Summer 8-15-2017

Designing Epigenome Editing Tools to Understand the Functional Role of DNA Methylation Changes in Cancer

James Mcdonald
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Genetics and Genomics

Dissertation Examination Committee:
John Edwards, Chair
Grant Challen
Sheila Stewart
Ting Wang
Jason Weber

Designing Epigenome Editing Tools to Understand the Functional Role of DNA Methylation Changes in Cancer

by
James McDonald

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2017
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

I thank my advisor, Dr. John Edwards, for his guidance that helped me grow as a scientist and his encouragement that helped me to keep a positive outlook concerning my work. He also provided strong support as I pursued careers goals that took met outside the lab. His sound advice has help me to become the scientist I am today.

I also thank all past and present members of the Edwards Lab. The ideas you have contributed in discussion and your constructive criticism of my work have helped to improve it. Your encouragement helped me to overcome the many challenges that accompany dissertation research. In addition, your friendship made the lab an enjoyable place to work.

I thank my thesis committee for their advice. Your guidance to abandon a failing project as well as the experiments you suggested have improved the quality of the work presented in this document.

I thank my family for their support and encouragement both before and during my dissertation research. Without my parents' early prodding to explore different career options, I may never have become a scientist. Having become a scientist, their encouragement has only helped to make me a better one.

I thank the Division of Biology and Biomedical Sciences for their administrative support. I especially thank the Molecular Genetics and Genomics program coordinator, Melanie Relich, for her help in meeting the procedural requirements of the program and for all the things she does for the program behind the scenes.

I thank Timothy H. Bestor and Bill Jack for the sequence of the M.SssI bacterial methyltransferase. I thank the Genome Engineering and iPSC Center at Washington University

School of Medicine for production and validation of sgRNAs. I thank Jessica Hoisington-Lopez for sequencing assistance. I also thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital for the use of the Siteman Flow Cytometry Core, which provided cell sorting and analysis services.

Various funding sources have made my dissertation research possible, especially the T32 CMB training grant that funded the first three years of my research (NIH 2T32GM007067-37) and the DBBS travel grant that allowed me to attend the 2016 Cold Spring Harbor Epigenetics and Chromatin meeting. Additional sources of funding include: National Institutes of Health [NIDDK R01DK102428], the Edward Mallinckrodt Jr. Foundation, the American Society of Hematology, Alex's Lemonade Stand Foundation for Childhood Cancer, the Children's Discovery Institute, the Sidney Kimmel Foundation for Cancer Research and the V Foundation for Cancer Research (to G.A.C.), the Siteman Cancer Center, U.S. Department of Defense Congressionally Directed Medical Research Program for Breast Cancer [W81XWH-11-1-0401], and the National Institute of General Medicine Sciences [NIGMS 5R01GM108811, NLM R21LM011199].

James McDonald

Washington University in St. Louis

August 2017

ABSTRACT OF THE DISSERTATION

Designing Epigenome Editing Tools to Understand the Functional Role

of DNA Methylation Changes in Cancer

by

James McDonald

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2017

Professor John Edwards, Chair

DNA methylation is known to silence gene expression in the context of imprinting, X-chromosome inactivation, and retrotransposon silencing. However, the role of DNA methylation in silencing gene expression outside of these contexts is not fully understood. This is especially true in diseases such as cancer, where normal DNA methylation patterns are significantly altered. In breast cancer as well as nearly all cancer types, most of the genome loses DNA methylation while small regions of the genome gain methylation. DNA methylation generally correlates with decreased gene expression when present at a gene promoter. Therefore, these regions of hypo- and hyper-methylation may contribute to cancer development and progression by activating oncogenes or silencing tumor suppressor genes. My work focuses on building tools to study the functional role of DNA methylation changes and exploring how methylation changes at a gene promoter promote resistance to treatment in breast cancer.

About 75% of breast cancers depend on estrogen signaling through the estrogen receptor (ERα). These tumors are effectively treated by aromatase inhibitors (AI) that prevent estrogen

production. However, almost all advanced cases of ERα positive breast cancer develop resistance to AI therapy. I therefore sought to identify methylation changes that promote this resistance. I studied *UCA1* and *PTGER4*, two genes identified by a screen for negatively correlated methylation and expression changes in a cell line model of AI resistance. *UCA1* is a long non-coding RNA that promotes growth and metastasis in bladder cancer. *PTGER4* encodes the prostaglandin E_2 receptor 4 (EP4), which supports the progression of multiple cancer types by altering cell signaling. While my experiments did not indicate that *UCA1* has a strong role in AI resistance, I found that hypomethylation of the *PTGER4* promoter correlates with increased expression and EP4 signaling. My data further suggest that the downstream effector of EP4 signaling, CARM1, promotes endocrine therapy resistance by increasing the ligand-independent transcription activity of ERα.

The effects of local DNA methylation changes are most often identified by correlating the methylation and expression levels from two samples. To show that methylation causes the expression change, these studies rely on non-specific tools that demethylate the whole genome: DNA methyltransferase (DNMT) inhibitors or by DNMT knockout/knockdown. To build a tool capable of inducing site-specific DNA methylation changes, I fused the human DNMT3A catalytic domain to the RNA-guided nuclease Cas9 (the Cas9 is nuclease dead). I used this tool to induce up to 53% DNA methylation on individual cytosines within 50 bp of the target site. When multiple sites within the *CDKN2A* or *ARF* promoters were targeted, the induced DNA methylation decreased the expression of the targeted gene. To determine the optimal DNMT catalytic domain to use in this system, I created alternative DNMT fusions that included human DNMT1, a fusion of mouse Dnmt3a to mouse Dnmt3L, human DNMT3B, and the bacterial methyltransferase M.SssI. While the Dnmt3a-Dnmt3L fusion increased methylation relative to

DNMT3A alone, it also induced more off-target methylation. The continued development of targeted DNA methylation technologies will increase our ability to identify functional methylation changes in tumors. As a result, we will learn the specific ways that methylation-induced gene expression changes contribute to cancer.

Chapter 1: Introduction

DNA methylation is an epigenetic mark that affects gene expression without changing the DNA sequence. The presence of DNA methylation at a gene's promoter correlates with decreased expression of that gene. DNA methylation is known to silence gene expression in the context of imprinting, X-chromosome inactivation, and retrotransposon silencing, but its role in silencing gene expression outside of these contexts is not fully understood (Edwards *et al.*, 2017). Indeed, it is not clear if DNA methylation initiates gene silencing or whether DNA methylation is added after gene silencing has already occurred. The work described in my thesis is aimed at further investigating this central question.

Almost all cancers develop aberrant focal hypermethylation in a background of genome-wide hypomethylation (Jones and Baylin, 2007). These regions of hypo- and hyper-methylation may contribute to cancer development and progression by activating oncogenes or silencing tumor suppressor genes, respectively (Ehrlich, 2009; Jones and Baylin, 2007). These DNA methylation changes can be used to detect cancer (Warton and Samimi, 2015) and identify specific cancer subtypes (Ciriello *et al.*, 2013; The Cancer Genome Atlas Network, 2012). Increased understanding of the specific ways that DNA methylation affects gene expression will allow physicians to refine the application of methylation information to the detection, diagnosis, and treatment of cancer (Warton *et al.*, 2016). Here I will review what is known concerning DNA methylation, the mechanisms for its establishment and maintenance, as well as its role in cancer.

1.1 DNA Methylation

My thesis focuses on the primary form of DNA methylation in humans, the addition of a methyl group (CH_3) to the 5 position of cytosine residues (5mC). Information on DNA methylation of other residues or in other organisms can be found in the following reviews: Luo *et al.* (2015); Suzuki and Bird (2008). Mammalian DNA methylation was first discovered in a calf thymus preparation by Hotchkiss in 1948 (Hotchkiss, 1948). In 1975, Holliday and Pugh as well as Riggs proposed models for the role of DNA methylation in regulating gene expression during development and X-inactivation, respectively (Holliday and Pugh, 1975; Riggs, 1975). Since that time, DNA methylation has been intensely studied as an epigenetic mark that participates with DNA sequence and other factors to regulate gene expression.

Human DNA methylation occurs most often on cytosine residues in the sequence context 5'–CG–3', which is known as a CpG site (Ramsahoye *et al.*, 2000; Ziller *et al.*, 2011). The symmetrical nature of these sites allows DNA methylation to be copied in a semi-conservative manner during DNA replication. The result is that DNA methylation can be stably maintained through cell division and its effects are thereby heritable (Holliday and Pugh, 1975; Riggs, 1975). There are about 28 million CpG sites in the human genome, and about 70% of these sites are methylated (Eckhardt *et al.*, 2006; Edwards *et al.*, 2010). CpG sites are unevenly distributed throughout the genome with a fraction of all CpGs occurring in ~1,000 bp clusters called CpG islands (CGI). This uneven distribution of CpG sites was created by the spontaneous deamination of methylated cytosine which creates a thymine residue (Bird, 1980; Coulondre *et al.*, 1978) as well as a T-G mismatch. The base excision repair machinery cannot identify the correct base, which results in a 50% chance for a C to T mutation if the G is excised for repair. Despite the

higher density of potential methylation sites, CGIs remain largely unmethylated and thereby suffer fewer C to T mutations (Bird *et al.*, 1985; Deaton and Bird, 2011). This then is the typical methylation pattern in human tissues: most of the genome is methylated but CpG poor while the CpG rich CGIs remain unmethylated.

1.1.1 DNA Methyltransferases

The enzymes primarily responsible for creating the DNA methylation patterns described above are the DNA methyltransferases (DNMTs). The catalytic domain of eukaryotic DNMTs adopt an AdoMet-dependent MTase fold: seven beta strand and alpha helix units that fold to place the beta sheet between the outward facing alpha helices (Jurkowska and Jeltsch, 2016). The methylation reaction proceeds roughly as follows. Upon recognizing a CpG target, the DNMT flips the base out (Klimasauskas *et al.*, 1994). The cytosine ring is then chemically altered to promote the methylation reaction. A conserved cysteine located in motif IV (Fig. 1.1) covalently

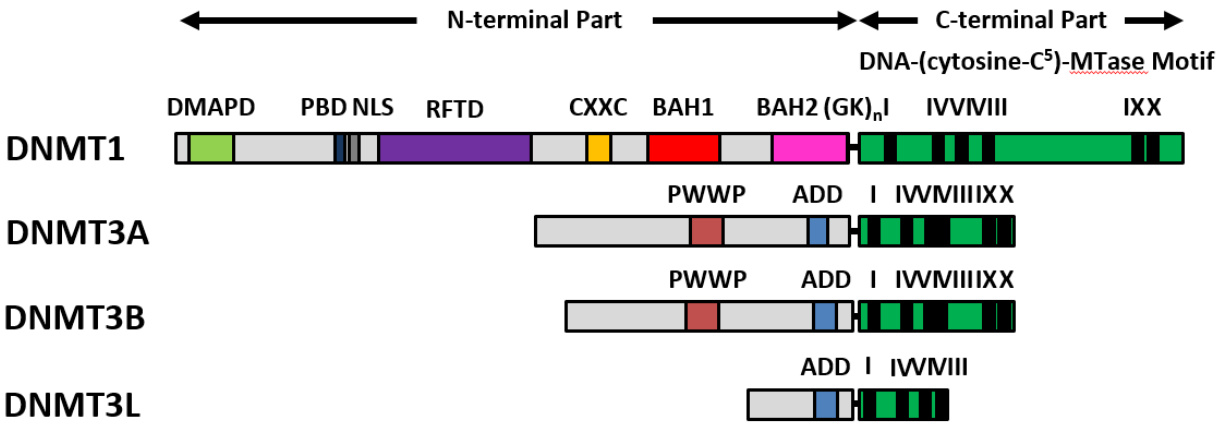


Figure 1.1. Domain structure of the mammalian DNA methyltransferases. Domain abbreviations: DMAPD, DNA methyltransferase associated protein 1 interacting domain; PBD, PCNA binding domain; NLS, nuclear localization signal; RFTD, replication foci targeting domain; CXXC, CXXC domain; BAH1 and BAH2, bromo-adjacent homology domains 1 and 2; GK_n, glycine-lysine repeats; PWWP, PWWP domain; ADD, ATRX-DNMT3-DNMT3L domain. Figure adapted from Jurkowska and Jeltsch (2016).

bonds to carbon six of the cytosine. In addition, a glutamate in motif VI (Fig. 1.1) donates its hydrogen to nitrogen three of the cytosine. These steps are supported by arginine residues located in motif VIII (Fig. 1.1, Lukashevich *et al.*, 2016). These steps cause electron density to accumulate on carbon five, which then forms bonds with the methyl group from S-adenosyl-L-methionine (AdoMet). Both the cysteine and glutamate are necessary for enzymatic activity, and their mutation creates a catalytically dead enzyme (Gowher *et al.*, 2006; Reither *et al.*, 2003). Several dissociation steps follow leaving a methylated base.

The DNMTs share a conserved catalytic domain, but their specific targets and functions are regulated by protein-protein interactions. For example, Dnmt1 is 4- to 75-fold more catalytically active at hemimethylated CpG sites where only one of the two cytosine residues is methylated (Okano *et al.*, 1998; Song *et al.*, 2011). Hemimethylated sites are created during DNA replication; therefore, DNMT1 is important for maintenance of DNA methylation patterns when cells divide. DNMT1 has several domains that support it in this role. The PCNA binding domain (PBD, Fig. 1.1) recruits DNMT1 to bind to replication forks (Chuang *et al.*, 1997). The replication foci targeting domain (RFTD, Fig. 1.1) interacts with Uhrf1, which also recruits Dnmt1 to replicated, hemimethylated DNA (Bostick *et al.*, 2007; Jeltsch, 2008; Sharif *et al.*, 2007). Uhrf1 also releases the autoinhibition of Dnmt1 caused by the interaction between the catalytic and RFT domains (Bashtrykov *et al.*, 2014; Berkyurek *et al.*, 2014; Takeshita *et al.*, 2011). When bound to unmethylated DNA, the CXXC domain (Fig. 1.1) blocks catalytic activity of a truncated Dnmt1 fragment (Song *et al.*, 2011), though this was not found to be the case with a full-length Dnmt1 (Bashtrykov *et al.*, 2012). In addition to these interactions, DNMT1's overall structure also supports its role in maintenance of DNA methylation patterns. The enzyme

wraps itself around the DNA, allowing it to be processive and focus on methylating one strand: the unmethylated strand produced during DNA replication (Hermann *et al.*, 2004; Song *et al.*, 2012). Information about additional interactions that affect DNMT1 function can be found in this excellent review (Jurkowska and Jeltsch, 2016).

The role of DNMT1 in maintenance methylation is confirmed by several other studies as well. DNMT1 has been cloned from both mouse (Bestor *et al.*, 1988) and human (Ramchandani *et al.*, 1998; Yen *et al.*, 1992; Yoder *et al.*, 1996). Knockout of DNMT1 is embryonic lethal in mice (Li *et al.*, 1992), though both mouse and human stem cells can survive without DNMT1 (Li *et al.*, 1992; Liao *et al.*, 2015; Tsumura *et al.*, 2006). Conditional knockout in differentiated cells also causes cell death (Fan *et al.*, 2001; Jackson-Grusby *et al.*, 2001; Sen *et al.*, 2010; Trowbridge *et al.*, 2009), including the HCT116 colorectal carcinoma cell line (Chen *et al.*, 2007). The importance of DNMT1 during DNA replication is further underscored by the finding that UHRF1 knockout produces a similar phenotype to DNMT1 knockout (Sharif *et al.*, 2007).

In contrast to DNMT1, the DNMT3 family of DNMTs methylate unmethylated CpG sites. This family contains three members: DNMT3A, DNMT3B, and DNMT3L. However, DNMT3L contains several amino acid substitutions in the catalytic domain that render it catalytically inactive (Jurkowska and Jeltsch, 2016). Instead, it acts a cofactor for DNMT3A in a manner described below. Additionally, a new member of the DNMT3 family, Dnmt3c, has been recently discovered (Barau *et al.*, 2016). This enzyme appears to have arisen from a rodent-specific duplication of Dnmt3b.

Consistent with their role as *de novo* DNMTs, Dnmt3a and Dnmt3b methylate both hemimethylated and unmethylated DNA (Aoki *et al.*, 2001; Gowher and Jeltsch, 2001; Okano *et*

al., 1998). The activity of both DNMT3A and DNMT3B is guided by the ADD and PWWP domains (Fig. 1.1). The PWWP domain binds to trimethylated histone 3 lysine 36 (H3K36me3), which targets DNMT3A and DNMT3B to methylate pericentric satellite repeats (Chen *et al.*, 2004; Ge *et al.*, 2004). Mutations in this domain prevent this interaction and have been linked with immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome (Ge *et al.*, 2004; Shirohzu *et al.*, 2002). The ADD domain is capable of myriad interactions that integrate the DNMTs with the cell's chromatin modification machinery. These interactions include unmethylated H3K4 (Guo *et al.*, 2015; Ooi *et al.*, 2007; Otani *et al.*, 2009; Zhang *et al.*, 2010b), SETDB1 (Li *et al.*, 2006), HP1β, and SUV39H1 (Fuks *et al.*, 2003). For a complete list of interactions, see the Jurkowska and Jeltsch (2016) review. Similar to the CXXC or RFTD domains of DNMT1, the ADD domain causes autoinhibition by blocking access to the catalytic domain (Guo *et al.*, 2015).

The DNMT3 family of proteins also form both homo- and hetero-multimeric structures that contribute to their function. The proteins interact at two regions: an arginine-aspartate (RD) interface near the DNA binding site and a phenylalanine-phenylalanine interface on the opposite side of the catalytic domain (Jia *et al.*, 2007). Dnmt3a binds to DNA in a non-specific manner (Rajavelu *et al.*, 2012) where it forms a homodimer on DNA via interaction at the RD interface. This initial binding event promotes cooperative binding (Emperle *et al.*, 2014; Jia *et al.*, 2007; Jurkowska *et al.*, 2008, 2011; Rajavelu *et al.*, 2012). These homodimers can methylate two CpG sites on opposite strands if the two sites are separated by about 1 helical turn apart (Jia *et al.*, 2007; Jurkowska *et al.*, 2008). While Dnmt3L lacks an RD domain (Jurkowska *et al.*, 2011), it retains an FF domain and binds on either side of the DNMT3A homodimer to create a Dnmt3L-

Dnmt3a-Dnmt3a-Dnmt3L heterotetramer (Jia *et al.*, 2007). This interaction increases Dnmt3a activity by increasing the affinity of Dnmt3a for both DNA and AdoMet (Chédin *et al.*, 2002; Gowher *et al.*, 2005; Kareta *et al.*, 2006; Suetake *et al.*, 2004). If a DNMT3A homodimer interacts with the FF interface of another DNMT3A homodimer, DNMT3A filaments can be formed on multiple parallel DNA strands. This interaction aids DNMT3A activity on organized heterochromatin. DNMT3L acts to break up the DNMT3A homomultimer filaments, perhaps potentially helping to redistribute it (Jurkowska *et al.*, 2011).

DNMT3B possesses similar interface sites, making it capable of forming RD interface homodimers like DNMT3A. However, mutation of the interacting domains in Dnmt3b does not affect the function of Dnmt3b and indicates multimerization is not important for its function. Instead, Dnmt3b may also methylate DNA in a processive manner (Norvil *et al.*, 2016). Two reports indicate processive activity for DNMT3A by pulse-chase experiments (Holz-Schietinger and Reich, 2010; Holz-Schietinger *et al.*, 2011). However, long turnover in pulse-chase experiments is consistent with slow dissociation of DNMT3A oligomers. Similarly, processive activity is not consistent with an increase in catalytic activity due to the addition of a catalytically dead Dnmt3a (Emperle *et al.*, 2014). The presence of a bias toward methylation of CpGs spaced every 8-10 bps is inconsistent with processive activity. The presence of this pattern *in vivo* (Jia *et al.*, 2007) therefore indicates that the cooperative mechanism for DNMT3A is correct.

Despite having similar targeting and interacting domains, DNMT3A and DNMT3B serve slightly different functions *in vivo*. Dnmt3a and Dnmt3b have been cloned (Okano *et al.*, 1998), and subsequent knockout studies have revealed the differences in function. Dnmt3b knockout mice die before birth, while Dnmt3a knockout mice die shortly after birth (Okano *et al.*, 1999).

In these mice, a Dnmt3b knockout specific loss of methylation at centromeric satellite DNA was observed. The specific role of Dnmt3a was revealed by Dnmt3L knockout. Dnmt3L is expressed in few tissues, and knockout causes almost no phenotype except faulty gametogenesis (Hata *et al.*, 2002). Male mice are sterile due to loss of methylation on and reactivation of retrotransposon sequences (Bourc'his and Bestor, 2004). Female mice fail to form the correct maternal imprints, which leads to the death of embryos *in utero* (Hata *et al.*, 2002). These are the same phenotypes as a germline specific Dnmt3a knockout (Kaneda *et al.*, 2004) and indicate that Dnmt3a and Dnmt3b are responsible for methylating different groups of repetitive elements. This difference in function could arise from the differences in the region N-terminal of the PWWP domain (Fig. 1), which is the least conserved part of the enzymes (Jurkowska and Jeltsch, 2016).

1.1.2 Maintenance and *de novo* Methylation are not Separate Roles

In the preceding section, I describe DNMT1 as solely responsible for maintenance methylation and DNMT3A/3B as solely responsible for *de novo* methylation. However, a review of the relevant data indicates that the truth is not so simple (Jeltsch and Jurkowska, 2014; Jones and Liang, 2009). The core of the problem is that neither group of enzymes does its job perfectly. DNMT1's preference for hemimethylated DNA is not strong enough to ensure perfect methylation transfer to the newly synthesized DNA strand (Jeltsch and Jurkowska, 2014; Jones and Liang, 2009). Similarly, the dimerization of DNMT3A and DNMT3B places the catalytic sites at two CpGs on opposite strands. Because of this, DNMT3A and DNMT3B frequently fail to fully methylate unmethylated CpG sites. While imperfect, the proficiencies of DNMT1 can compensate for the deficiencies of DMT3A/B and vice versa. Indeed, either *Dnmt1*$^{-/-}$ or *Dnmt3a*$^{-/-}$ *Dnmt3B*$^{-/-}$ double knockout mouse embryonic stem lose repetitive sequence

methylation, indicating that both systems are needed for consistent methylation maintenance (Liang *et al.*, 2002). All of these observations support the hypothesis that the combined activities of DNMT1 and DNMT3A/B cooperate to maintain the overall methylation level in a given region and not the specific methylation pattern (Jeltsch and Jurkowska, 2014; Jones and Liang, 2009).

1.1.3 Non-CpG Methylation

Despite their preference for CpG sites, DNMT3A and DNMT3B induce non-CpG methylation (Aoki *et al.*, 2001; Gowher and Jeltsch, 2001; Ramsahoye *et al.*, 2000). DNMT1 knockout does not affect non-CpG methylation levels indicating that the *de novo* methyltransferase are responsible (Ramsahoye *et al.*, 2000). DNMT3A/B induce non-CpG methylation preferentially: CpA > CpT > CpC (Aoki *et al.*, 2001; Gowher and Jeltsch, 2001; Laurent *et al.*, 2010; Ramsahoye *et al.*, 2000). Non-CpG methylation occurs most frequently in embryonic stem cells (Lister *et al.*, 2009; Ziller *et al.*, 2011), induced pluripotent stem cells (Lister *et al.*, 2011), neurons (Guo *et al.*, 2014; Lister *et al.*, 2013; Varley *et al.*, 2013), and oocytes (Shirane *et al.*, 2013); however, very low levels of non-CpG methylation can be detected in various differentiated tissues (Schultz *et al.*, 2015). There is some evidence that non-CpG methylation is functional and not noise. The *PGC-1α* and *PDK4* promoters gain non-CpG methylation in the skeletal muscle of type II diabetes patients, and this non-CpG hypermethylation correlates with decreased expression (Barrès *et al.*, 2009; Barres *et al.*, 2013). This might be accomplished through altering transcription factor binding; non-CpG methylation at Sp transcription factor binding sites of the *STY11* gene blocks transcription factor binding and prevents expression (Inoue and Oishi, 2005).

1.2 DNA Hydroxymethylation

5-hydroxymethylcytosine (5hmC) is an oxidized derivative of 5mC that plays a role in the demethylation of 5mC (Ito *et al.*, 2010; Tahiliani *et al.*, 2009). The Ten-Eleven Translocase (TET) family of enzymes catalyze the formation of 5hmC by the addition of a hydroxyl group to the methyl group of 5mC. TET enzymes are also capable of further oxidizing 5hmC to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC, He *et al.*, 2011; Ito *et al.*, 2011). 5hmC is present at <1% of all cytosine residues in most tissues (Globisch *et al.*, 2010; Ito *et al.*, 2011), and almost always in the CpG context (Lister *et al.*, 2013; Yu *et al.*, 2012). The greatest prevalence of 5hmC occurs in neurons at 40% the level of 5mC (Kriaucionis and Heintz, 2009). In other tissues, 5hmC is at levels about 4 to 6% that of 5mC (Globisch *et al.*, 2010; Ito *et al.*, 2011). 5fC and 5caC are almost completely absent from all tissues tested. Embryonic stem cell showed the highest levels of 5fC and 5caC at 20 or 3 modified residues per million cytosines, respectively (Ito *et al.*, 2011).

When it does occur, 5hmC is common at the promoters of poised genes, gene bodies, and enhancers and is rare within CGI promoters (Wu and Zhang, 2017). CGI promoters lack 5hmC because these promoters also lack 5mC, the TET enzymes' substrate. Instead, 5hmC is found on poised promoters which have both lower CpG density and higher 5mC (Williams *et al.*, 2011; Wu *et al.*, 2011). 5hmC concentration also increases along gene bodies with a bias toward the sense strand. This suggests that TET enzyme activity may be linked to active transcription (Wen *et al.*, 2014). Enhancer and insulator 5hmC occurs on cytosines flanking transcription factor and CTCF binding sites (Stroud *et al.*, 2011). This suggests that TET enzymes may be recruited by

transcription factors. However, it may also represent residual activity of TET enzymes at the promoter(s) that interact with the enhancer (Wu and Zhang, 2017).

1.2.1 The TET Enzymes

TET1, *TET2*, and *TET3* represent paralogs that all share a conserved and active C-terminal iron(II)- and 2-oxoglutarate (2OG)-dependent hydroxylase domain (Iyer *et al.*, 2009; Zhang *et al.*, 2010a). The TET enzymes also contain a conserved, N-terminal CXXC zinc finger DNA-binding domain. TET2, however, lacks the CXXC domain (Iyer *et al.*, 2009). This loss was apparently caused by a chromosomal inversion that allowed the *TET2* CXXC sequence to become its own gene, *IDAX*. IDAX aids TET2 binding to DNA, but in the process also triggers caspase based degradation of TET2 (Ko *et al.*, 2013). Interestingly, the Tet1 CXXC domain does not bind DNA and is not necessary for catalysis (Frauer *et al.*, 2011). Both mouse Tet1 and *Xenopus* Tet3 CXXC domains bind methylated and unmethylated CpGs (Xu *et al.*, 2011, 2012). This is unusual for a domain that typically binds unmethylated CpG sites and may be caused by the loss of a KFGG motif that is maintained in the CXXC domains of MLL and CFP1 (Xu *et al.*, 2012). Once bound to the DNA, however, the catalytic domain shows a preference for 5mC in the CpG context. However, the DNA binding interactions do not contact the methyl group. This allows TET proteins to bind multiple derivatives of 5mC and catalyze their oxidation (Hu *et al.*, 2013a). Nevertheless, constraints at the active site give TET enzymes render them more catalytically active on 5hmC than on either 5fC or 5caC (Hu *et al.*, 2015).

1.2.2 DNA Demethylation

The most straightforward way to demethylate DNA is to undergo multiple cell divisions while DNMT activity is inhibited. The oxidation of 5mC to 5hmC encourages this form of

demethylation. DNMT1 is 60-fold more active on hemimethylated DNA (hemi-5mC) DNA than on hemihydroxymethylated DNA (hemi-5hmC, Hashimoto *et al.*, 2012; Valinluck and Sowers, 2007). The result is inhibition of maintenance methylation and dilution of 5mC by cell division. Two mechanisms may counter this loss of activity. First, UHRF1 can bind to hemi-5hmC and recruit DNMT1 (Hashimoto *et al.*, 2012). Second, truncated isoforms of DNMT3A and DNMT3B remain active on hemi-5hmC DNA and can take over the maintenance role at these sites (Hashimoto *et al.*, 2012). This latter mechanism is supported by the observation that $Dnmt3a^{-/-}$ $Dnmt3b^{-/-}$ mouse ESCs gradually lose 5mC (Chen *et al.*, 2003). However, demethylation in the early embryo occurs in the presence of DNMT3A, suggesting that DNMT3A alone may not be sufficient to maintain methylation patterns (Wu and Zhang, 2014).

Alternatively, DNA methylation can be actively removed without DNA replication. Several mechanism have been proposed for this (Wu and Zhang, 2014); however, thymine-DNA-glycosylase (TDG) base excision and subsequent repair is most supported by the evidence (Rasmussen and Helin, 2016). In this pathway, TDG detects 5fC and 5caC and catalyzes the removal of the modified cytosine base (He *et al.*, 2011; Maiti and Drohat, 2011). The base excision repair pathway detects the damaged DNA and repairs it with an unmethylated cytosine residue. This mechanism is supported by the observation that overexpression of TDG in HEK293 cells depletes 5fC and 5caC (Nabel *et al.*, 2012). Initial immunological studies of early embryos found that DNA methylation of the paternal genome occurs before DNA replication, implying active demethylation (Mayer *et al.*, 2000; Oswald *et al.*, 2000). However, TDG expression is low in the early embryo (Tang *et al.*, 2011), and 5fC and 5caC are diluted over cell division (Inoue *et al.*, 2011). Altogether, it appears that DNA methylation is generally lost passively through DNA

replication, but specific regions may be targeted for active demethylation by TDG and base excision repair.

1.3 DNA Methylation and Gene Regulation

The presence of DNA methylation at a gene's promoter correlates with decreased expression of that gene. However, there are many genes that are still expressed despite the presence of DNA methylation (Weber *et al.*, 2007). The effect of DNA methylation on gene expression is position and context dependent. This can be demonstrated by the stronger correlations between methylation and expression in first exons (Schlosberg *et al.*, 2017; VanderKraats *et al.*, 2013) or at CGI shores (Irizarry *et al.*, 2009). A large part of this control is directed by the transcription factors and methyl binding domain proteins that directly interact with DNA methylation.

1.3.1 DNA Methylation Alters Transcription Factor Binding

The methyl group on methylated cytosine protrudes into the major groove of the DNA (Li and Zhang, 2014). DNA methylation is therefore able to alter the target site for transcription factors (TFs, Hu *et al.*, 2013b; Yin *et al.*, 2017) and exert direct control of gene expression. An early model was that methylation of TF binding sites (TFBS) blocked TF binding and caused gene repression (Tate and Bird, 1993). This principle has been demonstrated for MLTF (Watt and Molloy, 1988), YY1 (Gaston and Fried, 1995), and the E2F family of transcription factors (Campanero *et al.*, 2000). Indeed, TFBS overall tend to be unmethylated (Choy *et al.*, 2010), and SP1 protects the promoters to which it binds from being methylated (Brandeis *et al.*, 1994; Lienert *et al.*, 2011; Macleod *et al.*, 1994).

Further study has shown that the model described above applies in a limited number of cases. Of 519 TFs with motifs that contain CpGs, 202 (39%) of TFs bound irrespective of DNA methylation (Yin *et al.*, 2017). DNA methylation could not affect the expression of genes controlled by these TFs. The subset of 317 (61%) TF motifs affected by DNA methylation contained 117 TFs (37%) inhibited by DNA methylation, 175 TFs (55%) stimulated by DNA methylation, and 25 TFs (8%) with multiple effects due to CpG sites at multiple positions in their motif(s) (Yin *et al.*, 2017). This suggests that methylated motifs are capable of recruiting transcription activators as well as repelling them. Most importantly, the CpG sites that correlate with gene expression were found outside of TF motifs. This suggests that – on average – even if DNA methylation modifies TF binding, it will not affect gene expression (Medvedeva *et al.*, 2014). As a result, it is difficult to make a general statement about the effect of DNA methylation on gene expression via control of TF binding.

1.3.2 Methylation Binding Domain Proteins Directly Read Methylation Information

Methyl binding domain (MBD) protein family members represent a subset of DNA binding proteins that have a strong affinity for methylated CpG sites (Du *et al.*, 2015). This affinity is conferred by the 70 – 85 amino acid MBD. This family includes the founding member, MeCP2 (Meehan *et al.*, 1989), as well as MBD1, MBD2, MBD3, MBD4, MBD5, and MBD6, which were discovered by their homology to MeCP2 (Du *et al.*, 2015). The MBD domain preference for methylated CpG sites can range from about 3-fold up to over 100-fold (Fraga *et al.*, 2003). MBD proteins also prefer to bind methylated DNA with high CpG density (Baubec *et al.*, 2013; Cramer *et al.*, 2014; Fraga *et al.*, 2003). This suggests that MBD proteins are key readers of DNA methylation in the genome.

When bound to DNA, MBD proteins decrease gene expression by recruiting corepressors (Du *et al.*, 2015). For example, MeCP2 recruits both histone deacetylates (Jones *et al.*, 1998; Nan *et al.*, 1998) and HP1 (Agarwal *et al.*, 2007; Singleton *et al.*, 2011). HP1 in turn recruits SUV39H1, an H3K9 methyltransferase (Fujita *et al.*, 2003). Both HP1 and SUV39H1 are capable of interacting with the DNMT3 enzymes via their ADD domains (Fuks *et al.*, 2003; Jurkowska and Jeltsch, 2016). Similarly, MBD2 and MBD3 both recruit the NuRD complex to DNA (Guezennec *et al.*, 2006). The NuRD complex represses gene expression via nucleosome remodeling and histone deacetylation (Lai and Wade, 2011). The MBD2-NuRD complex reinforces the repression already established by DNA methylation at the *CDKN2A* (Magdinier and Wolffe, 2001) and *GSTP1* (Lin and Nelson, 2003) promoters. Altogether, these observations suggest that MBD proteins often serve to reinforce gene silencing at methylated regions of the genome by promoting the formation of heterochromatin.

Reinforcement of DNA methylation-induced gene silencing is not the only role of MBD proteins, however. MBD3 contains mutations that have eliminated its preference for 5mC (Saito and Ishikawa, 2002). Instead, MBD3 can bind 5hmC and repress genes in a 5hmC dependent manner (Yildirim *et al.*, 2011). MBD4 contains a thymine DNA glycosylase domain that allows it play a role in demethylation via repair of mismatches resulting for deaminated cytosine (Cunha *et al.*, 2014; Hendrich *et al.*, 1999). Lastly MeCP2 has roles in DNA looping that affect *H19/Igf2* imprinting (Kernohan *et al.*, 2010, 2014). These alternative functions of MBD proteins offer new and unexplored ways that DNA methylation can affect genome function.

1.3.3 Imprinting is Controlled by DNA methylation

The readers of DNA methylation play a direct role in control of monoallelic gene expression, also known as imprinting. The expression of imprinted genes is decided by the status imprinting control regions (ICRs), and DNA methylation regulates ICR activity (Peters, 2014). The *IGF2* and *H19* genes are imprinted and their ICR lies between them. A downstream enhancer can activate either gene. DNA methylation of the ICR controls the binding of the sequence-specific zinc finger CCCTC-binding factor (CTCF): CTCF binds the unmethylated maternal ICR but fails to bind the methylated paternal ICR. CTCF blocks the access of the enhancer to *IGF2* on the maternal allele and thereby activates maternal *H19* expression. On the other chromosome, ICR methylation prevents the enhancer from activating paternal *H19* expression, allowing it to activate *IGF2* (Peters, 2014). Dysregulation of methylation at the ICR of the *H19/IGF2* locus contributes to diseases such as Beckwith-Wiedemann syndrome (Bliek *et al.*, 2001; Weksberg *et al.*, 2001). This illustrates the role DNA methylation has in instigating gene silencing of imprinted genes.

1.3.4 DNA Methylation and Long-term Silencing During X-inactivation

In contrast to imprinting, DNA methylation is added after gene silencing during X-chromosome inactivation. As a form of dosage compensation for X-linked genes, mammalian female cells silence one of their two X chromosomes. The process begins when the long non-coding RNA, *XIST*, is upregulated and coats the X-chromosome (Clemson *et al.*, 1996). *XIST* recruits the PRC2 histone methyltransferase which applies the inactivating H3K27me3 mark (Plath *et al.*, 2003). Over time, the inactive X is bound up tightly in heterochromatin. Late in this process, DNA methylation accrues at the promoters of X-linked housekeeping genes (Norris *et*

al., 1991). Methylated genes, such as *Hprt*, can be reactivated by loss of DNA methylation (Bhatnagar *et al.*, 2014; Csankovszki *et al.*, 2001; Mohandas *et al.*, 1981). When *XIST* expression is lost after inactivation, the inactive X adopts a more active chromatin structure, but DNA methylation and gene silencing remain (Splinter *et al.*, 2011). Thus, DNA methylation serves as the lock that maintains the repression of many genes on the inactivated X chromosome.

1.3.5 DNA Methylation Silences Retrotransposon

DNA methylation also acts as a long-term lock that represses retrotransposon activity in the genome. Loss of *Dnmt3L* prevents *de novo* methylation of retrotransposons as well as increased transposition and chromosomal instability (Bourc'his and Bestor, 2004). In mice expressing a *Dnmt1* hypomorph, multiple long terminal repeat (LTR) intracisternal A particle (IAP) retrotransposons mobilized (Howard *et al.*, 2008). In the human genome, long interspersed nuclear element family 1 (LINE-1 or L1) elements are the transpositionally active retroelements (Konkel and Batzer, 2010). Transcription from L1 elements increases with global DNA hypomethylation, suggesting that hypomethylation reactivates retrotransposition (Konkel and Batzer, 2010).

Retrotransposons can alter gene expression by serving as alternative regulatory sequences. A well-known example is the hypomethylation of an IAP element upstream of a dominant *Agouti* allele in mice. This alternative promoter causes constitutive expression of the dominant allele, resulting in the development of a yellow coat, obesity, diabetes, and cancer (Michaud *et al.*, 1994; Morgan *et al.*, 1999). In addition, when unmethylated, the L1 element in intron 2 of the hepatocyte growth factor receptor, *MET*, produces a truncated transcript that is constitutively active and promotes tumors of various types (Hur *et al.*, 2013; Wallenius *et al.*,

2000; Wolff *et al.*, 2010). Transcriptome sequencing studies suggest that the above examples are not isolated incidents and that retrotransposons frequently serve as alternative promoters for coding genes (Faulkner *et al.*, 2009). Retrotransposons comprise 42% of the human genome (Lander *et al.*, 2001). Based on the above evidence, the control that DNA methylation has over retrotransposons therefore represents an important amount its control over gene expression.

1.3.6 The Effect of Gene Promoter Methylation Depends on CG Content

DNA methylation has different effects at gene promoters depending on whether the promoter has high, intermediate, or low CpG density (Jones, 2012). The promoters of 60% to 70% of human genes are present in CGI which have high/intermediate CpG content and remain largely unmethylated (Illingworth and Bird, 2009). There is some evidence that high CG content alone provides this protection (Krebs *et al.*, 2014). However, the chromatin context created by active transcription also provides protection. H3K4me3 marks active genes and repels DNMT3L (Ooi *et al.*, 2007). In systems with DNMT3L expression, *de novo* methylation is discouraged. Active promoters can be further protected by TF binding as described above. Despite these protections, some CGI promoters are methylated and silenced (Weber *et al.*, 2007). Interestingly, the CGIs at the methylated and silenced promoters generally contained intermediate CpG density. High CpG density CGI promoters remained unmethylated (Weber *et al.*, 2007).

Promoters that lack the protection of high CG content are more susceptible to DNA methylation (Weber *et al.*, 2007; Ziller *et al.*, 2013). One study found that DNA methylation silenced the *LAMB3* and *RUNX3* promoters in cancer (Han *et al.*, 2011). This was confirmed by a genome-wide study that found about 7500 out of 9000 methylated promoters were silenced (Sarda *et al.*, 2017). The observation that 1500 methylated promoters were still expressed had

been seen previously (Weber *et al.*, 2007), and appears to be due to transcription from alternative promoters that lie within CGI (Sarda *et al.*, 2017).

Part of the purpose of gene body methylation is to repress these alternative promoters. An actively transcribed gene body is marked by H3K36me3, which recruits DNMT3B (Baubec *et al.*, 2015). The result is that gene body methylation positively correlates with gene expression. This appears to contradict the role of DNA methylation in gene silencing. However, gene body methylation is instead silencing alternative promoters (Maunakea *et al.*, 2010; Neri *et al.*, 2017).

The role of DNA methylation at enhancers is less clear. The TFs that bind to enhancers both activate gene expression and recruit coactivators that mark active enhancers with H3K4me1 and H3K27ac (Rada-Iglesias *et al.*, 2011). The antagonism of TF binding by DNA methylation has been proposed as a means of shutting down enhancers (Calo and Wysocka, 2013). Initial studies found negative correlations between DNA methylation and active enhancer marks such as H3K4me1, H3K27ac, and TF binding (Stadler *et al.*, 2011; Thurman *et al.*, 2012). However, mouse CTCF and REST can bind enhancer regions in the presence of methylation and cause methylation loss (Stadler *et al.*, 2011). This argues against a regulatory role for DNA methylation at enhancers. The recent discovery of active bivalent enhancers that have both H3K27ac and DNA methylation (Charlet *et al.*, 2016) further indicates that DNA methylation interacts with enhancers differently than it does at promoters.

1.4 DNA methylation and Cancer

Global DNA hypomethylation and local hypermethylation are common occurrences in human cancers (Ehrlich and Lacey, 2013). Global hypomethylation was first observed in 1983 (Feinberg and Vogelstein, 1983; Gama-Sosa *et al.*, 1983). Focal hypermethylation was

discovered shortly thereafter at the calcitonin promoter (Baylin *et al.*, 1986) and later silencing the RB tumor suppressor in retinoblastoma (Greger *et al.*, 1989; Sakai *et al.*, 1991). In cancer, the protection of CGI promoters from methylation is not maintained (Illingworth *et al.*, 2010), and local hypermethylation promotes the silencing of tumor suppressor genes such as *RB* and *INK4A* (Feinberg and Tycko, 2004). The specific mechanisms by which global hypomethylation contribute to cancer are less clear. However, evidence is growing for several mechanisms by which hypomethylation promotes cancer including: increased genomic instability, altered gene regulation, and reactivation of retrotransposons (Ehrlich and Lacey, 2013). The precise mechanism(s) that cause these aberrant epigenetic profiles are not completely understood, but are sometimes due to mutations that affect epigenetic machinery (Plass *et al.*, 2013). Because of the reversible nature of epigenetic marks, they offer enticing therapeutic targets. In order to create targeted therapeutics, however, we will need a better understanding of the specific epigenetic contributions to carcinogenesis and progression.

1.4.1 The Effects of Genome-Wide Hypomethylation in Cancer

Given the preceding examples of methylation regulating silencing, genome-wide hypomethylation should upregulate gene expression (Ehrlich and Lacey, 2013). This prediction holds true at the cancer testis antigen PRAME (Ortmann *et al.*, 2008; Zhang *et al.*, 2016) and P-cadherin (Milicic *et al.*, 2008).

Hypomethylation can also reactivate repressed retrotransposon sequences. In mice expressing a *Dnmt1* hypomorph, multiple IAP retrotransposons mobilized and inserted at the *Notch1* locus, promoting lymphoma development (Howard *et al.*, 2008). In the human genome, active retrotransposons caused 194 somatic transposition events in 43 tumors from various

tissues (Lee *et al.*, 2012). Notably, these 194 insertions tended to downregulate the genes they inserted near, including some tumor suppressors (Lee *et al.*, 2012). However, most tumors had less than 10 transposition events, suggesting that these events are somewhat rare. Though rare, the epigenetically silenced LTR element located upstream of *CSF1R* promotes aberrant expression of the colony stimulating factor receptor and drives Hodgkin's lymphoma (Lamprecht *et al.*, 2010). In addition, when unmethylated, the L1 element in intron 2 of the hepatocyte growth factor receptor, *MET*, produces a truncated transcript that is constitutively active and promotes tumors of various types (Hur *et al.*, 2013; Wallenius *et al.*, 2000; Wolff *et al.*, 2010). Transcriptome sequencing studies suggest that the above examples are not isolated incidents and that retrotransposons frequently serve as alternative promoters for coding genes (Faulkner *et al.*, 2009).

Genome-wide hypomethylation does not only reactivate gene/retrotransposon activity. Hypomethylation also correlates with decreased transcription (Hansen *et al.*, 2011). Cancer-associated genome-wide hypomethylation occurs in large blocks, and 36% of expressed genes became silenced in these blocks. The mechanism is not clear, but significant overlap between these domains and lamin associated domains (Berman *et al.*, 2012) suggests that repression due to association with the nuclear lamina could be responsible (Hansen *et al.*, 2011).

In addition to its effects on gene expression, global hypomethylation also promotes chromosomal instability. Increased loss of heterozygosity and tumor formation was observed in $Nf1^{+/-}$ and $p53^{+/-}$ cells expressing a hypomorphic *Dnmt1* allele (Eden *et al.*, 2003). These mice developed T cell lymphomas at an early age, likely due to trisomy of chromosome 15 and the resulting overexpression of c-Myc. In contrast, hypomethylation can lead to deletion of *Apc* in

mouse liver tumors (Yamada *et al.*, 2005) and *MTHFR* in human glioblastomas (Cadieux *et al.*, 2006). Altogether, it is clear that global hypomethylation assists tumor formation and/or progression in multiple ways.

1.4.2 The Effects of Focal Hypermethylation in Cancer

Focal hypermethylation also assists tumor formation and progression by downregulating tumor suppressor genes. Most early study of DNA methylation and cancer focused on this phenomenon. As a result, several examples of this phenomenon have been discovered: *RB* (Greger *et al.*, 1989; Sakai *et al.*, 1991), *MLH1* (Miyakura *et al.*, 2004; Suter *et al.*, 2004), E-cadherin (Graff *et al.*, 1995), *VHL* (Herman *et al.*, 1994), and *CDKN2A* (Gonzalez-Zulueta *et al.*, 1995). Further research found that methylation induced silencing is a common occurrence in the cancers analyzed by the TCGA (Ciriello *et al.*, 2013).

As just one example, consider *CDKN2A* (also known as $p16^{INK4A}$). *CDKN2A* is a tumor suppressor that regulates cell cycle progression by preventing the activation of CDK4/6 (Zhao *et al.*, 2016). When all TCGA cancer types are considered, it is one of the most frequently methylated, deleted, and/or silenced genes (Ciriello *et al.*, 2013), and silencing strongly correlates with methylation in lung cancer (The Cancer Genome Atlas Research Network, 2012). *CDKN2A* expression can be reactivated by treatment with 5-azacytidine, a DNMT inhibitor (Hassler *et al.*, 2012), and monoallelic methylation results in monoallelic *CDKN2A* expression (Myöhänen *et al.*, 1998). Together these studies suggest that DNA methylation may cause the repression of *CDKN2A* expression. In mouse embryonic fibroblasts, *CDKN2A* appears to be regulated by PRC2 and hSNF5 (Wilson *et al.*, 2010). Unfortunately, this study failed to analyze DNA methylation making it impossible to determine whether DNA methylation initiated

silencing in these cells. As a result, we are left with the conclusion supported by the evidence from all the other studies: DNA methylation silences *CDKN2A* expression, and this is representative of other tumor suppressor silencing events in cancer.

1.4.3 DNA Methylation Based Cancer Therapeutics

Therapeutics based on focal hypermethylation in cancer have already been approved by the FDA. These drugs are the DNMT inhibitors 5-azacytidine and 5-aza-2-deoxycytidine (Jones *et al.*, 2016). Treatment with DNMTs caused 15% or more of myelodysplastic syndrome or acute myeloid leukemia patients to survive longer while present fewer cancerous blood cells and more healthy cells (Jones *et al.*, 2016). Part of the efficacy of these drugs arises from the reactivation of endogenous retrovirus sequences following 5-azacytidine treatment in cancer. The interferon response caused by increased expression of retrotransposon RNA draws the immune system to attack the tumor (Chiappinelli *et al.*, 2015). Nevertheless, by encouraging broad loss of methylation, it remains possible to trigger the pro-cancer effects described above. Treatments targeting a specific, local DNA methylation change could be used to avoid some of these side effects. However, both DNA methylation editing technologies and our understanding of aberrant cancer epigenetics must develop before targeted therapies become a reality.

1.5 Tools for DNA Methylation Analysis

A large problem in the DNA methylation field has been the lack of tools to make targeted DNA methylation changes. Many of the studies above that consider *CDKN2A* expression consider only the correlation between methylation and expression across multiple samples. The Hassler *et al.* (2012) study is the exception because it uses 5-azacytidine, a DNMT inhibitor, to demethylate and reactivate *CDKN2A* expression. That is one example of the most commonly

used approaches to demonstrate a causal relationship between methylation and expression: inhibition of DNMT activity by chemical inhibition, siRNA/shRNA knockdown, or gene knockout. This technique suggests a causal relationship if demethylation reactivates expression.

DNMT inhibition has downsides. The primary drawback is that demethylation occurs genome-wide and may cause undesirable off-target effects. 5-azacytidine has the additional drawback of inducing DNA and RNA damage (Cihák, 1974). To avoid these problems, it is possible to clone out the promoter of interest and then reintegrate methylated/unmethylated forms by site-specific recombination (Appanah *et al.*, 2007; Busslinger *et al.*, 1983; Lorincz *et al.*, 2004). This approach is very specific but has the drawback that the promoter is no longer being studied at the endogenous locus and the different chromatin context may bias the findings (Amabile *et al.*, 2016). The ideal solution would be targeted manipulation of DNA methylation.

1.5.1 A Brief History of Targeted DNA Methylation Editing

Targeted methylation technologies have existed for twenty years, but have not been widely adopted. The first demonstration of the technology was a fusion of the M.HhaI DNMT to the Zif268 zinc finger (ZF) protein (Xu and Bestor, 1997). The Zif268 protein successfully guided the methylation activity of M.HhaI to an oligo substrate *in vitro*. Since that proof of concept, several studies have fused DNMTs and targeting domains to study DNA methylation (Groote *et al.*, 2012). Early studies continued to use ZFs as the targeting domains for prokaryotic DNMTs M.SssI, M.HhaI, and M.HpaII (Groote *et al.*, 2012). These early studies targeted oligos *in vitro*, plasmid reporter systems, or DNA integrated into yeast (Groote *et al.*, 2012). Smith and Ford (Smith *et al.*, 2008) used a ZF-M.HhaI fusion to methylate a promoter integrated into the mouse genome. The integrated promoter was silenced and the methylation was stable for 16 days

(Smith *et al.*, 2008). The overall conclusion of the early studies was that DNA methylation silenced gene expression; however, these studies still did not consider the native locus.

It was not until 2012 that a mammalian DNMT was targeted to an endogenous locus (Rivenbark *et al.*, 2012). Rivenbark *et al.* (2012) targeted the mouse Dnmt3a catalytic domain to the promoter of *Maspin*, a tumor suppressor gene in human cancer cell lines. 50% methylation of the *Maspin* promoter induced *Maspin* repression and increased soft agar colony formation (Rivenbark *et al.*, 2012). Similar results have been described at the following promoter regions: *SOX2* (Stolzenburg *et al.*, 2015), EpCAM (Nunna *et al.*, 2014), and *CDKN2A* (Cui *et al.*, 2003).

Over time, the targeted DNA methylation field has incorporated new technologies and approaches. The known enhancement of DNMT3A activity by DNMT3L led Siddique *et al.* (2013) to use a ZF-mDnmt3a-mDnmt3L fusion. As expected, the new fusion induced more methylation than Dnmt3a alone (Siddique *et al.*, 2013). The field also adopted the more recent TALE proteins as a DNA targeting domain. When paired with the mDnmt3a-mDnmt3L fusion, this domain can methylate and silence *CDKN2A* in primary human fibroblasts (Bernstein *et al.*, 2015). The demethylation of DNA by targeted TET activity has also been performed with a ZF fusion (Chen *et al.*, 2013) and with a TALE fusion (Maeder *et al.*, 2013a) Despite adopting the new TALE technology, targeted methylation manipulation is still limited by the expense and time required to create custom DNA binding proteins. I believe this explains the fifteen-year delay between initial proof of concept and application of targeted methylation at an endogenous locus. Nevertheless, a novel targeting solution is needed to carry this technology into the future.

1.5.2 The CRISPR/Cas9 System Offers a More Efficient Targeting System

The CRISPR/Cas9 system offers an inexpensive and flexible DNA targeting system (Komor *et al.*, 2017) that can facilitate the use of targeted methylation strategies. The CRISPR/Cas9 system was discovered in bacteria where it degrades foreign DNA to protect the bacterial genome (Barrangou *et al.*, 2007; Garneau *et al.*, 2010). The bacteria store short, >20 bp fragments of foreign DNA called spacers between palindromic repeat sequences. This arrangement is known as the clustered regularly interspaced short palindromic repeat (CRISPR) array (Andersson and Banfield, 2008; Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). The long CRISPR array transcripts (pre-crRNA) is cleaved to free an individual spacer-repeat transcript. This is directed by a *trans*-activating CRISPR RNA (tracrRNA) that binds the repeat sequence of the pre-crRNA. The final crRNA/tracrRNA complex allows site-specific binding by Cas9 (Deltcheva *et al.*, 2011; Gasiunas *et al.*, 2012). This system has been optimized by fusing the two RNA components into one, called a small guide RNA or sgRNA (Fig. 1.2,

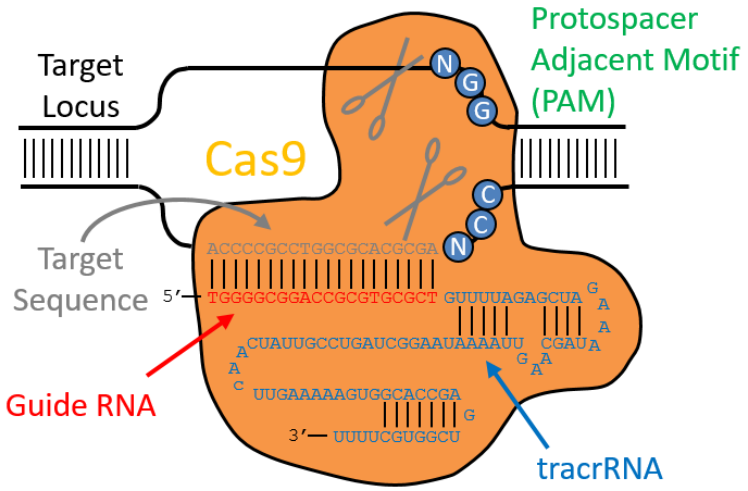


Figure 1.2. The CRISPR/Cas9 System. Figure adapted from https://www.systembio.com/genome-engineering-cas9-crispr-smartnuclease/overview.

Garneau *et al.*, 2010; Jinek *et al.*, 2012). As a result, the CRISPR/Cas9 system can be reprogrammed to target a new location by changing the spacer sequence. This requires a single cloning step, making it inexpensive and quick. Because of this ease of use, the technology has rapidly replaced ZF or TALE nuclease fusions that had previously been used for genome editing (Komor *et al.*, 2017).

Subsequent research has continued to optimize the CRISPR/Cas9 system for genome editing. One drawback of the CRISPR/Cas9 system over ZF- or TALE-based systems is increased off-target effects (Fu *et al.*, 2013; Hsu *et al.*, 2013). A few months ago, one group found that two mice treated with CRISPR/Cas9 showed a surprisingly high number of off-target mutations. Even more surprisingly, many of these were the same in both mice (Schaefer *et al.*, 2017). There is ongoing discussion on whether the authors conclusions were justified. Several approaches have been demonstrated to increase Cas9 specificity (Tsai and Joung, 2016). Increased specificity can be accomplished by requiring co-localization of two Cas9 nickase complexes (Guilinger *et al.*, 2014; Mali *et al.*, 2013; Redis *et al.*, 2013; Tsai *et al.*, 2014). The Cas9 protein can also be engineered for greater specificity (Kleinstiver *et al.*, 2016; Slaymaker *et al.*, 2016). The other drawback of the *S. pyogenes* CRISPR/Cas9 system is that all binding sites must contain a 5'-NGG-3' motif at the 3' end of the target site. This motif is called the protospacer adjacent motif (PAM) sequence (Jinek *et al.*, 2012). Engineered Cas9s can overcome this requirement (Kleinstiver *et al.*, 2015). However, it may be easier to utilize alternative Cas9 proteins that natively have different PAM sequence requirements (Komor *et al.*, 2017)

I believe that the Cas9 technology will enable the use of epigenome editing. The nuclease activity of Cas9 can be inactivated by two point mutations (Qi *et al.*, 2013). The nuclease dead

Cas9 (dCas9) thereby becomes a reprogrammable, RNA-directed DNA binding protein. This system has been used to guide the VP4 transcriptional activator (Maeder *et al.*, 2013b; Perez-Pinera *et al.*, 2013a) and the KRAB repressor (Gilbert *et al.*, 2013; Thakore *et al.*, 2015), the p300 histone acetyltransferase domain (Hilton *et al.*, 2015), and the LSD1 histone demethylase (Kearns *et al.*, 2015) to specific sites in the genome. The work described in Chapters 3 and 4 describes my work to create and optimize a dCas9-DNMT fusion for epigenome editing.

1.6 Summary

DNA methylation is an integral part of the regulatory machinery in the human genome, but many of its effects remain poorly understood. There are clear examples of gene silencing in the presence of DNA methylation such as imprinting and X-inactivation. However, there are many genes that remain expressed despite being methylated. It is therefore not clear which genes are controlled by DNA methylation. Moreover, during X-inactivation, DNA methylation is added after silencing occurs. In contrast, targeted methylation studies at the multiple promoters (see above) indicate that methylation induction is sufficient for gene silencing. It is therefore not clear when DNA methylation initiates gene silencing and when it is added afterward. My thesis seeks to gather data that helps to close these knowledge gaps.

In Chapter 2, I study the effects of genome-wide hypomethylation in breast cancer. About 75% of breast cancers depend on estrogen signaling through the estrogen receptor (ERα). These tumors are effectively treated by aromatase inhibitors (AI) that prevent estrogen production. However, almost all advanced cases of ERα positive breast cancer develop resistance to AI therapy. I therefore sought to identify methylation changes that promote this resistance. I studied

UCA1 and *PTGER4*, two genes identified by a screen for negatively correlated methylation and expression changes in a cell line model of AI resistance.

The DNA methylation field's inability to determine if and when DNA methylation causes gene expression change derives from the limitations of the tools available. The most commonly used tools demethylate the entire genome by DNMT inhibitors or by DNMT knockout or knockdown. Targeted methylation technologies have been demonstrated (Xu and Bestor, 1997). However, targeted approaches have not been widely adopted because they depend on custom ZF and TALE DNA binding proteins that are expensive and time consuming to produce. In the last few years the CRISPR/Cas9 system rose to prominence as a flexible, inexpensive, and easily implemented technology for targeting specific regions of the genome. In Chapters 3, I demonstrate that the fusion of a DNMT to dCas9 allows targeting of methylation to specific locations in the genome. In Chapter 4, I optimize the system by fusing dCas9 to the catalytic domain of alternative DNMTs.

Chapter 2: Promoter Hypomethylation Promotes Acquired Endocrine Therapy Resistance in Breast Cancer

2.1 Introduction

One of the hallmarks of cancer is the ability to sustain proliferative signaling (Hanahan and Weinberg, 2011). Estrogen receptor α (ERα) is a transcription factor that mediates the growth signaling necessary for normal human breast development. As a result, ~75% of breast tumors express ERα and depend on its transcriptional activity (Collins *et al.*, 2005; Harvey *et al.*, 1999; Johnston and Dowsett, 2003; Musgrove and Sutherland, 2009). Endocrine therapies that prevent ERα signaling succeed against 70% of ERα-positive breast cancers (Allred *et al.*, 2004). However, growth factor receptor signaling activates ERα-independent cell proliferation and cell survival, thereby promoting endocrine therapy resistance. Multiple growth factor receptor pathways can do this, but the contributions of the phosphatidylinositol-3 kinase (PI3K) and extracellular signal-regulated kinase (ERK) signaling pathways are best understood (Ma *et al.*, 2015). The mechanisms that drive PI3K or ERK signaling are therefore important to understanding and treating endocrine therapy resistant breast cancer.

Dysregulation of the genes in the PI3K or ERK signaling pathways can alter the signaling in that pathway. DNA methylation is an epigenetic mark that negatively correlates with gene expression, and the breast cancer genome is aberrantly hypomethylated genome-wide (Ehrlich and Lacey, 2013). This global loss of DNA methylation has the potential to alter the expression of cell signaling genes. For example, hypomethylation correlates with the expression of the *MET*

hepatocyte growth factor receptor oncogene (Wolff *et al.*, 2010), which affects the ERK signaling pathway (Xu and Yu, 2007). As a result, aberrant DNA methylation and epigenetic gene regulation may make significant contributions to the development of endocrine therapy resistance.

2.1.1 Breast Cancer Prevalence

Breast cancer remains the leading cause of cancer death in women worldwide, accounting for 25% of all cancers and 15% of all cancer deaths (Torre *et al.*, 2015). In addition, breast cancer incidence rose from 42.3 new diagnoses per 100,000 population worldwide in 2008 (1,384,000 total cases) to 43.3 new diagnoses per 100,000 population in 2012 (1,676,633 total cases, Youlden *et al.*, 2012, 2014). Breast cancer incidence is expected to continue increasing as more nations adopt breast cancer-prone habits: delayed childbearing, less breast feeding, and hormone therapy for postmenopausal women (Torre *et al.*, 2015). As a result, breast cancer represents a significant health issue for an increasing number of women.

Despite increasing breast cancer incidence, breast cancer mortality decreased from 13.9 deaths per 100,000 population in 2008 to 12.9 deaths per 100,000 population in 2012 (Youlden *et al.*, 2012, 2014). In addition, the five year survival rate for breast cancer can be nearly 90%, which is higher than the survival rate for most other cancers (Youlden *et al.*, 2012). These positive trends can be attributed – at least in part – to the success of endocrine therapies such as tamoxifen and aromatase inhibitors (AI, Youlden *et al.*, 2012). A five year tamoxifen regimen is capable of producing an absolute 39% reduction in breast cancer recurrence and a 33% reduction in breast cancer mortality over 15 years (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011). A five year aromatase inhibitor treatment provided another 2.9% and 1.1%

reduction in recurrence and mortality, respectively, when compared with tamoxifen (Dowsett *et al.*, 2010).

Despite their success, these treatments have limits. Endocrine therapies target signaling via ERα, and only ~75% of breast tumors express ERα (Collins *et al.*, 2005; Harvey *et al.*, 1999; Johnston and Dowsett, 2003; Musgrove and Sutherland, 2009). The remaining ~25% cannot be treated by endocrine therapies. Consistent with this, the above benefits of five years of endocrine therapy were not observed for ERα-negative cancers (Dowsett *et al.*, 2010; Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011). As a result, ERα activity is an important clinical indicator of endocrine therapy response and overall outcome. Unfortunately, even in ERα-positive cases that respond to endocrine therapy, 20% releapse. When breast cancer reaches advanced stages, nearly all cases acquire resistance to endocrine therapy (Ma *et al.*, 2015). Understanding the mechanisms that drive endocrine therapy resistance will therefore be important for continuing to more effectively treat breast cancer.

2.1.2 ERα Structure

ERα transcriptional activity and growth signaling cross-talk depend upon ERα structure (Zilli *et al.*, 2009) (Fig. 2.1). The ERα gene, *ESR1*, was first cloned in 1986 (Green *et al.*, 1986). *ESR1* maps to chromosome 6 and encodes a protein containing 595 amino acids and weighing about 66 kDa (Schiff *et al.*, 2008). Overall, the estrogen receptor shares homology with the nuclear receptor superfamily (Weatherman *et al.*, 1999). Receptors of this family are commonly subdivided into the five regions denoted by the letters A/B-F (Fig. 2.1), but these regions do not necessarily correlate with functional domains. Within the superfamily, ERα represents one of the steroid receptors, which in general have a more extensive activation function 1 (AF1) domain.

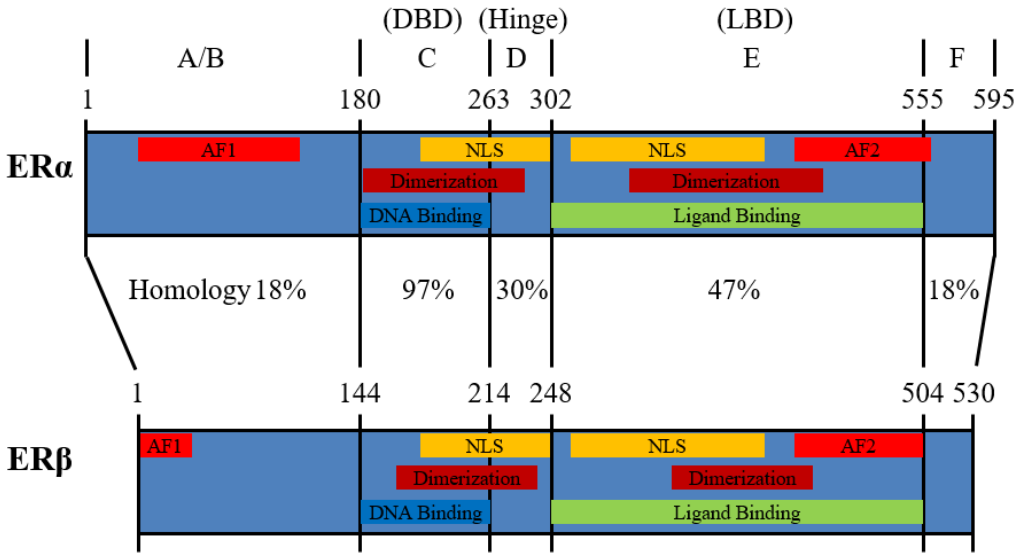


Figure 2.1: ERα domain structure. AF1 = Activating Function 1. AF2 = Activating Function 2. DBD = DNA Binding Domain. LBD = Ligand Binding Domain. NLS = Nuclear Localization Signal. Adapted from Zilli *et al.* (2009).

Subsequent research revealed another estrogen receptor, ERβ (Fig. 2.1, Mosselman *et al.*, 1996). ERβ maps to chromosome 14, representing a different receptor rather than a splice isoform (Schiff *et al.*, 2008). ERα and β have high sequence identity in their ligand- and DNA-binding domains, indicating similar functional responses to an estradiol binding event (Schiff *et al.*, 2008). The mere 18% similarity in the A/B region, however, indicates that ERα and ERβ have significantly different AF1 activities (Zilli *et al.*, 2009). The consequences of these differences and the role of ERβ in endocrine therapy resistance remain unclear.

2.1.3 ERα Signaling

ERα affects cell signaling through two pathways: genomic and non-genomic. Genomic signaling represents the ligand-dependent transcriptional activity of ERα through dimerization and DNA binding. Genomic signaling is initiated when estradiol binds to ERα and releases it from its heat-shock protein chaperones. Afterward, ERα localizes to the nucleus, dimerizes and

binds to estrogen response elements (EREs) near promoters. Once bound, coactivator and corepressor proteins cooperate with ERα to silence or activate ERα regulated genes. Active ERα may also alter the transcriptional activity of other transcription factors such as AP-1 at genes targeted by Jun/Fos. ERα phosphorylation, sumoylation, and variations in the ERE consensus sequence may all affect ERα conformation and activity in a complex regulatory web (Schiff *et al.*, 2008; Zilli *et al.*, 2009).

The above modifications affect the activity of the two ERα activating functions: activating function 1 (AF1) and activating function 2 (AF2). AF1 activity is controlled primarily by phosphorylation at S104, S106, S118, and S167 (Zilli *et al.*, 2009). Kinases downstream of PI3K and ERK target S167 and activate ERα (Mendoza *et al.*, 2011). In contrast, AF2 activity requires estradiol binding. Upon binding, helix twelve shifts to enclose the ligand and open the NR box motif for coactivator/corepressor binding (Zilli *et al.*, 2009). Estradiol binding also increases AF1 activity via promoting phosphorylation at S118. When active, ERα will recruit either histone acetyltransferases (HATs) through the NCoA1-3 coactivators of the Src family or histone deacetylases (HDACs) through the NCoR corepressors (Zilli *et al.*, 2009).

Non-genomic signaling occurs through cross talk with growth receptor pathways at the cell's plasma membrane. In order to reach the membrane, ERα undergoes palmitoylation. Once there, estradiol binding causes ERα to dimerize and interact with signaling proteins including the p85 subunit of PI3K, among others. Such interactions activate growth signaling through the PI3K and ERK pathways (Zilli *et al.*, 2009). These pathways activate AF1, which in turn upregulates *EGFR* and other genes that promote growth signaling (Mendoza *et al.*, 2011; Zilli *et*

al., 2009). The details about how this positive feedback loop can contribute to endocrine therapy resistance will be discussed below.

2.1.4 Endocrine Therapies

Endocrine therapies attack these oncogenic signaling mechanisms at different levels. Selective estrogen receptor modulators (SERMs) inhibit estradiol binding and promote inactive ERα conformations. There are three types of SERMs: triphenylethylene derivatives, steroidal derivatives, and non-triphenylethylene non-steroidal derivatives (Schiff *et al.*, 2008). Tamoxifen, the best known SERM, is a triphenyl ethylene derivative. It competitively inhibits estradiol binding and promotes helix 12 conformations of AF2 that preferentially recruit corepressors (Schiff *et al.*, 2008). The mechanisms of other SERM types are similar but not identical to that of tamoxifen. Notably, tamoxifen does not prevent ERα dimerization, which allows AF1 activation through ligand-independent phosphorylation at S167. When activated in this way, ERα promotes growth signaling through the non-genomic signaling pathway (Zilli *et al.*, 2009). Therefore, tamoxifen activates growth signaling pathways in tumors that express growth signaling receptors such as *EGFR* or *ERBB2*. Such signaling contributes to the resistance of breast tumors to endocrine therapy (Zilli *et al.*, 2009).

Selective estrogen receptor downregulators (SERDs) address the agonism of non-genomic signaling produced by SERMs. Fulvestrant is the best known SERD, and acts in a similar mechanism to tamoxifen by competitively inhibiting estradiol binding. However, the conformations induced by fulvestrant binding prevent dimerization and nuclear import. Instead of nuclear import, ERα localizes to the proteasome and is degraded. By inducing the degradation

of ERα, SERDs completely block both genomic and non-genomic ERα signaling. As a result, they do not promote endocrine therapy resistance through growth signaling (Zilli *et al.*, 2009).

Aromatase inhibitors utilize a different approach to inhibit ERα signaling. AIs inhibit the conversion of testosterone to estrogen by cytochrome P450 (*CYP19*), also known as aromatase. There are two classes of AI: steroidal derivatives that act as suicide inhibitors of aromatase and non-steroidal derivatives that only competitively inhibit aromatase. Both classes of inhibitor drastically decrease estrogen concentration and thereby inhibit all forms of ERα signaling (Zilli *et al.*, 2009). This may explain the clinically significant decreases in recurrence relative to tamoxifen because aromatase inhibitors do not promote the activity of growth factor signaling pathways (Dowsett *et al.*, 2010). However, AIs are largely restricted to use in postmenopausal women because aromatase activity in the ovaries of premenopausal women overcomes the effects of the inhibitors (Zilli *et al.*, 2009).

Inhibition of ERα signaling by the above therapies induces cell cycle arrest and apoptosis (Musgrove and Sutherland, 2009). These events are caused when ERα no longer upregulates *MYC* and cyclin D1 (*CCND1*). Decreased *MYC* expression leads to increased expression of the p21 cyclin dependent kinase inhibitor (*CDKN1A*). CDKN1A expression increases are compounded by decreased CCND1 expression that prevents the CCND1/CDK4 complexes from sequestering the CDKN1A CDK inhibitor. Combined, these events prevent cyclin E1/CDK2 phosphorylation of RB, and the end result is cell cycle arrest at G1 (Musgrove and Sutherland, 2009). This cell cycle arrest and sensitization to apoptosis acts to prevent breast cancer progression.

2.1.5 Mechanisms of Endocrine Therapy Resistance

Resistance to the above endocrine therapy effects develops via multiple mechanisms. Cell lines deprived of estrogen for several weeks become hypersensitive to estrogen (Sanchez *et al.*, 2011). Acquisition of activating ERα mutations could cause such hypersensitivity. Searches for point mutations have, however, yielded contradictory results. Roodi *et al.* (1995) found only 2 missense mutations in the same tumor out of 188 tumor samples. A later study of pre-cancerous breast hyperplasias, however, found a K303R mutation in 30% of these lesions (Fuqua *et al.*, 1995). A study of invasive breast cancer also found the K303R mutation in 50% of invasive breast carcinoma samples (Herynk *et al.*, 2007). The K303R mutation makes ERα more sensitive to estradiol binding and increases ERα affinity for the NCoA2 coactivator. Thus, the A908G SNP that causes the K303R allele represents a gain of function allele that correlates with poor prognosis (Herynk *et al.*, 2007). More recent studies appear to have solved this inconsistency by discovering that *ESR1* mutations are much more common in endocrine therapy resistant tumors (Li *et al.*, 2013). These mutation can occur in the ligand binding domain and activate ERα signaling (Merenbakh-Lamin *et al.*, 2013; Robinson *et al.*, 2013; Toy *et al.*, 2013).

Another mechanism by which endocrine therapy resistance could develop is the loss of dependence on ERα signaling. Independence from ERα can be accomplished by repressing ERα transcription (Zilli *et al.*, 2009). There are three ERα promoters: the most commonly used proximal promoter and two others further upstream. Transcription factors related to AP1, AP2, and SP1 bind to these sites. However, the interactions of these transcription factors and their control of ERα expression has not been well-defined (Schiff *et al.*, 2008).

ERα expression could also be silenced by an epigenetic mechanism. Hypermethylation of the CpG island containing the ERα promoter correlates with decreased ERα expression in cell lines and primary tumors (Iwase *et al.*, 1999; Nass *et al.*, 2000; Ottaviano *et al.*, 1994). In addition, DNA methyltransferase (DNMT) inhibition by 5-azacytidine induces ERα expression (Ferguson *et al.*, 1995). However, the ERα promoter remains hypermethylated in 35% of ERα-positive/progesterone receptor-positive cancers (Lapidus *et al.*, 1996). Furthermore, HDAC inhibitors can restore ERα expression without DNA demethylation (Zhou *et al.*, 2007). These data suggest that both DNA methylation and chromatin modification affect ERα gene regulation, but the controlling epigenetic mark is not clear.

Alterations of the expression or activity of proteins that interact with ERα can also promote endocrine therapy resistance. For example, tumors in which the ERα corepressor, NCoR1, is downregulated have a poorer prognosis relative to those with higher NCoR1 expression (Zilli *et al.*, 2009). Conversely, activation of AIB1, the ERα coactivator also known as NCoA3, by phosphorylation downstream of EGFR/HER2 signaling promotes HER2 expression via ERα and completes a positive feedback loop (Hurtado *et al.*, 2008). As a result, high AIB1 expression levels correlate with endocrine therapy resistance and poorer prognosis in tumors that express HER2 (Musgrove and Sutherland, 2009; Zilli *et al.*, 2009).

2.1.6 Growth Factor Receptor Signaling Promotes Endocrine Therapy Resistance

Stimulation of growth factor receptor signaling promotes resistance through multiple mechanisms in addition to the AIB1/HER2 feedback. In general, these additional mechanisms counteract the apoptosis and cell cycle arrest induced by endocrine therapies as described at the

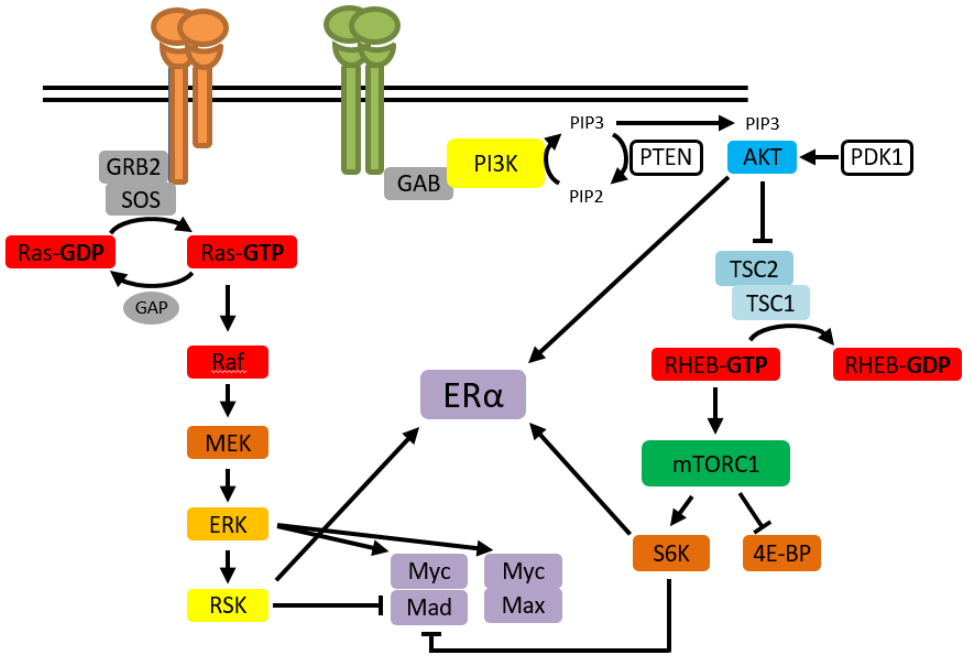


Figure 2.2: PI3K and ERK pathways components that promote AI therapy resistance. Figure adapted from Mendoza *et al.* (2011).

end of the Endocrine Therapies section above. Of the growth factor signaling pathways, PI3K and ERK are commonly used for this function in breast cancer. The PI3K pathway is activated by growth factor receptors, metabolic sensors, and other stimuli (Mendoza *et al.*, 2011). These receptors or sensors recruit PI3K to the membrane, where it generates phosphatidylinositol 3,4,5-triphosphate (PIP3). PIP3 recruits AKT to the membrane, where it is activated by 3-phosphoinositide dependent kinase 1 (PDK1). AKT phosphorylates TSC2, preventing the GAP activity of TSC1/2 complexes. GTP-bound TSC1/2 phosphorylate and activate mTOR complex 1, which phosphorylates the p70 ribosomal S6 kinase (S6K) and downstream effectors (Fig. 2.2) (Mendoza *et al.*, 2011). Of the above, AKT, mTOR, and S6K kinases are the most important contributors to endocrine therapy resistance.

The results of PI3K pathway activity include phosphorylation of ERα by AKT and S6K (Fig. 2.2, Mendoza *et al.,* 2011). This ERα phosphorylation induces its AF1 activity and restores proper regulation of *MYC*, and *CCND1*. The expression of *MYC* and *CCND1* is further aided by mTOR complex 1-mediated eIF4E binding protein (4E-BP) phosphorylation. Phosphorylated 4E-BP releases translation initiation factor, eIF4E, to promote translation of many proteins including MYC and cyclin D1. S6K activation goes even further by phosphorylating MAD1, an inactivating MYC binding partner. MAD1 phosphorylation promotes its degradation and allows MYC to bind its activating partner, Max (Mendoza *et al.*, 2011).

The ERK signaling pathway also has many activators that include EGFR, other RTKs, and G-protein coupled receptors (Mendoza *et al.,* 2011). These receptors lead to activation of RAS, which begins a phosphorylation cascade from Raf through MEK to ERK. ERK phosphorylates the p90 ribosomal S6 kinase (RSK) and other downstream effectors that promote cell growth and proliferation (Fig. 2.2, Mendoza *et al.*, 2011). Of these, the ERK and RSK kinases are the most important contributors to endocrine therapy resistance.

The ERK pathway targets ERα through RSK. This phosphorylation promotes the same effects as phosphorylation by AKT and S6K downstream the PI3K pathway (Fig. 2.2). ERα phosphorylation and activation upregulate EGFR, which in turn activates ERK signaling. This completes a positive feedback loop similar to that involving HER2 and AIB1. RSK also acts in concert with S6K to degrade MAD1 and promote MYC activity. ERK also promotes MYC activity by directly phosphorylating it. The above events therefore complement the PI3K pathway in counteracting the effects of endocrine therapies (Mendoza *et al.*, 2011).

ERK signaling also supports tumor metastasis by promoting cell motility and invasion (Vial and Pouysségur, 2004). Though increased motility and invasion do not necessarily contribute to endocrine therapy resistance, their link with ERK signaling means they will be highly correlated with resistance. ERK spurs motility through multiple mechanisms including the phosphorylation of myosin light chain kinase (MLCK). MLCK phosphorylates the myosin light chain, activating it to exert force on actin. The forces exerted by myosin are important for freeing the cell by dissolution of adherens junctions and forward movement (Vial and Pouysségur, 2004). Despite all we know about the downstream results of heightened growth factor receptor signaling, far less is known about the upstream causes (Ma *et al.*, 2015).

2.1.7 Endocrine Therapy Resistance is a Common Problem

Endocrine therapy resistance comes in two forms: intrinsic resistance and acquired resistance. For intrinsically resistant tumors, the above resistance mechanisms are active at the beginning of treatment and render the therapy ineffective. About 25% of all breast cancers have intrinsic resistance (Brinkman and El-Ashry, 2009). In tumors that acquire resistance, the above resistance mechanisms are activated during the course of treatment. Acquired resistance is very common: within fifteen years from the start of a five year tamoxifen treatment course, breast cancer will recur in 33% of women (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011). The frequency of endocrine therapy resistance makes it a significant clinical problem.

2.1.8 Functional Genomics Attempts to Solve Endocrine Therapy Resistance

The frequency of resistance makes identifying the proper treatment regimen for resistant tumors a challenge. Studies have shown that either fulvestrant or AI can effectively treat

tamoxifen-resistant tumors. This observation makes sense because tamoxifen partially activates ERα non-genomic signaling but fulvestrant and AI block ERα signaling completely through ERα degradation or estrogen deprivation, respectively. Other therapies, however, exhibit more complex interactions. For example, fulvestrant and AI remain effective in tumors that have developed resistance to the other drug. Furthermore, resistance to steroidal AIs does not indicate resistance to non-steroidal AIs (Zilli *et al.*, 2009). The adaptation of the cancer to only one endocrine therapy indicates that the molecular changes in resistant cancer are unique to each treatment.

Scientist and clinicians have attempted to use those unique molecular adaptations to determine the best therapy for a particular patient. Early attempts to determine proper treatment were based on biomarkers like ERα. These indicators are still used today and include the tumor grade, ERα, progesterone receptor (PR), and HER2 receptor tyrosine kinase (Sims *et al.*, 2008). Lower grade indicates a higher degree of tumor cell differentiation and better prognosis. PR and HER2 indicate intact ERα signaling and increased growth signaling, respectively. These markers can be used to assess the effect of endocrine or anti-HER2 therapies (Cui *et al.*, 2005; Sims *et al.*, 2008). However, these markers are not always sufficient to properly prescribe treatment. For example, approximately 30% of ERα-positive tumors do not respond to endocrine therapy (Allred *et al.*, 2004).

The insufficiency of single biomarkers suggested that the molecular signatures of tumors might be better diagnostic tools. This approach resulted in a recent study through The Cancer Genome Atlas (TCGA) Network (The Cancer Genome Atlas Network, 2012). This study included: copy number arrays, DNA methylation arrays, mRNA arrays, microRNA sequencing,

and reverse phase protein arrays. Cross-platform clustering identified four tumor subtypes: luminal A, luminal B, *ERBB2*-overexpressing, and basal (The Cancer Genome Atlas Network, 2012). Earlier studies had also identified similar clusters (Perou *et al.*, 2000; Sorlie *et al.*, 2003) and showed that the clusters correlate with clinical outcomes (Sorlie *et al.*, 2003). The molecular signature approach has also been applied to cell line models of acquired aromatase inhibitor resistance (Miller *et al.*, 2011). Ninety-nine genes underwent expression changes in three cell lines. This ninety-nine gene expression profile was able to separate luminal A and luminal B clusters, with the luminal B more closely resembling the resistant cell lines. Similarity with the endocrine therapy resistant cell lines also correlated with higher growth and recurrence rates as well as MYC expression (Miller *et al.*, 2011). These results are consistent with increased activity through growth signaling pathways as described above.

Cluster classification schemes are, however, of limited use for prescribing the proper treatment. This limitation arises from breast cancer heterogeneity. The TCGA Network study identified mutations in pathways that drive endocrine therapy resistance. These pathways include the PI3K pathway, RB pathway, and ERK pathway (The Cancer Genome Atlas Network, 2012). Mutations in the same pathways were also identified in a study of endocrine therapy resistant tumors (Ellis *et al.*, 2012). Mutations within the PI3K and p38 MAPK pathways tend to be mutually exclusive. However, each individual component of a pathway was affected with approximately equal frequency (Ellis *et al.*, 2012; The Cancer Genome Atlas Network, 2012). This variety of potential driver mutations hinders the application of cluster-based tools by introducing variation into the broad sub-classes. Identifying these driver events in each new tumor can be challenging.

The challenge to identify driving mutations is made even more complex by epigenetic gene regulation. Epigenetic mechanisms like those that silence ERα can either silence or activate genes and thereby act in the same direction as gain or loss of function alleles. Unlike genetic mutations, however, epigenetic events may occur simultaneously at multiple genes within a pathway (Jones and Baylin, 2007). This phenomenon increases the number of driver events that may occur in any given tumor. Moreover, the TCGA analysis is designed to cluster tumors into broad treatment groups, but they are not designed to identify the driving mutations within individual tumors. Therefore, a functional genomics approach is needed to identify driving genetic and epigenetic events that can be used to inform treatment of individual cancers.

2.1.9 Aberrant DNA Methylation Promotes Endocrine Therapy Resistance

Because of the role that DNA methylation plays in gene regulation (see Chapter 1: Introduction), aberrant DNA methylation could cause events that drive endocrine therapy resistance. The epigenetic silencing of *ESR1* described above is one possible mechanism by which this could occur (Zilli *et al.*, 2009). Because *HRAS* upregulates ERK growth signaling, its hypomethylation is another example of an epigenetic event promoting endocrine therapy resistance. Finally, hypermethylation correlated with repression of *PIK3R3*, the ERα-regulated p55 regulatory subunit of PI3K, in fulvestrant resistant MCF7 (Fan *et al.*, 2006). Interestingly, this study also showed that 86% of the promoters differentially methylated during the development of fulvestrant resistance in MCF7 were hypomethylated. This observation suggests that further demethylation occurs as breast cancer cells develop endocrine therapy resistance.

The genome-wide hypomethylation that occurs during carcinogenesis and continues during the development of endocrine therapy resistance demethylates many retrotransposons.

This occurs because retrotransposons represent a significant proportion of both methylated DNA and the human genome as a whole (Ehrlich, 2009). Hypomethylation at these loci is therefore almost inevitable. As discussed above, DNA methylation is important for preventing retrotransposon expression. Moreover, the significant contribution of retrotransposon promoters to the human transcriptome has been recognized (Faulkner *et al.*, 2009). By altering normal gene regulation, unmethylated retrotransposons could contribute to endocrine therapy resistance.

2.1.10 Functional Epigenomics Applied to Endocrine Therapy Resistance

We therefore adopted a functional epigenomics approach to explore the role of DNA methylation in endocrine therapy resistance. Specifically, we sought to identify annotated genes and/or retroelements whose gene expression strongly negatively correlated with altered DNA methylation. Candidate loci were identified based on data from Methyl-MAPS and RNA-seq. Methyl-MAPS determines methylation at single CpG resolution genome-wide via sequencing libraries generated from methylation-sensitive or methylation-resistant restriction enzyme digests of genomic DNA (Edwards *et al.*, 2010). It is capable of analyzing methylation at retroelements. These assays were performed on ERα-positive cell lines before and after they were grown in the absence of estradiol for several months (Fig. 2.3). These long-term estrogen deprived (LTED) cell lines were created by the Ellis lab to model acquired AI therapy resistance (Sanchez *et al.*, 2011). To date, we have performed both Methyl-MAPS and RNA-seq on the following LTED cell lines: MCF7, T47D, HCC1428, and MDA-MB-415.

These experiments identified 7 candidate genes in MCF7 cells (Table A.1) and 39 candidate genes in T47D cells (data not shown). These genes were selected by manual inspection and were chosen for correlated, strikingly clear methylation and expression changes. *WISP2* was

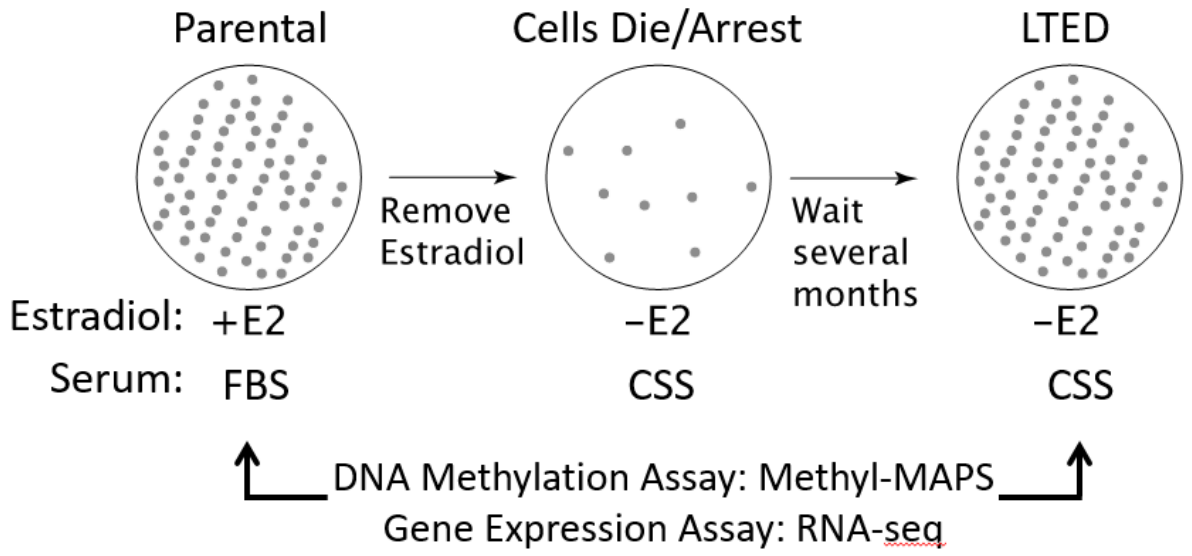


Figure 2.3: LTED cell line generation. An ERα positive breast cancer cell line (the parental cell line) is grown in estrogen deprived conditions. The cells that proliferate after overcoming the deprivation are the LTED cell line. FBS = fetal bovine serum. CSS = charcoal stripped serum.

a MCF7 candidate gene (Fig. A.1) known to promote estrogen-independent growth in MCF7 cells (Fritah *et al.*, 2008). Thus, our screen successfully identified a gene known to affect resistance to estrogen deprivation, the desired phenotype.

Urothelial cancer associated 1 (*UCA1*) was another candidate gene identified by our screen. In T47D LTED cells, *UCA1* is hypomethylated downstream of its promoter. This hypomethylation correlates with a clear increase in expression relative to the original T47D cell line. This candidate is particularly interesting because *UCA1* transcription begins in a LTR retrotransposon promoter on chromosome 19p13.12 (Wang *et al.*, 2008). Because retrotransposon promoters are normally methylated, *UCA1* therefore fits our hypothesis well: *UCA1* represents a locus where endocrine therapy-linked hypomethylation causes aberrant transcription which contributes to the development of endocrine therapy resistance. Moreover, *UCA1* is a long non-coding RNA (lncRNA) that promotes bladder cancer.

lncRNAs are transcripts >200 nt long without protein coding potential, and most lncRNA have no assigned function (Engreitz *et al.*, 2016). Of those with known function, however, many are linked to carcinogenesis or tumor progression. The lncRNA *HOTAIR*, for example, retargets Polycomb Repressor Complex 2 in breast cancer (Gupta *et al.*, 2010). The resulting altered chromatin structure creates an embryonic fibroblast-like transcription profile that promotes metastasis. Other examples include *HULC*, which promotes phosphorylation and activation of the CREB transcription factor by binding miR-372 and preventing it from downregulating CREB in liver cancers (Gibb *et al.*, 2011). In contrast, *H19* lncRNA transcripts are the source of miR-672 that targets the tumor suppressor *RB* in breast and other cancers (Gibb *et al.*, 2011). As a lncRNA, it is therefore quite possible that *UCA1* could support endocrine therapy resistance in breast cancer.

Furthermore, *UCA1* was first identified as a urine marker for bladder cancer (Wang *et al.*, 2006). The BLZ-211 bladder cancer cell line expressed three *UCA1* isoforms: 1.4 kb, 2.2 kb, and 2.7 kb (Wang *et al.*, 2008). Overexpression of either the 1.4 kb or 2.2 kb isoforms is sufficient to promote proliferation, colony formation, motility, invasion, and drug resistance in two different bladder cancer cell lines (Wang *et al.*, 2008, 2012). Further research has suggested that these phenotypes might be caused by increased PI3K and ERK signaling (Yang *et al.*, 2012). Based on this preliminary data, I hypothesized that *UCA1* contributed to acquired resistance to estrogen deprivation in our cancer cell line models. I tested this hypothesis by measuring the effect of *UCA1* knockdown/overexpression on cell growth, cell cycle progression, and apoptosis.

Prostaglandin E2 receptor 4 (PTGER4) was a third candidate identified by our screen. EP4, the *PTGER4* gene product, is a G-protein coupled receptor that activates the PI3K signaling

pathway as well as the adenylyl cyclase (AC) and protein kinase A (PKA) pathway in response to prostaglandin E_2 (Yokoyama *et al.*, 2013). EP4 contributes to breast cancer via several mechanisms. EP4 signaling contributes indirectly by inhibiting the activity of natural killer cells (Fulton and Heppner, 1985; Kundu *et al.*, 2009). More directly, EP4 activation promotes breast cancer metastasis (Ma *et al.*, 2006, 2012; Xin *et al.*, 2012), and downstream PI3K activation promotes tumor cell survival (Xin *et al.*, 2012). The activation of the AC pathway can also upregulate aromatase expression, stimulating estrogen production that will aid ERα positive tumors (Subbaramaiah *et al.*, 2008). These previous findings support our hypothesis that *PTGER4* upregulation may contribute to acquired estrogen deprivation resistance in our cell line model.

2.1.11 Epigenetic Activation of *PTGER4* Supports Acquired Endocrine Therapy Resistance

We have tested the hypothesis that epigenetic activation of *PTGER4* aids the acquisition of endocrine therapy resistance via downstream cell signaling, and published the results (Hiken *et al.*, 2017). I was interested in this project and performed follow-up experiments to bolster the main body of data. I also contributed significantly to the writing of the manuscript. Immediately following is the data gathered by my colleagues. The data I produced in the results section.

The *PTGER4* gene has a CpG island associated with its promoter that lost methylation in the transition from MCF7 to MCF7-LTED cells (Fig. 2.4A). Methylation changes were validated using Methyl-Screen (Fig. A.2A, Holemon *et al.*, 2007). This loss of methylation is accompanied with a greater than 20-fold gain in *PTGER4* expression as indicated by RNA-seq (Fig. 2.4A, Parental cpm = 1.9, 8.9; LTED cpm = 226.4, 175.5) and confirmed by RT-qPCR (Fig. A.2B).

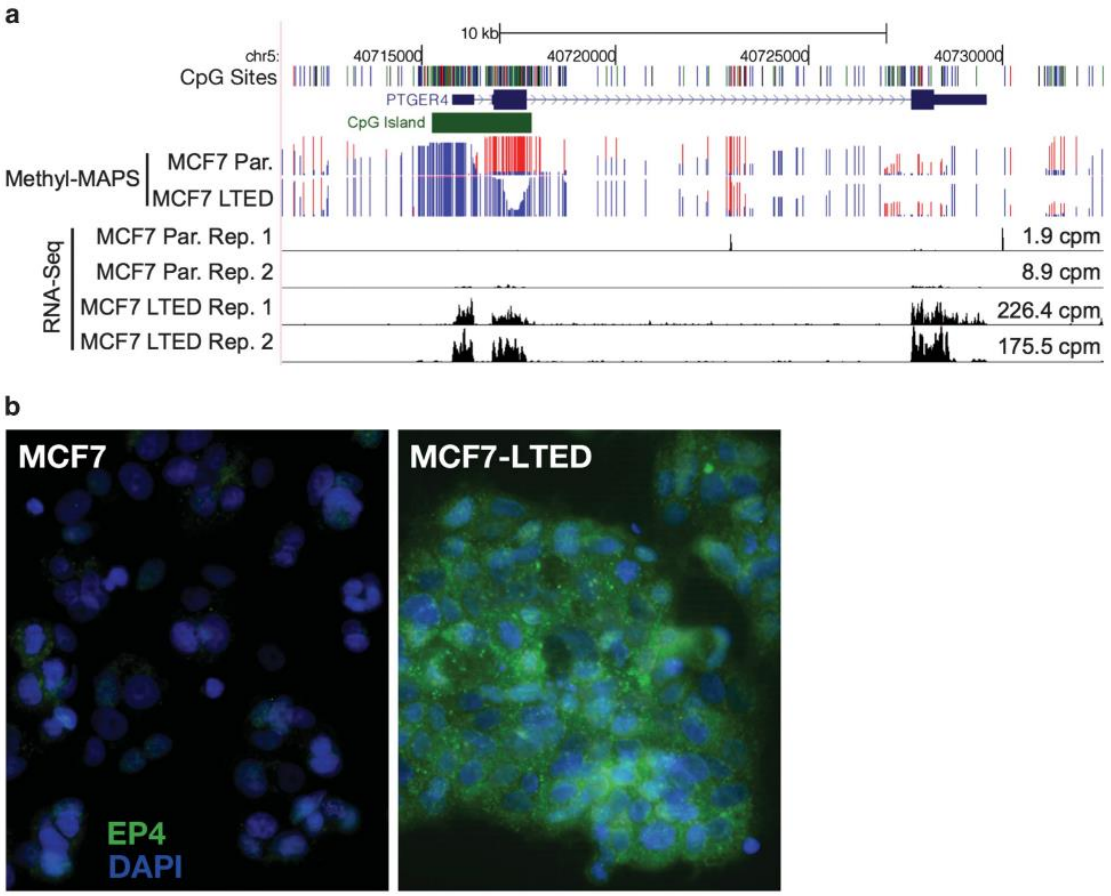


Figure 2.4: PTGER4 methylation and expression in MCF7-LTED cells. (A) Genome browser view of Methyl-MAPS methylation and RNA-seq expression data for *PTGER4*, the gene that encodes EP4. Red and blue lines indicate coverage of methylated and unmethylated fragments, respectively. Individual CpG sites are noted by tics in black at the top track. **(B)** Immunofluorescence staining of EP4 (green) and DAPI (blue) in MCF7 and MCF7-LTED cells.

Furthermore, MCF7 cells treated with 5-azacytidine expressed higher levels of *PTGER4* (Kim *et al.*, 2012), consistent with the hypothesis that DNA hypomethylation directly regulates expression. Increased mRNA levels in MCF7-LTED cells were accompanied with increases in EP4 protein expression as observed by immunofluorescence (Fig. 2.4B). Increased EP4 expression in MCF7 LTED cells was also accompanied by increased cAMP in response to the EP4 agonist L-902,688 (Fig. 2.6B).

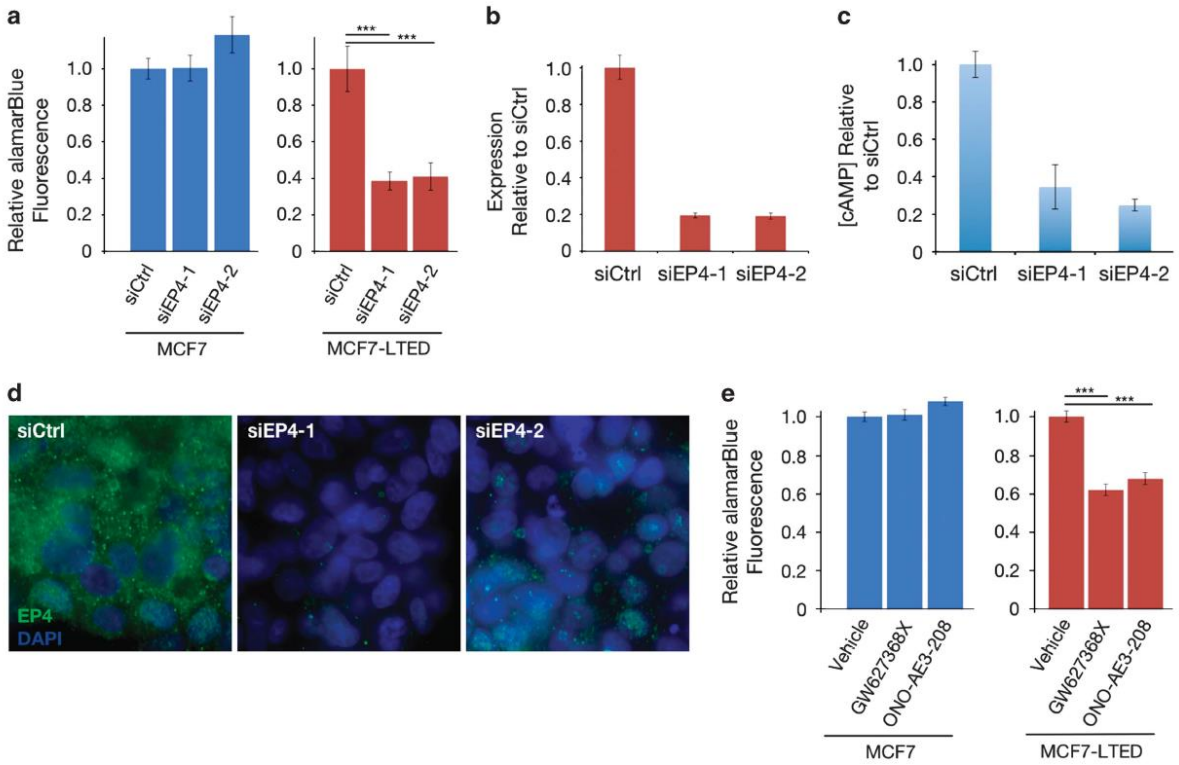


Figure 2.5: Knockdown or inhibition of EP4 signaling decreases estrogen-independent cell proliferation. (A) Proliferation of MCF7 cells, which express little EP4, and MCF7-LTED cells treated with EP4 siRNA relative to cells treated with negative control siRNA. **(B)** RT–qPCR analysis of *PTGER4* expression decreases in MCF7-LTED cells treated with two distinct EP4 siRNAs relative to control siRNA. Error bars are SD for three technical replicates. **(C)** cAMP levels in MCF7-LTED cells treated with EP4 agonist decrease in cells treated with EP4 siRNAs relative to control siRNA. **(D)** Immunofluorescence images of EP4 (green) and DAPI (blue) in MCF7-LTED cells treated with siRNAs targeting EP4 and siRNA controls. **(E)** Cell proliferation of MCF7 and MCF7-LTED cells treated with EP4 antagonists relative to cells treated with vehicle only. ***P<0.001. Error bars show SEM of three replicates.

EP4 inactivation decreases estrogen-independent cell growth in MCF7-LTED cells

We sought to determine whether EP4 knockdown would affect the ability of MCF7-LTED cells to grow in the absence of estrogen. EP4 was knocked down using two EP4-targeted small interfering RNAs (siRNAs) in both MCF7 and MCF7-LTED cells. Visual inspection indicated siEP4s drastically reduced cell proliferation in comparison with siRNA controls. We quantified the attenuated growth using reduction in alamarBlue to measure relative cellular proliferation

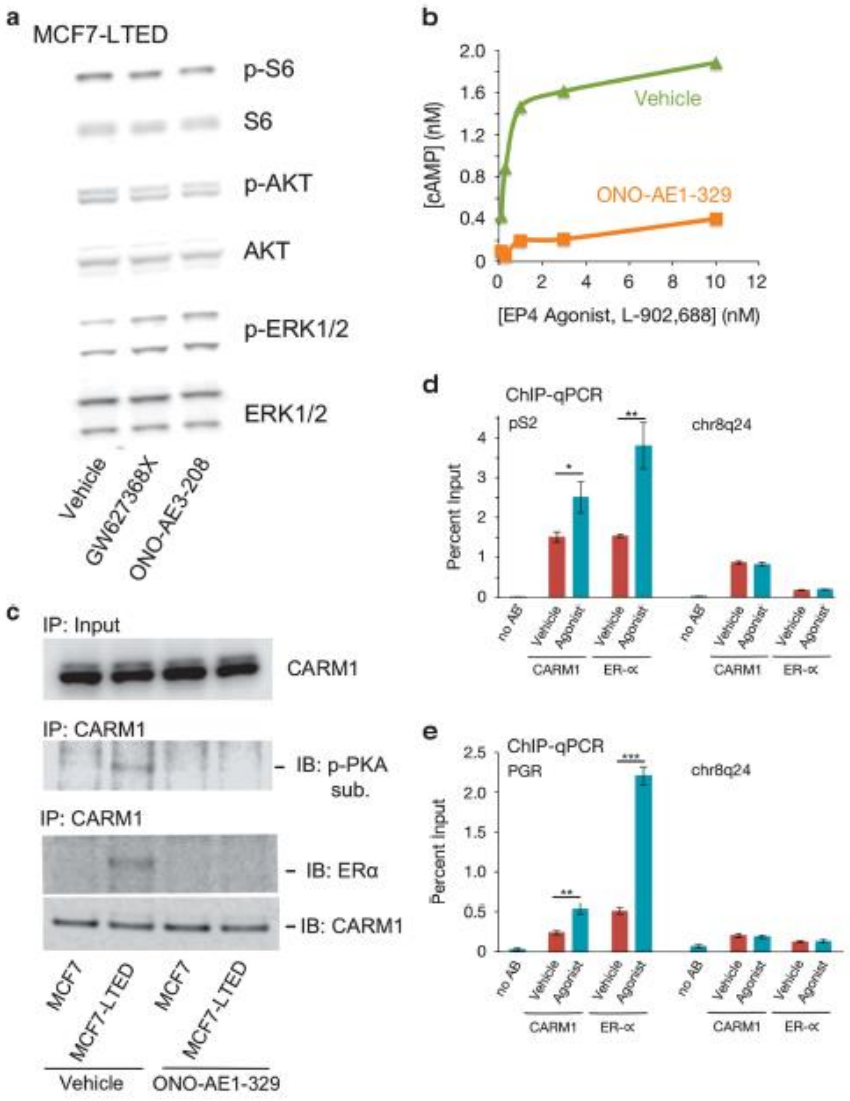


Figure 2.6: EP4 activates cAMP signaling and ligand-independent ER activation through CARM1. (A) Western blot analysis of AKT, mammalian target of rapamycin (mTOR) S6 kinase, and MAPK (ERK1/2) pathways in MCF7-LTED cells treated with vehicle or EP4 antagonists. **(B)** Dose response curve measuring induction of cAMP by EP4 agonist (L-902 688) in MCF7-LTED cells also treated with 10 nM EP4 antagonist (ONO-AE1-329) or vehicle. **(C)** Immunoblot of CARM1 in MCF7 and MCF7-LTED cells treated with EP4 antagonist or vehicle in the absence of estrogen. CARM1 immunoprecipitation (IP) followed by western blot analysis with antibodies for phosphorylated PKA substrate, ERα, or CARM1. **(D-E)** ChIP-qPCR of ERα targets (D) pS2 and (E) PGR with either no antibody (noAB), CARM1 antibody, or ERα antibody in MCF7-LTED cells treated with EP4 agonist or vehicle. Chr8q24 is a gene desert-negative control region. *P-value < 0.05, **P-value < 0.01, and ***P-value < 0.001. Error bars are SEM of three replicates. IB, immunoblot; p-, phosphorylated.

7 days after transfection (Fig. 2.5A). We verified knockdown at the mRNA level by RT-qPCR (Fig. 2.5B). Since appropriate antibodies could not be found for Western blot analysis, we validated a decrease in EP4 protein by immunofluorescence (Fig. 2.5D) and by measurement of the secondary messenger cAMP, which is produced as a result of EP4 activity (Fig. 2.5C). MCF7-LTED cells treated with siRNAs targeting EP4 mRNA showed a 61% and 59% average reduction in cell proliferation relative to control siRNAs ($p < 10^{-7}$ for each siRNA, Fig. 2.5A). These siRNAs showed no effect in MCF7 cells, which have negligible levels of EP4 expression at the mRNA and protein levels and thus provide a control for off-target effects.

We further validated the importance of EP4 for estrogen-independent growth in MCF7-LTED cells using two EP4-specific antagonists that do not inhibit other prostaglandin receptor subtypes: GW627368X and ONO-AE3-208. We observed an average 32% reduction in cellular proliferation for GW627368X-treated and an average 38% reduction for ONO-AE3-208-treated MCF7-LTED cells relative to vehicle alone ($p < 10^{-7}$ for each antagonist, Fig. 2.5E). Again, MCF7 cells, which have negligible EP4 expression, showed no significant proliferation reduction when treated with antagonists.

Functional analysis of EP4 signaling in MCF7-LTED cells

EP4 activates two pro-growth signaling pathways: PI3K and PKA; both of which could promote endocrine therapy resistance. MCF7-LTED cells show increased Akt and PI3K activity, and short-term estrogen deprived MCF7 cells are sensitive to the BGT226 and BKM120 PI3K inhibitors (Sanchez *et al.*, 2011). Additionally, EP4 activation can result in phosphorylation and activation of Akt (Yokoyama *et al.*, 2013). However, our MCF7-LTED cells are resistant to the effects of the BGT226 and BKM120 PI3K inhibitors (Sanchez *et al.*, 2011). Analysis of Akt and

PI3K phosphorylation also showed no changes after treatment of MCF7-LTED cells with EP4 antagonists (Fig. 2.6A). In addition, we did not find changes in mTOR or MAPK signaling. This suggests that other pathways are likely activated by EP4 in our MCF7-LTED model.

We next hypothesized that *PTGER4* might contribute to endocrine therapy resistance via AC. EP4 activates AC in addition to Akt, and AC activity can induce ERα phosphorylation. Once activated AC produces cAMP as its secondary messenger, and cAMP promotes proliferation through downstream effectors such as PKA (Yokoyama *et al.*, 2013). We observed that both the concentration of cAMP in LTED cells and cell proliferation decreased after EP4 reduction by siRNA knockdown (Fig. 2.5A,C). This suggested that EP4 promotes proliferation in LTED cells via AC and cAMP. To confirm this conclusion, we applied the EP4 specific agonist L-902,688 to MCF7-LTED cells and observed a dramatic increase in cAMP which went away upon inhibition of EP4 with antagonists (Fig. 2.6B).

A recent report showed that cAMP activates CARM1 via phosphorylation by PKA and drives ligand-independent ERα transcriptional activity to promote resistance to tamoxifen (Carascossa *et al.*, 2010). CARM1 is an estrogen-dependent ERα co-activator and phosphorylated CARM1 binds directly to ERα (Chen *et al.*, 1999). We thus tested whether a similar mechanism caused ligand independent ER signaling in MCF7-LTED cells. CARM1 proteins levels remained unchanged in MCF7-LTED and parental cells in the absence of estrogen even after treatment with EP4 antagonist (Fig. 2.6C). We found however that CARM1 acted as a PKA substrate in MCF7-LTED cells in the absence of estrogen and this interaction went away after treatment with EP4 antagonist. We further found that ERα showed a strong interaction between CARM1 and ERα in MCF7-LTED cells based on immunoprecipitation of

CARM1 followed by immunoblotting with an ERα-specific antibody (Fig. 2.6C). Treatment with EP4 antagonist removed this interaction, suggesting that the CARM1-ERα interaction is dependent on EP4 signaling. As has been reported previously (Chen *et al.*, 1999), in estrogen deprived conditions MCF7 cells showed no activation of CARM1 or association with ERα.

We next used chromatin-immunoprecipitation (ChIP)-qPCR to interrogate the relationship between EP4 and ligand independent binding of ERα in MCF7-LTED cells at two ERα binding sites, the pS2 and PGR promoters. MCF7-LTED cells were treated with agonist or vehicle and subjected to ChIP-qPCR with antibodies to CARM1 or ERα at the pS2 promoter, the PGR promoter, and a negative control gene desert region on chr8q24 (Fig. 2.6D,E). Both CARM1 and ERα were found to have increased binding upon treatment with EP4-specific agonist relative to cells treated with vehicle. No such gain was observed in the gene desert region. This implies that EP4 signaling can activate ligand-independent ER activation mediated through CARM1.

ER signaling in MCF7-LTED cells

To understand whether ligand-independent ER signaling is active in MCF7-LTED cells, we compared gene expression in MCF7-LTED cells relative to MCF7 cells that recovered with the addition of estradiol after short-term estrogen deprivation (1 day) using RNA-seq. Using MCF7 cells treated with estradiol we defined a signature of ER signaling based on genes up-regulated two-fold or more. Gene Set Enrichment Analysis (GSEA, Subramanian *et al.*, 2005) shows that the majority of ER responsive genes are also up-regulated in MCF7-LTED ($p < 10^{-7}$, Fig. A.3). This is consistent with the finding that our MCF7-LTED cells are responsive to the

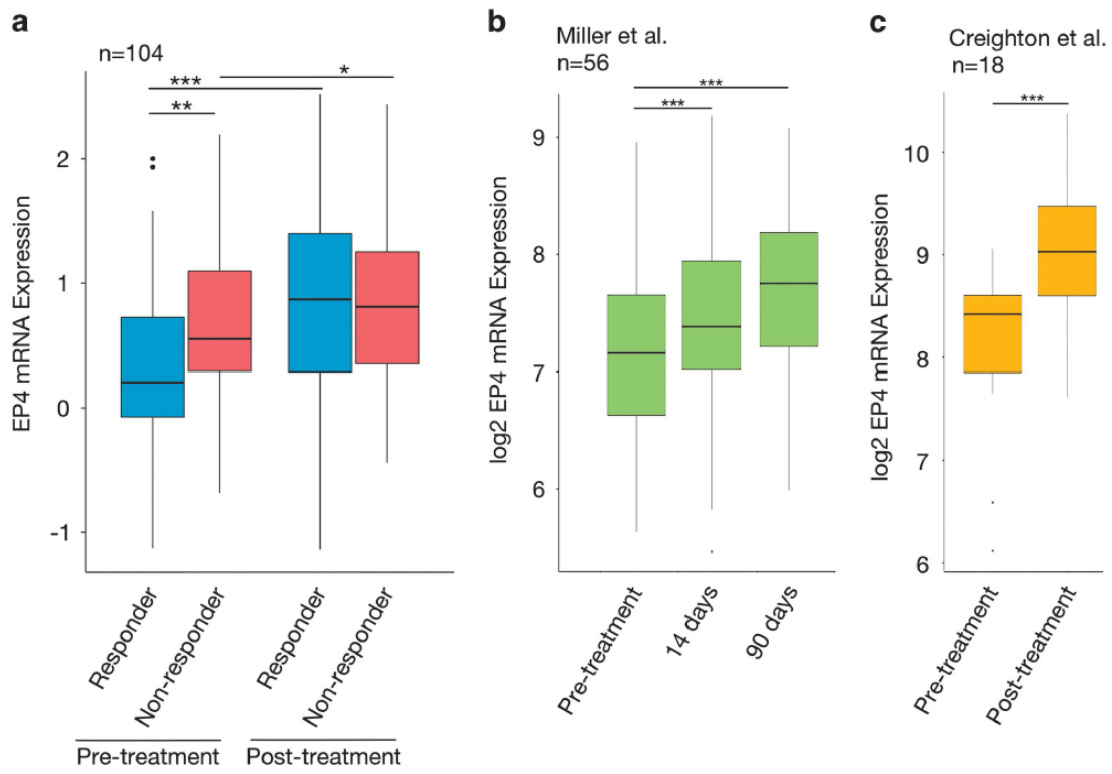


Figure 2.7: EP4 expression is higher in patients that fail to respond to neoadjuvant endocrine therapy. (A) Boxplots of EP4 expression from 104 TCGA tumor samples. Responders and non-responders were defined based on Ki67 levels after neoadjuvant therapy. Non-responders had Ki67 levels greater than 10%. **(B)** Expression data from Miller *et al.* (2011) showing an increase in EP4 expression during prolonged neoadjuvant therapy. **(C)** Expression data from Creighton *et al.* (2009) for pre- and post-neoadjuvant endocrine therapy. *P-value < 0.05, **P-value < 0.01 and ***P-value < 0.001. Error bars are SEM.

ER-antagonist fulvestrant (Sanchez *et al.*, 2011) and indicates that ligand-independent ER-signaling is active in MCF7-LTED cells at a genomic level.

EP4 mRNA expression increased in ER+ breast cancer in the neoadjuvant setting

To assess the potential clinical relevance of our findings, we tested the hypothesis that EP4 expression increases are associated with initial sensitivity to aromatase inhibitor endocrine therapies. Expression profiling was performed for 104 patients before and after neoadjuvant aromatase inhibitor therapy. 74 tumors were determined as ‘aromatase-inhibitor-resistant’ and 30 as ‘aromatase-inhibitor-sensitive’ based on the presence of the proliferation marker Ki67 after

treatment (see details in methods). Ki67 is a well-established indicator of clinical outcome in the neoadjuvant setting (Ellis *et al.*, 2008). We observed that EP4 expression was higher in patients that demonstrated resistance to AI-therapy versus patients that responded to AI-therapy (Fig. 2.7A), which is consistent with EP4 expression playing a role in AI-resistance. Interestingly, EP4 expression also significantly increased during neoadjuvant therapy in both responder and non-responder tumor samples. Data from two smaller studies examining mRNA changes during neoadjuvant therapy verify this finding (Fig. 2.7B,C; Fig. A.4, Creighton *et al.*, 2009; Miller *et al.*, 2011). This suggests that in addition to promoting resistance to AIs, EP4 activation may be a response to the loss of ER signaling in ER+ tumors.

2.2 Results

2.2.1 *PTGER4* Promotes Acquired Endocrine Therapy Resistance

The *PTGER4* experiments described in the introduction led to the conclusion that *PTGER4* becomes hypomethylated and upregulated during LTED treatment and promotes ERα signaling via phosphorylation of CARM1. That body of work represented a complete story; however, there were still several unknowns. For example, we did not show that *PTGER4* expression increase occurred early enough in the adaptation process to aid the adaptation to LTED conditions. Similarly, we showed validated methylation data that is negatively correlated with expression but do not establish that the DNA methylation causes the change in expression. Finally, we claimed that CARM1 is the downstream effector of the EP4 activity and necessary

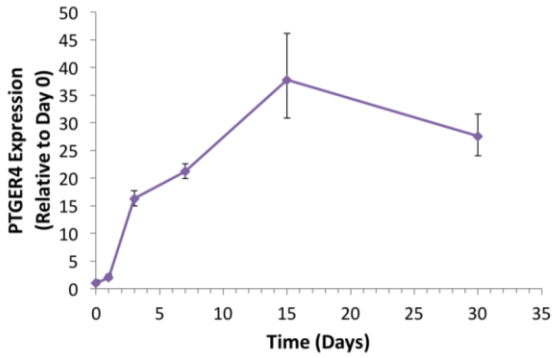


Figure 2.8: Expression of PTGER4 in MCF7 cells after short-term withdrawal of estrogen. RT-qPCR expression data for PTGER4 in MCF7 cells grown in LTED media for 30-days. Expression data is relative to the sample from MCF7 parental cells on Day 0 before changing to LTED media. Error bars are the SD of three technical replicates.

for T47D LTED proliferation, but we do not demonstrate that this is the case. I therefore set out to investigate these additional aspects of the project.

I began by studying the expression of *PTGER4* during short-term estrogen deprivation (STED, the first 30 days after estrogen withdrawal). Within three days, MCF7 parental cells upregulate *PTGER4* about 16-fold at the transcript level (Fig. 2.8). This indicates that *PTGER4* upregulation occurs rapidly and can aid the development of endocrine therapy resistance.

To support our claim that DNA methylation regulates *PTGER4* expression, I analyzed the DNA methylation levels at the *PTGER4* exon 2. Exon 2 is the location of methylation loss in MCF7 LTED cells (Fig. 2.4), and I expected the increased expression of *PTGER4* to correlate with a decrease in DNA methylation. Interestingly, hypomethylation did not occur during STED (Fig. A.5A). It therefore became important to test whether the demethylating agent 5-azacytidine could increase *PTGER4* expression. 5-azacytidine induced a three- to four-fold increase in *PTGER4* expression (Fig. 2.9). This indicates that DNA methylation controls *PTGER4* expression. Again, however, DNA methylation levels did not decrease (Fig. A.5B).

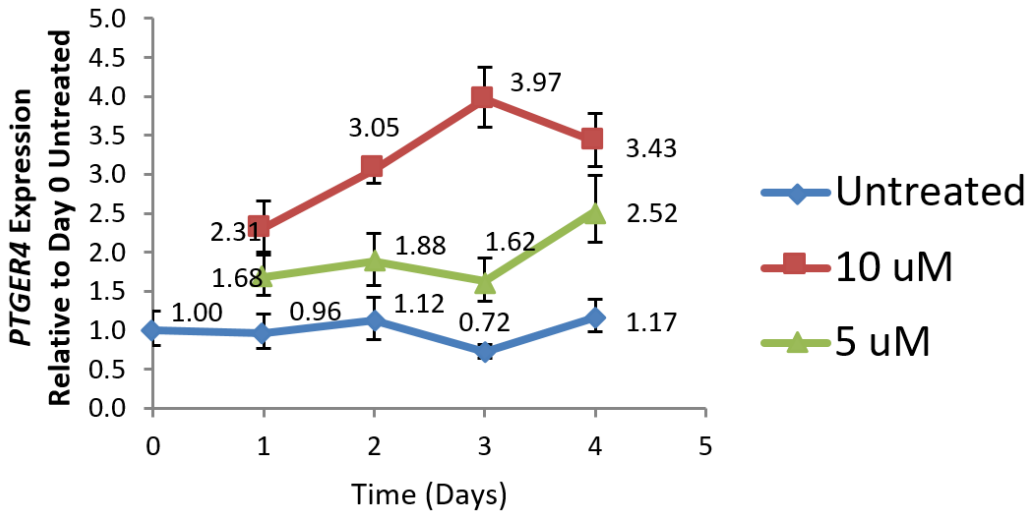


Figure 2.9: 5-azacytidine treatment increases PTGER4 expression. RT-qPCR data for *PTGER4* expression in MCF7 cells treated with 5 uM and 10 uM 5-azacytidine for four days. Expression data is relative to the untreated sample from Day 0. Error bars are +/- 1 SD of three technical replicates.

The lack of a decrease in DNA methylation levels is consistent with what we know about development of resistance to LTED conditions. After withdrawal of estrogen, most of the cells die or arrest growth (Fig. 2.3). The estrogen withdrawal selects from the heterogenous mix of struggling cells that can overcome the challenge. The clonal expansion of the cell(s) gives rise to the LTED line. This explains the STED data. During the 30-day STED experiment, a small number of cells have lost methylation and overexpress *PTGER4*. The expression increase is a much stronger signal and is detected, but the loss of methylation remains masked by the methylation signal from the other cells.

Lastly, we claimed EP4 activity promoted endocrine therapy by phosphorylating and activating CARM1. CARM1 in turn binds to ERα and promotes endocrine therapy resistance via ligand-independent ERα activity. To confirm the role of CARM1, I showed that *CARM1* knockdown caused decreased proliferation in T47D cells (Fig. 2.10).

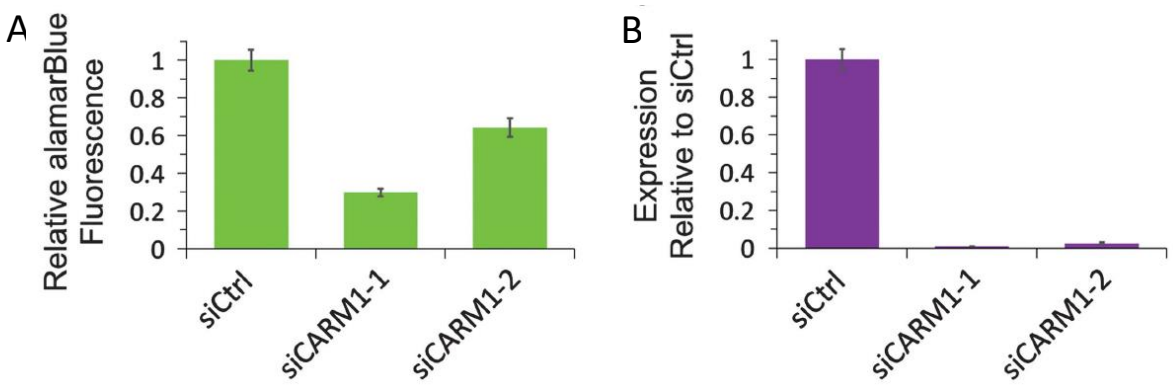
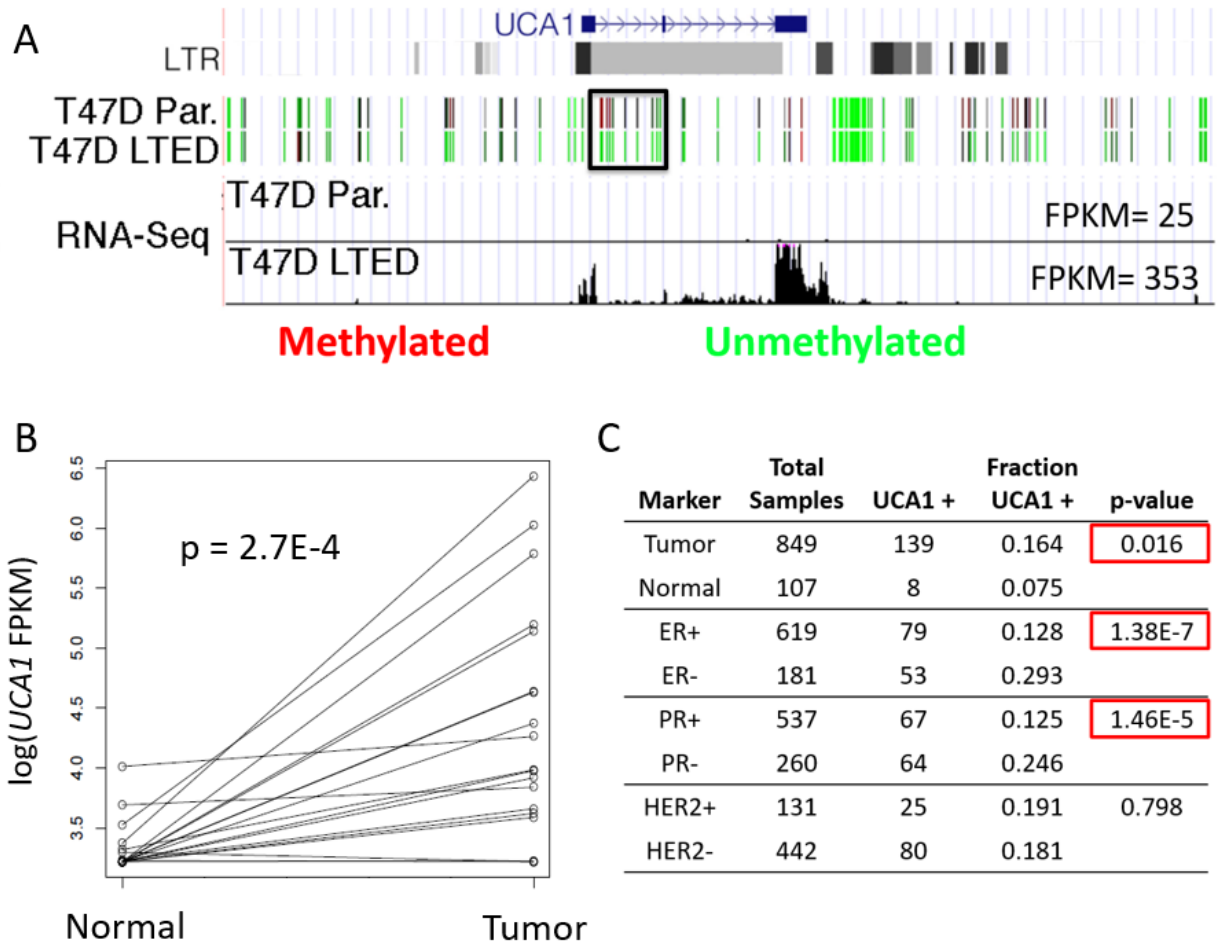


Figure 2.10: CARM1 knockdown inhibits proliferation. (A) Proliferation of MCF7-LTED cells treated with *CARM1* siRNA relative to MCF7-LTED cells treated with negative control siRNA. **(B)** RT–qPCR analysis of *CARM1* expression after *CARM1* knockdown in MCF7-LTED, matching data from (A). Error bars = +/- 1 SD of three technical replicates.

2.2.2 *UCA1* Support for Acquired Endocrine Therapy Resistance is Inconclusive

The potential role for *UCA1* in acquired endocrine therapy resistance was suggested by its hypomethylation and upregulation in T47D cells (Fig. 2.11A). *UCA1* was also hypomethylated and upregulated in the tumorigenic HCC1954 cell line relative to the non-tumorigenic HMEC cell line (Fig. A.6). *UCA1* was also upregulated 6.7-fold in HCC712 LTED cells (Fig. A.7). Though a much smaller increase compared to *UCA1* upregulation in T47D LTED, it suggests that *UCA1* activity is not limited to a single cell line model. *UCA1* is more highly expressed in matched breast cancer tumor/normal pairs from The Cancer Genome Atlas (Fig. 2.11B). *UCA1* expression is also higher in ERα positive and progesterone receptor positive cells (Fig. 2.11C). This *in vivo* data tenuously suggests that *UCA1* may be functional in ERα positive tumors. If confirmed, the hypothesized role may therefore be active in real tumors and could be studied in multiple cell line models.



C

Marker	Total Samples	UCA1 +	Fraction UCA1 +	p-value
Tumor	849	139	0.164	0.016
Normal	107	8	0.075	
ER+	619	79	0.128	1.38E-7
ER-	181	53	0.293	
PR+	537	67	0.125	1.46E-5
PR-	260	64	0.246	
HER2+	131	25	0.191	0.798
HER2-	442	80	0.181	

Figure 2.11: UCA1 expression correlates with methylation changes and tumor status. (A) Genome browser shot of DNA methylation and expression at the *UCA1* promoter. Long terminal repeat (LTR) track is shown for reference. **(B)** Expression level of *UCA1* in matched tumor/normal breast cancer samples from The Cancer Genome Atlas (TCGA). Data for 19 of 107 matched samples where *UCA1* > 25 FPKM. P-value from Wilcoxon signed-rank test. **(C)** *UCA1* expression correlated with clinical markers. All TCGA breast cancer samples included. Clinical marker status determined by the TCGA. *UCA1* positive if > 25 FPKM. P-value from proportion test.

To test my hypothesis, I first assayed the effect of *UCA1* overexpression and knockdown on the proliferation of T47D cells. A *UCA1* overexpression construct was introduced into T47D parental cells with an empty vector control. The over 100-fold higher expression of *UCA1* did not increase parental T47D proliferation in the presence of estrogen (Fig. 2.12A). In estrogen-deprived conditions, there was also no difference (Fig. 2.12A). This lack of change suggests

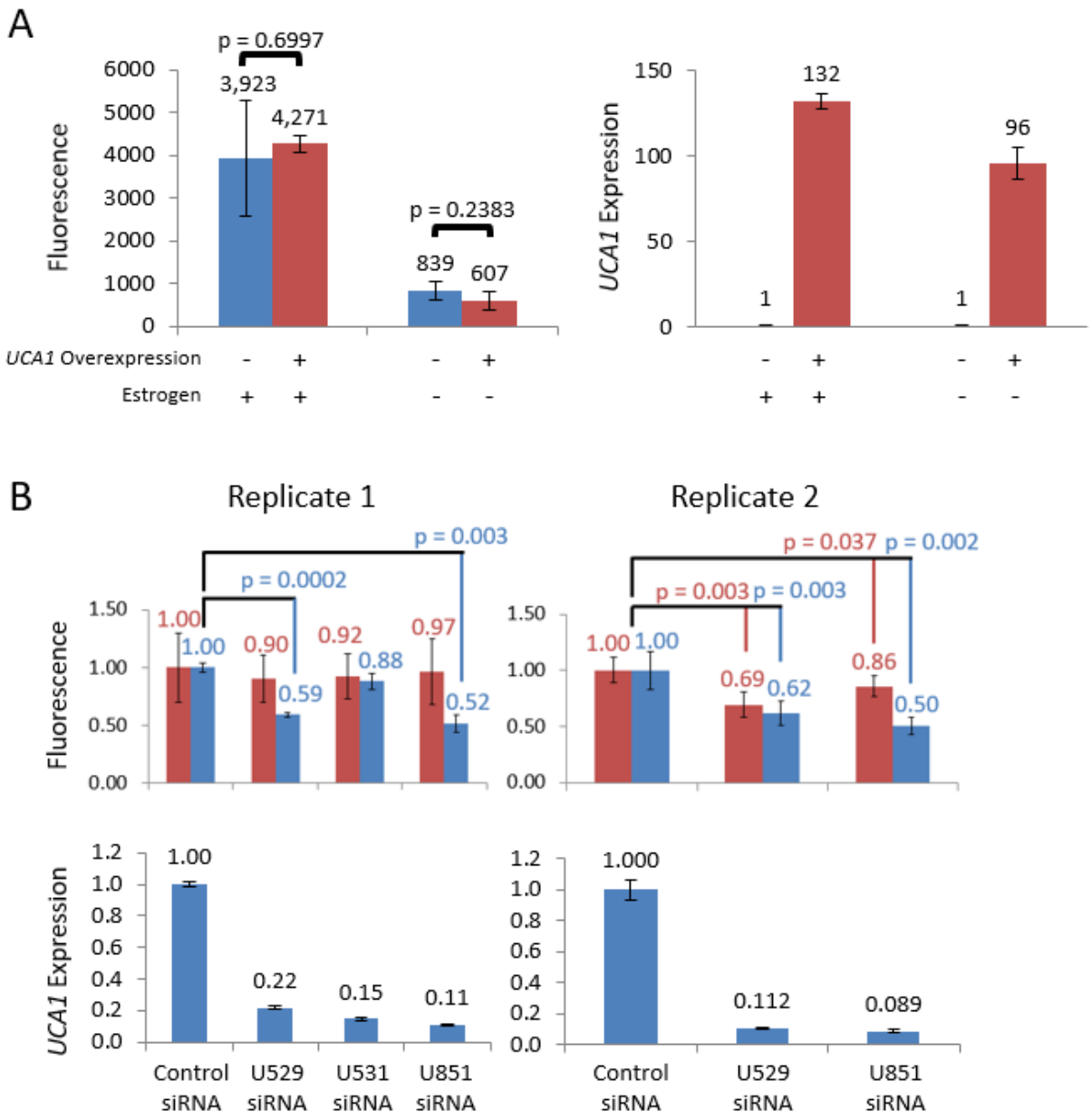


Figure 2.12: UCA1 overexpression and knockdown cause inconsistent changes in T47D proliferation. (A) Left: AlamarBlue fluorescence data from T47D parental cells treated with empty overexpression vector or *UCA1* overexpression vector in both with and without estrogen in the media. Data represent average of three repetitions Error bars are +/- 1 SD. P-value from Welch's test. Right: representative RT-qPCR data from one of the three runs averaged in the alamarBlue data. Error bars are +/- 1 SD of two technical replicates. **(B)** Top: AlamarBlue fluorescence data T47D Parental and LTED cells treated with control siRNA or *UCA1* siRNA. All samples normalized to control. Error bars are +/- 1 SD. Bottom: matching RT-qPCR data from LTED cells. All samples normalized to control. Error bars are +/- 1 SD of two technical replicates.

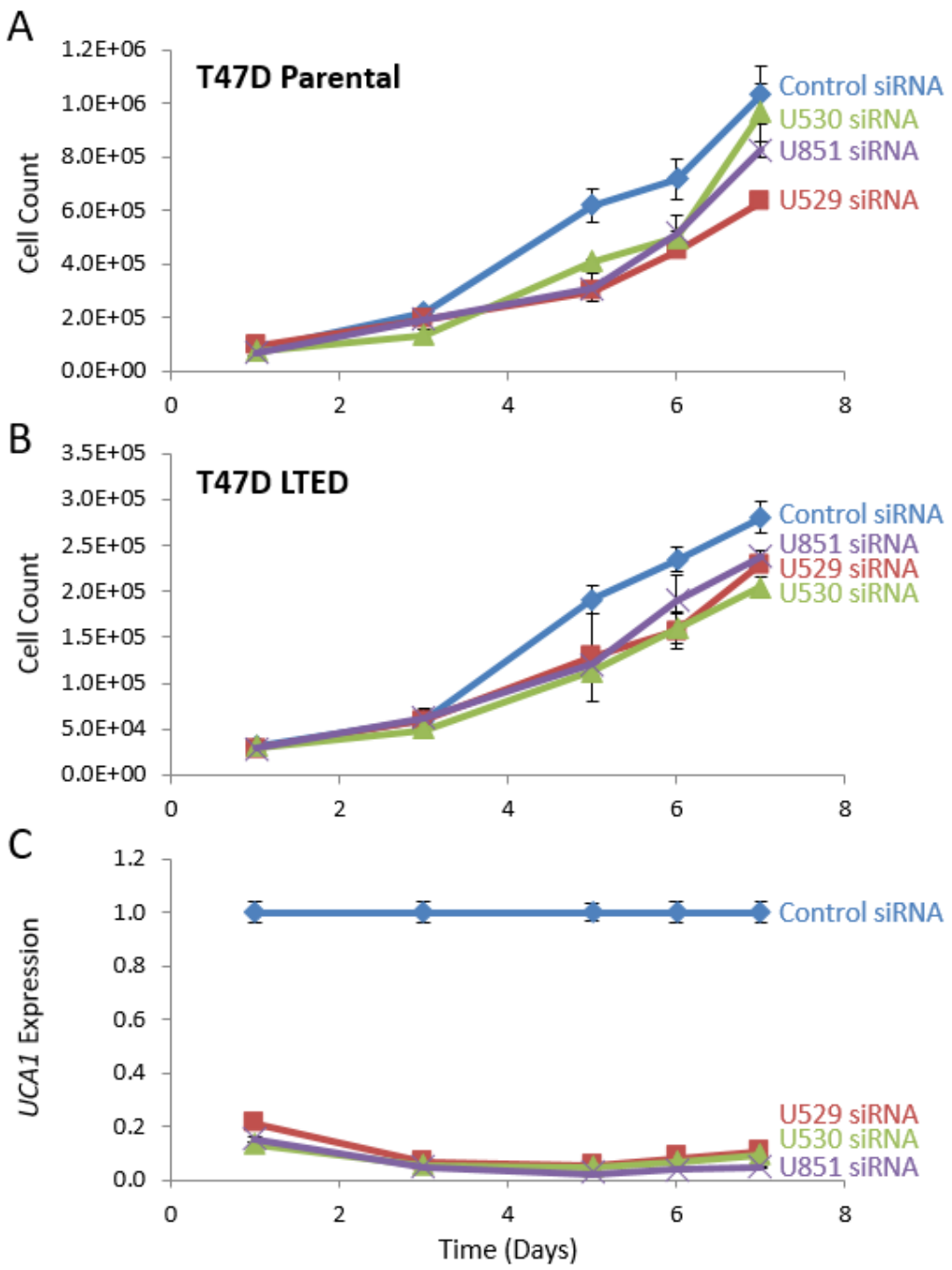


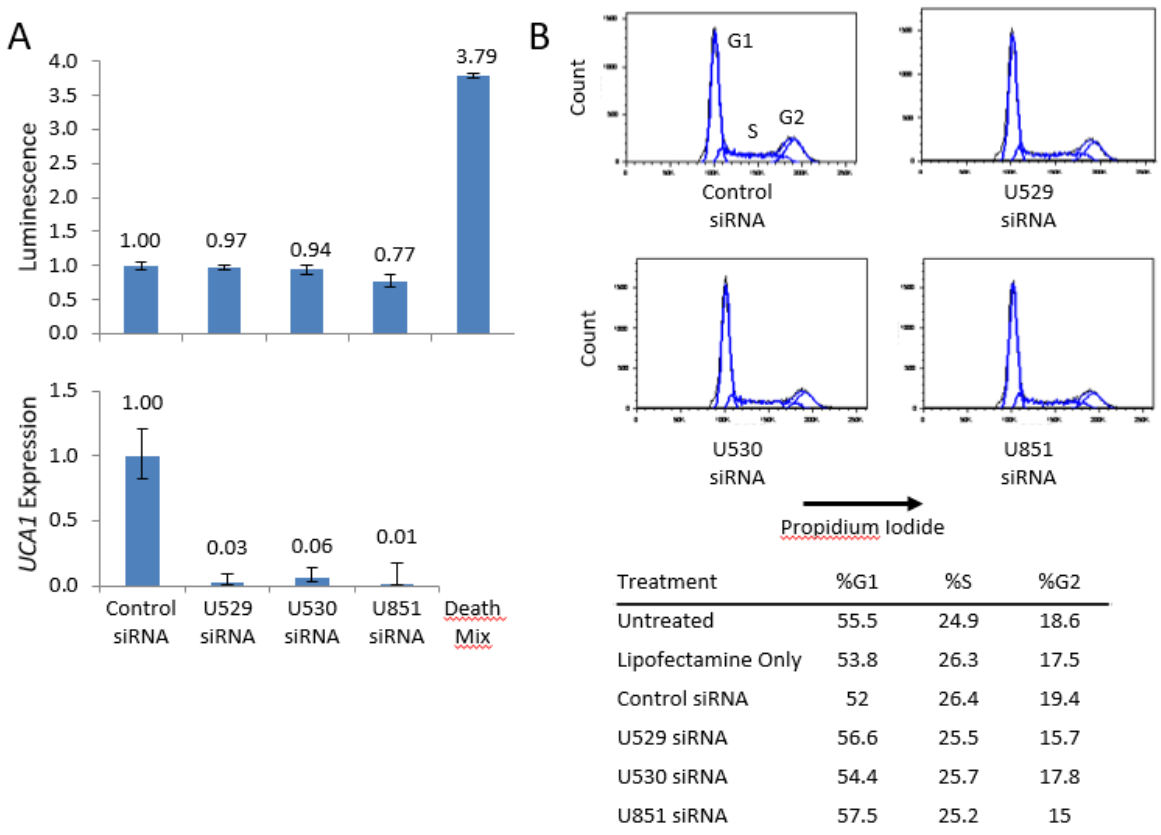
Figure 2.13: Small T47D proliferation decrease measured by cell counting after UCA1 knockdown. (A) Cell count over 7-day time course for *UCA1* knockdown in T47D parental cells. **(B)** Cell count over 7-day time course for *UCA1* knockdown in T47D LTED cells. **(C)** *UCA1* expression measured by RT-qPCR in T47D LTED cells over the same 7-day time course. Each data point is normalized to control expression. Error bars are +/- 1 SD of three technical replicates.

UCA1 does not aid growth in estrogen deprived conditions. However, T47D LTED cells were produced after nine months of growth in estrogen-deprived conditions (Sanchez *et al.*, 2011). Therefore, the short-term nature of the experiment may mean that *UCA1* expression alone is unable to overcome estrogen withdrawal in that time.

UCA1 knockdown in T47D LTED cells caused a 40 to 50% decrease in cell proliferation relative to the control siRNA. The knockdown effects were also sometimes observed in the parental cells (Fig. 2.12B) despite extensive attempts at optimizing the transfection conditions and siRNA concentration. I confirmed that knockdown of *UCA1* occurred in both the cytoplasm and the nucleus (Fig. A.8); therefore, inconsistent knockdown was not the source of the problem.

These inconsistencies could be due to our use of the alamarBlue fluorescence assay, which depends on the assumption that alamarBlue reduction is proportional to total cell count. I therefore attempted to show the effects of knockdown by manually counting cells with a hemocytometer. *UCA1* knockdown produced a small decrease in proliferation; however, this decrease was observed in both parental and LTED cells (Fig. 2.13). Because of the inconsistent effect size and variability in the control parental T47D cells, I could not make a strong conclusion that *UCA1* is necessary for the LTED phenotype.

In addition to increased proliferation, *UCA1* promotes bladder cancer cell survival (Wang *et al.*, 2012) and cell cycle progression (Yang *et al.*, 2012). I therefore tested whether *UCA1* supported these functions in T47D LTED cells. *UCA1* knockdown did not increase apoptosis as measured by a caspase activity assay (Fig. 2.14A). Similarly, *UCA1* knockdown failed to cause a shift in cell cycle progression (Fig. 2.14B).



Treatment	%G1	%S	%G2
Untreated	55.5	24.9	18.6
Lipofectamine Only	53.8	26.3	17.5
Control siRNA	52	26.4	19.4
U529 siRNA	56.6	25.5	15.7
U530 siRNA	54.4	25.7	17.8
U851 siRNA	57.5	25.2	15

Figure 2.14: UCA1 knockdown does not increase T47D apoptosis or cell cycle progression. (A) Top: luminescence data from Caspase-Glo assay, normalized to control. Death mix is positive control treatment provided in Caspase-Glo kit. Error bars are +/- 1 SD. Bottom: matching RT-qPCR data demonstrating *UCA1* knockdown. Error bars are +/- 1 SD. (B) Top: Histograms of propidium iodide signal from T47D LTED cells after *UCA1* knockdown counted by FACS. Bottom: percent of cells in each cell cycle stage calculated by Watson Pragmatic model in FlowJo.

In summary, we sometimes observe the expected proliferation decrease after *UCA1* knockdown. However, *UCA1* knockdown failed to alter apoptosis or cell cycle progression, and *UCA1* did not induce a proliferation increase when overexpressed. Altogether, these data do not provide strong evidence that *UCA1* contributes to acquired endocrine therapy resistance.

2.3 Discussion

In this chapter, I explored whether DNA methylation changes at the promoters of the genes *UCA1* and *PTGER4* promote endocrine therapy resistance in breast cancer. Genome-wide

hypomethylation is common in cancer and can alter gene expression in ways that favor carcinogenesis and metastasis (Ehrlich and Lacey, 2013). *UCA1* could not be conclusively linked to increased survival and proliferation under estrogen withdrawal. In contrast, I was able to provide evidence that *PTGER4* expression is regulated by DNA methylation and contributes to proliferation in low estrogen conditions via activation of CARM1.

The role of EP4 in endocrine therapy resistance is consistent with its contributions to the proliferation of several cancer types including colon, lung, prostate, ovarian and breast (Yokoyama *et al.*, 2013). Antagonists of EP4 have been shown to inhibit metastasis in hormone-resistant murine mammary tumor cells (Ma *et al.*, 2006). Further, the contribution of EP4 to endocrine therapy resistance via cell signaling is consistent with its role in the development of castration-resistant prostate cancer via PKA and cAMP (Terada *et al.*, 2010).

Mutations are a possible but unlikely alternative to epigenetics for activation of EP4. According to the COSMIC database, EP4 mutations are exceedingly rare in breast cancer, occurring in 0.33% of patients and copy number gains in 1.5% (Forbes *et al.*, 2015). Sanger sequencing of *PTGER4* in both MCF7 and MC7-LTED cells failed to uncover any mutations that could contribute to the increased activation of EP4. However, analysis of TCGA data indicates that *PTGER4* methylation appears to accompany gene silencing in many breast tumors (Fig. A.9). Interestingly, *PTGER4* is unmethylated and expressed in normal breast tissue. MCF7 cells treated with demethylating agents increase *PTGER4* mRNA levels, consistent with DNA methylation playing a direct role in its transcriptional regulation (Kim *et al.*, 2012). This supports a mechanism whereby an upstream epigenetic change in the EP4 promoter regulates *PTGER4* expression. Namely an increase in DNA methylation that silences *PTGER4* as the tumor forms

followed by a decrease in methylation that reactivates the gene to promote resistance to estrogen withdrawal in long-term estrogen deprived cells (Fig. 2.4).

Increased EP4 levels induce increased cAMP production, which drives ligand-independent activation of ER through the PKA-mediated activation of the co-activator CARM1. Prior work has also indicated that PKA can activate ERα by direct phosphorylation of serines 236 (S236) and 305 (S305). When S236 is phosphorylated, ERα dimerization and DNA binding is still dependent on ligand (Chen *et al.*, 1999) and thereby S236 phosphorylation is unlikely to activate ERα in MCF7-LTED cells. Studies in HeLa cells using a phosphomimetic S305E ERα suggested that S305 phosphorylation was sufficient to induce ligand-independent ERα binding to transcriptional targets in the absence of ERα dimerization (Tharakan *et al.*, 2008). However, other studies have instead suggested that S305 phosphorylated ERα is insufficient for ERα activation and that cofactors such as CARM1 (Carascossa *et al.*, 2010) in MCF7 cells and CREB (Lazennec *et al.*, 2001) in CHO were required for ERα binding to DNA and activation of response elements. While we cannot rule out the direct activation of ERα by PKA nor the activation of additional cofactors, our observations suggest that in estrogen deprived conditions EP4 promotes ligand independent activation of ERα through the PKA-mediated activation of the co-activator CARM1.

Another potential mechanism of ligand-independent ER activation is through genetic alterations of the *ESR1* locus. While any individual event is quite rare, as a whole these alterations are becoming an important theme in resistance. Metastatic tumors gain activating *ESR1* ligand binding domain mutations, especially after endocrine therapy (Li *et al.*, 2013; Merenbakh-Lamin *et al.*, 2013; Robinson *et al.*, 2013; Toy *et al.*, 2013; Zhang *et al.*, 1997).

Translocations of the *ESR1* locus can fuse with activating coding sequence or constitutive promoters that activate *ESR1* in the absence of estrogen (Li *et al.*, 2013). *ESR1* amplification has also been noted in both MCF7-LTED cells (Aguilar *et al.*, 2010) and a xenograft line from a patient tumor resistant to AIs (Li *et al.*, 2013). Rather than contradict our findings, however, this indicates that MCF7-LTED cells remain dependent on ERα signaling. Indeed, MCF7-LTED cells are sensitive to fulvestrant, an ER antagonist that accelerates proteasomal degradation of ER, indicating that these cells remain reliant on ER signaling (Sanchez *et al.*, 2011). GSEA analysis of RNA-seq data from MCF7-LTED cells further shows that MCF7-LTED cells show increased expression of ER target genes. Further, we did not observe a dramatic shift in DNA methylation changes at ER binding sites, which are a surrogate for altered ER binding.

Altogether, these data suggest that increased EP4 expression represents a viable means of developing resistance to aromatase inhibitor therapy. Expression increases in *PTGER4* were associated with decreases in DNA methylation, which is consistent with the idea that DNA methylation has a regulatory role in *PTGER4* expression. Our data indicate that EP4 activation is necessary for estrogen-independent growth. While we cannot completely rule out activation of alternative signaling pathways, our results support the conclusion that EP4 acts through PKA and the co-activator CARM1 to drive ligand-independent ERα activation. EP4 signaling presents a potential therapeutic target for the treatment of AI-resistant breast cancer.

While my work does not implicate *UCA1* in resistance to AI therapy, several other studies have identified ways that *UCA1* does contribute to breast cancer. *UCA1* decreases expression of the p27 tumor suppressor (Huang *et al.*, 2014). *UCA1* has also been found to sequester the miR-143, that acts as a tumor suppressor in breast cancer (Tuo *et al.*, 2015). By

absorbing miR-18a, *UCA1* allows expression of cell cycle regulators in a manner that promotes resistance to tamoxifen (Li *et al.*, 2016). *UCA1* was also shown to be necessary for proliferation, migration, and tumorigenicity of tamoxifen resistant derivatives of MCF7 and T47D cells. This effect was mediated by Wnt/β-catenin signaling (Liu *et al.*, 2016a). The Wnt pathway promotes stemness in breast cancer cells, which are resistant to treatment due to increased signaling through the PI3K pathway, among other reasons (Ma *et al.*, 2015). Because *UCA1* was not upregulated in MCF7-LTED cells, and it does not seem to strongly support the LTED phenotype in T47D, I suspect *UCA1*'s mechanism of support for tamoxifen resistance is not sufficient to overcome AI resistance.

2.4 Materials and Methods

Cell Culture

MCF7 and T47D cells were from the ATCC and maintained in RPMI 1640 (Gibco, Thermo Fisher Scientific) supplemented with 5% FBS (Gibco), 10 mM HEPES (Corning) , 4.5 g/L glucose (Corning), 2 mM L-glutamine (Gibco), 1 mM sodium pyruvate (Corning), and 50 µg/ml gentamicin (Gibco) in a humidified 37°C incubator containing 5% CO_2. T47D-LTED cells were previously derived from MCF7 and T47D cells, respectively (Sanchez *et al.*, 2011), and maintained in the same media except with 5% charcoal stripped FBS (Gibco) and RPMI 1640 without phenol red (Gibco).

Methyl-MAPS genome-wide methylation analysis

Methyl-MAPS analysis was performed as in Edwards *et al.* with custom barcoded adaptors (Edwards *et al.*, 2010). Libraries were made with AB-SOLiD and paired-end

sequenced. Sequencing reads were demultiplexed and analyzed using custom perl scripts (Edwards *et al.*, 2010). Sequencing statistics are in Table A.2 and Table A.3.

RNA-seq

RNA-seq libraries were prepared using NEB Next RNA-seq kit (NEB) with custom barcodes and sequenced with an Illumina HiSeq. Reads were demultiplexed using custom perl scripts. Reads were mapped to the human genome (hg18) using TopHat (v1.4) (Trapnell *et al.*, 2009). HT-seq was used to assign RefSeq annotations the reads (Anders *et al.*, 2015). Statistical analysis was performed using EdgeR (Robinson *et al.*, 2010). Ontology analysis was performed with Metacore.

2.4.1 *UCA1* Methods

TCGA Data Analysis

On April 9th, 2013, I downloaded a data archive of RNA-seq data for 856 tumor samples and 108 matched normal samples with the corresponding biotab clinical data from The Cancer Genome Atlas (https://cancergenome.nih.gov/). Custom perl scripts were used to enforce a floor of > 25 FPKM for all samples. Statistical tests were performed in R.

Overexpression Plasmid Construction and siRNAs

UCA1 overexpression constructs were created by cloning *UCA1* out of T47D mRNA and into the pBABE-puro expression vector. The overexpression plasmid or empty vector was stably integrated into T47D cells by lentiviral transduction. Silencer Select siRNAs were from Ambion (Thermo Fisher Scientific; Control siRNA, Cat. #4390843; U529, ID n272529; U530, ID n272530; U531, ID n272531; U851, ID s227851). siRNAs were transfected with Lipofectamine RNAiMax (Invitrogen, Thermo Fisher Scientific).

AlamarBlue Assay

AlamarBlue assays were performed five days after the start of the experiment. One-tenth volume of alamarBlue reagent (AbD Serotec BUF012, Bio-Rad) was added to each replicate. AlamarBlue fluorescence was measured after 1 hour.

RT-qPCR

RNA was extracted with the Zymo Research Quick mRNA Miniprep kit. RNA concentration was measured with the Qubit RNA BR kit. Reverse transcription was performed using the Bio-Rad iScript Reverse Transcriptase kit. *UCA1* mRNA expression was assessed was performed with the Bio-Rad iTaq Universal with SYBR Green reagent on an Applied Biosystems Viia7 instrument. The thermocycler protocol was the following: (1) 95°C, 20 s; (2) 95°C, 3 s; (3) 60°C, 20 s; for 40 cycles. Primer information is in Table A.4. Except where indicated, all qPCR measurements were performed in triplicate. *RPL0* was used as the endogenous control.

Caspase Assay

UCA1 knockdowns were prepared as for AlamarBlue assay. On day three, caspase activity was assayed using the Caspase Glo-3/7 Kit (Promega) according to the manufacturer protocol. Luminescence was read using a Molecular Devices SpectraMax M5e.

Cell Cycle Analysis

Harvest and pellet 1 x 10^6 cells in a 12 x 75 mm FACS tube. Resuspend in 1 mL ice cold 70% ethanol, and store at -20°C overnight or until needed. To continue, pellet fixed cells at 500xg for 10 min at room temperature. Aspirate 70% ethanol and wash twice with 1 mL1% BSA (Cell Signaling) in DPBS (Gibco). After washing, resuspend in 100 µL 1% BSA in DPBS. Add

500 μL of propidium iodide and RNAseA (24 μL propidium iodide, Invitrogen P3566; 12 uL RNaseA, Thermo EN0531; 464 uL DPBS). Final concentrations: propidium iodide = 40 μg/mL, RNaseA = 200 μg/mL. Incubate at 37°C for 30 min. Break up any clumps with a p1000 before pipetting through a 100 μm filter (smaller filters may be used if clumps remain and clog the FACS intake). Count cells on a BD FACScan machine.

2.4.2 *PTGER4* Methods

Immunofluorescence

Images were captured on a Zeiss AxioImager Z1 with AxioCam MRc and Axiovision software. The primary antibody was Cayman Chemical Company (Ann Arbor, MI) 101775. The secondary antibody was Jackson ImmunoResearch (West Grove, PA) 111-545-003.

RT-qPCR

PTGER4 mRNA expression was assessed using Applied Biosystems Viia7 with SYBR green. Primer information is in Table A.4. All qPCR measurements were performed in triplicate.

Methyl-Screen

Methyl-screen analysis was performed as in Holemon *et al.* (2007) with slight modification. Genomic DNA was mock digested, AciI (NEB) digested, McrBC (NEB) digested, or digested with both AciI and McrBC. Real-time quantitative PCR was performed with *PTGER4*-specific primers: 5'-GCAGCTTTGTCTCTCTTC-3' and 5'-TACCGAGACCCATGTTG-3'. Unmethylated control gDNA was produced by whole genome amplification of MCF7 gDNA with the REPLI-g kit (Qiagen). Methylated control DNA was produced by treating amplified gDNA with M.SssI (Zymo Research).

Cell proliferation

Silencer Select siRNAs were from Ambion (Thermo Fisher Scientific; siCtrl, 4390846; siEP4, 4427037; siEP4-1, ID s60396; siEP4-2, ID s11456). siRNAs were transfected with Lipofectamine RNAiMax (Invitrogen, Thermo Fisher Scientific). When specified, EP4 antagonist GW627368X was used at 3.3 μM and ONO-AE3-208 was used at 10 μM. AlamarBlue assays were performed at 7 days for MCF7-LTED and 4 days for MCF7 cells. One-tenth volume of alamarBlue reagent (AbD Serotec BUF012, Bio-Rad) was added to each replicate. AlmarBlue fluorescence was measured after 1 hour. Eight replicates were performed for MCF7-LTED cells and 4 replicates for MCF7 cells. p-values were computed using a one-way ANOVA followed post-hoc by Tukey's HSD (honest significant difference) test. Bartlett's test of the homogeneity of variances was insignificant ($\alpha = 0.05$) for all comparisons under the null-hypothesis of unequal variances.

ChIP-qPCR

Cells were grown in 10 cm dishes to 85% confluence. The cells were starved, by exchanging growth media for reduced-serum media containing 0.5% charcoal stripped FBS, for 18 hours before activation of EP4 with 10nM L902,688 EP4 agonist (Cayman Chemical) for 2 hours. An ethanol-treated sample served as a vehicle control. Afterward, cross-linking was initiated by drop-wise addition of 37% formaldehyde to a final concentration of 1% at room temperature. Cross-linking ran for 10 minutes and was quenched for 5 minutes at room temperature by adding 0.5ml of 2.5M glycine. Quenched cells were washed twice with 10ml ice-cold PBS, removed from the plate by scraping in ice-cold DPBS (Gibco) with protease inhibitors, and pelleted. Cell pellets were resuspended in 350 ul ChIP lysis buffer (1% SDS, 10

mM EDTA, 50 mM Tris-HCl pH 8.0) with protease inhibitors. Lysed cells were sonicated and 100 ug of protein were brought to a final volume of 50ul lysis buffer with protease inhibitors. 9 ul of CARM1 antibody (CST, (3H2) Mouse mAb #12495) or 2ul ER antibody (Santa Cruz sc-7207X) was added and incubated overnight at 4°C.

Antibody bound proteins were purified by adding samples to 25µl Protein A/G magnetic beads (Pierce) that were prewashed twice with ChIP Dilution Buffer. The samples were incubated with the beads 2 hours at 4 °C. Afterward the beads were washed with 0.5 ml of each of the following ice-cold buffers: 1) low salt immune complex wash buffer (EMD-Millipore 20-154), one wash; 2) high salt immune complex wash buffer (EMD-Millipore 20-155), one wash; 3) LiCl immune complex wash buffer (EMD-Millipore 20-156), one wash; 4) 1X TE-T, two washes (10mM TrisHCl, 1mM EDTA, 0.1% Triton, pH 8). Elution buffer (1% SDS + 0.1M NaHCO3) was prepared fresh from 10X stocks. 200 µl room temperature elution buffer was added, and the samples were incubated at room temperature for 30 min to elute the protein. Afterward, 1 µl of 10 mg/mL RNase A was added and the samples were incubated overnight at 65°C. The Qiagen PCR Purification kit was used to purify the DNA in a final volume of 60µl Buffer EB.

Real-time quantitative PCR was performed on an Applied Biosystems Viia7 using SYBR green. Primers are in Table A.4. Values are reported as percent relative to input. All experiments were performed in triplicate and p-values are computed using Student's t-test.

cAMP measurements

cAMP measurements were made using the Cyclic AMP XP Assay Kit (Cell Signaling Technologies) according to the manufacturer's instructions. MCF7-LTED cells were washed

twice with warm PBS and then pre-treated with 0.5 mM IBMX in serum-free medium for 30 min prior to addition of L-902,688, at the indicated concentrations, or 1 μM forskolin. For combined antagonist/agonist treatments, cells were pre-treated 10 min with ONO-AE3-208 before addition of L-902,688. For EP4 knock down experiments, cAMP measurements were made 2 days post transfection. cAMP measurements were quantitated as follows: %Activity = 100 x [(A – A_{basal})/(A_{max} – A_{basal})], where A is the absorbance of the agonist treated sample, A_{max} is the absorbance of the forskolin treated sample, and A_{basal} is the absorbance of the vehicle treated sample.

CARM1-IP – ER blot

MCF7-LTED cells were grown to confluence in 10 cm dishes. Before harvesting, cells were grown in media with low serum concentration (0.5%). The next day, the cells were incubated with 10 mL of RPMI (no serum) containing 0.5 mM IBMX for 30 minutes. The cells were then treated with 10 nM L-902,688 in 10 mL of RPMI + 0.5 mM IBMX and harvested after 0, 5, 15, or 30 minutes. After washing with ice-cold PBS, cells were lysed with 1 mL of RIPA buffer (1 M tris pH 8.0, 5 M NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, and 2 mM EDTA with freshly added PMSF to 1 mM final and 1X Halt protease and phosphatase inhibitor cocktail [ThermoFisher Scientific]). Cells were scraped loose, the cell lysate was transferred to a microcentrifuge tube, and the tube was put on a rotator for 30 minutes at 4°C. The supernatant was isolated by centrifugation at 12,000xg for 20 minutes at 4°C.

Immunoprecipitation was performed by addition of mouse anti-CARM1 (Cell Signaling Technology #12495) and precipitation overnight at 4°C. Precipitated proteins were collected with the Dynabeads Protein G magnetic beads. 25 uL of beads were washed with 500 uL of

RIPA buffer before adding the immunoprecipitated proteins. This was then incubated at 4°C for 3 hours. The protein-bound beads were then washed three times with 0.5 mL RIPA with 10 mM NaF and 1 mM sodium pyrophosphate and resuspended in 0.5 mL of wash buffer. The protein was eluted in 40 uL of 2x NuPAGE LDS sample buffer at 90°C for 10 minutes. Final analysis of immunoprecipitated proteins occurred by Western blot on 4-12% gradient PAGE gels (Life Tech.). ERα was identified using a rabbit polyclonal antibody (Santa Cruz Biotech sc-7207) to minimize signal from the mouse anti-CARM1 IgG used for immunoprecipitation. The blot was stripped using Restore™ Western Blot Stripping Buffer (ThermoFisher), and reblotted with rabbit anti-CARM1 antibody (Cell Signaling Technologies #3379). Alexa Fluor 680 or Alexa Fluor 790 labeled secondary antibodies were from Jackson ImmunoResearch. Blots were imaged using the Odyssey Fc (LI-COR Biosciences).

The Cancer Genome Atlas data analysis

Infinium 450k methylation, RNA-seq expression and clinical data were downloaded from the The Cancer Genome Atlas (TCGA) data portal. P-values were computed using Mann-Whitney U.

Expression Analysis of ER+ breast cancer in neoadjuvant setting

Expression analysis was performed for patients enrolled in the POL (Olson *et al.*, 2009) and ACOSOG (American College of Surgeons Oncology Group) Z1031 (Ellis *et al.*, 2011) trials. Postmenopausal women with ER+ Stage II/III breast cancer were treated with letrozole, anastrozole, or exemestane for 16-18 weeks. Biopsies were performed before and after treatment. Patients were labeled as resistant if 10% or more of malignant cells stained positive for the proliferative marker Ki67 at 16 weeks. Expression analysis was performed using Agilent

Human Gene Expression 4x44K v2 Microarray. P-values for *PTGER4* expression were computed using Mann-Whitney U.

2.5 Supplementary Material

Supplementary material is in Appendix A.

Chapter 3: Reprogrammable CRISPR/Cas9-based system for inducing site specific DNA methylation

The work described in this chapter represents my contribution to a collaborative project between the Edwards and Challen labs. Our combined efforts have been published (McDonald *et al.*, 2016). I have included here only parts of the project that I had a direct hand in such as the text and the data produced by my laboratory. For the data produced by the Challen lab, please see the complete publication.

3.1 Introduction

DNA methylation of CpG dinucleotides is a prominent epigenetic modification of the mammalian genome that can influence gene expression, and aberrant distribution of DNA methylation is associated with a spectrum of human disorders including cancers (Egger *et al.*, 2004). Despite intensive study, it remains unclear which CpG dinucleotides must change methylation state in order to alter transcription. Genome-wide analyses have found associations between DNA methylation and reduced gene expression that occur both in the proximal promoter and downstream of the gene's transcription start site (TSS, (Bell *et al.*, 2011; Bock *et al.*, 2012; Lou *et al.*, 2014; Lund *et al.*, 2014; VanderKraats *et al.*, 2013). However, evidence supports both that DNA methylation can cause a loss of expression, and that expression changes can alter DNA methylation patterns (Bestor *et al.*, 2015; Busslinger *et al.*, 1983). Here, we sought to develop tools for locus-specific epigenetic remodeling to directly address the role of DNA methylation in regulating gene expression.

Targeted DNA methylation approaches have been attempted by fusing DNA methyltransferase enzymes (DNMTs) to DNA binding proteins such as zinc finger proteins (ZFPs) (Siddique *et al.*, 2013), and transcriptional activator-like effector (TALE) (Bernstein *et al.*, 2015). However, engineering custom proteins for each targeted sequence is laborious and requires specialized expertise. Moreover, in these studies, induced DNA methylation of the targeted loci was relatively poor, with substantial off-target activity. An engineered form of the clustered, regularly interspaced, short palindromic repeat (CRISPR) system has emerged as an alternative for achieving site-specific DNA targeting (Jinek *et al.*, 2012). Here, the Cas9 endonuclease is directed to genomic targets by engineered short guide RNAs (sgRNAs) (Jinek *et al.*, 2012). Because the sgRNA is the DNA sequence-specific component of the system, it allows for efficient targeting of multiple regions due to the ease of design and synthesis of new sgRNAs (relative to engineering new custom proteins for each target site). A Cas9 mutant (D10A and H840A; henceforth referred to as dCas9) that lacks endonuclease activity but can still be recruited by sgRNA(s) (Jinek *et al.*, 2012) has recently been used to target genes in mammalian cells for transcriptional activation (Maeder *et al.*, 2013b; Mali *et al.*, 2013; Perez-Pinera *et al.*, 2013a, 2013b). Here, we demonstrate an easily reprogrammable CRISPR/dCas9 DNMT fusion capable of inducing site-specific DNA methylation.

3.2 Results

To design a flexible system to target DNA methylation, we fused dCas9 to the catalytic domain of the *de novo* DNA methyltransferase DNMT3A (Fig. 3.1A). To test this system, we targeted DNA methylation to the tumor suppressor gene *CDKN2A* (cyclin dependent kinase 2A),

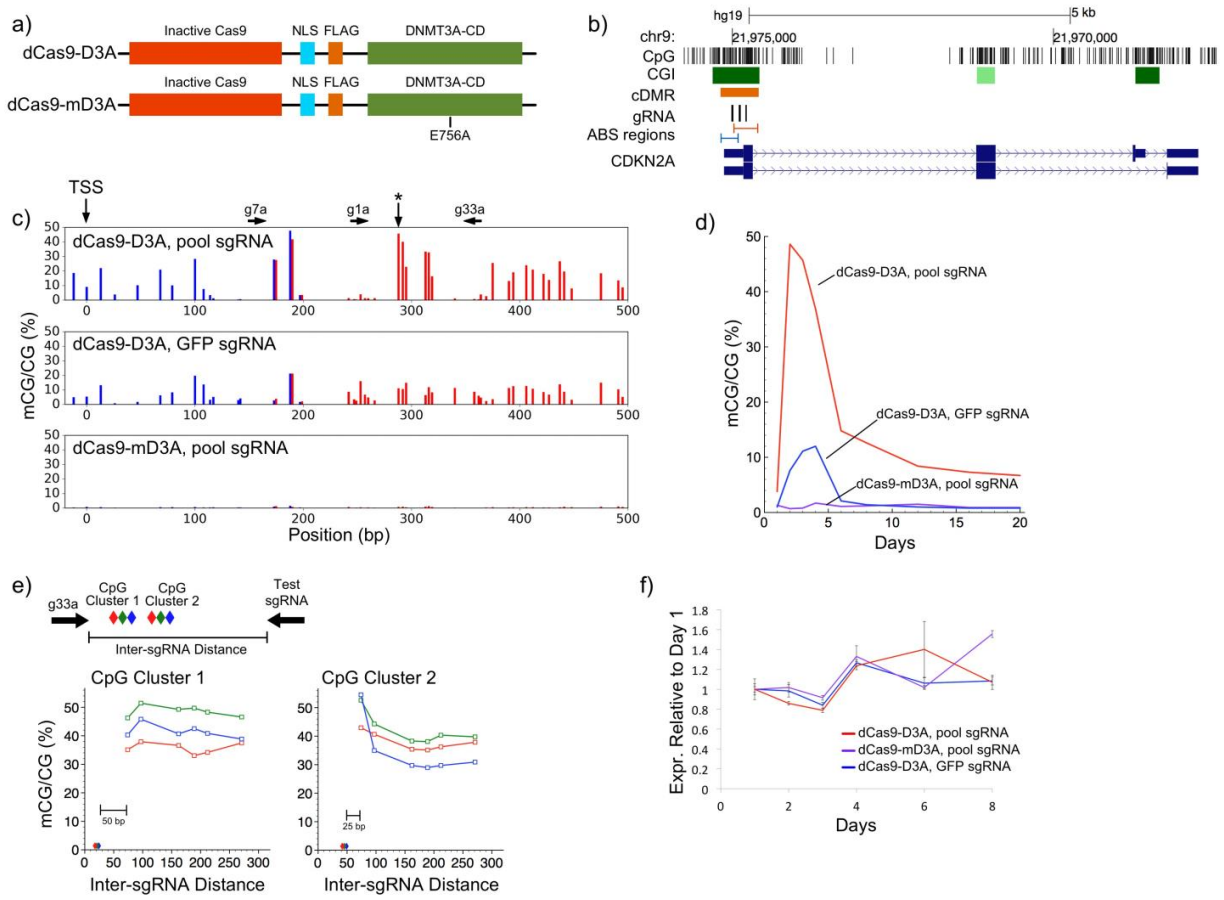


Figure 3.1: Site-specific induction of DNA methylation using a CRISPR-Cas9 DNMT fusion. (A) dCas9-DNMT3A CD fusion constructs. The E756A mutation inactivates the DNMT3A-CD. **(B)** UCSC genome browser view showing the locations of the three *CDKN2A* sgRNA (g1a, g7a and g33a). The three sgRNA were validated to ensure they targeted this locus (Figure 1-figure supplement 2). cDMR indicates the region from the literature where methylation changes are associated with expression changes (Fig. B.1). **(C)** Induced DNA methylation at the *CDKN2A* promoter three days post-transfection. Colors correspond to the red and blue ABS regions in (B). Three CpGs were independently measured in both amplicons. sgRNA target sites are indicated above the graphs. Pool sgRNA indicates g1a, g7a, g33a were used simultaneously. Sanger sequencing validation is presented in Fig. B.3, and non-CpG methylation data is presented in FigB.4. **(D)** Time course of the percent methylation data for the CpG marked with an asterisk in (C). Additional CpGs are shown in FigB.5. **(E)** Methylation induced by a pair of sgRNA decreases with increasing intervening distance. Distance is calculated relative to the 3' end of the g33a sgRNA. Diamonds indicate the location each CpG monitored for methylation; whose color corresponds to appropriate line in the graph. Additional data from individual and paired sgRNA is presented in Fig. B.6. **(F)** *CDKN2A* expression for samples with induced methylation. Expression is normalized to day one for each respective sample. Error bars represent standard error of the mean (n=3). NLS, nuclear localization signal; FLAG, FLAG tag domain.

which inhibits progression through the cell cycle (Liggett and Sidransky, 1998). *CDKN2A* is one of the most frequently hypermethylated genes in The Cancer Genome Atlas (Ciriello *et al.*, 2013), and numerous clinical studies show a negative correlation between *CDKN2A* methylation and expression in colorectal cancer (Shima *et al.*, 2011). While it is generally assumed that *CDKN2A* methylation induces gene silencing, it has also been suggested that DNA methylation occurs after the loss of expression (Hinshelwood *et al.*, 2009). From a literature search, we identified 17 publications that associate *CDKN2A* methylation with expression and/or cancer (Fig. B.1). Overwhelmingly, these papers studied the differentially methylated region (cancer DMR, cDMR) on the 3′ end of the CpG island that overlapped the first exon of *CDKN2A* (Fig. 3.1B).

We computationally designed three sgRNAs (g1a, g7a and g33a) to target this region and test whether DNA methylation was sufficient to induce gene silencing. We validated the ability of each sgRNA to target the *CDKN2A* locus by transfecting them with active CRISPR/Cas9 and measuring the ability of Cas9 to cleave the locus (Fig. B.2, Table B.1). We then transfected HEK293T cells with the pool of three sgRNAs along with either a normal dCas9-DNMT3A catalytic domain (CD) fusion (dCas9-D3A) or one with a DNMT3AE756A mutation (dCas9-mD3A, Fig. 3.1A), which abolishes DNA methyltransferase activity (Reither *et al.*, 2003). Transfection efficiencies were >80-90% for all experiments as measured by cotransfection with GFP-containing plasmids. We analyzed DNA methylation levels for 20 days post-transfection using Illumina sequencing of two amplicon regions (amplicon bisulfite sequencing, ABS). ABS results were validated using Sanger bisulfite sequencing (R2=0.83; Fig. B.3), and DNA methylation levels at CpGs analyzed in both of two independent ABS amplicons showed

strong correspondence (R2=0.98; Fig. 3.1C). All CpGs had >100× sequencing coverage, with a median coverage of 15,200.

Over the 20-day time course we observed an increase in DNA methylation at the *CDKN2A* target locus that ranged from 20-43% at its peak on day three. Background methylation from transfection with dCas9-mD3a was consistently less than 1.5%, while background methylation from an off-target sgRNA, which controls for DNMT3A-CD overexpression, was less than 14% on day three. Induced DNA methylation levels were highest over a set of eight CpGs directly between the g33a and g7a sgRNA target sites (Fig. 3.1C). Increases in CHG and CHH methylation were minimal (Fig. B.4). DNA methylation decreased rapidly after passaging the cells on day four, but stabilized 20 days post-transfection at 6-10% (Fig. 3.1D; Fig. B.5). Despite the literature support for a negative correlation between expression and DNA methylation in this region, we did not observe a measurable effect on *CDKN2A* gene expression by RT-qPCR (Fig. 3.1F). This suggests that a limited increase in methylation in the region 100-400 bp downstream of the *CDKN2A* TSS is insufficient to trigger gene silencing.

Spatially, the induced DNA methylation spiked near the sgRNA target sites and dropped quickly toward background levels at surrounding CpG sites. Analysis of DNA methylation induced by single sgRNAs indicates that methylation occurs primarily within 50 bp of the sgRNA binding site (Fig. B.6). Higher DNA methylation levels were often observed 3′ of the sgRNA binding site (Fig. B.6). Our initial data from the three pooled sgRNAs suggested that CpG methylation was higher between pairs of sgRNAs. To investigate this effect, we transfected pairs of sgRNA with varying intervening distances and monitored methylation of six clustered CpGs between the sgRNA pairs using ABS (Fig. 3.1E). DNA methylation of the three CpGs

(cluster 1) within 20 bp of the fixed sgRNA (g33a) did not change with addition of a second sgRNA 77 bp or further away. However, the methylation level increased from 30-40% to 42-53% when sgRNAs were paired within 80 bp and both sgRNAs were within 50 bp of CpG cluster 2 (Fig. 3.1E). This suggests the DNMT3A-CD activity at the target locus is additive.

We next tested whether we could use our approach to methylate an entire CpG-island (CGI). We designed 17 sgRNAs (Fig. 3.2A; Table B.1) to target DNA methylation across the *CDKN2A* CGI, which spans the TSS. We applied three combinations of sgRNAs (Set 1, 2, All) to test whether inducing DNA methylation of the entire CGI could decrease gene expression (Fig. 3.2A). ABS analysis of eight amplicons (minimum per CpG sequencing depth of 100) showed that the DNA methylation level increased to an average of 22% across the entire region with a peak of 54% (Fig. 3.2B). As an off-target negative control, we used three sgRNAs targeted to the ARF promoter located ~20 kb away. The average background methylation at *CDKN2A* after treatment with off-target ARF sgRNAs was 9% (Fig. 3.2B). The other two sgRNA sets (Set 2 and All) induced similar increases of methylation across the CGI overlapping the TSS of *CDKN2A* (Fig. B.7).

Analysis of *CDKN2A* expression by RT-qPCR in all three sgRNA targeting experiments (Set 1, 2, All) indicated an average 39% decrease in *CDKN2A* mRNA expression after targeting with dCas9-D3A (Fig. 3.2C). Cells transfected with dCas9-mD3A showed a 16-26% reduction in *CDKN2A* expression, likely due to CRISPR inhibition (Fig. 3.2C). Across the three replicates (Set 1, 2, All) expression decreased by an average of 17% in dCas9-D3A relative to dCas9-mD3A ($P<0.01$ paired-t-test). This indicated that DNA methylation directly decreased *CDKN2A* expression, but targeting of the entire CGI was required to trigger this effect. Our results are

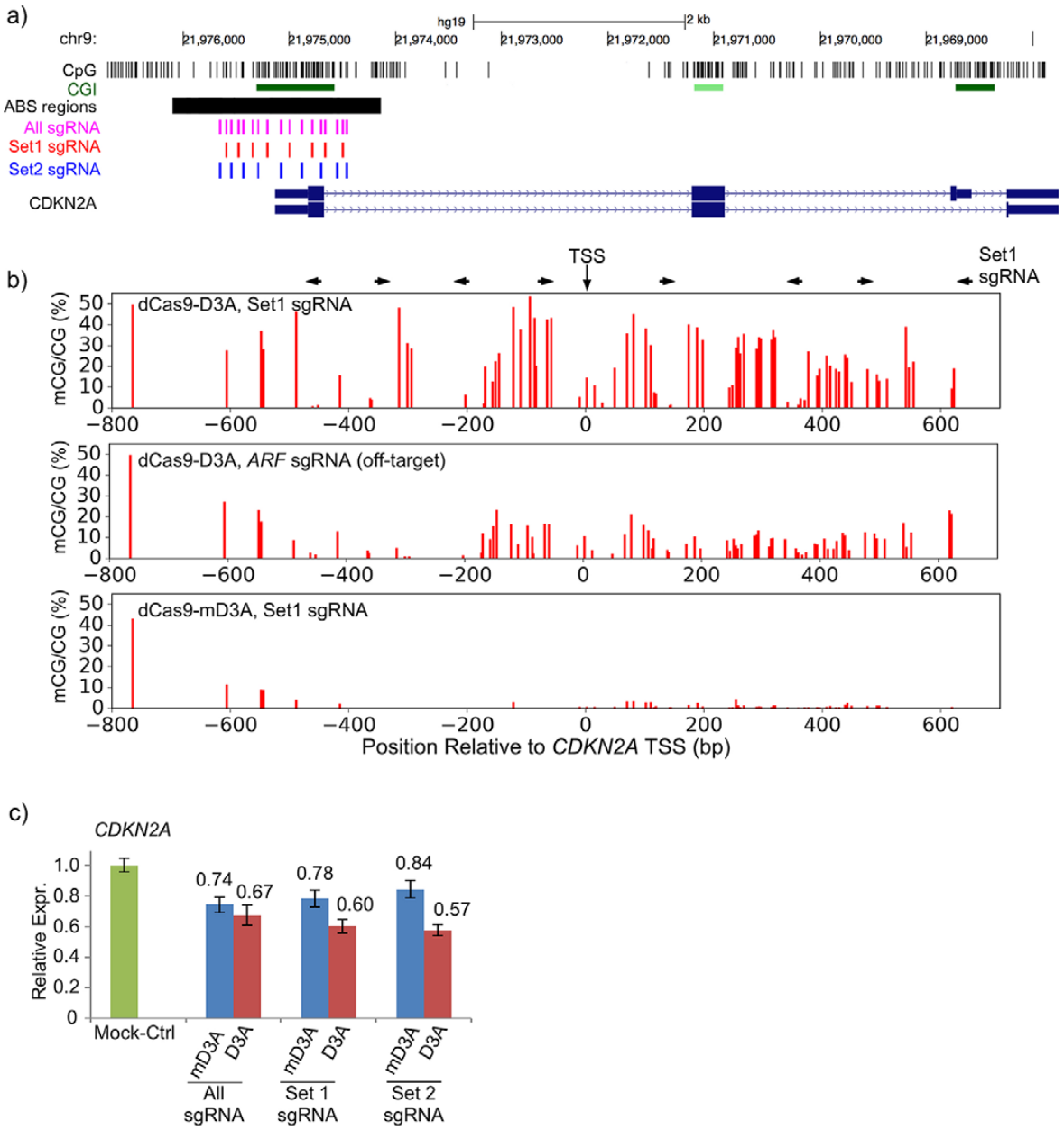


Figure 3.2: *CDKN2A* promoter methylation decreases expression in a context dependent manner. (A) Locations of the 17 sgRNAs used in Sets 1 and 2 and regions sequenced in B. sgRNA coordinates are in Table B.1. **(B)** Percent methylation is plotted for regions designated in A. Set 1 sgRNA target sites are indicated above the graphs. Data for All sgRNA, Set 2 sgRNA, and off-target SgRNA is presented in Fig. S2. **(C)** Methylation induced by *CDKN2A*-targeted sgRNA Set 1, Set 2, and All sgRNA, decreases gene expression. Relative expression of *CDKN2A* is normalized to a mock-treated control. Error bars=mean±s.e.m. (All, n=2; Set 1, n=1; Set 2, n=1; all performed in technical triplicate, paired t-test).

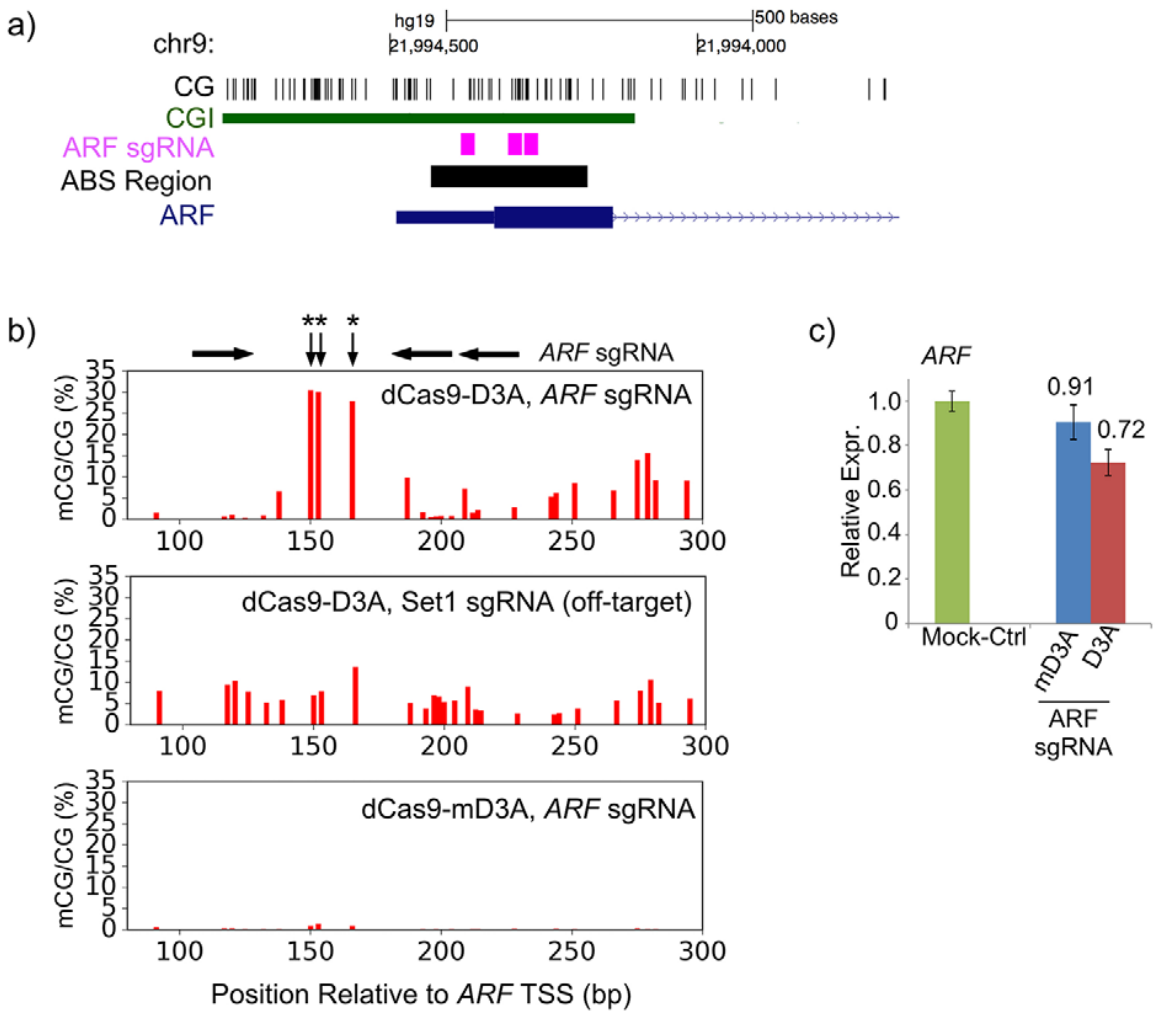


Figure 3.3: Induced methylation decreases *ARF* expression. **(A)** Locations of the three sgRNA used to induce DNA methylation at the ARF promoter. **(B)** Percent methylation is plotted for a region of exon 1 of the ARF locus. ARF sgRNA target sites and three targeted CpGs (asterisks) are indicated above. **(C)** Methylation induced by ARF targeted sgRNA decreases gene expression. Relative expression of ARF is normalized to a mock-treated control. Error bars = mean ± SEM (n=1 biological replicate performed in technical triplicate).

consistent with other studies that find a similar reduction in gene expression after inducing methylation at the *CDKN2A* promoter using ZFP- and TALE-based systems (Bernstein *et al.*, 2015; Cui *et al.*, 2015).

To verify the effect of our system with a separate locus, we designed three inwardly-directed (5′ to 3′) sgRNAs that bracketed three CpG sites in the *ARF* promoter located 150-170

bp downstream of the TSS (Fig. 3.3A). *ARF*-targeted sgRNAs increased the DNA methylation level to 27-30% at these three CpG sites with less than 15% methylation induced in adjacent sites (Fig. 3.3B). Induced methylation of the *ARF* promoter was associated with a 19% decrease in its expression (Fig. 3.3C).

3.3 Discussion

We provide an outline for using a modified CRISPR/dCas9 system to evaluate the functional relevance of DNA methylation at specific CpGs and described guidelines for its use. DNA methylation induction occurs within ~50 bp of a sgRNA target site and is strongest between two adjacent and inwardly directed sgRNA binding sites. Based on our design criteria, we designed sets of sgRNAs that induced methylation at the human *CDKN2A* and *ARF* promoters, and the mouse *Cdkn1a* promoters with similar efficiency. Induced methylation was sufficient to decrease expression of all three genes. Methylation increases and changes in expression were highly significant and reproducible either by using multiple distinct sgRNA combinations in the case of *CDKN2A* or at the clonal levels as observed for Cdkn1a. Moreover, the reduction in *Cdkn1a* expression clearly had functional consequences (increased proliferation) for the transduced cells. Though modest, the expression decreases caused by induced methylation are consistent with previously published results using ZFP and TALE fusions (Bernstein *et al.*, 2015; Cui *et al.*, 2015).

The effects of the induced methylation also appeared to be context dependent. While methylation of the entire CGI at the *CDKN2A* promoter repressed gene expression, inducing DNA methylation of a region 100-400 bp downstream of the *CDKN2A* TSS alone was insufficient to affect expression despite the frequent observation of a negative correlation

between methylation and expression in this region. This indicates the importance of the flexibility to target multiple regions offered by our CRISPR/dCas9 DNMT fusion system.

Although approaches for targeted DNA methylation have been previously described, our method is advantageous for several reasons including; (1) the ease of designing new sgRNAs for targeting, (2) higher levels of induced DNA methylation, (3) little off-target activity. This approach can be used to interrogate the effects of DNA methylation on only a few CpG sites by bracketing them with sgRNAs, or can be used to test the effects of broader increases in DNA methylation by using many sgRNAs simultaneously. Further, the method described here provides a robust, reprogrammable approach to allow researchers to easily and thoroughly explore the functional roles of DNA methylation changes in development and disease.

3.4 Materials and Methods

dCas9 fusion protein design and construction

The catalytic domain (CD) of human DNMT3A (amino acids 602 to 912 of NP_783328.1) both with and without the E756A mutation was cloned between the NheI and AgeI sites of pCMV_dCas9_VP64 (Addgene plasmid #49015, Cambridge, MA) with a NLS and FLAG tag linker. Plasmids sequences were validated by Sanger sequencing and prepared for transfection using a Qiagen Maxiprep kit. All plasmids are available in Addgene (#78256, 78257), and detailed information is available at http://epigenomics.wustl.edu/epigenomeEditing.

sgRNA design

Target sequences were entered into the MIT sgRNA design software (http://crispr.mit.edu/), the BROAD sgRNA design tool (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design-v1) (Doench *et al.*, 2014),

and the sgRNAcas9 tool (version 2.0.10) (Xie *et al.*, 2014). The intersection of sgRNA target sites produced by all tools was taken for further analysis. sgRNA sequences that failed the BROAD test (score<0.2) were excluded. sgRNA were selected based on high BROAD scores and location relative to other sgRNAs. sgRNA coordinates and sequences are in Tables B.1 and Table B.2. Oligonucleotides corresponding to the target sites were annealed and cloned into MLM3636 (Addgene plasmid #43860).

Cell culture

HEK293T cells were acquired from ATCC (CRL-3216) and grown in DMEM supplemented with 10% FBS (Gibco), 1× Penicillin/Streptomycin (Gibco), and 2 mM GlutaMax (Gibco). For transfection experiments, 3×10^5 HEK293T cells were plated in a 60 mm dish. The next day, the cells were transfected with Lipofectamine LTX (Thermo Fisher Scientific). The Lipofectamine:DNA ratio was 3.5, with a total of 5.5 μg of plasmid DNA. The mass of Cas9-DNMT3A CD fusion plasmid was equal to the total mass of the sgRNA plasmids. Since HEK293T cells incorporate either all plasmids or none, 0.5 or 0.7 μg of pMaxGFP was co-transfected in order to indicate the transfection efficiency. The plasmid DNA was first diluted in Gibco OptiMEM, then Lipofectamine LTX was added and mixed in by inversion. After 30 min, the transfection mixture was added dropwise to the cells and they were placed back in the incubator.

Sanger bisulfite sequencing

Genomic DNA (gDNA) was isolated using the Zymo Research Quick gDNA MiniPrep kit and quantified with the Qubit dsDNA broad range assay (Thermo Fisher Scientific). gDNA was bisulfite converted with the Zymo Research EZ DNA methylation kit according to the

manufacturer's instructions. All samples underwent bisulfite conversion with a high efficiency of at least 98% as determined by conversion of unmethylated, non-CpG cytosines. For *CDKN2A*, the target regions were amplified with the Qiagen PyroMark PCR kit and CDKN2A_A primers in Table B.3. PCR products were cloned into the Promega pGEM-T Easy plasmid and transformed into NEB 10β competent cells. PCR products from individual colonies were sequenced by Sanger. Sanger bisulfite sequencing analysis was performed using BIQ Analyzer (Bock *et al.*, 2005).

Amplicon bisulfite sequencing

gDNA was extracted, bisulfite converted and PCR amplified as above. A tenfold molar excess of Illumina sequencing Y-adapters was then annealed to 100 ng of PyroMark PCR product (total Y-adapter mass varies with PCR product length) with NEB Quick Ligase for 15 min. The ligations were purified over a 1.5% agarose II (ISC BioExpress) gel in 1× TBE in order to remove incompletely ligated DNA. Custom 13 bp barcode index sequences were added via PCR using NEB Phusion. The standard Phusion PCR protocol was followed but with the following primer concentrations (for a 50 µl reaction): 1 µl of Illumina Primer 1.0 (25 µM), 1 µl of Illumina PCR Primer 2.0 (diluted fresh to a final concentration of 0.5 µM), and 1 µl of the index primer (25 µM). The thermocycler protocol was the following: (1) 98°C, 30 s; (2) 98°C, 10 s; (3) 64°C, 30 s; (4) 72°C, 30 s; (5) Return to step two 11 times; (6) 72°C 5 min; (7) 4°C hold. The PCR products were again purified over a 1.5% agarose II gel in 1× TBE, and their concentration was measured with the Qubit dsDNA high sensitivity kit. Individually indexed samples were pooled and submitted for sequencing. Amplicon bisulfite sequencing data were checked for quality using fastQC, adaptor and poor quality sequence (quality less than 20) was

trimmed using fqtrim, and the trimmed sequences were mapped to the target sequences using Bismark (Krueger and Andrews, 2011).

Expression analysis

RNA was extracted with the Zymo Research Quick mRNA Miniprep kit. RNA concentration was measured with the Qubit RNA BR kit. RNA integrity was determined by visualizing rRNA bands using agarose gel electrophoresis. Reverse transcription was performed using the Bio-Rad iScript Reverse Transcriptase kit. Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was performed with the Bio-Rad iTaq Universal with SYBR Green reagent on an Applied Biosystems Viia7 instrument. The thermocycler protocol was the following: (1) 95°C, 20 s; (2) 95°C, 3 s; (3) 60°C, 20 s; for 40 cycles. qPCR primers are listed in Table B.4. A melt curve was performed to indicate there was not off-target amplification. Data was analyzed as described by Hellemans *et al.* (2007) using the geometric mean of *ACTB, GAPDH,* and *RPL0* as an internal control. The All sgRNA sample represents data from two independent transfection experiments. The data for all remaining samples derives from technical replication using the same RNA sample. P-values were calculated with paired sample t-tests on the normalized levels of gene expression.

Western blot

Cells were lysed in complete RIPA buffer containing protease inhibitors (Santa Cruz Biotechnology). 20 μg of protein lysates were separated on 10% SDS-PAGE gels and transferred to nitrocellulose membranes (Millipore). Membranes were subsequently probed to detect fusion proteins using primary antibodies recognizing Cas9 (Active Motif) or β-actin (Santa Cruz) and

detection was performed using horseradish-peroxidase-conjugated secondary mouse antibody (Santa Cruz) and chemiluminescence (Millipore).

3.5 Supplementary Material

Supplementary material is in Appendix B.

Chapter 4: Optimizing the dCas9-DNMT system with Alternative DNMT Catalytic Domains

4.1 Introduction

DNA methylation of CpG dinucleotides is a prominent epigenetic modification of the mammalian genome that influences both chromatin state and gene expression. Alteration of DNA methylation patterns has been observed in numerous human diseases and disorders, including cancer (Schübeler, 2015). Despite intensive study, however, it remains difficult to predict the effect of a specific methylation change on gene expression. There is a global negative correlation between the presence of DNA methylation at a gene promoter and decreased gene expression, but the correlation is a weak one (Bock, 2012). Analysis of genomic data sets only finds a strong relationship between methylation and expression at a small subset of genes (Bell *et al.*, 2011; Bock *et al.*, 2012; Lou *et al.*, 2014; Schlosberg *et al.*, 2017; VanderKraats *et al.*, 2013). Additionally, while there is evidence that DNA methylation can cause decreased expression (Busslinger *et al.*, 1983), there is also evidence that the reverse is true (Bestor, 2000).

The difficulty in defining the relationship between methylation and expression partly derives from the lack of tools capable precisely manipulating DNA methylation. The most commonly used approaches include DNA methyltransferase (DNMT) knockout, knockdown, or chemical inhibition. Unfortunately, all of these approaches alter methylation genome-wide and thereby complicate the analysis of specific methylation changes. Targeted DNA methylation approaches are therefore preferable and suggested as early as 1997 (Xu and Bestor, 1997). However, the targeting technologies available at the time, zinc fingers (ZF), were time-

consuming and expensive to create. As a result, relatively few investigators have chosen this approach. I attempted to solve this problem by adapting the flexible and inexpensive CRISPR/Cas9 system to make targeted methylation changes, and I was one of the first to publish a working system (McDonald *et al.*, 2016; Vojta *et al.*, 2016). Several other labs have published similar systems. Stepper *et al.* (2016) showed that the Dnmt3a-Dnmt3L fusion outperforms Dnmt3a, as previously observed with a zinc finger targeting domain (Siddique *et al.*, 2013). dCas9-DNMT fusions have also been used to induce long-term gene silencing (Amabile *et al.*, 2016) as well as to methylate CTCF binding sites and alter chromosomal looping (Liu *et al.*, 2016b). Of these studies only Amabile *et al.* (2016) achieved an average of >80% methylation at an endogenous gene target, and they did so by simultaneously altering chromatin with the KRAB repressor domain. This is the first report to achieve average levels of methylation that imitate the >70% methylation level observed at silenced CpG island (CGI) promoters *in vivo* (Weber *et al.*, 2007). This is an important achievement, but it highlights the fact that the dCas9-DNMT system alone appears incapable of imitating *in vivo* methylation levels at silenced genes.

I therefore sought to induce higher levels of methylation by using alternative DNMT catalytic domains. Targeted methylation studies using mammalian DNMTs have almost exclusively relied on DNMT3A-based fusions (Amabile *et al.*, 2016; Bernstein *et al.*, 2015; Cui *et al.*, 2015; Groote *et al.*, 2012; Liu *et al.*, 2016b; McDonald *et al.*, 2016; Rivenbark *et al.*, 2012; Siddique *et al.*, 2013; Stepper *et al.*, 2016; Vojta *et al.*, 2016). I hypothesized that other DNMTs or combinations of DNMTs might induce more methylation. The combination of DNMT3A and DNMT3L is known to enhance DNMT3A activity (Siddique *et al.*, 2013; Stepper *et al.*, 2016). Like DNMT3A, DNMT3B is a *de novo* DNMT and might also induce high levels of

methylation. In contrast to DNMT3A/B, DNMT1 is considered the maintenance methyltransferase that targets hemimethylated sites during DNA replication (Jurkowska and Jeltsch, 2016). However, Dnmt1 can methylate DNA better than both Dnmt3a and Dnmt3b at unmethylated sites *in vitro* despite its overall preference for hemimethylated sites (Okano *et al.*, 1998). In addition, the *de novo* and maintenance methyltransferase activity might synergize and increase the induced methylation levels (Jeltsch and Jurkowska, 2014; Jones and Liang, 2009). Lastly, the bacterial DNMT, M.SssI, can methylate DNA 100% *in vitro* and induces strong methylation of open chromatin (Kelly *et al.*, 2010). Though bacterial DNMTs have been used for targeted methylation before (Groote *et al.*, 2012), they have not been used to target an endogenous promoter.

To test these alternative DNMTs, I cloned the following DNMTs: human DNMT1, a mouse Dnmt3a-mDnmt3L fusion (Siddique *et al.*, 2013), human DNMT3B, and M.SssI. I then performed transient transfection of HEK293T cells with sgRNA to the *CDKN2A* and *ARF* promoters as previously published (McDonald *et al.*, 2016). The mDnmt3a-mDnmt3L fusion increased the methylation levels by a difference of 10% to 20% compared to human DNMT3A. However, this increase was accompanied by higher background methylation. In contrast, DNMT3A produced high levels of methylation with relatively low background. DNMT3A therefore appears to be the optimal enzyme for use in most cases.

4.2 Results

To further expand the induced methylation tools available, we cloned several DNA methyltransferases in place of the DNMT3A catalytic domain used in our previous publication (McDonald *et al.*, 2016). These alternative DNMT catalytic domains include: amino acids (aa)

656 – 1616 of human DNMT1 (hDNMT1 – aa656), aa 730 – 1616 of human DNMT1 (hDNMT1 – aa730, a mouse Dnmt3a-Dnmt3L fusion (mDnmt3a-mDnmt3L), human DNMT3B (hDNMT3B), and a humanized M.SssI (hM.SssI, see Fig. 4.1A). The hDNMT1 aa730 clone excludes the CXXC domain that was found to support DNMT1's preference for hemimethylated DNA (Song *et al.*, 2011). I hypothesized that this might make DNMT1 more catalytically active in my system. The hM.SssI construct was the only construct cloned in its entirety; the humanized designation indicates it was codon optimized for expression in human cells. See methods for further cloning details.

As an initial trial of my new fusions, I targeted the *CDKN2A* promoter with three sgRNA (Fig. 4.1B, Table C.3). These sgRNA were shown to target the *CDKN2A* exon 1 in our previous study (McDonald *et al.*, 2016). The m3Dnmt3a-mDnmt3L fusion induced the most methylation. The difference of the average methylation between mDnmt3a and mDnmt3L and hDNMT3A was 10% to 20% (Fig. 4.1C). While hM.SssI performed almost equivalently to hDNMT3A, and hDNMT3B performed worse (Fig. 4.1C). Interestingly, neither hDNMT1 clone induce any methylation.

To further explore the properties of the new constructs, I varied the mass of the dCas9-DNMT fusions included in the transfection. I found that DNA methylation levels increase with the amount of the dCas9-DNMT fusion transfected and reach a maximum level of induction at 2.4 ug (Fig. 4.2B). Interestingly, hM.SssI construct induces its peak methylation at 1 ug and induced methylation decreases at higher concentrations.

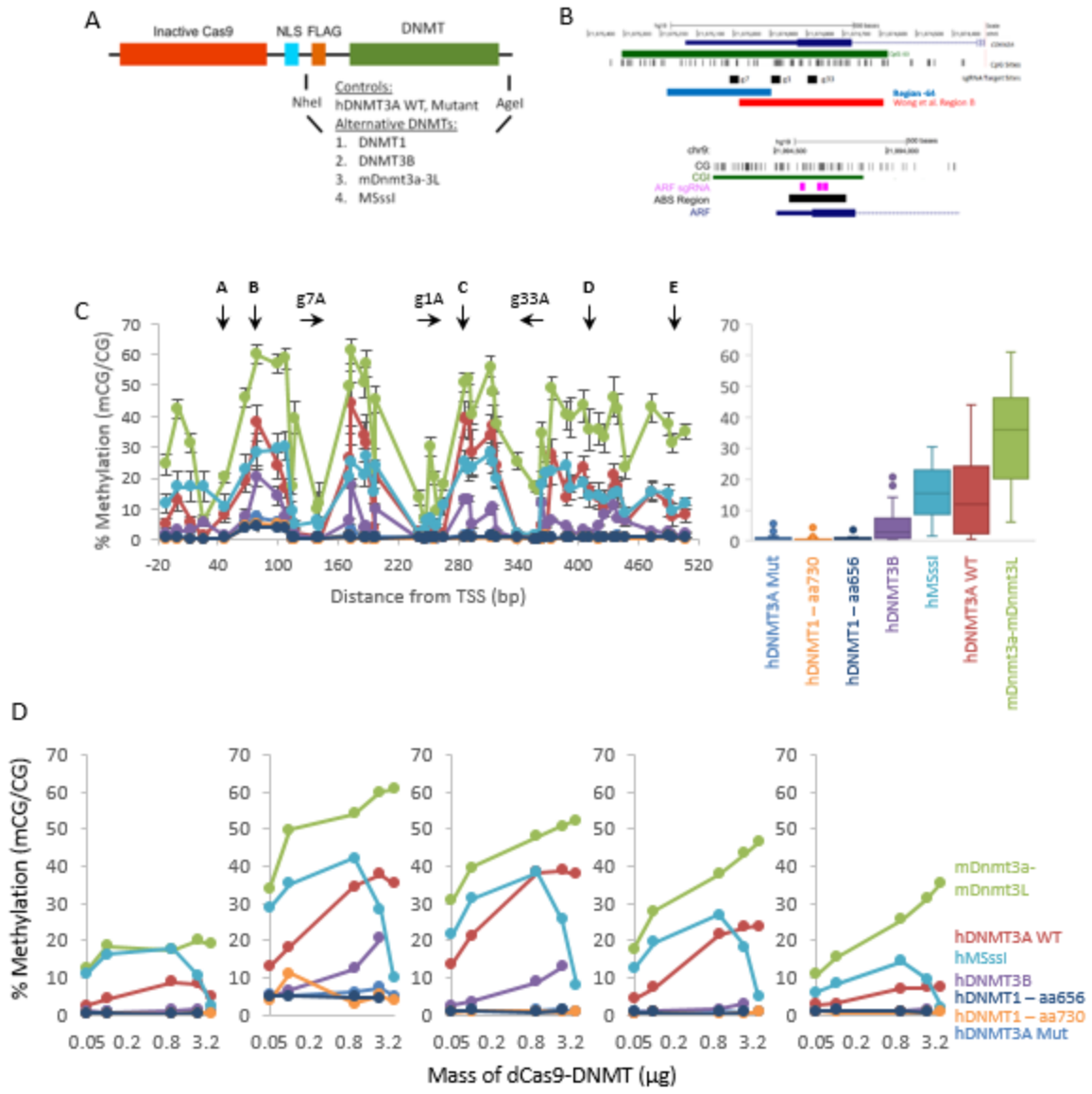


Figure 4.1: Alternative dCas9-DNA methyltransferase fusions induce site-specific methylation. **(A)** Cloning scheme for alternative DNMTs. **(B)** Genome browser shots. Top: *CDKN2A* promoter; black boxes indicate sgRNA; blue and red boxes indicate sequenced regions. Bottom: *ARF* promoter; pink boxes indicate sgRNA; the black box indicates the region sequenced. **(C)** Methylation levels induced by each DNMT at the *CDKN2A* promoter. 2.4 ug each Cas9-DNMT fusion was transfected. Boxplot shows data from scatter plot. N = 3 except hDNMT3B, hDNMT1 – aa730, and hDNMT1 – aa656 where N = 2. Error bars = +/- 1 SD for N = 3; range for N = 2. **(D)** Methylation level for each CpG indicated in (C) above across the concentration gradient of dCas9-DNMT. Each line represents a different enzyme.

When transfecting 2.4 ug of hDNMT3A, I previously observed background methylation up to 14% in the off-target controls. I also previously observed an average 9% methylation at the

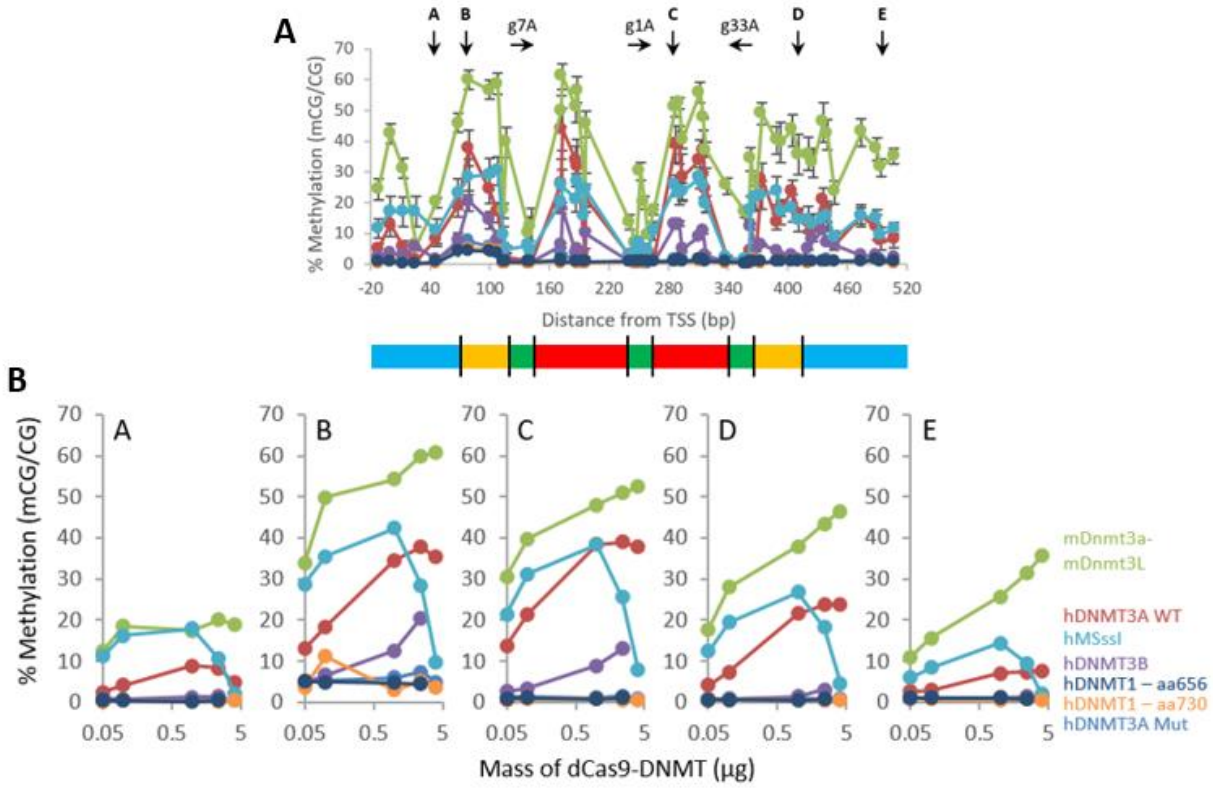


Figure 4.2: Alternative dCas9-DNA methyltransferase fusions induce site-specific methylation. (A) Methylation levels induced by each DNMT at the *CDKN2A* promoter. 2.4 ug each Cas9-DNMT fusion was transfected. Boxplot shows data from scatter plot. N = 3 except hDNMT3B, hDNMT1 – aa730, and hDNMT1 – aa656 where N = 2. Error bars = +/- 1 SD for N = 3; range for N = 2. **(B)** Methylation level for each CpG indicated in (A) above across the concentration gradient of dCas9-DNMT. Each line represents a different enzyme.

CDKN2A promoter when targeting methylation to the *ARF* promoter (McDonald *et al.*, 2016). I therefore set out to determine 1) the background methylation levels induced by the alternative DNMT constructs and 2) whether the background could be controlled by the varying the mass of the dCas9-DNMT fusion that I transfected. On-target or signal methylation was measured at the *CDKN2A* promoter when using *CDKN2A* sgRNA. Off-target or noise methylation was measured by plotting the methylation levels at the *CDKN2A* promoter when using sgRNA targeted to the *ARF* promoter ~20 kb away. mDnmt3a-mDnmt3L produced more methylation both on-target

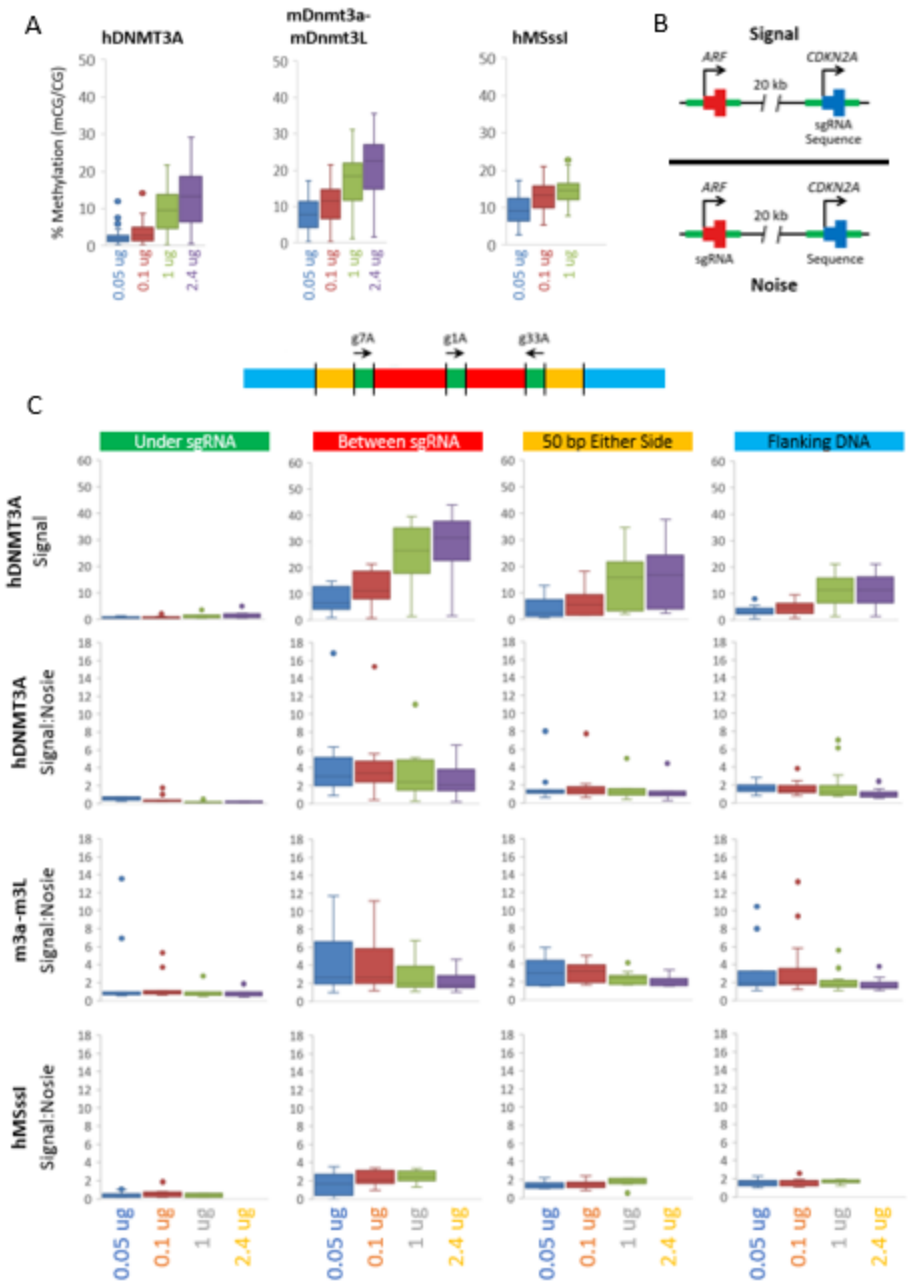


Figure 4.3: hDNMT3A induces methylation with the highest signal-to-noise ratio. (A) Distribution of DNA methylation levels across the *CDKN2A* promoter when using sgRNA to the *ARF* promoter at the specified concentrations. **(B**) Diagram of signal-to-noise calculation. Signal at each CpG in the *CDKN2A* promoter when using *CDKN2A* sgRNA is divided by the noise at the same CpG when using *ARF* sgRNA targeting 20 kb upstream. **(C)** Distribution of DNA methylation levels broken out by distance from the sgRNA as shown at the top. The first row is the signal when using hDNMT3A. The remaining rows show the signal-to-noise for hDNMT3A, mDnmt3a-mDnmt3L, and hM.SssI.

(Fig. 4.1C) and off-target (Fig. 4.3A) than hDNMT3A. The off-target methylation could be controlled to an extent by decreasing the amount of the dCas9-DNMT in the transfection (Fig. 4.3A). Nevertheless, mDnmt3a-mDnmt3L produced more off-target methylation at all concentrations.

I also tested whether the off-target methylation increased proportionally to the on-target methylation. To do this, I calculated the signal-to-noise ratio for hDNMT3A and mDnmt3a-mDnmt3L (Fig. 4.3B). The signal-to-noise was calculated for each CpG by dividing the percent methylation at the *CDKN2A* promoter treated with *CDKN2A* sgRNA by the percent methylation at the *CDKN2A* promoter treated with *ARF* sgRNAs. The highest signal-to-noise ratios occur between sgRNA (Fig. 4.3C). Within 50 bp of the sgRNA, the signal-to-noise ratio drops toward one. This indicates strong on-target methylation and a rapid loss of signal outside the target site. The signal-to-noise ratios also decrease as the amount of dCas9-DNMT fusion increases, indicating the noise increases more than the signal as more of the dCas9-DNMT construct is transfected.

The multimerization of the mDnmt3a-mDnmt3L fusion along a DNA strand has been shown to be important for strong induction of methylation over a 1.2 kb region of the EpCAM promoter (Stepper *et al.*, 2016). This raises the question of whether the background methylation represents spreading from sites bound by dCas9-DNMT or methylation induced as the DNMT transiently interacts with DNA. To explore this issue, I used ChIP-seq to identify the sites bound by the hDNMT3A fusion. I also attempted to simultaneously measure DNA methylation at the ChIPed loci by bisulfite converting and sequencing the precipitated DNA. The data I obtained after bisulfite conversation appeared to be only background (data not shown). Using *CDKN2A*

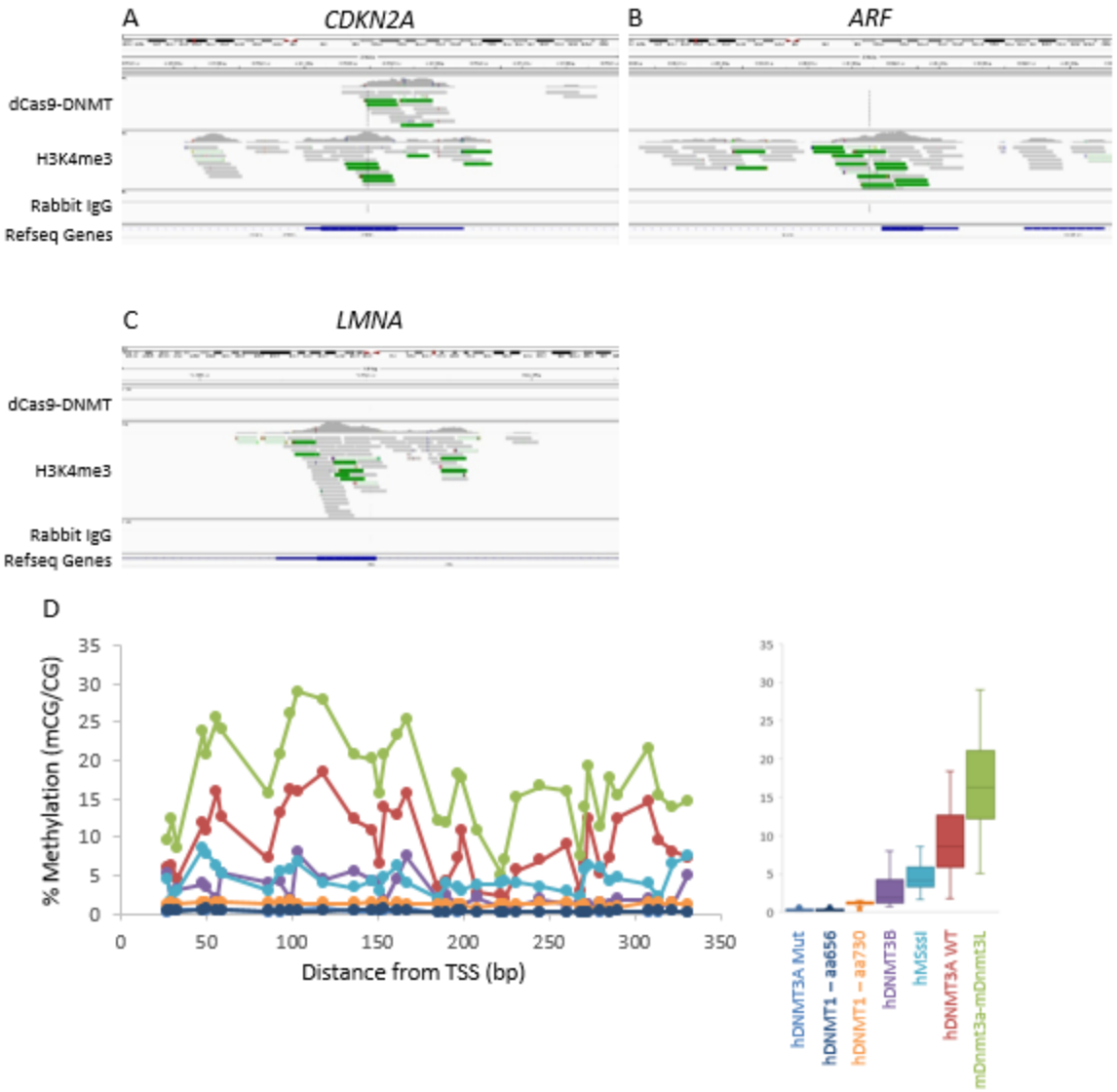


Figure 4.4: Analysis of background methylation induced by dCas9-DNMT constructs. (A) IGV browser shot of mapped reads at the *CDKN2A* Promoter: anti-FLAG ChIP of dCas9-DNMT3A fusion (top), anti-H3K4me3 ChIP (middle), rabbit IgG control (bottom). Refseq Gene track shown for reference. **(B)** Same as (A) but at the *ARF* promoter. **(C)** Same as (A) but at the *LMNA* promoter. **(D)** Methylation levels induced by each DNMT at the *LMNA* promoter. 2.4 ug each Cas9-DNMT fusion was transfected. Boxplot shows data from scatter plot. N = 1 for all samples except hDNMT1 – aa656 and hDNMT3B where N = 2.

sgRNA, dCas9-DNMT3A bound to the *CDKN2A* promoter but not the *ARF* promoter (Fig. 4.4A,B). Because the dCas9-DNMT3A does not strongly bind at the *ARF* promoter, it is possible that the methylation spreads from its target at *CDKN2A* to the *ARF* promoter. However, the

spreading hypothesis is contradicted by the methylation of the *LMNA* promoter despite the fact that it is both not bound by dCas9-DNMT3A and that it is located on a separate chromosome (Fig. 4.4C,D). Note, these libraries are sequenced at extremely low coverage of ~600,000 total reads. Additional peaks may be discovered with deeper sequencing.

To address this gap in my work, I propose to measure methylation with RRBS. Because RRBS enriches for CpG rich DNA, I expect to analyze data that overlaps several of the dCas9-DNMT ChIP-seq peaks as well as non-bound regions. gDNA from mock treated cells would set the baseline methylation level. If the methylation levels remain high outside of a 100 bp window centered on a ChIP-seq peak, that will indicate the spreading mechanism. Spreading is particularly expected with the mDnmt3a-mDnmt3L fusion because of oligomerization along the DNA strand (Stepper *et al.*, 2016). If non-bound RRBS regions--especially those distal from any bound site--also contain methylation, that will indicate the degree of off-target methylation from transient interactions. The creation of these RRBS libraries is underway and will be published in a paper containing the work described in this chapter (currently under preparation).

A very small amount of CHG methylation was induced by the hDNMT3A construct at the *CDKN2A* promoter in our previous study (McDonald *et al.*, 2016). Unexpectedly, the mDnmt3a-mDnmt3L construct induced much higher levels of CHG methylation. None of the other enzymes do this (Fig. 4.5), including hDNMT3B which is known to induce non-CpG methylation (Aoki *et al.*, 2001). The peak doublet represents methylation of two CHG sites directly flanking a SP1 transcription factor binding site, suggesting that these increases may affect gene expression. When the enzymes are targeted to the *ARF* promoter, this CHG

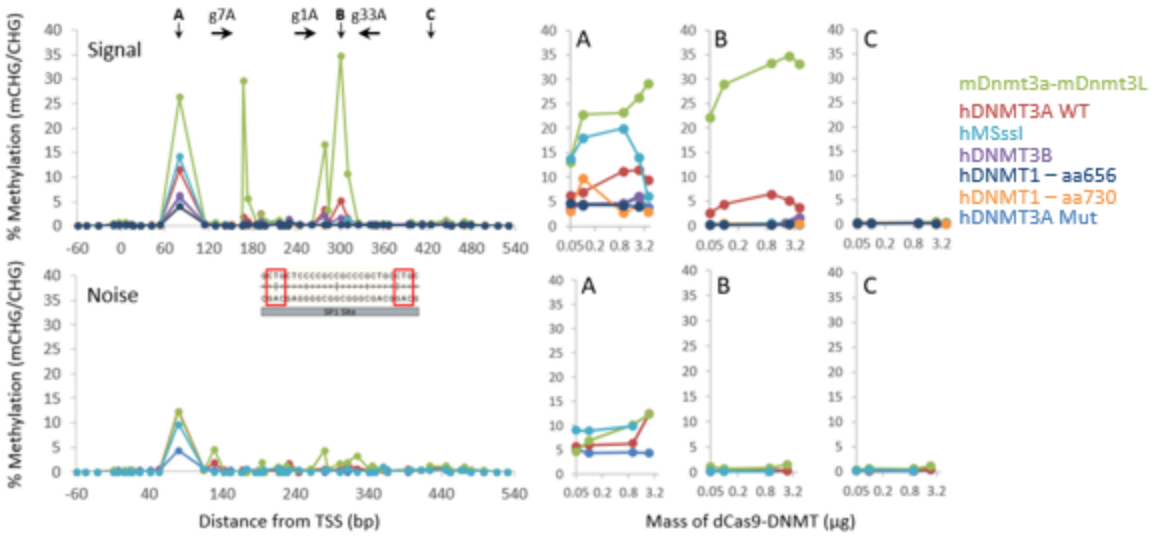


Figure 4.5: The mDnmt3a-mDnmt3L fusion induces a high level of non-CpG methylation. CHG methylation data at the *CDKN2A* promoter. Top: signal from using *CDKN2A* sgRNA. Bottom: noise from using *ARF* sgRNA. Data in plots at left are all at the 2.4 ug concentration. Plots for the labeled cytosines are presented at right across the concentration gradient.

methylation is lost (Fig. 4.5). However, one CHG site in the *ARF* promoter gains methylation (data not shown). This suggests that observation of non-CpG induced methylation must occur in close proximity to the bound dCas9-DNMT fusion.

I next sought to further increase the amount of induced methylation by either 1) combining hDNMT3A with hDNMT1 and 2) combining DNMTs with the KRAB repressor. Because DNMT3A multimerization favors hemimethylated sites (Jia *et al.*, 2007; Jurkowska et al., 2008), we hypothesized that combined treatment with DNMT1 – which preferentially methylates hemimethylated sites – might increase methylation levels. The KRAB repressor recruits SETDB1, HP1, the NuRD complex, and DNMTs, which cooperate to form a repressive chromatin state (Ecco *et al.*, 2017). We expected that the heterochromatin formation encouraged by the KRAB repressor would aid methylation as it did for Amabile *et al.* (2016). However, neither approach increased the amount of induced DNA methylation (Fig. 4.6).

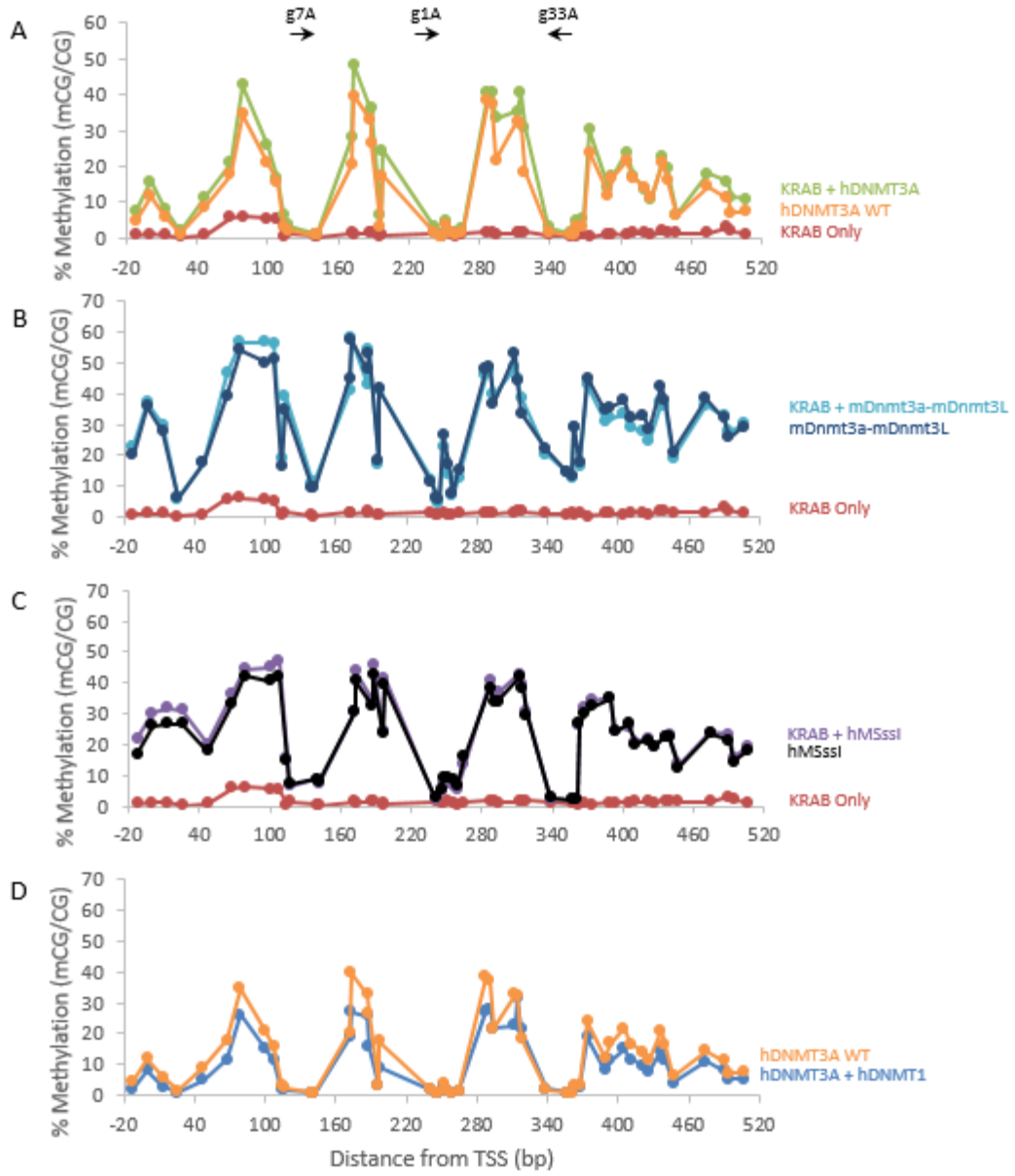


Figure 4.6: Attempts to increase the level of induced methylation. DNA methylation levels after targeted methylation treatment with **(A)** KRAB + hDNMT3A, **(B)** KRAB + mDnmt3a-mDnmt3L, **(C)** KRAB + hM.SssI, and **(D)** hDNMT3A + hDNMT1. The DNMT/KRAB combinations are listed at right. sgRNA locations are indicated by arrows. All increase attempts, N = 1. Reference samples for hDNMT3A, mDnmt3a-mDnmt3L, and hM.SssI, N = 2.

I also targeted the *ARF* promoter with the three most active DNMTs. I obtained similar levels of DNA methylation (Fig. 4.7A). I previously observed a 19% decrease in *ARF* expression following methylation induction with hDNMT3A (McDonald *et al.*, 2016). I therefore checked

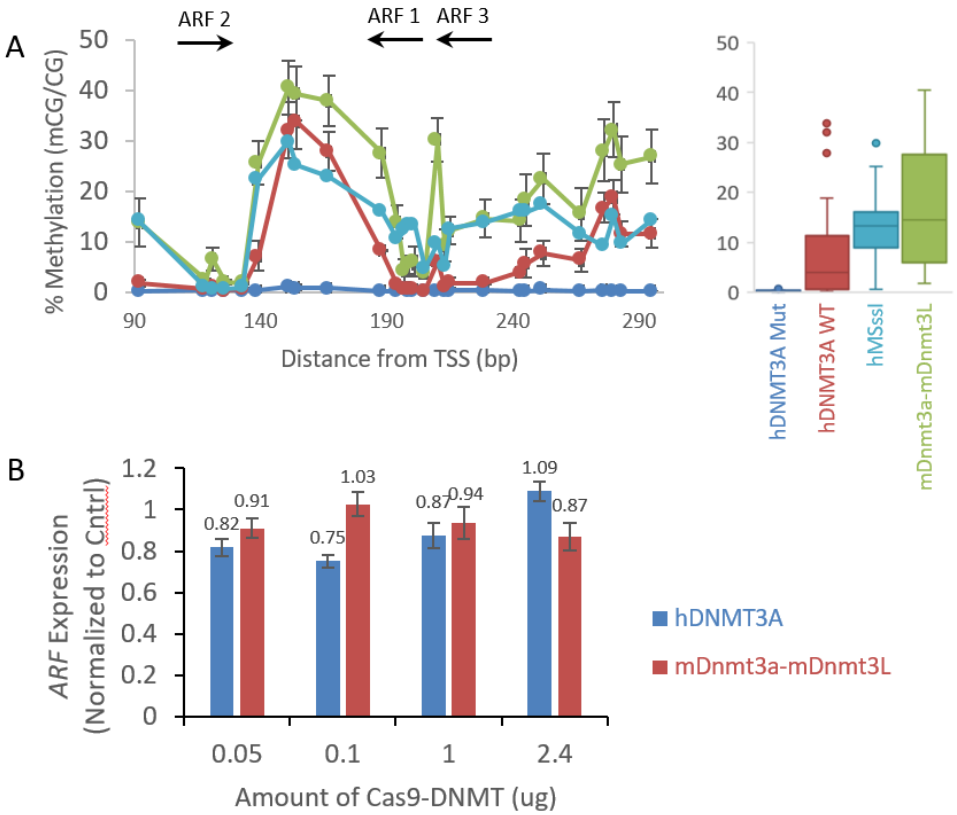


Figure 4.7: DNA methylation and expression analysis of ARF after alternative dCas9-DNMT treatment. (A) Average methylation levels induced at the *ARF* promoter. Boxplot shows data from scatter plot. Arrows indicate the location of *ARF* sgRNA. N = 3 for mDnmt3a-mDnmt3L, N = 2 for hDNMT3A WT and Mutant, N = 1 for hM.SssI. Error bars = +/- 1 SD for N = 3. Error bars = range for N = 2. **(B)** *ARF* expression after treatment with hDNMT3A and mDnmt3a-mDnmt3L fusions across a concentration gradient. Each point is normalized to a matching sample from a catalytically inactive dCas9-DNMT fusion. N = 2. Error bars = +/- 1 SE.

ARF expression after treatment with the mDnmt3a-mDnmt3L and hDNMT3A fusions (Fig. 4.7B). hDNMT3A produced an average 16.8% decrease in *ARF* expression in the 0.05 ug, 0.1 ug, and 1 ug. However, the knockdown was not observed in the 2.4 ug samples. Additionally, there was no additional knockdown from increasing the amount of transfected dCas9-DNMT fusion plasmid and thereby the methylation level. Interestingly, I observed almost no knockdown with the mDnmt3a-mDnmt3L fusion (Fig. 4.7B), despite its slightly higher level of DNA methylation. In the latter case, the increased background induced by the mDnmt3a-mDnmt3L

construct may interfere with methylation induced silencing of *ARF*. However, these observations may also indicate that DNA methylation only loosely controls *ARF* expression.

4.3 Discussion

In this study, I created and characterized dCas9-DNMT fusions containing the hDNMT1, mDnmt3a-mDnmt3L, hDNMT3B, and hM.SssI DNMTs. I showed that these DNMTs are capable of inducing targeted DNA methylation. Only the mDnmt3a-mDnmt3L construct induced more methylation than the original hDNMT3A construct. This is expected because the DNMT3A-DNMT3L interaction is known to enhance DNMT3A catalytic activity (Suetake *et al.*, 2004). However, these gains came at the cost of higher off-target methylation and higher non-CpG methylation. These effects could be viewed as drawbacks, but we believe that they may also represent opportunities to study niche applications. For example, the mDnmt3a-mDnmt3L construct may be useful for methylating a large area or for inducing non-CpG methylation – as long as increased background methylation can be tolerated. We have therefore expanded the number of usable dCas9-DNMT constructs.

Methylated and silenced CpG island promoters can have average methylation levels above 70% (Weber *et al.*, 2007). Neither the DNMT/KRAB nor the hDNMT3A/hDNMT1 combination increased the amount of induced DNA methylation. In contrast to my findings, Amabile *et al.* (2016) were able to induce >80% methylation on average in a 2.5 kb region using a combination of Dnmt3a, Dnmt3L, and KRAB in HEK293T cells (Amabile *et al.*, 2016). We believe this difference could be explained by the different amounts of time that methylation was induced. In this study, we analyzed methylation after 4 days while Amabile *et al.* induced methylation for 30 days (Amabile *et al.*, 2016). Induced methylation proves to be unstable in our

system and decreases quickly after the cells are passaged on day 4. In contrast, we found that DNA methylation induced by a lentiviral construct for 12 days was much more stable (McDonald *et al.*, 2016). This indicates that sustained induction of DNA methylation is an important factor when designing an induced methylation study.

The background methylation can be controlled by the DNMT domain and the amount of dCas9-DNMT transfected. Previous studies that report lower levels of DNA methylation achieve this by transfecting plasmid amounts roughly equivalent to our 100 ng sample (Stepper *et al.*, 2016; Vojta *et al.*, 2016). Therefore, control of the dCas9-DNMT fusion construct expression appears to be the key to controlling off-target DNA methylation. The use of inducible expression from an integrated lentiviral plasmid would be a way to accomplish both control and the extended expression needed to reach high methylation levels.

The induction of non-CpG methylation by the mDnmt3a-mDnmt3L construct creates the possibility to study the effects of directed non-CpG methylation. Interestingly, hDNMT3A and hDNMT3B can both catalyze non-CpG methylation (Jurkowska and Jeltsch, 2016) but they do not do so in this context. This suggests that the stimulatory effect of the Dnmt3L could cause this effect. This opportunity to capitalize on this increased activity is highlighted by the fairly specific methylation of CHG sites flanking an SP1 transcription factor binding site. Non-CpG methylation is known to inhibit Sp transcription factor binding, indicating that this methylation event could alter transcription (Inoue and Oishi, 2005). Furthermore, non-CpG methylation was observed at this region in breast cancer samples, but not on the same sites (Woodcock *et al.*, 1999). Though CHG methylation is more common in specific sequence contexts (Aoki *et al.*, 2001), there are too many sites that remain unmethylated for sequence preference to explain the

specificity. The main unifying factor is that all three sites occur within ~20-30 bp of the 3' end of a sgRNA binding site. I would like to suggest that the CHG methylation is an uncommon misfire by the hyperactive mDnmt3a-mDnmt3L fusion and thereby only occurs within proximity to the sgRNA binding sites. However, this still does not explain why many other CHG sites in this range are excluded. This phenomenon requires further study before this tool can be used to reliably induce non-CpG methylation. Until induced non-CpG methylation can be controlled and/or prevented, it will remain an artefact that hinders interpretation of CpG methylation data.

The DNA methylation field is no longer dependent on techniques such as DNMT knockout or inhibition to manipulate methylation events. The use of the reprogrammable CRISPR/Cas9 system to target DNA methylation activity to specific regions of the genome has facilitated the study of specific changes in DNA methylation. In this study, we have described the effects of targeted methylation with multiple previously unused DNMTs. Most DNMTs were functional and provided varying levels of DNA methylation. hDNMT3A produced the most methylation while maintaining the highest signal-to-noise ratio. mDnmt3a-mDnmt3L produced the most methylation overall at the cost of increased background and non-CpG methylation. These constructs may prove useful in nice applications where high background can be tolerated or if it is necessary to study the effect of a DNMT other than DNMT3A.

4.4 Materials and Methods

Plasmid Construction

The dCas9-DNMT3A (Addgene #78256) and E756A mutant dCas9-DNMT3A (Addgene #78257) were constructed previously (McDonald *et al.*, 2016). Novel DNMT catalytic domains were cloned between the NheI and AgeI sites of pCMV-dCas9-D3A (Addgene #78256,

(McDonald *et al.*, 2016). The catalytic domains were the following: DNMT1, amino acids (aa) 656-1616 and 730-1616 from NP_001124295.1; mDnmt3a-mDnmt3L, aa 612-912 of NP_783328.1 and aa 1069-1716 of NP_037501.2 as one piece amplified from pET-Dnmt3a3l-sc27 (Addgene #71827, (Siddique *et al.*, 2013); DNMT3B, aa 535-854 of NP_008823.1; M.SssI, cloned in its entirety based on the sequence of a construct generously given to us by Timothy Bestor. IDT codon optimization was performed on the M.SssI sequence before it was synthesized via IDT gBlocks. To create the mutant mDnmt3a-mDnmt3L control, I synthesized the NheI to BstEII fragment with the E752A mutation via IDT gBlocks and used this to replace the WT sequence. With the exception of M.SssI, each DNMT was amplified by a nested PCR strategy that also added the FLAG tag as part of the 5' end of the forward primer in the final/inner PCR (see primers in Table C.1). sgRNA plasmids were identical to those used previously (Table C.2, (McDonald *et al.*, 2016).

Cell Culture and Transfection

HEK293T cells were acquired from ATCC (CRL-3216) and grown in DMEM supplemented with 10% FBS (Gibco), 1× Penicillin/Streptomycin (Gibco), and 2 mM GlutaMax (Gibco). For transfection experiments, 3×10^5 HEK293T cells were plated in a 60 mm dish. The next day, the cells were transfected with Lipofectamine LTX (Thermo Fisher Scientific). The Lipofectamine:DNA ratio was 3.5, with a total of 5.5 μg of plasmid DNA. 2.4 ug of the dCas9-DNMT fusion plasmid and 2.4 ug of the sgRNA plasmids (800 ng each) were transfected. dCas9-DNNMT masses transfected in titration experiments were: 50 ng, 100 ng, 1 μg, 2.4 μg, and 4 μg. When the dCas9-DNMT mass was decreased below 2.4 μg, a filler plasmid was used to hold overall DNA mass constant. 0.7 μg of pMaxGFP was co-transfected in order to indicate

the transfection efficiency. The plasmid DNA was first diluted in Gibco OptiMEM, then Lipofectamine LTX was added and mixed in by inversion. After 30 min, the transfection mixture was added dropwise to the cells and they were placed back in the incubator. This procedure was scaled up for the 150 mm dishes used in ChIP-seq (see below). Further details can be found at http://epigenomics.wustl.edu/edwardsLab/index.php/5mc-editting.

Amplicon Bisulfite Sequencing

Genomic DNA (gDNA) was isolated using the Zymo Research Quick gDNA MiniPrep kit and quantified with the Qubit dsDNA broad range assay (Thermo Fisher Scientific). 1 ug of gDNA was bisulfite converted with the Zymo Research EZ DNA Methylation kit according to the manufacturer's instructions. Target regions were amplified with the Qiagen PyroMark PCR kit using primers locate in Table C.3. A tenfold molar excess of Illumina sequencing Y-adapters was then annealed to 100 ng of PyroMark PCR product (total Y-adapter mass varies with PCR product length) with NEB Quick Ligase for 15 min. The ligations were purified over a 1.5% agarose II (ISC BioExpress) gel in 1×TBE in order to remove incompletely ligated DNA. Custom 13 bp barcode index sequences were added via PCR using NEB Phusion. The standard Phusion PCR protocol was followed but with the following primer concentrations (for a 50 µl reaction): 1 µl of Illumina Primer 1.0 (25 µM), 1 µl of Illumina PCR Primer 2.0 (diluted fresh to 0.5 µM final concentration), and 1 µl of the index primer (25 µM). The thermocycler protocol was the following: (1) 98°C, 30 s; (2) 98°C, 10 s; (3) 64°C, 30 s; (4) 72°C, 30 s; (5) Return to step two 11 times; (6) 72°C 5 min; (7) 4°C hold. The PCR products were again purified over a 1.5% agarose II gel in 1× TBE, and their concentration was measured with the Qubit dsDNA high sensitivity kit. Individually indexed samples were pooled and submitted for sequencing.

Amplicon bisulfite sequencing data were checked for quality using FastQC, adaptor and poor-quality sequence (quality less than 20) was trimmed using fqtrim, and the trimmed sequences were mapped to the target sequences using Bismark (Krueger and Andrews, 2011).

RT-qPCR

RNA was extracted with the Zymo Research Quick mRNA Miniprep kit. RNA concentration was measured with the Qubit RNA BR kit. RNA integrity was determined by visualizing rRNA bands using agarose gel electrophoresis. Reverse transcription was performed using the Bio-Rad iScript Reverse Transcriptase kit. Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was performed with the Bio-Rad iTaq Universal with SYBR Green reagent on an Applied Biosystems Viia7 instrument. The thermocycler protocol was the following: (1) 95°C, 20 s; (2) 95°C, 3 s; (3) 60°C, 20 s; for 40 cycles. qPCR primers are listed in Table C.4. A melt curve was performed to indicate there was not off-target amplification. Data was analyzed as described by Hellemans *et al.* (2007) using the geometric mean of *ACTB*, *GAPDH*, and *RPL0* as an internal control.

ChIP-seq

Cells in 150 mm dishes were grown in 30 mL of media. To cross-link, 1 m L of 37% Formaldehyde (1% final concentration) was added directly to the cells. The cross-linking reaction ran for 10 min with constant rocking at room temperature. 3.5 mL of 1.25 M Glycine (0.125 M final concentration) was added to quench the cross-linking reaction. Quenching proceeded for 5 min with constant rocking at room temperature. The cells were placed on ice and scraped off the dish into 4°C DPBS (Gibco). The cells were pelleted at 500 xg for 5 min at 4C, washed once with DPBS, flash frozen on dry ice and stored at -80C. Frozen cell pellets were

thawed on ice and resuspended in 300 uL SDS Lysis Buffer supplemented with protease inhibitors (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8, 1X cOmplete Protease Inhibitor Cocktail from Roche) for every ~$1x10^7$ cells in the pellet. Chromatin was sheared in 300 uL, ~$1x10^7$ cell aliquots using a BioRuptor Pico: 30 cycles of 30 sec ON, 30 sec OFF. Cell debris was pelleted at 14,000 rpm and 4°C for 10 min. The supernatant was diluted to 1 mL with ChIP Dilution Buffer supplemented with protease inhibitors (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8, 167 mM NaCl, 1X cOmplete Protease Inhibitor Cocktail from Roche) and placed on a rotator at 4°C for 30 min. 100 uL of sheared DNA from each sample was aliquoted and stored at -20°C overnight as an input sample. 5 ug of antibody was added to the remaining chromatin and this mixture was incubated overnight on a rotator at 4°C. Antibodies used were: mouse monoclonal anti-FLAG M2 clone (Sigma F1804), mouse monoclonal anti-Cas9 (Active Motif 61578), rabbit monoclonal anti-H3K4me3 clone MC315 (Millipore 04-745), and rabbit IgG (Vector I-1000).

30 uL of Invitrogen Protein G Dynabeads were washed twice and resuspended in 200 uL of DPBS plus 0.01% Tween-20. The chromatin was then added to the beads and precipitation performed for 4 hours on a rotator at 4°C. The beads were pelleted on a magnet and washed for 5 min at 4°C on a rotator in the following series of buffers: Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl), High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 500 mM NaCl), LiCl Wash Buffer (0.25 M LiCl, 1% Igepal CA-630, 1% Sodium Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8), and twice in TE (10 mM Tris-HCl pH 8, 1 mM EDTA). DNA was eluted by adding 125 uL of fresh Elution Buffer (1% SDS, 0.1 M $NaHCO_3$), vortexing, and rotating for 15 min at

room temperature. After the supernatant was recovered, the wash/elution steps were repeated and the eluates were combined for 250 uL total. Input samples were thawed and diluted to 250 uL total with ChIP Dilution Buffer supplemented with protease inhibitors. 10 uL of 5 M NaCl was added to each sample and cross-linking was reversed at 65°C overnight.

To begin cleanup of the DNA, 15 uL of Proteinase K (Thermo Scientific EO0491) was added to each sample. Digestion was performed for 2 hours at 45°C. The DNA was extracted with 550 uL of Phenol-Chloroform-Isoamyl Alcohol (Sigma 77617). The samples were vortexed and spun at 14,000 rpm and 4°C for 5 min. The aqueous layer was transferred to a fresh tube, the extraction was repeated, and the aqueous layers were combined. To precipitate the DNA, 1 uL of glycogen and 1 mL of 100% ethanol were added to each sample. Samples were stored at -80°C overnight. The DNA was pelleted by thawing the precipitation on ice, pelleting the DNA at 14,000 rpm and 4°C for 20 min. The pellet was washed with 500 uL of ice cold 70% ethanol. The pellet was air dried until no ethanol remained and resuspended in 30 uL of 10 mM Tris-HCl. Multiple DNA samples, each corresponding to ~$1x10^7$ cells can be pooled if necessary at this step. The final DNA concentration was measured using the Qubit dsDNA HS kit.

ChIP-seq libraries were produced following the protocol of the NEBNext Ultra DNA Library Prep Kit for Illumina (E7370) using the methylated Illumina adapter found in the NEBNext Multiplex Oligos for Illumina (E7535). Bisulfite conversion was carried out with the Zymo Research EZ DNA Methylation kit. Final libraries were quantified by Qubit dsDNA HS kit and submitted for sequencing.

4.5 Supplementary Material

Supplementary material is in Appendix C.

Chapter 5: Conclusions and Future Directions

The role of DNA methylation in inducing gene silencing has been and continues to be an active area of research. DNA methylation is known to silence imprinted genes, retrotransposons, and X-linked genes on the inactive X-chromosome. However, the role for silencing genes outside these contexts is not fully understood. In Chapter 2, 5-azacytidine treatment increased the expression of *PTGER4*. This validated the negative correlation between methylation and expression observed genomic data from a cell line model of breast cancer aromatase inhibitor (AI) resistance. It also demonstrated how a single DNA methylation change could contribute to cancer progression. In Chapter 3, I designed and demonstrated a dCas9-DNMT3A catalytic domain fusion that could induce targeted methylation changes. These changes were sufficient to decrease the expression of *CDKN2A* and *ARF*. In Chapter 4, I optimized the dCas9-DNMT system by measuring the effectiveness of alternative DNMTs. Overall the data presented in the preceding chapters support the hypothesis that DNA methylation can induce gene silencing and play an important functional role in disease.

5.1 Studying how DNA Methylation and Chromatin Modifications Interact during Gene Silencing

Research in the DNA methylation field shows that DNA methylation induced silencing is highly dependent on context. As a general example, promoter methylation favors gene silencing, but gene body methylation marks actively transcribed genes (Jones, 2012). The chromatin context also helps to determine DNA methylation location and function. H3K4me3 prevents the activity of DNMT3A/3L at active promoters (Ooi *et al.*, 2007) while H3K36me3 attracts DNMT3B to methylate gene bodies (Baubec *et al.*, 2015). My work supports this context

dependence. DNA methylation induced *CDKN2A* silencing, but only when the entire promoter region was methylated (Fig. 3.1, 3.2). In contrast, *ARF* silencing occurred after methylation of a much smaller region (Fig. 3.3). While my own work does not examine the chromatin context, I believe this is an important direction for the field.

The interactions of various epigenetic regulators are just beginning to be understood. Amabile *et al.* (2016) show that DNMT3A, DNMT3L, and KRAB act synergistically to silence gene expression. Similarly, both the G9A H3K9 methyltransferase and v-ErbA histone deacetylase silence *VEGF-A* expression when targeted there. When combined, they synergize to more strongly silence *VEGF-A* (Snowden *et al.*, 2002). In contrast, both DNA methylation and H3K9me3 silenced *VEGF-A* individually, but neither mark recruited the other, which indicates a lack of synergy (Kungulovski *et al.*, 2015). Epigenetic silencing can also be position dependent. Amabile *et al.* (2016) also discovered position-dependent silencing when the randomly integrated reporter was more efficiently silenced than at the *AAVS1* site. The LSD1 histone demethylase silenced genes when targeted to an enhancer but not the promoter. In contrast, the KRAB repressor was active when targeted to both enhancers and promoters (Kearns *et al.*, 2015). Repressors such as KRAB can further complicate the study of context dependence by recruiting different epigenetic modifiers. Finally, the repression of a gene *B2M* promoter was weaker in HEK293T cells than in K562 cells in all tested conditions (Amabile *et al.*, 2016). This indicates that cell line specific expression of epigenetic readers and writers may significantly impact the effects of epigenetic editing. The resulting variability will make it difficult to generalize results from epigenome editing studies and may hinder the adoption of targeted epigenetic therapies.

The dCas9 system allows me to address the mechanisms behind epigenetic context dependence. For example, I could test the hypothesis that two or more epigenetic modifiers synergize. I could also test a hypothesis that a specific member of a large complex drives the effects of that complex. The system I would construct to test these hypotheses is the following. The dCas9 and each of my desired modifiers would be stably integrated into the genome under individual inducible promoters. Note that the dCas9 and each effector domain would be separate. Instead, the interaction between dCas9 and each effector domain would be mediated by two proteins that tightly bind each other. Because it is based on only a few plasmids, the system could be easily implemented in multiple cell lines. Location dependent effects could be easily studied by using a different set of sgRNA. Stable integration also allows for the study of both short- and long-term effects of epigenetic silencing as well as to observe how the epigenetic context is altered during repeated cycles of activation and repression.

For a specific example, let's consider further study of *CDKN2A* regulation. In addition to DNA methylation, *Cdkn2a* has been suggested to be silenced PRC2 and reactivated by hSNF5 in mouse embryonic fibroblasts (MEF, Wilson *et al.*, 2010). Wilson *et al.*, however, used conditional knockouts that affected the expression of many genes and they did not explore the role of DNA methylation in silencing *Cdkn2a*. With my new system, I could study the effects of targeted PRC2 or hSNF5 activity and consider additional silencing factors, such as DNMTs.

I would first demonstrate that hSNF5 can expel PRC2 as demonstrated by Wilson *et al.* (2010). I would then extend the reach of my experiments by including DNA methylation. Using both PRC2 and DNMTs simultaneously, I would assess any synergy between these two marks. I would test if pre-treatment of the *Cdkn2a* promoter with DNA methylation is able prevent

hSNF5 upregulation when PRC2 cannot. If DNA methylation can, I would further unravel the mechanism by finding the specific marks that inhibited hSNF5 activity. It may prove difficult to target the entire PRC2 complex, and in that case, I would just target the catalytic core, EZH2. This approach is also slightly constrained by the number of individually inducible promoters, but at least eight inducible promoters already exist (Ede *et al.*, 2016). These limitations can be overcome in time, and I believe that this type of experiment will drive our understanding of epigenetic regulatory logic in years to come.

5.2 Identifying Functional Methylation Changes to Refine DNA Methylation Based Cancer Therapeutics

DNA methylation based therapies have already been approved by the FDA. These treatments use DNMT inhibitors, 5-azacytidine and 5-aza-2-deoxycytidine, to demethylate the genome and reactivate abnormally silenced genes. Treatment with 5-azacytidine caused 15% or more of myelodysplastic syndrome (MDS) or acute myeloid leukemia (AML) patients to survive longer while present fewer cancerous blood cells and more healthy cells (Jones *et al.*, 2016). However, both intrinsic and acquired resistance to these drugs is common (Prébet *et al.*, 2011; Qin *et al.*, 2011), and the mechanism of acquired resistance is currently not explained. It is possible that effects of hypomethylation itself could be promoting resistance through one of the mechanisms described in the Introduction. After relapse, overall survival is 5.6 months (Prébet *et al.*, 2011). Understanding if and how DNA methylation contributes to resistance could direct the course of treatment after relapse to improve patient outcomes.

This situation is similar to the hypothetical contribution of DNA methylation changes to acquired aromatase inhibitor therapies I explored in Chapter 2. Unlike in Chapter 2, however, I

have the advantage of the dCas9-DNMT system. This system can easily target many regions simultaneously. This approach can rapidly screen for functional changes at sites suggested correlation-based analysis of genomic data sets. Identifying these functional epigenetic changes in turn better inform the use of epigenetic therapies in cancer.

To study whether hypomethylation contributes to 5-azacytidine resistance in myelodysplastic syndrome (MDS), I would use bioinformatic tools to identify hypomethylated genes with altered expression in resistant patient samples. A previous study identified 394 genes hypomethylated at relapse in two out of four myelodysplastic syndrome patients (Qin *et al.*, 2011). However, this study did not include expression data. My lab has built tools to predict the genes that are controlled by DNA methylation from analysis of DNA methylation and expression data (Schlosberg *et al.*, 2017; VanderKraats *et al.*, 2013). These tools helped to identify a set of genes that shows a strong correlation between 5-azacytidine induced methylation loss and upregulation in acute myeloid leukemia (AML, Lund *et al.*, 2014). I would focus these bioinformatic tools on a comparison of high resolution DNA methylation data from MDS patient samples before and after the development of resistance. The combination of the Qin *et al.* (2011) MDS genes, the Lund *et al.* (2014) AML genes, and the genes from my MDS analysis would form the core target genes for my lentiviral screen.

The lentiviral screen would be based in a MDS cell line with the dCas9-DNMT3A fusion stably integrated under the control of an inducible promoter. I would then transduce a library of paired sgRNA at a MOI = 1. The sgRNA pairs would be designed to target sites within 50 to 100 bp of each other and therefore likely to give a strong signal at the target site. Each pair of sgRNA would be expressed from a single plasmid under the control of a pair of divergent U6 promoters.

I would monitor which methylation changes are selected for/against in the cell population by monitoring the sgRNA sequences that remain in living cells. I am specifically looking for hypermethylated sites that are selected against, indicating that methylation prevents growth. In MDS, hypomethylation of these sites due to 5-azacytidine might therefore encourage growth. It would be important to validate these findings using another demethylating approach such as targeted TET activity or DNMT knockdown.

The knowledge of functional methylation changes that the above approach could produce would improve our ability to apply DNA methylation information in the clinic. DNA methylation changes can be used to detect cancer (Warton and Samimi, 2015) and identify specific cancer subtypes (Ciriello *et al.*, 2013; The Cancer Genome Atlas Network, 2012). However, the results of the experiment described above would allow for improvement in diagnosis and treatment. Knowledge of specific methylation events that drive carcinogenesis could help predict which patients would respond well to epigenetic therapies such as 5-azacytidine. Similarly, knowledge of the epigenetic events that lead to resistance would allow us to detect emerging resistance and proactively modify the course of treatment.

The dCas9-DNMT system facilitates targeted studies that require the nuanced manipulation of the epigenome. With further development, the extreme flexibility of this technology would also allow the study of complex interactions at the chromatin level and the high-throughput identification of function epigenetic changes in diseases such as cancer. The creation of the dCas9-DNMT technology therefore represents an exciting step forward for the DNA methylation field and opens many avenues for further study.

References

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M.C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. Nucleic Acids Res. *35*, 5402–5408.

Aguilar, H., Solé, X., Bonifaci, N., Serra-Musach, J., Islam, A., López-Bigas, N., Méndez-Pertuz, M., Beijersbergen, R.L., Lázaro, C., Urruticoechea, A., *et al.* (2010). Biological reprogramming in acquired resistance to endocrine therapy of breast cancer. Oncogene *29*, 6071–6083.

Allred, D.C., Brown, P., and Medina, D. (2004). The origins of estrogen receptor alpha-positive and estrogen receptor alpha-negative human breast cancer. Breast Cancer Res. *6*, 240.

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell *167*, 219–232.e14.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andersson, A.F., and Banfield, J.F. (2008). Virus Population Dynamics and Acquired Virus Resistance in Natural Microbial Communities. Science *320*, 1047–1050.

Aoki, A., Suetake, I., Miyagawa, J., Fujio, T., Chijiwa, T., Sasaki, H., and Tajima, S. (2001). Enzymatic properties of *de novo*-type mouse DNA (cytosine-5) methyltransferases. Nucleic Acids Res. *29*, 3506–3512.

Appanah, R., Dickerson, D.R., Goyal, P., Groudine, M., and Lorincz, M.C. (2007). An Unmethylated 3′ Promoter-Proximal Region Is Required for Efficient Transcription Initiation. PLoS Genet *3*, e27.

Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Hérault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science *354*, 909–912.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science *315*, 1709–1712.

Barrès, R., Osler, M.E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., and Zierath, J.R. (2009). Non-CpG Methylation of the PGC-1α Promoter through DNMT3B Controls Mitochondrial Density. Cell Metab. *10*, 189–198.

Barres, R., Kirchner, H., Rasmussen, M., Yan, J., Kantor, F.R., Krook, A., Näslund, E., and Zierath, J.R. (2013). Weight Loss after Gastric Bypass Surgery in Human Obesity Remodels Promoter Methylation. Cell Rep. *3*, 1020–1027.

Bashtrykov, P., Jankevicius, G., Smarandache, A., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2012). Specificity of Dnmt1 for Methylation of Hemimethylated CpG Sites Resides in Its Catalytic Domain. Chem. Biol. *19*, 572–578.

Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2014). The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism. J. Biol. Chem. *289*, 4106–4115.

Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. Cell *153*, 480–492.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247.

Baylin, S.B., Höppener, J.W.M., Bustros, A. de, Steenbergh, P.H., Lips, C.J.M., and Nelkin, B.D. (1986). DNA Methylation Patterns of the Calcitonin Gene in Human Lung Cancers and Lymphomas. Cancer Res. *46*, 2917–2922.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. *12*, R10.

Berkyurek, A.C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirakawa, M., and Tajima, S. (2014). The DNA Methyltransferase Dnmt1 Directly Interacts with the SET and RING Finger-associated (SRA) Domain of the Multifunctional Protein Uhrf1 to Facilitate Accession of the Catalytic Center to Hemi-methylated DNA. J. Biol. Chem. *289*, 379–386.

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P.E., Dijk, C.M. van, Tollenaar, R.A.E.M., *et al.* (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat. Genet. *44*, 40–46.

Bernstein, D.L., Le Lay, J.E., Ruano, E.G., and Kaestner, K.H. (2015). TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J. Clin. Invest. *125*, 1998–2006.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum. Mol. Genet. *9*, 2395–2402.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the

mammalian enzymes is related to bacterial restriction methyltransferases. J. Mol. Biol. *203*, 971–983.

Bestor, T.H., Edwards, J.R., and Boulard, M. (2015). Notes on the role of dynamic DNA methylation in mammalian development. Proc. Natl. Acad. Sci. 201415301.

Bhatnagar, S., Zhu, X., Ou, J., Lin, L., Chamberlain, L., Zhu, L.J., Wajapeyee, N., and Green, M.R. (2014). Genetic and pharmacological reactivation of the mammalian inactive X chromosome. Proc. Natl. Acad. Sci. *111*, 12591–12598.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. *8*, 1499–1504.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91–99.

Bliek, J., Maas, S.M., Ruijter, J.M., Hennekam, R.C., Alders, M., Westerveld, A., and Mannens, M.M. (2001). Increased tumour risk for BWS patients correlates with aberrant H19 and not KCNQ1OT1 methylation: occurrence of KCNQ1OT1 hypomethylation in familial cases of BWS. Hum. Mol. Genet. *10*, 467–476.

Bock, C. (2012). Analysing and interpreting DNA methylation data. Nat. Rev. Genet. *13*, 705–719.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics *21*, 4067–4068.

Bock, C., Beerman, I., Lien, W.-H., Smith, Z.D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D.J., and Meissner, A. (2012). DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells. Mol. Cell *47*, 633–647.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science *317*, 1760–1764.

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

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Names, A., Temper, V., Razin, A., and Cedar, H. (1994). Spl elements protect a CpG island from *de novo* methylation. Nature *371*, 435–438.

Brinkman, J., and El-Ashry, D. (2009). ER Re-expression and Re-sensitization to Endocrine Therapies in ER-negative Breast Cancers. J. Mammary Gland Biol. Neoplasia *14*, 67–78.

Busslinger, M., Hurst, J., and Flavell, R.A. (1983). DNA methylation and the regulation of globin gene expression. Cell *34*, 197–206.

Cadieux, B., Ching, T.-T., VandenBerg, S.R., and Costello, J.F. (2006). Genome-wide Hypomethylation in Human Glioblastomas Associated with Specific Copy Number Alteration, Methylenetetrahydrofolate Reductase Allele Status, and Increased Proliferation. Cancer Res. *66*, 8469–8476.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a mechanism for the regulation of E2F activity. Proc. Natl. Acad. Sci. U. S. A. *97*, 6481–6486.

Carascossa, S., Dudek, P., Cenni, B., Briand, P.-A., and Picard, D. (2010). CARM1 mediates the ligand-independent and tamoxifen-resistant activation of the estrogen receptor α by cAMP. Genes Dev. *24*, 708–719.

Charlet, J., Duymich, C.E., Lay, F.D., Mundbjerg, K., Dalsgaard Sørensen, K., Liang, G., and Jones, P.A. (2016). Bivalent Regions of Cytosine Methylation and H3K27 Acetylation Suggest an Active Role for DNA Methylation at Enhancers. Mol. Cell *62*, 422–431.

Chédin, F., Lieber, M.R., and Hsieh, C.-L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. Proc. Natl. Acad. Sci. U. S. A. *99*, 16916–16921.

Chen, D., Pace, P.E., Coombes, R.C., and Ali, S. (1999). Phosphorylation of Human Estrogen Receptor α by Protein Kinase A Regulates Dimerization. Mol. Cell. Biol. *19*, 1002–1015.

Chen, H., Kazemier, H.G., Groote, M.L. de, Ruiters, M.H.J., Xu, G.-L., and Rots, M.G. (2013). Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. Nucleic Acids Res. gkt1019.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. Mol. Cell. Biol. *23*, 5594–5605.

Chen, T., Tsujimoto, N., and Li, E. (2004). The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin. Mol. Cell. Biol. *24*, 9048–9058.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat. Genet. *39*, 391–396.

Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., *et al.* (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell *162*, 974–986.

Choy, M.-K., Movassagh, M., Goh, H.-G., Bennett, M.R., Down, T.A., and Foo, R.S. (2010). Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC Genomics *11*, 519.

Chuang, L.S.-H., Ian, H.-I., Koh, T.-W., Ng, H.-H., Xu, G., and Li, B.F.L. (1997). Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for $p21^{WAF1}$. Science *277*, 1996–2000.

Cihák, A. (1974). Biological effects of 5-azacytidine in eukaryotes. Oncology *30*, 405–422.

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. Nat. Genet. *45*, 1127–1133.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Collins, L.C., Botero, M.L., and Schnitt, S.J. (2005). Bimodal Frequency Distribution of Estrogen Receptor Immunohistochemical Staining Results in Breast Cancer An Analysis of 825 Cases. Am. J. Clin. Pathol. *123*, 16–20.

Coulondre, C., Miller, J.H., Farabaugh, P.J., and Gilbert, W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. Nature *274*, 775–780.

Cramer, J.M., Scarsdale, J.N., Walavalkar, N.M., Buchwald, W.A., Ginder, G.D., and Williams, D.C. (2014). Probing the Dynamic Distribution of Bound States for Methylcytosine-binding Domains on DNA. J. Biol. Chem. *289*, 1294–1302.

Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I., *et al.* (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc. Natl. Acad. Sci. U. S. A. *106*, 13820–13825.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist Rna, DNA Methylation, and Histone Hypoacetylation in Maintaining X Chromosome Inactivation. J. Cell Biol. *153*, 773–784.

Cui, C., Gan, Y., Gu, L., Wilson, J., Liu, Z., Zhang, B., and Deng, D. (2015). P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. Genome Biol. *16*, 252.

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., and Feinberg, A.P. (2003). Loss of IGF2 Imprinting: A Potential Marker of Colorectal Cancer Risk. Science *299*, 1753–1755.

Cui, X., Schiff, R., Arpino, G., Osborne, C.K., and Lee, A.V. (2005). Biology of Progesterone Receptor Loss in Breast Cancer and Its Implications for Endocrine Therapy. J. Clin. Oncol. *23*, 7721–7735.

Cunha, S., Lin, Y.-C., Goossen, E.A., DeVette, C.I., Albertella, M.R., Thomson, S., Mulvihill, M.J., and Welm, A.L. (2014). The RON Receptor Tyrosine Kinase Promotes Metastasis by Triggering MBD4-Dependent DNA Methylation Reprogramming. Cell Rep. *6*, 141–154.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. *32*, 1262–1267.

Dowsett, M., Cuzick, J., Ingle, J., Coates, A., Forbes, J., Bliss, J., Buyse, M., Baum, M., Buzdar, A., Colleoni, M., *et al.* (2010). Meta-Analysis of Breast Cancer Outcomes in Adjuvant Trials of Aromatase Inhibitors Versus Tamoxifen. J. Clin. Oncol. *28*, 509–518.

Du, Q., Luu, P.-L., Stirzaker, C., and Clark, S.J. (2015). Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics *7*, 1051–1073.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. The Lancet *378*, 771–784.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719–2729.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., *et al.* (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. *38*, 1378–1385.

Ede, C., Chen, X., Lin, M.-Y., and Chen, Y.Y. (2016). Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. ACS Synth. Biol. *5*, 395–404.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal Instability and Tumors Promoted by DNA Hypomethylation. Science *300*, 455–455.

Edwards, J.R., O'Donnell, A.H., Rollins, R.A., Peckham, H.E., Lee, C., Milekic, M.H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., *et al.* (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res. *20*, 972–980.

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

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature *429*, 457–463.

Ehrlich, M. (2009). DNA hypomethylation in cancer cells. Epigenomics *1*, 239–259.

Ehrlich, M., and Lacey, M. (2013). DNA Hypomethylation and Hemimethylation in Cancer. In Epigenetic Alterations in Oncogenesis, A.R. Karpf, ed. (Springer New York), pp. 31–56.

Ellis, M.J., Tao, Y., Luo, J., A'Hern, R., Evans, D.B., Bhatnagar, A.S., Chaudri Ross, H.A., von Kameke, A., Miller, W.R., Smith, I., *et al.* (2008). Outcome Prediction for Estrogen Receptor–Positive Breast Cancer Based on Postneoadjuvant Endocrine Therapy Tumor Characteristics. JNCI J. Natl. Cancer Inst. *100*, 1380–1388.

Ellis, M.J., Suman, V.J., Hoog, J., Lin, L., Snider, J., Prat, A., Parker, J.S., Luo, J., DeSchryver, K., Allred, D.C., *et al.* (2011). Randomized Phase II Neoadjuvant Comparison Between Letrozole, Anastrozole, and Exemestane for Postmenopausal Women With Estrogen Receptor–Rich Stage 2 to 3 Breast Cancer: Clinical and Biomarker Outcomes and Predictive Value of the Baseline PAM50-Based Intrinsic Subtype—ACOSOG Z1031. J. Clin. Oncol. *29*, 2342–2349.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., Tine, B.A.V., Hoog, J., Goiffon, R.J., Goldstein, T.C., *et al.* (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature *486*, 353–360.

Emperle, M., Rajavelu, A., Reinhardt, R., Jurkowska, R.Z., and Jeltsch, A. (2014). Cooperative DNA Binding and Protein/DNA Fiber Formation Increases the Activity of the Dnmt3a DNA Methyltransferase. J. Biol. Chem. *289*, 29602–29613.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat. Rev. Mol. Cell Biol. *17*, 756–770.

Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiewicz, D., Bates, B., Lee, P.P., Kuhn, R., *et al.* (2001). DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. J. Neurosci. Off. J. Soc. Neurosci. *21*, 788–797.

Fan, M., Yan, P.S., Hartman-Frey, C., Chen, L., Paik, H., Oyer, S.L., Salisbury, J.D., Cheng, A.S.L., Li, L., Abbosh, P.H., *et al.* (2006). Diverse Gene Expression and DNA Methylation Profiles Correlate with Differential Adaptation of Breast Cancer Cells to the Antiestrogens Tamoxifen and Fulvestrant. Cancer Res. *66*, 11954–11966.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., *et al.* (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. *41*, 563–571.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. Nat. Rev. Cancer *4*, 143–153.

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

Ferguson, A.T., Lapidus, R.G., Baylin, S.B., and Davidson, N.E. (1995). Demethylation of the Estrogen Receptor Gene in Estrogen Receptor-negative Breast Cancer Cells Can Reactivate Estrogen Receptor Gene Expression. Cancer Res. *55*, 2279–2283.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., *et al.* (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. *43*, D805–D811.

Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., and Esteller, M. (2003). The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res. *31*, 1765–1774.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F., *et al.* (2011). Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1. PLoS ONE *6*, e16627.

Fritah, A., Saucier, C., De Wever, O., Bracke, M., Bièche, I., Lidereau, R., Gespach, C., Drouot, S., Redeuilh, G., and Sabbah, M. (2008). Role of WISP-2/CCN5 in the Maintenance of a Differentiated and Noninvasive Phenotype in Human Breast Cancer Cells. Mol. Cell. Biol. *28*, 1114–1123.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *advance online publication*.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003). Methyl-CpG Binding Domain 1 (MBD1) Interacts with the Suv39h1-HP1

Heterochromatic Complex for DNA Methylation-based Transcriptional Repression. J. Biol. Chem. *278*, 24132–24138.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. *31*, 2305–2312.

Fulton, A.M., and Heppner, G.H. (1985). Relationships of Prostaglandin E and Natural Killer Sensitivity to Metastatic Potential in Murine Mammary Adenocarcinomas. Cancer Res. *45*, 4779–4784.

Fuqua, S. a. W., Wiltschke, C., Castles, C., Wolf, D., and Allred, D.C. (1995). A role for estrogen-receptor variants in endocrine resistance. Endocr. Relat. Cancer *2*, 19–25.

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

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. *109*, E2579–E2586.

Gaston, K., and Fried, M. (1995). CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Res. *23*, 901–909.

Ge, Y.-Z., Pu, M.-T., Gowher, H., Wu, H.-P., Ding, J.-P., Jeltsch, A., and Xu, G.-L. (2004). Chromatin Targeting of *de novo* DNA Methyltransferases by the PWWP Domain. J. Biol. Chem. *279*, 25447–25454.

Gibb, E.A., Brown, C.J., and Lam, W.L. (2011). The functional role of long non-coding RNA in human carcinomas. Mol. Cancer *10*, 38.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., *et al.* (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. PLOS ONE *5*, e15367.

Gonzalez-Zulueta, M., Bender, C.M., Yang, A.S., Nguyen, T., Beart, R.W., Van Tornout, J.M., and Jones, P.A. (1995). Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. Cancer Res. *55*, 4531–4535.

Gowher, H., and Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites1. J. Mol. Biol. *309*, 1201–1208.

Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J. Biol. Chem. *280*, 13341–13348.

Gowher, H., Loutchanwoot, P., Vorobjeva, O., Handa, V., Jurkowska, R.Z., Jurkowski, T.P., and Jeltsch, A. (2006). Mutational Analysis of the Catalytic Domain of the Murine Dnmt3a DNA-(cytosine C5)-methyltransferase. J. Mol. Biol. *357*, 928–941.

Graff, J.R., Herman, J.G., Lapidus, R.G., Chopra, H., Xu, R., Jarrard, D.F., Isaacs, W.B., Pitha, P.M., Davidson, N.E., and Baylin, S.B. (1995). E-Cadherin Expression Is Silenced by DNA Hypermethylation in Human Breast and Prostate Carcinomas. Cancer Res. *55*, 5195–5199.

Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature *320*, 134–139.

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

Groote, M.L. de, Verschure, P.J., and Rots, M.G. (2012). Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. Nucleic Acids Res. *40*, 10596–10613.

Guezennec, X.L., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A.M., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different Biochemical and Functional Properties. Mol. Cell. Biol. *26*, 843–851.

Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. *32*, 577–582.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., *et al.* (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. *17*, 215–222.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., *et al.* (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature *517*, 640–644.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

Han, H., Cortez, C.C., Yang, X., Nichols, P.W., Jones, P.A., and Liang, G. (2011). DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. Hum. Mol. Genet.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., *et al.* (2011). Increased methylation variation in epigenetic domains across cancer types. Nat. Genet. *43*, 768–775.

Harvey, J.M., Clark, G.M., Osborne, C.K., and Allred, D.C. (1999). Estrogen Receptor Status by Immunohistochemistry Is Superior to the Ligand-Binding Assay for Predicting Response to Adjuvant Endocrine Therapy in Breast Cancer. J. Clin. Oncol. *17*, 1474–1474.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

Hassler, M.R., Klisaroska, A., Kollmann, K., Steiner, I., Bilban, M., Schiefer, A.-I., Sexl, V., and Egger, G. (2012). Antineoplastic activity of the DNA methyltransferase inhibitor 5-aza-2′-deoxycytidine in anaplastic large cell lymphoma. Biochimie *94*, 2297–2307.

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

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al.* (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science *333*, 1303–1307.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. *8*, R19.

Hendrich, B., Hardeland, U., Ng, H.-H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301–304.

Herman, J.G., Latif, F., Weng, Y., Lerman, M.I., Zbar, B., Liu, S., Samid, D., Duan, D.S., Gnarra, J.R., and Linehan, W.M. (1994). Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc. Natl. Acad. Sci. U. S. A. *91*, 9700–9704.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites. J. Biol. Chem. *279*, 48350–48359.

Herynk, M.H., Parra, I., Cui, Y., Beyer, A., Wu, M.-F., Hilsenbeck, S.G., and Fuqua, S.A.W. (2007). Association between the Estrogen Receptor α A908G Mutation and Outcomes in Invasive Breast Cancer. Clin. Cancer Res. *13*, 3235–3243.

Hiken, J.F., McDonald, J.I., Decker, K.F., Sanchez, C., Hoog, J., VanderKraats, N.D., Jung, K.L., Akinhanmi, M., Rois, L.E., Ellis, M.J., *et al.* (2017). Epigenetic activation of the prostaglandin receptor EP4 promotes resistance to endocrine therapy for breast cancer. Oncogene *36*, 2319–2327.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *advance online publication*.

Hinshelwood, R.A., Melki, J.R., Huschtscha, L.I., Paul, C., Song, J.Z., Stirzaker, C., Reddel, R.R., and Clark, S.J. (2009). Aberrant *de novo* methylation of the p16INK4A CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. Hum. Mol. Genet. *18*, 3098–3109.

Holemon, H., Korshunova, Y., Ordway, J.M., Bedell, J.A., Citek, R.W., Lakey, N., Leon, J., Finney, M., McPherson, J.D., and Jeddeloh, J.A. (2007). MethylScreen: DNA methylation density monitoring using quantitative PCR. BioTechniques *43*, 683–693.

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

Holz-Schietinger, C., and Reich, N.O. (2010). The Inherent Processivity of the Human *de novo* Methyltransferase 3A (DNMT3A) Is Enhanced by DNMT3L. J. Biol. Chem. *285*, 29091–29100.

Holz-Schietinger, C., Matje, D.M., Harrison, M.F., and Reich, N.O. (2011). Oligomerization of DNMT3A Controls the Mechanism of *de novo* DNA Methylation. J. Biol. Chem. *286*, 41479–41488.

Hotchkiss, R.D. (1948). The Quantitative Separation of Purines, Pyrimidines, and Nucleosides by Paper Chromatography. J. Biol. Chem. *175*, 315–332.

Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene *27*, 404–408.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., *et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P., and Xu, Y. (2013a). Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. Cell *155*, 1545–1555.

Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., Lou, Z., Zhang, L., Li, W., Gong, W., *et al.* (2015). Structural insight into substrate preference for TET-mediated oxidation. Nature *527*, 118–122.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., *et al.* (2013b). DNA methylation presents distinct binding sites for human transcription factors. ELife *2*.

Huang, J., Zhou, N., Watabe, K., Lu, Z., Wu, F., Xu, M., and Mo, Y.-Y. (2014). Long non-coding RNA UCA1 promotes breast tumor growth by suppression of p27 (Kip1). Cell Death Dis. *5*, e1008.

Hur, K., Cejas, P., Feliu, J., Moreno-Rubio, J., Burgos, E., Boland, C.R., and Goel, A. (2013). Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut *Published Online First: 23-05-2013. doi: 10.1136/gutjnl-2012-304219*.

Hurtado, A., Holmes, K.A., Geistlinger, T.R., Hutcheson, I.R., Nicholson, R.I., Brown, M., Jiang, J., Howat, W.J., Ali, S., and Carroll, J.S. (2008). Regulation of ERBB2 by oestrogen receptor–PAX2 determines response to tamoxifen. Nature *456*, 663–666.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands – 'A rough guide.' FEBS Lett. *583*, 1713–1720.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. PLOS Genet. *6*, e1001134.

Inoue, S., and Oishi, M. (2005). Effects of methylation of non-CpG sequence in the promoter region on the expression of human synaptotagmin XI (syt11). Gene *348*, 123–134.

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. *21*, 1670–1676.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., *et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

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.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300–1303.

Iwase, H., Omoto, Y., Iwata, H., Toyama, T., Hara, Y., Ando, Y., Ito, Y., Fujii, Y., and Kobayashi, S. (1999). DNA methylation analysis at distal and proximal promoter regions of the oestrogen receptor gene in breast cancers. Br. J. Cancer *80*, 1982–1986.

Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle Georget. Tex *8*, 1698–1710.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., *et al.* (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. *27*, 31–39.

Jeltsch, A. (2008). Reading and writing DNA methylation. Nat. Struct. Mol. Biol. *15*, 1003–1004.

Jeltsch, A., and Jurkowska, R.Z. (2014). New concepts in DNA methylation. Trends Biochem. Sci. *39*, 310–318.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. Nature *449*, 248–251.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Johnston, S.R.D., and Dowsett, M. (2003). Aromatase inhibitors for breast cancer: lessons from the laboratory. Nat. Rev. Cancer *3*, 821–831.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jones, P.A., and Baylin, S.B. (2007). The Epigenomics of Cancer. Cell *128*, 683–692.

Jones, P.A., and Liang, G. (2009). Rethinking how DNA Methylation Patterns are Maintained. Nat. Rev. Genet. *10*, 805–811.

Jones, P.A., Issa, J.-P.J., and Baylin, S. (2016). Targeting the cancer epigenome for therapy. Nat. Rev. Genet. *17*, 630–641.

Jones, P.L., Jan Veenstra, G.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat. Genet. *19*, 187–191.

Jurkowska, R.Z., and Jeltsch, A. (2016). Enzymology of Mammalian DNA Methyltransferases. In DNA Methyltransferases - Role and Function, A. Jeltsch, and R.Z. Jurkowska, eds. (Springer International Publishing), pp. 87–122.

Jurkowska, R.Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., and Jeltsch, A. (2008). Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res. *36*, 6656–6663.

Jurkowska, R.Z., Rajavelu, A., Anspach, N., Urbanke, C., Jankevicius, G., Ragozin, S., Nellen, W., and Jeltsch, A. (2011). Oligomerization and Binding of the Dnmt3a DNA Methyltransferase to Parallel DNA Molecules. J. Biol. Chem. *286*, 24200–24207.

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

Kareta, M.S., Botello, Z.M., Ennis, J.J., Chou, C., and Chédin, F. (2006). Reconstitution and mechanism of the stimulation of *de novo* methylation by human DNMT3L. J. Biol. Chem. *281*, 25893–25902.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods *12*, 401–403.

Kelly, T.K., Miranda, T.B., Liang, G., Berman, B.P., Lin, J.C., Tanay, A., and Jones, P.A. (2010). H2A.Z Maintenance During Mitosis Reveals Nucleosome Shifting on Mitotically Silenced Genes. Mol. Cell *39*, 901–911.

Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R.W., and Bérubé, N.G. (2010). ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. Dev. Cell *18*, 191–202.

Kernohan, K.D., Vernimmen, D., Gloor, G.B., and Bérubé, N.G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic Acids Res. *42*, 8356–8368.

Kim, J.H., Kang, S., Kim, T.W., Yin, L., Liu, R., and Kim, S.J. (2012). Expression Profiling after Induction of Demethylation in MCF-7 Breast Cancer Cells Identifies Involvement of TNF-α Mediated Cancer Pathways. Mol. Cells *33*, 127–133.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. *33*, 1293–1298.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). Hhal methyltransferase flips its target base out of the DNA helix. Cell *76*, 357–369.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Äijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., *et al.* (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature *497*, 122–126.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20–36.

Konkel, M.K., and Batzer, M.A. (2010). A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin. Cancer Biol. *20*, 211–221.

Krebs, A., Dessus-Babus, S., Burger, L., and Schübeler, D. (2014). High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. ELife e04094.

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

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

Kundu, N., Ma, X., Holt, D., Goloubeva, O., Ostrand-Rosenberg, S., and Fulton, A.M. (2009). Antagonism of the prostaglandin E receptor EP4 inhibits metastasis and enhances NK function. Breast Cancer Res. Treat. *117*, 235–242.

Kungulovski, G., Nunna, S., Thomas, M., Zanger, U.M., Reinhardt, R., and Jeltsch, A. (2015). Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. Epigenetics Chromatin *8*, 12.

Lai, A.Y., and Wade, P.A. (2011). Cancer biology and NuRD: a multifaceted chromatin remodelling complex. Nat. Rev. Cancer *11*, 588–596.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M.A., Richter, J., Soler, E., *et al.* (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat. Med. *16*, 571–579.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lapidus, R.G., Ferguson, A.T., Ottaviano, Y.L., Parl, F.F., Smith, H.S., Weitzman, S.A., Baylin, S.B., Issa, J.P., and Davidson, N.E. (1996). Methylation of estrogen and progesterone receptor gene 5' CpG islands correlates with lack of estrogen and progesterone receptor gene expression in breast tumors. Clin. Cancer Res. *2*, 805–810.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., *et al.* (2010). Dynamic changes in the human methylome during differentiation. Genome Res. *20*, 320–331.

Lazennec, G., Thomas, J.A., and Katzenellenbogen, B.S. (2001). Involvement of cyclic AMP response element binding protein (CREB) and estrogen receptor phosphorylation in the synergistic activation of the estrogen receptor by estradiol and protein kinase activators. J. Steroid Biochem. Mol. Biol. *77*, 193–203.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., *et al.* (2012). Landscape of Somatic Retrotransposition in Human Cancers. Science *337*, 967–971.

Li, E., and Zhang, Y. (2014). DNA Methylation in Mammals. Cold Spring Harb. Perspect. Biol. *6*, a019133.

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

Li, H., Rauch, T., Chen, Z.-X., Szabó, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. J. Biol. Chem. *281*, 19489–19500.

Li, S., Shen, D., Shao, J., Crowder, R., Liu, W., Prat, A., He, X., Liu, S., Hoog, J., Lu, C., *et al.* (2013). Endocrine-Therapy-Resistant ESR1 Variants Revealed by Genomic Characterization of Breast-Cancer-Derived Xenografts. Cell Rep. *4*.

Li, X., Wu, Y., Liu, A., and Tang, X. (2016). Long non-coding RNA UCA1 enhances tamoxifen resistance in breast cancer cells through a miR-18a-HIF1α feedback regulatory loop. Tumor Biol. *37*, 14733–14743.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. Mol. Cell. Biol. *22*, 480–491.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., *et al.* (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat. Genet. *47*, 469–478.

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.

Liggett, W.H., and Sidransky, D. (1998). Role of the p16 tumor suppressor gene in cancer. J. Clin. Oncol. *16*, 1197–1206.

Lin, X., and Nelson, W.G. (2003). Methyl-CpG-binding Domain Protein-2 Mediates Transcriptional Repression Associated with Hypermethylated ***GSTP1*** CpG Islands in MCF-7 Breast Cancer Cells. Cancer Res. *63*, 498–504.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., *et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., *et al.* (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., *et al.* (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. Science *341*.

Liu, H., Wang, G., Yang, L., Qu, J., Yang, Z., and Zhou, X. (2016a). Knockdown of Long Non-Coding RNA UCA1 Increases the Tamoxifen Sensitivity of Breast Cancer Cells through Inhibition of Wnt/β-Catenin Pathway. PLOS ONE *11*, e0168406.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016b). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233–247.e17.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat. Struct. Mol. Biol. *11*, 1068–1075.

Lou, S., Lee, H.-M., Qin, H., Li, J.-W., Gao, Z., Liu, X., Chan, L.L., KL Lam, V., So, W.-Y., Wang, Y., *et al.* (2014). Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. Genome Biol. *15*.

Lukashevich, O.V., Cherepanova, N.A., Jurkovska, R.Z., Jeltsch, A., and Gromova, E.S. (2016). Conserved motif VIII of murine DNA methyltransferase Dnmt3a is essential for methylation activity. BMC Biochem. *17*, 7.

Lund, K., Cole, J.J., VanderKraats, N.D., McBryan, T., Pchelintsev, N.A., Clark, W., Copland, M., Edwards, J.R., and Adams, P.D. (2014). DNMT inhibitors reverse a specific signature of aberrant promoter DNA methylation and associated gene silencing in AML. Genome Biol. *15*, 406.

Luo, G.-Z., Blanco, M.A., Greer, E.L., He, C., and Shi, Y. (2015). DNA N6-methyladenine: a new epigenetic mark in eukaryotes? Nat. Rev. Mol. Cell Biol. *16*, 705–710.

Ma, C.X., Reinert, T., Chmielewska, I., and Ellis, M.J. (2015). Mechanisms of aromatase inhibitor resistance. Nat. Rev. Cancer *15*, 261–275.

Ma, X., Kundu, N., Rifat, S., Walser, T., and Fulton, A.M. (2006). Prostaglandin E Receptor EP4 Antagonism Inhibits Breast Cancer Metastasis. Cancer Res. *66*, 2923–2927.

Ma, X., Kundu, N., Collin, P.D., Goloubeva, O., and Fulton, A. (2012). Frondoside A inhibits breast cancer metastasis and antagonizes prostaglandin E receptors EP4 and EP2. Breast Cancer Res. Treat. *132*, 1001–1008.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. *8*, 2282–2292.

Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., *et al.* (2013a). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. *advance online publication*.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013b). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods *10*, 977–979.

Magdinier, F., and Wolffe, A.P. (2001). Selective association of the methyl-CpG binding protein MBD2 with the silent p14/p16 locus in human neoplasia. Proc. Natl. Acad. Sci. *98*, 4990–4995.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. J. Biol. Chem. *286*, 35334–35338.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., *et al.* (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253–257.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Embryogenesis: Demethylation of the zygotic paternal genome. Nature *403*, 501–502.

McDonald, J.I., Celik, H., Rois, L.E., Fishberger, G., Fowler, T., Rees, R., Kramer, A., Martens, A., Edwards, J.R., and Challen, G.A. (2016). Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol. Open bio.019067.

Medvedeva, Y.A., Khamis, A.M., Kulakovskiy, I.V., Ba-Alawi, W., Bhuyan, M.S.I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A.R., Bajic, V.B., *et al.* (2014). Effects of cytosine methylation on transcription factor binding sites. BMC Genomics *15*, 119.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499–507.

Mendoza, M.C., Er, E.E., and Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. Trends Biochem. Sci. *36*, 320–328.

Merenbakh-Lamin, K., Ben-Baruch, N., Yeheskel, A., Dvir, A., Soussan-Gutman, L., Jeselsohn, R., Yelensky, R., Brown, M., Miller, V.A., Sarid, D., *et al.* (2013). D538G Mutation in Estrogen Receptor-α: A Novel Mechanism for Acquired Endocrine Resistance in Breast Cancer. Cancer Res. *73*, 6856–6864.

Michaud, E.J., Vugt, M.J. van, Bultman, S.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. (1994). Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes Dev. *8*, 1463–1472.

Milicic, A., Harrison, L.-A., Goodlad, R.A., Hardy, R.G., Nicholson, A.M., Presz, M., Sieber, O., Santander, S., Pringle, J.H., Mandir, N., *et al.* (2008). Ectopic expression of P-Cadherin correlates with promoter hypomethylation early in colorectal carcinogenesis and enhanced intestinal crypt fission in mice. Cancer Res. *68*, 7760–7768.

Miller, T.W., Balko, J.M., Ghazoui, Z., Dunbier, A., Anderson, H., Dowsett, M., González-Angulo, A.M., Mills, G.B., Miller, W.R., Wu, H., *et al.* (2011). A Gene Expression Signature from Human Breast Cancer Cells with Acquired Hormone Independence Identifies MYC as a Mediator of Antiestrogen Resistance. Clin. Cancer Res. *17*, 2024–2034.

Miyakura, Y., Sugano, K., Akasu, T., Yoshida, T., Maekawa, M., Saitoh, S., Sasaki, H., Nomizu, T., Konishi, F., Fujita, S., *et al.* (2004). Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. Clin. Gastroenterol. Hepatol. *2*, 147–156.

Mohandas, T., Sparkes, R.S., and Shapiro, L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393–396.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. J. Mol. Evol. *60*, 174–182.

Morgan, H.D., Sutherland, H.G.E., Martin, D.I.K., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nat. Genet. *23*, 314–318.

Mosselman, S., Polman, J., and Dijkema, R. (1996). ERβ: Identification and characterization of a novel human estrogen receptor. FEBS Lett. *392*, 49–53.

Musgrove, E.A., and Sutherland, R.L. (2009). Biological determinants of endocrine resistance in breast cancer. Nat. Rev. Cancer *9*, 631–643.

Myöhänen, S.K., Baylin, S.B., and Herman, J.G. (1998). Hypermethylation Can Selectively Silence Individual *p16^{ink4A}* Alleles in Neoplasia. Cancer Res. *58*, 591–593.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat. Chem. Biol. *8*, 751–758.

Nan, X., Ng, H.-H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–389.

Nass, S.J., Herman, J.G., Gabrielson, E., Iversen, P.W., Parl, F.F., Davidson, N.E., and Graff, J.R. (2000). Aberrant Methylation of the Estrogen Receptor and E-Cadherin 5′ CpG Islands Increases with Malignant Progression in Human Breast Cancer. Cancer Res. *60*, 4346–4348.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77.

Norris, D.P., Brockdorff, N., and Rastan, S. (1991). Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. Mamm. Genome Off. J. Int. Mamm. Genome Soc. *1*, 78–83.

Norvil, A.B., Petell, C.J., Alabdi, L., Wu, L., Rossie, S., and Gowher, H. (2016). Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface. Biochemistry (Mosc.).

Nunna, S., Reinhardt, R., Ragozin, S., and Jeltsch, A. (2014). Targeted Methylation of the Epithelial Cell Adhesion Molecule (EpCAM) Promoter to Silence Its Expression in Ovarian Cancer Cells. PLOS ONE *9*, e87703.

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

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for *De novo* Methylation and Mammalian Development. Cell *99*, 247–257.

Olson, J.A., Budd, G.T., Carey, L.A., Harris, L.A., Esserman, L.J., Fleming, G.F., Marcom, P.K., Leight, G.S., Giuntoli, T., Commean, P., *et al.* (2009). Improved Surgical Outcomes for Breast Cancer Patients Receiving Neoadjuvant Aromatase Inhibitor Therapy: Results from a Multicenter Phase II Trial. J. Am. Coll. Surg. *208*, 906–916.

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.

Ortmann, C.A., Eisele, L., Nückel, H., Klein-Hitpass, L., Führer, A., Dührsen, U., and Zeschnigk, M. (2008). Aberrant hypomethylation of the cancer-testis antigen PRAME correlates with PRAME expression in acute myeloid leukemia. Ann. Hematol. *87*, 809–818.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr. Biol. *10*, 475–478.

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

Ottaviano, Y.L., Issa, J.-P., Parl, F.F., Smith, H.S., Baylin, S.B., and Davidson, N.E. (1994). Methylation of the Estrogen Receptor Gene CpG Island Marks Loss of Estrogen Receptor Expression in Human Breast Cancer Cells. Cancer Res. *54*, 2552–2555.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., *et al.* (2013a). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods *10*, 973–976.

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013b). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat. Methods *10*, 239–242.

Perou, C.M., Sørlie, T., Eisen, M.B., Rijn, M. van de, Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet. *15*, 517–530.

Plass, C., Pfister, S.M., Lindroth, A.M., Bogatyrova, O., Claus, R., and Lichter, P. (2013). Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. Nat. Rev. Genet. *14*, 765–780.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., Cruz, C.C. de la, Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of Histone H3 Lysine 27 Methylation in X Inactivation. Science *300*, 131–135.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653–663.

Prébet, T., Gore, S.D., Esterni, B., Gardin, C., Itzykson, R., Thepot, S., Dreyfus, F., Rauzy, O.B., Recher, C., Adès, L., *et al.* (2011). Outcome of High-Risk Myelodysplastic Syndrome After Azacitidine Treatment Failure. J. Clin. Oncol. *29*, 3322–3327.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Qin, T., Castoro, R., Ahdab, S.E., Jelinek, J., Wang, X., Si, J., Shu, J., He, R., Zhang, N., Chung, W., *et al.* (2011). Mechanisms of Resistance to Decitabine in the Myelodysplastic Syndrome. PLOS ONE *6*, e23372.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rajavelu, A., Jurkowska, R.Z., Fritz, J., and Jeltsch, A. (2012). Function and disruption of DNA Methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. Nucleic Acids Res. *40*, 569–580.

Ramchandani, S., Bigey, P., and Szyf, M. (1998). Genomic structure of the human DNA methyltransferase gene. Biol. Chem. *379*, 535–540.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. *97*, 5237–5242.

Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. *30*, 733–750.

Redis, R.S., Siewerts, A.M., Look, M.P., Tudoran, O., Ivan, C., Spizzo, R., Zhang, X., de Weerd, V., Shimizu, M., Ling, H., *et al.* (2013). CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. Oncotarget *4*.

Reither, S., Li, F., Gowher, H., and Jeltsch, A. (2003). Catalytic Mechanism of DNA-(cytosine-C5)-methyltransferases Revisited: Covalent Intermediate Formation is not Essential for Methyl Group Transfer by the Murine Dnmt3a Enzyme. J. Mol. Biol. *329*, 675–684.

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

Rivenbark, A.G., Stolzenburg, S., Beltran, A.S., Yuan, X., Rots, M.G., Strahl, B.D., and Blancafort, P. (2012). Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics *7*, 350–360.

Robinson, D.R., Wu, Y.-M., Vats, P., Su, F., Lonigro, R.J., Cao, X., Kalyana-Sundaram, S., Wang, R., Ning, Y., Hodges, L., *et al.* (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nat. Genet. *45*, 1446–1451.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roodi, N., Bailey, L.R., Kao, W.-Y., Verrier, C.S., Yee, C.J., Dupont, W.D., and Parl, F.F. (1995). Estrogen Receptor Gene Analysis in Estrogen Receptor-Positive and Receptor-Negative Primary Breast Cancer. J. Natl. Cancer Inst. *87*, 446–451.

Saito, M., and Ishikawa, F. (2002). The mCpG-binding Domain of Human MBD3 Does Not Bind to mCpG but Interacts with NuRD/Mi2 Components HDAC1 and MTA2. J. Biol. Chem. *277*, 35434–35439.

Sakai, T., Toguchida, J., Ohtani, N., Yandell, D.W., Rapaport, J.M., and Dryja, T.P. (1991). Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. Am. J. Hum. Genet. *48*, 880–888.

Sanchez, C.G., Ma, C.X., Crowder, R.J., Guintoli, T., Phommaly, C., Gao, F., Lin, L., and Ellis, M.J. (2011). Preclinical modeling of combined phosphatidylinositol-3-kinase inhibition with endocrine therapy for estrogen receptor-positive breast cancer. Breast Cancer Res. BCR *13*, R21.

Sarda, S., Das, A., Vinson, C., and Hannenhalli, S. (2017). Distal CpG islands can serve as alternative promoters to transcribe genes with silenced proximal promoters. Genome Res. *27*, 553–566.

Schaefer, K.A., Wu, W.-H., Colgan, D.F., Tsang, S.H., Bassuk, A.G., and Mahajan, V.B. (2017). Unexpected mutations after CRISPR-Cas9 editing *in vivo*. Nat. Methods *14*, 547–548.

Schiff, R., Selever, J., and Fuqua, S.A.W. (2008). The Importance of Estrogen Receptors in Breast Cancer. In Breast Cancer: Prognosis, Treatment, and Prevention, (Informa Health Care), pp. 83–111.

Schlosberg, C.E., VanderKraats, N.D., and Edwards, J.R. (2017). Modeling complex patterns of differential DNA methylation that associate with gene expression changes. Nucleic Acids Res. *45*, 5100–5111.

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

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., *et al.* (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212–216.

Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L., and Khavari, P.A. (2010). DNMT1 Maintains Progenitor Function in Self-Renewing Somatic Tissue. Nature *463*, 563–567.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., *et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912.

Shima, K., Nosho, K., Baba, Y., Cantor, M., Meyerhardt, J.A., Giovannucci, E.L., Fuchs, C.S., and Ogino, S. (2011). Prognostic significance of CDKN2A (p16) promoter methylation and loss of expression in 902 colorectal cancers: Cohort study and literature review. Int. J. Cancer *128*, 1080–1094.

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. PLOS Genet. *9*, e1003439.

Shirohzu, H., Kubota, T., Kumazawa, A., Sado, T., Chijiwa, T., Inagaki, K., Suetake, I., Tajima, S., Wakui, K., Miki, Y., *et al.* (2002). Three novel DNMT3B mutations in Japanese patients with ICF syndrome. Am. J. Med. Genet. *112*, 31–37.

Siddique, A.N., Nunna, S., Rajavelu, A., Zhang, Y., Jurkowska, R.Z., Reinhardt, R., Rots, M.G., Ragozin, S., Jurkowski, T.P., and Jeltsch, A. (2013). Targeted Methylation and Gene Silencing of VEGF-A in Human Cells by Using a Designed Dnmt3a–Dnmt3L Single-Chain Fusion Protein with Increased DNA Methylation Activity. J. Mol. Biol. *425*, 479–491.

Sims, A.H., Clarke, R.B., Howell, A., and Howell, S.J. (2008). The Cellular Origins of Breast Cancer Subtypes. In Breast Cancer: Prognosis, Treatment, and Prevention, (Informa Health Care), pp. 71–81.

Singleton, M.K., Gonzales, M.L., Leung, K.N., Yasui, D.H., Schroeder, D.I., Dunaway, K., and LaSalle, J.M. (2011). MeCP2 is required for global heterochromatic and nucleolar changes during activity-dependent neuronal maturation. Neurobiol. Dis. *43*, 190–200.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Smith, A.E., Hurd, P.J., Bannister, A.J., Kouzarides, T., and Ford, K.G. (2008). Heritable Gene Repression through the Action of a Directed DNA Methyltransferase at a Chromosomal Locus. J. Biol. Chem. *283*, 9878–9885.

Snowden, A.W., Gregory, P.D., Case, C.C., and Pabo, C.O. (2002). Gene-Specific Targeting of H3K9 Methylation Is Sufficient for Initiating Repression *In vivo*. Curr. Biol. *12*, 2159–2166.

Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. Science *331*, 1036–1040.

Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-Based Mechanistic Insights into DNMT1-Mediated Maintenance DNA Methylation. Science *335*, 709–712.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., *et al.* (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. *100*, 8418–8423.

Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J.G., Zhu, Y., Kaaij, L.J.T., van IJcken, W., Gribnau, J., Heard, E., *et al.* (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. *25*, 1371–1383.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Erik van Nimwegen, Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., *et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Stepper, P., Kungulovski, G., Jurkowska, R.Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T.P. (2016). Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. Nucleic Acids Res. gkw1112.

Stolzenburg, S., Beltran, A.S., Swift-Scanlan, T., Rivenbark, A.G., Rashwan, R., and Blancafort, P. (2015). Stable oncogenic silencing *in vivo* by programmable and targeted *de novo* DNA methylation in breast cancer. Oncogene.

Stroud, H., Feng, S., Kinney, S.M., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. *12*, R54.

Subbaramaiah, K., Hudis, C., Chang, S.-H., Hla, T., and Dannenberg, A.J. (2008). EP2 and EP4 Receptors Regulate Aromatase Expression in Human Adipocytes and Breast Cancer Cells EVIDENCE OF A BRCA1 AND p300 EXCHANGE. J. Biol. Chem. *283*, 3433–3444.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., *et al.* (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. J. Biol. Chem. *279*, 27816–27823.

Suter, C.M., Martin, D.I.K., and Ward, R.L. (2004). Germline epimutation of MLH1 in individuals with multiple cancers. Nat. Genet. *36*, 497–501.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet. *9*, 465–476.

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.

Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A., and Tajima, S. (2011). Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc. Natl. Acad. Sci. *108*, 9055–9059.

Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B.B., Heard, E., Lao, K., and Surani, M.A. (2011). Deterministic and Stochastic Allele Specific Gene Expression in Single Mouse Blastomeres. PLoS ONE *6*.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. *3*, 226–231.

Terada, N., Shimizu, Y., Kamba, T., Inoue, T., Maeno, A., Kobayashi, T., Nakamura, E., Kamoto, T., Kanaji, T., Maruyama, T., *et al.* (2010). Identification of EP4 as a potential target for the treatment of castration-resistant prostate cancer using a novel xenograft model. Cancer Res. *70*, 1606–1615.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.

Tharakan, R., Lepont, P., Singleton, D., Kumar, R., and Khan, S. (2008). Phosphorylation of estrogen receptor alpha, serine residue 305 enhances activity. Mol. Cell. Endocrinol. *295*, 70–78.

The Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

The Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature *489*, 519–525.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., *et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. CA. Cancer J. Clin. *65*, 87–108.

Toy, W., Shen, Y., Won, H., Green, B., Sakr, R.A., Will, M., Li, Z., Gala, K., Fanning, S., King, T.A., *et al.* (2013). ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. Nat. Genet. *45*, 1439–1445.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trowbridge, J.J., Snow, J.W., Kim, J., and Orkin, S.H. (2009). DNA Methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell *5*, 442–449.

Tsai, S.Q., and Joung, J.K. (2016). Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat. Rev. Genet. *17*, 300–312.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. *32*, 569–576.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R., *et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805–814.

Tuo, Y.-L., Li, X.-M., and Luo, J. (2015). Long noncoding RNA UCA1 modulates breast cancer cell growth and apoptosis through decreasing tumor suppressive miR-143. Eur. Rev. Med. Pharmacol. Sci. *19*, 3403–3411.

Valinluck, V., and Sowers, L.C. (2007). Endogenous Cytosine Damage Products Alter the Site Selectivity of Human DNA Maintenance Methyltransferase DNMT1. Cancer Res. *67*, 946–950.

VanderKraats, N.D., Hiken, J.F., Decker, K.F., and Edwards, J.R. (2013). Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. Nucleic Acids Res. *41*, 6816–6827.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., *et al.* (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. *23*, 555–567.

Vial, E., and Pouysségur, J. (2004). Regulation of Tumor Cell Motility by ERK Mitogen-Activated Protein Kinases. Ann. N. Y. Acad. Sci. *1030*, 208–218.

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., Klasić, M., and Zoldoš, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. gkw159.

Wallenius, V., Hisaoka, M., Helou, K., Levan, G., Mandahl, N., Meis-Kindblom, J.M., Kindblom, L.-G., and Jansson, J.-O. (2000). Overexpression of the Hepatocyte Growth Factor (HGF) Receptor (Met) and Presence of a Truncated and Activated Intracellular HGF Receptor Fragment in Locally Aggressive/Malignant Human Musculoskeletal Tumors. Am. J. Pathol. *156*, 821–829.

Wang, F., Li, X., Xie, X., Zhao, L., and Chen, W. (2008). UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. FEBS Lett. *582*, 1919–1927.

Wang, X.-S., Zhang, Z., Wang, H.-C., Cai, J.-L., Xu, Q.-W., Li, M.-Q., Chen, Y.-C., Qian, X.-P., Lu, T.-J., Yu, L.-Z., *et al.* (2006). Rapid Identification of UCA1 as a Very Sensitive and Specific Unique Marker for Human Bladder Carcinoma. Clin. Cancer Res. *12*, 4851–4858.

Wang, Y., Chen, W., Yang, C., Wu, W., Wu, S., Qin, X., and Li, X. (2012). Long non-coding RNA UCA1a(CUDR) promotes proliferation and tumorigenesis of bladder cancer. Int. J. Oncol. *41*, 276–284.

Warton, K., and Samimi, G. (2015). Methylation of cell-free circulating DNA in the diagnosis of cancer. Front. Mol. Biosci. *2*.

Warton, K., Mahon, K.L., and Samimi, G. (2016). Methylated circulating tumor DNA in blood: power in cancer prognosis and response. Endocr. Relat. Cancer *23*, R157–R171.

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.

Weatherman, R.V., Fletterick, R.J., and Scanlan, T.S. (1999). Nuclear-Receptor Ligands and Ligand-Binding Domains. Annu. Rev. Biochem. *68*, 559–581.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. *39*, 457–466.

Weksberg, R., Nishikawa, J., Caluseriu, O., Fei, Y.L., Shuman, C., Wei, C., Steele, L., Cameron, J., Smith, A., Ambus, I., *et al.* (2001). Tumor development in the Beckwith-Wiedemann syndrome is associated with a variety of constitutional molecular 11p15 alterations including imprinting defects of KCNQ1OT1. Hum. Mol. Genet. *10*, 2989–3000.

Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., *et al.* (2014). Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. Genome Biol. *15*, R49.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., *et al.* (2010). Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. Cell Stem Cell *7*, 532–544.

Wolff, E.M., Byun, H.-M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., and Liang, G. (2010). Hypomethylation of a LINE-1 Promoter Activates an Alternate Transcript of the MET Oncogene in Bladders with Cancer. PLoS Genet *6*, e1000917.

Woodcock, D.M., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1999). DNA methylation in the promoter region of the p16 (CDKN2/MTS-1/INK4A) gene in human breast tumours. Br. J. Cancer *79*, 251–256.

Wu, H., and Zhang, Y. (2014). Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. Cell *156*, 45–68.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *advance online publication*.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. *25*, 679–684.

Xie, S., Shen, B., Zhang, C., Huang, X., and Zhang, Y. (2014). sgRNAcas9: A Software Package for Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites. PLoS ONE *9*, e100448.

Xin, X., Majumder, M., Girish, G.V., Mohindra, V., Maruyama, T., and Lala, P.K. (2012). Targeting COX-2 and EP4 to control tumor growth, angiogenesis, lymphangiogenesis and metastasis to the lungs and lymph nodes in a breast cancer model. Lab. Invest.

Xu, G.-L., and Bestor, T.H. (1997). Cytosine methylation targetted to pre-determined sequences. Nat. Genet. *17*, 376–378.

Xu, K.-P., and Yu, F.-S.X. (2007). Cross Talk between c-Met and Epidermal Growth Factor Receptor during Retinal Pigment Epithelial Wound Healing. Invest. Ophthalmol. Vis. Sci. *48*, 2242–2248.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., *et al.* (2011). Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. Mol. Cell *42*, 451–464.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., *et al.* (2012). Tet3 CXXC Domain and Dioxygenase Activity Cooperatively Regulate Key Genes for Xenopus Eye and Neural Development. Cell *151*, 1200–1213.

Yamada, Y., Jackson-Grusby, L., Linhart, H., Meissner, A., Eden, A., Lin, H., and Jaenisch, R. (2005). Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. Proc. Natl. Acad. Sci. U. S. A. *102*, 13580–13585.

Yang, C., Li, X., Wang, Y., Zhao, L., and Chen, W. (2012). Long non-coding RNA UCA1 regulated cell cycle distribution via CREB through PI3-K dependent pathway in bladder carcinoma cells. Gene *496*, 8–16.

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.

Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell *147*, 1498–1510.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., *et al.* (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*, eaaj2239.

Yoder, J.A., Yen, R.-W.C., Vertino, P.M., Bestor, T.H., and Baylin, S.B. (1996). New 5′ Regions of the Murine and Human Genes for DNA (Cytosine-5)-methyltransferase. J. Biol. Chem. *271*, 31092–31097.

Yokoyama, U., Iwatsubo, K., Umemura, M., Fujita, T., and Ishikawa, Y. (2013). The Prostanoid EP4 Receptor and Its Signaling Pathway. Pharmacol. Rev. *65*, 1010–1052.

Youlden, D.R., Cramb, S.M., Dunn, N.A.M., Muller, J.M., Pyke, C.M., and Baade, P.D. (2012). The descriptive epidemiology of female breast cancer: An international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. *36*, 237–248.

Youlden, D.R., Cramb, S.M., Yip, C.H., and Baade, P.D. (2014). Incidence and mortality of female breast cancer in the Asia-Pacific region. Cancer Biol. Med. *11*, 101–115.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., *et al.* (2012). Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome. Cell *149*, 1368–1380.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010a). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Res. *20*, 1390–1393.

Zhang, Q.-X., Borg, Å., Wolf, D.M., Oesterreich, S., and Fuqua, S.A.W. (1997). An Estrogen Receptor Mutant with Strong Hormone-independent Activity from a Metastatic Breast Cancer. Cancer Res. *57*, 1244–1249.

Zhang, W., Barger, C.J., Eng, K.H., Klinkebiel, D., Link, P.A., Omilian, A., Bshara, W., Odunsi, K., and Karpf, A.R. (2016). PRAME expression and promoter hypomethylation in epithelial ovarian cancer. Oncotarget *7*, 45352–45369.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., *et al.* (2010b). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. *38*, 4246–4253.

Zhao, R., Choi, B.Y., Lee, M.-H., Bode, A.M., and Dong, Z. (2016). Implications of Genetic and Epigenetic Alterations of CDKN2A (p16INK4a) in Cancer. EBioMedicine *8*, 30–39.

Zhou, Q., Atadja, P., and Davidson, N.E. (2007). Histone deacetylase inhibitor LBH589 reactivates silenced estrogen receptor alpha (ER) gene expression without loss of DNA hypermethylation. Cancer Biol. Ther. *6*, 64–69.

Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., *et al.* (2011). Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. PLoS Genet. *7*.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., *et al.* (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *advance online publication*.

Zilli, M., Grassadonia, A., Tinari, N., Di Giacobbe, A., Gildetti, S., Giampietro, J., Natoli, C., and Iacobelli, S. (2009). Molecular mechanisms of endocrine resistance and their implication in the therapy of breast cancer. Biochim. Biophys. Acta BBA - Rev. Cancer *1795*, 62–81.

Appendix A: Supplementary Material for Chapter 2

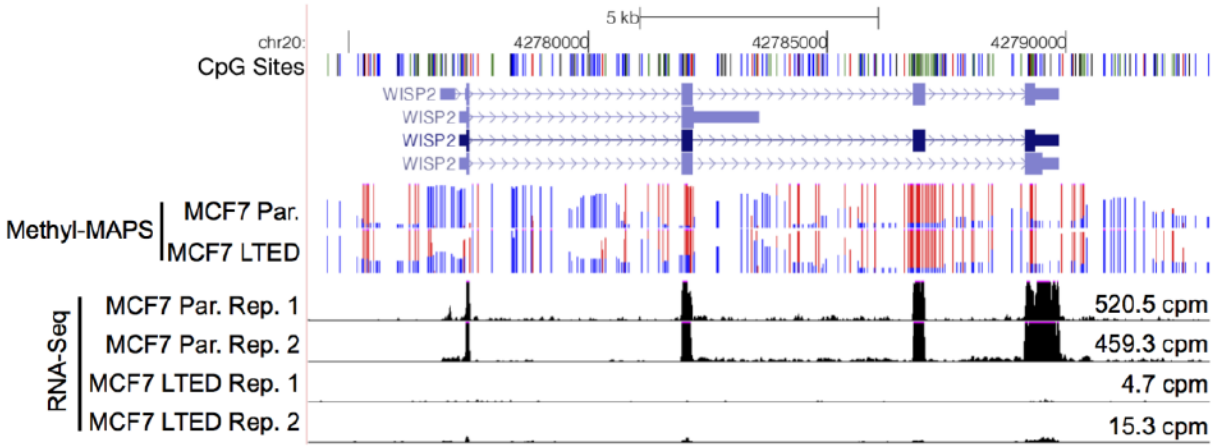


Figure A.1: Methylation and expression at WISP2 in MCF7-LTED cells. Genome browser view of Methyl-MAPS methylation and RNA-seq expression data for *WISP2*. Red and blue lines indicate coverage of methylated and unmethylated fragments, respectively. Individual CpG sites are noted by tics in black at the top track.

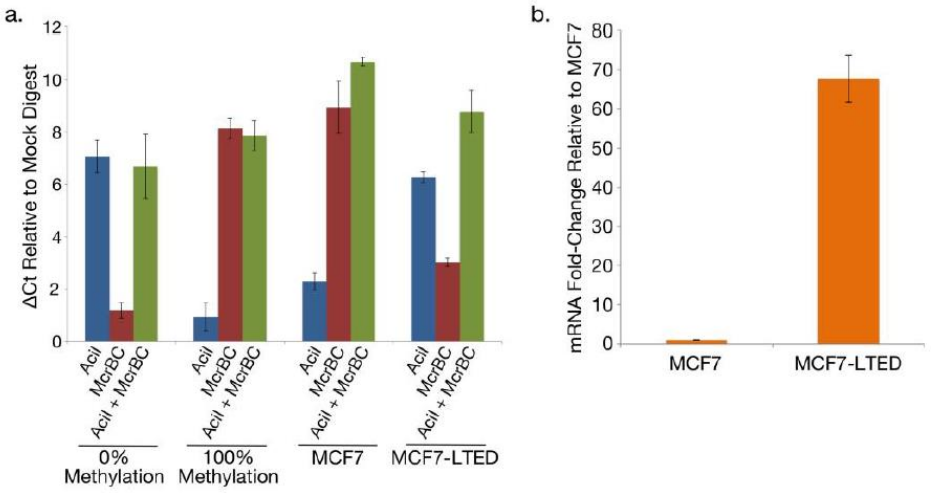


Figure A.2: Validation of PTGER4 methylation and expression changes in MCF7-LTED cells. (A) *PTGER4* methylation was assessed by Methyl-Screen. MCF7 and MCF7-LTED cells are compared to 0% and 100% methylated control DNA. **(B)** RT-qPCR analysis of *PTGER4* expression shows a 68-fold increase in MCF7-LTED cells compared to MCF7 cells. Error bars are SD for three technical replicates.

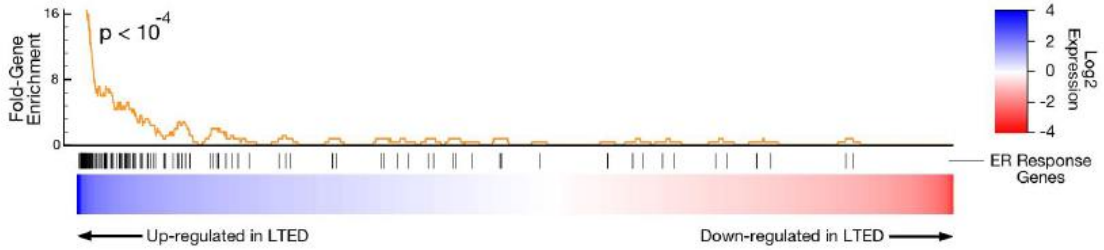


Figure A.3: Gene set enrichment analysis of ER target genes in MCF7-LTED cells. Differentially expressed genes are ranked from highest to lowest expression change. ER-response genes are marked.

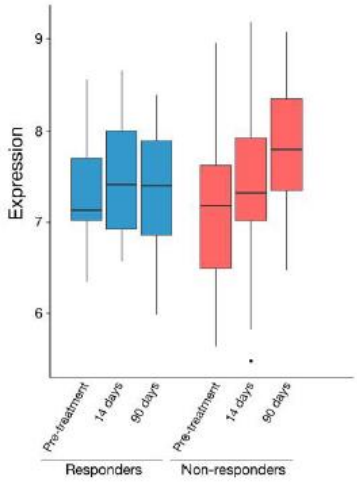


Figure A.4: EP4 expression in responders and non-responders to AI-therapy. Expression data from 42 responders and 14 non-responders to neoadjuvant AI-therapy. Data from Miller *et al.* (2011).

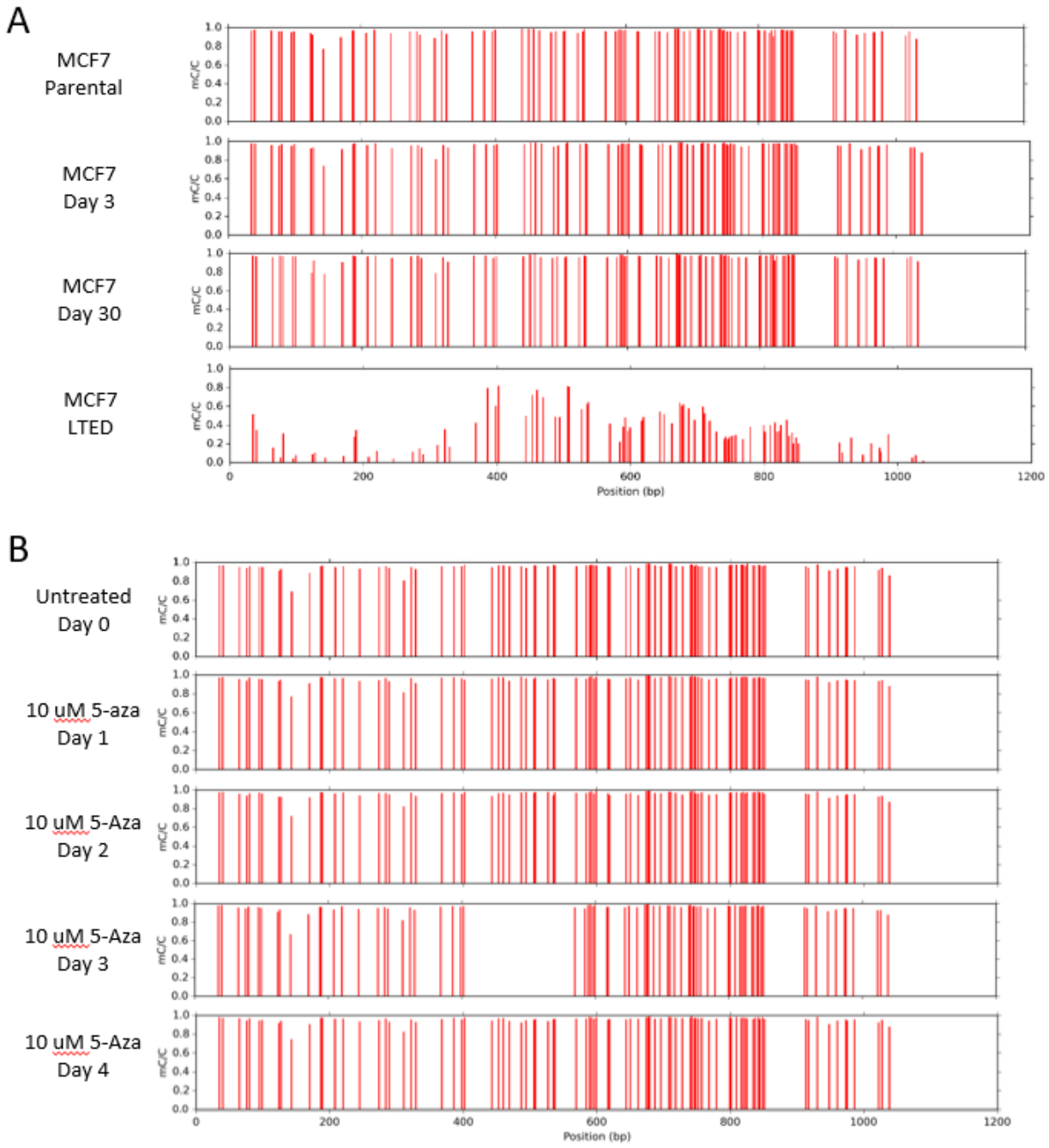


Figure A.5: DNA Methylation analysis of PTGER4 exon 2. (A) Bisulfite sequencing of *PTGER4* exon 2 in samples from the STED time course. **(B)** Bisulfite sequencing of *PTGER4* exon 2 in samples from the 5-azacytidine time course. Missing sequence in Day 3 sample is caused by low sequencing coverage of that region.

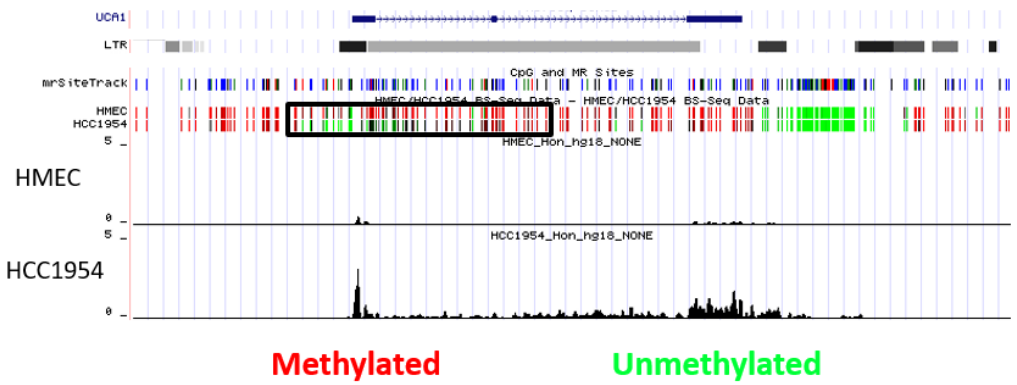


Figure A.6: UCA1 is hypomethylated and upregulated in HCC1954. Genome browser shot showing of the methylation and expression of *UCA1* in comparison between HMEC and HCC1954 cells. Long terminal repeat (LTR) track shown for reference.

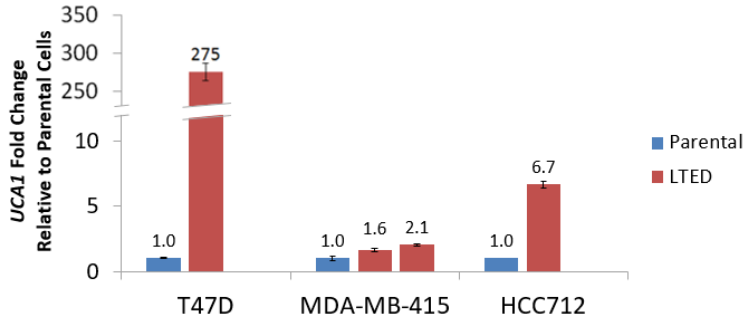


Figure A.7: UCA1 is upregulated in MDA-MB-415 and HCC712 cell lines. RT-qPCR data for *UCA1* expression in two additional cell lines. Error bars are +/- 1 SD.

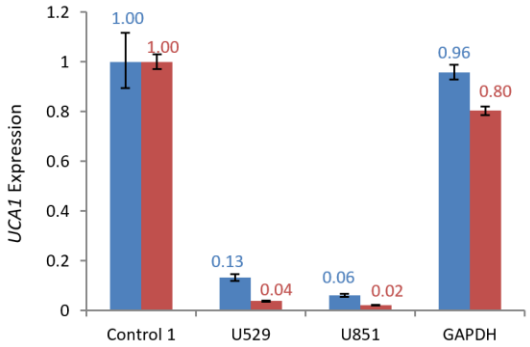


Figure A.8: UCA1 knockdown occurs in both the nucleus and the cytoplasm. RT-qPCR data for *UCA1* expression in T47D RNA separated into nuclear and cytoplasmic fractions. Error bars are +/- 1 SD.

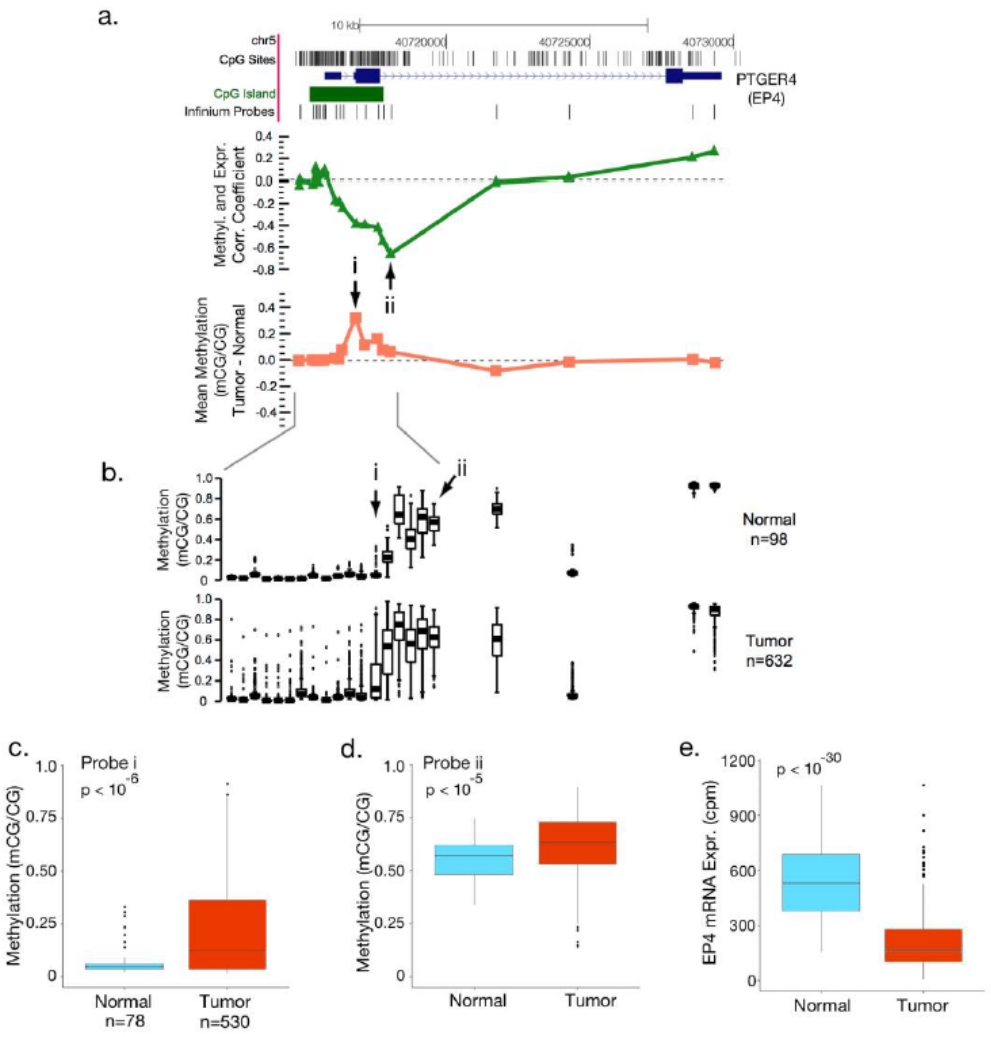


Figure A.9: Methylation analysis of EP4 in TCGA data. (A) Genome browser view of EP4 (*PTGER4*) showing CpG sites and Infinium probe locations. The green line indicates the correlation coefficient between methylation and expression across each probe for 730 TCGA normal and carcinogenic breast samples. The orange line indicates the mean methylation difference between 632 carcinogenic and 98 normal breast samples at each probe. The arrow marked (i) indicates the most differentially methylated probe and the arrow marked (ii) indicates the most anti-correlated probe. **(B)** Boxplots of methylation levels (mCG/CG) at each Infinium probe in tumor and normal TCGA breast samples. Most differentially methylated probe (i) and most anti-correlated probe (ii) are indicated. **(C-E)** Boxplots comparing methylation (C, D) and expression (E) data for normal versus tumor tissue.

Table A.1: Genes with differential expression and DNA methylation in MCF7.

GeneID	Gene	Log2 Expr. Fold-Change	Description	Locus (hg18)
Up-regulated				
ENSG00000171522	**PTGER4**	4.56	PTGER4 prostaglandin E receptor 4, subtype EP4	chr5:40715789-40729594
ENSG00000131016	AKAP12	3.38	AKAP12 A kinase (PRKA) anchor protein 12 isoform 2	chr6:151602827-151721390
ENSG00000119514	GALNT12	3.21	UDP-N-acetyl-alpha-D-galactosamine:polypeptide	chr9:100609802-100652184
ENSG00000141756	FKBP10	2.04	FK506 binding protein 10, 65 kDa	chr17:37222488-37232995
ENSG00000114993	RTKN	2.03	rhotekin isoform c	chr2:74506496-74521218
Down-regulated				
ENSG00000064205	**WISP2**	-5.11	WNT1 inducible signaling pathway protein 2	chr20:42777299-42789866
ENSG00000132329	RAMP1	-3.29	receptor activity-modifying protein 1 precursor	chr2:238432926-238485498

Anchor

Anchor

Table A.2: Raw sequencing statistics from MCF7 and MCF7-LTED Methyl-MAPS analysis.

Sample	Library	F3 Raw Reads	R3 Raw Reads	AAA Pairs
MCF7	McrBC	72,138,738	71,799,998	34,427,093
	RE	80,625,832	80,235,306	37,871,732
MCF7-LTED	McrBC	100,022,230	99,469,120	36,963,847
	RE	92,579,974	92,095,440	36,860,673

Table A.3: Number of CpGs retained in analysis at different levels of coverage in Methyl-MAPS data.

Sample	MR Sites with 1x Coverage	MR Sites with 7x Coverage	% Coverage at 7x	MR Sites with 10x Coverage	% Coverage at 10x
MCF7	8,171,301	7,402,593	88.7%	6,895,024	82.6%
MCF7-LTED	8,160,032	7,234,459	86.7%	6,698,537	80.3%

Anchor

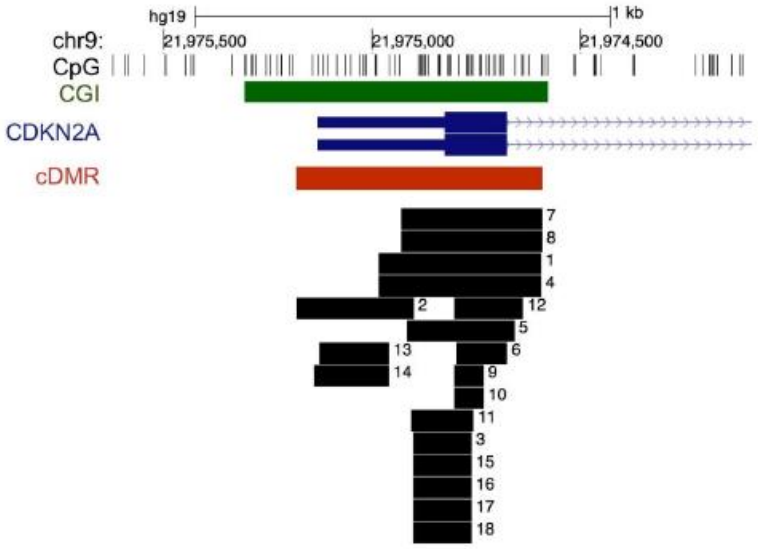
Table A.4: ChIP-qPCR and RT-qPCR primer sequences.

Locus	Purpose	Forward Primer Sequence	Reverse Primer Sequence
pS2	ChIP-qPCR	GGCCATCTCTCACTATGAATCACTTCTGC	GGCAGGCTCTGTTTGCTTAAAGAGCG
PGR1	ChIP-qPCR	GCCTGACCTGTTGCTTCAAT	GCAGGACGACTTCTCAGACC
chr8q24	ChIP-qPCR	AATGCTGGGCTTCCAAGGA	GACCTTGGTGACTGTTGAGGAAAC
PTGER4	RT-qPCR	AGGACAAGGTGAAAGCAGG	GAGTGGACATGATAGTGGCTG
RPL0	RT-qPCR	AGACTGGAGACAAAGTGGGA	CAGACAGACACTGGCAACA
CARM1	RT-qPCR	TCGCCCTCTACAGCCATGA	CACACGGCTGCACTCTGTCT
UCA1	RT-qPCR	TTAGGCTGGCAACCATCAGATCCT	TAAGCTGAGGCTGGCAAAGAGTGA

Anchor

Appendix B: Supplementary Material for Chapter 3

a)



b)

Region	Genomic Coordinate (hg19)	Reference
1	chr9:21974985-21974593	Wong et al. 1999
2	chr9:21974899-21975181	Wong et al. 1999
3	chr9:21974762-21974902	Herman et al. 1996
4	chr9:21974594-21974985	Foster et al. 1998
5	chr9:21974657-21974917	Huschtscha et al. 1998
6	chr9:21974677-21974798	Gonzalez-Zulueta et al. 1995
7	chr9:21974591-21974930	Merlo et al. 1995
8	chr9:21974591-21974930	Herman et al. 1995
9	chr9:21974732-21974802	Widschwendter et al. 2004
10	chr9:21974732-21974802	Shima et al. 2011
11	chr9:21974757-21974906	Barault et al. 2008
12	chr9:21974638-21974802	Bihl et al. 2012
13	chr9:21974959-21975128	Kim et al. 2010
14	chr9:21974959-21975140	Shen et al. 2007
15	chr9:21974762-21974902	Esteller et al. 2001
16	chr9:21974762-21974902	Ishiguro et al. 2006
17	chr9:21974762-21974902	Liang et al. 1999
18	chr9:21974762-21974902	Ward et al. 2003

Fig. B.1: CDKN2A promoter methylation analysis literature review. (A) UCSC genome browser view showing the regions of *CDKN2A* with either a reported relationship between methylation and expression or reported to be cancer-specific changes. The number beside each region corresponds to the reference in (B). **(B)** Table containing the hg19 coordinates of each sequencing region a3s well as the relevant reference.

References For Fig B.1:

Barault, Ludovic, Céline Charon-Barra, Valérie Jooste, Mathilde Funes de la Vega, Laurent Martin, Patrick Roignot, Patrick Rat, *et al.* 2008. "Hypermethylator Phenotype in Sporadic Colon Cancer: Study on a Population-Based Series of 582 Cases." Cancer Research 68 (20): 8541–46. doi:10.1158/0008-5472.CAN-08-1171.

Bihl, Michel P., Anja Foerster, Alessandro Lugli, and Inti Zlobec. 2012. "Characterization of CDKN2A(p16) Methylation and Impact in Colorectal Cancer: Systematic Analysis Using Pyrosequencing." Journal of Translational Medicine 10 (1): 173. doi:10.1186/1479-5876-10-173.

Esteller, Manel, Paul G. Corn, Stephen B. Baylin, and James G. Herman. 2001. "A Gene Hypermethylation Profile of Human Cancer." Cancer Research 61 (8): 3225–29.

Foster, Scott A., David J. Wong, Michael T. Barrett, and Denise A. Galloway. 1998. "Inactivation of p16 in Human Mammary Epithelial Cells by CpG Island Methylation." Molecular and Cellular Biology 18 (4): 1793–1801.

Gonzalez-Zulueta, M, C M Bender, A S Yang, T Nguyen, R W Beart, J M Van Tornout, and P A Jones. 1995. "Methylation of the 5' CpG Island of the p16/CDKN2 Tumor Suppressor Gene in Normal and Transformed Human Tissues Correlates with Gene Silencing." Cancer Research 55 (20): 4531–35.

Herman, J. G., J. R. Graff, S. Myöhänen, B. D. Nelkin, and S. B. Baylin. 1996. "Methylation-Specific PCR: A Novel PCR Assay for Methylation Status of CpG Islands." Proceedings of the National Academy of Sciences 93 (18): 9821–26.

Herman, J G, A Merlo, L Mao, R G Lapidus, J P Issa, N E Davidson, D Sidransky, and S B Baylin. 1995. "Inactivation of the CDKN2/p16/MTS1 Gene Is Frequently Associated with Aberrant DNA Methylation in All Common Human Cancers." Cancer Research 55 (20): 4525–30.

Huschtscha, L I, J R Noble, A A Neumann, E L Moy, P Barry, J R Melki, S J Clark, and R R Reddel. 1998. "Loss of p16INK4 Expression by Methylation Is Associated with Lifespan Extension of Human Mammary Epithelial Cells." Cancer Research 58 (16): 3508–12.

Ishiguro, Atsushi, Takenori Takahata, Masato Saito, Gen Yoshiya, Yoshihiro Tamura, Mutsuo Sasaki, and Akihiro Munakata. 2006. "Influence of Methylated p15 INK4b and p16 INK4a Genes on Clinicopathological Features in Colorectal Cancer." Journal of Gastroenterology and Hepatology 21 (8): 1334–39. doi:10.1111/j.1440-1746.2006.04137.x.

Kim, Jin C., Jin S. Choi, Seon A. Roh, Dong H. Cho, Tae W. Kim, and Yong S. Kim. 2010. "Promoter Methylation of Specific Genes Is Associated with the Phenotype and Progression of Colorectal Adenocarcinomas." Annals of Surgical Oncology 17 (7): 1767–76. doi:10.1245/s10434-009-0901-y.

Liang, Jin-Tung, King-Jen Chang, Jeng-Chang Chen, Chun-Chung Lee, Yung-Ming Cheng, Hey-Chi Hsu, Ming-Shiang Wu, Shih-Ming Wang, Jaw-Town Lin, and Ann-Lii Cheng. 1999. "Hypermethylation of the p16 Gene in Sporadic T3N0M0 Stage Colorectal Cancers: Association with DNA Replication Error and Shorter Survival." Oncology 57 (2): 149–56. doi:10.1159/000012023.

Merlo, Adrian, James G. Herman, Li Mao, Daniel J. Lee, Edward Gabrielson, Peter C. Burger, Stephen B. Baylin, and David Sidransky. 1995. "5′ CpG Island Methylation Is Associated with Transcriptional Silencing of the Tumour Suppressor p16/CDKN2/MTS1 in Human Cancers." Nature Medicine 1 (7): 686–92. doi:10.1038/nm0795-686.

Shen, Lanlan, Paul J. Catalano, Al B. Benson, Peter O'Dwyer, Stanley R. Hamilton, and Jean-Pierre J. Issa. 2007. "Association between DNA Methylation and Shortened Survival in Patients with Advanced Colorectal Cancer Treated with 5-Fluorouracil–Based Chemotherapy." Clinical Cancer Research 13 (20): 6093–98. doi:10.1158/1078-0432.CCR-07-1011.

Shima, Kaori, Katsuhiko Nosho, Yoshifumi Baba, Mami Cantor, Jeffrey A. Meyerhardt, Edward L. Giovannucci, Charles S. Fuchs, and Shuji Ogino. 2011. "Prognostic Significance of CDKN2A (p16) Promoter Methylation and Loss of Expression in 902 Colorectal Cancers: Cohort Study and Literature Review." International Journal of Cancer 128 (5): 1080–94. doi:10.1002/ijc.25432.

Ward, Robyn Lynne, Kay Cheong, Su-Lyn Ku, Alan Meagher, Terence O'Connor, and Nicholas John Hawkins. 2003. "Adverse Prognostic Effect of Methylation in Colorectal Cancer Is Reversed by Microsatellite Instability." Journal of Clinical Oncology 21 (20): 3729–36. doi:10.1200/JCO.2003.03.123.

Widschwendter, Martin, Allison Jones, and Andrew E. Teschendorff. 2012. "Epigenetics Makes Its Mark on Women-Specific Cancers—an Opportunity to Redefine Oncological Approaches?" Gynecologic Oncology, no. 0. Accessed November 7. doi:10.1016/j.ygyno.2012.09.027.

Wong, David J., Scott A. Foster, Denise A. Galloway, and Brian J. Reid. 1999. "Progressive Region-Specific *De novo* Methylation of the p16 CpG Island in Primary Human Mammary Epithelial Cell Strains during Escape from M0 Growth Arrest." Molecular and Cellular Biology 19 (8): 5642–51.

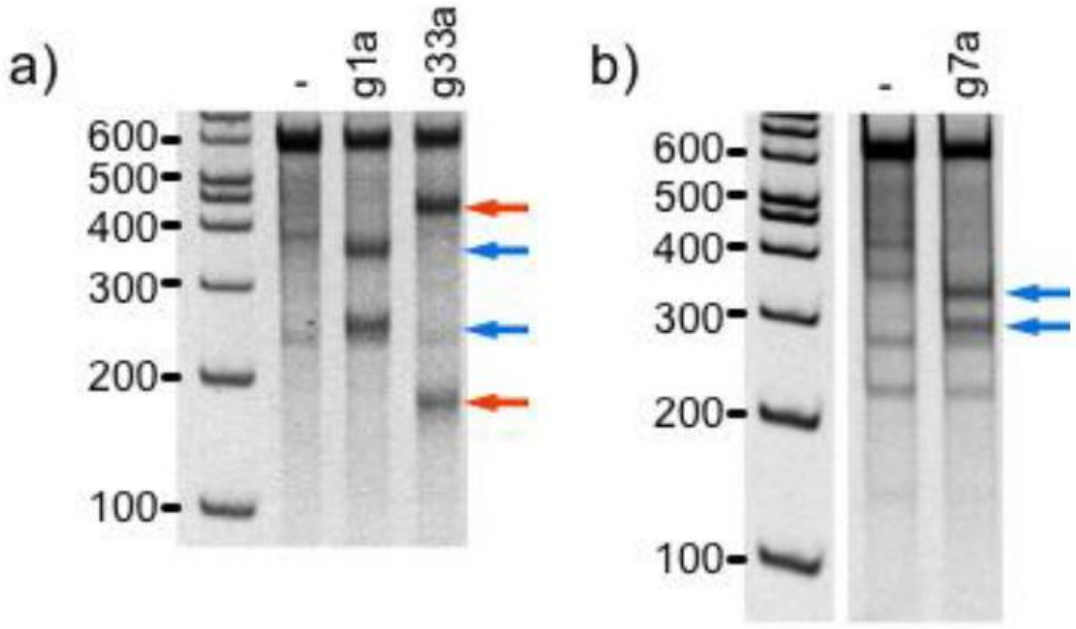


Fig. B.2: sgRNA target validation. Validation of g1a and g33a **(A)** and for g7a (**B)** using a mismatch detection assay. Cells were transfected with active CRISPR/Cas9 and the indicated sgRNA. The parental band is 655 for (A) and 755 for (B). Cleavage products are 384 and 326 for g7a (blue arrows), 384 and 271 for g1a (blue arrows), and 472 and 183 for g33a (red arrows). NHEJ frequencies are estimated at 28% for both g1a and g33a, and at 24% for g7a.

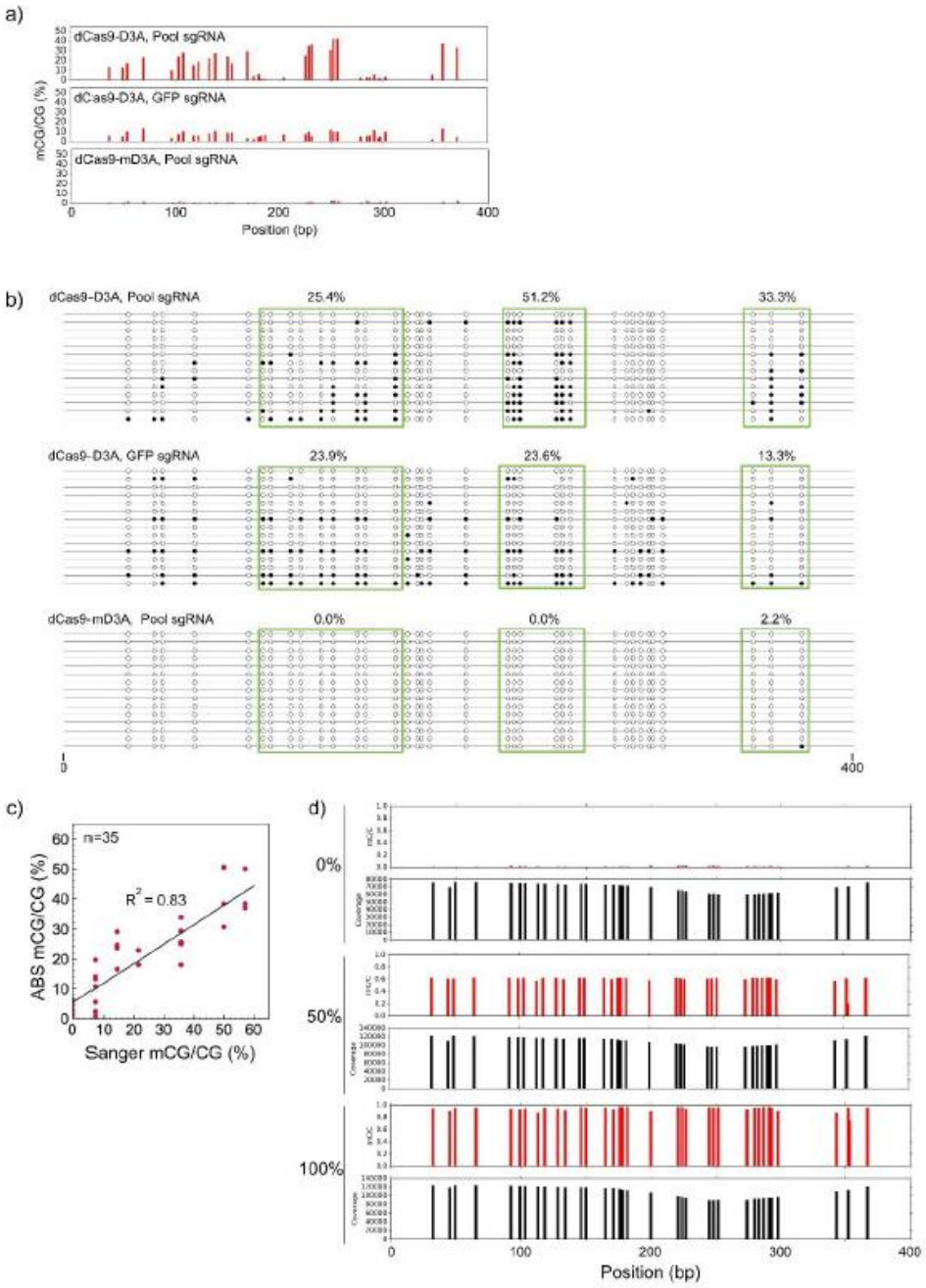


Fig. B.3: Validation of the Amplicon Bisulfite Sequencing (ABS) method. ABS **(A)** and Sanger bisulfite sequencing **(B)** data for the red region in Figure 3.1C. For Sanger data, each line represents one read. Filled circles represent a methylated CpG; empty circles represent an unmethylated CpG. The percent methylation of all CpGs in the green boxes is included on top of each box. **(c)** Scatter plot of the Sanger sequencing versus ABS data showing high correspondence between each method. **(D)** ABS sequencing of the same amplicon in (A) for genomic DNA artificially methylated using M.SssI at 0%, 50% and 100% levels.

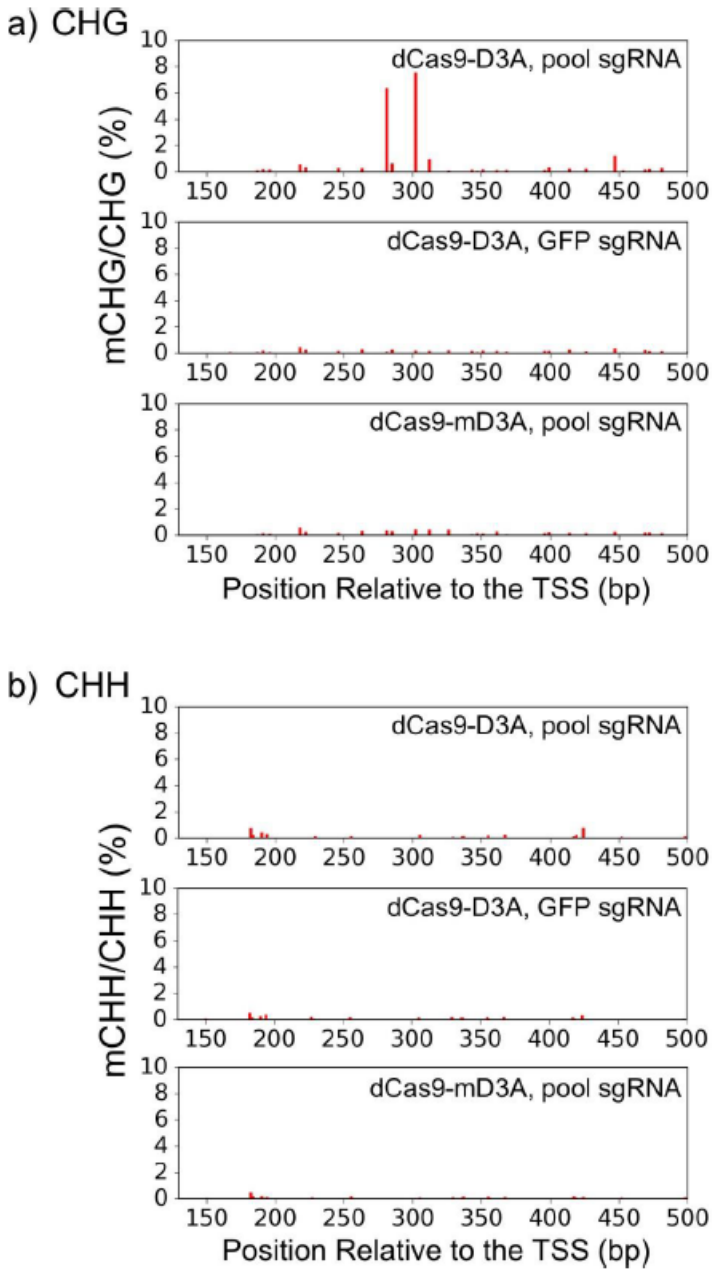


Fig. B.4: dCas9-DNMT3A Induces Non-CpG Methylation. CHG methylation **(A)** and CHH methylation **(B)** after transfection of CDKN2A locus with three pooled sgRNA (g1a, g7a, and g33a). Methylation data is from the red region in Figure 3.1A. Peak methylation is observed at the same six CpG cluster where we observe peak CpG methylation (Figure 3.1C, near position 300 bp).

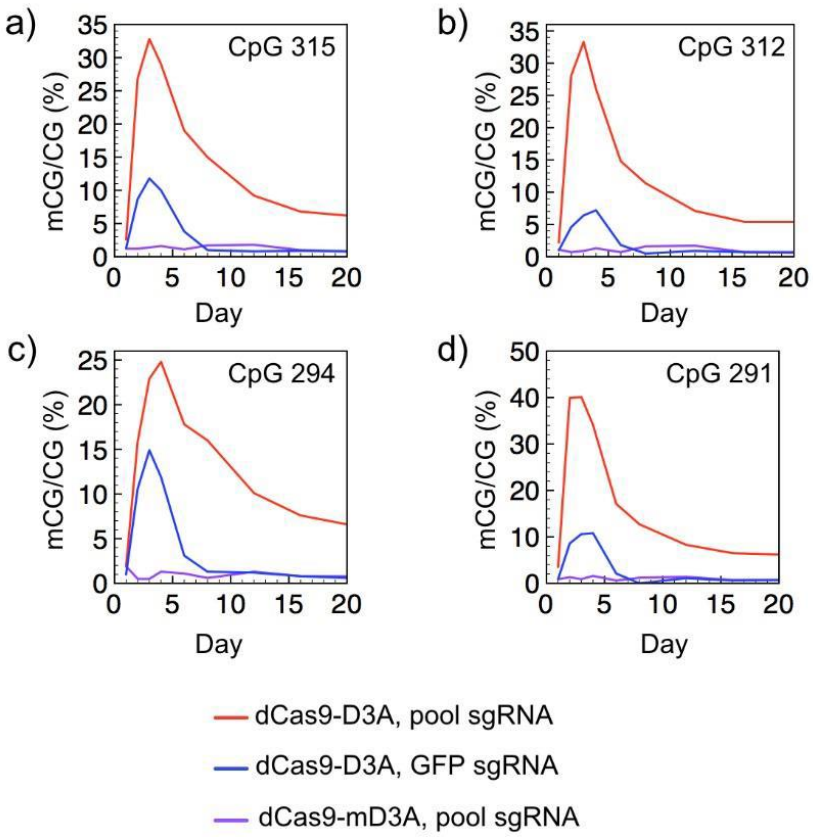


Fig. B.5: Percent methylation for additional CpGs over the 20-day time course. This data supplements Figure 3.1D. CpG locations indicated in each panel are relative to CDKN2A's TSS as in Figure 3.1C.

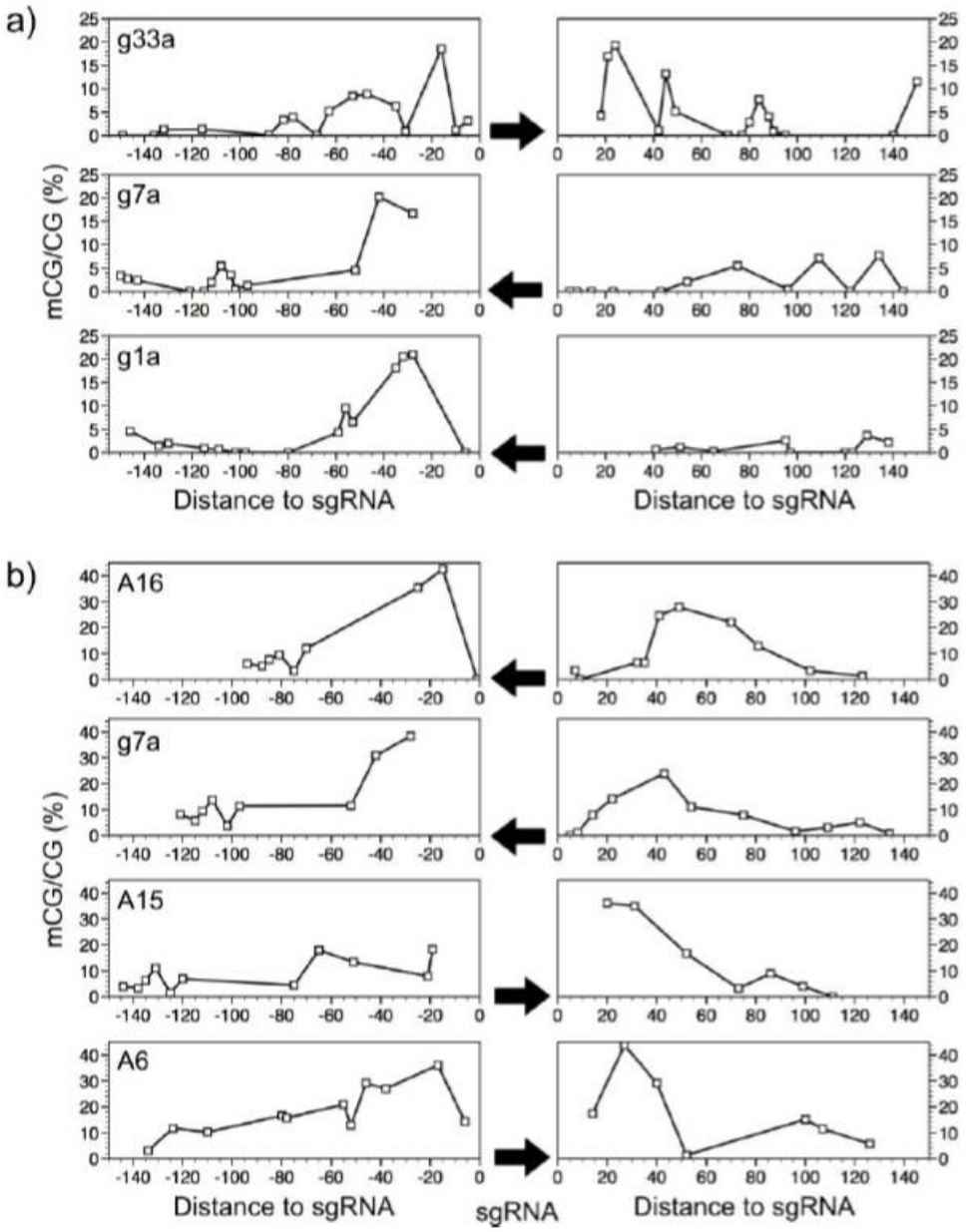


Fig. B.6: sgRNA directed methylation is strand specific and decreases after 50 bp. Arrows indicate the 5' to 3' direction of the sgRNA. **(A)** Percent methylation in each direction on either side of a sgRNA site for cells transfected with a single sgRNA. **(B)** Percent methylation relative to the sgRNA target site. Data comes from cells transfected with both g33a and the indicated sgRNA (Figure 3.1E). CpGs that fall within 50 bp of the g33a sgRNA are removed to avoid confounding effects from the second sgRNA. Methylation levels are background-subtracted using methylation levels from off-target sgRNAs as the background (Figure 3.2B).

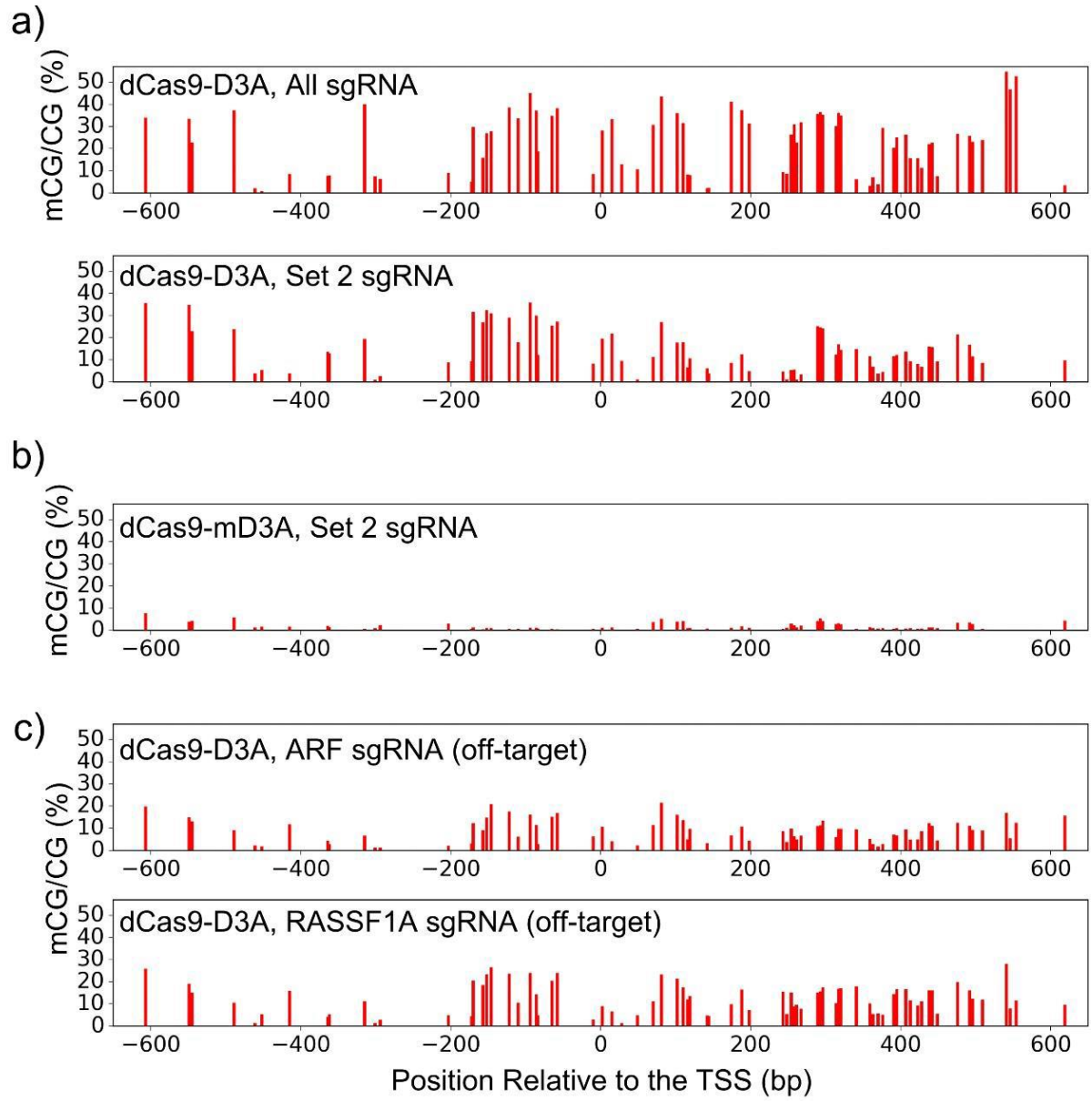


Fig. B.7: Additional ABS data supplemental to Fig 3.2. (A) Methylation at the *CDKN2A* promoter after treatment with the dCas9-D3A and either All or Set 2 *CDKN2A* sgRNA. **(B)** Methylation induced at the *CDKN2A* promoter after treatment with catalytically inactive dCas9-mD3A and Set 2 sgRNA. **(C)** Methylation induced at the *CDKN2A* promoter after treatment with sgRNA to the *ARF* or *RASSF1A* promoters as an off-target control. Sequencing of the *RASSF1A* locus is not shown, since it was highly methylated in all samples, including those with catalytically inactive DNMT3A.

Table B.1. *CDKN2A* targeted sgRNA.

sgRNA	sgRNA Sequence	Genomic Coordinates (hg19)	3 sgRNA Pool	Subset 1	Subset 2	Distance Series
g33A	GTGGCCAGCCAGTCAGCCGAAGG	chr9:21974774-21974796	x	x		x
g1A	TGGGGCGGACCGCGTGCGCTCGG	chr9:21974873-21974895	x		x	x
g7A	CCTTGCCTGGAAAGATACCGCGG	chr9:21974988-21975010	x	x		x
A1	TCGAGAGATCTGTACGCGCGTGG	chr9:21974449-21974471			x	
A2	CTTGGCTTCCCAAGCCCCCAGGG	chr9:21974488-21974510		x		
A3	ACCCCTTCTGAAAACTCCCCAGG	chr9:21974540-21974562			x	
A4	GGTGGGTAGAGGGTCTGCAGCGG	chr9:21974655-21974677		x		
A5	CCCAACGCACCGAATAGTTACGG	chr9:21974693-21974715			x	
A6	GGCACCCCTGAAGTCGCCCCAGG	chr9:21975070-21975092			x	x
A7	CTGGGAGCAGGGAGGCCGGAGGG	chr9:21975193-21975215		x		
A8	CGGAGGAAGGAAACGGGGCGGGG	chr9:21975282-21975304			x	
A9	TCGGGGTGTTTGGTGTCATAGGG	chr9:21975335-21975357		x		
A10	GGTGATTTCGATTCTCGGTGGGG	chr9:21975421-21975443			x	
A11	CTGAAAATCAAGGGTTGAGGGGG	chr9:21975469-21975491		x		
A12	GAGCAGTCCGACTCTCCAAAAGG	chr9:21975535-21975557			x	
A13	GCGTGTAAAACGGCTGTCTGGGG	chr9:21975584-21975606		x		
A14	GAGCAGACTCTGAAAAGATGAGG	chr9:21975643-21975665			x	
A15	GGACGCCGTGAGCGAGTGCTCGG	chr9:21975011-21975033				x
A16	TCCAGAGGATTTGAGGGACAGGG	chr9:21974961-21974983				x
A17	GGGGCTGGCTGGTCACCAGAGGG	chr9:21974896-21974918				x

Table B.2. ARF and RASSF1A targeted sgRNA.

sgRNA	sgRNA Sequence	Genomic Coordinates (hg19)
ARF_1	GCACGCGCGCCGAATCCGGAGGG	chr9:21994286-21994308
ARF_2	GGTGCCAAAGGGCGGCGCAGCGG	chr9:21994363-21994385
ARF_3	ACGAAAACCCTCACTCGCGGCGG	chr9:21994260-21994282
RASSF1A_1	GCACCCAGGTTTCCATTGCGCGG	chr3:50378369-50378391
RASSF1A_2	GAAACACGGGTATCTCCGCGTGG	chr3:50378524-50378546
RASSF1A_3	TGTGAGGAGGGGACGAAGGAGGG	chr3:50378461-50378483

Table B.3. Amplicons for bisulfite sequencing.

Amplicon	Genomic Coordinate	Amplicon Length (bp)	Primer Sequence
CDKN2A_B	chr9:21974899-21975181	283	F: GTAGGTGGGGAGGAGTTTAGTT
			R: TCTAATAACCAACCAACCCCTCC
CDKN2A_A	chr9:21974594-21974985	392	F: TTTTTAGAGGATTTGAGGGATAGG
			R: CTACCTAATTCCAATTCCCCTACA
CDKN2A_1	chr9:21974140-21974555	215	F: TTGATTTTTTAAAGTGAAGGGATTATAAGG
			R: ATTTTCAAAAAAAATTTATAATCACAAACCTCC
CDKN2A_2	chr9:21974523-21974920	398	F: GAGGTTTGTGATTATAAATTTTTTTTG
			R: AAAAACTAACTAATCACCAAAAAATAAAAC
CDKN2A_3	chr9:21974890-21975191	301	F: GTTTTATTTTTTGGTGATTAGTTAGTTTTT
			R: ATATAAAAAAACAAATAAAAAAAAACCCAATC
CDKN2A_4	chr9:21975160-21975525	365	F: GATTGGGTTTTTTTTATTTGTTTTTTATAT
			R: ACTAAAATTTCTAACTTAATAAACCCC
CDKN2A_5	chr9:21975512-21975862	351	F: TAGAAATTTTAGTTTAAAGGATTTTTTTTGGAGAG
			R: TAAAAAATATAAAAAAAACAAAACAATATTTAAAAC
CDKN2A_6	chr9:21975634-21975976	343	F: TTGGTATAAGAGTAGATTTTGAAAAGATGAGG
			R: ATTAATACATCATTTAAACTCACAACAACCCC
ARF	chr9:21994178-21994434	257	F: TGGGTTTTAGTTTGTAGTTAAGGG
			R: CCTCAATAACATCAACAC

Table B.4. RT-qPCR primers.

Gene Name	Forward Primer	Reverse Primer
GAPDH	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG
CDKN2A	GCTGCCCAACGCACCGAATAGTTA	TGCACGGGTCGGGTGAGAG
ARF	GATGCTACTGAGGAGCCAGCGTCTA	GCAGCAGCTCCGCCACTC
ACTB	CGACAGGATGCAGAAGGAG	TCCTGCTTGCTGATCCACAT
RPL0	AGACTGGAGACAAAGTGGGA	CAGACAGACACTGGCAACA

Appendix C: Supplementary Material for Chapter 4

Table C.1. Alternative DNMT Cloning Primers

Name	Sequence
DNMT1_aa656_Outer_F	GACAGAGAAGACAAGGAGAAC
DNMT1_aa730_Outer_F	TCACCCAAAAAAATGCACCA
DNMT1_Outer_R	GGCTTTGGCCAACATACAAA
DNMT1_aa656_NheI_FLAG_F	ATCGAGGCTAGCGACTACAAAGACGATGACGACAAGGGAAGCGACAGAGAAGACAAGGAGAACG
DNMT1_aa730_NheI_FLAG_F	ATCGAGGCTAGCGACTACAAAGACGATGACGACAAGGGAAGCTCACCCAAAAAAATGCACCAG
DNMT1_AgeI_R	ATCGAGACCGGTGGCTTTGGCCAACATACAAAG
DNMT3B_Outer_F	GTCCTGCGGCGCCGGA
DNMT3B_Outer_R	CTATTCACATGCAAAGTAGT
DNMT3B_NheI_FLAG_F	ATCGAGGCTAGCGACTACAAAGACGATGACGACAAGGGAAGCGTCCTGCGGCGCCGGAA
DNMT3B_AgeI_R	ATCGAGACCGGTCTATTCACATGCAAAGTAGTC
mDnmt3a-3l_NheI_FLAG_F	ATCGAGGCTAGCGACTACAAAGACGATGACGACAAGGGAAGCCATATGAACCATGACCAGGAA
mDnmt3a-3l_AgeI_R	ATCGAGACCGGTTTAAAGAGGAAGTGAGTTTTGAG

Red Text = Restriction Site
Blue Text = FLAG Tag Sequence

Table C.2. sgRNA Sequences

sgRNA	sgRNA Sequence	Genomic Coordinates (hg19)
g33A	GTGGCCAGCCAGTCAGCCGAAGG	chr9:21974774-21974796
g1A	TGGGGCGGACCGCGTGCGCTCGG	chr9:21974873-21974895
g7A	CCTTGCCTGGAAAGATACCGCGG	chr9:21974988-21975010
ARF_1	GCACGCGCGCCGAATCCGGAGGG	chr9:21994286-21994308
ARF_2	GGTGCCAAAGGGCGGCGCAGCGG	chr9:21994363-21994385
ARF_3	ACGAAAACCCTCACTCGCGGCGG	chr9:21994260-21994282

Table C.3. Bisulfite Sequencing Regions and Primers

Amplicon	Genomic Coordinate (hg19)	Amplicon Length (bp)	Primer Sequence
CDKN2A_B	chr9:21974594-21974985	392	F: TTTTTAGAGGATTTGAGGGATAGG
			R: CTACCTAATTCCAATTCCCCTACA
CDKN2A_-64	chr9:21974890-21975190	301	F: GTTTTATTTTTTGGTGATTAGTTAGTTTTT
			R: ATATAAAAAAACAAATAAAAAAAAACCCAATC
ARF	chr9:21994178-21994434	257	F: TGGGTTTTAGTTTGTAGTTAAGGG
			R: CCTCAATAACATCAACAC
LMNA	chr1:156084418-156084817	400	F: TTGGGATTGTTTTTTTAAGAGTAGTG
			R: CTACAAATCCTCCTTCTCCTACAACC

Table C.4. RT-qPCR Primers

Gene	Forward Primer	Reverse Primer
CDKN2A	GCTGCCCAACGCACCGAATAGTTA	TGCACGGGTCGGGTGAGAG
ARF	GATGCTACTGAGGAGCCAGCGTCTA	GCAGCAGCTCCGCCACTC
ACTB	CGACAGGATGCAGAAGGAG	TCCTGCTTGCTGATCCACAT
GAPDH	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG
RPL0	AGACTGGAGACAAAGTGGGA	CAGACAGACACTGGCAACA