequenom Epityper technique (Ito et al., 2010). In short, the technique combines bisulfite treatment, with PCR, *in vitro* transcription, RNA nucleotide specific cleavage, and mass spectrometry. Bisulfite treatment of DNA was performed as described previously (Park et al., 2012). Treated DNA was amplified using primers that contained both a region of homology (not underlined) and either a tag (double underlined), or a T7 (underlined) primer sequence (on the reverse primer). The following primers were used to amplify the RC1 allele: F-<u>TGGCCTTGCT</u>GTTGTTGTTTTTATATTTATT and R-<u>CAGTAATACGACTCACTATAGGGGAGAAGGCT</u>CAAAAACAACAATAATAACA AAAACAAAAACAATAT to produce a 416bp product. The following primers were used to amplify the BJR3 allele: F-<u>TGGCCTTGCT</u>GTTGTTGTTTTTATATTTATT and R-<u>CAGTAATACGACTCACTATAGGGAGAAGGCT</u>CCACCCCTCCACCCCTCTCC TAAAAAAA, which produced a 381bp product. After amplification PCR products were submitted to the Cornell Core Facility where the remaining steps were completed.

II.4 Results:

II.4.1 Small RNAs at *Rasgrf1* (collaborator contributions)

We hypothesized that establishment of 5mC at the *Rasgrf1* DMR was dependent on ncRNAs. Interestingly, the DMR sequence is a degenerate sequence derived from a retrotransposon called RMER4B. As mentioned previously, transposon silencing in mouse gametes is dependent on PIWI pathway mediated DNA methylation (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008).

To test the hypothesis that the *Rasgrf1* DMR methylation was dependent on the PIWI pathway our collaborators surveyed small RNA libraries made from embryonic testes, searching for 24-30nt sequences with *Rasgrf1* DMR homology. Interestingly, numerous unique reads were identified that mapped to the *Rasgrf1* DMR sequence (**Fig II.3A**). To confirm these small ncRNAs were piRNAs, small RNA libraries were made from *MitoPLD* mutant testes. *MitoPLD* is a nuclease domain-containing protein known to be involved in primary piRNA production and thought to have RNase activity (Watanabe et al., 2011a). In the *MitoPLD* mutant library, the *Rasgrf1*-specific small RNAs were considerably depleted (**Fig II.3B**), indicating that they were indeed piRNAs and their presence was dependent on components of the PIWI piRNA processing machinery. Since piRNAs are essential for for establishment of 5mC in retrotransposons, our collaborators
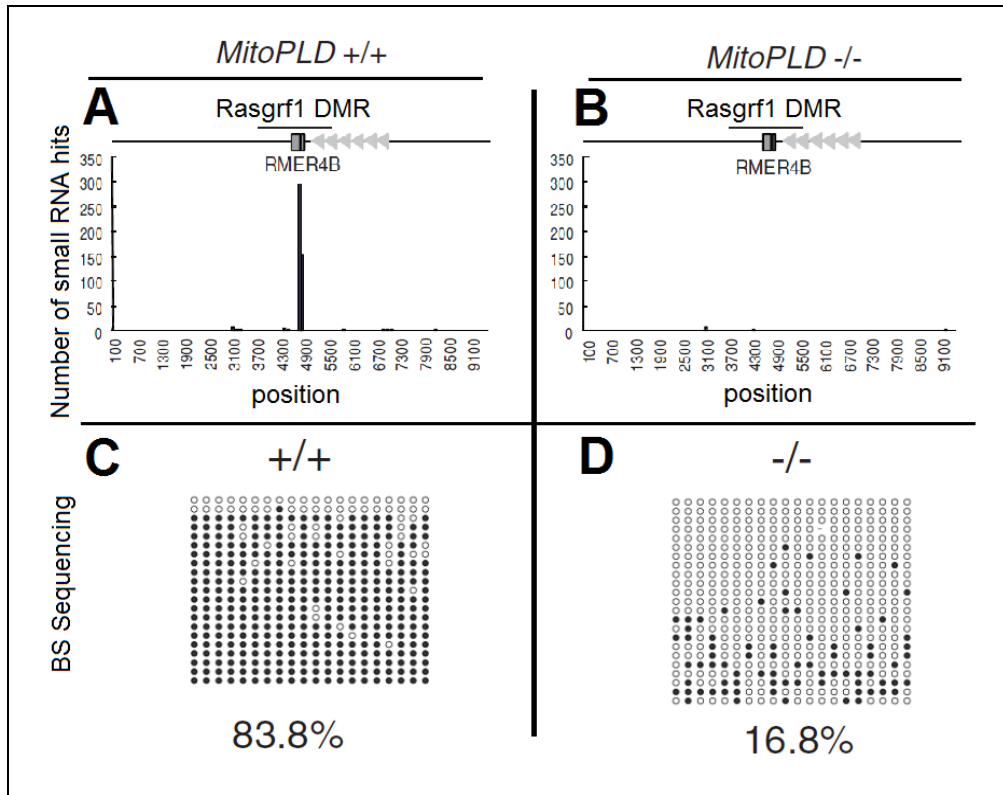
Fig II.3 - The PIWI pathway member *MitoPLD* controls methylation at the *Rasgrf1* DMR.

A + B) Numbers and locations of Small RNA abundance from *MitoPLD+/+* (A) and *MitoPLD–/–* testes (B) from 16.5 embryonic testes, mapping to the *Rasgrf1* DMR (top). Results for small RNAs mapping was relaxed to include up to two mismatches including indels.

C + D) Bisulfite sequencing to determine methylation level in *MitoPLD+/+* (C) and *MitoPLD–/–* (D) spermatogonia. Each row corresponds to a given sequencing run where circles represent a single CpG. Open circles depict lack of methylated, and filled in circles depict methylated CpGs. Loss of methylation correlates with decreased abundance of small RNAs.

reasoned that *Rasgrf1*-specific 5mC may also be impaired in the *MitoPLD*

mutants. To test this, they performed BS-Seq of the locus. There was a 67%

reduction in 5mC over the *Rasgrf1* DMR in *MitoPLD* mutants (**Fig II.3C+D**),

indicating that like retrotransposons, *Rasgrf1* is dependent on the piRNA

pathway for proper establishment of DNA methylation. The reduction in

methylation observed in the *MitoPLD* mutant was similar to the levels observed

when the *Rasgrf1* repeat region was deleted. We therefore wondered: Does the

repeat region somehow function in accordance with the piRNA pathway to

regulate imprinted 5mC?

II.4.2 Repeat region is a promoter for pitRNA

During piRNA ping-pong biogenesis it is hypothesized that transcription

from two different genomic regions is required for proper silencing. The model

proposes that two long ncRNAs, one from the target locus and the other from a

piRNA cluster region, are sequentially processed to generate two small 24-30bp

piRNAs that have a 10bp overlapping region of complementarity (Aravin et al.,

2007a, 2007b). It also predicts that removal of one of the long ncRNAs would

result in failed piRNA processing, and derepression of the target locus. We

therefore reasoned that the repeat region is required for transcribing a noncoding

DMR spanning RNA. To test this hypothesis, we extracted RNA from e16.5

testes, made cDNA and PCR amplified using primers that partially overlapped

with the DMR (**Fig II.4A**). PCR was performed using cDNA made from both Wt
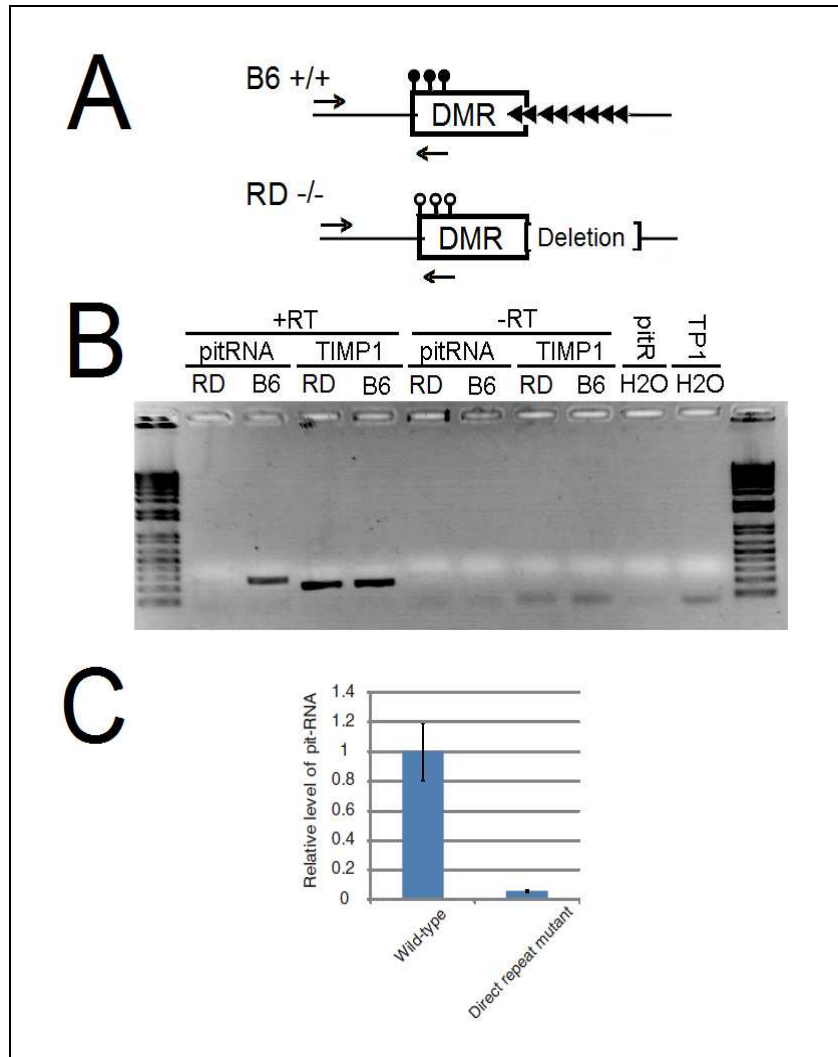
Fig II.4 - Detection of pitRNA trascript from *Rasgrf1* in embryonic day 16.5 testes

   A) A schematic illustration of the *Rasgrf1* locus (similar to figure II.1) primer locations are marked for both Wt (B6) *Rasgrf1* and direct repeat deletion (RD) mutant mice. Arrows correspond to primer binding sites.

   B) Detection of pit-RNA by reverse transcription polymerase chain reaction (RT-PCR) in germ cells from both mutant repeat deletion (RD) and Wt (B6) embryonic day 16.5 testes. *Timp1* shown as an RNA loading control.

   C) RT-PCR was preformed similar to B, however levels were quantified using qPCR. Error bars represent standard deviation (n = 3). The level of *Oct4* mRNA was used as a reference.

and mutant BJR3 e16.5 testes RNA. We anticipated that if the repeats are required from expression of the *Rasgrf1* ncRNA, we would see a PCR product only from Wt cDNA and not from BJR3 cDNA, which is precisely what we observed (**Fig II.4B**). In fact, the BJR3 homozygous mutant had <10% of the Wt pitRNA transcript levels (**Fig II.4C**). We concluded from these experiments that the repeat region functions as a promoter for a ncRNA, which is transcribed through the DMR, and is processed into smaller piRNAs. Furthermore, based on what has been observed at other retrotransposons, we believe the *Rasgrf1* specific RNA is targeted by other piRNAs and functions to recruit epigenetic modifiers, like DNMTs or HMTs. It is therefore referred to as the piRNA-targeted ncRNA (pitRNA). Future research will ultimately determine how these factors are recruited.

Because there is no methylation on the maternally inherited *Rasgrf1* DMR, the pitRNA must mediate deposition of 5mC to the paternal allele in an exclusive manner. It is likely this occurs because pitRNA expression is limited to male gametes (**Fig II.5A**). Although there was a moderate level of allelic bias when the mother was on the B6 background, reciprocal crosses of Wt B6 to FVB and sequencing of RT-PCR products, confirmed that there is no imprinted expression of the piRNA cluster region, and regulation of imprinted 5mC is likely limited to the *Rasgrf1* locus itself (**Fig II.5B**).
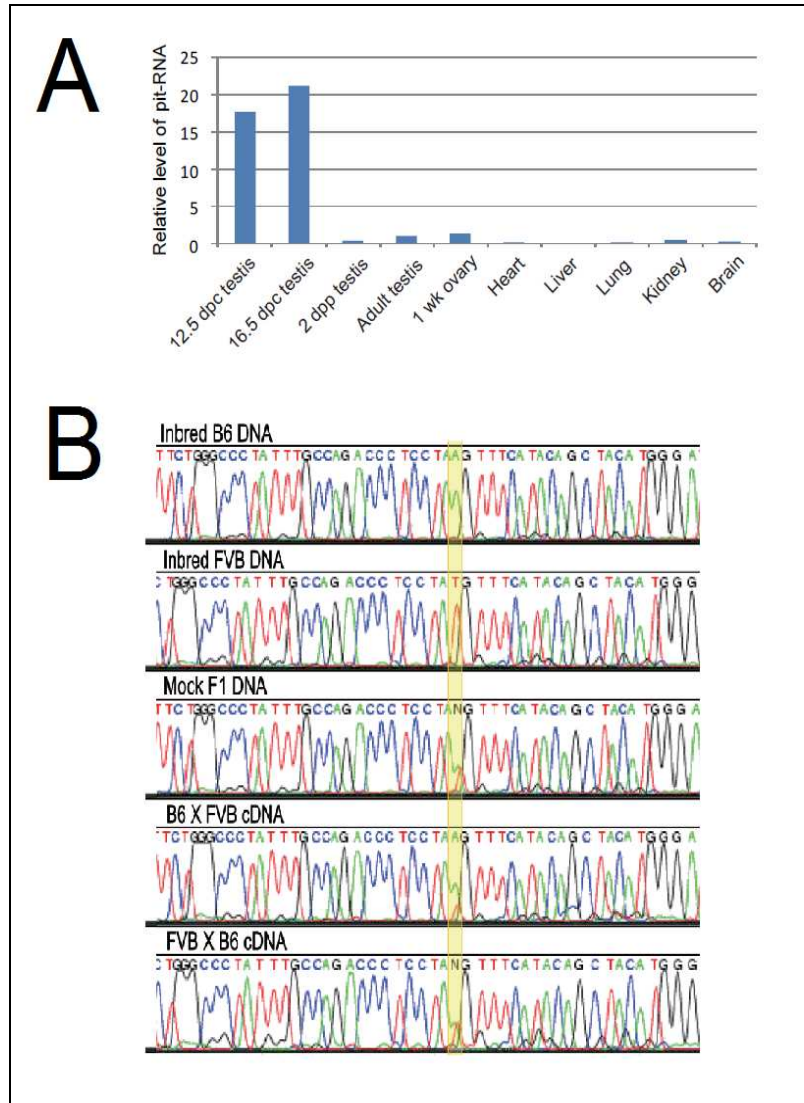
Fig II.5 - Characterization of PIWI pathway RNA transcritpion

A) Fetal testis-specific expression of pit-RNA, using RT-PCR and primers similar to Fig II.4. qPCR was performed with total RNAs from indicated tissues. The expression level of pit-RNA was normalized using that of beta-actin.

B) Allelic expression status of the RNA in the chr7 piRNA cluster. Sequence traces of the RT-PCR products generated from embryonic day 16.5 testes RNA with primers flanking the single nucleotide polymorphism (highlighted in yellow). Control samples were from inbred FVB/n mice (FVB) (top), B6 mice (second from top) or a 1:1 mixture of the two inbred samples to simulate a scenario where both alleles were equally expressed (middle). The F1 test progeny were from B6 mothers and FVB fathers (second from bottom), or from the reciprocal cross (bottom).

II.4.3 *Cis* mediated regulation of 5mC by the pitRNA

It is unknown if transcription at the targeted locus is necessary for 5mC deposition. It is possible that pitRNA function is limited to production and processing of the piRNA species during ping-pong biogenesis. In this scenario we anticipate, when all the necessary machinery is present, that the piRNAs would have the capacity to act in *trans* at the BJR3 locus, which cannot transcribe the pitRNA. Alternatively, if transcription of the *Rasgrf1* pitRNA is necessary for targeting of DNMTs to the DMR, we expect that the piRNAs would act exclusively in *cis*, and DNMTs can only be recruited to the DMR which is actively transcribed.

To distinguish between these two scenarios, we decided to assay a mutant mouse that possessed a transgenic copy of the entire *Rasgrf1* locus (RC1). Aside from repeat-flanking LoxP sites, RC1 is identical and unlinked to the endogenous allele. Previous work has demonstrated that patterns of imprinting at RC1 are also identical to *Rasgrf1*, including establishment and maintenance of 5mC on the paternally inherited allele, and its imprinted expression in neonatal brain (Park et al., 2012). To confirm that the transgenic RC1 pitRNA is expressed in a similar fashion as *Rasgrf1* , allele-specific RT-PCR was performed (**Fig II.6A**). Additionally, to confirm transgenic pitRNA expression, the transgenic allele was crossed through both the material and paternal lineage into the BRJ3 homozygous mutant background. Then, using primers specific to

Fig II.6 – Proper expression of the pitRNA from the RC1 transgene
A) Pedigree showing the crosses used to generate mouse embryos that were homozygous for the BJR3 mutation and hemizygous for the RC1 transgene.
B) RT-PCR to assay for pitRNA expression in RC1 containing, and wild type mice. *Rpl32* primers were used to control for RNA input. When necessary primers were specific to the endogenous allele. Primers were either specific for the transgene derived pitRNA (transgenic) or did not distinguish between pitRNA from the transgenic and endogenous alleles (total).
C) RT-PCR to assay for pitRNA expression. Data demonstrate that the BJR3 mutation and lineage have no effect on pitRNA expression. Embryos were derived from mating where fathers inherited RC1 either maternally or paternally. All mice are BJR3 homozygous mutant at the *Rasgrf1* allele. "Paternal Genotype" refers to RC1 status of male parent, and "RC1 Transmission" refers to genotype of the mice analyzed.

the transgene, we assayed for RC1 specific pitRNA expression in both the Wt

Rasgrf1 (**Fig II.6B)** and homozygous BJR3 mutant (**Fig II.6C**) backgrounds. The

pitRNA was expressed independent of the mutant background, confirming the

RC1 behaves identically to endogenous locus.

These crosses produced double mutants, which were hemizygous for the

RC1 allele and homozygous for the BJR3 allele, allowing us to determine what

role the pitRNA has in establishment of 5mC. As previously mentioned, if the

piRNAs have the capacity to act in *trans* and deposit 5mC at an untranscribed

locus, then the activity of the pitRNA is not limited to the locus from which it was

transcribed and it is likely the pitRNA's function is limited to piRNA production,

which is sufficient for 5mC deposition at the DMR. However, if the piRNAs act

exclusively in *cis*, then aside from piRNA production, transcription of

the pitRNA is required for targeting of DNMTs to the allele from which it was

transcribed. These two scenarios can be distinguished by measuring 5mC levels

over the DMR at both the endogenous BJR3 allele and the transgenic RC1 allele.

To assay 5mC, we performed a modified version of BS-Seq using the Sequenom

Epityper (Ehrich et al., 2005). Although imprinted 5mC patterns were as expected

at the transgenic locus (**Fig II.7B**), there was no acquisition of methylation at

either the endogenous maternal or paternal alleles (**Fig II.7A**). This indicated that

the pitRNA, when expressed from the repeats, can only function in *cis* and not in

trans, and must have a role outside of ping-pong biogenesis in targeting of
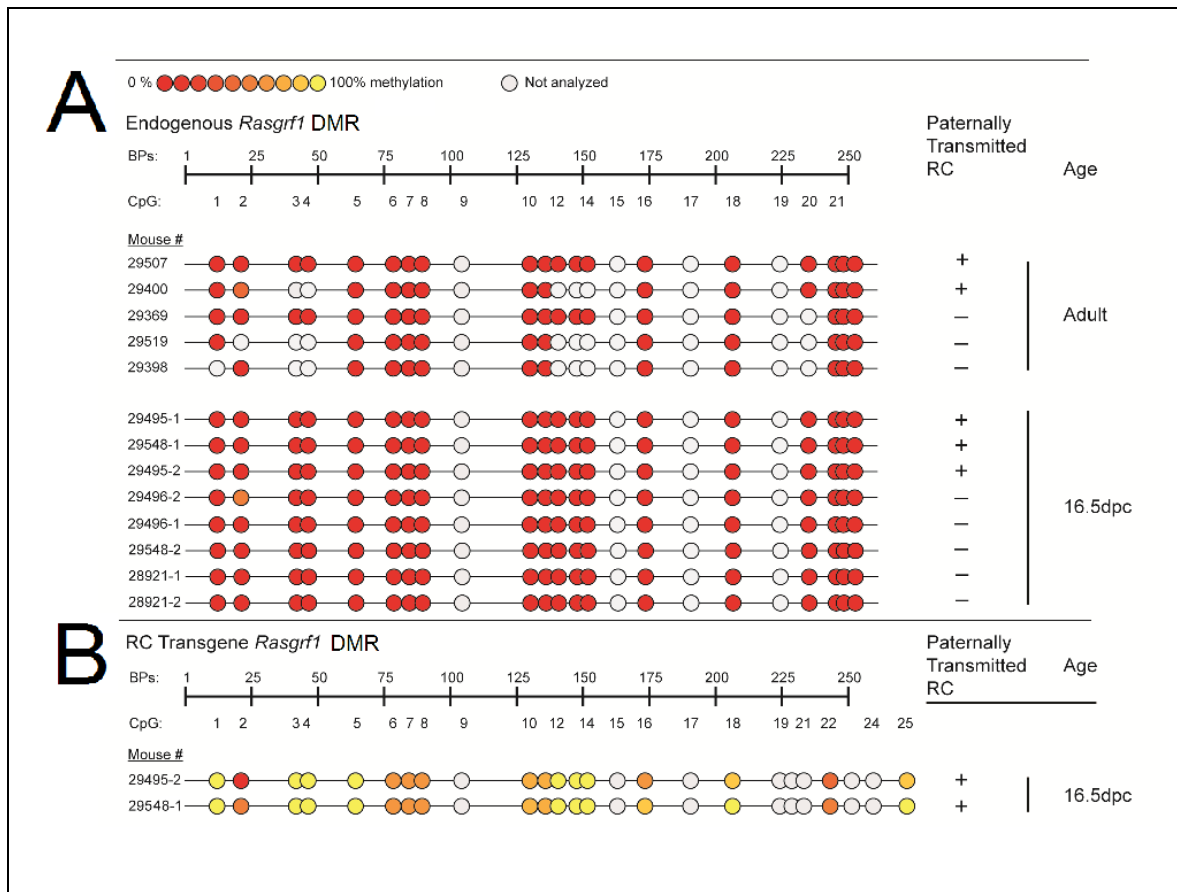
Fig II.7 - pit-RNA transcribed by the *Rasgrf1* repeats controls DNA methylation only in *cis* and not in *trans*

DNA methylation of the endogenous (A) and transgenic (B) copies of the *Rasgrf1* DMD was analyzed by Sequenom MassARRAY. Circles represent a given CpG dinucleotide. Color corresponds to methylation level (top). All mice were homozygous for a deletion of the endogenous copies of the *Rasgrf1* repeats and contained (+) or lacked (−) a paternally inherited copy of the RC1 transgene. Bisulfite PCR assays were specific for the endogenous (A) or RC1-derived (B) DMR. The first 18 CpG within 210 base pairs are shared between the two copies of the DMR; other CpGs are specific to the alleles. DNAs came from adult or embryonic somatic tissues. Robust DNA methylation and pit-RNA expression characteristic of the paternal transgene recapitulates what was seen at the wild type endogenous locus. Because the endogenous locus failed to acquire methylation, the pit-RNA made by RC1 could impart methylation only at the RC1 DMR and not the endogenous locus, indicating pit-RNAs function in *cis* when transcribed from the repeat.

54

DNMTs. We cannot exclude to possibility that chromosomal location of RC1 limits its ability to function in *trans*.

II.5 Discussion

In mammals, two *de novo* methyltransferases, DNMT3a and DNMT3b, are responsible for establishment of 5mC throughout the genome. Targeting of methyltransferases to specific genomic regions is likely to be regulated by a variety of mechanisms. Interestingly, DNMT3 single mutants maintain proper 5mC at the *Rasgrf1* DMR, indicating that the DNMTs have redundant function there. There is reasonable evidence to support a model where histone modifications (Ooi et al., 2007; Rai et al., 2010), and/or DNA sequence context (Lienert et al., 2011), are critical for localization of 5mC. Because only a portion of genomic DNA methylation can be explain by these data, presumably undefined mechanisms exist and crosstalk between them could help to establish genome wide 5mC patterns (Denis et al., 2011).

Earlier work the at *Rasgrf1* locus determined that imprinted expression patterns are dependent on 5mC over an enhancer blocking sequence (Yoon et al., 2005). Methylation of this sequence was shown to be dependent on the presence of a proximally located repeat region, which functions during a critical window where mouse epigenetic developmental reprogramming takes place (Yoon et al., 2002; Holmes et al., 2006). The precise mechanism by which the

repeat region regulates 5mC remained unknown however. Interestingly, there

was compelling evidence from the Soloway lab and others that suggested

ncRNAs were involved in this process (Herman et al., 2003; Aravin et al., 2007b,

2008).

We define a novel function for the *Rasgrf1* repeat region, and show that

the repeats are essentially a promoter for a pitRNA, which is a precursor for

smaller piRNA. We demonstrate: 1) These piRNAs, and the PIWI pathway are

essential for establishment of 5mC to the *Rasgrf1* locus, and 2) Transcription of

the pitRNA is not only required for piRNA biogenesis, but is also for proper

targeting of 5mC to the DMR in *cis*. These findings not only redefine the model

for regulation of imprinted *Rasgrf1* expression, but also describe a previously

undocumented mechanism for targeting of DNMTs to euchromatic loci.

We propose a model where targeting of the nascent pitRNA by piRNA-

containing complexes occurs in a co-transcriptional manner, and is an important

step in the sequence specific methylation of the *Rasgrf1* DMR. This process

takes place during mouse embryonic development, when germ cells' epigenetic

marks are reprogrammed. In order to establish new parent-specific modifications

during migration of male primordial germ cells at embryonic day 11.5, 5mC is

removed. Removal of 5mC relieves genomic transposon silencing genome wide,

which presumably includes derepression of the retrotransposon present at

Rasgrf1 (RMER4B). Subsequently, nascent transcription begins at the repeat region and pitRNA expression is initiated. PIWI pathway members then process the pitRNA into a primary piRNA, which has complementarity to an RNA transcribed on chromosome 7 and located within a piRNA cluster. This produces secondary piRNAs and begins the ping-pong cycle piRNA biogenesis cycle, leading to rapid amplification of both primary and secondary piRNA species.

Although we have determined the pitRNA is required for the recruitment of DNMTs, its unclear how this process takes place. Perhaps, piRNA in mammals function similar to siRNAs in *S.pombe*, where siRNAs are able to target repressive histone modifiers, like histone 3 lysine 9 methyltransferase, to specific genomic sequences (Verdel and Moazed, 2005). If this were the case, it is conceivable that H3K9me would be deposited prior to 5mC at the *Rasgrf1* DMR, and DNMT3a or DNMT3b would be recruited in an H3K9me dependent manner. Additionally, this model suggests H3K9me3 would be specific to the paternal allele, and its abundance would decrease when the repeats are deleted in a paternal allele specific manner. Previous studies have demonstrated both of these predictions to be the case (Delaval et al., 2007; Lindroth et al., 2008). There is precedent for this mechanism in other systems, including *Arabidopsis*, Zebra fish, and at pericentric heterochromatin in mouse (Johnson et al., 2002; Lehnertz et al., 2003; Rai et al., 2010). It is also possible that members the piRNA processing machinery, like *MitoPLD* for instance, directly recruits DNMTs

during biogenesis, and the DNA methylation is established prior to deposition of histone modifications. Future research in this area will likely uncover a more detailed mechanism, and refine this model further.

We have demonstrated *cis*-mediated deposition of 5mC at the Wt *Rasgrf1* locus, however, there are exceptions. When the repeats were replaced with the R2 sequence from *Igf2* there was acquisition of methylation on the maternal allele in *trans.* These results are in direct opposition to the aforementioned model. It is possible the R2 sequence causes improper timing of expression and/or increased piRNA transcript abundance. Either of these may lead to the observed *trans* effect. Perhaps over-expression causes a high molarity of piRNAs, enabling them to outcompete the DNA sense strand, bind the complementary antisense maternal DNA stand, and establish aberrant *trans* 5mC. This hypothesis can be tested by creating mice where the DMR-spanning pitRNA is regulated by an inducible expression promoter. Controlling the temporal expression, spatial localization, and transcription level of the pitRNA in this manner may allow for proper assessment of aberrant methylation phenomena.

The experiments described focus on defining the function of the repeat region and its associated ncRNA during establishment of 5mC in gametic tissue. Interestingly, the repeats have also been shown to function after fertilization to

maintain imprinted 5mC in the developing embryo (Holmes et al., 2006). It is

possible repeat mediated pitRNA transcription plays a role during zygotic

development similar to its role during gametogenesis. Since piRNAs and PIWI

proteins are not known to function at this time, it is likely the repeats are

interacting with an additional pathway and have a third yet to be defined role.

Various pathways have been recently reported to function in preserving 5mC at

Rasgrf1 in the post fertilization embryo (Quenneville et al., 2011; Messerschmidt

et al., 2012; Nakamura et al., 2012). Two proteins identified, ZFP57 and Kap1,

function in the same molecular pathway and, similar to the repeat deletion

mutant, there is a decrease in paternal 5mC and H3K9me over the DMR when

either of the genes encoding these proteins are mutated. This genetic interaction

suggests the repeat region functions in the same pathway as ZFP57 and Kap1 to

maintain imprinted 5mC at *Rasgrf1* in the developing zygote. It would be

interesting to determine if recruitment of these factors is dependent on the repeat

region, or on its transcription. It is likely that future research will focus on defining

the direct relationship between these proteins the *Rasgrf1* repetitive region.

Because aberrant localization of 5mC has been implicated in a variety of

diseases including cancer (Doi et al., 2009; Lendvai et al., 2012), defining the

means by which DNMTs are targeted is of utmost importance. Modulating DNMT

targeting or inhibiting their activity has even been suggested as a means to treat

patients (Foulks et al., 2012). Here we made use of the highly characterized

59

mouse *Rasgrf1* locus to study the complicated mechanisms of ncRNA mediated

DNMT targeting in a controlled and well defined system, and the work presented

has helped to redefine the mechanism by which 5mC is established and targeted

to specific loci during male gametic epigenetic reprogramming. Furthermore,

these data, and our conclusions, will inspire new research and help to address

previously unresolved questions relating to mammalian RNA mediated DNA

methylation, and maintenance of 5mC in the early embryo.

III. Designing and optimizing a platform to perform single molecules epigenetic analysis

III.1 Abstract

Epigenetic mark coordination occurs on a genome wide scale in order to synchronize the expression changes that are necessary for cellular differentiation and development (Ernst et al., 2011; Ram et al., 2011). Abnormal epigenetic mark placement has been associated with both initiation and progression of cancer (Feinberg, 2007). Additionally, epigenetic state can be influenced by environmental factors that include diet (Waterland and Jirtle, 2003), environmental toxins (Anway et al., 2005), and maternal behavior (Weaver et al., 2004). Therefore, it is of great interest to identify epigenetic marks genome-wide. Developing a system that would enable rapid, quantitative measurement of epigenetic marks on a genome wide scale would prove useful for characterizing epigenomic regulatory mechanisms, for discovery of disease biomarkers, and potentially for clinical assessment of patient response. Though tools for this type of analysis exist, they have significant limitations. For example, no technique is capable of quantifying genomic epigenetic mark abundance from very few cells, and no currently available assay can survey for more than two epigenetic marks simultaneously. If a newly designed platform were capable of detecting multiple epigenetic marks from rare populations of cells, it would enable researchers to survey coordination of multiple epigenetic marks, and potentially identify novel

mechanisms in tissue types where epigenetic analysis is currently unavailable. Here we describe the development and optimization of a platform capable of doing just this. By collaborating with Harold Craighead and his research group, we were able to engineer a system that combines single molecule nanofluidics and fluorescence correlation spectrometry (FCS) with real-time automated electrical sorting. We then use our system to assay for and purify epigenetic marks on single molecules.

We demonstrate quantitative detection of both individual chromatin molecules, and in-vitro methylated DNA. We go on to sort and purify for 5mC from a mixed population of both in-vitro methylated and unmethylated DNAs. These experiments provide a proof of principle that it is possible to sort molecules based on epigenetic state, and therefore, establish groundwork for future biologically significant studies.

III.2 Introduction

III.2.1 Limitations of current techniques

Typically, ChIP and BS-Seq based techniques (Barski et al., 2007; Lister et al., 2008) are used to study epigenetic modifications. Although these methods are very powerful, and have produced a wealth of data, they are not without limitations. For example, for studies simply seeking to determine the genomic

abundance of a given epigenetic feature, both techniques are cumbersome, requiring many days for sample preparation, data collection, and analysis. Typical ChIP protocols require around 10^6 cells, making experiments that assay for epigenetic marks on chromatin from tissue that cannot easily be harvested, or from rare cell types, difficult, if not impossible. The limitations to ChIP and BS-Seq are primarily due to the incredibly low yield and high background level which may lead to misrepresentation of the true chromatin structures (Pondugula and Kladde, 2008).

Occasionally, multiple epigenetic marks are found together at a given locus, and can act combinatorially to control gene expression patterns. ChIP experiments aimed to assay for coincidence of multiple epigenetic marks traditionally rely on two independent immuno-precipitations. The co-occurrence of epigenetic marks is then inferred by overlapping the datasets. Methods to overcome this limitation exist and will be discussed further in chapter 4.

III.2.2 Single molecule approaches

Single molecule based approaches have the ability to overcome many of the limitations imposed by bulk, ensemble methods. The most obvious advantage is the ability to directly inspect each molecule rather than infer their characteristics indirectly or from population averaging. The average state may not be representative of the discrete properties for any single chromatin molecule

within the population. In one example, data have been generated using traditional macro scale experiments, which argued that centromeric nucleosomes are actually tetrameric hemisomes rather than octamers (Dalal et al., 2007a). However, others cite structural data (Tachiwana et al., 2011) to argue that tetramers did not truly exist. When studies were performed using single molecule platforms, which allowed for direct observation, the existence of centromeric hemisome structures was confirmed (Wang et al., 2008b).

In a second example, DNA repair enzyme function is typically studied *in vitro*, using either naked DNA or reconstituted chromatin. It is assumed based on the results of these studies that the repair enzyme can efficiently navigate through the nucleosome uninhibited. However, because this process had never been directly observed, some questioned how, or if, it was actually occurring *in vivo*. Recent single molecule studies have allowed for direct detection, and observation of the DNA repair enzyme as it traversed through nucleosomes, confirming earlier inferences (Gorman et al., 2010). There were similar questions with regard to RNA polymerase and its ability to navigate through nucleosomes. Again, single molecule based direct detection methods proved useful. They demonstrated that not only can the polymerase efficiently navigate the nucleosome, but data suggest trailing polymerases can assists with progression and alleviate polymerase pausing and backtracking behavior (Jin et al., 2010).

III.2.3 Nanofluidic single molecule approaches

The previously mentioned methods made observations from immobilized single molecule entities. There are additional advantages when molecules are flowed through nanofluidic channels and assayed in real-time. With the addition of real-time fluidic detection, single molecule studies allow for direct detection of each individual molecule in a high throughput manner, and promote accuracy and precision combined with high sensitivity (Moerner and Fromm, 2003). For example, the size and relative proportion of individual DNA fragments has been determined from only 76 femto grams of material (Foquet et al., 2002), the expression level of 45 different human microRNAs was determined from only femto molar concentrations of RNA (Neely et al., 2006), and the abundance of a beta-actin transcript has been quantified from as low as 100 femto moles of RNA (Nolan et al., 2003). Generally speaking, single molecule nanofluidic techniques are capable of assaying very small amounts of biologically relevant material at near physiological concentrations. Due to the very small molecular confinement volumes, we can reliably perform direct detection of single molecules at high throughput speed in a highly sensitive manner. In these assays, as a sample flows through a nanoscale channel, each molecule passes through what is called an "inspection volume." In our case, the inspection volume is the space created when a laser beam intersects with a nano-channel (**Fig III.1A+B**). The Poisson

$$C = x \cdot V \cdot N_A$$

$$m = \frac{molecules\ within}{focal\ volume}$$

$$P_c(m) = \frac{c^m e^{-c}}{m!}$$
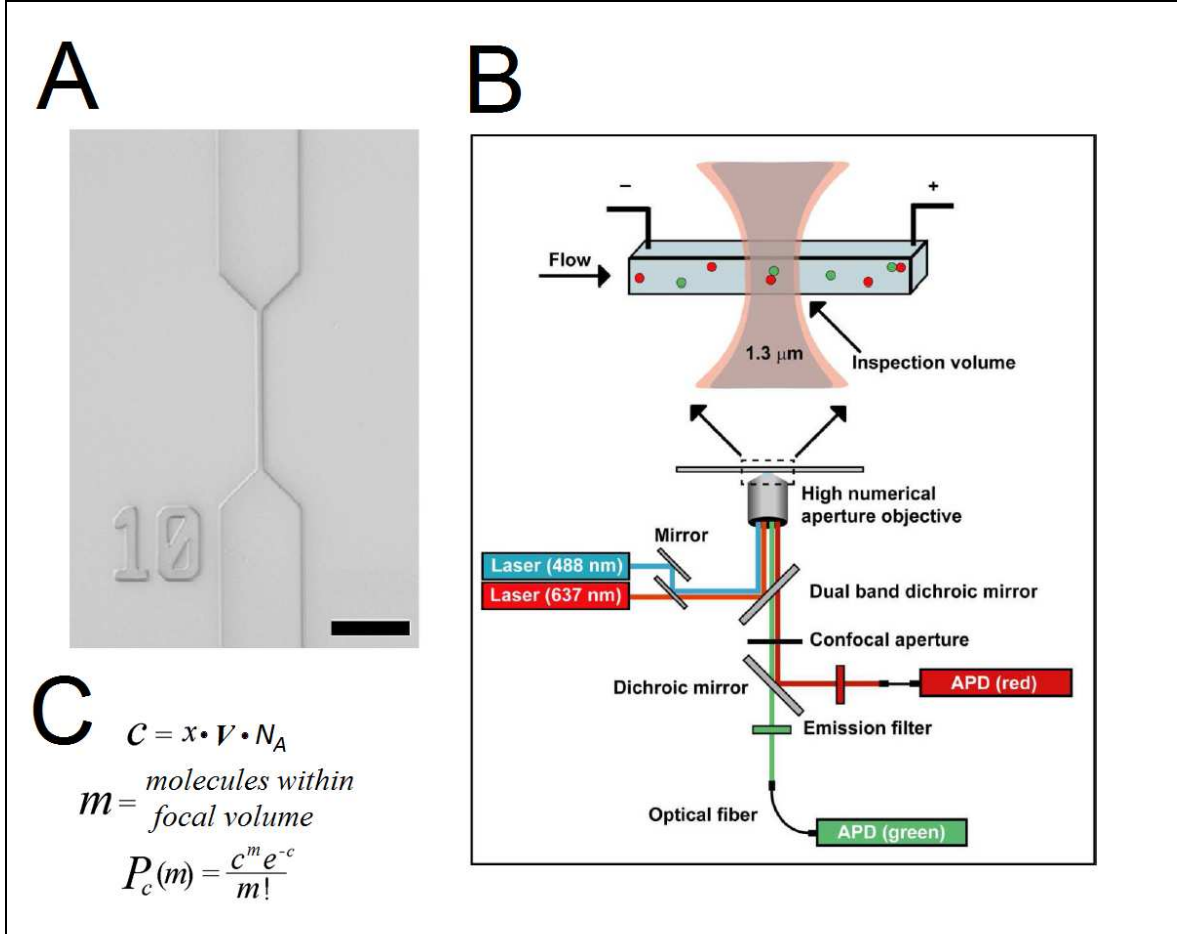
Fig III.1 - Tools for single molecule analysis.
A) A micrograph image of nanoscale channel is displayed. The inner most portion of the channel, which is used for single molecule laser inspection, has dimensions of 500nm wide, 250nm deep and 10um long. The scale bar is 10 um.
B) A diagram of a SCAN device mounted on a confocal fluorescence microscope. Two overlapped lasers illuminated a 1.3 μm length of the nanofluidic channel and formed an inspection volume of ~150attoL. Collection of fluorescent emission was achieved using a confocal apertures and avalanche photodiodes (APDs).
C) The probability a molecule resides within the inspection volume at any given time is determined by the Poisson statistic. P = probability, m = molecules within inspection volume at any given time. c = input concentration, where V = volume, x = molecules in solution, and NA = Avogadro's Number.

statistic determines the probability a molecule has of passing through the inspection volume at any given time (Fig III.C).

III.2.4 Real-time detection of ensemble molecules by FCS

   Although, as previously mentioned, single molecule studies alone are useful for analyzing a variety of biological phenomena, when combined with FCS one gains the ability to monitor the behavior of molecules in real-time rather than in post-process analysis. Simply, FCS is used to measure the fluctuations and diffusion of fluorescent single molecules (Magde et al., 1974). From FCS based detection, one can observe both the auto correlation and the cross correlation. The autocorrelation defines the flow (or the diffusion) characteristics of a single fluorophores; and the cross correlation defines the coordinated flow (diffusion) characteristics of multiple fluorophores. When a solution carries multiple fluorophores, that are behaving in a correlated manner are likely to exist together in a complex. From FCS one is able to determine: 1) estimated flow rate and consistency for each population of molecules, and 2) if different colored molecules are flowing as a correlated unit (or complex) within the larger population of molecules. However, it's important to note that FCS based detection is not single molecule based analysis and the conclusions represent characteristics of ensemble data. It is, therefore, only a tool used in parallel to single molecule studies, in order to facilitate real-time monitoring and detection. For the purpose of our research FCS has proven to be a useful tool when

operated in parallel with single molecule analysis. It enables us to monitor flow rate, discern channel clogging, detect when two different colored fluorophores are bound to one another, and observe fluorescence crossover bleed-through, which occurs when a given fluorescent emission is inappropriately ascribed to a different fluorophore emitting at a longer wavelength.

III.2.5 A novel single molecule fluorescence detection technique

Although single molecule nanofluidic based fluorescent detection platforms have been used for assaying a number of biological substrates (Foquet et al., 2002; Nolan et al., 2003; Neely et al., 2006), they have never been used for detection of chromatin and, more importantly, up until this point no one has detected epigenetic marks using these type of tools. Here we present the development of a novel technique called SCAN (Single Chromatin Analysis at the Nanoscale) that combines single molecule nanofluidics with multicolor fluorescence microscopy to detect DNA, 5mC, or histones on individual chromatin fragments. Furthermore, we go on to present an advanced version of this platform that enables real-time detection and automated sorting of individual fluorescent molecules. In proof of principle experiments, we use this new platform to sort and purify 5mC from a mixture of both methylated and unmethylated DNA samples. We envision this tool will eventually become an alternative to ChIP, and allow for detection and purification of material that currently cannot be analyzed. We also foresee a modified version of this

68

technique will be used to assess histone modifications and potentially to determine if multiple epigenetic marks are actually coincident on the same chromatin fragments through out the genome.

III.3 Materials and methods

III.3.1 Collaborator contributions

Fabrication of nanofluidic channels, fluorescence confocal microscopy, real-time photon counting, qPCR and post experimental data analysis were performed by collaborators in the Craighead and Soloway labs and are described in our collaborative publications (Cipriany et al., 2010, 2012). The methods which I performed, are described below.

III.3.2 HeLa cell culture

HeLa cells constitutively expressing green fluorescent protein (GFP) on histone H2B (H2B–GFP) were provided by Geoffrey M. Wahl at The Salk Institute for Biological Studies, USA (Kanda et al., 1998). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal calf serum, and were passaged when 80% confluence was reached.

III.3.3 Chromatin preparation

Cells were harvested from two 15cm dishes, when a density of 1×10^6

cells/ml was reached. Harvesting was by gentle scraping, followed by washing in

5mL RSB buffer (10 mM Tris pH 7.6, 15 mM NaCl, 1.5 mM $MgCl_2$). Pellets were

then resuspended in 5 mL RSB buffer containing 1% Triton-X 100 and

homogenized with eight strokes of a loose fitting pestle in a 7mL Dounce

homogenizer. Samples were then centrifuged at 4000g (units of gravitational

force) for 5min at room temperature (RT) in a swing bucket rotor. The nuclei

containing pellet was resuspended in 1.5 ml Buffer A (15 mM Tris pH 7.6, 15 mM

NaCl, 60 mM KCl, 0.34 M Sucrose, 0.5 mM Spermidine, 0.15 mM Spermine,

0.25 mM PMSF, 0.1 M $CaCl_2$, and 0.1% β-mercaptoethanol) and micrococcal

nuclease (MNase) were added at a concentration of 266 gel-units/mL of buffer.

At various time points (ranging from 0 to 90 minutes) 250uL aliquots were taken

and reactions were stopped using 5 μL of 0.5 M EDTA. Digests were centrifuged

at 10,000g and pellets were suspended in 450 μL 10 mM EDTA, 50 μL 5 M NaCl

to solubilize native chromatin. To confirm size ranges and chromatin integrity,

DNA was extracted and analyzed by agarose gel electrophoresis. Concentration

of extracted DNA, was measured by Nanodrop, and used as a proxy to estimate

chromatin concentration. Typically concentration was 150-600ng/ul.

III.3.4 Methyl binding domain (MBD) protein synthesis and labeling

Plasmid for bacterial expression of 1xMBD (pET31b) was provided by

Adrian P. Bird at The Wellcome Trust Centre for Cell Biology, University of

Edinburgh, UK (Jørgensen et al., 2006). Recombinant His-tagged MBD was

purified, from 200 mL of IPTG induced BL21(DE3) bacterial cultures, on Ni-NTA

agarose (Qiagen) using denaturation and on-column refolding. Briefly, cells were

grown to log phase and inducted with 1mL of 1 Molar IPTG, which took place

when optical density reached 0.5. After 3 hours, cultures were pelleted and lysed

in 40mL denaturing buffer (6 M Guanidinium HCl, 0.3 M NaCl, 0.1 M Sodium

Phosphate, 0.01 M Tris-HCl, 10 mM 2-ME, 10 mM imidazole) and sonicated

using a microtip with 12 cycles of 10 seconds on and 10 seconds off at mid level

power output. Lysed culture was then centrifuged at 10,000g. The soluble

fraction was added to 20mL of 50% nickel-NTA slurry, and allowed to bind over

night at 4C on a rotator. The next day the entire sample was loaded into an

empty column, and washed 12 times with 5mL of various buffers that gradually

transitioned from denaturing to native buffer conditions (0.3 M NaCl, 0.1 M

Sodium Phosphate, 0.01 M Tris-HCl, 10 mM 2-ME, 10 mM imidazole). Once

refolded, MBD was eluted using 5mL PBS containing 250 mM imidazole. The

sample was then dialyzed into PBS to remove imidazole and an equal volume of

100% glycerol was added prior to storage at -20 Celsius. Following purification

the protein was labeled with Alexa-488 (green) using Invitrogen's Microscale

Labeling kit (A30006), which labels free amines. Three additional rounds of resin

purification (included within kit) were performed to better purify away free label.

Absorbance measurements indicated an average degree of labeling of 1.5 dye

per MBD. Southwestern blotting was used to confirm MBD activity. In subsequent

experiments, instead of resin based purification, serial filter based size exclusion

spin columns were used in an attempt to better eliminate free dye (Ambicon UFC501008).

III.3.5 Southwestern blotting

Nitrocellulose membranes were spotted with various quantities (ranging from 5ng to 500ng) of either methylated or unmethylated Lambda DNA. Prior to probing with MBD, membranes were blocked with 10mL of tris-buffered saline (TBS) containing 1% BSA and 0.1% triton X- 100 (TBST-B). The blots were then submerged in 3mL of TBST-B solution containing 12nM concentrations of Alexa-488 labeled MBD. After overnight incubation on a rocker, samples were washed 4 times for 15 minutes each with 10mL of TBS containing 0.1% triton X- 100. They were then imaged using a Typhoon Fluorescent scanner.

III.3.6 Intercalator labeling

All DNAs analyzed were labeled with TOTO-3 DNA intercalator (red) (Invitrogen). The labeling reaction was conducted by mixing DNA and the dye at a 1:5 dye to base pair ratio. Based on the DNA mass within a given sample the appropriate amount of intercalator dye stock was added directly to each sample and immediately vortexed. After mixing, samples were protected from light and incubated at RT for 1 hour. Following 1 hour RT incubation samples were moved to 4C and left for at least on additional hour. Oftentimes samples remained at 4C overnight, prior to analysis. There was a significant increase in fluorescence

intensity following intercalator binding, eliminating the need to purify samples further.

III.3.7 pML4.2, pUC19, and Lambda preparation and *in-vitro* methylation

Lambda DNA was purchased from Promega, and had been grown in a methylation deficient host (D1521). pUC19 and pML4.2 (Herman et al., 2003) were grown in dam-/dcm- Escherichia coli (New England Biolabs—C2925) and purified using a QIAGEN Plasmid Midi Kit (12143). Lambda and pML4.2 were *in vitro* methylated with *Sss*I methyltransferase. Efficiency of methylation was assessed by resistance to digestion using methylation sensitive enzymes *Hpa*II, and *Hha*I.

III.3.8 MBD-DNA affinity binding reaction

1.9uM MBD was mixed with 75nM TOTO-3 labeled DNA and incubated in 20uL total binding buffer for two hours at RT with moderate agitation, then stored at 4 ℃ in the dark. The binding buffer consisted o f TBS at pH 8.0 in 0.5% BSA and 0.1% triton X- 100. The following morning all biological samples were diluted into flow buffer (10 mM Tris, and 1 mM EDTA, 0.1% Triton X-100, 0.3% polyvinylpyrrolidone, buffered to pH 8.0) such that total concentration of molecules was below 1nM.

III.4 Results:

III.4 Nanofluidic system

The nanofluidic fluorescence based detection system used is similar to one developed previously to size and elongate DNA molecule (Foquet et al., 2002; Levy and Craighead, 2010). The nanofluidic channels were 500 nano Meters (nm) wide and 250nm deep. The length varied depending on whether we were sorting, or simply counting molecules (**Fig III.1A**). Channels were fabricated by our collaborators in fused silica substrates using photolithography and reactive ion etching. Two 10 centimeter (cm) silica discs were bonded, where channels were etched in the bottom disc, and a second disc was placed over the first to form a roof for the channels. Prior to bonding the two, the second disc was sandblasted to provide access ports to fill the channels. Devices were designed in arrays of 16 parallel channels, with each array having a single input and a single output reservoir on each end. These reservoirs served as sample loading ports, and where were the electrodes made contact with the fluid sample. The nanofluidic device was mounted on a confocal microscope and a series of mirrors and fiber optics were used to direct light from a series of lasers through the objective lens and onto the stage. When the laser beams intersect the nanoscale channel, it creates a 150 attoL inspection volume (**Fig III.1B**). When a fluorescently labeled molecule is present within the inspection volume the laser is able to excite the fluorophore, and a longer wavelength light is emitted. Emitted

light is then directed back through the objective and is detected using avalanche photo diodes (**Fig III.1B**). Due to the 150attoL inspection volume created by the laser intersection, the Poisson statistic indicates that if molecules are below 1nM concentration, there is greater than a 99.9% probability that one or fewer molecules are being inspected at any given time. In other words, these parameters establish the limits we must stay within for each experiment to generate meaningful data, and to ensure that we are working at single molecule resolution.

III.4.2 Detection of intact native chromatin

For our first proof of principle experiments, we needed to determine if chromatin could remain intact as it passed through the nanoscale channels. To do this, we chose to use chromatin from HeLa cells that expressed a transgenic H2B-GFP fusion protein. These cells were chosen specifically to extract chromatin from because no post extraction histone labeling needed to be performed in order for them to be fluorescent. As a second means for detection, we chose to use the DNA intercalator TOTO-3 (red), which has excitation and emission profiles that are spectrally distinct from GFP. Both FCS and post-process color correlation analysis (which will be discussed later) rely on two color fluorescence. Using two distinct flours (TOTO-3 is red and GFP is green) allowed us to eliminate confounding variables like FRET (Föster's Resonance Energy Transfer), which occurs when emission energy from one fluorophore can excite

neighboring fluorophores. FRET effects can cause quenching of emission from the first fluorophore. Bleed-through occurs when the first fluorophore has emission energy that is so intense that it is mistakenly detected as the second, longer wavelength, fluorophore (Andrews, 1989; Periasamy, 2001). The two color scheme we designed allowed us to determine if DNA and histones could be detected, and if chromatin remained intact as molecules flowed though the inspection volume.

After extracting native chromatin from the GFP expressing HeLa cells, we confirmed chromatin was intact prior to being loaded into the input port by purifying DNA from a small portion of extracted chromatin and running it on an agarose gel. Following MNase digestion, if chromatin extraction was successful and had remained intact, we anticipate that extracted DNA would form a 160bp ladder (**Fig III.2A**). Next TOTO-3 (red) was bound to our chromatin and samples were loaded into the nanoscale channels. After a voltage was applied, fluorescence emissions intensity and time of occurrence were collected using a single molecule photon counter contained within our FCS system (**Fig III.2B**). FCS allowed us to monitor flow rate, aggregation, and clogging of channels in real-time. For both the red TOTO-3 molecules, and for the green GFP molecules, no clogging or aggregation was observed. As molecules entered the inspection volume their time of entry was recorded, and if the time between two different color fluorescent bursts was below a single molecule's anticipated residence time
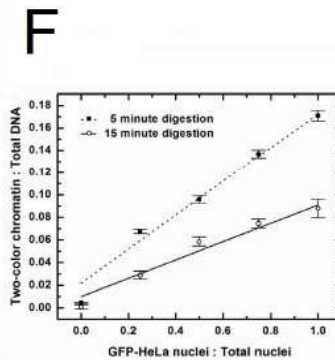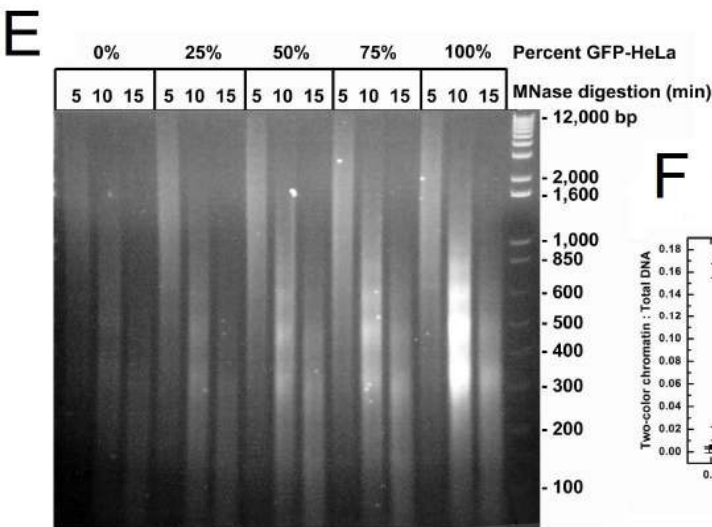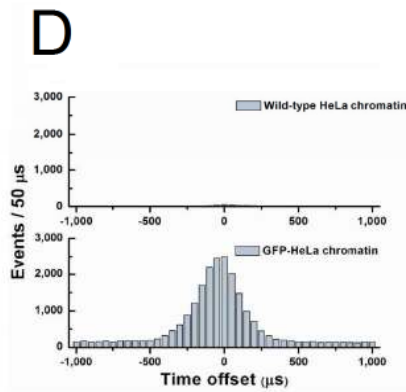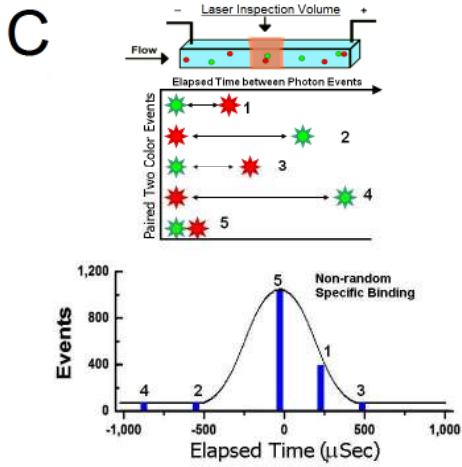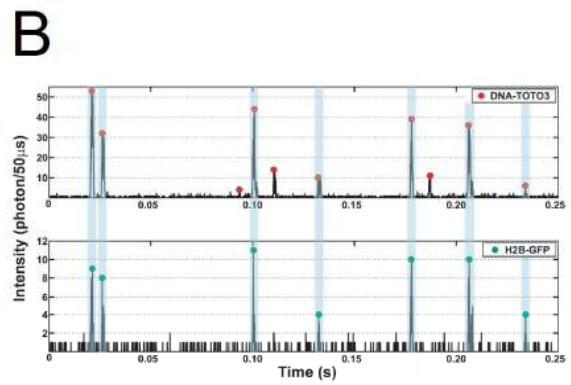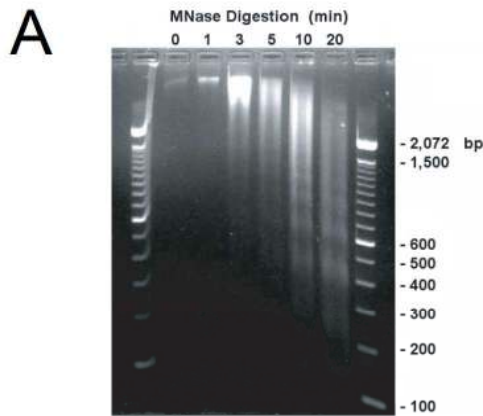
Fig III.2- Chromatin remains intact as it flows through nanoscale channels
A) Chromatin was extracted from GFP containing HeLa nuclei and digested using MNase. With increased time of MNase treatment, we observed a decrease in chromatin fragment sizes, as verified by gel electrophoresis. 160bp laddering patterns are an indication of chromatin intactness.
B) Process of Single Molecule Detection (SMD) and Two-Color Coincidence Analysis. Time-trace record of photon bursts observed by each APD, showing 0.25 seconds of a 15 minute nanofluidic SCAN. A burst with a sum of 10 or more photons satisfied a threshold condition and was designated a SMD event, shown here by a red or green marker identifying DNA and histone H2B respectively. Intact chromatin fragments, highlighted in blue, were identified by time-coincident detection of both a red and green event.
C) Coincident Detection Illustration. As molecules passed through the inspection volume (top) the elapsed time between adjacent color events was precisely measured (middle). If binding between antibody and chromatin occurred, the elapsed time approached zero. Binding events (bottom) appear as a central peak centered at time zero on the frequency plot.
D) A TCH (time coincidence histogram) illustrates the absence of coincident two-color SMD events when analyzing chromatin from wild-type HeLa nuclei. With GFP-HeLa chromatin, a central Gaussian peak, corresponding to intact chromatin molecules emitting two fluorescent colors, emerged from a background of uncorrelated events.
E) Chromatin was extracted from both GFP containing, and Wt HeLa nuclei and digested using MNase (similar to A). In some cases GFP containing nuclei were mixed with Wt HeLa nuclei prior to digestion. The portion of GFP nuclei is indicated.
F) The proportion of two-color chromatin molecules increased in direct proportion with GFP-HeLa nuclei content, as described by a linear fit with R-squared =0.98 and R-squared =0.95 for the 5 and 15 minute digestion assays, respectively. Error bars represent the error from both the bound and unbound molecules.

within the inspection volume, the event was called as a single two color "coincident event". This initial characterization is somewhat arbitrary and true coincident burst detection requires post analysis processing. During post-processing, the time between different colored bursts for all coincident events is measured (elapsed time) and plotted as a histogram (**Fig III.2C**). We anticipate that with intact chromatin, there would be an enrichment of molecules with an elapsed time approaching zero. As a control, we repeated the experiment using chromatin extracted from HeLa cells that did not express GFP (Wt HeLa). We expected there to be no coincidence in this sample. We observed precisely what we expected for both the GFP expressing and the Wt HeLa chromatin samples (**Fig III.2D**), indicating that as molecules flow through the nanoscale inspection volume they remain as intact chromatin. To confirm the level of coincident detection corresponded to the abundance of GFP chromatin, we mixed Wt HeLa nuclei with GFP expressing HeLa nuclei and prepared chromatin (**Fig III.2E**). For two different MNase digest times the level of coincidence was proportional to the input fraction of GFP containing nuclei (**Fig III.2F**), which not only confirmed our chromatin was intact, but also indicated the coincident detection level can be used for relative quantification of chromatin.

III.4.3 Detection of 5mC on Lambda DNA

Our next goal for SCAN was to detect true epigenetic marks. DNA methylation was chosen as the first epigenetic mark that would be queried. To
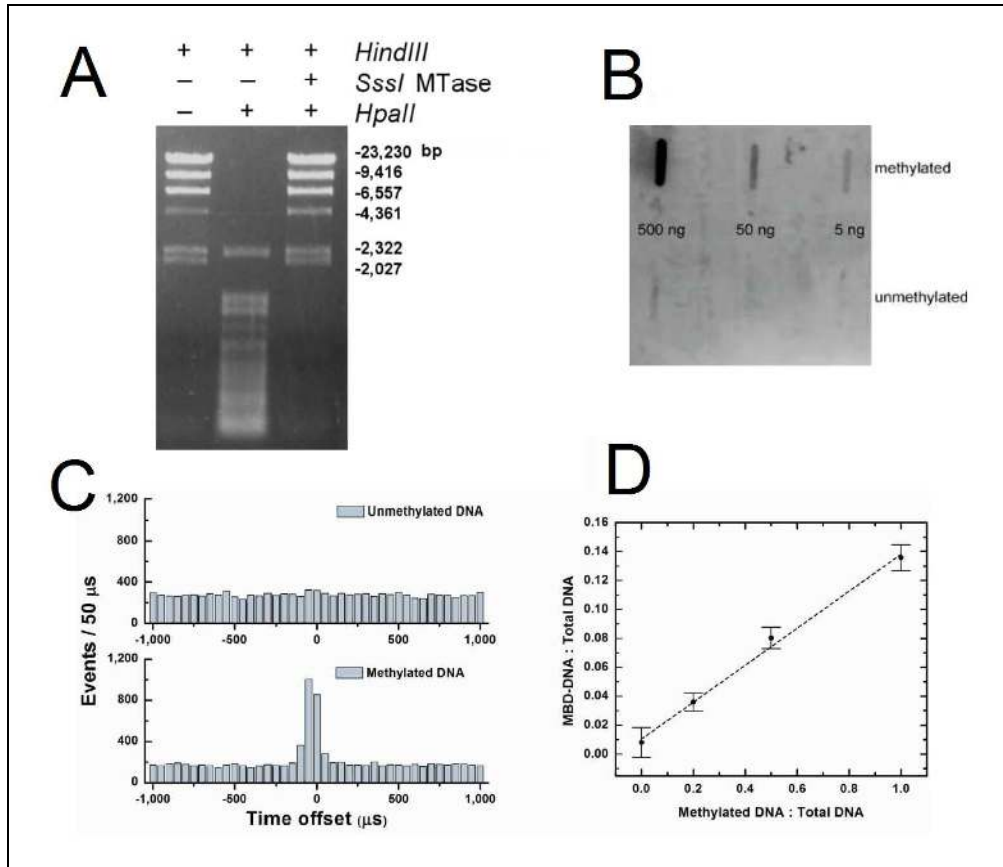
Fig III.3 - Detection of methylated DNA using MBD
A) Lambda DNA was either methylated using *Sss*I or left unmethylated. As a test
   for efficacy of the methylation reaction, aliquots of DNAs were then digested
   with *Hind*III or with both *Hind*III and the methylation sensitive enzyme *Hpa*II.
   Resistance to digestion by *Hpa*II is evidence for DNA methylation.
B) Southwestern blot analysis of MBD-1 specificity binding to methylated DNA.
   Both unmethylated and in-vitro methylated lambda DNA were bound to a
   nitrocellulose membrane in varying quantities using a slot blotting apparatus.
   The Alexa 488 labeled MBD protein was then used to probe the entire blot.
   Following over night incubation at 4°C the blot wa s washed and scanned with
   Typhoon imager to detect the Alexa Fluor 488 label.
C) Detection of DNA methylation using SCAN. Unmethylated (top) and
   methylated (bottom) DNA samples labeled with TOTO-3 were both incubated
   with a molar excess of MBD-488 protien then analyzed for 15 minutes using
   SCAN. MBD-1 binds to only the methylated sample.
D) Mixtures of methylated and unmethylated DNA were assayed using SCAN (as
   in C). MBD binding frequency (y-axis) increases as the proportion of the
   methylated DNA in the mixture increases (x-axis). Linear fit with R-squared
   =0.99. Error bars represent the error from both the bound and unbound
   molecules.

optimize conditions amenable to 5mC detection, we needed to first generate a

substrate for detection, Lambda DNA was purchased and *in vitro* methylated

using *Sss*I methyltransferase. This enzyme is able to methylate efficiently all CG

dinucleotide pairs, of which 3,113 exist on the Lambda sequence. To confirm the

methylation level of Lambda, 5mC sensitive restriction digests were performed

 (**Fig III.3A**). Next, to generate a reagent that was able to bind specifically to

methylated DNA, we chose to use a recombinant methyl binding domain (MBD)

protein. The histidine tagged MBD was purified from *E.coli* using IPTG inducible

expression, and nickel-NTA based purification. The refolded protein was labeled

using AlexaAlexa-488 (green) dye and the activity of the reagent was tested

using a southwestern blotting technique (**Fig III.3B**). In this assay both

methylated and unmethylated DNA was immobilized on a membrane at various

concentrations. The entire blot was then incubated in the presence of the labeled

MBD protein. Following incubation and a series of washes, to eliminate

nonspecific signal, the blots were imaged using a fluorescent scanner. Using this

assay we observed specific binding of the MBD only to the methylated samples,

and there was no binding to unmethylated samples (**Fig III.3B**). Once we had

confirmed the MBD's activity, we used it in combination with either methylated or

unmethyated Lambda, to test our SCAN device. We anticipated that we would

see two color coincidence only in the sample that was methylated, and there

would be no coincidence in the unmethylated sample. We observed precisely

what was expected (**Fig III.3C**). Additionally, similar to the GFP HeLa samples,

when we added increasing concentrations of methylated Lambda to an unmethylated sample, there was a direct increase in the level of coincidence detected, indicating that in a binding reaction SCAN can be used to quantify the relative abundance of 5mC (**Fig III.3C**). In the next chapter I will discuss the application of these methods to more complex systems using chromatin instead of naked DNA. I will demonstrate detection and quantification of 5mC and histone modifications, and perform measurements using chromatin from genetically modified or pharmacologically treated cells.

III.4.4 Devising a method for sorting

Eventually we plan to use the material we detect in our single molecule platform for sequencing, and map the genomic location of various epigenetic marks. In order to do this, it is vital that we devise a means to sort molecules based on their ability to bind a fluorescent probe that is specific for a given epigenetic mark. To engineer a sorting device, we modified the nanofluidic device by adding a bifurcation, which created two outflow channels (**Fig III.4A**). By applying a voltage switch molecules could be sorted into one or the other branch of the bifurcation. Furthermore, to confirm molecules were being sorted correctly, instead of only a single inspection volume, two additional inspection volumes were added. This was done by creating an elongated laser beam that stretched across all three of the channel regions, including the input channel, and both output channels (**Fig III 4.B**). This design allows for coincidence to be
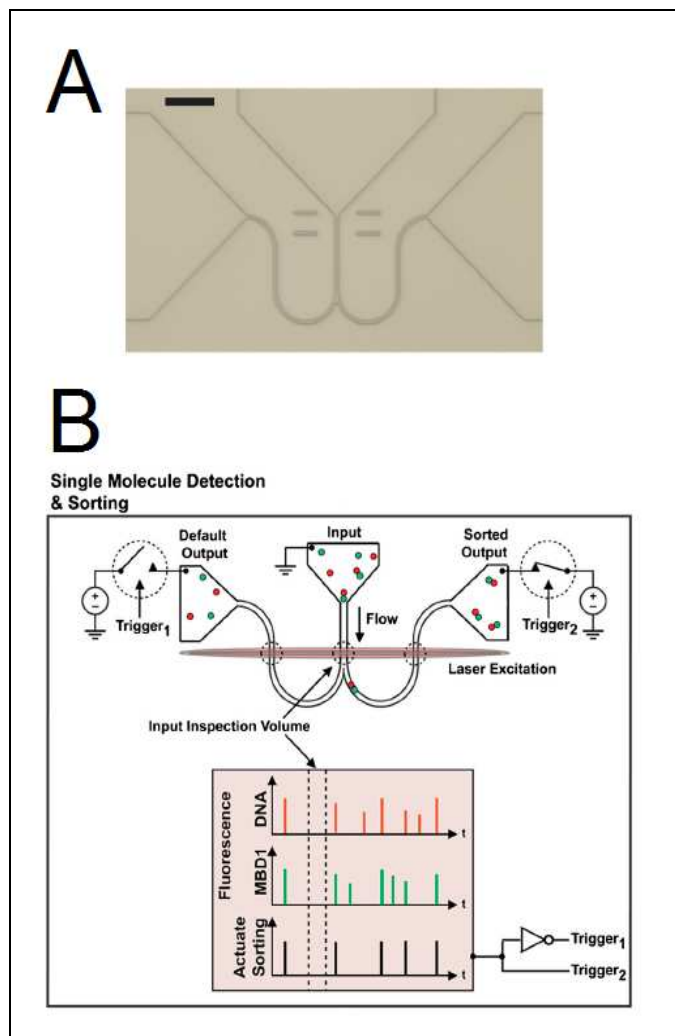
Fig III.4 - Tools for purifying epigenetically modified DNA
A) A bright field photomicrograph of a bifurcated nanofluidic sorting channel with cross-section 500 nm wide by 250 nm deep. The scale bar is 10 µm.
B) (Top) Single molecule detection and sorting schematic. Samples were loaded into the input of a bifurcated nanofluidic device. An applied voltage flowed molecules through the device. As each fluorescently labeled molecule passed through the input inspection volume its fluorescence was detected and then evaluated in real time. In this example, a green MBD protein is bound to methylated DNA, and identified by its two-color coincident fluorescence. These type of molecules were then directed to the sorted output channel. The sorted and default channel inspection volumes, which were used to confirm sorting, are also illustrated.(Bottom) An example of time resolved fluorescent detection and data collection is illustrated. The two color coincident fluorescent molecules cause an actuated sort trigger and a pair of opposing switches to direct the molecule toward the sorted output.

assessed at three times, first in the input channel prior to sorting, then again

subsequent to sorting, in both the sorted and default channels. The platform

design facilitates the ability to confirm that molecules have been sorted correctly.

In order for molecules to be sorted properly, coincidence analysis needs to be

performed in real-time, and an automated electrical switch needs to be applied to

modulate voltage and sort those molecules that are called as coincident (**Fig

III.4B**). All of these modifications were made by our collaborators.

III.4.5 Sorting methylated DNA

With the newly fabricated device, we designed an experiment to test the

performance of the system. Our goal was to sort methylated DNA. Because

intercalator dye fluorescence intensity directly corresponds to DNA length, we

thought it useful to perform experiments using two different sized species of

plasmid DNA. This choice was made specifically to allow for an orthogonal

means to confirm sorting accuracy in real-time. In other words, by choosing a

large methylated fragment and a small unmethylated fragment, which bind

different amounts of intercalator and hence have different intensities of

fluorescence emission, we could predict which should be sorted based only on

intensity. We chose to use pUC19 (2.7kb) and pML4.2 (15.2kb). To test the

integrity of the automated sorting device, we mixed our two plasmids, loaded

them into the input port, applied a voltage, and asked the automated program to

sort based only on fluorescent intensity (**Fig III.5A+B**). The voltage switch was

84
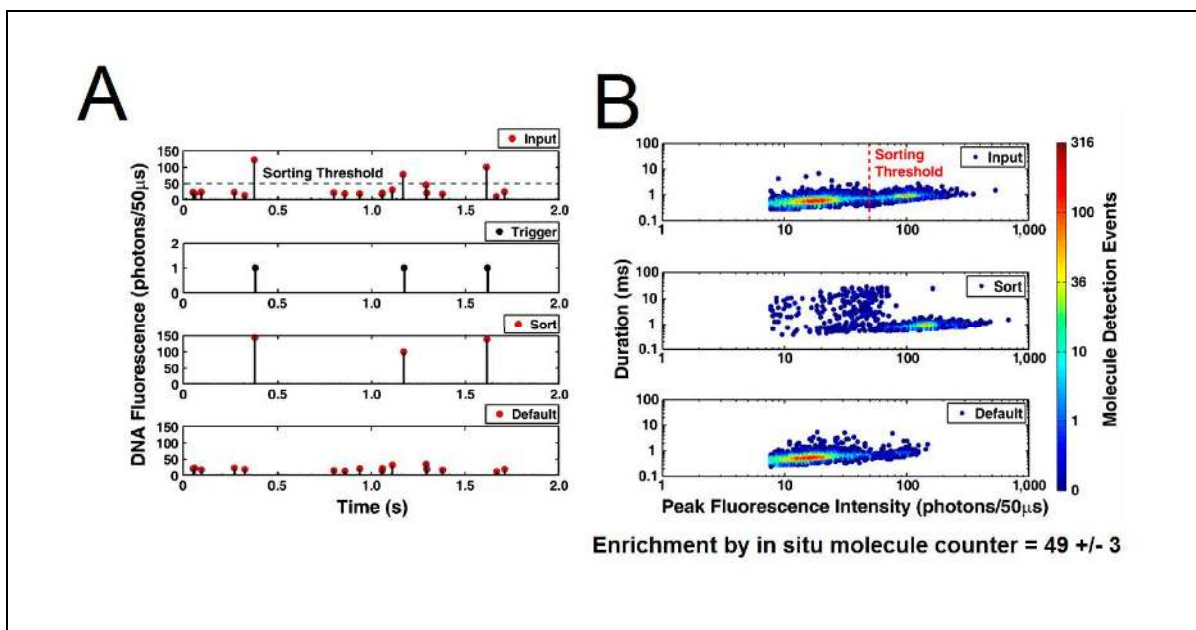
Fig III.5 – Sorting based on fluorescence intensity
A) A time-resolved record of single molecule events. Fluorescence intensity was monitored over a 2 second window of time (top). As a molecule passes though the input laser inspection volume its time of entry (x-axis) and fluorescence intensity (y-axis) was recorded. When a given molecule met the intensity threshold requirement a pulse was initiated to trigger sorting (second from top). Once sorted, molecules were monitored again in both the sorted channel and in the default channel. In this example three bright molecules were correctly sorted.
B) Following a given sorting experiment, where a mixture of 15,848 molecules were sorted, fluorescence intensity and occupancy time duration were measured as each molecule passed through the input (top), sort (middle) and default (bottom) laser inspection volumes. The molecules were plotted in accordance with their intensity (x-axis) and time duration (y-axis). The degree of event overlap is illustrated using a heat map (right).

85

applied each time a bright molecule (with an intensity above a predetermined threshold) passed through the laser inspection volume (**FigIII.5A**). By measuring the fluorescence intensity of each molecule, we observed two sub-populations (**Fig III.5B**). The first population, those molecules that were bright, represents the pML4.2 molecules, and the second, those that were dim, represent the pUC19 molecules. Fluorescence intensity measurements were also made in both the output ports. After counting molecules and classifying them as either bright or dim, we observed a 49-fold enrichment for pML4.2 in the sorted channel (**Fig III.5B**). Next, in order to sort for methylation, the longer of the two plasmids (pML4.2) was *in vitro* methylated in a way similar to what had been previously done for Lambda DNA. We confirmed the methylation level of both pUC19 and pML4.2 using 5mC sensitive restriction digest (**Fig III.6A+B**). Because in previous experiments MBD binding had a high level of background false coincidence (**Fig III.3C**), we suspected that free Alexa-488 (green) dye remained in our sample. To eliminate excess free dye, we performed serial purification in size exclusion filter-based spin columns, instead of resin based purification methods used in the past. Because this required new purification, and new protein labeling, we needed to re-confirm the MBD's activity. Similar to what was described previously, southwestern blotting was performed, and results indicated that the new MBD was able to specifically bind methylated DNA (**Fig III.6C**). Next MBD was bound to methylated pML4.2, and flowed into the nanoscale sorting devices. In order to calibrate the
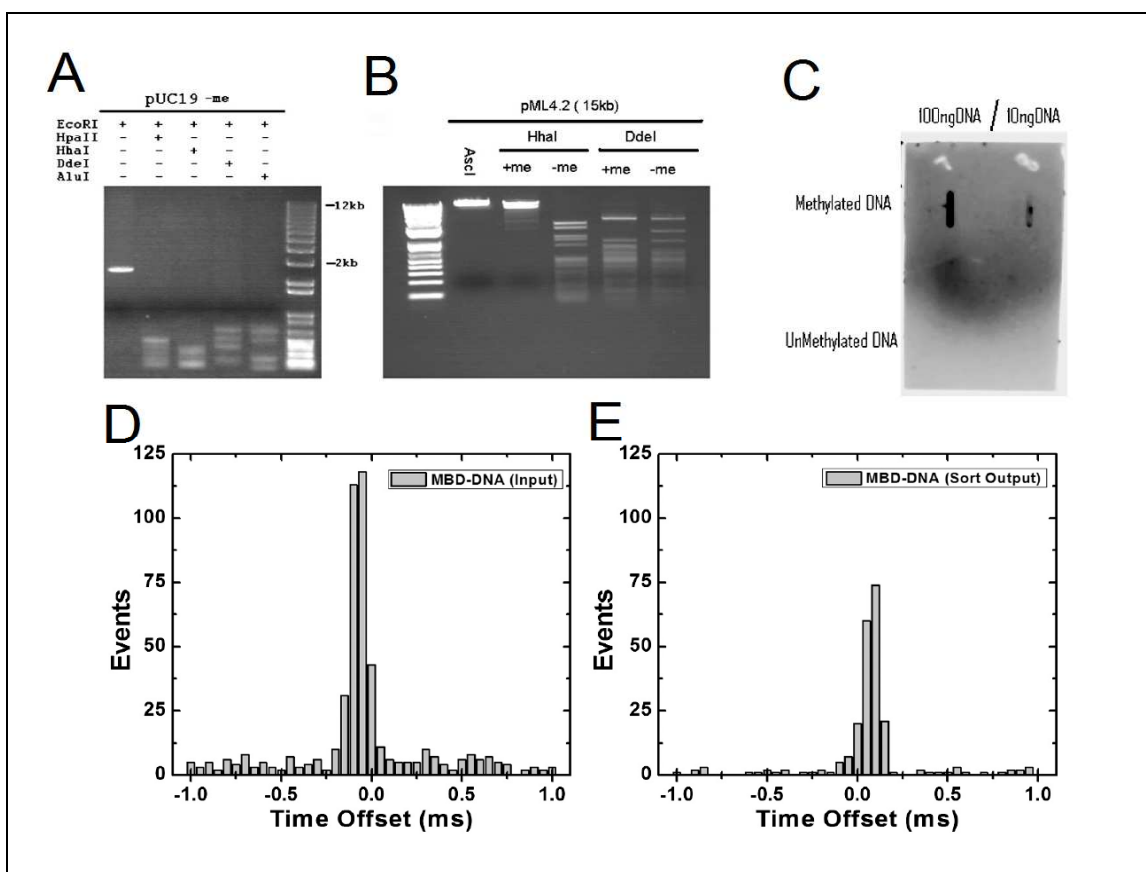
Fig III.6 - Detection of methylated DNA using new MBD and sorting device
A + B) pUC19 (A) and pML4.2 (B) were linearized with *Eco*RI and *Asc*I,
respectively, and methylation-sensitive restriction enzymes, *Hha*I and *Hpa*II,
were used to confirm proper *in vitro* methylation of pML4.2 and the
unmethylated state of pUC19. Methylation insensitive enzymes *Dde*I and *Alu*I
were used as controls to cut pUC19.
C) Southwestern blotting was preformed similar to fig III.3B except with the newly
purified and labeled MBD.
D + E) Coincidence analysis (TCH) of MBD bound pML4.2 molecules at two
separate places within sorting channel. Analysis was performed first in the
input inspection volume (D) then again in the sorted output channel (E) to
confirm proper sorting.

automated sorter, and determine the occupancy time of bound molecules, coincident time offset analysis was performed. This analysis was performed in both the input inspection volume, and in the sorted output channel. Results indicated there was abundant binding of the MBD to the methylated DNA, with an occupancy time less than one millisecond (ms) and background false coincidence levels lower than were previously observed (**Fig III.6D+E**), indicating the new experiment yielded a higher sensitivity.

Once the system was fully calibrated, we began to test its ability to sort for methylated DNA. For this experiment we mixed the methylated pML4.2 with the unmethylated pUC19 plasmid. The mixed sample was then bound to MBD overnight and the following morning TOTO-3 (red) DNA intercalator was added. The sample was then loaded into the sorting device, and the automated electrical voltage switching system was applied. When molecules passed through the input laser inspection volume red fluorescence intensity, which reported the presence of DNA molecules and their length, was recorded. We observed two populations of molecules, one having 5-fold higher fluorescence intensity than the other (**Fig III.7A**). The 5-fold difference corresponds to the size difference between pUC19 and pML4.2. Since pML4.2 was methylated, and pUC19 was unmethylated, we anticipated that if the automated sorting device was working properly, MBD would be bound and coincident only with the bright molecules, and those bright molecules, if properly sorted, would be detected in the sorted output channel. We

A

B

C

	Data analyzed by qPCR			Data analyzed by in situ molecule counter		
	Input Channel *	Sorted Channel *	Enrichment	Input Channel *	Sorted Channel *	Enrichment
Experiment 1	0.036 +/- 0.002	0.41 +/- 0.09	11 +/- 3	0.085 +/- 0.002	7.5 +/- 0.1	88 +/- 5
Experiment 2	0.48 +/- 0.02	1.5 +/- 0.2	3.1 +/- 0.5	0.27 +/- 0.01	13 +/- 1	49 +/- 3

* = Ratio of pML4.2 to pUC19
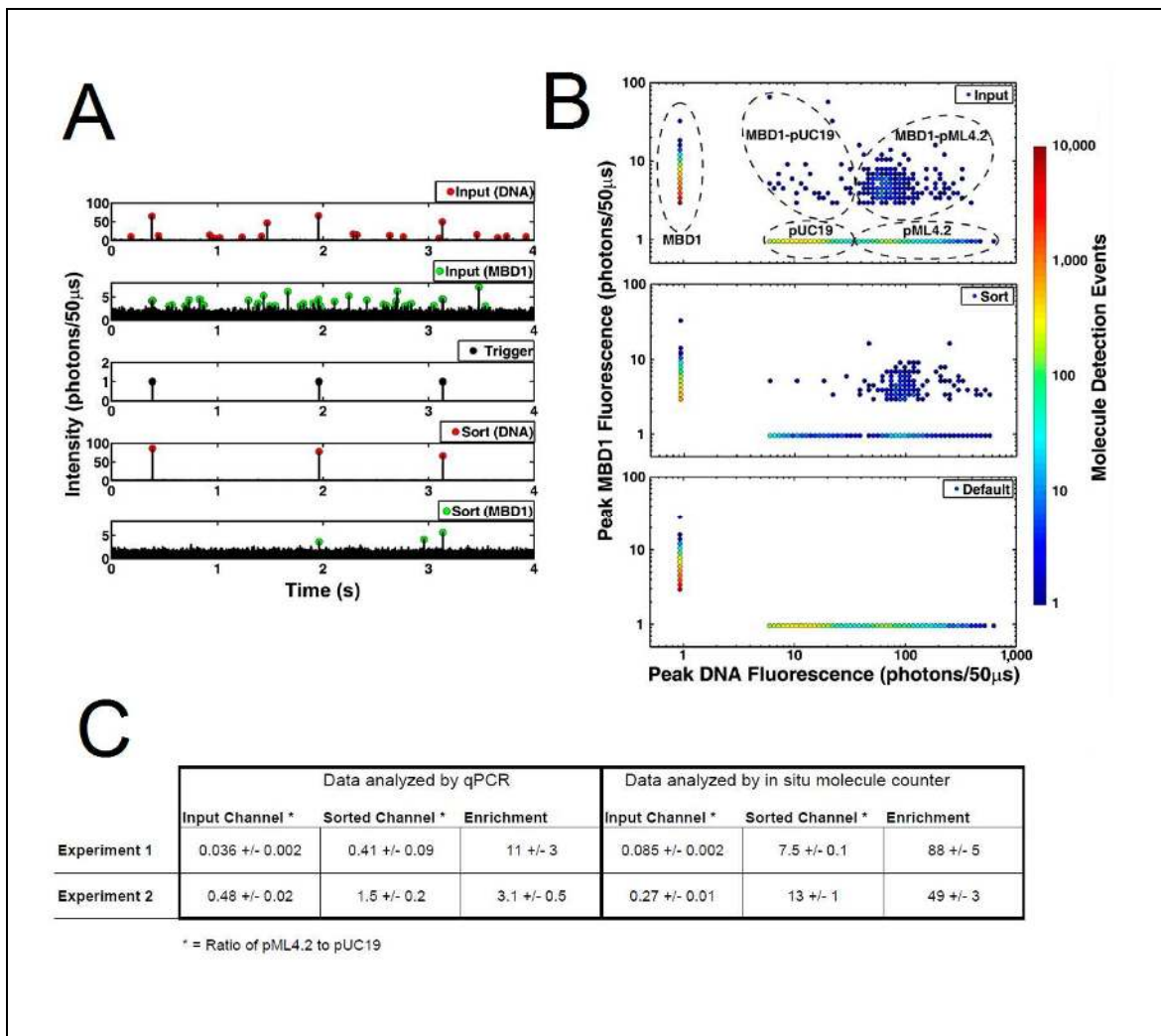
Fig III.7 - Sorting methylated DNA
A) Four seconds of time-resolved fluorescent detection are shown for both red colored events and green colored events. Three two color coincident molecules were identified, which triggered an actuated sorting event. After sorting, the fluorescence from two of the three MBD molecules remained paired with DNA, while the third had fluctuated below the intensity threshold.
B) During a sorting experiment 5,723 DNA molecules were analyzed. Fluorescence intensity data, for both the red color (x-axis) and the green color (y-axis), were collected in the input, sort, and default channels.
C) Following sorting, samples were removed from the reservoir and DNA levels were quantified using qPCR. For two different automated sorting experiments there was between 11 and 3.1 fold enrichment and a standard error of 3 and 0.5 respectively.

observed precisely what was expected (**Fig III.7A**). Following 20 minutes of continuous automated sorting, data were analyzed to characterize the fluorescence intensity of all molecules surveyed. By measuring the two color fluorescence intensity of each molecule, we observed five sub-populations (**Fig III.7B**). The first population, those molecules that were only green, represented the unbound MBD, and those that were only red represented the unbound DNA. Among the molecules that were red, there were dim red molecules, which have fluorescent properties of pUC19, and there were bright red molecules, which had fluorescent properties of pML4.2. Those that were both red and green were the MBD bound DNA molecules, included both pML4.2 and pUC19. In the sorted output channel we anticipated the unmethylated pUC19 molecules would be purified away and less abundant. Again, our observations were precisely as anticipated. (**Fig III.7B**). MBD-DNA complexes include 27 MBD-pUC19 and 270 MBD-pML4.2. *In situ* molecule tracking demonstrated significant enrichment of methylated pML4.2 at the sorted output, which demonstrated specific binding of MBD and enrichment of methylated DNA (**Fig III.7C**). From these data the false positive rate was determined to be 5.6%. To confirm we were actually purifying the methylated DNA, samples were removed from the input, sort, and default reservoirs, and were analyzed using qPCR to quantify the abundance of each plasmid. From this analysis, after two separate experiments, we observed between 3 and 11-fold enrichment of methylated DNA post sorting (**Fig III.7C**),

which confirmed the sorting device was working properly and enriching for the molecules of interest.

III.5 Discussion

We have described the development of two nanofluidic devices that can be used for epigenetic analysis. One, called SCAN, was used to analyze individual fluorescently labeled chromatin molecules and to characterize methylation on DNA fragments. The other, a nano-sorter, was used to detect and purify methylated DNA in an automated fashion. Using the SCAN device, we demonstrated that chromatin can passage through a nanoscale channel based on electrokinetic flow, and remain intact. By effectively diluting wild-type HeLa chromatin into GFP containing HeLa chromatin, we were able to confirm the authenticity of coincidence detection. We also show that 5mC can be reliably detected using the SCAN device, and sorted for using the automated nano-sorter. Using the nanoscale sorting device, we demonstrated that methylated DNA can be detected, and purified, and samples can be recovered for down stream analysis.

We plan to use this device to sort for epigenetic marks on native chromatin samples. We also envision sequencing the DNA that has been sorted, and mapping epigenetic marks to the genome. Although in principle this process is quite possible, with our current throughput and recovery method, it is wholly

impractical. Throughput may be increased in several ways. For example, by removing the post sort laser inspection volumes, we can increase the concentration of molecules by 20-fold and still remain under single molecule analytical conditions. Molecules that are part of bound complexes make up the very small minority of all molecules that pass through the laser inspection volume. Essentially, this means there is a large amount of time wasted, while the sorting device is idling through uninformative molecules. To solve this problem, and increase throughput, we envision performing a pre-purification of our MBD bound material and remove unbound fluorescent probes. Finally, if the sorting devices were fabricated to operate in parallel arrays, multiple nano-sorters can potentially function simultaneously, and greatly increase total yield. Our current method of simply pipetting material out of the reservoir is highly inefficient and leads to considerable sample loss and variability. In the future, we plan to perform whole genome amplification within our sorted output reservoir to eliminate this problem. We also envision fusing our sorting device with single molecule sequencing platforms (Braslavsky et al., 2003; Levene et al., 2003); this would potentially eliminate sample loss and greatly increase sensitivity.

In chapter 4, we will discuss the use of our SCAN device for analysis of *bona fide* epigenetic marks on native chromatin molecules. For these experiments antibodies were used and a cross-linking step was implemented following the binding reaction. We will also discuss SCAN as a means for

detecting simultaneous presence of epigenetic marks. In fact, with the addition of fluorescently labeled antibodies, the SCAN technique could potentially allow for simultaneous detection of numerous marks, and if proper modifications were made to our fluorescent microscope setup, including additional color filters and inspection volumes, there would be no limit to the number of marks that could be simultaneously queried and sorted.

We plan for this technique to operate at very high throughput, by making use of multiple channel arrayed devices, and eventually sort whole genomes for actual epigenetic marks, and sequence the DNA from our sorted material. If this goal is achieved, we will have developed a technology that is an alternative to ChIP-seq. More importantly, if we can detect and sort for multiple epigenetic marks simultaneously, we will have potentially created a technology that supersedes ChIP-seq, and allows for investigations of the histone code at a level that is more extensive than has ever been preformed.

IV. Single-molecule Analysis of Combinatorial Epigenomic States in Normal and Tumor Cells

IV.1. Abstract

Proper placement of epigenetic marks on DNA and histones is fundamental to normal development, and perturbations contribute to a variety of disease states. Combinations of marks act together to control gene expression, therefore, detecting their colocalization is important, but because of technical challenges, such measurements are rarely reported. Instead, measurements of epigenetic marks are typically performed one at a time in a population of cells, and their colocalization is inferred by association. In other words, multiple independent chromatin states could exist in a single population of cells being analyzed, and genomic location could vary from cell to cell. Comparing across ChIP datasets will not allow one to determine if two epigenetic marks truly reside on the same chromatin molecule, or if they simply map to the same genetic location and are never actually on the same piece of chromatin. Sequential ChIP can be used to overcome this limitation; however it is wholly impractical when small inputs are assayed, or when more than two epigenetic modifications are queried. Mass spectrometry can also assay for multiple epigenetic marks, but only if they reside in close proximity on the same histone tail (Britton et al., 2011). All other epigenetic mark combinations cannot be assayed using mass spectrometry. Here we describe a novel single-molecule analytical approach that

can perform direct detection of multiple epigenetic marks simultaneously, and use it to identify mechanisms coordinating H3K9me3, H3K27me3, and cytosine methylation (5mC) placement in the normal and cancer genome. We show that 5mC and H3K9me3 are present together on individual chromatin fragments in mouse embryonic stem cells (ESC) and that half of the H3K9me3 marks require 5mC for their placement. In contrast to H3K9me3, H3K27me3 is antagonized by 5mC in both ESC and primary mouse fibroblasts (MF), indicating this antagonism is shared among primary cells. However, antagonism is lost upon immortalization of MF and persists after tumorigenic transformation. Importantly, human promyelocytic cells show the same loss of H3K27me3 antagonism by 5mC. Because aberrant placement of gene silencing marks like H3K27me3 and 5mC correlate with tumor suppressor silencing and tumor progression, loss of H3K27me3 antagonism by 5mC upon immortalization measured by these methods is likely to be fundamental to cancer. Our platform can enable other studies involving coordination of epigenetic marks, and leverage efforts to discover disease biomarkers and epigenome-modifying drugs.

IV.2. Introduction

IV.2.1. Coordination and co-occupancy of epigenetic marks

Epigenetic marks are responsible for controlling the temporal and spatial pattern of gene expression throughout the genome. In a number of instances,

these marks have been shown to act combinatorially (Bernstein et al., 2006; Murr, 2010). Co-occurance of epigenetic marks have been implicated in a variety of important processes including developmental differentiation (Bernstein et al., 2006), gametogenesis (Ooi et al., 2007), and DNA replication (Eaton et al., 2011). Additionally, examples exist where epigenetic marks can directly promote or inhibit the presence of one another (Tamaru et al., 2003; Zilberman et al., 2008). Consequently, reliably detecting epigenetic mark colocalization is an essential step for advancing a host of biological studies. Histone modifications and 5mC are traditionally assayed by ChIP and BS-Seq respectively. Typically, one assay is performed at a time and colocalization of marks is inferred by association. As previously mentioned in chapter 3, with this approach, it remains unknown if the inferred combinatorial states actually exist (Ernst et al., 2011). reChIP can detect combinations of histone modifications, but its low efficiency requires an abundant source of chromatin, and it is impractical for assaying more than two modifications; BS-Seq of ChIP DNA can report coincidence of histone modifications and 5mC (Brinkman et al., 2012). As discussed in the previous chapter, mass spectrometry can quantify combinations of histone marks, but if they reside nearby on the same histone (Johnson et al., 2004). Each method is labor intensive and difficult to use when quantitative data are needed. Here we describe a novel single-molecule analytical approach that can rapidly and quantitatively assay combinations of epigenomic marks.

IV.2.2. Improved SCAN

We previously described SCAN (Single Chromatin fragment Analysis in Nanochannels), a nanofluidic approach that enabled high throughput fluorescent measurements of single DNA and chromatin molecules (Cipriany et al., 2010). When used to analyze native chromatin from H2B-GFP expressing HeLa cells, we showed that molecules bound with a fluorescent DNA intercalator also carried GFP, demonstrating the chromatin remained intact during the analysis. When we analyzed mixtures of methylated and unmethylated DNAs that were combined with a fluorescently tagged MBD protein, we observed specific detection of methylated DNA. These results suggested SCAN could be used for rapid, quantitative epigenomic measurements, and that it could be used to detect the presence of combinations of epigenetic features on individual chromatin molecules. Here, we reduce this objective to practice and apply SCAN to demonstrate the interdependence of histone modifications on DNA methylation status. We show that 5mC is needed for proper H3K9me3 placement and that it antagonizes H3K27me3 in primary cells, however, the antagonism is lost in immortalized and transformed cells. This loss of antagonism might be a mechanism for aberrant placement of gene silencing marks on tumor suppressors and disease progression.

IV.3. Materials and methods

IV.3.1 Collaborator contributions

       Fabrication of nanofluidic channels and design of post experimental data analysis software were performed by collaborators in the Craighead lab. All other experiments, which I performed, are described below.

IV.3.2. Chromatin preparation

       Native chromatin was prepared as described (Cipriany et al., 2010). Briefly, $3X10^7$ cells where homogenized using a Dounce homogenizer and triton X-100 containing PBS buffer. Isolated nuclei were treated with micrococcal nuclease, and chromatin was solubilized in a high salt EGTA containing buffer. Presence of nucleosome ladders was verified and sizes estimated using agarose gels loaded with DNA purified from chromatin (**Fig IV.1**). Samples used for analysis had size distributions centered on 2kbp. Concentration was assessed spectrophotometrically.

IV.3.3. Labeling epigenetic probes

       Antibodies were purchased from Active Motif (61013, 39155, 61037, 39763) and the MBD probe was prepared as previously described (Cipriany et al., 2010). All reagents were then labeled using Invitrogen's Microscale Protein Labeling kits (A10238, A30009, A30006) according to manufacturer's instructions
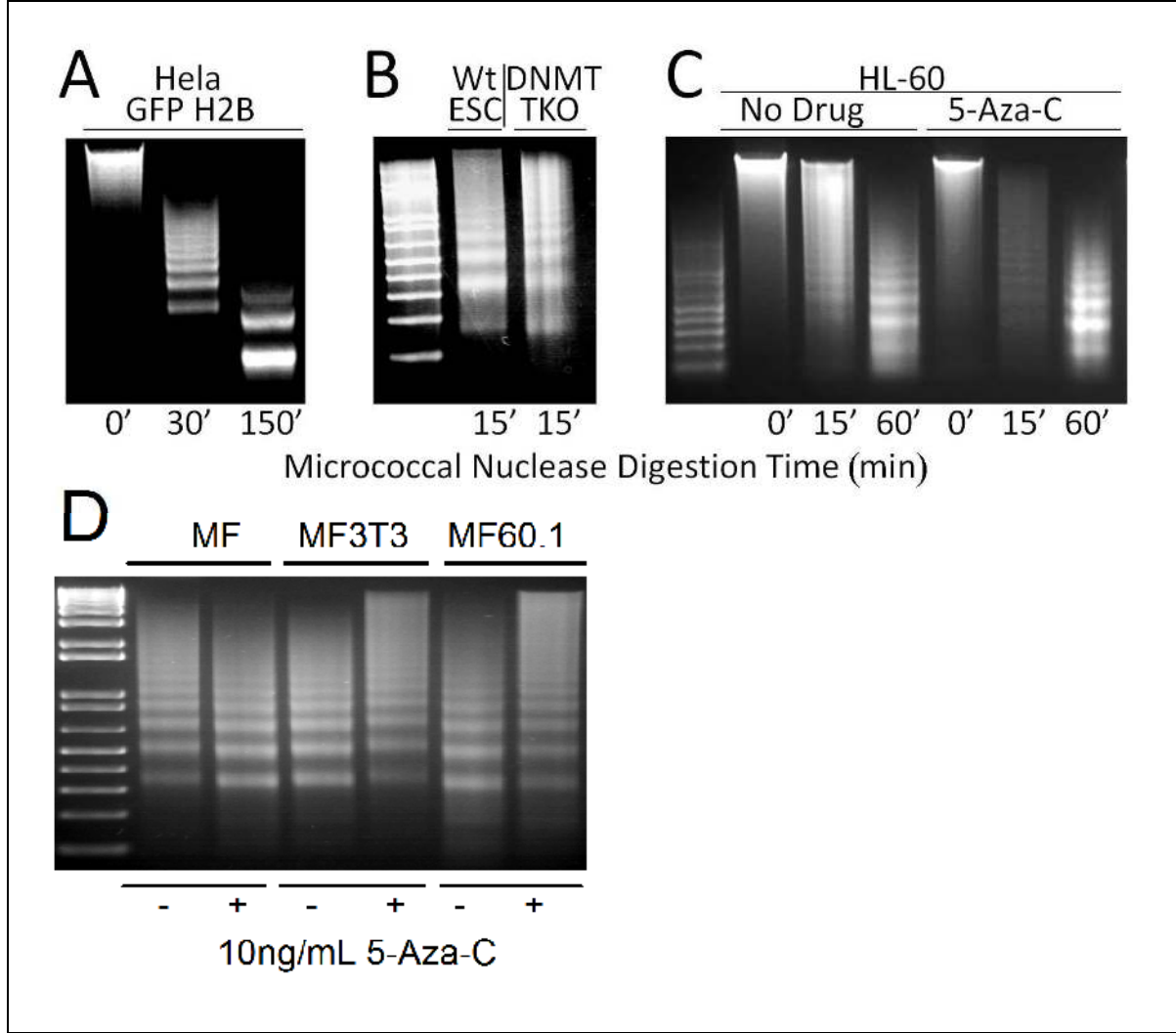
Fig IV.1 - Native Chromatin Purification Laddering
DNA was purified by phenol chloroform extraction and ethanol precipitation from native chromatin isolated from (A) HeLa cells expressing GFP-tagged H2B; (B) Wild type mouse ESC or DNMT TKO ESC deficient for the three DNA methyltransferases, and (C) HL-60 cells cultured without or with 5-Aza-2′-deoxycytidine, and then analyzed by agarose gel electrophoresis. (D) Following 72 hours of culture in media containing 10ng/mL of 5-Aza-C, cells were harvested and native chromatin was extracted. The above chromatin samples were selected for analysis based on a fragment size distribution that was similar to previously analyzed from ESC and HL-60 samples. Micrococcal nuclease incubation times were varied. 1 kbp+ ladders are shown.

and purified from free dye by seven cycles of 10 fold dilution into PBS and re-concentration in Amicon Ultra Centrifuge concentration columns (UFC501008, UFC51008). Labeling was assayed using fluorometry and absorbance; protein concentrations were confirmed using a BCA assay.

IV.3.4. Binding reactions

Native chromatin (3-10nM) and fluorescent antibodies (125-400nM) were mixed such that there was a 15-40 fold molar excess of antibody or MBD and then left to incubate over night at 4°C on a rotato r. The following morning 0.75% formaldehyde was added and samples incubated at RT for 15min, after which, 160mM glycine was added to quench the crosslinking reaction. Once crosslinked, fluorescent DNA intercalator dye (TOTO-3 or YOYO-1, Invitrogen) was added such that there was 1 molecule of dye for every 5 DNA base pairs. Samples were left to incubate for 1 hour at RT followed by 1 hour at 4°C. Before SCAN, binding reactions were diluted into flow buffer such that the total concentration of fluorescent molecules was below 1nM.

IV.3.5. SCAN and data analysis

SCAN data were collected as described (Cipriany et al., 2010). To measure the abundance of epigenetic marks, we determined the fraction of intercalator-stained material bound to the epigenetic probe, and normalized this

value to the fraction bound to anti-H2B for both test and control cells, and then

took the ratio of normalized values for the two cell types according to the formula:

$$
NAE = \frac{\left[\left(\frac{B_e}{I}\right)\Big/\left(\frac{B_2}{I}\right)\right]_T}{\left[\left(\frac{B_2}{I}\right)\Big/\left(\frac{B_e}{I}\right)\right]_C}
$$

where *NAE* = Normalized Abundance of Epigenetic marks, B_e = count of

molecules bound to both intercalator and epigenetic probes, I= count of

intercalator bound molecules, B_2 = count of molecules bound to both intercalator

and anti-H2B, *T* = Test cells, *C* = Control cells. All data analysis was performed

using MATLAB 2010b (Mathworks). Raw intensity versus time traces were

autocorrelated and fit to the 1D FCS equation with the addition of intersystem

crossing and directed flow terms, to determine the characteristic transit time of

molecules through the focal volume. A noise level of random background

photons was established and used to set a baseline for the intensity traces.

Single molecule events were identified above the baseline by applying a signal

threshold and locating local maxima among contiguous points above the

threshold. Coincident molecules were identified by temporally translating one

channel trace relative to a second, over a time window whose size was

determined from the characteristic transit time of the single molecule events.

The number of overlapping local maxima from both channels was determined at

each time offset generating a time distribution of coincident events. By analyzing

the time distribution of coincident events from multiple datasets a signal to noise ratio was determined and used to calculate background levels and correct the values for abundance of epigenetic mark detection. Due to an increased noise level from falsely detected coincident events, for experiments where two different colored probes were used, a similar background correction was applied by uniformly subtracting the bulk of background events from total events, effectively shifting the y-axis on coincidence plots.

IV.3.6. Cell culture and drug treatment

A variety of cell types were used for the described experiments and culture conditions varied. The conditions used for HeLa cell culture were described in chapter 3. Mouse embryonic stem cells were V6.5 and cultured in DMEM supplemented with 15% fetal calf serum and 1,000 units per mL of leukemia inhibitory factor (LIF). Mouse fibroblasts, 3T3 fibroblasts, and 60.1 (Soloway et al., 1996) fibroblasts were cultured in DMEM supplemented with 10% fetal calf serum. HL-60 cells were cultured in RPMI media supplemented with 5% fetal calf serum. During drug treatment, $5X10^7$ cells were cultured in media containing 1.1ng/mL (HL-60) or 10ng/mL (MF lines) 5-Aza-2′-deoxycytidine (Sigma A3656) for 72 hours with the addition of fresh drug containing media every 24 hours. Chromatin was prepared as described above. Concentrations were chosen that permitted cell accumulation at a level equal to 40-60% of untreated cells (**Fig IV.2**).
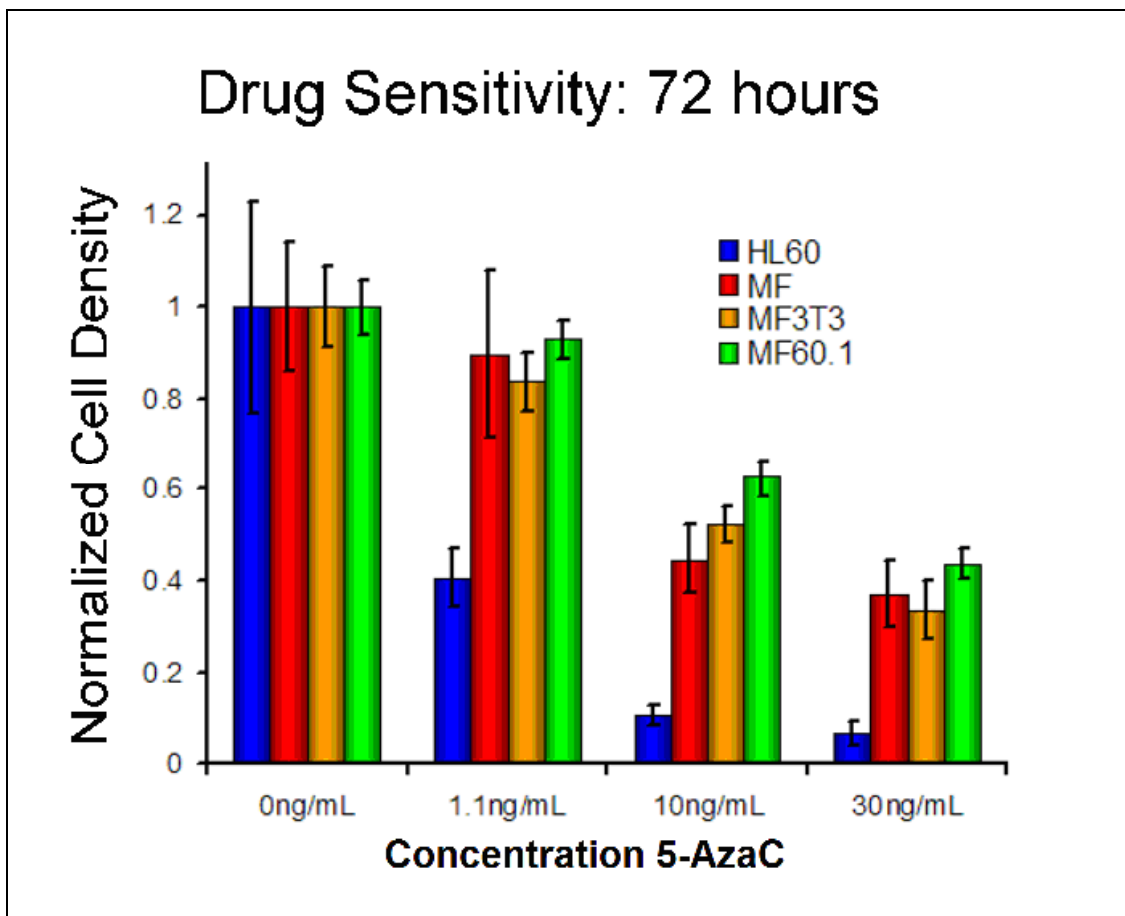
Fig IV.2 - 5-Aza-2′-deoxycytidine Kill Curve
Four different concentrations of 5-Aza-2′-deoxycytidine were added to HL-60 or mouse fibroblast (MF) cells, which included primary fibroblasts (MF), immortalized fibroblasts prepared by groWth under 3T3 conditions (MF3T3) or fully tumorigenic fibroblasts transformed with Ha-*ras* and v-*myc* containing virus (MF60.1). Fresh drug was added every 24 hours and after 72 hours cells were harvested and counted. Three biological replicates were used for each cell line and three technical replicate counts were made per dish. The graph corresponds to the average across three biological replicates, ± standard deviation. Experiments in fig. 4 used HL-60 cells grown in 1.1ng/mL 5-Aza-2′-deoxycytidine and MF cells grown in 10ng/mL 5-Aza-2′-deoxycytidine.

IV.4. Results

IV.4.1. Detection of epigenetic marks on native chromatin

We first established conditions for binding fluorescent MBD and antibodies recognizing histone features to chromatin with high specificity (**Fig IV.3A**). As an initial test, we labeled an antibody recognizing the unmodified N-terminal tail of histone-H3 (anti-H3) with AlexaFluor647, bound it to native chromatin isolated from HeLa cells expressing an H2B-GFP fusion protein, and then analyzed the mixture by SCAN. We ensured the total concentration of fluorophores in our analyte remained at or below 1nM so that the probability of detecting only a single molecule or complex was greater than 99.5% (**Fig IV.3B**). If SCAN could detect antibody-chromatin complexes, we should detect individual molecular complexes that emit both green (GFP) and red (Alexa647) photons. This is precisely what we observed demonstrating labeled antibodies could be used to detect chromatin features on single molecules using SCAN (**Fig IV.4A+B**).

Because most chromatin samples of interest will not come from cells with GFP-tagged histones, we extended this anti-H3 binding test using native chromatin from wild type mouse embryonic stem cells (Wt-ESC) labeled with the fluorescent intercalator YOYO-1, and then used SCAN to detect complexes carrying YOYO-1 and anti-H3AlexaFluor647. We could easily detect these

104
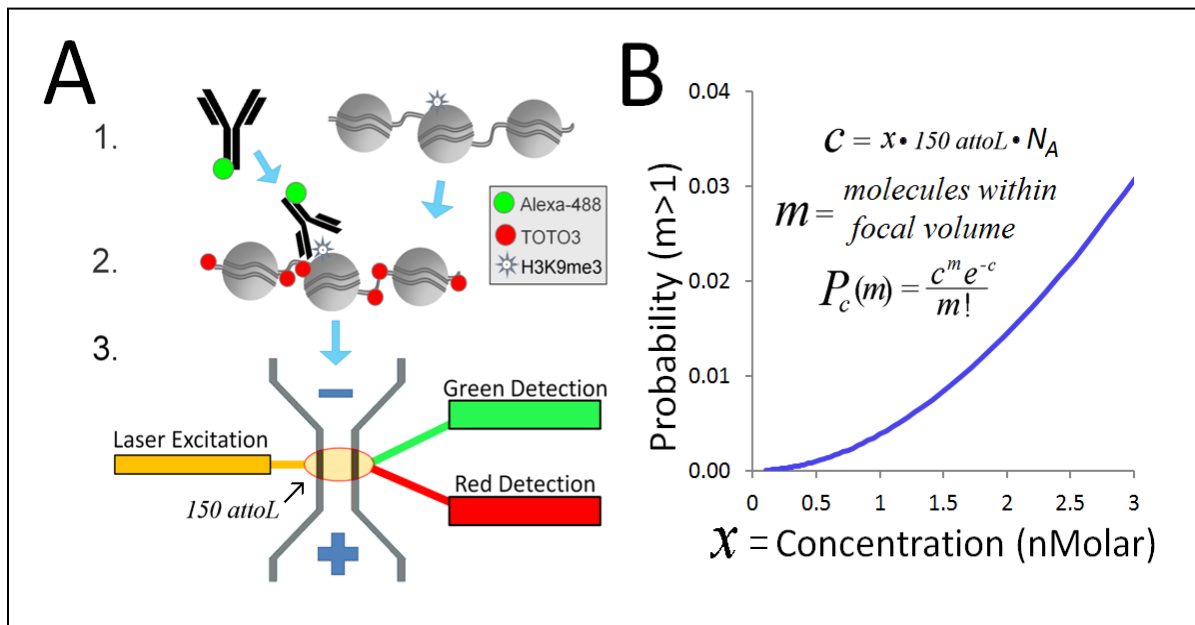
Fig IV.3 - Single chromatin molecule analysis at the nanoscale (SCAN) workflow

A) Native chromatin bearing epigenetic marks is mixed with fluorophore (e.g. Alexa Fluor-488) labeled antibody specific to a given mark. After binding, the chromatin is labeled with an intercalator (e.g. TOTO-3). Finally, the chromatin is driven by voltage through a nanoscale channel fabricated in fused silica and fluorescent measurements of individual molecules are taken in a 150 aL inspection volume. A more detail schematic of laser setup can be found in our previous publication (Cipriany et al., 2010).

B) Probability of erroneously interrogating more than a single molecule increases with analyte concentration according to a Poisson distribution, and is less than 0.5% at the concentrations used here (≤1 nM). Pc(m), probability of m molecules residing in the 150 aL interrogation volume at any one time, given concentration c of fluorescent molecules in analyte, expressed as molecule count x in 150 aL, where NA is Avogadro's Number. The curve shows the probability of more than 1 molecule is in the inspection volume.
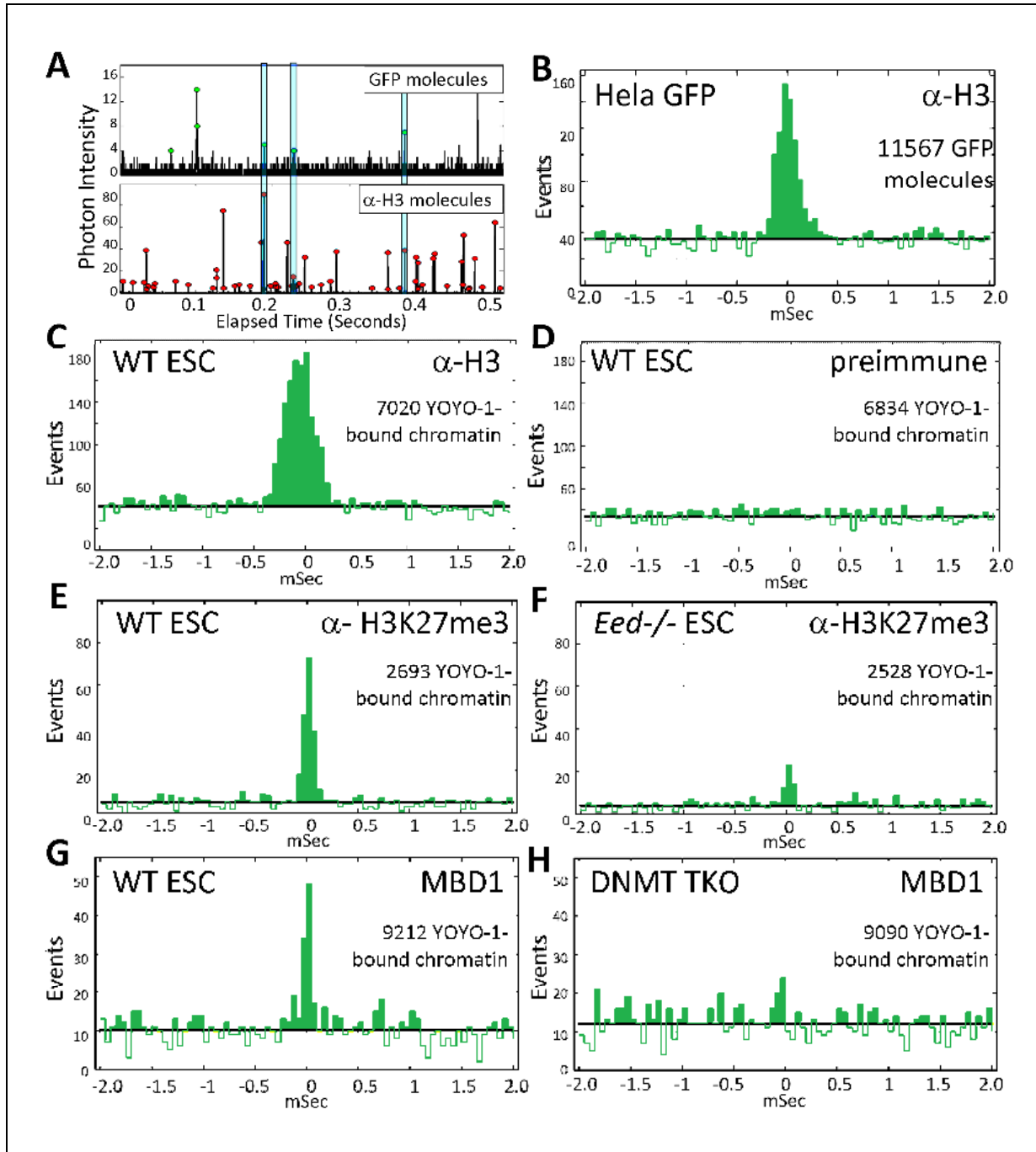
Fig IV.4 – SCAN detects chromatin features with high specificity
AlexaFluor dye-tagged antibodies or MBD protein was bound to chromatin before
analysis by SCAN.
A+B) Chromatin was from H2B-GFP expressing HeLa cells, and antibody was
H3-specific.
A) 0.5 s of time-resolved photon counts reporting GFP (top) and H3 (bottom)
fluorescent emissions, depicted as colored peaks. Blue shading identifies
antibody-bound chromatin complexes emitting both fluorophores nearly
simultaneously.
B) Time offset histogram for 7,065 GFP chromatin molecules analyzed in A.
GFP fluorescence from coincident events is placed at time 0 and the time
offset identifies how soon before or after the GFP emission, Alexa dye was
detected.   The peak centered at time 0 with width less than the transit time
for molecules passing through the inspection volume (~1 ms) identifies
antibody-chromatin complexes. A more detailed illustration of the
coincidence calculation is provided in supplemental methods and in Fig. 2.
C+ D) Chromatin was from wild type mouse ES cells (Wt ESC) bound to anti H3
antibody (C), or pre-immune mouse serum (D).
E + F) Antibody probe was specific for H3K27me3 and chromatin was from Wt
ESC (E) or Eed– ESC, which are deficient for H3K27me3 (F).
G + H) MBD probe was specific for mC in duplex DNA, and chromatin was from
Wt ESC (G) or DNMT TKO ESC, which are deficient for mC (H). Chromatin
in C-F was detected by labeling with the intercalator YOYO-1, whose
emission defined the 0 time in the offset plots.

complexes, demonstrating the utility of SCAN for a variety of chromatin sources (**Fig IV.4C**). To test the effects of non-specific crosslinking of antibody to chromatin, as a negative control, we repeated this test using AlexaFluor647 labeled pre-immune mouse serum in place of anti-H3 and observed no antibody-chromatin complexes, demonstrating specificity of antibody binding (**Fig IV.4D**). Similar experiments were performed at various formaldehyde concentrations, and over a range of incubation times. From these tests, optimal cross linking time was determined to be 15 minutes, and optimal formaldehyde concentration was 0.75%. However, we observed no increase in binding or in nonspecific crosslinking when formaldehyde concentration was increased to 1%, or when incubation time was increased to 20 minutes. (data not shown) Presumably there is still room to improve the crosslinking protocol with further optimization. In additional pilot experiments, MBD was bound to chromatin from ES cell chromatin isolated from both Wt and TKO (mutant for all three of the DNA methyltansferases) cells at various concentrations ranging from 312nm to 2.5um. Although there was abundant binding in Wt chromatin, there was no detectable binding of the MBD in the TKO chromatin when concentrations were at or below 625nM. However, when concentrations of MBD exceeded 625nM there was binding in both the Wt and DNMT TKO samples (data not shown). This result indicated there are concentration dependent effects on crosslinking specificity. For this reason, the cross linking reaction was always performed at concentrations below 500nM.

Anti-H3 should bind most if not all chromatin molecules in a mixture of complex chromatin. To determine if we could use SCAN to detect less common epigenomic features of particular interest, we bound AlexaFluor647 labeled anti-H3K27me3 specific antibodies to native chromatin from Wt-ESC, labeled it with YOYO-1, and then performed SCAN. We could readily detect YOYO-1 labeled chromatin bound to anti-H3K27me3 antibodies (**Fig IV.4E**). As a negative control for anti-H3K27me3 binding specificity, we repeated this test using chromatin from ESC homozygous for a mutation in *Eed,* which encodes a PRC2 component needed for efficient H3K27me3 (Montgomery et al., 2005). In *Eed*-deficient cells, the number of YOYO-1 labeled molecules bound to antibody was greatly diminished (**Fig IV.4F**). Coincidence in *Eed-* chromatin was <25% of Wt levels. Therefore, detection of H3K27me3 using SCAN and anti-H3K27me3 antibodies is specific. The moderate level of coincidence in the *Eed-* chromatin samples could be the result of either nonspecific antibody binding, or residual H3K27me3 in the absence of EED. We performed a similar test using fluorescently-tagged MBD protein to detect DNAs harboring 5mC in Wt-ESC, and included as a negative control chromatin from DNMT TKO ESC that are deficient for the three mammalian DNA methyltransferases, which have ~2% of the 5mC content of Wt ESC (Meissner et al., 2005). The MBD bound chromatin from Wt ESC, but binding to DNMT TKO ESC chromatin was greatly diminished (**Fig IV.4G+H**). Like H3K27me3, detection of 5mC by SCAN and MBD is specific.

IV.4.2. Simultaneous detection of two epigenetic marks

We next used SCAN to detect combinations of epigenetic features on chromatin. For these experiments, we bound to chromatin two different probes recognizing epigenetic features, both labeled with spectrally distinct fluorophores. In our first tests we bound antibodies recognizing H3 and H2B to Wt ESC, which are expected to bind virtually all chromatin fragments. As a negative control for antibody aggregation, we performed an identical binding reaction without chromatin. Only in the binding reactions that included chromatin could we detect complexes with both antibodies, demonstrating that we can use SCAN for simultaneous detection of multiple chromatin features (**Fig IV.5A+B**). We extended this test incrementally by substituting anti-H3K27me3 for the anti-H3 antibody and were able to detect chromatin molecules binding both antibodies (**Fig IV.5C**). Finally, we performed binding reactions using anti-H3K9me3 and MBD and detected chromatin molecules bearing two combined epigenetic features (**Fig IV.5D**). To our knowledge, this is the first direct detection of a combination of epigenetic features on single chromatin molecules. As combinations of epigenetic features are fundamental to genomic regulation (Ruthenburg et al., 2007), their multiplexed detection with tools like SCAN can provide new insights into the epigenome.

Having shown that 5mC and H3K9me3 were commonly detected on the same individual chromatin molecules, we wondered if H3K9me3 was dependent
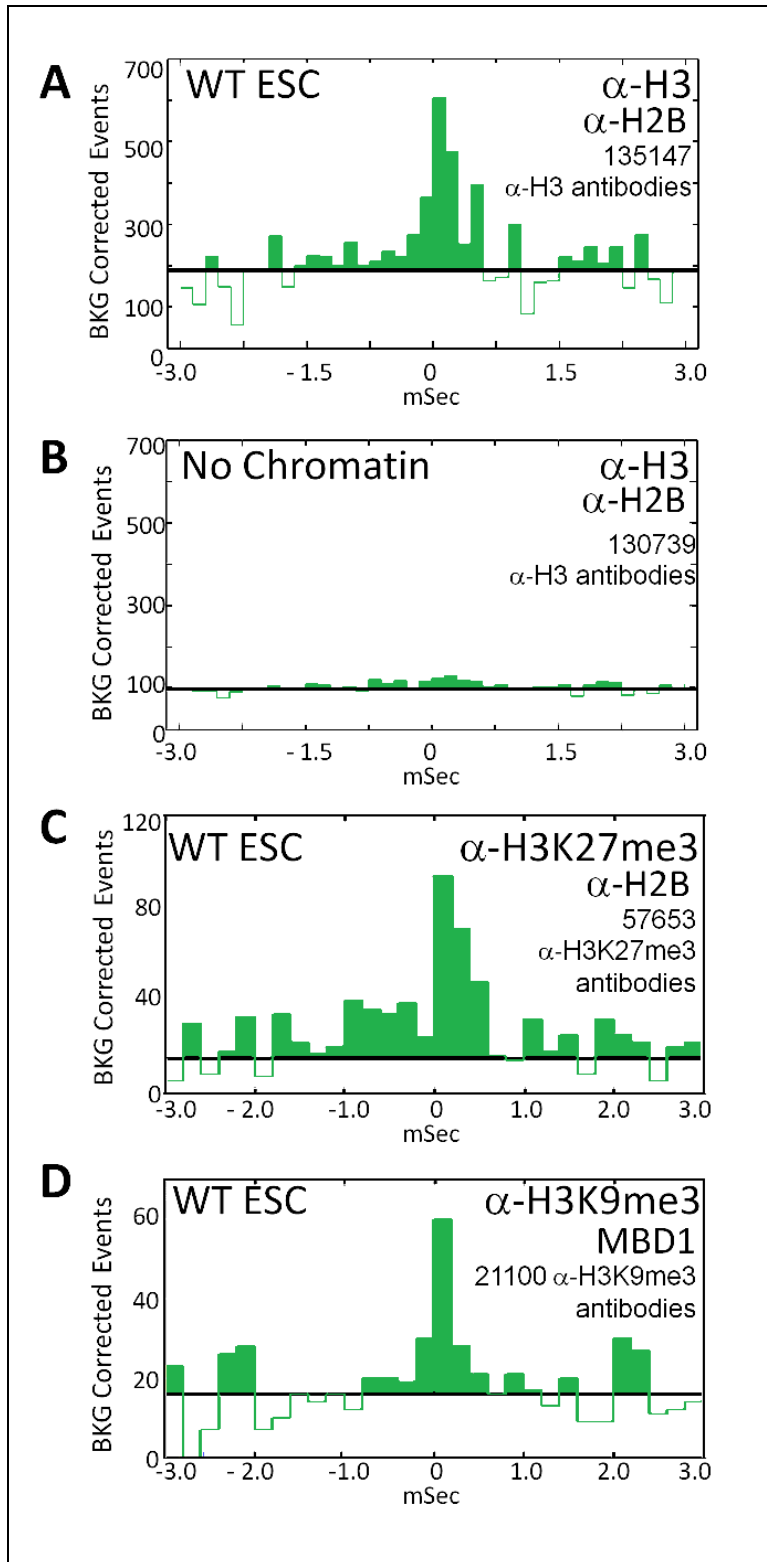
Fig IV.5 - Detection of two epigenetic marks simultaneously
A + B) Antibodies against H3 and H2B were labeled with spectrally distinct fluorophores and bound to chromatin (A), or included in a binding reaction without chromatin as a negative control (B).
C+ D) Binding reactions were performed as in (A) using antibodies against H2B and H3K27me3 (C), or using MBD protein and an antibody against H3K9me3 (D). Chromatin was from Wt ESC; the central peak identifies chromatin molecules bound to two fluorescent probes as described in Figure 2. Due to increased false coincident detection, events (y-axis) were background corrected for A-D.

on 5mC for its placement. There is precedent for cross regulation of the two marks: H3K9me3 is needed for normal 5mC deposition in *Neurospora crassa* and mice (Lehnertz et al., 2003; Tamaru et al., 2003); conversely, 5mC positively affects H3K9me3 placement at normally silenced loci in *Arabidopsis thaliana* (Johnson et al., 2002) however in some human cell cultures, 5mC antagonizes H3K9me3 placement (Komashko and Farnham, 2010). In none of these studies was the magnitude of these effects quantified genome wide. We used quantitative SCAN to measure the relative abundance of H3K9me3 on chromatin in Wt ESC *vs.* DNMT TKO cells, using H2B antibody to normalize the signals from the two cell types. Our results showed that 60% of the H3K9me3 levels in ESC depend on 5mC placement (**Fig IV.6A**).

IV.4.3. Coordination of epigenetic mark placement

 H3K27me3, like H3K9me3 and 5mC, is commonly associated with gene silencing; therefore, we also measured changes in H3K27me3 levels when 5mC was diminished. In contrast to H3K9me3, H3K27me3 levels rose approximately 220% when 5mC was lost (**Fig IV.6A**). This is consistent with previous reports that 5mC antagonizes H3K27me3 placement in primary mouse fibroblasts at *Rasgrf1* (Lindroth et al., 2008) and is consistent with global suppression of H3K27me3 in mouse ESC by 5mC (Hagarman et al., 2013). Interestingly, the antagonism between 5mC and H3K27me3 seen in primary cells breaks down
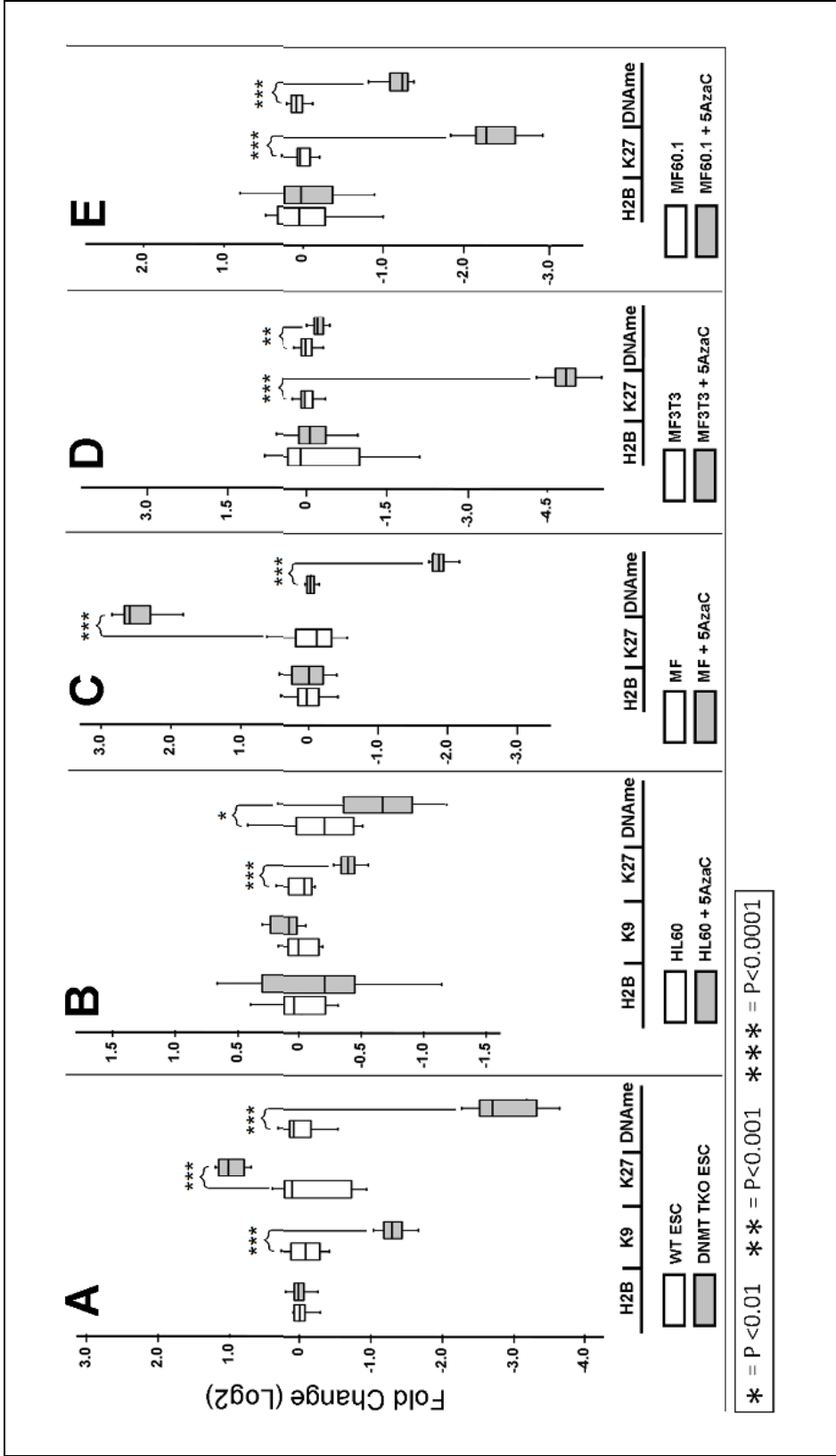
Fig IV.6 - DNA methylation state controls histone modification states
ESC (A), HL-60 (B) MF primary mouse fibroblasts (C), MF3T3 immortal mouse fibroblasts (D) or MF60.1 transformed oncogenic mouse fibroblasts (E) had normal mC, or mC impaired by mutation (A) or 5-Aza-2′-deoxycytidine (5AzaC) treatment (B-E). Chromatin from cells was analyzed by SCAN after labeling with an intercalator (TOTO-3 or YOYO-1) and fluorescent MBD protein or antibody recognizing H2B, H3K9me3 (K9) or H3K27me3 (K27). H2B signals were used to normalize relative abundance of H3K9me3, H3K27me3, and mC in cells with normal and impaired mC. In box plots, whiskers represent 5th and 95th percentile.

in transformed cells. For example, 5mC requires H3K27me3 at several loci in HeLa cells (Vire et al., 2005). Additionally, 5mC is not found at many PRC2 binding sites in primary cells but 5mC is aberrantly acquired at PRC2 binding sites in cancerous cells (Keshet et al., 2006; Ohm et al., 2007, 2010; Schlesinger et al., 2007; Widschwendter et al., 2007; Doi et al., 2009; 5mCCabe et al., 2009). Generally, coordination between 5mC and H3K27me3 in transformed *vs*. normal cells varies by genomic location (Hon et al., 2012). To compare the influence of 5mC on H3K27me3 in transformed cells with its effect in primary ESC, we treated the human acute promyelocytic leukemia cell line HL-60 with the demethylating agent, 5-Aza-2′-deoxycytidine (5AzaC) (**Fig IV.6)**, and then used quantitative SCAN to measure H3K27me3 abundance. In contrast to primary ESC where 5mC depletion leads to increases in H3K27me3 placement, depletion of 5mC in HL-60 cell caused a 24% decrease in chromatin associated H3K27me3, consistent with the hypothesis that antagonism between 5mC and H3K27me3 that exists in primary cells breaks down in cancer (**Fig IV.6B**). As a further test of this hypothesis, and to define when during oncogenic transformation antagonism between 5mC and H3K27me3 is lost, we used 5AzaC to demethylate mouse fibroblasts from primary cultures (MF), from cells immortalized during 3T3 culture selection (MF3T3) and from tumorigenic fibroblasts (MF60.1), fully transformed with Ha-*ras* and *v*-myc (Soloway et al., 1996), and then used SCAN to quantify the effects of 5mC depletion on H3K27me3 abundance. Similar to mouse ESC, depletion of 5mC in primary MF

116

caused 570% increase in the abundance of H3K27me3, indicating the antagonism of H3K27me3 by 5mC seen in primary ESC was also seen in other primary cells (**Fig IV.6C**). However, in immortalized MF3T3 and fully tumorigenic MF60.1 cells depleted of 5mC, H3K27me3 levels decreased by 98% and 81% respectively, demonstrating antagonism of H3K27me3 by 5mC was lost (**Fig IV.6D + E**). Therefore, upon immortalization and extending into oncogenic transformation the influence of 5mC on H3K27me3 switches from inhibitory to stimulatory, leading to an overall increase in the abundance of H3K27me3 mark placement (**Fig IV.7**). Although there was a significant decreases in the levels of 5mC following drug treatment, the degree of decrease varied across cell types. This variation is presumably do to non-uniform drug susceptibility from one cell type to another. Alternatively, observed variation could be a reflection of our detection method's limits on sensitivity. The observed switch in antagonism might be fundamental to epigenetic chaos associated with cancer and that contributes to tumor suppressor silencing. That it occurs upon immortalization and before full transformation suggests loss of H3K27me3 antagonism by 5mC is an early biomarker for cancer.

IV.5. Discussion

Here we described a single molecule analytical approach to characterize epigenetic states. This allowed us to detect combinations of epigenetic features,

117
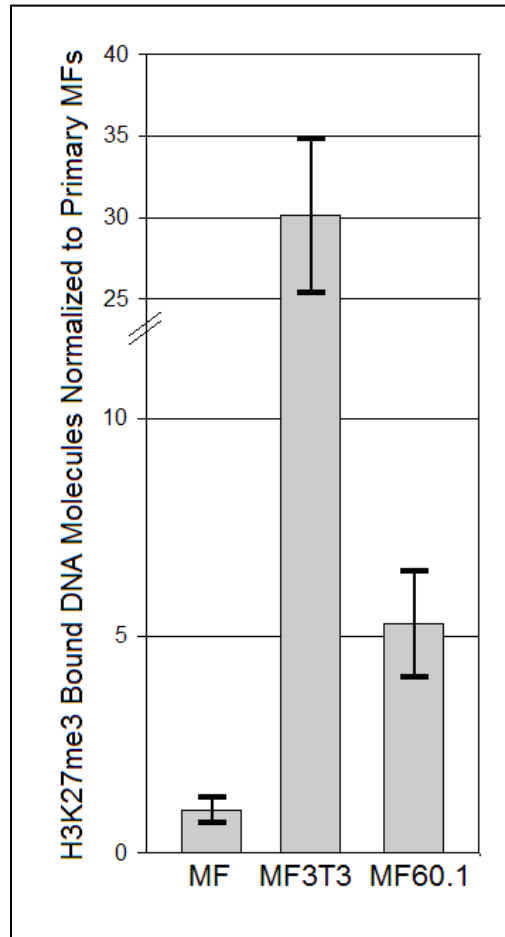
Fig IV.7 - H3K27me3 abundance in mouse fibroblasts
H3K27me3 abundance was quantified for primary mouse fibroblasts (MF), immortalized mouse fibroblast (MF3T3), and tumoregenic mouse fibroblasts (MF60.1). Similar to figure 4, abundance was determined by measuring antibody binding in chromatin from each cell type. Counts are first normalizing to H2B binding abundance then displayed as compared to H3K27me3 levels in primary fibroblasts. Error bars represent standard deviation.

measure epigenetic mark abundance rapidly, identify coordinated regulation among 5mC, H3K9me3 and H3K27me3, and monitor effects of epigenome modifying drugs on cancer cells.

Our platform directly detected the colocalization of H3K9me3 and 5mC on individual molecules in ESC.   The colocalization is relevant mechanistically: H3K9me3 requires 5mC for its normal placement in the genome. The requirement is not absolute as nearly half of H3K9me3 is placed in the absence of 5mC. It is possible that without 5mC, the histone deacetylase-containing NuRD complex is not recruited to 5mC by the MBD2 protein in the complex, and H3K9Ac persists preventing H3K9me3 placement.

In contrast to H3K9me3, 5mC antagonizes H3K27me3 in primary cells. The antagonism can be explained by the fact that 5mC impairs binding of PRC2, which is needed for H3K27me3 (Wu et al., 2010). This antagonism breaks down in immortalized mouse cells and fully transformed mouse and human cells. Further characterization of H3K27me3 antagonism by 5mC will be necessary to determine if its loss is a reliable biomarker for different cancers.

How immortalization and subsequent transformation causes H3K27me3 to become dependent on 5mC is unknown. It is possible that the composition of PRC2 or PRC2 associated factors, or posttranslational modifications of any of

119

these changes upon immortalization, enabling recruitment of PRC2 to 5mC

rather than exclusion as occurs in primary cells.   Alternatively, the failure of 5mC

to exclude H3K27me3 might be a consequence of altered energy metabolism in

cancerous cells. Increased glycolysis and reduced tricarboxylic acid (TCA) cycle

activity occurs in many cancer types (Cao et al., 2010; Cardaci and Ciriolo, 2012;

Mullen and Deberardinis, 2012), which can lead to accumulation of succinate,

which in turn has been shown to directly inhibit the H3K27 demethylase JMJD3

(Cervera et al., 2009). It is possible that in normal cells, antagonism of

H3K27me3 by 5mC entails targeting of JMJD3 to methylated DNA sequences,

and increased succinate in transformed cells might impair removal of H3K27me3,

causing it to accumulate at sites where 5mC is present (**Fig IV.7**). The

mechanisms controlling H3K27me3 placement in primary cells are fundamentally

different from immortal and tumorigenic cells and loss of H3K27me3 antagonism

by 5mC during transformation might contribute to cancer-associated phenotypes

such as tumor suppressor silencing. Interfering with this aberrant epigenetic

mechanism might be beneficial therapeutically.

Our studies of coordination among 5mC, H3K9me3 and H3K27me3

represent just one application of our single molecule analytical platform. Other

applications for the SCAN platform include enabling discovery of additional

epigenome based cancer biomarkers, facilitating quantitative assays for the

effects of epigenome modifying agents in a drug discovery pipeline, and serving

in clinical assays to measure patient responses to such drugs. There are several

additional embodiments possible for our system that can expand its capabilities,

including performing parallel analyses in multiple channels to increase

throughput; adding a third fluorescent color, which would allow for absolute

quantitation of epigenetic marks; and implementing a recently developed method

to sort single DNA molecules based on epigenetic state (Cipriany et al., 2012)

which would allow for downstream sequencing. Aside from epigenetic marks, in

the future we anticipate using SCAN to access the genomic localization of

transcription factors, which would likely be accomplished by analyzing *in vivo*

cross linked chromatin. Because individual molecules are queried, only small

amounts of cellular material are required, raising the possibility of epigenomic

analyses of cells that are rare, impossible to culture or even of single cells

(Benítez et al., 2012).

V. Expanded discussion

Waddington originally used the term epigenetics to describe the combined effects and interaction of genes with their products as they contribute to a developmental phenotype. In Waddington's day the mechanisms controlling epigenetic phenomena were largely a mystery. Today we have a deep understanding of what epigenetic modifications are, and where they are placed throughout the genome, but generally, even after 75 years of research, we only have a superficial understanding of what mechanisms determine which genes are chemically modified, and how these epigenetic modifications lead to an altered cell fate.

This dissertation investigated the mechanism underlying epigenetic mark localization and mark coordination during transformation. In doing so, we have defined a mechanism that functions to control placement of 5mC in the genome during gametic epigenetic mark establishment. Additionally, we have developed a new method to assay for epigenetic phenomena, and use it to determine that genome wide antagonism of H3K27me3 and 5mC breaks down during immortalization and tumorigenic progression. The findings presented here have not only helped to characterize epigenetic mark regulation, but will ultimately lead to new questions and innovative research projects.

The characterization of 5mC establishment during mouse epigenetic reprogramming was accomplished using the imprinted *Rasgrf1* model locus. Our data indicate that a repeat region, located 3' the DMR functions as a promoter for a pitRNA. They also indicate that 24-30bp piRNAs are produced from a pitRNA transcript and are processed by PIWI proteins. Together, PIWI proteins, the repeat region, the pitRNA, and piRNAs are necessary for deposition of 5mC. Additionally, we determined that *cis* and not *trans* mediated transcription of the pitRNA is required for proper establishment of 5mC. For this reason, we propose a model where deposition of methylation takes place in a co-transcriptional manner.

The means by which DNMTs are targeted to the DMR is unclear, and the mediator proteins are unknown. Presumably by RNA-Immuno precipitation (RIP), and by gel shift assays these mediator proteins can be identified. An inducible pitRNA promoter system, in combination with a variety of techniques like ChIP and RIP, may also be useful to confirm that methylation is being placed co-transcriptionally. The inducible expression system may also enable one to determine during at what developmental time, at what expression level, and in which tissue, the pitRNA is required.

Maintaining 5mC in the prelimplantation embryo also requires the repeat region (Holmes et al., 2006). It is possible that in the embryo, similar to their role

in embryonic gametes, the repeats drive expression of a ncRNA. However

PIWI/piRNA mechanisms have not been described in early embryonic somatic

tissue, and have not been linked to maintenance of 5mC. Perhaps there is an

unreported low level of expression from genes within the PIWI pathway in the

zygote, and piRNA mediated silencing is required for maintenance of 5mC.

Alternatively, independent of piRNAs, the repeats may facilitate binding of

proteins, or histone modifying agents that protect against demethylation of the

DMR. In either case the inducible promoter system could help to address this

question. By regulating the pitRNA expression at various developmental time

points with a promoter distinct from the repeats, and monitoring 5mC of histone

modifications, one may be able to determine when, and in what tissues pitRNA

transcription is required for 5mC deposition on the DMR. By creating a system

where *trans*-acting factors are tethered directly to the DMR (Quenneville et al.,

2012), one may be able to determine which factors and/or histone modifications

are necessary for maintenance of 5mC over the *Rasgrf1* DMR. It is likely that

future research will focus on defining this mechanism and refine the model to

explain both establishment and maintenance of 5mC at the imprinted *Rasgrf1*

DMR. In general, mechanisms to explain genomic acquisition of 5mC are not well

established, so research in this area will be important for the field.

Interestingly, earlier work at *Rasgrf1* led to the hypothesis that H3K27me3

and 5mC are antagonistic on a genome wide scale (Lindroth et al., 2008).

Understanding coordinate regulation of epigenetic marks, and identifying complex relationships between various modifications is important for a number of reasons (Li et al., 2005; Bernstein et al., 2006; Mikkelsen et al., 2007; Ooi et al., 2007; Zilberman et al., 2008; Eaton et al., 2011). Unfortunately, prior to our work, no technique could assay for multiple epigenetic marks from rare populations of cells. For this reason, our goal was to develop a technology capable of doing just that. Here we have described the optimization of single molecule nanofluidic based techniques that are capable of detecting intact chromatin, assaying for epigenetic marks, purifying DNA based on 5mC, quantifying relative epigenetic mark abundance, and detecting multiple simultaneous epigenetic mark presence on single chromatin molecules. Importantly, we have used this new technology to determine that normal antagonism of H3K27me3 and 5mC breaks down during cellular immortalization. We have not, however, demonstrated this device's utility to analyze very rare populations of cells.

Recently we have begun to isolate small amounts of genetic material from fewer than 100 cells using microscale extraction devices (Benítez et al., 2012). In this system, cells are loaded directly into a single stream cell-capture and DNA extraction system. The capture device consists of a micro-fluidic channel in the shape of a bottle neck. Cells are deposited into the narrow end of the channel and, as the device loads, they tumble down towards the wide end of the channel. Through-out the device are a series of randomly positioned pillars that are meant

125

to trap, and immobilize cells. Once trapped, cells are chemically lysed, and DNA (or chromatin) is then immobilized and elongated on the pillars. At this point DNA can be assayed using either *in situ* hybridization probes, or intercalator dyes. Alternatively, DNA can be released by nucleases, and collected in the outflow reservoirs for downstream analysis. We foresee the nanofluidic epigenetic sorting device will be fused to the microscale cell capture device, and following extraction and sorting, epigenetic mark containing molecules will be sequenced. Thus, we envision a system able to perform genome wide multiple epigenetic analyses from fewer than 100 cells. Because our nanoscale channels require sub attogram quantities of material, it's conceivable that we would be able to sample the epigenetic state of a single cell a number of times during an individual experiment. However, to perform whole genome analysis, we would need to increase the throughput of sorting by 4 to 5 orders of magnitude. Strategies to accomplish this were discussed in chapter 3. Because this new and improved version of SCAN would allow for chromatin to be purified and analyzed within a very short amount of time and with relatively little sample loss, we foresee its use in a clinical setting, as a means for analyzing the epigenome of rare populations of cells, or for patient diagnosis.

In this dissertation I have described two mechanisms by which epigenetic marks are regulated. The first occurs during embryonic gamete development while *in utero* and during preimplantaiton development, and the second occurs in

normal tissue, but breaks down during carcinogenesis. The latter observations were made using a novel technology developed through the thesis work described. Determining the mechanisms that control epigenetic state is fundamental to our understanding of normal development and disease. Next generation technologies, like SCAN, will enable new kinds of studies to explore epigenomic mechanisms, and will provide tools for diagnoses and treatment of disease.

Previously, I discussed the carcinogenic effects in rodents from *in utero* BPA exposure (Doshi et al., 2012). Frighteningly, this chemical is used extensively in a variety of house hold products, including toys, drinking bottles, food containers, and plumbing (Welshons et al., 2006). Recent studies have found that human BPA exposure is significantly higher than originally predicted, and such levels could have significant effects on human phenotype and disease state (Welshons et al., 2006). There are other examples. From 1940 to 1970, diethylstilbestrol (DES) was prescribed to women in order to reduce the risk of pregnancy complications. DES has subsequently been found to cause an increased risk of hypospadias in both male and female offspring (Brouwers et al., 2006), and effects have been shown to persist in generations following initial exposure, indicating trans-generational inheritance of the an environmentally induced phenotype. This drug has since been banned from consumption. Its conceivable BPA use and exposure will also be restricted.

Aside from drug induced effects, diet and behavior can also impact successive generations. For example, humans conceived during the Dutch famine of 1944-1945 had higher rates of adult obesity (men at age 19 and women at age 50) than those conceived before or after the famine (Ravelli et al., 1999). Obesity has also been shown to be correlated with decreased DNA methylation over the imprinted *IGF2* gene for those conceived from fathers who were *in utero* during the famine (Heijmans et al., 2008). Earlier we discussed how maternal behavior in rats can have a significant impact on 5mC in offspring, and lead to inappropriate stress response (Weaver et al., 2004). Similarly, in humans, abuse as children can lead to increased levels of cytosine methylation over a neuron-specific glucocorticoid receptor promoter and a greater risk of suicide (McGowan et al., 2009).

Epigenetic modifications not only control ES cell differentiation, cell fate determination, and carcinogenesis, but when we consider humans, epigenetic alterations can have broad and long lasting effects on both the individual and the offspring. Since epigenetic modifications can be reliably passed from parent to offspring, the effect of our actions could have a significant impact on multiple future generations. It is therefore supremely important for us, as a scientific community, to study both epigenetic effects and their causal mechanisms. The work presented here may have only helped us to understand small details of

epigenetics as a whole, but hopefully, others will build upon this work, and our

ultimate findings will together lead to better diagnosis, therapeutic treatment,

disease prevention, and human life style alterations.

References:

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P.B., et al. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J. *18*, 1923–1938.

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731–734.

Allis, C.D., Chicoine, L.G., Richman, R., and Schulman, I.G. (1985). Deposition-related histone acetylation in micronuclei of conjugating Tetrahymena. Proc Natl Acad Sci U S A *82*, 8048–8052.

Andrews, D.L. (1989). A unified theory of radiative and radiationless molecular energy transfer. Chemical Physics *135*, 195–201.

Anway, M.D., Cupp, A.S., Uzumcu, M., and Skinner, M.K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science *308*, 1466–1469.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007a). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell *31*, 785–799.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007b). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744–747.

Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K., and Schapira, M. (2012). Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov *11*, 384–400.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. J Exp Med *79*, 137–158.

Barakat, T.S., and Gribnau, J. (2012). X chromosome inactivation in the cycle of life. Development *139*, 2085–2089.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153–155.

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

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396.

Benítez, J.J., Topolancik, J., Tian, H.C., Wallin, C.B., Latulippe, D.R., Szeto, K., Murphy, P.J., Cipriany, B.R., Levy, S.L., Soloway, P.D., et al. (2012). Microfluidic extraction, stretching and analysis of human chromosomal DNA from single cells. Lab Chip.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Birger, Y., Shemer, R., Perk, J., and Razin, A. (1999). The imprinting box of the mouse Igf2r gene. Nature *397*, 84–88.

Blair, R.H., Goodrich, J.A., and Kugel, J.F. (2012). Single-Molecule Fluorescence Resonance Energy Transfer Shows Uniformity in TATA Binding Protein-Induced DNA Bending and Heterogeneity in Bending Kinetics. Biochemistry *51*, 7444–7455.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2003). Nucleosomes unfold completely at a transcriptionally active promoter. Mol. Cell *11*, 1587–1598.

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123–1134.

Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S.R. (2003). Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A *100*, 3960–3964.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. *22*, 1128–1138.

Britton, L.-M.P., Gonzales-Cope, M., Zee, B.M., and Garcia, B.A. (2011). Breaking the histone code with quantitative mass spectrometry. Expert Rev Proteomics *8*, 631–643.

Brouwers, M.M., Feitz, W.F.J., Roelofs, L.A.J., Kiemeney, L.A.L.M., De Gier, R.P.E., and Roeleveld, N. (2006). Hypospadias: a transgenerational effect of diethylstilbestrol? Hum. Reprod. *21*, 666–669.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843–851.

Cao, Z., Song, J.H., Kang, Y.W., Yoon, J.H., Nam, S.W., Lee, J.Y., and Park, W.S. (2010). Analysis of succinate dehydrogenase subunit B gene alterations in gastric cancers. Pathol. Int. *60*, 559–565.

Cardaci, S., and Ciriolo, M.R. (2012). TCA Cycle Defects and Cancer: When Metabolism Tunes Redox State. Int J Cell Biol *2012*, 161837.

Castanotto, D., Tommasi, S., Li, M., Li, H., Yanow, S., Pfeifer, G.P., and Rossi, J.J. (2005). Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells. Molecular Therapy *12*, 179–183.

Cattanach, B.M., and Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. , Published Online: 06 June 1985; | Doi:10.1038/315496a0 *315*, 496–498.

Cerf, A., Tian, H.C., and Craighead, H.G. (2012). Ordered arrays of native chromatin molecules for high-resolution imaging and analysis. ACS Nano *6*, 7928–7934.

Cervera, A.M., Bayley, J.-P., Devilee, P., and McCreath, K.J. (2009). Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells. Mol. Cancer *8*, 89.

Chung, J.H., Whiteley, M., and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell *74*, 505–514.

Cipriany, B.R., Murphy, P.J., Hagarman, J.A., Cerf, A., Latulippe, D., Levy, S.L., Benítez, J.J., Tan, C.P., Topolancik, J., Soloway, P.D., et al. (2012). Real-time analysis and selection of methylated DNA by fluorescence-activated single molecule sorting in a nanofluidic channel. Proc Natl Acad Sci U S A *109*, 8477–8482.

Cipriany, B.R., Zhao, R., Murphy, P.J., Levy, S.L., Tan, C.P., Craighead, H.G., and Soloway, P.D. (2010). Single Molecule Epigenetic Analysis in a Nanofluidic Channel. Anal. Chem. *82*, 2480–2487.

Crick, F. (1956). Ideas on Protein Synthesis. (University College London),.

Crick, F. (1958). On Protein Synthesis. Cambridge University Press, The Symposia of the Society for Experimental Biology *12*, 138–163.

Dalal, Y., Furuyama, T., Vermaak, D., and Henikoff, S. (2007a). Structure, dynamics, and evolution of centromeric nucleosomes. Proc Natl Acad Sci U S A *104*, 15974–15981.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007b). Tetrameric Structure of Centromeric Nucleosomes in Interphase Drosophila Cells. PLoS Biol *5*, e218.

Delaval, K., Govin, J., Cerqueira, F., Rousseaux, S., Khochbin, S., and Feil, R. (2007). Differential histone modifications mark mouse imprinting control regions during spermatogenesis. EMBO J. *26*, 720–729.

Denis, H., Ndlovu, 'Matladi N, and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO Rep *12*, 647–656.

Dobyns, W.B., Filauro, A., Tomson, B.N., Chan, A.S., Ho, A.W., Ting, N.T., Oosterwijk, J.C., and Ober, C. (2004). Inheritance of most X-linked traits is not dominant or recessive, just X-linked. American Journal of Medical Genetics Part A *129A*, 136–143.

Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nature Genetics *41*, 1350–1353.

Dolinoy, D.C., Huang, D., and Jirtle, R.L. (2007). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc. Natl. Acad. Sci. U.S.A. *104*, 13056–13061.

Dorus, E. (1983). X-chromosome inactivation and the study of X-linked dominant transmission of bipolar illness. Arch. Gen. Psychiatry *40*, 698–699.

Doshi, T., D'Souza, C., Dighe, V., and Vanage, G. (2012). Effect of neonatal exposure on male rats to bisphenol a on the expression of DNA methylation machinery in the postimplantation embryo. J. Biochem. Mol. Toxicol. *26*, 337–343.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat. Struct. Mol. Biol. *13*, 713–719.

Eaton, M.L., Prinz, J.A., MacAlpine, H.K., Tretyakov, G., Kharchenko, P.V., and MacAlpine, D.M. (2011). Chromatin signatures of the Drosophila replication program. Genome Res. *21*, 164–174.

Ehrich, M., Nelson, M.R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C.R., Field, J.K., and Van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A *102*, 15785–15790.

Elson, E.L. (2011). Fluorescence Correlation Spectroscopy: Past, Present, Future. Biophysical Journal *101*, 2855–2870.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Essien, K., Vigneau, S., Apreleva, S., Singh, L.N., Bartolomei, M.S., and Hannenhalli, S. (2009). CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. Genome Biol *10*, R131.

Feinberg, A.P. (2007). Phenotypic plasticity and the epigenetics of human disease. Nature *447*, 433–440.

Feinberg, A.P., and Vogelstein, B. (1983a). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89–92.

Feinberg, A.P., and Vogelstein, B. (1983b). Hypomethylation of ras oncogenes in primary human cancers. Biochem. Biophys. Res. Commun. *111*, 47–54.

Ferguson-Smith, A.C., Cattanach, B.M., Barton, S.C., Beechey, C.V., and Surani, M.A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. Nature *351*, 667–670.

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M., and Surani, M.A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. , Published Online: 22 April 1993; | Doi:10.1038/362751a0 *362*, 751–755.

Foquet, M., Korlach, J., Zipfel, W., Webb, W.W., and Craighead, H.G. (2002). DNA fragment sizing by single molecule detection in submicrometer-sized closed fluidic channels. Anal. Chem. *74*, 1415–1422.

Foulks, J.M., Parnell, K.M., Nix, R.N., Chau, S., Swierczek, K., Saunders, M., Wright, K., Hendrickson, T.F., Ho, K.-K., McCullar, M.V., et al. (2012). Epigenetic drug discovery: targeting DNA methyltransferases. J Biomol Screen *17*, 2–17.

Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., et al. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nature Genetics *37*, 391–400.

Gorman, J., Plys, A.J., Visnapuu, M.-L., Alani, E., and Greene, E.C. (2010). Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. Nature Structural & Molecular Biology *17*, 932–938.

Götz, F., Schulze-Forster, K., Wagner, H., Kröger, H., and Simon, D. (1990). Transcription inhibition of SV40 by in vitro DNA methylation. Biochim. Biophys. Acta *1087*, 323–329.

Griffith, J.S., and Mahler, H.R. (1969). DNA ticketing theory of memory. Nature *223*, 580–582.

Hagarman, J.A., Motley, M., Kristjansdottir, K., and Soloway, P. (2013). Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells. PLOS ONE *(in press)*,.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc. Natl. Acad. Sci. U.S.A. *105*, 17046–17049.

Herman, H., Lu, M., Anggraini, M., Sikora, A., Chang, Y., Yoon, B.J., and Soloway, P.D. (2003). Trans allele methylation and paramutation-like effects in mice. Nat. Genet. *34*, 199–202.

Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. Genes Dev. *6*, 2288–2298.

Hogart, A., Lichtenberg, J., Ajay, S.S., Anderson, S., Margulies, E.H., and Bodine, D.M. (2012). Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. Genome Res. *22*, 1407–1418.

Holliday, R., and Pugh, J.E. (1975). DNA Modification Mechanisms and Gene Activity during Development. Science *187*, 226–232.

Holmes, R., Chang, Y., and Soloway, P.D. (2006). Timing and Sequence Requirements Defined for Embryonic Maintenance of Imprinted DNA Methylation at Rasgrf1. Mol Cell Biol *26*, 9564–9570.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA Hypomethylation Coupled to Repressive Chromatin Domain Formation and Gene Silencing in Breast Cancer. Genome Res. *22*, 246–258.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129–1133.

Jackson, V. (1978). Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. Cell *15*, 945–954.

Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat. Genet. *22*, 94–97.

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

Jin, J., Bai, L., Johnson, D.S., Fulbright, R.M., Kireeva, M.L., Kashlev, M., and Wang, M.D. (2010). Synergistic action of RNA polymerases in overcoming the nucleosomal barrier. Nature Structural & Molecular Biology *17*, 745–752.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science *316*, 1497–1502.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr. Biol. *12*, 1360–1367.

Johnson, L., Mollah, S., Garcia, B.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Jacobsen, S.E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res. *32*, 6511–6518.

Jones, L., Hamilton, A.J., Voinnet, O., Thomas, C.L., Maule, A.J., and Baulcombe, D.C. (1999). RNA–DNA Interactions and DNA Methylation in Post-Transcriptional Gene Silencing. Plant Cell *11*, 2291–2301.

Jørgensen, H.F., Adie, K., Chaubert, P., and Bird, A.P. (2006). Engineering a high-affinity methyl-CpG-binding protein. Nucleic Acids Res *34*, e96.

Kalscheuer, V.M., Mariman, E.C., Schepens, M.T., Rehder, H., and Ropers, H.-H. (1993). The insulin–like growth factor type–2 receptor gene is imprinted in the mouse but not in humans. Nature Genetics *5*, 74–78.

Kanda, T., Sullivan, K.F., and Wahl, G.M. (1998). Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. Curr. Biol. *8*, 377–385.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H., et al. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat. Genet. *38*, 149–153.

Komashko, V.M., and Farnham, P.J. (2010). 5-azacytidine treatment reorganizes genomic histone modification patterns. Epigenetics *5*, 229–240.

Kriaucionis, S., and Heintz, N. (2009). The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. Science *324*, 929–930.

Kuo, M.H., Zhou, J., Jambeck, P., Churchill, M.E., and Allis, C.D. (1998). Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. Genes Dev. *12*, 627–639.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development *131*, 839–849.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. *22*, 908–917.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E., and Eisenman, R.N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell *89*, 349–356.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. *13*, 1192–1200.

Lendvai, A., Johannes, F., Grimm, C., Eijsink, J.J.H., Wardenaar, R., Volders, H.H., Klip, H.G., Hollema, H., Jansen, R.C., Schuuring, E., et al. (2012). Genome-wide methylation profiling identifies hypermethylated biomarkers in high-grade cervical intraepithelial neoplasia. Epigenetics *7*,.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. Science *299*, 682–686.

Levy, S.L., and Craighead, H.G. (2010). DNA manipulation, sorting, and mapping in nanofluidic systems. Chem Soc Rev *39*, 1133–1152.

Li, B., Pattenden, S.G., Lee, D., Gutiérrez, J., Chen, J., Seidel, C., Gerton, J., and Workman, J.L. (2005). Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. PNAS *102*, 18385–18390.

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

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. *43*, 1091–1097.

Lindroth, A.M., Park, Y.J., McLean, C.M., Dokshin, G.A., Persson, J.M., Herman, H., Pasini, D., Miró, X., Donohoe, M.E., Lee, J.T., et al. (2008). Antagonism between DNA and H3K27 Methylation at the Imprinted Rasgrf1 Locus. PLoS Genet *4*, e1000145.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523–536.

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

Magde, D., Elson, E.L., and Webb, W.W. (1974). Fluorescence correlation spectroscopy. II. An experimental realization. Biopolymers *13*, 29–61.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Martin, D.I.K., Cropley, J.E., and Suter, C.M. (2011). Epigenetics in disease: leader or follower? Epigenetics *6*, 843–848.

McCabe, M.T., Lee, E.K., and Vertino, P.M. (2009). A multifactorial signature of DNA sequence and polycomb binding predicts aberrant CpG island methylation. Cancer Res. *69*, 282–291.

McGowan, P.O., Sasaki, A., D'Alessio, A.C., Dymov, S., Labonté, B., Szyf, M., Turecki, G., and Meaney, M.J. (2009). Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat. Neurosci. *12*, 342–348.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

Mendel, G. (1865). Versuche über Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd (Experiments in Plant Hybridisation) (Abhandlungen).

Meneghini, M.D., Wu, M., and Madhani, H.D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell *112*, 725–736.

Messerschmidt, D.M., Vries, W. de, Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition. Science *335*, 1499–1502.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Moerner, W.E., and Fromm, D.P. (2003). Methods of single-molecule fluorescence spectroscopy and microscopy. Review of Scientific Instruments *74*, 3597–3619.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. Curr. Biol. *15*, 942–947.

Mullen, A.R., and Deberardinis, R.J. (2012). Genetically-defined metabolic reprogramming in cancer. Trends Endocrinol. Metab. *23*, 552–559.

Murr, R. (2010). Interplay between different epigenetic modifications and mechanisms. Adv. Genet. *70*, 101–141.

Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature *486*, 415–419.

Neely, L.A., Patel, S., Garver, J., Gallo, M., Hackett, M., McLaughlin, S., Nadel, M., Harris, J., Gullans, S., and Rooke, J. (2006). A single-molecule method for the quantitation of microRNA gene expression. Nat. Methods *3*, 41–46.

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

Nolan, R.L., Cai, H., Nolan, J.P., and Goodwin, P.M. (2003). A Simple Quenching Method for Fluorescence Background Reduction and Its Application to the Direct, Quantitative Detection of Specific mRNA. Anal. Chem. *75*, 6236–6243.

Ohm, J.E., Mali, P., Van Neste, L., Berman, D.M., Liang, L., Pandiyan, K., Briggs, K.J., Zhang, W., Argani, P., Simons, B., et al. (2010). Cancer-related epigenome changes associated with reprogramming to induced pluripotent stem cells. Cancer Res. *70*, 7662–7673.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., et al. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat. Genet. *39*, 237–242.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects

unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ONand OFF states by trxG and PcG proteins. Genes Dev. *20*, 2041–2054.

Park, Y.J., Herman, H., Gao, Y., Lindroth, A.M., Hu, B.Y., Murphy, P.J., Putnam, J.R., and Soloway, P.D. (2012). Sequences Sufficient for Programming Imprinted Germline DNA Methylation Defined. PLoS One *7*,.

Pasquardini, L., Potrich, C., Quaglio, M., Lamberti, A., Guastella, S., Lunelli, L., Cocuzza, M., Vanzetti, L., Pirri, C.F., and Pederzolli, C. (2011). Solid phase DNA extraction on PDMS and direct amplification. Lab Chip *11*, 4029–4035.

Periasamy, A. (2001). Fluorescence resonance energy transfer microscopy: a mini review. J. Biomed. Opt *6*, 287–291.

Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., et al. (1996). Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat. Genet. *14*, 106–109.

Pondugula, S., and Kladde, M.P. (2008). Single-Molecule Analysis of Chromatin: Changing the View of Genomes One Molecule at a Time. J Cell Biochem *105*, 330–337.

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., and Trono, D. (2012). The KRAB-ZFP/KAP1 System Contributes to the Early Embryonic Establishment of Site-Specific DNA Methylation Patterns Maintained during Development. Cell Rep *2*, 766–773.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., et al. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol. Cell *44*, 361–372.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201–1212.

Rai, K., Jafri, I.F., Chidester, S., James, S.R., Karpf, A.R., Cairns, B.R., and Jones, D.A. (2010). Dnmt3 and G9a cooperate for tissue-specific development in zebrafish. J. Biol. Chem. *285*, 4110–4121.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011). Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-wide Location Analysis in Human Cells. Cell *147*, 1628–1639.

Ravelli, A.C., Van Der Meulen, J.H., Osmond, C., Barker, D.J., and Bleker, O.P. (1999). Obesity at the age of 50 y in men and women exposed to famine prenatally. Am. J. Clin. Nutr. *70*, 811–816.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-Wide Location and Function of DNA Binding Proteins. Science *290*, 2306–2309.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. Cell *147*, 1408–1419.

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

Robertson, K.D. (2005). DNA methylation and human disease. Nature Reviews Genetics *6*, 597–610.

Rundlett, S.E., Carmen, A.A., Kobayashi, R., Bavykin, S., Turner, B.M., and Grunstein, M. (1996). HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. Proc. Natl. Acad. Sci. U.S.A. *93*, 14503–14508.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983–994.

Sanders, S.L., Portoso, M., Mata, J., Bähler, J., Allshire, R.C., and Kouzarides, T. (2004). Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. Cell *119*, 603–614.

Scarano, E. (1971). The control of gene function in cell differentiation and in embryogenesis. Adv Cytopharmacol *1*, 13–24.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007).

Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet. *39*, 232–236.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

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

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. , Published Online: 19 July 1990; | Doi:10.1038/346240a0 *346*, 240–244.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810–813.

Soloway, P.D., Alexander, C.M., Werb, Z., and Jaenisch, R. (1996). Targeted mutagenesis of Timp-1 reveals that lung tumor invasion is influenced by Timp-1 genotype of the tumor but not by that of the host. Oncogene *13*, 2307–2314.

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

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.-W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., et al. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. Mol. Cell. Biol. *22*, 1298–1306.

Surani, M.A., Barton, S.C., and Norris, M.L. (1987). Influence of parental chromosomes on spatial specificity in androgenetic----parthenogenetic chimaeras in the mouse. Nature *326*, 395–397.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., et al. (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232–235.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930–935.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. , Published Online: 21 August 1975; | Doi:10.1038/256640a0 *256*, 640–642.

Tamaru, H., Zhang, X., McMillen, D., Singh, P.B., Nakayama, J., Grewal, S.I., Allis, C.D., Cheng, X., and Selker, E.U. (2003). Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in Neurospora crassa. Nat. Genet. *34*, 75–79.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016–1028.

Valleix, S., Vinciguerra, C., Lavergne, J.-M., Leuer, M., Delpech, M., and Negrier, C. (2002). Skewed X-chromosome inactivation in monochorionic diamniotic twin sisters results in severe and mild hemophilia A. Blood *100*, 3034–3036.

Verdel, A., and Moazed, D. (2005). RNAi-directed assembly of heterochromatin in fission yeast. FEBS Letters *579*, 5872–5878.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Eynde, A.V., Bernard, D., Vanderwinden, J.-M., et al. (2005). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.

Voigt, P., LeRoy, G., Drury III, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically Modified Nucleosomes. Cell *151*, 181–193.

Waddington, C.H. (1952). The epigenetics of birds (University Press).

Waddington, C.H. (Conrad H. (1939). An introduction to modern genetics (New York, The Macmillan company).

Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P., and Zhang, Y. (2001). Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Mol. Cell *8*, 1207–1217.

Wang, H., Dalal, Y., Henikoff, S., and Lindsay, S. (2008a). Single-epitope recognition imaging of native chromatin. Epigenetics & Chromatin *1*, 10.

Wang, H., Dalal, Y., Henikoff, S., and Lindsay, S. (2008b). Single-epitope recognition imaging of native chromatin. Epigenetics & Chromatin *1*, 10.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. *22*, 1680–1688.

Watanabe, T., Chuma, S., Yamamoto, Y., Kuramochi-Miyagawa, S., Totoki, Y., Toyoda, A., Hoki, Y., Fujiyama, A., Shibata, T., Sado, T., et al. (2011a). MITOPLD Is a Mitochondrial Protein Essential for Nuage Formation and piRNA Biogenesis in the Mouse Germline. Developmental Cell *20*, 364–375.

Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A., et al. (2011b). Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. Science *332*, 848–852.

Waterland, R.A., and Jirtle, R.L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol. Cell. Biol. *23*, 5293–5300.

Watson, J.D., and Crick, F.H.C. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. *2*, 1136–1143.

Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. Nat. Neurosci. *7*, 847–854.

Weinhouse, C., Anderson, O.S., Jones, T.R., Kim, J., Liberman, S.A., Nahar, M.S., Rozek, L.S., Jirtle, R.L., and Dolinoy, D.C. (2011). An expression microarray approach for the identification of metastable epialleles in the mouse genome. Epigenetics *6*, 1105–1113.

Welshons, W.V., Nagel, S.C., and Saal, F.S. vom (2006). Large Effects from Small Exposures. III. Endocrine Mechanisms Mediating Effects of Bisphenol A at Levels of Human Exposure. Endocrinology *147*, s56–s69.

White, A.K., VanInsberghe, M., Petriv, O.I., Hamidi, M., Sikorski, D., Marra, M.A., Piret, J., Aparicio, S., and Hansen, C.L. (2011). High-throughput microfluidic single-cell RT-qPCR. PNAS.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., et al. (2007). Epigenetic stem cell signature in cancer. Nat. Genet. *39*, 157–158.

Wolff, G.L., Kodell, R.L., Moore, S.R., and Cooney, C.A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. FASEB J. *12*, 949–957.

Wu, A.R., Hiatt, J.B., Lu, R., Attema, J.L., Lobo, N.A., Weissman, I.L., Clarke, M.F., and Quake, S.R. (2009). Automated microfluidic chromatin immunoprecipitation from 2,000 cells. Lab Chip *9*, 1365–1370.

Wu, A.R., Kawahara, T.L.A., Rapicavoli, N.A., Van Riggelen, J., Shroff, E.H., Xu, L., Felsher, D.W., Chang, H.Y., and Quake, S.R. (2012). High throughput automated chromatin immunoprecipitation as a platform for drug screening and antibody validation. Lab Chip *12*, 2190–2198.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science *329*, 444–448.

Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F., and Barlow, D.P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature *389*, 745–749.

Yen, R.W., Vertino, P.M., Nelkin, B.D., Yu, J.J., el-Deiry, W., Cumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P., and Baylin, S.B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Res *20*, 2287–2291.

Yokoyama, A., Kitabayashi, I., Ayton, P.M., Cleary, M.L., and Ohki, M. (2002). Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. Blood *100*, 3710–3718.

Yoon, B., Herman, H., Hu, B., Park, Y.J., Lindroth, A., Bell, A., West, A.G., Chang, Y., Stablewski, A., Piel, J.C., et al. (2005). Rasgrf1 Imprinting Is Regulated by a CTCF-Dependent Methylation-Sensitive Enhancer Blocker. Mol Cell Biol *25*, 11184–11190.

Yoon, B.J., Herman, H., Sikora, A., Smith, L.T., Plass, C., and Soloway, P.D. (2002). Regulation of DNA methylation of Rasgrf1. Nature Genetics *30*, 92–96.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750–756.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125–129.