NORTHWESTERN UNIVERSITY


Perturbations of Epigenetic Events in Prostate Cancer: The Role of TET1, A DNA Demethylase


A DISSERTATION


SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS


for the degree


DOCTOR OF PHILOSOPHY


Field of Driskill Graduate Training Program in Life Sciences


By


Yeqing (Angela) Yang


EVANSTON, ILLINOIS


June 2017

**ABSTRACT**

FOXA1 is a FKHD family protein that plays pioneering roles in lineage-specific enhancer activation and gene transcription. Through genome-wide location analyses, here we show that FOXA1 expression and occupancy are, in turn, required for the maintenance of these epigenetic signatures, namely DNA hypomethylation and histone 3 lysine 4 methylation. Mechanistically, this involves TET1, a 5-methylcytosine dioxygenase. We found that FOXA1 induces TET1 expression via direct binding to its cis-regulatory elements. Further, FOXA1 physically interacts with the TET1 protein through its CXXC domain. TET1 thus co-occupies FOXA1-dependent enhancers and mediates local DNA demethylation and concomitant histone 3 lysine 4 methylation, further potentiating FOXA1 recruitment. Consequently, FOXA1 binding events are markedly reduced following TET1 depletion. Together, our results suggest that FOXA1 is not only able to recognize but also remodel the epigenetic signatures at lineage-specific enhancers, which is mediated, at least in part, by a feed-forward regulatory loop between FOXA1 and TET1.

Continuing our endeavor to better understand TET1's role in context of prostate cancer, we found that there exists a transcript isoform for TET1 in prostate cells, which has not been characterized before. Driven by an alternative promoter, the isoform lacks the first 2 exons of the full length gene but has an extra exon from the intronic region in front of exon 3, and is expressed more than 20-fold higher than full length. Functional experiments revealed that the short isoform (later referred to as TET1s) is important for regulating AR expression and thus AR signaling program, therefore making TET1s crucial for prostate cancer cell growth and survival.

# ACKNOWLEDGEMENTS

It has been said that obtaining a PhD degree is not a sprint but a marathon. In my case, this marathon has taken a bit short of 6 years to complete, from Fall 2011 to Spring 2017. The journey is long and not without obstacles, but I am fortunate to have received help and support along the way from many.

First of all, I would like to thank my advisor, Dr. Jindan Yu, for seeing the potential in me and letting me join her lab. Coming to Northwestern, I had held a strong interest in studying epigenetics as a result of my undergraduate research work, particularly DNA methylation. And it is in her lab that I found an exciting project and a drive to pursue answers in a fascinating field. Jindan has always encouraged my professional development, by generously providing resources and opportunities for me to attend conferences, form collaborations, and write review articles, all of which contributed to my better understanding of how to be a scientist. Moreover, I have to thank current and past members in the Yu lab, Longtao Wu, Jung Kim, Jonathan Zhao, Will Fong to name a few, for spending their time and sharing their expertise to allow me to continually learn and acquire new skills. Especially Jung, both a peer mentor and a friend in lab, I feel fortunate to have her as a model whose qualities and achievements I try to learn from. Additionally, I am also grateful to my committee members, Drs. John Crispino, Debu Chakravarti, and Jon Licht, for their support from the beginning and many helpful suggestions to my project, in addition to their kind responses to requests for recommendation letters. And I also want to thank the DGP program, Steve Anderson and Pam Carpentier, for all the support and resources in not only academic but also career development aspects, provided to graduate students.

Outside of lab, I also want to thank my friends and fellow PhD students, Sali Liu, Xiaojun Bao, Yiwen Zhu, Xiaobao Li, Sandy Shan, with whom I shared lots of memories exploring various corners of Chicago and venturing into different cuisines and cultures the city has to offer. Last but not least, I have to thank my parents, for their never-ending support, in both an emotional and an economical sense. My dad, who obtained his PhD in his forties, has taught me to be unafraid and always encouraged me to do what I set my mind on; though he made publishing papers and finishing the degree seem much easier than what I experienced. My mom, has always been patient and always providing unconditional support, including

her hand-knitted sweaters and scarves which proved to be indispensable for helping me get through the many cruel winters in Chicago (often referred to as Chiberia).

*To Mom and Dad.*

**TABLE OF CONTENTS**

**LIST OF FIGURES AND TABLES**

**CHAPTER 1: INTRODUCTION**

**I. DNA (Hydroxy)Methylation and Their Implications in Disease**

<u>DNA Methylation</u>

In the early 19th century, the French biologist Jean Baptiste Lamarck proposed an insightful idea that organisms have the capability to alter their characteristics in response to environmental changes and subsequently pass them on to future progeny. This suggestion of a potential ability of self adaptation, which leads to direct evolution, is known as the Lamarckian inheritance of acquired traits[1]. In contrast, Darwin spoke of an indirect evolutionary process that involves stochastic mutations giving rise to advantageous phenotypes to account for viability and survival of the individual. Through the remainder of the 19th century, as Darwin's idea on natural selection received approval and recognition in the academic world, Lamarck's theory was largely disregarded by the scientific community. However, in the past several decades, Lamarck's theory has been attracting more interest, as there is new evidence that phenotypes can be modified by environmental influences[2]. Currently, studies performed in the field of epigenetics may provide insight into the integrative effects of genetics and environment, offering a new perspective for understanding the etiologies of human disorders. The term "epigenetics" literally means "above the genetic sequence," and it refers to the regulation of genes through processes that do not involve alteration to the DNA sequence[3]. Epigenetic changes are thought to be reversible and heritable through mitosis and meiosis, and they play a crucial role in cellular differentiation and development[4]. In other words, this non-sequence based information, which is essential for normal cellular function, is preserved during DNA replication and cell division and can even be passed on from one generation to the next.

Epigenetic regulation of gene expression is generally thought to occur in two ways: DNA methylation and histone modifications (e.g. post-translational transformations that include acetylation, methylation, phosphorylation and ubiquitylation)[5], both of which act together to influence the architecture of chromatin, and ultimately the expression and function of genes[4]. DNA methylation refers to the addition of a methyl group (CH3) onto the C5 position of a cytosine ring, usually in CpG dinucleotides, which are regions in a linear sequence of DNA where a cytosine nucleotide is adjacent to a guanine. This covalent

transformation is performed and maintained by a family of enzymes named DNA methyltransferases (DNMTs)[5] (**Figure 1.1A**). It has been found that, in humans, around 80% of all CpG dinucleotides are methylated[6]. The remaining unmethylated sites are mostly located in CpG islands, which are long stretches of DNA (of at least 500 base pairs) that contain clusters enriched in CpG sites[4]. In humans, CpG islands are normally located on the promoter region of genes[7]. DNA methylation in the CpG islands triggers a complex series of events downstream, recruiting various chromatin modifying complexes (such as the methyl-CpG-binding proteins and histone deacetylases), which interact to remodel chromatin conformation[8]. The structure of chromatin becomes more condensed with the formation of heterochromatin, which might eventually diminish the amount of transcription in that region. At the same time, the methyl groups on CpG sites also interfere with the recruitment of transcription factors, which may also reduce transcription to a large extent and consequently result in downregulation of genes[9].

DNA Hydroxymethylation

More recently, it is discovered that 5-methylcytosine (5mC) can be further oxidized to generate, in a stepwise manner, 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[10] (**Figure 1.1A**). These covalent modifications to cytosines are recognized as additional epigenetic events, and their functions in mammalian genome have been widely studied in recent years. While 5mC is quite constant across adult tissues, marking approximately 4-5% of all cytosines, 5hmC level is estimated to vary between 0.1% and 0.7% and is the highest in tissues of the central nervous system[11,12] as well as in embryonic stem cells (~0.4%)[10,13]. Subsequent substrates 5fC and 5caC are found to have much lower abundance, 0.02% of all cytosines for 5fC[13] and more than 6-fold less for 5caC[14]. The oxidized 5mC derivatives have been demonstrated to lead to both passive and active DNA demethylation[15], which will be discussed further below. Moreover, due to 5hmC's higher abundance in specific tissue types, it is also believed to behave as a stable epigenetic mark[16], and its exact function is under extensive research.

**Figure 1.1**

**A.**
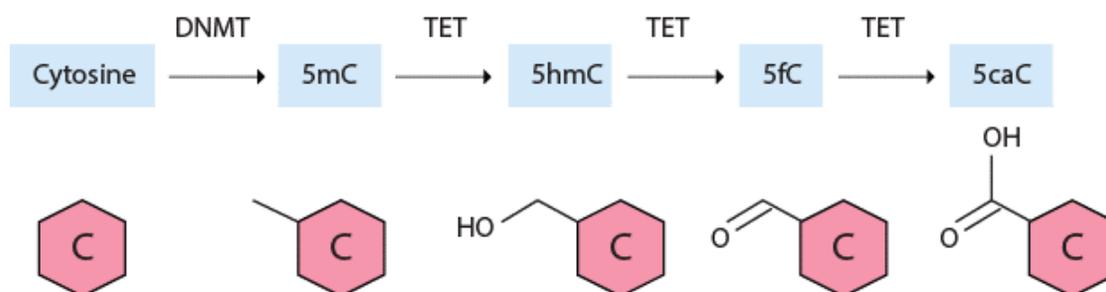


**B.**

**Figure 1.1 DNA hydroxymethylation can be carried out by TET family proteins.** (**A**) Step-wise oxidation reactions conducted by TET family proteins on methylated cytosine nucleotides in DNA. (**B**) Schematic illustration of 3 members of the TET1 family enzymes.

In 2009 it was first reported that hydroxymethylation as well as downstream oxidation reactions are catalyzed by a family of 3 enzymes called TET (Ten Eleven Translocation) proteins, also known as methylcytosine dioxygenases[10]. All members of the TET family, TET1, TET2, and TET3, are oxoglutarate- and iron-dependent enzymes that have the capacity to successively catalyze 5mC oxidation. As depicted in **Figure 1.1B**, they contain a C-terminus catalytic domain that is high in cysteine residues, which is thought to be important for DNA binding[17], and also consists of predicted domains, including a double-stranded β-helix for binding of Fe(II) and 2-oxoglutarate (2OG or α-ketoglutarate), which are essential cofactors in the oxidation process[11]. Furthermore, TET1 and TET3 contain a CXXC zinc finger motif in their N-terminus, which is found in many chromatin-associated proteins, including DNMT1 and MLL, etc., and thought to be involved in recognizing CpG sites in the genome[15].

There has been an increasing amount of evidence indicating that TET proteins and TET-mediated 5mC oxidation can ultimately lead to DNA demethylation, through a couple of pathways. During DNA replication in cell division, maintenance methylation carried out by DNMT1 has to occur so that newly synthesized strands of DNA will be symmetrically methylated as dictated by the complementary template. During this process, DNMT1 interacts with and is recruited by UHRF1, which binds to hemimethylated CpG dinucleotides[18]. However, *in vitro* assays demonstrated that affinity of UHRF1 to hemi-5hmC is 10-fold less than that of hemi-5mC[19], thus dramatically reducing the activity of recombinant DNMT1 by 12-fold[19] to 50-fold[20] on hydroxymethylated DNA. Thus, this 5hmC-dependent inhibition of maintenance DNA methylation during DNA replication is considered as passive demethylation. Moreover, TET proteins have also been shown to "actively" demethylate" DNA, through crosstalk with the base excision repair pathway (BER). Several reports showed that 5fC and 5caC bases can be recognized and cleaved by thymine DNA glycosylase (TDG)[21-23], consequently leading to their removal and replacement by a normal cytosine nucleotide. Both electrophoretic gel mobility assays (EMSAs) and structure of TDG-5caC reveal that TDG binds to 5caC:G mismatches much more efficiently than its conventional substrate T:G[23].

Genomic Distributions of 5mC and 5hmC and Ways to Study Them

Understanding the biological functions of 5mC and its oxidative derivatives relies significantly on how well we know their genomic locations, therefore extensive efforts have been dedicated to map precisely where these marks reside. Techniques such as bisulfite sequencing, in which treatment of DNA with sodium bisulfite converts unmethylated cytosines into uracils[24]), and methyl-DNA immunoprecipitation (MeDIP) sequencing[25], in which a 5mC-specific antibody is used to enrich methylated DNA fragments, have been instrumental in aiding our knowledge of genome-wide distribution of 5mC. It has been well illustrated that methylation in close proximity to transcription start sites (TSSs) are associated with gene silencing due to blockage of transcription initiation, whereas methylation in the gene bodies may play a role in stimulating transcription elongation and thus tend to be positively correlated with gene expression[26]. In addition, methylation is also found to be enriched in repeat regions such as centromeres, where it is critical for maintaining chromosomal stability and suppressing transposable elements[27].

With the discovery of 5hmC bases, researchers have developed a number of approaches to determine their exact locations in the genome. A technique similar to MeDIP but adapted for using an antibody that specifically recognizes 5hmCs, termed hMeDIP, is one widely used method for genome-wide probing. However, concerns have been raised for this approach as the antibody has a tendency to preferentially recognize regions with high density of 5hmC[28] and CA repeats[29], resulting in enrichment bias which could obscure our interpretation of 5hmC function. To overcome this problem, a chemical labeling method was developed by Song et al.[30], which takes advantage of the efficient and specific binding between biotin and streptavidin to eliminate the modification density bias. This approach uses β-glucosyltransferase (βGT) to transfer an azide-modified glucose group onto 5hmC, and then biotin can be added to the azide group through click chemistry, and subsequently streptavidin beads are used to specifically pull down fragments containing 5hmC labeled with biotin[30]. Facilitated by next generation sequencing, these profiling methods have revealed that 5hmC is enriched in gene promoters in many cell types, including human and mouse embryonic stem cells (ESCs)[28,31-34], mouse neural progenitor cells[35],

mouse neurons[35] and cerebellum[36], and associated genes are expressed at lower levels compared to other genes without hydroxymethylated promoters. However this observation is not indicative of a repressive role for 5hmC, since its deposition has to rely on pre-existing 5mC at that particular region. In other cell types that were studied, including primordial germ cells (PGCs)[37], adult nervous tissue[36], liver cells[38] and benign nevi[39], 5hmC is found to be depleted at TSSs. More intriguingly, 5hmC is shown to be enriched in gene bodies in almost all mammalian cell types studied[31,33,35,36,38,39], where gene expression seems to be positively correlated with gene body hydroxymethylation. Moreover, enhancers, which typically have low CpG density and are depleted of DNA methylation[40], are shown to have high levels of 5hmC[31,32,38,40,41]. It has been reported that 5hmC levels increase significantly at active enhancers in differentiating ESCs, and that the gain in 5hmC coincides with onset of differentiation and occurs either before or with histone 3 lysine 27 acetylation (H3K27ac), a histone modification known for active enhancers[42]. Based on these results, 5hmC has been thought to be strongly associated with enhancer activation and lineage specification, and TET proteins are postulated to play a role in transcription factor occupancy at these active enhancer sites. However, exact mechanisms underlying this phenomenon await to be further elucidated.

Aberrations of DNA (Hydroxy)Methylation in Disease

Epigenetic changes are very pertinent to human health and disease. Thus, alongside the study of 'traditional' DNA sequence variations and environmental factors, epigenetic mechanisms add a new perspective in the investigation of disease etiology.

Studies have shown strong evidence that suggests epigenetic mechanisms are closely associated with a variety of human diseases, especially complex diseases with non-Mendelian patterns of inheritance, where the proportions of various observed phenotypes do not match the expected values predicted using Mendel's Laws[43]. For instance, there has been compelling evidence that suggests several types of cancer, such as breast cancer and prostate cancer, are associated with epigenetic alteration[44,45]. Other human diseases, including Prader-Willi and Angelman syndrome, are also shown to have causes related to epigenetics[43]. These disorders are thought to be imprinting disorders caused by the

misregulation of imprinted genes. Imprinting refers to the regulation of genes by epigenetically silencing one copy of either the paternal or the maternal gene by DNA methylation, which results in mono-allelic expression of that particular gene. Around 90 imprinted genes have been identified so far, and they play an important role in helping us understand relevant human pathologies due to their susceptibility to epigenetic changes[46]. It has also been suggested that epigenetic factors may be involved in a number of psychiatric disorders, including schizophrenia, bipolar disorder and major depression[47-49]. Several key genes that are thought to be related to the etiology of psychiatric disorders have been found to be differentially methylated between affected and control subjects[50]. Variations in methylation patterns have been identified in loci associated with glutamatergic and GABAergic neurotransmission pathways, which is consistent with previous findings on the pathogenesis of such psychotic diseases[51]. It is believed that epigenetic changes in prostate cancer appear to be manifested earlier than genetic changes, thus are likely to be more closely related to disease etiology[52]. Studies have shown that abnormalities in DNA methylation in prostate cancer can be detected at the earliest stages of transformation[53]. An extensive list of genes has been reported to exhibit hypermethylation specifically in prostatic tumor cells[54]. Overall a combination of several of them, e.g. GSTP1 and APC, can be subjected to hypermethylation assays to allow a clear and robust discrimination between benign and cancerous prostate cells[55].

The role of 5hmC in disease has not been clearly understood yet, and there is contrasting evidence regarding 5hmC's global level in various types of disorders. Some reports show that 5hmC levels exhibit a global decrease in hematological malignancies, which may be connected to impaired TET activity[56-58]. Similarly, overall reduction in 5hmC has also been reported in many solid tumor cancers, including breast cancer, prostate cancer, colon cancer and melanoma[39,59-61]. However, increases in 5hmC are also seen in certain types of malignancies, where TET proteins are proposed to carry out an oncogenic function[62,63]. As a result, currently there is no consensus on 5hmC global patterns in disease, and comprehensive studies are needed to describe how 5hmC could be altered in pathological situations, not only at the global level, but in a more gene-specific manner to infer more about mechanistic consequences to disease development and progression.

Therapeutic Targeting of DNA Methylation

Through the study of epigenetics of human disorders we are able to gather large amounts of information, which can be highly relevant in many applications including the determination of diagnostic markers as well as the identification and development of novel therapeutic targets and agents. For example, hypermethylation of tumor suppressor genes, resulting in their silencing, is found to occur in many types of cancers, some of which include cancer of the prostate, bladder, ovary etc. Thus, these genes that are silenced through epigenetic modifications can be used as diagnostic markers, enabling both the screening and prognosis of cancer patients[64]. One class of drugs called DNA methylation inhibitors has been developed to reverse the irregular methylation of various genes. One such example is azacytidine, a nucleoside analogue that can be incorporated into replicating DNA to interfere with the methylation process, shown effective in treating myelodysplastic syndrome and leukemias[65]. These agents can potentially be targeted to tumor suppressor genes, reversing their methylation to reactivate their function in pathological situations[5]. Additionally, another class of therapeutics termed histone deacetylase (HDAC) inhibitors can be used in combination with DNA methylation inhibitors to kill cancer cells in a synergistic manner[65]. Epigenetic therapy may also shed new light on the treatment of neuropsychiatric disorders. For example, a known HDAC inhibitor, valproic acid, is a drug that is developed to treat schizophrenia[66]. Currently, the therapeutic effects of epigenetic drugs are short-lasting and relapse occurs in some cases[5].

Specifically, in prostate cancer, there has been growing interest in the development of epigenetic modulators as therapeutic strategies. Several FDA-approved DNA methylation inhibitors have been studied in prostate cancer treatment. One nucleoside analogue used as DNMT inhibitor, 5-azacytidine, has been tested in Phase II clinical trials. In one completed clinical trial[67], 5-azacytidine was shown to increase the overall median PSA doubling time in 36 PCa patients, and while one patient underwent 30% decrease in PSA, 14 patients had a slight PSA decline. It was observed that *LINE-1*, a repetitive element, showed decreased methylation in plasma. However, grade 3 toxicities including fatigue and neutropenia were reported and 4 patients had to stop treatment. In another clinical trial with metastatic CRPC patients,

more than 50% decline in PSA from baseline was seen in 10 out of 19 evaluable patients, and significant decrease in GADD45A methyation was observed[68]. Although nucleoside analogues such as 5-azacytidine may show promising clinical potential, due to their mechanistic dependence on incorporation into DNA and active DNA synthesis, these drugs are most effective in hyperproliferative cancers and thus may have limited success in solid tumors[69].

In addition, most epigenetic drugs tend to be non-specific regarding their targets. Thus, advancements should be made to increase the effect duration and improve the specificity of these novel therapeutics[5]. In conclusion, the study of epigenetics has cultivated new approaches to examine the etiological factors contributing to disease phenotypes, especially those that exhibit a non-Mendelian inheritance pattern. However, a great deal of complexities regarding the exact mechanisms has yet to be elucidated. Therefore, future expansion of the research in this field will allow us to fully map the epigenome and acquire a more complete picture of the underlying epigenetic events that contribute to disease development. It is expected that with this kind of etiological knowledge in the future, we will be able to screen patients for epigenetic disorders with a higher degree of sensitivity and specificity. Furthermore, careful investigations at the molecular level of the integrative effects of epigenetic processes, DNA sequence variation, and environmental factors are likely to offer useful tools for generating prevention and intervention strategies.

**II. The Role of Chromatin Conformation in Cancer**

Chromatin Organization and Gene Transcription

Closely linked to epigenetic events, the structure of chromatin has been well-known to associate with the status of gene transcription. As early as the 1980s, scientists were able to demonstrate that the mere presence of nucleosomes can inhibit initiation by ribonucleic acid polymerase II (RNAPII) and thus stall transcription[70]. The mechanisms for regulation of the chromatin structure with respect to gene transcription are diverse, and may involve histone displacement, histone variant incorporation,

posttranslational modifications, chromosome territories, and DNA looping[71] (**Figure 1.2**). Each of these mechanisms has its unique influence on chromatin conformation, which in turn dictates gene transcription status.

While packaging of the DNA into nucleosomes can inhibit transcription in vitro, this stereochemical constraint may be relieved by structural changes in nucleosomes[70]. Histones have been observed to exhibit high turnover properties from the core nucleosome. It is reported that histone dimers of H2A and H2B are relatively more susceptible to displacement when compared to H3 and H4[72]. Results from biochemical and genetic studies consistently reinforce the notion that histone eviction from the nucleosome typically occurs at promoters during gene activation, and such process may be mediated by events including but not limited to adenosine triphosphate (ATP)-dependent chromatin remodeling, as well as histone chaperones[73].

For instance, chromatin remodeling complexes, such as switch/sucrose nonfermentable (SWI/SNF)[74-76] and chromatin structure remodeling (RSC) complex[77,78], and additionally active RNAP II[79] can all take part in evicting H2A and H2B to assist nucleosome unraveling. Thus, in a stepwise manner, these chromatin remodeling complexes can mediate repositioning[80] or ejection[81] of nucleosomes at promoters to initiate transcription activation. Moreover, histone chaperone proteins (Asf1[82,83], Nap1[84], and nucleophosmin[85]), which act by sequestering the evicted histones to prevent their reincorporation into the nucleosome, are also an indispensable component for proper histone displacement and ultimately gene transcription.

**Figure 1.2**
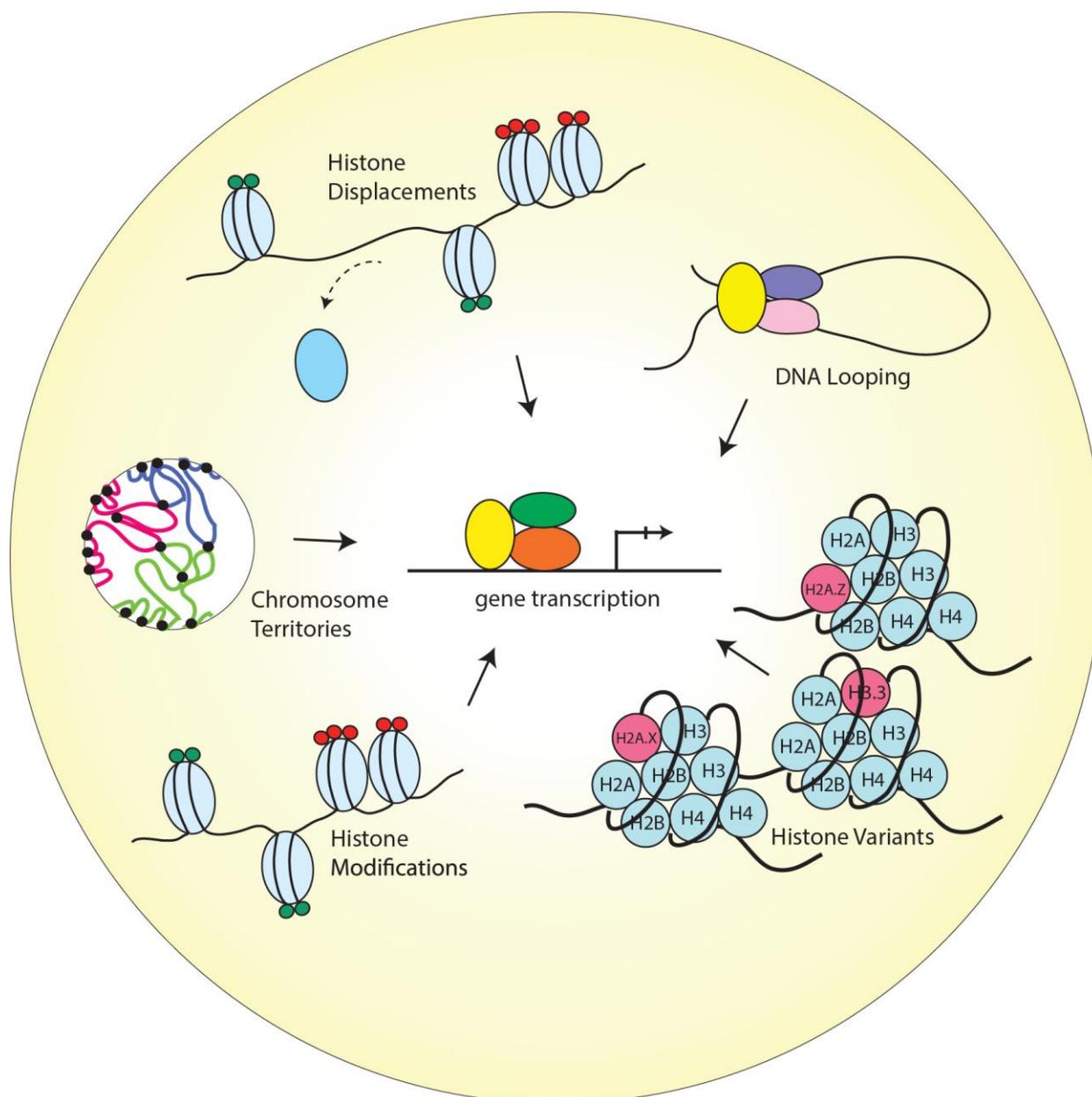
**Figure 1.2 Different chromatin remodeling regulates gene transcription.** Various architectures of the chromatin, histone displacement, DNA looping, histone variants, histone modification, and chromosome territories, regulate gene transcription. This figure was generated by Jung Kim.

In addition to the physical exchange of histone proteins, the incorporation of variant histones can also lead to modifications in chromatin structure and transcriptional regulation. Unlike canonical core histones, generally these unconventional histone proteins are distinguished by the fact that they are expressed outside of the S phase and their deposition into the nucleosome is deemed DNA replication-independent[86].

As a result of changes in their amino acid sequence, variant forms of histones could acquire divergent biophysical properties predisposing them to localize in specific regions of the genome. One prominent histone variant is H2A.Z, which is an alternative form of H2A, and differs from its counterpart in that its N-terminal tail sequence and several key internal residues, which can effectively alter its ability to interact with H2B as well as the H3/H4 tetramer that eventually manifests in reduced nucleosome stability[87,88]. The deposition of H2A.Z is reportedly carried out by ATP-dependent histone exchange reactions through SWR1[89], or by the aforementioned chaperone protein Nap1[90]. Another well-studied histone variant is H3.3, and in spite of the fact that it only differs in four amino acids from its canonical form H3, H3.3 has its distinct deposition pattern where it is preferentially enriched in transcriptionally active chromatin and regulatory sites[91,92].

On the other hand, certain variants, such as macroH2A (mH2A), have the ability to repress gene transcription by remodeling the chromatin to impede RNAPII binding. The name mH2A is derived from the structural feature of this histone variant, which contains a large nonhistone region (NHR), known as the macro domain, on its N-terminus[93]. As a consequence, the NHR of mH2A alters nucleosome structure and interferes with the transcription machinery[94].

Furthermore, a significant category of mechanisms contributing to chromatin organization is posttranslational modifications (PTMs) on histone proteins. There has been extensive research conducted to compile and characterize existing histone modifications, depicting a close relationship between histone PTMs and chromatin structure. Some of the most widely studied histone PTMs include acetylation, methylation, phosphorylation, ubiquitination, and sumoylation. They covalently modify the N-terminal and/or the C-terminal histone tails, while affecting the globular domains at a lesser extent[95].

These various forms of histone marks generate a code that can be interpreted by specialized proteins to regulate gene expression or to mediate DNA repair[96]. Modifications that reflect in active transcription have been elucidated and include acetylation of H3 and H4, and di- or trimethylation of H3 at lysine position 4 (H3K4me2 or me3). In contrast, modifications that instigate inactivation of transcription include methylation at H3K9 and H3K27[71].

In eukaryotes, individual chromosomes can occupy spatially defined territories in the interphase nucleus, and repositioning of these genomic regions has an impact on the regulation of gene expression. FISH analysis has shown that chromosome territories adjoin at their borders to create boundaries between chromatin domains. More recently, it is demonstrated that TADs are enclosed by sharp boundaries enriched for the insulator-binding protein CTCF, as well as the heterochromatin mark H3K9me3[97]. Since boundaries of these topological domains display properties of classical insulator and barrier features, it is therefore suggested that TADs may be linked to transcriptional control.

Concordantly, another study reported that the positions of TADs align with repressive epigenetic marks, as well as lamina-associated domains, and disrupting a TAD boundary can lead to the long-range deregulation in gene expression during X-chromosome inactivation[98]. Therefore, the evidence is convincing that TADs indeed play a role in shaping transcriptional landscapes by clearly defining which sequences belong to the same regulatory network.

Last, as DNA is packaged inside the nucleus, long-range chromatin interactions inevitably occur and – as a result – form loop structures, a majority of which take place between cis-regulatory elements and promoters. It is reported that the dynamic alterations of chromatin looping can either activate or suppress gene expression by facilitating the interactions between enhancers or silencers and their target genes.

One study revealed that only approximately 7% of looping is bridging its nearest gene, reflecting that this chromatin structure is not restrained by genomic proximity and is capable of engaging promoters with distal sites to form complex networks[99]. At the same time, these long-range interactions are not

inhibited by CTCF and cohesin occupancy[99], which argues against previous notions that CTCF's binding to insulator sequences may prevent promoter-enhancer interactions.

Moreover, evidence suggests that the enhancer-promoter loop interactions are formed, in a cell type-specific manner, prior to the binding of transcription factors, indicating their critical role in laying the groundwork for transcriptional control during lineage specification[100]. Furthermore, in terms of thermodynamic properties of DNA looping, it is understood that this mechanism of bringing together multiple components into one functional unit serves to simultaneously increase specificity and affinity and reduce transcriptional noise[101].

## Abnormalities in Chromatin Conformation in Cancer

Due to the crucial role chromatin structure has on determining gene transcription, it is intuitive that chromatin conformation could be manipulated during oncogenic transformation of cancerous cells. It has been demonstrated that under the employment of tumor cells, these chromatin organization machineries become deregulated, disrupting the 3D architecture and undermining the genomic integrity. One of the most recurring phenomena that is associated with cancer development is chromosomal translocations[102]. In the past several decades, a copious number of translocation events have been identified to play pivotal roles in development of a wide range of hematological malignancies as well as solid tumors, which have in turn been utilized as valuable diagnostic and prognostic markers. Aside from chromosomal translocations, a myriad of events have been implicated in cancer, most of which are deviations from the physiological occurrences of chromatin organization discussed previously. Here, we will catalog the most significant aberrations pertinent to chromatin topology that contribute to cancer development, with a particular emphasis on prostate cancer (**Figure 1.3**).

The Philadelphia chromosome is recognized as one the most prominent cancer-associated cytogenetic abnormality that was first reported by Nowell and Hungerford in 1960[103]. It is a highly frequent oncogenic event found in more than 90% of chronic myelogenous leukemia. The translocation is characterized by a reciprocal interchange between chromosome 9 and chromosome 22, which inopportunely generates a BCR-ABL tyrosine kinase gene fusion product[104].

**Figure 1.3**
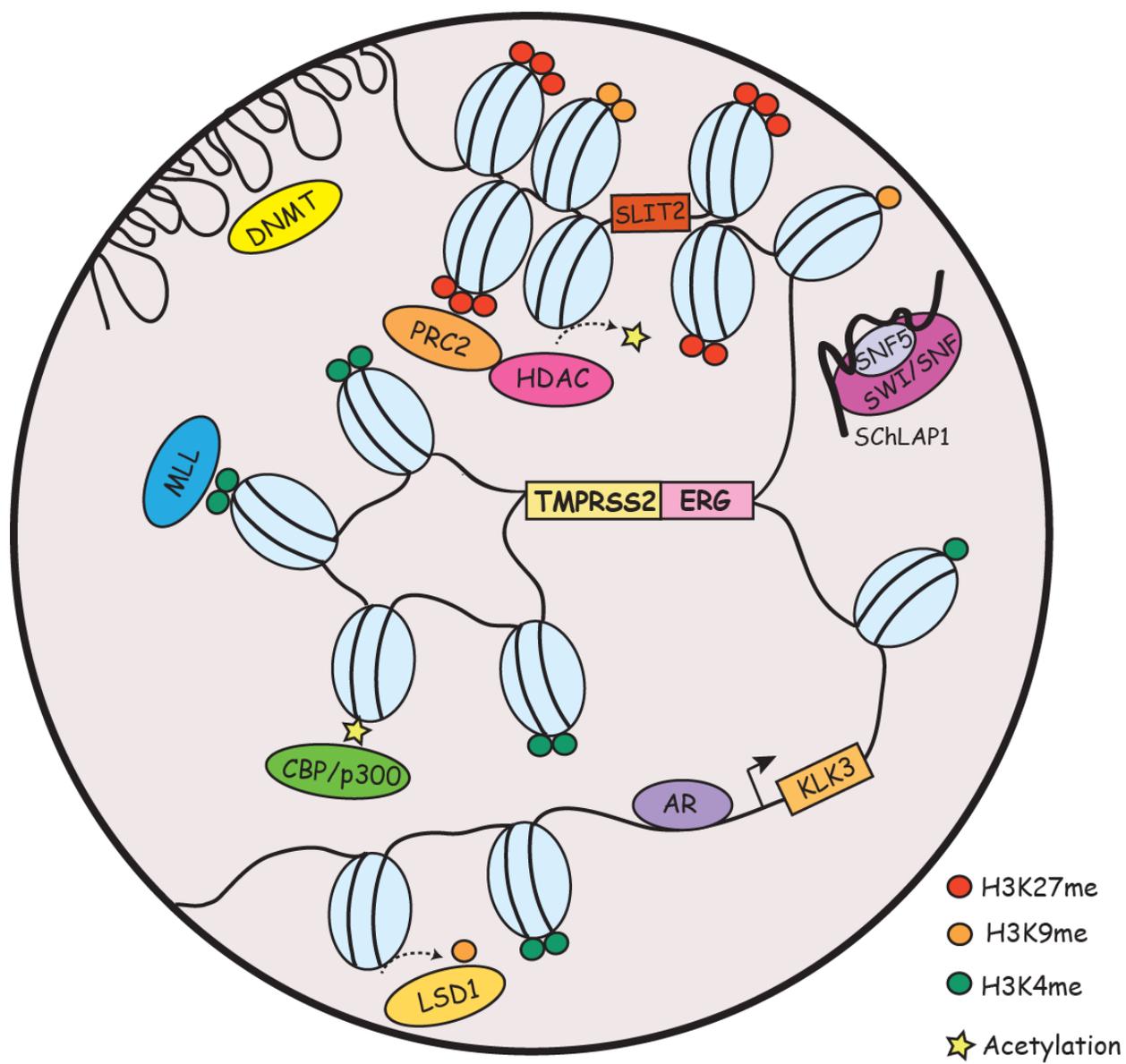
**Figure 1.3 Chromatin organization aberrations in prostate cancer.** Chromatin organizations are altered in prostate cancer through DNA looping, histone PTMs, ncRNAs, and chromosomal translocations, which differentially regulate gene expression. This figure was generated by Jung Kim.

As a result of juxtaposing the breakpoint cluster region (BCR) promoter with the coding region of the ABL gene, the hyperactive BCR-ABL fusion protein confers myeloproliferative properties and leads to leukemogenesis[105]. Clinical successes obtained through pharmacological therapies directly inhibiting the activity of BCR-ABL (eg, imatinib mesylate) have provided a promising paradigm in which chromosomal organization could be a critical target for cancer development and, certainly, cancer treatment.

It was not until recently, however, that chromosomal translocations have been identified in solid tumors. In 2005, Tomlins et al made the breakthrough discovery of the fusion of the TMPRSS2 and ERG genes in prostate cancer[106]. According to their study, a striking proportion of 50% of prostate cancers were found to contain a merged product of the 5′ untranslated region of TMPRSS2(21q22), an androgen-regulated gene, and the protein-coding sequences of ERG (21q22), an erythroblast transformation-specific (ETS) transcription factor (**Figure 1.3**). The TMPRSS2-ERG rearrangement has been confirmed to be present in 36%–78% of prostate cancers[107]. In addition, other members of the ETS family, including ETV1 (7p21), ETV4 (17q21), and ETV5 (3q28), were also uncovered as fusion partners with TMPRSS2 in prostate cancer, but they were detected in lower frequency[108]. Unlike the BCR-ABL translocation, the fusion between TMPRSS2 and ETS genes does not generate a chimeric protein, but instead it promotes the overexpression of oncogenic factors directed by a corrupted promoter element. While solely TMPRSS2 has been identified as a fusion partner of ERG, other 5′ partners of ETS genes have also been observed. These include androgen-induced genes SLC45A3, KLK2, CANT1, and NDRG1, and an endogenous retroviral element HERV-K_22q11.23, which are functionally comparable to TMPRSS2, as well as androgen-repressed gene C15orf21[109-111]. It was also reported that rearrangements in the rapidly accelerated fibrosarcoma (RAF) pathway also occur in advanced prostate cancer (SLC45A3-BRAF, ESRP1-RAF1), which can be targeted by RAF kinase inhibitors[112]. Moreover, a recent study was able to identify a median of 90 rearrangements in seven prostate cancer tumor samples[113]. Examples of disrupted genes due to rearrangement include CADM2, which is a cell adhesion molecule, and phosphatase and tensin homolog (PTEN), a well-established tumor suppressor, as well as a PTEN-interacting protein, MAGI2. These findings depict a

convoluted network of genomic rearrangements and chromatin conformation, which synergistically confer deregulated gene expressions and contribute to tumorigenesis.

In addition to chromosomal translocations, modifications to histone could also place a huge impact on the 3D structure of chromatin and has been widely implicated in cancer. In prostate cancer, H3K4 methylation and H3K27 methylation are among the most extensively investigated histone PTMs; while the former is generally associated with activation of proto-oncogenes, the latter is associated with silencing of tumor suppressors. The repressive epigenetic PTM, H3K27 trimethylation (H3K27me3), has been found to be significantly enriched in promoters of numerous tumor suppressor genes (eg, ADRB2[114], SLIT2[115], DAB2IP[116,117], etc.), in metastatic prostate cancer. Meanwhile, H3K9me1 and me2, generally accompanied by heterochromatin assembly[118], are also implicated in prostate cancer. Demethylation of H3K9 has been reported to reflect in derepression of AR-regulated genes[119]. H3K4 mono- and dimethylation (H3K4me1, H3K4me2) have been thought of as markers for enhancer sites in directing the androgen receptor (AR) transcriptional program, by facilitating AR binding directly or indirectly through the recruitment of coactivators, such as FOXA1, GATA2, and MED1[120]. Moreover, an endeavor combining high-resolution nucleosome positioning with histone marks mapping showed strong evidence that H3K4me2-containing nucleosomes spaced 250–450 bp (base pair) apart can flank binding sites of AR prior to its ligand-mediated activation, while the binding site is occluded by a well-positioned nucleosome. Following AR activation, nucleosomes with altered H3K4me2 marks become destabilized at AR binding sites and are comparably more stable at the two flanking loci[121].

In addition, the study revealed that the labile H2A.Z variant was more likely to be present in the central nucleosome relative to the flanking nucleosomes, which further contributes to reduced stability of the nucleosome occupied at the AR binding site. Also, it has been shown that androgen treatment can increase the level of H2A.Z and that the incorporation of H2A.Z in enhancer and proximal promoter sites of the AR-induced gene prostate-specific antigen (PSA; or KLK3) can poise the gene for activation by AR[122].

Established and maintained by protein–protein interaction between transcription factors bound at enhancers and at promoters[123], DNA looping and chromatin compartmentalization are essential

processes governing gene transcription; hence, they are a frequent target for disruption during cancer development. In the case of prostate cancer, AR-mediated chromatin looping has been a longtime research interest in the field, and extensive efforts have been devoted to elucidate the process of how AR signaling may lead to changes in chromatin conformation during prostate tumorigenesis. Studies using chromatin immunoprecipitation (ChIP) techniques showed a striking feature of AR genome-wide binding pattern that, approximately 86%–95% of AR localization occurs in non-promoter regions[120,124]. This evidence strongly indicates that AR, as a transcription factor, is able to direct its specific transcriptional program from a distance – sometimes, even hundreds of kilobases –away from its target gene. Therefore, it is plausible to presume that a looping model is the mechanism by which AR can regulate its targets from afar. In fact, this model has been proven to be true through 3C-based assays, which demonstrated that distal AR enhancer regions form long-range physical contacts with transcription start sites of AR-regulated genes, such as PSA and TMPRSS2[125,126], as well as UBE2C, which is a critical enzyme involved in promoting growth of castration-resistant prostate cancer[120,127].

**III. FOXA1 is a Pioneer Factor and Often Deregulated in Prostate Cancer**

The forkhead box A1 (FOXA1; previously termed as hepatocyte nuclear factor 3α, HNF-3α) protein belongs to a superfamily of winged helix transcription factors[128,129]. The name of "forkhead box" gene family is originally derived from a prominent phenotypic feature of developmental defects observed in Drosophila with the *fork head* gene mutant, which manifests in the foregut and hindgut being replaced by ectopic head structures[130]. Like other forkhead (FKHD) family proteins, FOXA1 controls gene transcription by directly binding to its consensus sequence, the FKHD motif. In addition, FOXA1 has been shown capable of opening surrounding chromatin and subsequently allowing other transcription factors, such as androgen receptor (AR), to come in close proximity to their target sites and thus exert transcriptional control of gene expression[131-134]. Although this transcription regulatory effect of FOXA1 is quite well understood, important new developments have been made recently concerning the functional roles of FOXA1 in prostate cancer.

FOXA1 in Development

FOXA1 was initially discovered approximately 25 years ago as an important liver-enriched transcriptional regulator of hepatic differentiation, since it was found to occupy the promoters of liver genes α1-antitrypsin and transthyretin[135]. Subsequent mouse studies have shown that Foxa1 expression can be observed in endoderm-, mesoderm- and ectoderm-derived tissues of adult mice[136]. It has been reported that detectable Foxa1 mRNA could first be observed at E7 in the late primitive streak stage in the midline endoderm of mouse embryos, following that the expression could be seen in the notochord, neural plate and floor plate of the neural tube, indicating that Foxa1's roles can range from establishment of definitive endoderm to formation of neural tube patterning[137-139].

Although Foxa1 null mice don't exhibit discernible morphological defects, they display severe growth retardation and die between postnatal days 2 and 14 (P2 and P14), which is resulted from a combination of phenotypes including dehydration and hypoglycemia[140,141]. Therefore, these observations indicate that FOXA1 plays a pivotal role in the maintenance of glucose homeostasis and pancreatic islet function. Tissue-specific deletion of *Foxa1* in the pancreas shows that FOXA1 and FOXA2 jointly regulate the expansion of pancreatic primordial, specification of endocrine and exocrine compartments, and maturation of islet cells[142]. Similarly, there is also evidence that FOXA1 is important for lung development by regulating respiratory epithelial differentiation[143], and that it acts in a complementary manner with FOXA2 to ensure proper branching morphogenesis of the lung[144]. Moreover, it has been demonstrated that both FOXA1 and FOXA2 in conjunction are required for initiating the onset of hepatogenesis and hepatic specification[145]. More recently, a study utilizing conditional knockout of *Foxa1* and *Foxa2* in dopamine neurons reports that both factors are required for dopamine neuron maintenance and that their loss can give rise to locomotor deficits resembling the manifestations of Parkinson's disease[146]. Taken together, mice studies corroborate the notion that FOXA1 has critical influence on organogenesis.

In particular, a number of papers have demonstrated the significance of FOXA1 during development of the prostate and mammary glands. It has been said that the mammary ductal morphogenesis, but not the alveolar lineage, is dependent on FOXA1, and that while Foxa1-null glands

can form milk-producing alveoli, they have lost ERα expression and functional activity, which ultimately result in compromised ductal lineage specification[147]. Likewise, in the prostate, FOXA1 deficiency leads to abolished differentiation and maturation of luminal epithelial cells[148]. Initially derived from the hindgut endoderm, the mouse prostate epithelium has persistent Foxa1 expression throughout the processes of prostate development, growth, and adult differentiation[149]. The origin of the prostate is the urogenital sinus, which is a midline structure composed of an endoderm-derived epithelial layer and a mesoderm-derived mesenchymal layer[150]. In the mouse, at approximately E17.5, prostatic morphogenesis starts to take place, prompted by responsiveness to circulating androgens and induction of AR activity[150]. During the course of development, Foxa1 expression was characterized in all lobes of the murine prostate, and is specifically enriched in AR-expressing epithelial cells. FOXA1 plays a critical role in modulating AR-regulated transcriptional signaling in prostate epithelial cells[133], and concordantly Foxa1-deficient prostate has severely impaired ductal pattern formation, due to inhibition of ductal canalization and epithelial cytodifferentiation[148]. As a consequence, the Foxa1-null prostate lacks structural maturity as well as secretory activities. Taken together, there is compelling evidence that FOXA1 is critically involved in growth and differentiation of prostatic cells and is required for prostate glandular morphogenesis.

FOXA1 Deregulation in Prostate Cancer

As FOXA1 is highly involved in developmental processes and lineage specification in several organs, when expressed at aberrant levels it may disrupt normal physiological events and lead to formation of cancer. Molecular and genetic studies have shown that FOXA1 is often found to be abnormally expressed in a number of cancer types, including acute myeloid leukemia (AML), lung, esophageal, thyroid, breast and prostate cancers[151-157]. At present, the prevailing views on FOXA1 expression in prostate cancer have not reached a consensus, with contrasting evidence seen in different cohorts of cancer patients. Analyses of human prostate cancer specimens have revealed that FOXA1 is overexpressed in metastatic as well as castration-resistant prostate cancer (CRPC) patients, but its expression is lower in normal and neoplastic transitional zone tissues[158]. In addition, the level of FOXA1 may be positively correlated with conventional parameters indicative of cancer progression (including

tumor stage and Gleason scores), and negatively correlated with relapse-free survival times[157,158]. In other words, high FOXA1 level is associated with poor prognosis. However, other studies have also demonstrated that low FOXA1 levels are found in metastatic and CRPC tumors and may in fact denote unfavorable prognostic outcome in advanced prostate cancer[159,160]. In order to reconcile these conflicting findings, the function of FOXA1 should be carefully dissected with respect to cellular context, taking into consideration the status of AR program and androgen responsiveness, to fully understand how FOXA1 may fit as a piece of jigsaw in the prostate cancer puzzle.

In addition to deregulation at expression level, mutations in the FOXA1 gene have also been uncovered in prostate tumors (**Figure 1.4**), as reported in recent literature[161] as well as in TCGA (The Cancer Genome Atlas). Recurrent FOXA1 gene mutations had recently been identified and characterized in 5 of 147 prostate cancers, including both localized as well as castration-resistant cases[162]. Moreover, 4 of these 5 mutations are located in the C-terminal transactivation domain, and mutated FOXA1 was demonstrated to repress androgen signaling and augment tumor growth[162]. Another independent study also reported 3 different non-silent mutations residing in or close to the forkhead domain in FOXA1[163], which are anticipated to disrupt DNA binding, but to what extent and how it may be related to prostate carcinogenesis will require further studies. Moreover, a recent study reported that, by adopting the methodology of 3D organoid culture system, the genetic heterogeneity of prostate cancer could be recapitulated, and FOXA1 was among the most mutated genes in the organoid CRPC lines[164].

Another level of FOXA1 deregulation in prostate cancer is reflected in somatic mutations of its cis-regulatory elements, which in turn affects FOXA1 transcriptional activity. It has been described that in prostate cancer there exists single nucleotide polymorphism (SNP) within the consensus forkhead motif, which is recognized and bound by the FOXA1 protein (**Figure 1.4**). A prevalent SNP was identified in LNCaP prostate cancer cell line, which locates in the proximal promoter of the gene encoding UDP glucuronosyltransferase 2B17 (UGT2B17)[165]. This G to A polymorphism, UGT2B17 – 155 G/A, also appearing in NCBI SNP database as rs59678213, was shown to have a notable impact on FOXA1 binding, with the A-containing allele being 13-fold more active in luciferase assays[165].

**Figure 1.4**
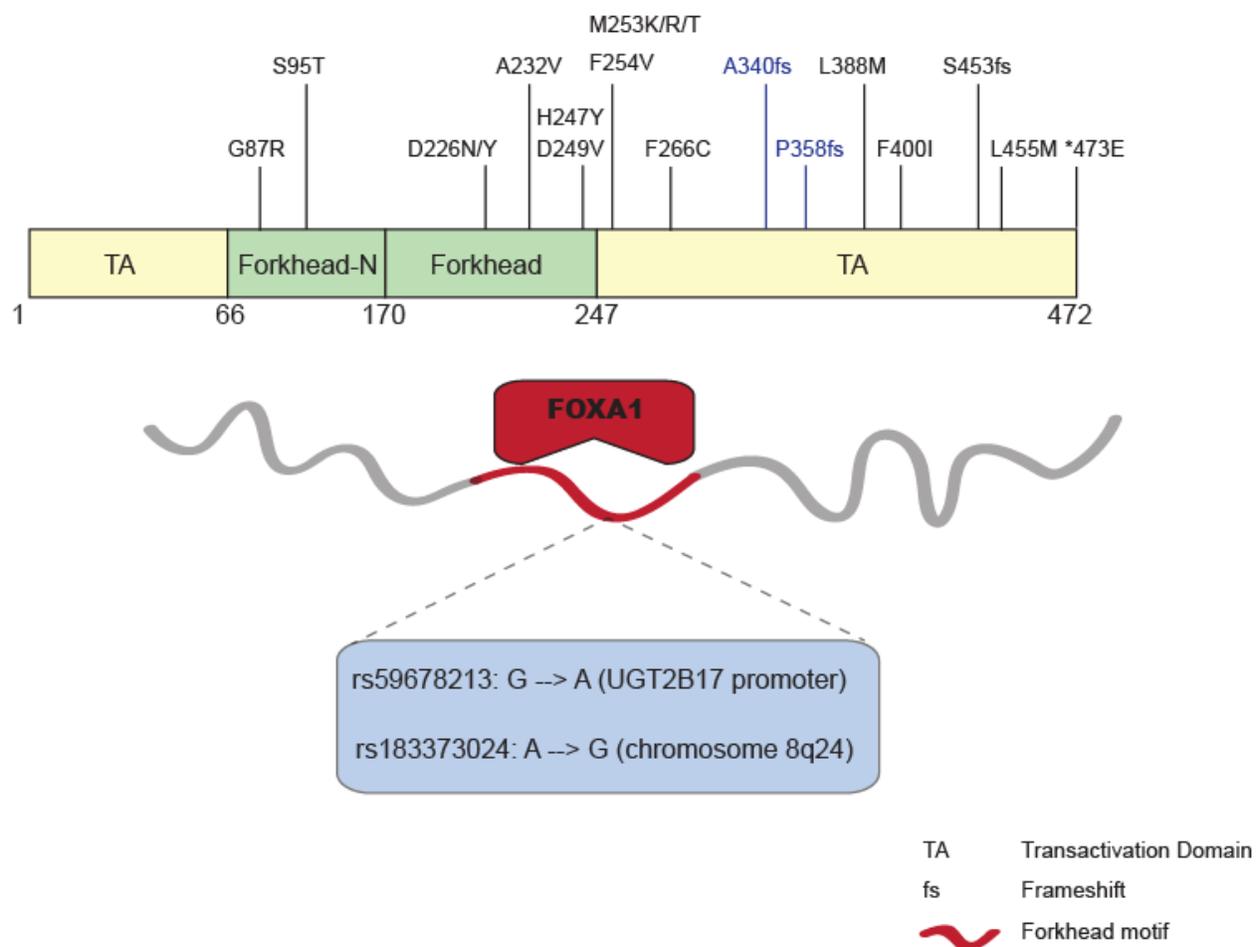
**Figure 1.4 FOXA1 mutations in prostate cancer.** Somatic mutations of the FOXA1 gene that have been identified in localized and CRPC tumors, as well as in prostate cancer cell lines (the latter shown in blue, namely F266C, A340fs, P358fs). SNPs associated with prostate cancer risk have also been reported occurring within the consensus sequence of FOXA1 binding motifs.

Moreover, another study revealed an SNP significantly associated with risk for prostate cancer in the chromosome 8q24 region (rs183373024), where it disrupts the FOXA1 recognition motif[166]. As ChIP-seq data have reported AR and FOXA1 binding at this region in cell lines, it is predicted that this particular SNP in prostate cancer may cause disruption of FOXA1 and/or AR binding and thus lead to deregulation of some tumor suppressor genes. Several candidate genes were presented to be potential targets for this putative SNP-containing FOXA1 enhancer, however chromatin conformation capture (3C)-based techniques or CRISPR assays may be needed to assure the target gene.

Role of FOXA1 in regulating prostate cancer cell growth and motility

Overall, both oncogenic and tumor-suppressive roles have been reported for FOXA1, which suggests that its precise contribution to cancer development or progression may be depended on disease stage, context, and treatment histories. In conjunction with AR signaling in the presence of androgen, FOXA1 is known to promote prostate cancer proliferation by inducing expression of cell cycle genes[159,167]. However, under androgen-depleted conditions, FOXA1 was shown to rather inhibit cell proliferation and its loss led to androgen-independent prostate cancer cell growth, being consistent with its regulation of AR signaling[168]. In support with this tumor suppressor role, in mice with prostate-specific *Foxa1* gene deletion, progressive hyperplasia can be observed, and *Foxa1* knockout epithelial cells exhibit increased proliferation and altered morphology[169]. Further, following castration, the number of *Foxa1*-positive cells was significantly reduced, supporting *Foxa1* loss as a potential mechanism to castration resistance. Thus, like its modulation of AR signaling, FOXA1 regulation of prostate cancer cell growth is context-dependent.

Through analyses of genome-wide gene expression profiling, it has been discovered that FOXA1 may also possess AR-independent functions in inhibiting cell motility and epithelial-to-mesenchymal transition (EMT)[159]. In prostate cancer cells lacking AR expression, ectopic introduction of FOXA1 is sufficient to impede cell invasion and migration[159]. On the other hand, loss of FOXA1 in LNCaP cells increases cell invasiveness, both in androgen-containing and -deprived conditions. Both cases demonstrated the AR-independent function of FOXA1 in inhibiting prostate cancer cell motility.

Meanwhile, it is also found that FOXA1 can negatively regulate EMT, and loss of FOXA1 in LNCaP cells results in an astrocyte-like, fusiform, or fibroblastic phenotype characteristic of mesenchymal and neuroendocrine cells. Further analysis revealed that among direct transcriptional targets of FOXA1, *SLUG* was identified to be a key repressed gene that confers the anti-motility properties associated with FOXA1[159]. Similar functions of FOXA1 in preventing metastasis have been reported in other forms of cancer as well, such as lung cancer and pancreatic cancer[170,171], supporting the idea that this anti-EMT role is AR-independent. Being concordant with these functionalities, expression profiling datasets of prostate tumors confirm that FOXA1 is upregulated from benign tissue to localized tumor, but downregulated in metastatic CRPC tumors compared to localized ones[159,160]. However, there exists contrasting histological evidence that FOXA1 level is high in metastatic prostate cancer[172]. FOXA1 expression level may need to be more carefully looked at taking into considerations of disease stage, hormone deprivation treatment history, and relative AR level.

FOXA1 Defines Prostate Lineage-Specific AR Cistrome

Like other forkhead proteins, FOXA1 encompasses a winged helix domain that is composed of three α-helices, three β-sheets and two loops[173]. This unique structure, which closely resembles that of linker histones[174], has imparted to FOXA1 the ability of binding to highly compacted chromatin and subsequently prying it open[175]. In doing so, FOXA1 creates an open and easily accessible chromatin conformation to facilitate hormonal transcription factors such as estrogen receptor (ER) and AR to bind their cis-regulatory elements[176]. Comparison of FOXA1 cistromes between breast and prostate cancer cells illustrates very distinct, lineage-specific profiles, where less than 40% of binding sites are shared in between[177]. And yet, FOXA1 occupies a majority of the binding sites of AR and ER in prostate and breast cells, respectively, suggesting that FOXA1 may be critical in determining lineage-specific hormonal factor chromatin-targeting. It has been shown that FOXA1 is essential for AR-mediated prostatic gene activation, corroborated by the facts that FOXA1 regulatory elements are found in the core enhancer of the prostate-specific antigen (PSA) gene, a prototypical AR target, adjacent to the androgen response elements (AREs), and that perturbations in the FOXA1 motif can significantly abolish induction of PSA by androgens[133]. Through bioinformatic and biochemical analyses, it has been discovered that the FKHD

motif is enriched in AR cistromes[126,178], and that FOXA1 can physically interact with AR[133]. Thus, upon stimulation by androgens, AR translocates into the nucleus and preferentially binds cis-regulatory sequences that are largely pre-occupied by FOXA1, potentially under the recruitment by the FOXA1 protein. Expression profiling studies showed that FOXA1 indeed positively regulates prostatic gene expression induced by androgen[168].

The mechanisms by which FOXA1 recognizes lineage-specific enhancers have also been investigated. It has been reported that FOXA1 binding can be guided by specific chromatin marks, namely mono- and di-methylation of histone 3 lysine 4 (H3K4me1, me2)[177]. Both are epigenetic signatures typically associated with enhancers, H3K4me1 and H3K4me2 genomic distribution was thought to provide a blueprint for directing differential FOXA1 binding in a lineage-specific manner. Furthermore, DNA methylation has also been shown to play a part in defining FOXA1 binding and enhancer activation. Genome-wide interrogation of DNA methylation reveals that FOXA1-bound enhancers are generally hypomethylated compared to juxtaposing genomic regions, in a pattern which is also correlated with cell type-specific FOXA1 binding[179]. To better understand the sequence of events occurring at FOXA1-activated enhancers, kinetics study showed that binding of FOXA1 to chromatin could be detected prior to significant induction of H3K4me2 and DNA demethylation, which suggests that hypomethylation may be an epigenetic phenomenon succeeding FOXA1 binding rather than a pre-established mark[179]. Although DNA hypomethylation does not seem to be required for FOXA1 binding, it is important for turning on FOXA1 transcriptional activity, as shown in luciferase reporter assays which exhibited reduced enhancer activation when constructs were methylated[179].

With the current knowledge in mind, we began our quest to delve further into the epigenetic events underlying FOXA1's recruitment to hypomethylated DNA regions, and to identify novel players which may play a role in the process. We hope the findings will provide more insights on molecular mechanisms utilized by FOXA1 to carry out its transcriptional activities and shed light on how epigenetic changes could be important for transcription factor occupancy and enhancer activation.

**CHAPTER 2: FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function**

**I. Introduction**

FOXA1 is critical in directing hormone receptor-dependent transcriptional programs to regulate prostate- or breast-specific gene expression and cell differentiation[147,148]. FOXA1 acts as a 'pioneer transcription factor' that can associate with compact chromatin to increase local chromatin accessibility and facilitate the recruitment of other transcription factors including nuclear receptors to these sites[180]. Genome-wide location analyses have reported that FOXA1 preferentially recognizes and binds lineage-specific enhancers that are demarcated by active histone modifications including histone H3 lysine 4 mono- and di-methylation (H3K4me1, me2)[177], histone 27 acetylation (H3K27ac)[181], as well as local DNA hypomethylation[179]. On the other hand, enforced expression of FOXA1 and its subsequent recruitment to enhancers lead to DNA demethylation and *de novo* gain of H3K4me1, suggesting that FOXA1 is able to remodel heterochromatic regions[179,182]. However, the molecular mechanisms by which FOXA1 imposes this chromatin remodeling have not been characterized.

As discussed earlier, through catalyzing DNA demethylation, TET proteins play important roles in embryonic stem cell maintenance and in regulating appropriate lineage differentiation of these cells. These activities can be linked to the ability of DNA demethylation in modulating transcription factor occupancy and *vice versa*[183,184]. During neural and adipocyte differentiation, dynamic hydroxmethylation has been associated with lineage-specific distal regulatory regions and represents an early event of enhancer activation[42]. Concordantly, a separate study has demonstrated that deletion of Tet2 led to extensive loss of 5hmC and gain of DNA hypermethylation at enhancers and modulates enhancer activity of differentiation-related genes[185]. However, the roles of TET proteins in FOXA1 recruitment and regulation of prostate lineage-specific enhancers are yet to be delineated.

Here, we show that TET1 is a direct target of FOXA1-mediated transcriptional activation. Further, TET1 physically interacts with the FOXA1 protein and modulates local DNA demethylation that in turn facilitates and stabilizes the recruitment of FOXA1. FOXA1 and TET1 thus form a feed-forward loop that activates lineage-specific enhancers. Not only does this mechanism provide a new perspective on the dynamic functional significance of the newly discovered TET1 DNA hydroxylase, but also offer insight into the molecular details underlying FOXA1's ability to fine-tune and modulate lineage-specific enhancer activation. As FOXA1 is a critical regulator and a top mutated gene in multiple cancers such as breast and prostate cancers[161], our study thus forms the framework for future understanding of the roles of TET1 in lineage-specific gene expression and cancer progression.

**II. Results**

FOXA1 expression contributes to lineage-specific enhancer activation

To determine the correlation between FOXA1 and active enhancer marks, we re-analyzed previously published FOXA1 (GSE37345), H3K4me2 and H3K27ac ChIP-seq data (GSE27823)[160,168] and confirmed that FOXA1 binding sites (FXBS) are indeed enriched for H3K4me2 and H3K27ac (**Figure 2.1A**). Further, we performed MeDIP for 5mC and chemical labeling of 5hmC followed by deep sequencing, namely MeDIP-seq and hMe-Seal-seq[30,186], respectively, to map their genomic landscapes in LNCaP cells which express FOXA1. Bioinformatic analysis revealed that FXBS are depleted of 5mC, but enriched for 5hmC, being consistent with previous reports[179] (**Figure 2.1**). In addition, we found that this correlation was much weaker in two other prostate cell lines namely PrEC and PC-3M, wherein FOXA1 expression is low, suggesting that FOXA1 expression and occupancy might contribute to DNA demethylation at local chromatin (**Figure 2.2**). Since it has been previously suggested that transcription factor binding sites can demonstrate the low 5mC high 5hmC signature in embryonic stem cells[184], we looked at DNA methylation profiles in LNCaP cells for two other transcription factors CTCF and AR and observed similar patterns for 5mC and 5hmC (**Figure 2.3**). As a measure of negative control, genomic regions 20 kb downstream from the FOXA1 peaks, which will be referred to as non-peak sites throughout this paper, were examined for epigenetic signatures but did not exhibit any distinct pattern (**Figure 2.4**).

To further elaborate on this, we depleted FOXA1 in LNCaP cells through lentiviral shRNA transduction (**Figure 2.5A**) and performed pulldown and deep sequencing of 5mC and 5hmC. Interestingly, although the average intensity of 5mC around all FOXA1-occupied sites was not hugely affected upon FOXA1 depletion (**Figure 2.1B**), there was a significant decrease in 5hmC (**Figure 2.1C**), whereas no changes were seen in either 5mC or 5hmC for non-peak sites (**Figure 2.4B-C**). Concordantly, active enhancer marks H3K4me2 and H3K27ac were decreased around FXBS following FOXA1 knockdown, supporting reduced enhancer activities (**Figure 2.5B-C**).
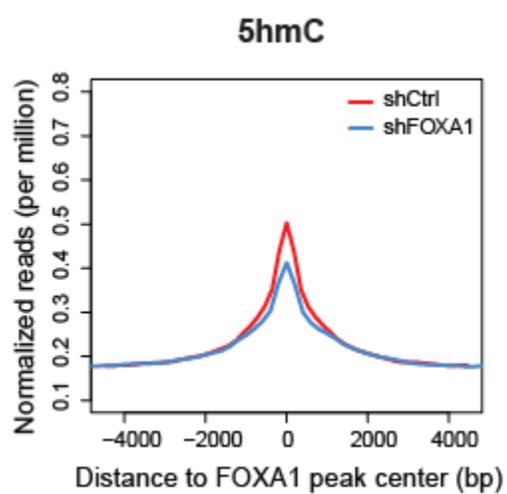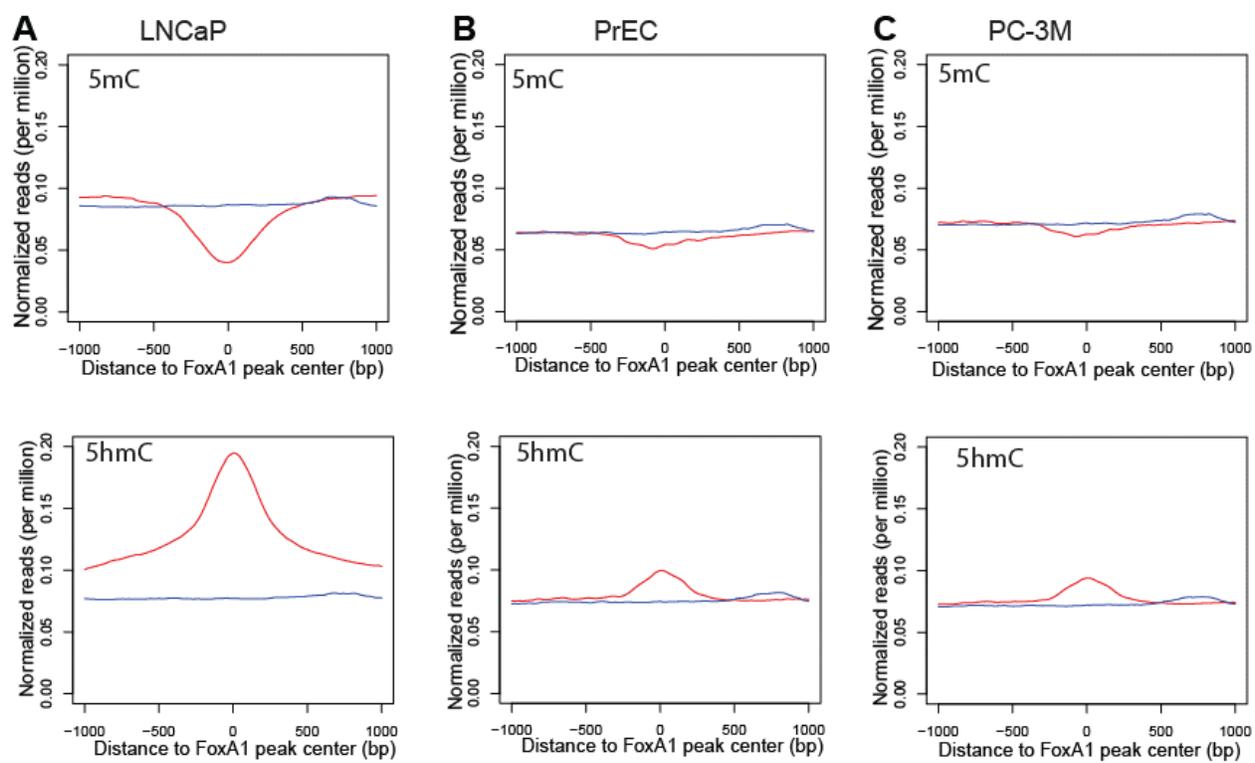
**Figure 2.1**

**A.**



**B.**



5mC

**C.**
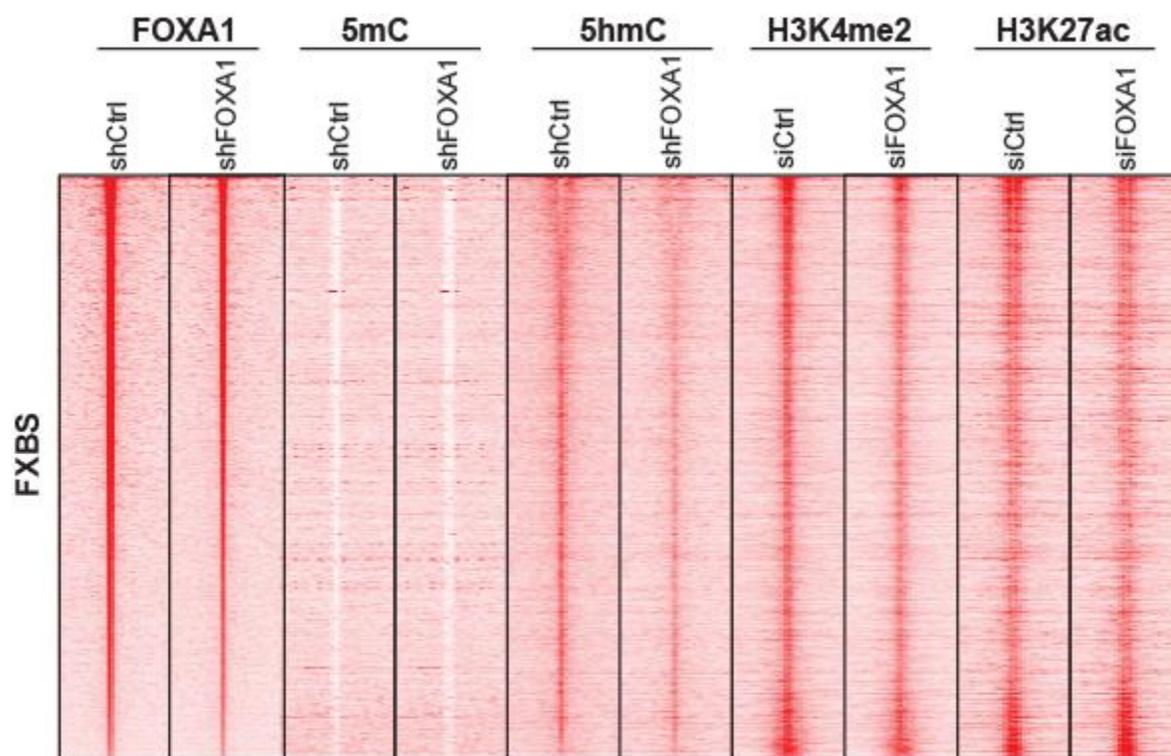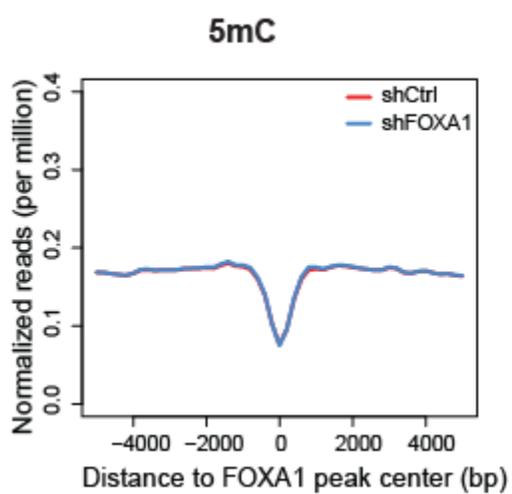


5hmC

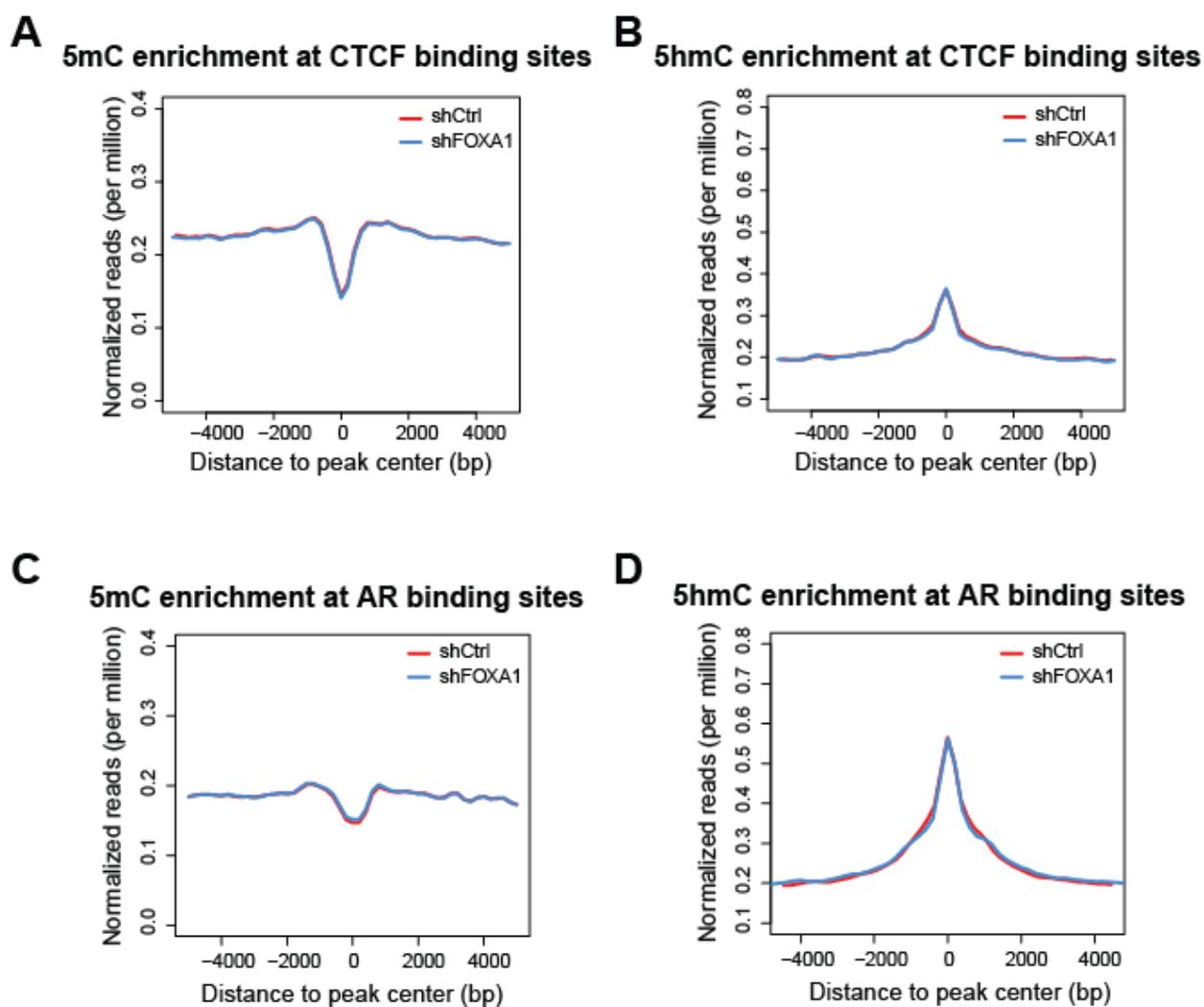**Figure 2.1 FOXA1 contributes to enhancer activation through epigenetic modifications.**

(**A**) Epigenetic signatures of FOXA1 binding sites (FXBS) in control and shFOXA1 LNCaP cells. FOXA1 and H3K4me2/H3K27ac ChIP-seq data were obtained from publicly available datasets GSE37345 and GSE27823, respectively. Genomic landscapes of 5mC and 5hmC were determined by MeDIP and hMe-Seal, respectively, followed by deep sequencing. ChIP-seq read intensities of indicated epigenetic marks around (±5 kb) FXBS or non-peak regions in control (shCtrl) and FOXA1-knockdown (shFOXA1) cells were presented in heatmap format, ranked by read intensity of FOXA1 occupancy. (**B** and **C**) Average intensity plots of 5mC (B) and 5hmC (C) enrichment around all FXBS shown in A. These figures were generated by Jonathan Zhao.

**Figure 2.2**

**Figure 2.2 Epigenetic signatures at FOXA1 binding sites.** Intensity plots showing 5mC and 5hmC

enrichment around FOXA1 binding sites (±1 kb) in LNCaP (A), PrEC (B), and PC-3M cells (C). 5mC and

5hmC chemical labeling, or TAmC and hMe-Seal were performed using genomic DNA extracted from

LNCaP, PrEC and PC-3M cell lines. Enriched DNA was made into libraries and subjected to deep

sequencing. The read intensities of TAmC- and hMe-Seal-seq in different cell lines were evaluated

relative to FOXA1 binding sites in LNCaP cells. These figures were generated by Jonathan Zhao.
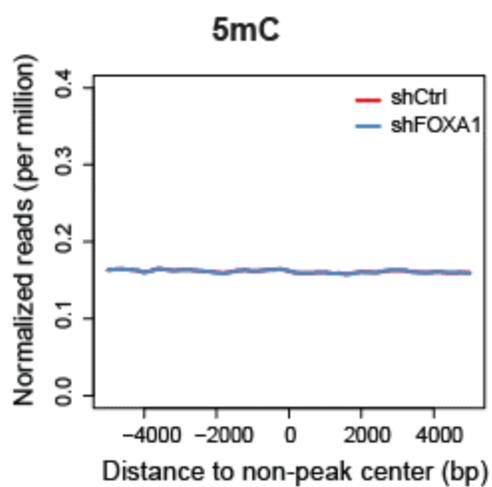
**Figure 2.3**



**A** 5mC enrichment at CTCF binding sites

**B** 5hmC enrichment at CTCF binding sites

**C** 5mC enrichment at AR binding sites

**D** 5hmC enrichment at AR binding sites

**Figure 2.3 Epigenetic signatures at CTCF and AR binding sites.** Intensity plots showing 5mC and

5hmC enrichment around binding sites (±4 kb) of CTCF (A-B) and AR (C-D) in LNCaP. CTCF peaks were

obtained from publicly available dataset GSM947528, and AR peaks were obtained from GSM353644

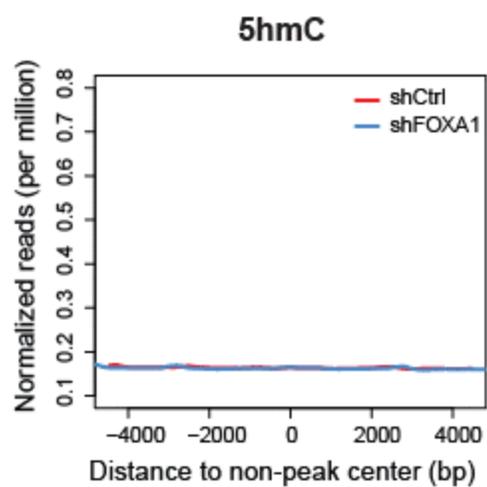previously published by our lab. These figures were generated by Jonathan Zhao.
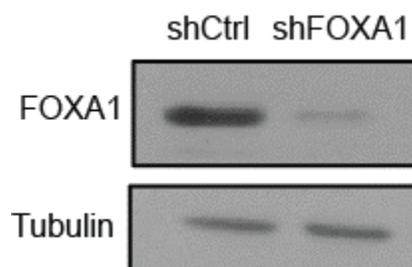
**Figure 2.4**

**A.**



**B.**



5mC

**C.**
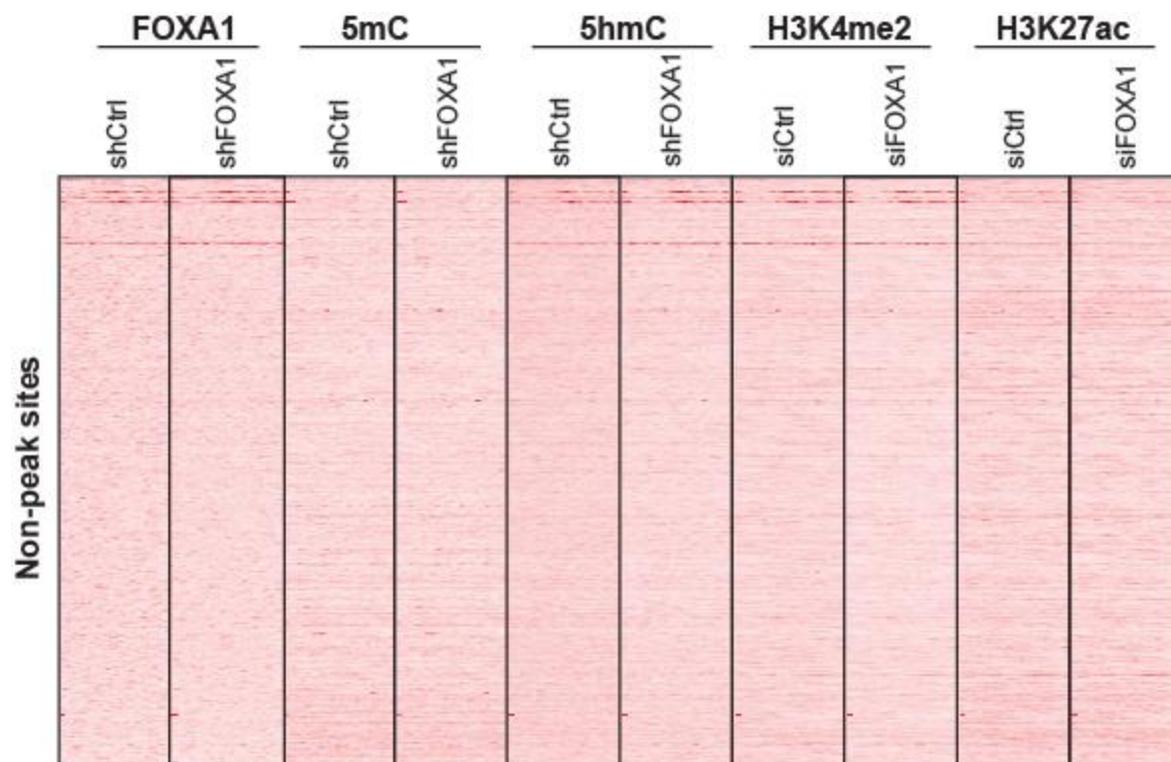


5hmC

**Figure 2.4 Epigenetic marks at non-peak control regions.** Epigenetic signatures of non-peak sites

taken 20 kb downstream (**A**) in control and shFOXA1 LNCaP cells. FOXA1 and H3K4me2/H3K27ac

ChIP-seq data were obtained from publicly available datasets GSE37345 and GSE27823, respectively.

Genomic landscapes of 5mC and 5hmC were determined by MeDIP and hMe-Seal, respectively, followed

by deep sequencing. ChIP-seq read intensities of indicated epigenetic marks around non-peak regions in

control (shCtrl) and FOXA1-knockdown (shFOXA1) cells were presented in heatmap format. (**B** and **C)**

Average intensity plots of 5mC (**B**) and 5hmC (**C**) enrichment around all non-peak sites shown in

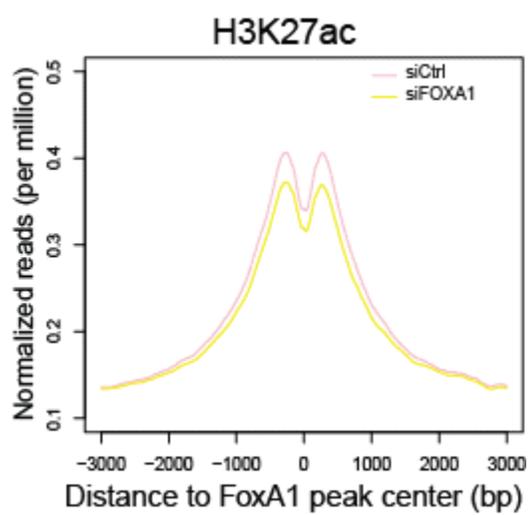A. These figures were generated by Jonathan Zhao.

**Figure 2.5**

**A.**
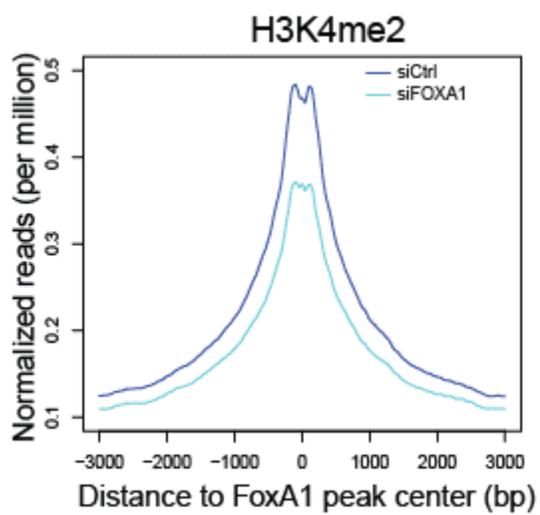


**B.**                                                        **C.**

**Figure 2.5 FOXA1 knockdown affects active enhancer histone marks.** (**A**) Confirmation of FOXA1 knockdown by western blot. LNCaP cells were infected with shCtrl or shFOXA1 lentivirus, and protein lysates were subjected to western analysis using anti-FOXA1 and anti-Tubulin.

(**B, C**) Intensity plots of H3K4me2 (**B**) and H3K27ac (**C**) enrichment around FOXA1 binding sites in control and FOXA1-depleted LNCaP cells. ChIP-seq results for active enhancer marks H3K4me2 and H3K27ac were obtained from publicly available datasets (GSE27823). Enrichment of both histone marks around FOXA1-occupied sites (±3 kb) is shown for control and FOXA1 knockdown LNCaP cells. B and C were generated by Jonathan Zhao.

To ensure the reliability of this genome-wide phenomenon, as well as to examine the changes with a more sensitive method, we performed MeDIP and hMe-Seal followed by qPCR for individual genes. Expectedly, 5hmC was greatly reduced across a number of FXBS (**Figure 2.6A**). On the other hand, despite the fact that 5mC showed no obvious change on a global scale, MeDIP-PCR revealed moderate increases in 5mC upon FOXA1 knockdown (**Figure 2.6B**). Taking into consideration that 5hmC abundance represents only ~10% of 5mC in embryonic stem cells[10], it is reasonable to observe a more significant change in 5hmC rather than 5mC. It can be inferred from these results that FOXA1 may be functioning to alter DNA methylation specifically at regions where it occupies to achieve a demethylated state while accumulating 5hmC marks, thus potentiating enhancer activation.

**Figure 2.6**

**A.**
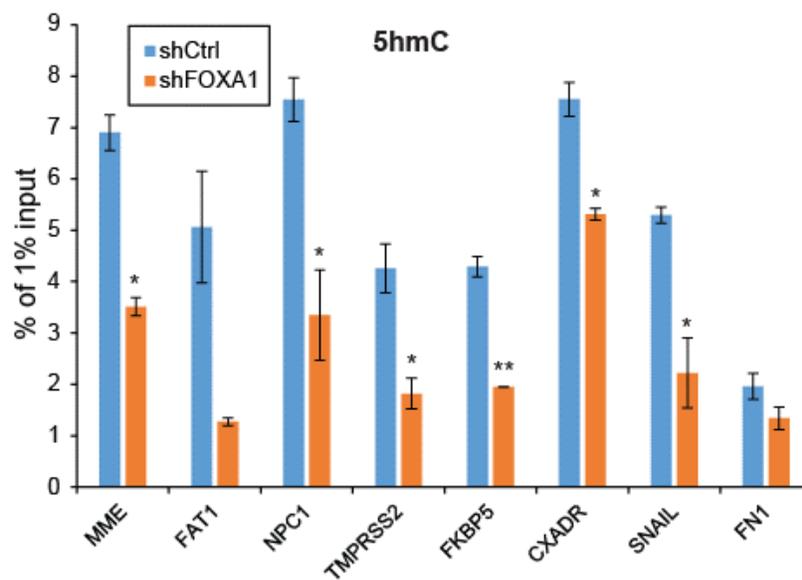


**B.**
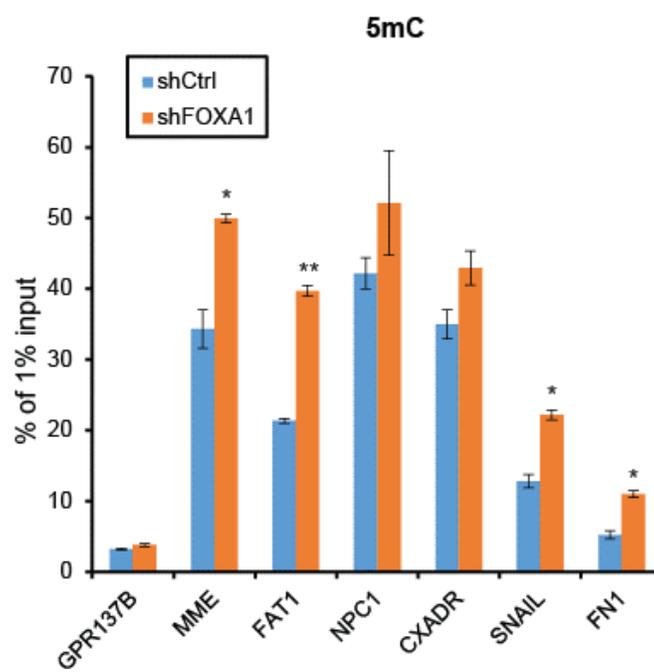
**Figure 2.6 FOXA1 depletion affects 5hmC and 5mC enrichment.** (**A**) Locus-specific change in 5hmC by qPCR of hMe-Seal at representative FXBS for control and shFOXA1 LNCaP cells. (**B**) MeDIP-PCR was performed with DNA from control and shFOXA1 LNCaP cells. Moderate increases were seen at several FOXA1-occupied sites. Data shown is mean ± SEM of technical replicates from one representative experiment out of two. *$P < 0.05$ and **$P < 0.01$.

FOXA1 positively regulates TET1 gene expression

As DNA demethylation has recently been shown to be catalyzed by the TET proteins, we next examined whether TET gene expression is associated with FOXA1. We first performed qRT-PCR analysis of FOXA1 and TET1 transcript across a panel of 12 prostate cell lines (**Figure 2.7A-B**). Interestingly, like FOXA1, TET1 is in general expressed at a much higher level in AR-positive prostate cancer cell lines such as C4-2B and VCaP cells than in AR-negative cells including DU145 and RWPE. Further analysis showed that TET1 expression level is highly correlated ($r = 0.96$, $P < 0.001$) with that of FOXA1 (**Figure 2.7C**). This positively correlated expression between FOXA1 and TET1 was confirmed in three large prostate cancer patient datasets (**Figure 2.8A-C**). As the correlation between FOXA1 and other TET proteins is relatively weaker, we decided to focus on TET1 in this study.

Since TET1 exhibited a similar expression pattern to FOXA1, we asked whether FOXA1 regulates TET1 gene expression. To test this, we first examined TET1 level in LNCaP cells with control or FOXA1 knockdown. Importantly, both TET1 transcript and protein levels were markedly decreased in LNCaP cells following FOXA1 knockdown (**Figure 2.9A**). As demonstrated in **Figure 2.8D**, Western blots of different exposure times were included to show that TET1 was detected much more strongly at 150 kDa, while also giving a very weak band above 250 kDa, both of which were depleted upon shRNA knockdown. Although the predicted size of TET1 is 235 kDa, the 150 kDa band is consistently much more strongly detected and specifically targeted by shRNA in prostate cancer cells. Here the 150 kDa band will be used to reflect TET1 level in prostate cells, and the presence of this seemingly truncated protein will be further investigated and characterized in the next Chapter. Concordantly, depletion of FOXA1 in another independent prostate cancer cell lines C4-2B also resulted in a decrease in TET1 expression (**Figure 2.9B**). On the other hand, when FOXA1 was overexpressed in 22Rv1 cells through adenovirus infection, TET1 expression was augmented (**Figure 2.10A**), which was further validated in another prostate cancer cell line DU145 that contained low endogenous FOXA1 level (**Figure 2.10A**). To visualize the inductive effect of FOXA1 on TET1 at the cellular level, we performed immunofluorescence staining. TET1 was barely detectable in control DU145 cells infected with empty vector adenovirus (**Figure 2.10B,**

**top panel**). However, upon infection with adenoviral FOXA1 (Flag-tagged, shown in red), TET1 staining (shown in green) was significantly enhanced (middle panel). Specifically, TET1 was stained positively in the majority of cells that had FOXA1 infection and overexpression, but not in the uninfected cells, as further illustrated in the zoomed-in microscopy images (**Figure 2.10B, bottom panel**). Taken together, our data support that FOXA1 positively regulates TET1 gene transcription.

**Figure 2.7**

**A.**
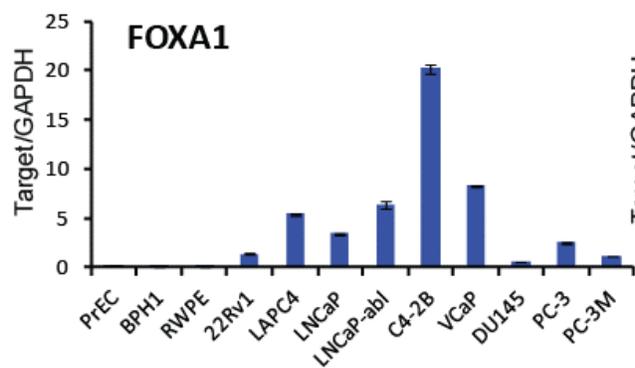


**B.**
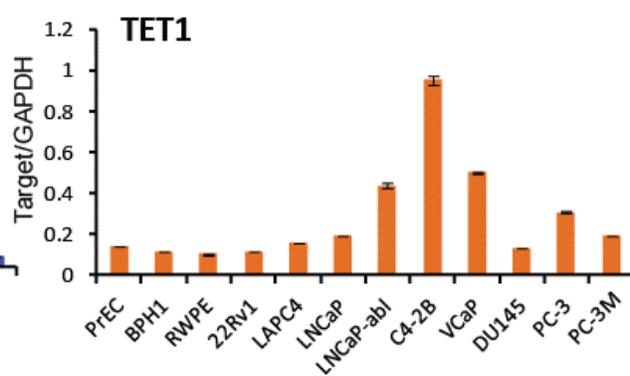
**C.**
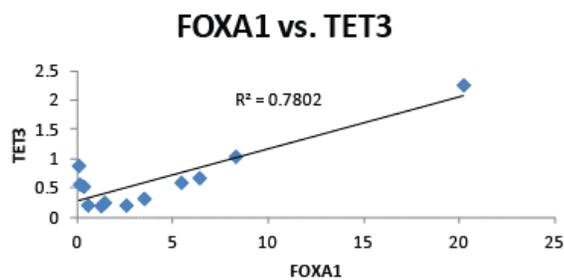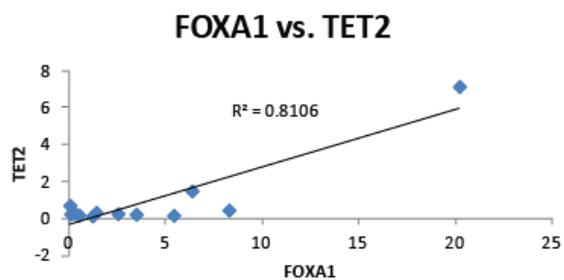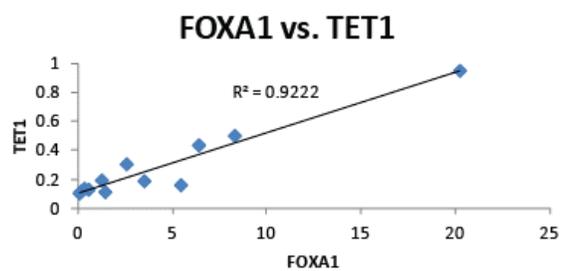
**Figure 2.7 FOXA1 expression correlates with TET1 in prostate cells.** (**A and B**) Correlated FOXA1 and TET1 gene expression in prostate cells. RNA was extracted from a panel of 12 prostate cell lines and analyzed by qRT-PCR for FOXA1 (**A**) and TET1 (**B**) gene expression. Data shown are mean ± SEM of technical replicates from one representative experiment out of three. (**C**) Scatter plots of FOXA1 and TET1, TET2, TET3 transcript level, measured by qRT-PCR, in 12 prostate cell lines.

**Figure 2.8**
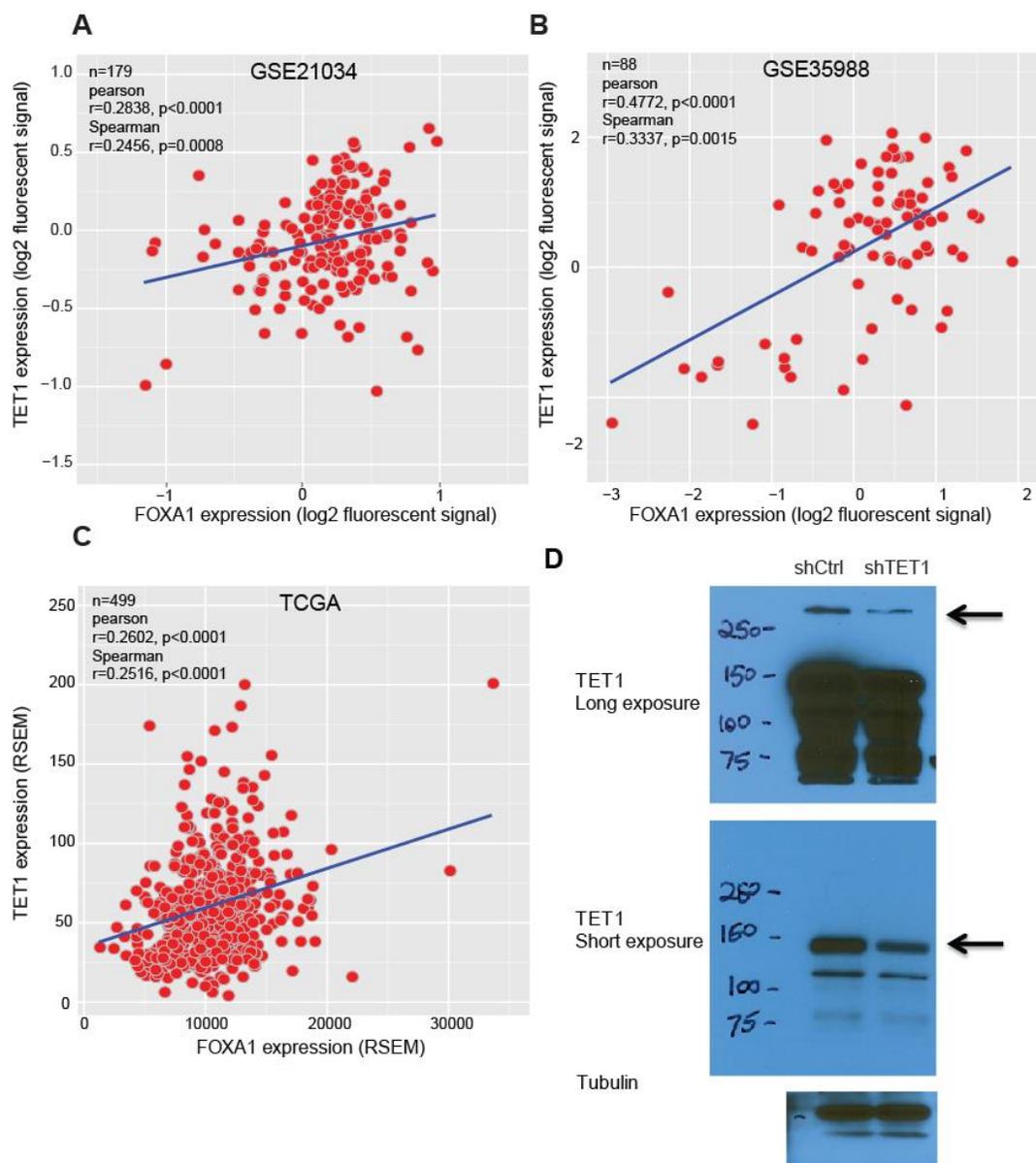
**Figure 2.8 FOXA1 and TET1 are positively correlated in patient datasets.** (**A-C**) Correlation analyses were performed using publicly available gene expression datasets for prostate cancer patients, GSE21032 (**A**), GSE35988 (**B**) and TCGA (**C**). All three scatter plots demonstrated statistically significant positive correlation between the two genes in patient specimens. (**D**) Western blot of control and shTET1 samples in LNCaP cells. The predicted size of TET1 is 235kDa. Both a faint band and a strong band could be detected with a reduction in shTET1, above 250kDa marker and at 150kDa marker, respectively. As the 150kDa band is much easier to detect and showed a similar decrease by TET1 knockdown, in the subsequent experiments we used this band to represent endogenous TET1. A-C were generated by Jung Kim.

**Figure 2.9**

**A.**
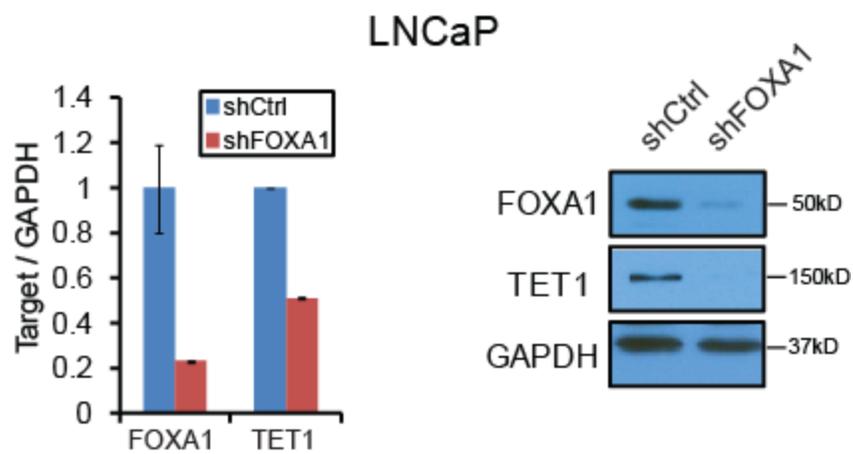


**B.**
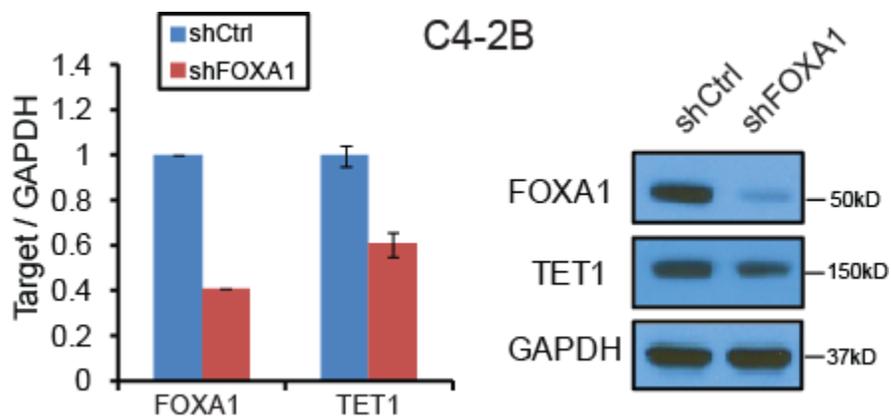
**Figure 2.9 FOXA1 induces TET1 gene expression.** (**A**) TET1 transcript and protein are downregulated following FOXA1 knockdown in LNCaP cells. LNCaP cells were infected with shCtrl or shFOXA1 lentivirus and subsequently subjected to qRT-PCR and western blot analysis. Data shown are one representative out of triplicate experiments. (**B**) TET1 is downregulated by FOXA1 knockdown in C4-2B cells. C4-2B cells were infected with shCtrl or shFOXA1 lentivirus for 8 h followed by puromycin selection for 4 days, and subsequently subjected to qRT-PCR and western blot analysis. Data shown are one representative out of triplicate experiments.

**Figure 2.10**

**A.**
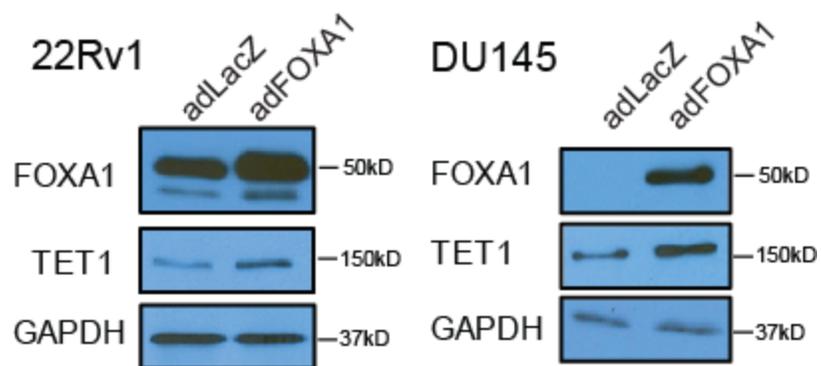


**B.**
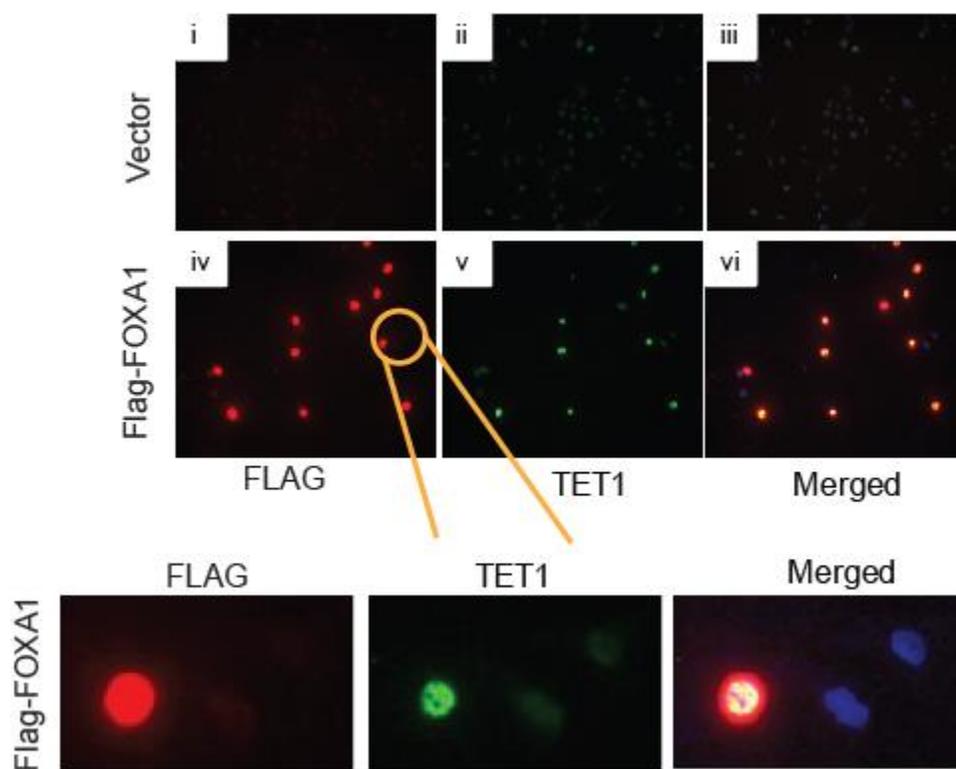
**Figure 2.10 FOXA1 induces TET1 gene expression.** (**A**) TET1 is upregulated following FOXA1

overexpression. The 22Rv1 and DU145 cells were infected with LacZ or FOXA1 adenovirus for 48 h and

immunoblot was performed to assess FOXA1 and TET1 protein levels. (**B**) Positive TET1 staining in

FOXA1-expressing cells. DU145 cells were infected with LacZ control (i–iii) or Flag-tagged FOXA1 (iv–vi)

adenovirus for 48 h and then subjected to Immunofluorescence co-staining of FOXA1 and TET1. Bottom

panel shows zoomed-in region containing both FOXA1-uninfected and -infected cells.

TET1 is a direct transcriptional target of FOXA1

To determine how FOXA1 transcriptionally controls TET1 expression, we examined FOXA1

ChIP-seq data previously obtained from LNCaP cells[168], and observed a strong FOXA1 binding event

within the intragenic region, between exons 3 and 4, of the TET1 gene (**Figure 2.11A**). On the other

hand, no comparable FOXA1 peaks were seen near TET2 and TET3 genes (**Figure 2.11B**). Being

consistent with FOXA1 as an enhancer regulator that modulates target genes through enhancer–

promoter looping, we also found a weak FOXA1 binding event at the TET1 promoter. To validate the

results of ChIP-seq, we performed ChIP-qPCR in LNCaP cells and found that FOXA1 is enriched at the

TET1 enhancer for nearly 170-fold relative to IgG control, an enrichment level comparable to that at the

Prostate-Specific Antigen (PSA, or KLK3) gene enhancer, and for about 10-fold at the TET 1 promoter

(**Figure 2.11C**). A similarly strong enrichment of FOXA1 at the TET1 enhancer and promoter was also

observed in an additional FOXA1-expressing cell line C4-2B (**Figure 2.11E**). Moreover, upon lentiviral

knockdown, FOXA1 binding to its target site for the PSA gene was greatly diminished as expected, and

similarly for TET1 enhancer and promoter, confirming that the ChIP enrichment signal was specific for

FOXA1 (**Figure 2.11D**). Next, to examine whether FOXA1 occupancy at the TET1 enhancer and

promoter leads to regulation of their transcriptional activities, we cloned these regions into reporter

constructs. Luciferase assays showed that FOXA1 overexpression indeed significantly increased,

whereas FOXA1 knockdown decreased, TET1 enhancer and promoter activities (**Figure 2.12A-B**). To

further demonstrate that this regulation is due to FOXA1 occupancy at the TET1 enhancer and promoter,

we analyzed the DNA sequences around the FOXA1 binding peaks for FKHD motifs within the TET1

enhancer as well as promoter. Through mutagenesis assays, we generated TET1 enhancer and promoter

constructs with mutations to highly conserved FKHD motifs (**Figure 2.11A, bottom panels**). Importantly,

luciferase assays revealed that mutations to the FKHD motifs abolished FOXA1 regulation of TET1

enhancer as well as promoter activities (**Figure 2.12C**). Taken together, our data support that FOXA1

directly binds to the regulatory elements of TET1 gene to induce its transcription. As FOXA1 contributes

to local DNA demethylation (**Figure 2.1**) and TET1 is a known DNA demethylase, we hypothesized that
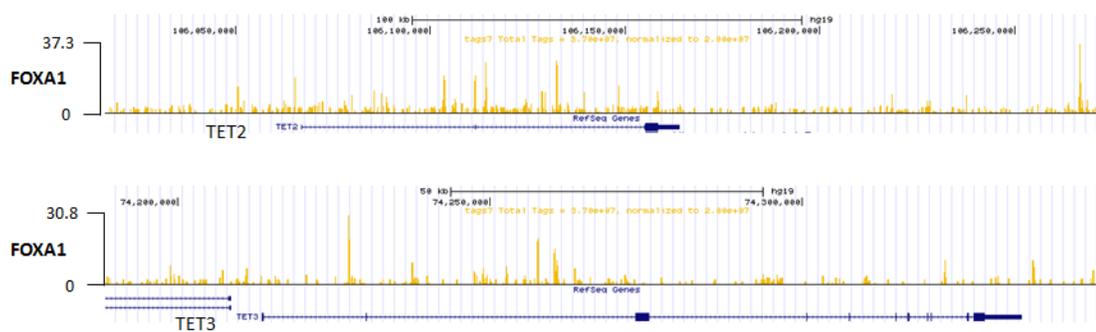
TET1 may be attributable for DNA demethylation around the FXBS. To test this hypothesis, we started

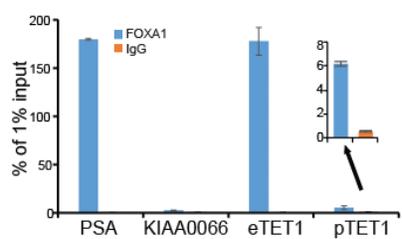out by examining potential interactions between the FOXA1 and TET1 proteins.

**Figure 2.11**

**A.**
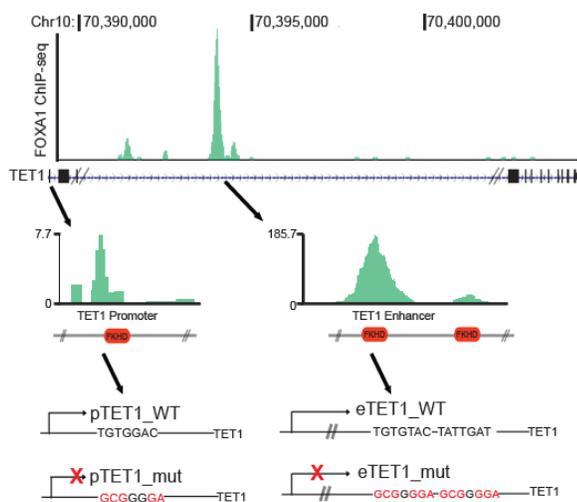


**B.**



**C.**                              **D.**                              **E.**
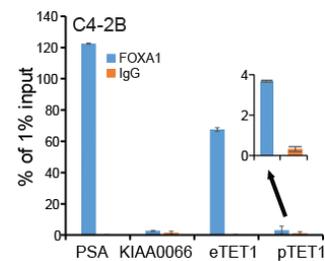
**Figure 2.11 TET1 is a direct transcriptional target of FOXA1.** (**A**) ChIP-seq showing FOXA1 binding

events at TET1 promoter and enhancer. FOXA1 ChIP-seq was conducted in LNCaP cells and FOXA1

binding events were identified by HOMER and visualized in UCSC Genome Browser. FKHD motifs

(indicated by red box) near FXBS were determined by JASPAR. DNA fragments containing FXBS at the

TET1 promoter (pTET1) and enhancer (eTET1) were each cloned into pGL4 luciferase reporter construct

with wild-type (WT) or mutated (mut) FKHD motif (mutated nt shown in red at the bottom panel). (**B**)

ChIP-seq showing no prominant FOXA1 binding sites close to TET2 and TET3 genes. (**C**) ChIP-PCR

validation of FOXA1 binding to TET1 enhancer and promoter in LNCaP cells. ChIP was performed using

anti-FOXA1 and anti-IgG antibodies in LNCaP cells. ChIP-qPCR was performed using primers flanking

the FOXA1 binding peaks at the TET1 enhancer (eTET1) and promoter (pTET). PSA is used as a positive

control while KIAA0066 a negative control. Data shown are mean ± SEM of technical replicates from one

representative experiment out of three. (**D**) FOXA1 occupancy at TET1 promoter and enhancer was

decreased by FOXA1 knockdown. ChIP-qPCR using anti-FOXA1 antibody was carried out in control and

FOXA1-depleted LNCaP cells. Data shown are mean ± SEM of technical replicates from one

representative experiment out of three. *P < 0.05 and **P < 0.01. (**E**) ChIP experiments were done in

FOXA1-containing C4-2B cells, using anti-FOXA1 and anti-IgG antibodies.
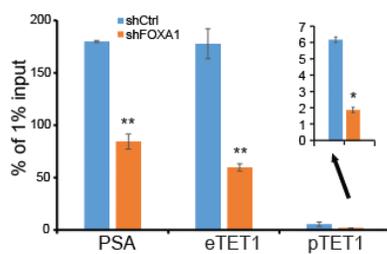
**Figure 2.12**

**A.**
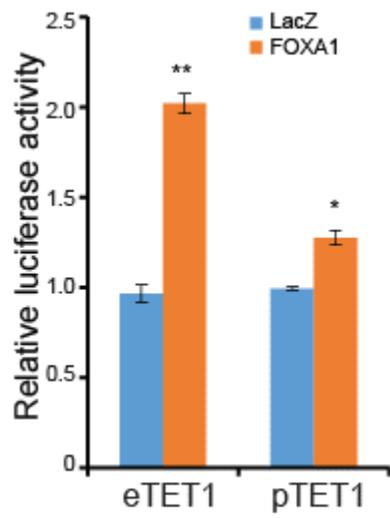


**B.**
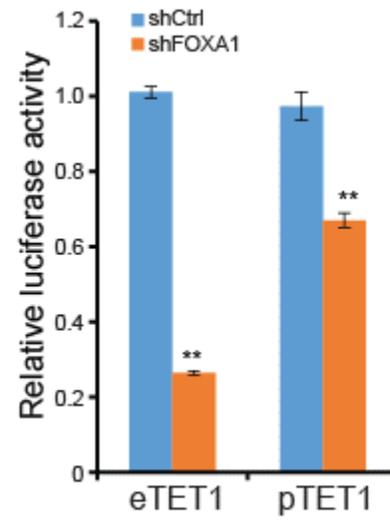


**C.**
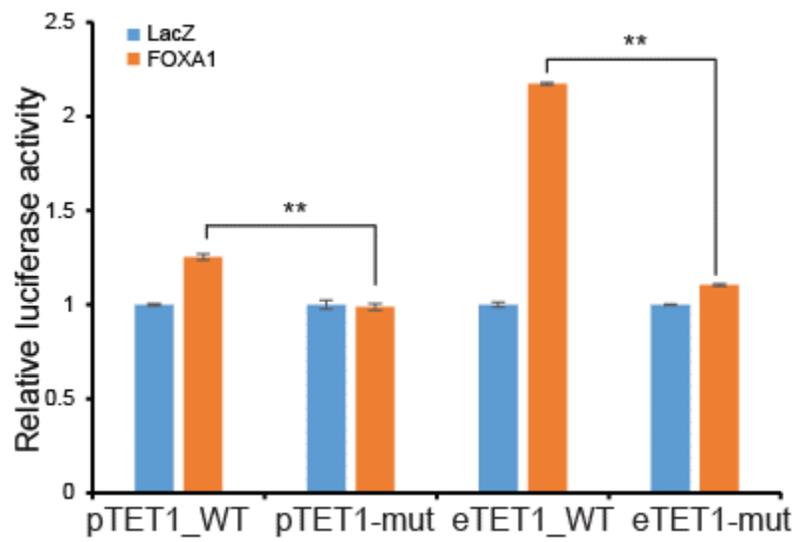
**Figure 2.12 TET1 is a direct transcriptional target of FOXA1.** (**A** and **B**) FOXA1 positively regulates TET1 enhancer and promoter activities. TET1 enhancer and promoter reporter constructs were transfected into LNCaP cells with control or FOXA1 overexpression (**A**) and LNCaP cells with control or FOXA1 knockdown (**B**) for 48 h. Luciferase activities were determined and normalized to internal control Renilla reporter. Data shown are mean ± SEM of two independent experiments. *$P < 0.05$ and **$P < 0.01$. (**C**) FKHD motif is required for FOXA1-induced TET1 promoter and enhancer luciferase activities. Control and FOXA1-overexpressing LNCaP cells were transfected with either WT or mutated (depicted in **Figure 2.11A**) TET1 promoter and enhancer reporter constructs. Luciferase activities were determined and normalized to internal control Renilla reporter. Data shown are mean ± SEM of two independent experiments. *$P < 0.05$ and **$P < 0.01$.

FOXA1 and TET1 proteins physically interact

By use of overexpression systems in 293T cells, we conducted co-IP experiments to assess whether physical interaction is present between ectopic FOXA1 and TET1 proteins. The 293T cells were co-transfected with Flag-tagged TET1 along with FOXA1 or empty vector. Successful expression of the ectopic proteins was confirmed by western blot analysis of the input lysate. IP using an anti-FOXA1 antibody followed by immunoblotting confirmed successful pulldown of FOXA1 itself as well as the TET1 protein, the latter only in the cells expressing both TET1 and FOXA1 (**Figure 2.13A**). To demonstrate the interaction through reversal co-IP, we cloned TET1 into the SFB-tagged expression vector, which enabled pulldown of the TET1 protein using S-protein agarose beads and detection by anti-Flag antibodies[187]. Either SFB-vector control or SFB-TET1 was co-transfected with FOXA1 into 293T cells and their expression was confirmed by western blot analysis of the input lysate. S-protein pulldown followed by western blotting using anti-Flag validated successful enrichment of SFB-tag only or SFB-TET1 (of different sizes) in the corresponding lysates, while immunoblotting using anti-FOXA1 revealed FOXA1 pulldown only in the SFB-TET1-expressing cells (**Figure 2.13B**), supporting physical interaction between ectopic FOXA1 and TET1 proteins.

Next, we attempted to confirm this interaction between endogenous FOXA1 and TET1 proteins. LNCaP cell nuclear lysate was subjected to IP using rabbit anti-TET1, anti-FOXA1 and IgG control followed by western blotting with mouse anti-TET1 or anti-FOXA1 antibodies. Our results demonstrated that TET1 and FOXA1 antibodies are able to pull down each other, supporting strong protein interactions (**Figure 2.13C**). To address the potential involvement of DNA in mediating this interaction, we performed co-IP in the presence or absence of ethidium bromide. Notably, our results demonstrated persistent interaction between FOXA1 and TET1 proteins in the presence of ethidium bromide, thus indicating that DNA was not required for their association (**Figure 2.13E**). Moreover, this interaction between endogenous FOXA1 and TET1 proteins was also confirmed in C4-2B cells (**Figure 2.13D**).

73

**Figure 2.13**

**A.**
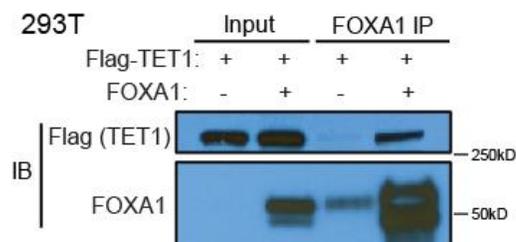
**B.**
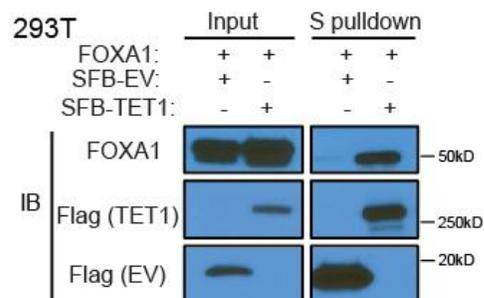
**C.**

**D.**
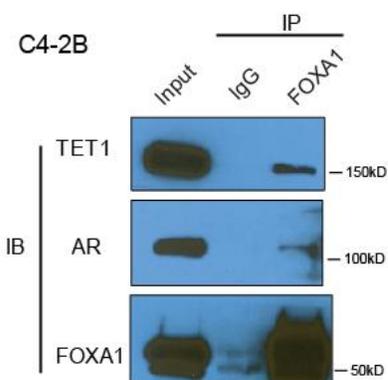
**E.**
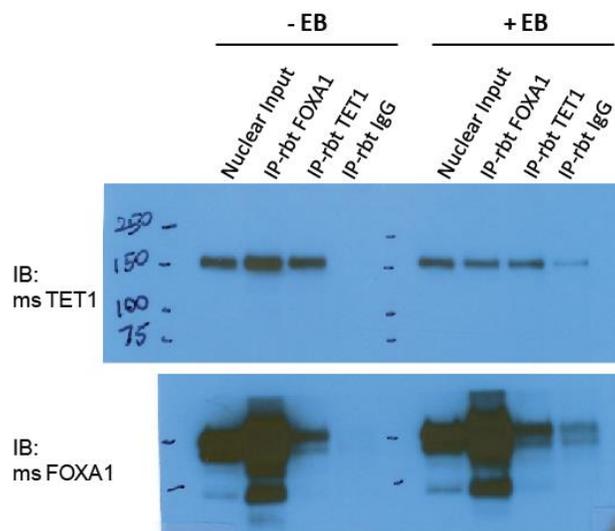
**Figure 2.13 FOXA1 and TET1 proteins physically interact.** (**A**) Immunoprecipitation of ectopic FOXA1 pulled down TET1 protein. The 293T cells were transfected with Flag-TET1, either alone or together with FOXA1, for 48 h and then subjected to immunoprecipitation using an FOXA1 antibody. Whole cell (Input) and IP-enriched lysates were then analyzed by western blotting using anti-Flag (TET1) and anti-FOXA1 antibodies. (**B**) Ectopic TET1 immunoprecipitation pulled down FOXA1 protein. The 293T cells were co-transfected with FOXA1 and SFB-tagged empty vector (EV) or TET1 for 48 h before immunoprecipitation using S beads, which will pull down SFB-EV or SFB-TET1. The input and IP-enriched cell lysates were then subjected to western blotting using anti-FOXA1 and anti-Flag (for SFB-EV or SFB-TET1) antibodies. (**C**) Endogenous FOXA1 and TET1 proteins interact in LNCaP cells. LNCaP cells were subjected to immunoprecipitation using anti-FOXA1, anti-TET1 and IgG control, followed by western blotting of FOXA1 and TET1 proteins. (**D**) Endogenous Co-IP was performed in C4-2B nuclear proteins. AR, which is known to have interaction with FOXA1, is shown as a positive control. (**E**) LNCaP endogenous Co-IP was performed in absence and presence of Ethidium bromide (EB). Nuclear proteins of LNCaP cells were used for IP with anti-FOXA1, anti-TET1 and anti-rabbit IgG antibodies. Ethidium bromide treatment (50ug/ml for 30min) was done to abrogate any potential DNA-mediated protein-protein interaction. Western blot analysis was subsequently used to look at whether FOXA1-TET1 protein interaction is dependent on DNA association.

To further determine which domains of the TET1 protein are important for its interaction with FOXA1, we generated four Myc-tagged TET1 domain constructs, namely the N-terminal, CXXC, middle and CD domains, which were co-transfected with SFB-tagged FOXA1 into 293T cells. S-protein pulldown followed by western blot analysis showed that only the TET1 fragment containing the CXXC module was able to bind FOXA1 (**Figure 2.14A**). On the other hand, we attempted to map out the FOXA1 domain that is responsible for its interaction with the TET1 protein. Similarly, we created three Flag-tagged FOXA1 domain constructs, namely N-terminal, Forkhead (FH) and C-terminal domains, which were co-transfected with SFB-tagged TET1-CXXC domain into 293T cells. Western blot analysis confirmed the expression of various FOXA1 domains of different sizes as expected (**Figure 2.14B**). S-protein pulldown of TET1 followed by western blotting revealed that only the FH-containing domain of FOXA1 protein is able to interact with the TET1-CXXC domain. Moreover, we also performed *in vitro* pulldown assay utilizing purified TET1-CXXC and FOXA1-FH domain proteins, which confirmed that the two proteins directly interact (**Figure 2.14C**). As the CXXC zinc finger module in Tet3 protein has been shown critical for specific chromatin targeting, while its enzymatic domain modulates its biological function[188], we hypothesized that TET1 interaction with FOXA1 through its CXXC domain may be important for its recruitment to FXBS where it carries out hydroxylation on methylated CpG's closeby through its CD domain. Therefore, we next asked whether TET1 regulates DNA demethylation and alters epigenetic modifications around FXBS.
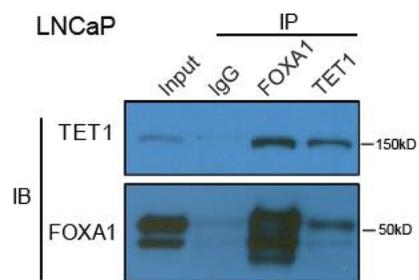
**Figure 2.14**

**A.**
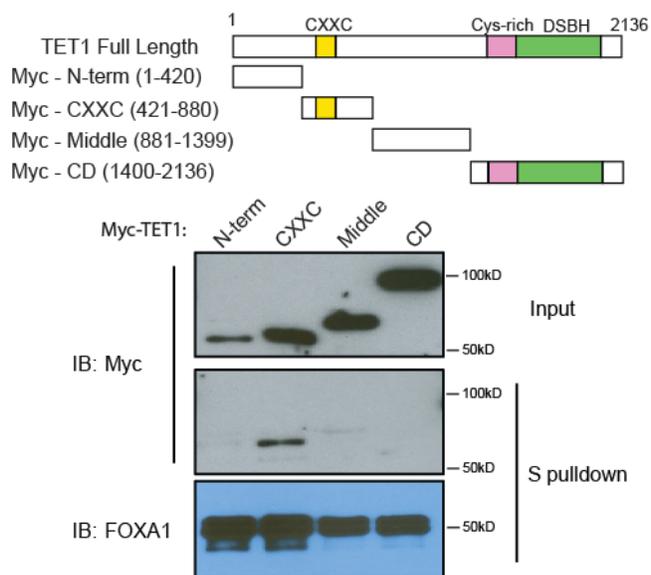


**B.**
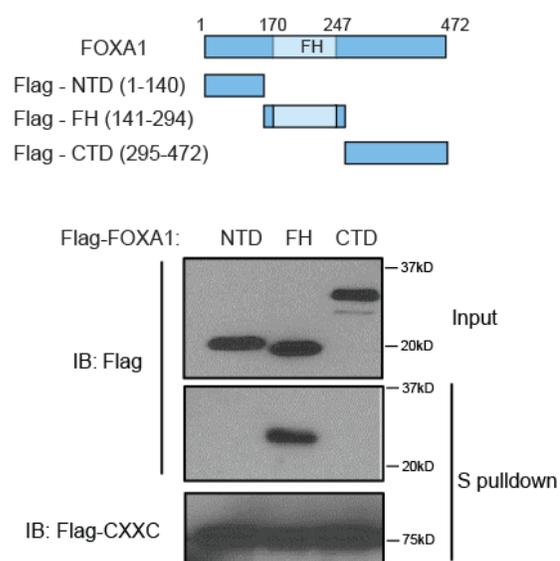


**C.**
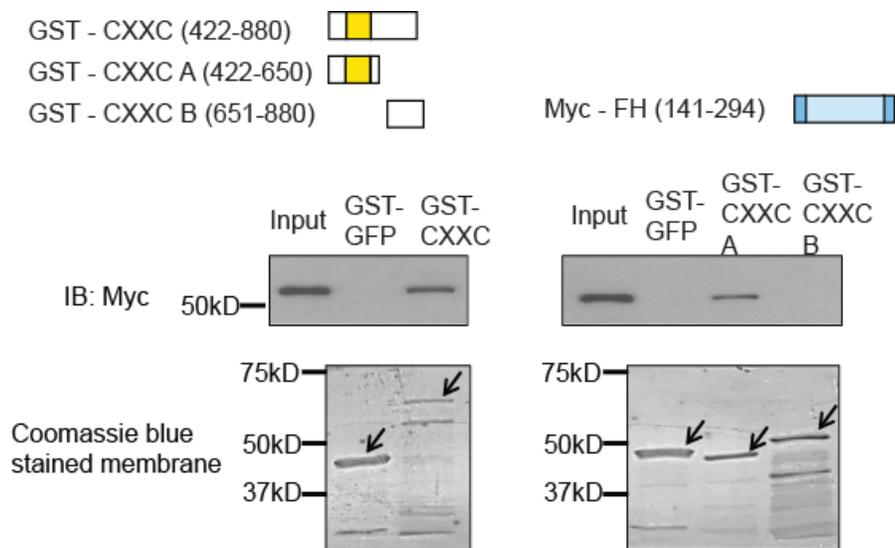
**Figure 2.14 FOXA1 and TET1 proteins physically interact.** (**A**) TET1 CXXC domain interacts with the FOXA1 protein. 293T cells were co-transfected with SFB-FOXA1 along with various Myc-tagged TET1 domain constructs. The expression of TET1 domains in whole cell lysate (input) was confirmed by western blotting using anti-Myc. Cell lysates were then subjected to S pull down (of FOXA1) and subsequently western blot analysis using anti-FOXA1 and anti-Myc antibodies. (**B**) FOXA1 FH (Forkhead-containing) domain interacts with TET1 CXXC domain. 293T cells were co-transfected with SFB-CXXC along with various Flag-tagged FOXA1 domain constructs and subjected to S pull down (of TET1-CXXC) followed by western blotting using an anti-Flag antibody. (**C**) *In vitro* interaction assay was conducted using purified proteins of TET1 CXXC domain and FOXA1 Forkhead domain. CXXC domain was tagged with GST and further subdivided into fragments A and B (the 'C-X-X-C' motif was located in residues 590–609 in fragment A), and FH domain was tagged with Myc. Arrows point to expression of proteins according to their expected size.  C was performed by Will Fong.

TET1 mediates active epigenetic modification at FOXA1-dependent enhancers

To determine whether TET1 affects the epigenetic environment at FOXA1-occupied enhancers, we first tested whether TET1 is able to co-occupy FOXA1-bound genomic regions. As human anti-TET1 antibody has not been well-established for ChIP, we transfected HA-tagged TET1 into LNCaP prostate cancer cells, validated by western blot in **Figure 2.15A** and performed ChIP using ChIP-grade anti-HA antibody. ChIP-qPCR confirmed much stronger HA (TET1) enrichment at FXBS in cells expressing HA-TET1 than cells transfected with HA-control vector (**Figure 2.15B**). Next, to examine how TET1 alters DNA methylation around these FOXA1-bound regions, we performed TET1 knockdown using shRNA (**Figure 2.16A**), and specific depletion of endogenous TET1 at both 250 kDa and 150 kDa can be observed (**Figure 2.8D**). As TET1 is a DNA demethylase that catalyzes 5mC–5hmC, we next sought to determine the level of 5hmC and 5mC in TET1-knockdown cells. Dot blot experiment confirmed significant reduction of total 5hmC abundance in shTET1 cells (**Figure 2.16B**). Further, hMe-Seal-seq revealed a remarkable decrease of total 5hmC-enriched regions following TET1 knockdown (**Figure 2.16C**). By contrast, 5mC as measured by MeDIP-seq was increased nearly 33% (**Figure 2.16C**). Average intensity view of all peaks showed that hMe-Seal signals were significantly decreased, while MeDIP signals increased upon TET1-knockdown (**Figure 2.16D**). Focused analysis of these epigenetic modifications around FXBS confirmed an overall decrease of 5hmC and increase of 5mC following TET1 depletion, suggesting that TET1 is critical for the maintenance of the demethylated state of these enhancers (**Figure 2.17A-B**). As DNA methylation has been shown to inhibit enhancer activation[179], we next asked whether TET1 knockdown prohibits enhancer activation at FXBS. ChIP-qPCR showed that indeed H3K4me2 and H3K27ac were both significantly reduced following TET1 depletion (**Figure 2.17C-D**). ChIP-seq further confirmed a global decrease of H3K4me2 level in TET1-knockdown (**Figure 2.17E**). Taken together, our data support that TET1 expression contributes to the activation of FOXA1-target enhancers through mediating active DNA demethylation.

**Figure 2.15**

**A.**
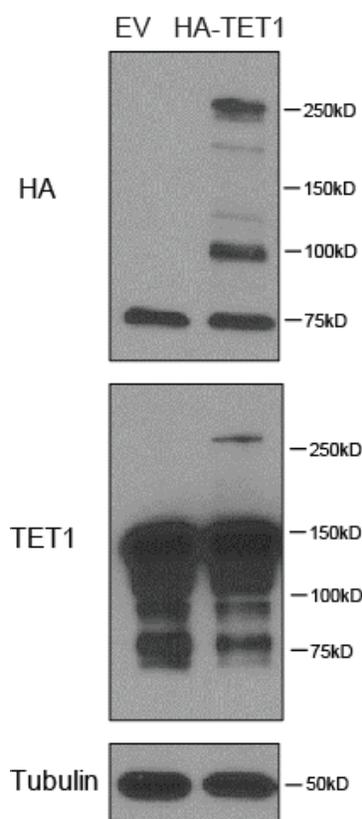


**B.**
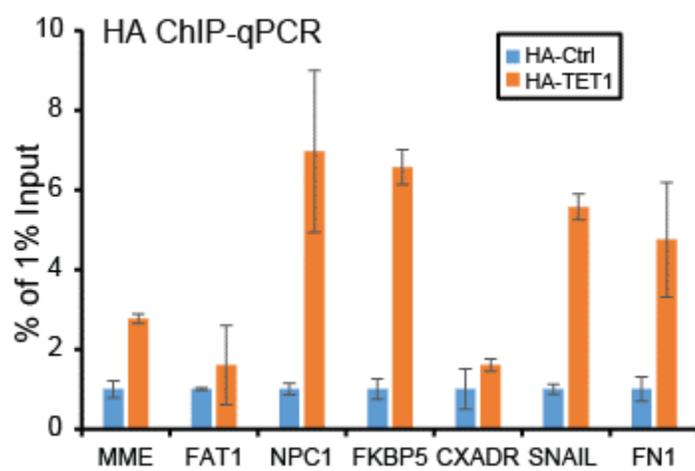
**Figure 2.15 TET1 co-occupies FOXA1 binding sites.** (**A**) Western blot of control empty vector (EV) and HA-TET1 transfected in LNCaP cells. Both HA and TET1 antibodies were used to confirm overexpression. (**B**) LNCaP cells were transfected with HA-tagged empty vector or TET1 constructs and were subsequently used for ChIP with anti-HA antibody. HA ChIP-qPCR was performed using primers flanking a number of FXBS. Data shown are mean ± SEM of technical replicates from one representative experiment out of three.

**Figure 2.16**

**A.**



**B.**



**C.**



**D.**
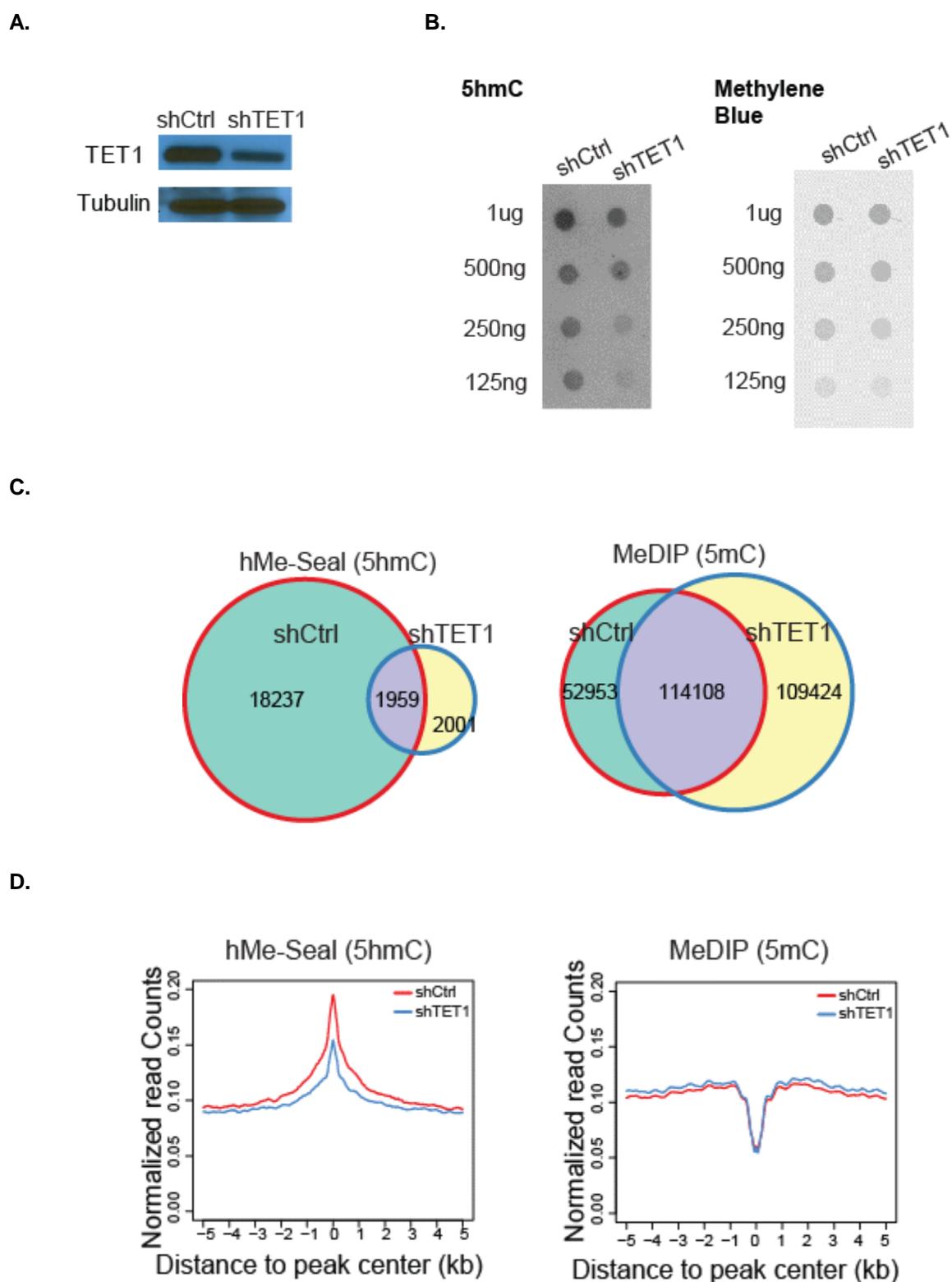
**Figure 2.16 TET1 knockdown decreases 5hmC.** (**A**) Western blots confirming TET1 knockdown.

LNCaP cells were infected with either scramble or shTET1 lentivirus followed by puromycin selection for 4

days before western blot analysis. Tubulin is used as a loading control. (**B**) Genomic DNA extracted from

LNCaP control and shTET1 cells was serially diluted and used for dot blot with anti-5hmC antibody.

Methylene blue staining was included as loading control. (**C**) Venn Diagrams showing alterations in global

genomic regions enriched for 5hmC and 5mC following TET1 knockdown. LNCaP cells with control or

shTET1 were subjected to hMe-Seal-seq and MeDIP-seq for genome-wide location analysis of 5hmC and

5mC, respectively, which were subsequently compared between control and TET1-depleted cells. (**D**)

Average intensity plot of normalized hMe-Seal-seq and MeDIP-seq reads around (±5 kb) FXBS. C and D

were generated by Jonathan Zhao.

**Figure 2.17**

**A.**
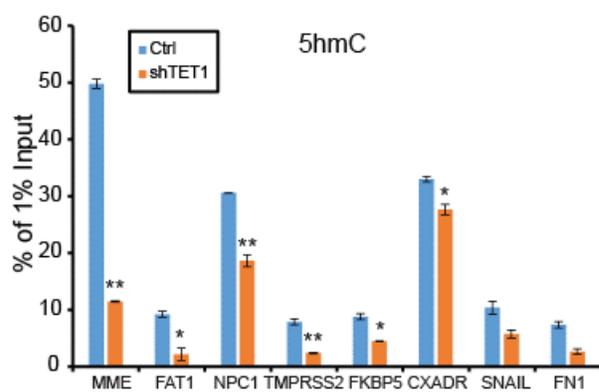


**B.**
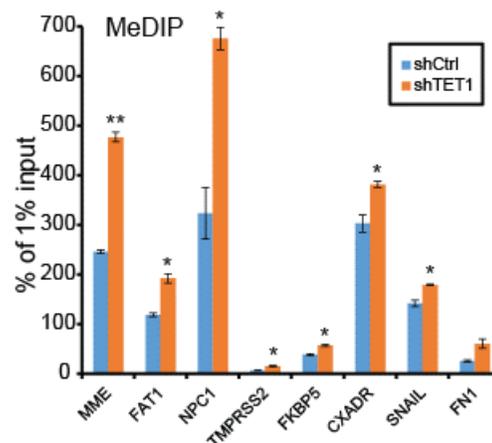


**C.**
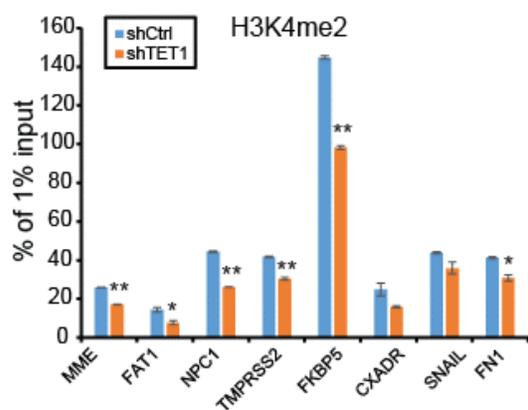


**D.**
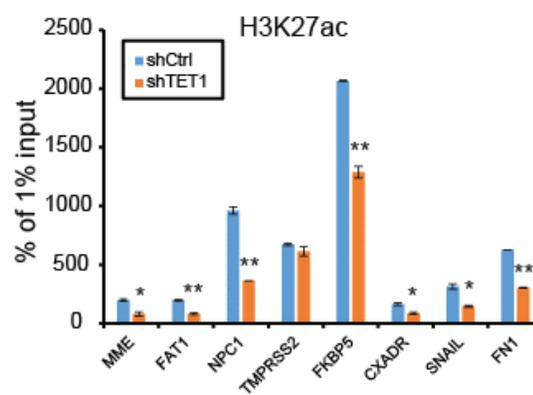


**E.**
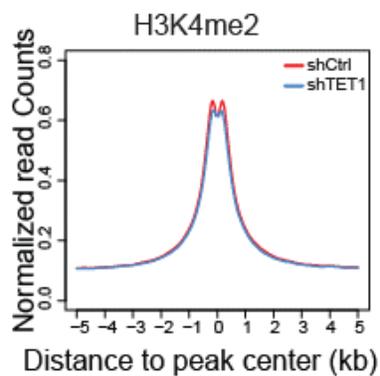
**Figure 2.17 TET1 mediates active epigenetic modification at FOXA1-bound enhancers.** (**A-D**) TET1 knockdown led to altered epigenetic signatures at FXBS. LNCaP cells with control or TET1 knockdown were subjected to hMe-Seal (**A**) and MeDIP (**B**) and ChIP using anti-H4K4me2 (**C**) and anti-H3K27ac (**D**) antibodies, followed by qPCR analysis with site-specific primers. Data shown are mean ± SEM of technical replicates from one representative experiment out of two. *$P$ < 0.05 and **$P$ < 0.01. (**E**) Average intensity plots of normalized H3K4me2-seq reads around (±5 kb) FXBS. E was generated by Jonathan Zhao.

TET1 expression is required for FOXA1 recruitment to target enhancers

Since it has been reported that DNA methylation and removal of H3K4me2 could impair FOXA1 binding[177,179], the changes in DNA methylation and histone modification events observed following TET1 depletion were suggestive of disrupted FOXA1 recruitment to these regions. To test this, we performed FOXA1 ChIP-seq in control and TET1-knockdown LNCaP cells to determine whether TET1 depletion is able to regulate FOXA1 chromatin targeting. A global assessment of the total binding events before and after TET1 knockdown demonstrated that a significant proportion of FOXA1 binding events were lost upon TET1 depletion (**Figure 2.18A**). The total number of FXBS was decreased from 76 000 to 55 000. In addition, the average intensity of FOXA1 binding events appeared to be much weaker even for the sites that were not fully abolished (i.e. shared sites) following TET1 knockdown (**Figure 2.18B**). Genome browser view of several FOXA1-dependent enhancers further illustrated significant loss of FOXA1 occupancy in TET1-depleted cells (**Figure 2.18C**). Meanwhile, DNA methylation at these enhancers was increased as indicated by enhanced 5mC but reduced 5hmC signals, while active enhancer mark H3K4me2 was decreased, being concordant with the genome-wide switch to repressive chromatin state. Moreover, ChIP-qPCR confirmed that TET1 knockdown significantly decreased FOXA1 occupancy at multiple target enhancers (**Figure 2.19A**).

As TET1 interacts with the FOXA1 protein through its CXXC domain but is known to carry out enzymatic activities through its CD domain, we next attempted to understand mechanistically whether CD-mediated DNA demethylation is sufficient to facilitate FOXA1 recruitment to target enhancers. A recent study has reported an interesting and important observation that the CD domain of TET proteins induces massive global DNA demethylation, whereas the function of full-length TET1 is much restricted to unmethylated CpG islands[189]. We thus predict that CD domain may be able to restore FOXA1 recruitment in TET1-knockdown cells. To test this, we overexpressed the CD domain in LNCaP cells with TET1 knockdown (western in **Figure 2.19B**). ChIP-qPCR confirmed that FOXA1 binding at target enhancers was decreased by TET1 knockdown, which, importantly, can be fully rescued by concomitant CD domain

overexpression (**Figure 2.19C**). Taken together, our data support that TET1 facilitates FOXA1 recruitment to target enhancers through active demethylation.

**Figure 2.18**

**A.**
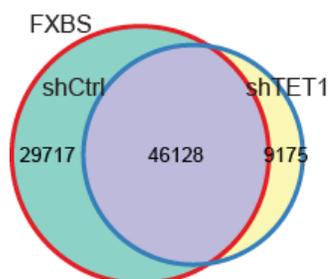


**B.**
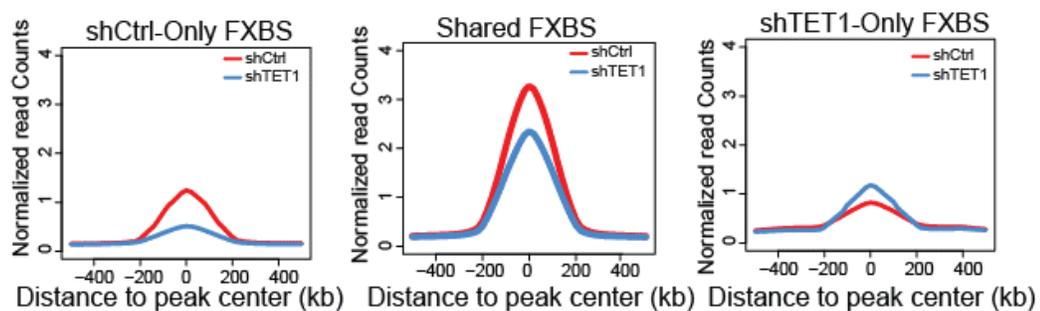


**C.**
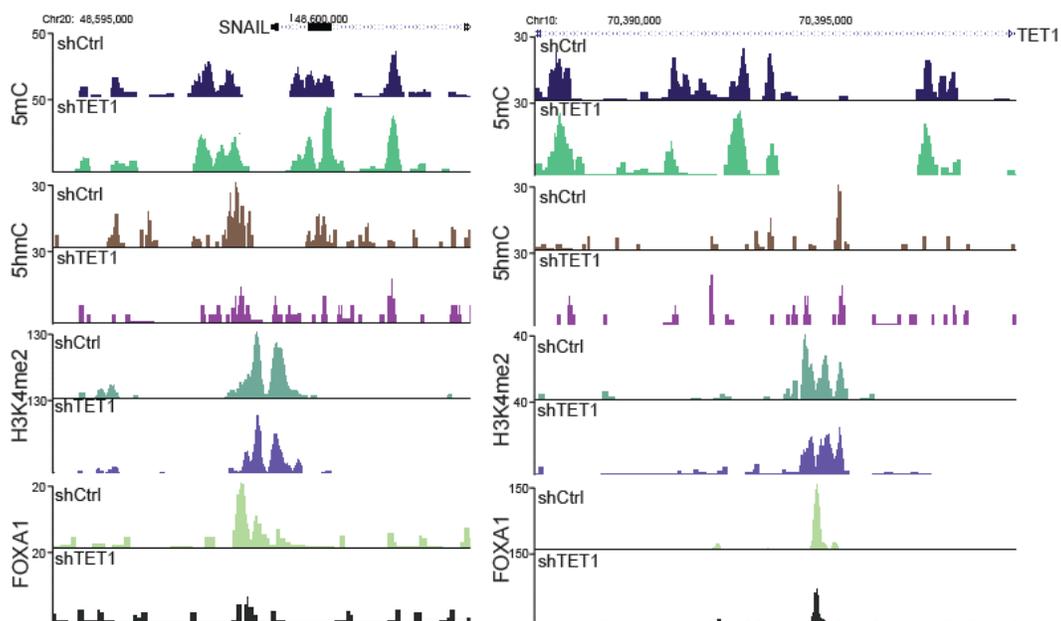
**Figure 2.18 TET1 is required for FOXA1 recruitment to lineage-specific enhancers.** (**A**) Venn

diagram showing overlap of FXBS in control and FOXA1-knockdown LNCaP cells. (**B**) Average FOXA1

ChIP-seq read intensity around (±500 bp) shCtrl-only, shared and shTET1-only FXBS identified from

overlap Venn diagram in (**A**). (**C**) Genome browser views of epigenetic modifications at the regulatory

regions of FOXA1-target genes SNAIL and TET1 itself. MeDIP-seq (5mC), hMe-Seal-seq (5hmC),

H3K4me2 and FOXA1 ChIP-seq were performed in control and TET1-knockdown LNCaP cells. For each

mark, the shCtrl and shTET1 tracks are shown on the same scale (Y-axis) for visual comparison of

enrichment. A and B were generated by Jonathan Zhao.

**Figure 2.19**

**A.**
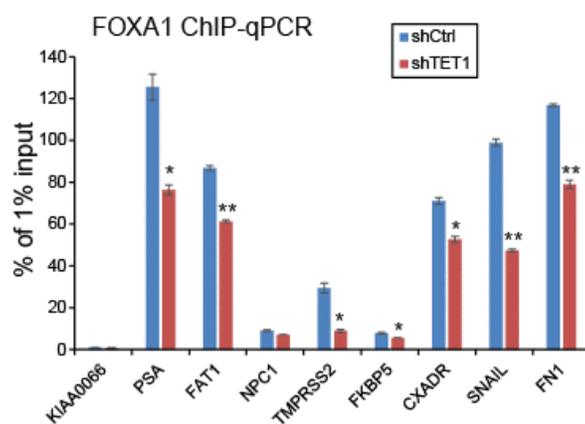


**B.**
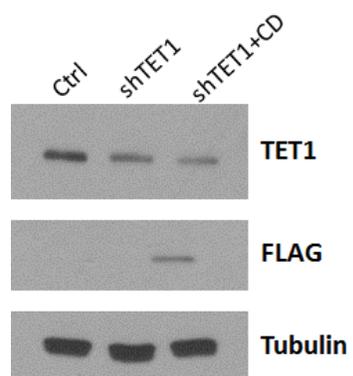


**C.**
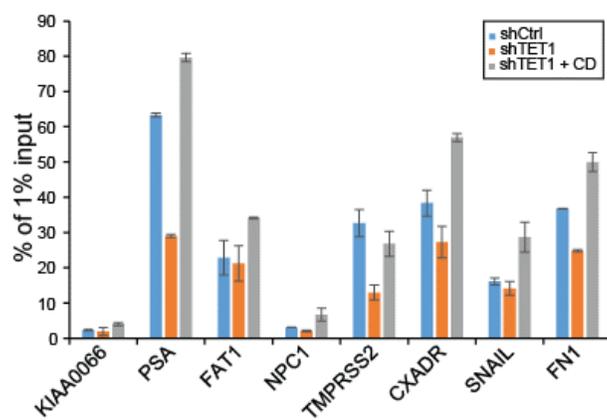
**Figure 2.19 TET1 catalytic function is required for FOXA1 recruitment to lineage-specific**

**enhancers.** (**A**) TET1 depletion attenuates FOXA1 recruitment to target enhancers. ChIP-qPCR was

performed in control and shTET1 LNCaP cells using anti-FOXA1 antibody. Data shown are mean ± SEM

of technical replicates from one representative out of triplicate experiments. *$P$ < 0.05 and **$P$ < 0.01. (**B**)

LNCaP cells were subjected to control or TET1 knockdown with or without concomitant TET1 CD

overexpression. TET1 knockdown and CD domain (Flag-CD) overexpression were confirmed by western

blot analysis. (**C**) Cells were subsequently used for ChIP with an anti-FOXA1 antibody followed by qPCR

analysis, showing that impaired FOXA1 recruitment in TET1-depleted cells was restored by TET1 CD

overexpression. Data shown are mean ± SEM of technical replicates from one representative of duplicate

experiments.

**III. Discussion**

FOXA1 is a critical regulator of hormone-mediated gene expression in prostate and breast cancers. Much efforts have been devoted to understand the molecular basis for FOXA1's activity as a pioneer factor and studies in the past two decades have helped to paint a clearer picture of how FOXA1 activity is dependent on a number of epigenetic signatures that exhibit lineage specificity[177]. Although FOXA1 has been shown to impose some effects on the epigenetic signatures around target enhancers[177], the molecular mechanisms by which FOXA1 remodels heterochromatin remain largely unknown. In the present study, we show that FOXA1 is able to directly regulate the transcription of *TET1* gene. Further, FOXA1 physically interacts with the TET1 protein, leading to DNA demethylation and H3K4me2/H3K27ac modifications at FOXA1-target enhancers. These changes in the epigenetic environment on the other hand enhance FOXA1 recruitment. Therefore, our data support a model wherein FOXA1 is not only able to recognize and bind enhancer regions, but contributes to *de novo* gain of H3K4 methylation and enhancer activation. The latter is mediated by, at least in part, a feed-forward loop between FOXA1 and TET1 where FOXA1 induces TET1 expression and binding at lineage-specific enhancers, which in turn facilitates and stabilizes FOXA1 recruitment through catalyzing DNA demethylation (**Figure 2.20**). Accompanying changes in DNA methylation are also reductions in H3K4me2 and H3K27ac upon FOXA1 depletion. Whether these are secondary to DNA demethylation or FOXA1/TET1 may regulate histone methyltransferases such as MLL are areas for future investigation.

TET1 has been implicated in the regulation of enhancer activation and lineage differentiation through DNA demethylation[42,185], the underlying mechanism of which, however, remains elusive. In this study, using prostate cancer cells as a model system, we demonstrated that TET1 contributes to FOXA1 recruitment to prostate-specific enhancers by modulating local epigenetic switch. In future studies, it will be interesting to investigate and compare how TET1 regulates epigenetic marks and FOXA1 recruitment in breast cancer, since FOXA1 has been shown to bind distinct, lineage-specific enhancers in prostate and breast cells[177]. In addition, this study will pave the way to further investigation of how TET1, through

modulation of FOXA1-dependent enhancer activation, regulates hormone-dependent gene expression and prostate and breast cancer progression.

The CXXC domain of TET proteins has been shown critical for specific chromatin targeting, while the CD domain modulates its enzymatic activity[188]. Further, a recent study has reported that the full-length TET1 protein preferably binds to unmethylated CpG islands through its CXXC domain[189]. Being consistent with these reports, we found that FOXA1 interacts with TET1 protein through its CXXC domain. Such interaction may be critical for targeting TET1 to prostate-specific enhancers denoted by FOXA1, which may be interesting lines for further investigation utilizing various TET1 deletion constructs and ChIP-seq experiments. Moreover, TET1 might similarly interact with other lineage-defining transcription factors and get recruited to distinct, lineage-specific enhancers in different cell types. By contrast, overexpression of the TET1 CD domain alone has been shown to induce massive global DNA demethylation[189]. Indeed, in our study we found overexpression of CD domain is able to rescue the effects of TET1 knockdown on FOXA1 recruitment to target enhancers. Therefore, through interaction with other transcription factors, TET1 achieves its specificity to bind selected enhancers, where it carries out its role in the maintenance of hypomethylated landscape and regulation of lineage differentiation.

In conclusion, FOXA1 is a multipotent pioneer transcription factor, which is impressively capable of chromatin remodeling through not only histone displacement but also DNA demethylation by employing the DNA hydroxylase TET1. Collectively, through regulation of TET1 expression and function, FOXA1 is able to control the epigenetic signatures present at its cis-regulatory elements through a feed-forward loop, ultimately giving rise to chromatin relaxation and enhancer activation.

**Figure 2.20**
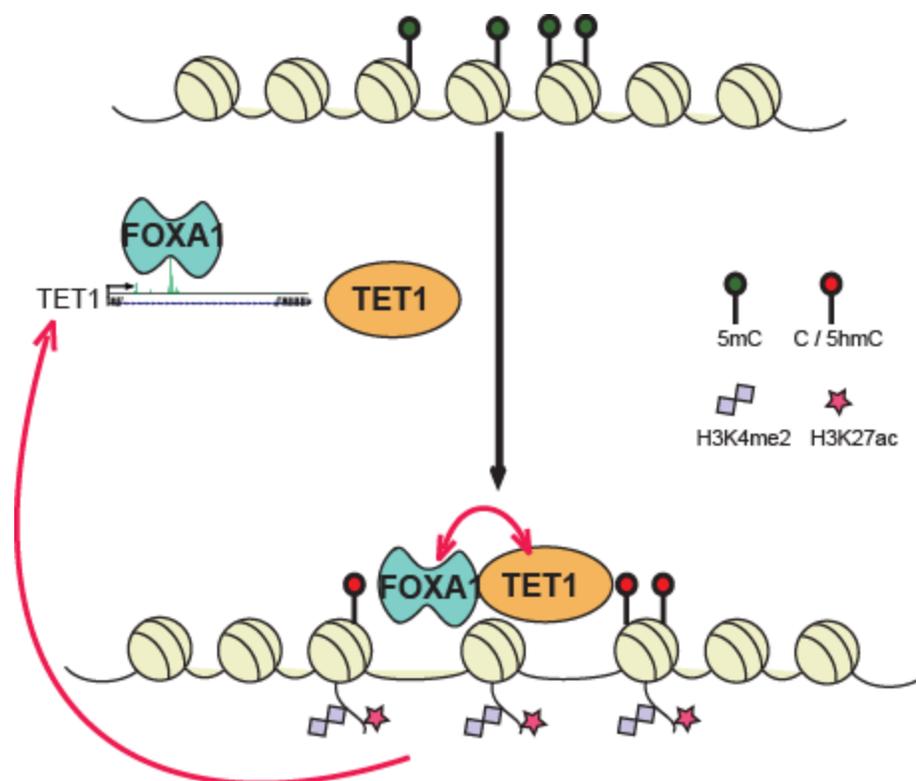
**Figure 2.20 Schematic model depicting feed-forward regulation between FOXA1 and TET1 in lineage-specific enhancer activation.** FOXA1 protein occupies at an intragenic enhancer of the TET1 gene to induce TET1 expression. Through direct interaction with FOXA1 protein, TET1 modulates DNA demethylation and subsequently H3K4 methylation and H3K27 acetylation at FOXA1-target enhancers, which in turn facilitates FOXA1 recruitment. Thus, FOXA1 and TET1 form a positive feedback loop in lineage-specific enhancer activation. FOXA1 is not only capable of recognizing but also modifying epigenetic signatures at lineage-specific enhancers.

**CHAPTER 3: TET1 has a transcript isoform in prostate cancer cells**

**I. Introduction**

As shown earlier in **Figure 2.8D**, western blot using an anti-TET1 antibody displays 2 distinct bands that are both diminished upon shRNA-mediated TET1 knockdown, in the prostate cancer cell line LNCaP. While the predicted size of TET1 is 235kDa, detection at this size is overwhelmingly weaker compared to a band at 150kDa, which points to the possibility that there may exist a truncated form of the protein.

The TET1 family proteins only came to light in 2009, studies since then have focused on their functional role in mediating DNA demethylation, and many aspects of these proteins remain to be further investigated to provide us more thorough understanding of their regulatory mechanisms.

In 2014, it was reported that Tet proteins in mouse ESCs can be subject to degradation by a specific family of proteases called calpains, which are calcium-dependent proteases[190]. Another report in 2013 showed that IDAX and CXXC5 can interact with TET2, which is the only TET family protein lacking the CXXC domain, and regulate TET2 stability through caspase-mediated degradation[191]. Interestingly, *IDAX* was derived as a result of chromosomal gene inversion during evolutionary and was originally encoded within the ancestral *TET2* gene, which was split into a segment containing the catalytic domain (*TET2*) and another segment containing the CXXC domain (*IDAX*)[17,192]. The same report also revealed that TET3, but not TET1, exerts autoinhibitory regulation of its own stability via its N-term CXXC domain. In addition to these findings on post-translational regulation of the TET family proteins, a recent study showed that in the mouse genome, at the *TET1* gene locus, there is an alternative promoter that regulates expression of a truncated isoform[193]. Since all these studies were conducted in mouse cells with mouse proteins, it is important to elucidate whether these events also take place in human, which may help provide answers to the puzzling presence of the shorter version of the TET1 protein seen in prostate cancer cells.

**II. Results**

<u>Detection of TET1 isoform in prostate cell lines</u>

We began our investigation on the truncated TET1 protein by expanding western experiment to include 293T cell as well as normal prostate and prostate cancer cells. As shown in the top panel of **Figure 3.1A**, the smaller protein of approximately 150kDa is not only abundantly detected in LNCaP cells, but also in AR-positive cell lines VCaP, C4-2B and 22Rv1. Upon longer film exposure time (middle panel), it can be observed that while this protein has much less significant abundance in other AR-negative cell lines PC3, DU145, PrEC and BPH1 (the latter 2 considered normal or benign cells), it is exclusively expressed in prostate cells but not 293T. Consistent with previous results, at the 235kDa predicted size of TET1, only a very faint band can be seen with longer exposure time. Additionally, level of this larger protein in 293T is higher than most prostate cells and comparable to only the highest-expressing prostate cell line C4-2B. This result suggests that there may be tissue-specific expression of a smaller form of the TET1 protein that is found in higher levels in prostate cancer cells.

To rule out the possibility that the smaller protein is a product of degradation, we used a proteosome inhibitor MG132 in LNCaP and 293T cells, and examined cytoplasmic and nuclear fractions for TET1 level following 24hr of treatment with 10uM MG132 or DMSO control. If the main band at 150kDa in LNCaP cells we observe is a result of proteasome-mediated protein degradation from the full length, we would expect to see less degradation upon MG132 treatment but restoration of the 250kDa protein. However, western blot showed that in LNCaP cells TET1 is found in both cytoplasmic and nuclear compartments, and MG132 treatment did not reduce the smaller protein amount (**Figure 3.1B**), indicating that its presence is not a mere effect of protein degradation. On the other hand, 293T cell contained the full length protein only in the nucleus and a protein around 100kDa in the cytoplasm, which suggests that there may be cell type-specific expression of different forms of the TET1 protein.

**Figure 3.1**

A.
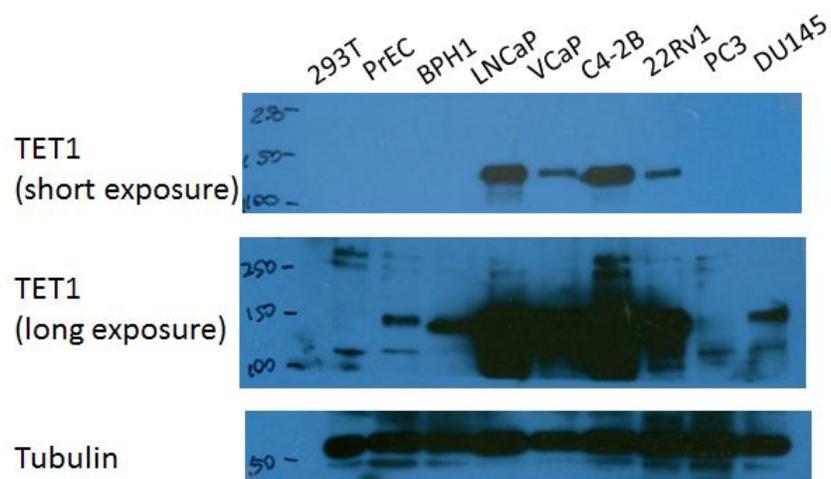


B.
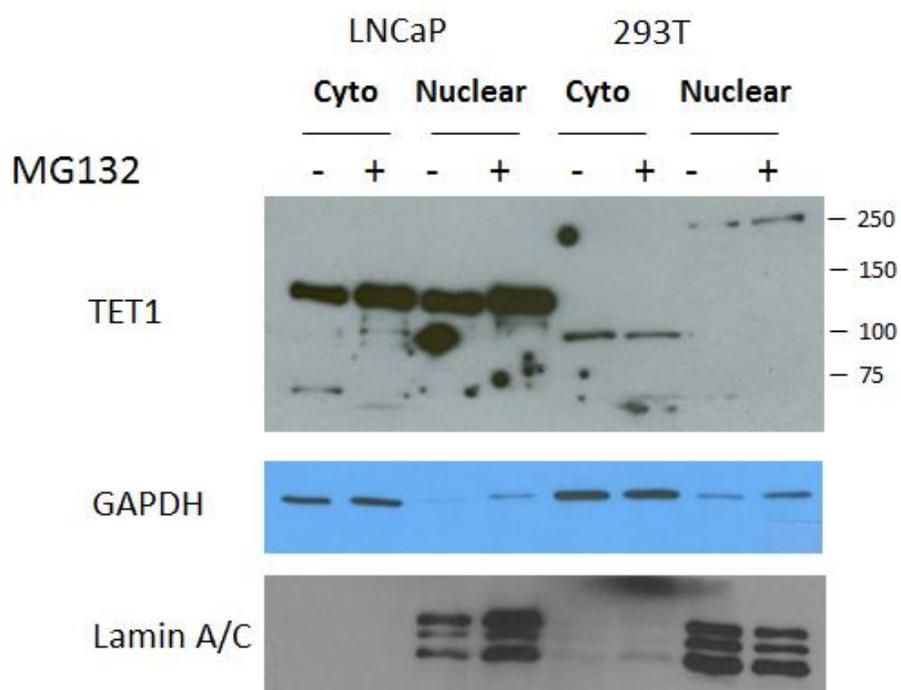
**Figure 3.1 TET1 of different sizes detected in a panel of cell lines.** (**A**) Western blot using an anti-TET1 antibody was done with lysates from several prostate cell lines and 293T cells. (**B**) Western blot of TET1 in cytoplasmic and nuclear portions of LNCaP and 293T cells upon treatment with vehicle control DMSO or 10uM MG132 for 24hr. GAPDH and Lamin A/C are used as loading controls for cytoplasmic and nuclear proteins, respectively.

The human TET1 gene localizes in chromosome 10, and is consisted of 12 exons. To date there have been no reports published on alternative splicing events or gene variants. In order to explore the likelihood that the smaller protein may be encoded by a shorter transcript, we utilized qPCR to look at whether primers spanning different exons would result in differential expression. To this end, 2 sets of primers were used, where one set spans across exons 1 and 2 and the other spans across exons 11 and 12. The results showed that strikingly, primers targeting the first 2 exons yielded very minimal expression in prostate cells, whereas primers targeting the last 2 exons gave much more detectable expression that may correlate with the western results (**Figure 3.2A**). It appears that a transcript containing the first 2 exons is barely detectable in prostate cell lines, unlike in 293T where it's 15- to 180-fold higher (relative to C4-2B and LNCaP, respectively). Thus, we postulated that there may be a major form of TET1 transcript in prostate cancer as well as 293T cells which does not include exons 1 and 2.

With the help of MiTranscriptome[194], which is an extensive compilation of RNA-seq data from over 6.500 samples, comprehensively cataloging human RNA transcripts in diverse tissue and cancer types (many previously unannotated), we found that indeed there are multiple transcript variants for TET1, among which 2 are lacking the first 2 exons (**Figure 3.2B**). In the genome browser view of the TET1 locus, there are 2 shorter transcripts that start slightly upstream of exon 3, with a cryptic exon that resides in the intronic region. These 2 shorter transcripts differ by the inclusion of another cryptic exon between exons 7 and 8 (indicated by dotted black box on the right). This further supports our hypothesis that the truncated TET1 protein observed in prostate cell lines may be encoded by a variant transcript shorter than the full length gene. Next, we proceeded to verify and characterize this isoform transcript in LNCaP cells.

**Figure 3.2**

**A.**



TET1 isoforms
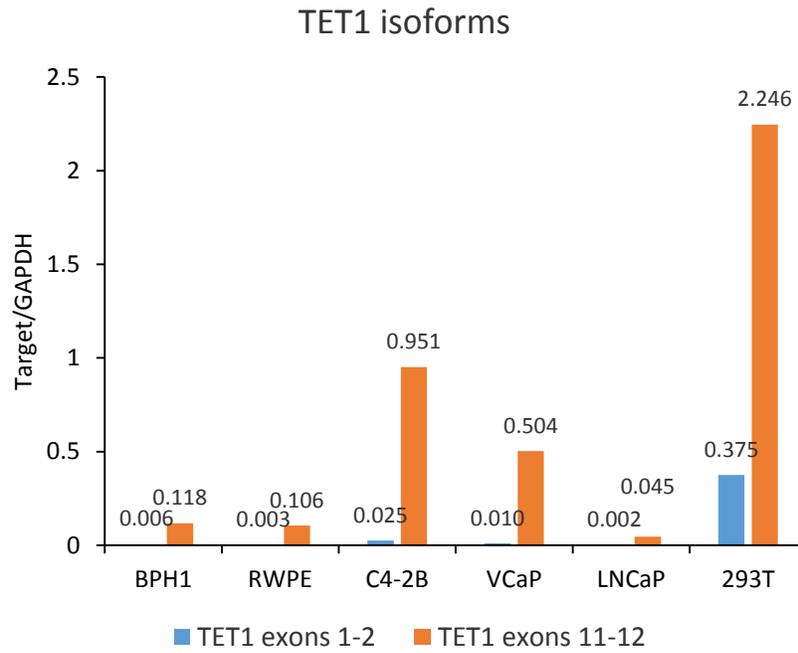
**B.**
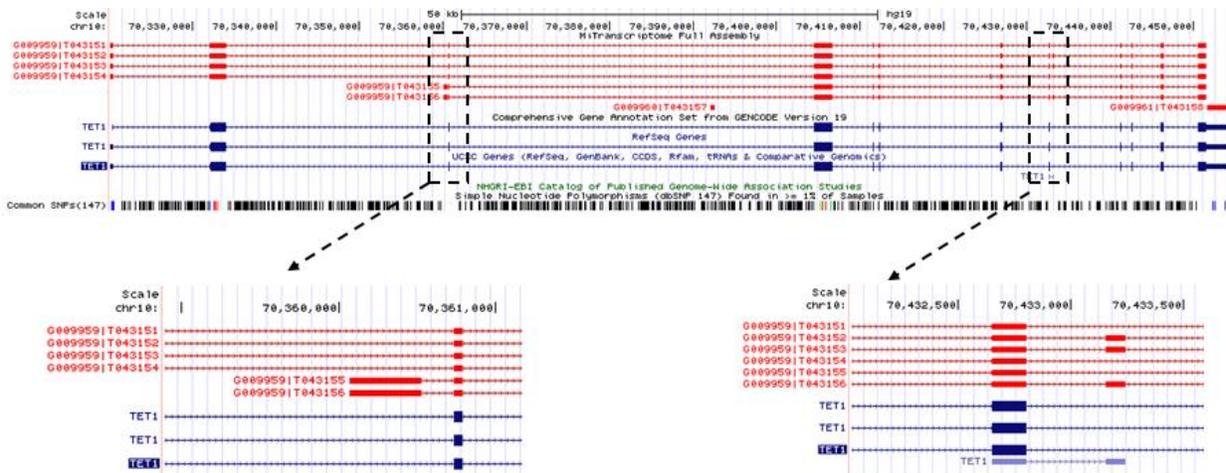
**Figure 3.2 There may be alternative gene transcripts for TET1.** (**A**) qPCR was performed in a number of cell lines with 2 sets of primers targeting either 5' or 3' end of the TET1 gene. (**B**) Genome browser view of MiTranscriptome data showing the presence of a shorter TET1 transcript.

Characterization of TET1 isoform in LNCaP cells

We performed 5' rapid amplification of cDNA ends (RACE), with an anchor primer (GSP1) designed to be within exon 4, as illustrated in **Figure 3.3A**. This primer will anneal to mRNA and reverse transcribe it into cDNA from 3' to 5' in terms of the original transcript. After RNA is degraded, the cDNA strand will be tailed with dCTP on its 3' end (or 5' of the RNA template), which is then used for PCR with a universal primer on the 3' end and GSP2 on the 5'. The amplicon is then subjected to another nested PCR, and the final PCR products are used for visualization on agarose gel as well as Sanger sequencing. Using RNA from 293T, LNCaP and 22Rv1 cells, we observed that there are fragments of different lengths in the cell types (**Figure 3.3B**), indicative of variant forms of transcripts amplified from the same region in TET1 exon 4. While a band of around 2000bp was seen in 293T, presumably the full length gene, it was not detected in either LNCaP or 22Rv1. In particular, only one band was seen in LNCaP, of around 600bp, and the PCR product was subsequently sequenced to determine the exact location of 5' start site. As a result, consistent with MiTranscriptome data, a ~300bp cryptic exon was identified upstream of exon 3 in TET1 (**Figure 3.3C**), giving rise to the truncated TET1 transcript seen in LNCaP.

In order to confirm that all exons are included in the isoform transcript, we designed a series of primers spanning across adjacent exons, as depicted in **Figure 3.4A**, for PCR analysis. Bands of expected sizes are seen with all sets of primers, suggesting that neighboring exons are indeed joined together. It should be noted that the isoform transcript is predicted to give rise to a smaller protein with 1465 residues encoded by the overlapping region with the full length gene, and it would retain the catalytic domain on its C-terminus but not the CXXC domain. In agreement with our previous qPCR results (**Figure 3.2A**), the full length transcript containing exons 1 and 2 is much more abundant in 293T than LNCaP. Therefore, using primers specific for TET1 full length (FL) or short isoform (Iso), we also performed qPCR to determine their relative expression levels in a panel of cell lines (**Figure 3.4B**). Consistent with western blot result (**Figure 3.1A**), TET1 FL is significantly lower than the shorter transcript in prostate cells, and it also appears that Iso is more abundant in AR-positive cells, including C4-2B and VCaP, suggesting it may play a more important role in those cells compared to FL.

**Figure 3.3**

**A.**
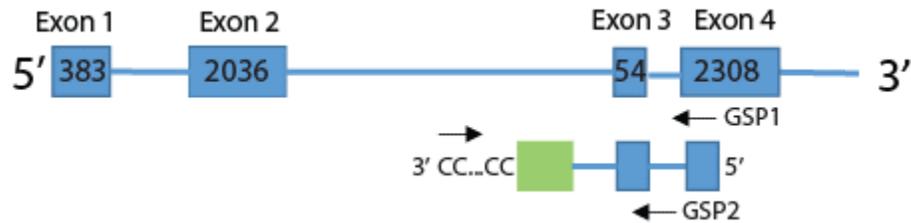


**B.**
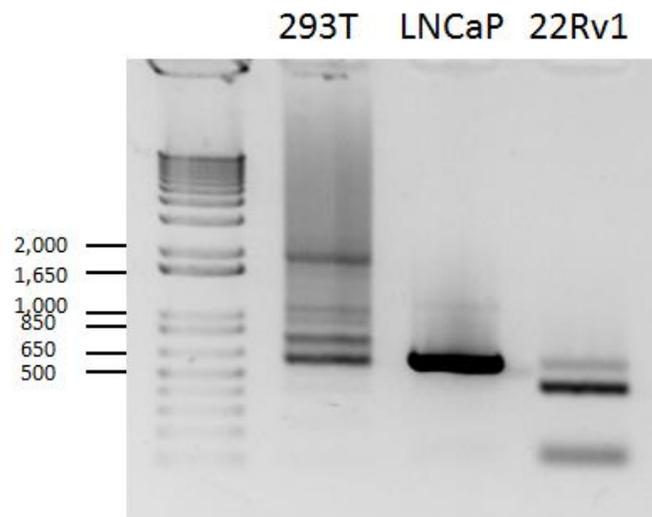


**C.**
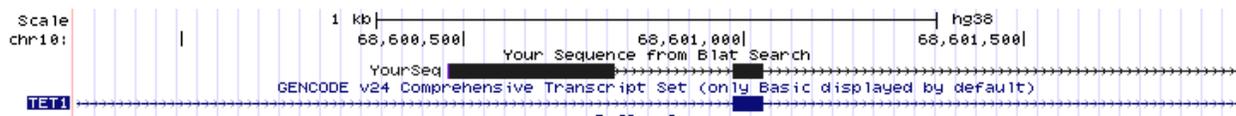


Additional exon sequence:

GATCCTTTGTCTAGGAAGGGCTAGGCTGAGAAGATCACTTTGGAGTTGCTAAGCTCTGCA
TTGCCATTGGCTGCACTCTTTGTGGGGCTGTTTTTGCAAAAGGCATGCCCAGCACGTCTA
CAGTGCTGGCAGGCCAGGGCTGCCTTTAAACACCATGATAATAATACCAAAAAGCAGTTT
GCAGAACTGTTCCACTGTAGAGGCTTCCTCTGCAGAAGCAAGCAAGATGGCTACCTCGT
GGGCTTCGCTGTGAAGAATGCAGAGGCTAATTAAGACAGACTTTAAGGGGAAAAAAGG

**Figure 3.3 Characterization of TET1 transcript isoform.** (**A**) Schematic of 5' rapid amplification of cDNA ends (RACE) experimental design. (**B**) Transcript fragments of different sizes detected in 293T, LNCaP and 22Rv1 cells. (**C**) Sanger sequencing of RACE product in LNCaP reveals a hidden exon 5' of exon 3.

**Figure 3.4**

**A.**
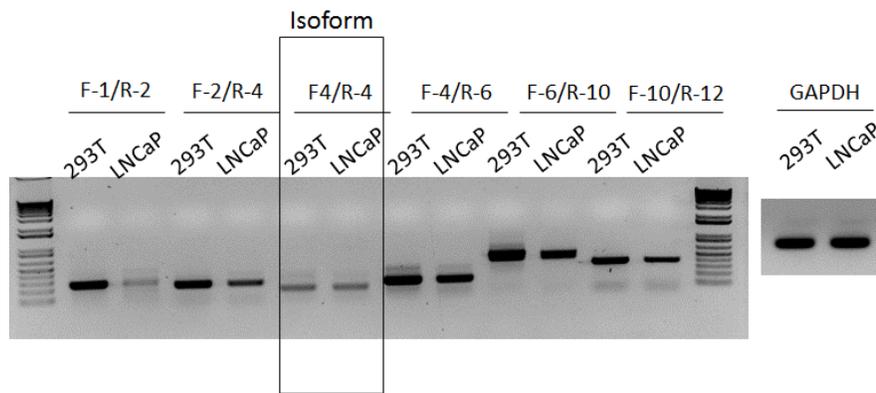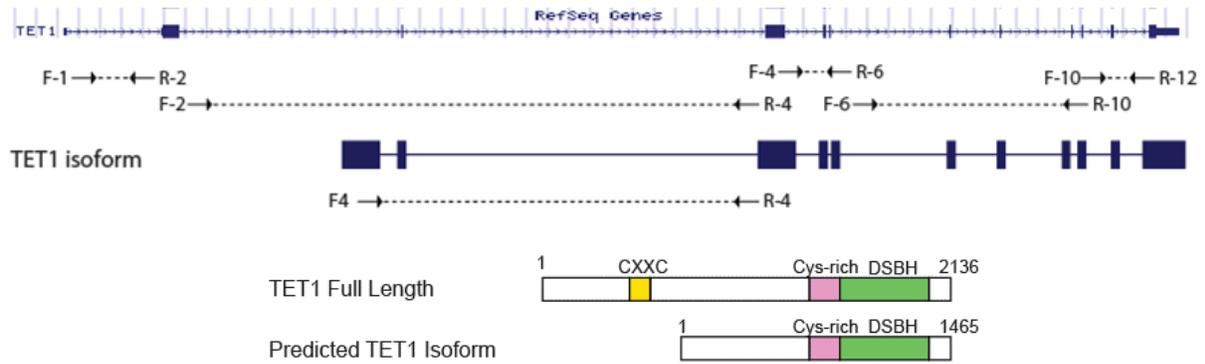


**B.**
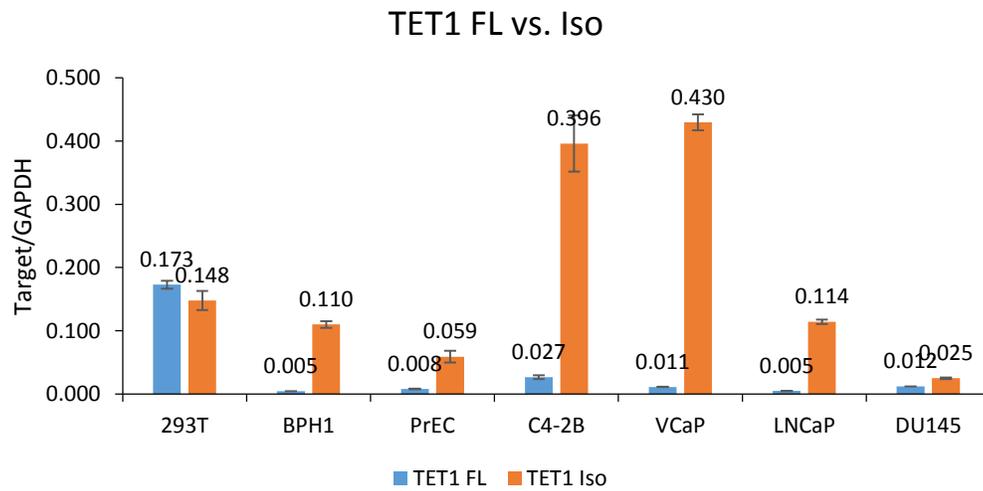
**Figure 3.4 Quantitative comparison between TET1 full length and transcript variant.** (**A**) Primer design for PCR analysis of exon junction for full length and isoform genes. The full length wildtype protein and predicted isoform protein containing the catalytic C-terminus are indicated below. (**B**) qPCR using primers specific for TET1 full length or isoform transcripts in various cell lines. Data shown is mean ± SEM of technical replicates.

Regulation of TET1 isoform

Next, we asked how TET1 isoform is regulated in LNCaP cells. As discussed earlier, FOXA1 can transcriptionally induce TET1's expression level in a number of cell lines. **Figure 2.9A** in previous chapter showed that FOXA1 knockdown in LNCaP cells significantly decreased TET1 RNA and protein levels. Notably, according to western blot, reduced TET1 was detected at 150kDa, thus suggesting that FOXA1 does exert its transcriptional regulation on TET1 isoform. In addition, qPCR primers used in the early experiment spanned across exons 11 and 12 in the TET1 gene, and since isoform is more than 20-fold higher than full length gene, therefore the 50% reduction of TET1 transcript observed in FOXA1-depleted cells was mainly due to regulation on the isoform.

To further examine how FOXA1 transcriptionally regulates the isoform gene, we looked at FOXA1 ChIP-seq data in LNCaP and found that it has a binding site immediately upstream of TET1 exon 3, where presumably the alternative promoter for TET1-iso lies (**Figure 3.5A**). Using primers designed under the peak for qPCR, we were able to validate that FOXA1 indeed binds to the promoter region of TET1-iso, in both LNCaP and C4-2B cells (**Figure 3.5B**). Moreover, RNA Pol II (phospho-Ser5) ChIP in control or shFOXA1 LNCaP cells showed that there was significantly more Pol II occupancy on the promoter of TET1-iso compared to FL, and more importantly it was decreased upon FOXA1 knockdown. Taken together, these results indicate that TET1-iso has its own alternative promoter that is also bound by FOXA1, which has the ability to transcriptionally turn on TET1-iso expression in prostate cancer cells including LNCaP and C4-2B.

**Figure 3.5**

**A.**



**B.**



FOXA1 ChIP on TET1-iso promoter

**C.**



P-Pol II ChIP
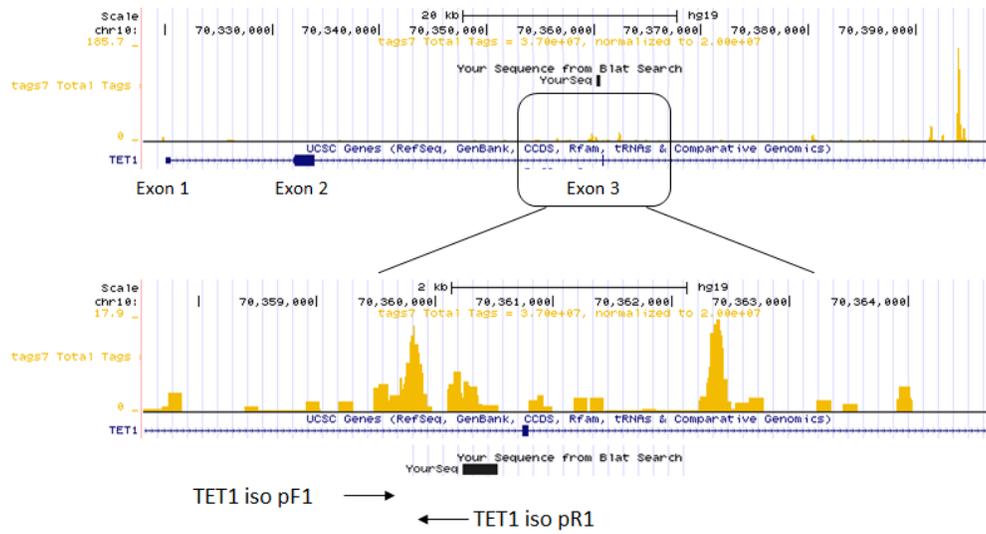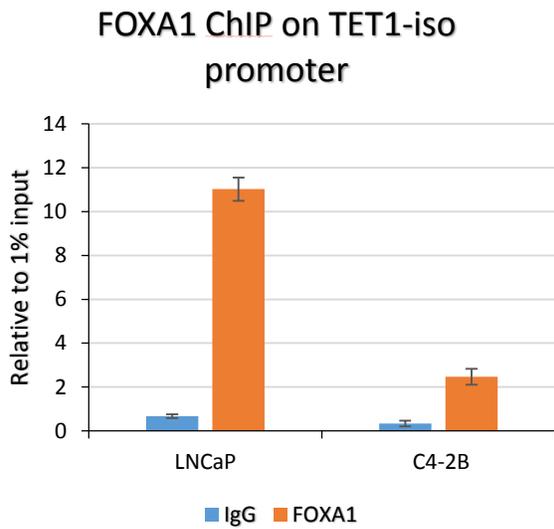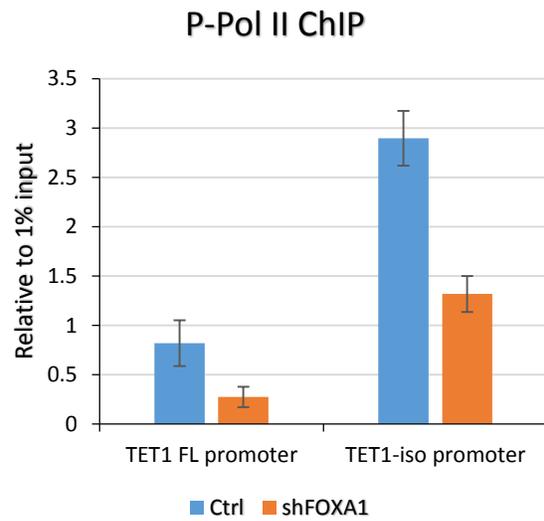
**Figure 3.5 FOXA1 regulates TET1 isoform expression.** (**A**) FOXA1 ChIP-seq demonstrates its binding site in promoter region of TET1 isoform transcript. (**B**) FOXA1 ChIP-qPCR validating its occupancy on TET1-iso promoter in both LNCaP and C4-2B cells. IgG is used as a negative control for ChIP. (**C**) RNA p-Pol II ChIP in control and shFOXA1 cells showed decreased Pol II binding on both promoter regions of TET1 FL and iso upon FOXA1 knockdown. Data shown is mean ± SEM of technical replicates.

Both TET1 full length and isoform are important for prostate cancer cell growth and survival

In order to fully understand the functions of TET1 FL and Iso in prostate cancer cells, we designed shRNAs specifically targeting each, and used lentiviral-mediated infection to study whether they have similar or disparate roles in prostate cancer. The sequences for FL- and Iso-specific shRNAs locate in the unique regions of each transcript (in exon 2 and the cryptic exon in front of exon 3, respectively), thus would only recognize and deplete either FL or Iso without affecting the other. After lentiviral transduction, it was observed that there was an immediate effect of increased cell death with shTET1-FL, but not shTET1-Iso. **Figure 3.6A** shows images of LNCaP control, shTET1-FL and shTET1-Iso cells taken 4 days post-infection, and it can be seen that there are apoptotic features demonstrated with shTET1-FL, but not shTET1-Iso. However, colony formation assays performed with the cells for 2 weeks exhibited drastically reduced cell viability for not only shTET1-FL cells but also shTET1-Iso as well (**Figure 3.6B**), which could indicate that loss of FL and Iso proteins may ultimately lead to the same phenotype, but their underlying mechanisms may potentially differ.

Both TET1 full length and isoform regulate AR expression

To better understand how the effects of inhibited cell growth and survival are mediated by TET1 FL and Iso depletion, we used qRT-PCR and western blot to look at changes in AR and its downstream targets. Here, an additional "shTET1 total" was included which targets both FL and Iso, that had been used in previous experiments (Chapter 2). First of all, by PCR with primers specific for FL or Iso, we were able to confirm that knockdown was achieved at 20% (**Figure 3.7A**). Similarly at the protein level, we observed that while both the 250kDa and 150kDa TET1 proteins were decreased with shTET1 Total, only the corresponding protein was knocked down with specific shTET1's (**Figure 3.7B**). Interestingly and unexpectedly, there appeared to be a reciprocal increase on the non-targeted protein, leading to a slightly higher level of TET1 Iso or FL upon shRNA infection with shTET1 FL or Iso, respectively.

**Figure 3.6**

**A.**

Ctrl
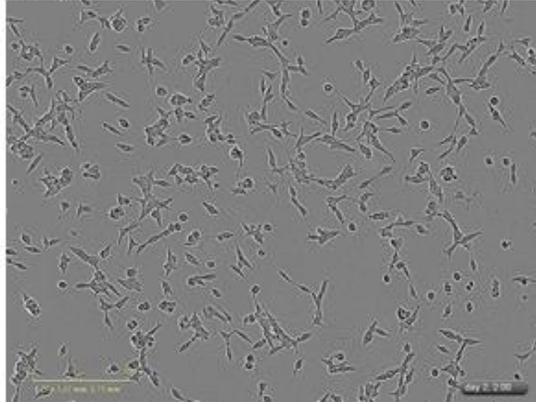


shTET1 FL
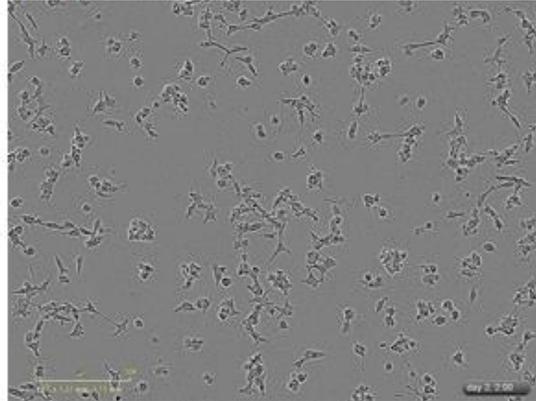


shTET1 Iso
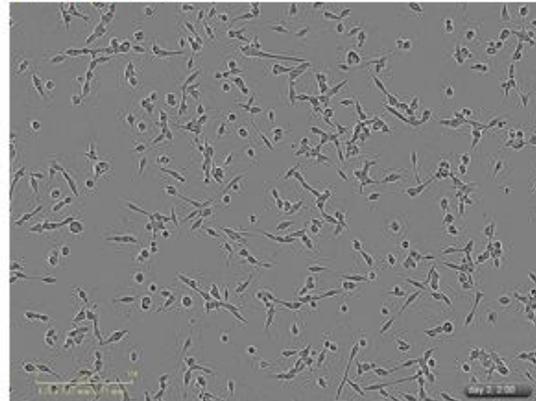
**B.**



Ctrl
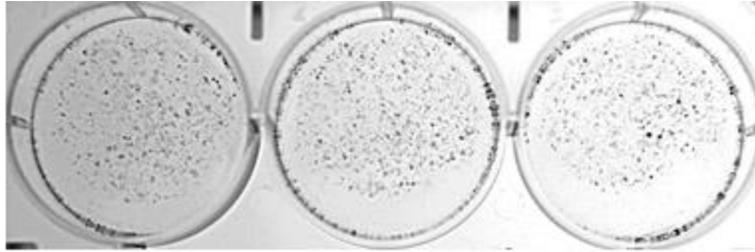
shTET1 FL
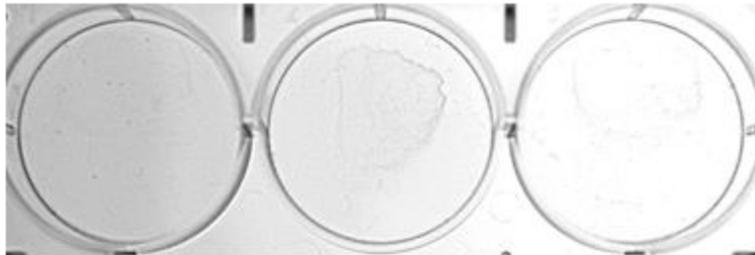
shTET1 Iso
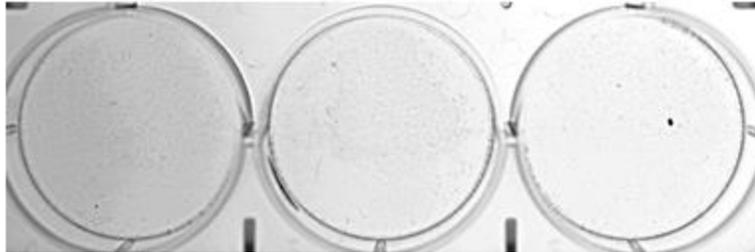
**Figure 3.6 Effects of TET1 FL or Iso knockdown on LNCaP cell growth.** (**A**) Images of cells infected with LNCaP control, shTET1 FL or shTET1 Iso lentiviruses for 4 days. (**B**) Images of colony formation assays of LNCaP control, shTET1 FL or shTET1 Iso cells 2 weeks post-infection.

**Figure 3.7**

**A.**
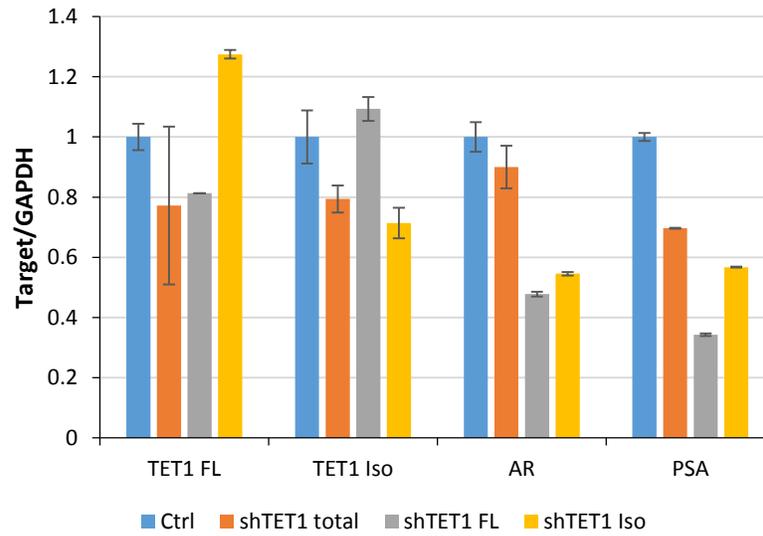


**B.**
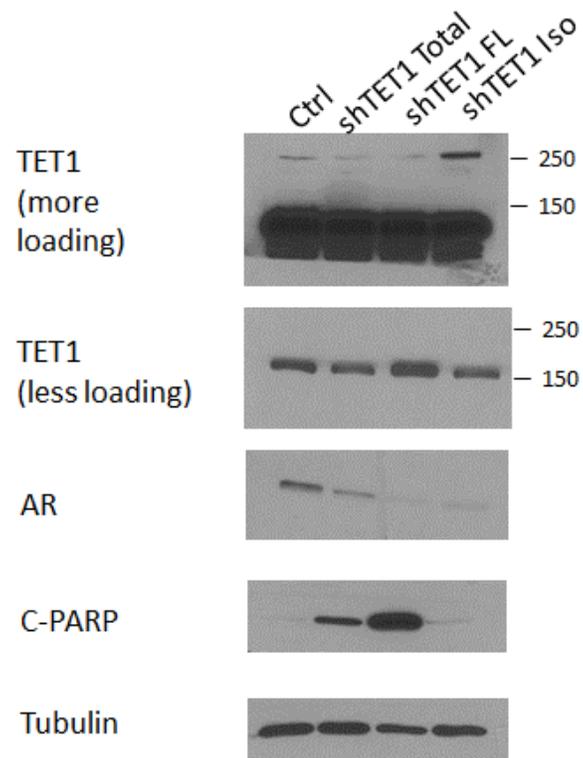
**Figure 3.7 Both TET1 FL and Iso regulate AR expression.** (**A**) qRT-PCR was done using LNCaP cells subjected to control, shTET1 total, shTET1 FL or shTET1 Iso lentiviral-infection for 4 days. Primers unique to TET1 FL or Iso were used to confirm knockdown efficiency and specificity. Data shown is mean ± SEM of technical replicates from one representative experiment out of two. (**B**) The same cells were also used for western analysis. Tubulin was used as loading control.

Next, we went on to examine changes in AR level after TET1 knockdown, to see if the most critical signaling pathway in prostate cells can be affected by loss of TET1 FL or Iso. Reflected by both qRT-PCR and western blot (**Figure 3.7A-B**), AR was decreased upon knockdown with all 3 shRNA's. However, compared to shTET1 total, the single-targeting shTET1's were more effective in decreasing AR at both the transcript and protein level. In addition, a canonical AR-induced gene, PSA, was shown to be significantly downregulated upon loss of TET1 by all 3 shRNA constructs (**Figure 3.7A**), suggesting that AR-mediated transcriptional program could be severely impaired due to loss of AR. Moreover, consistent with observed phenotypes and cell images taken in **Figure 3.6A**, the level of cleaved PARP, a marker for apoptosis, was also dramatically increased in shTET1 FL cells (**Figure 3.7B**).

Using RNA from LNCaP cells infected with either control or shTET1 total, we had performed gene expression microarray for triplicate samples. Subsequently, Gene Set Enrichment Analysis (GSEA) was carried out to look at how AR-regulated genes were affected by total knockdown of TET1. As shown in **Figure 3.8** (left), AR-induced genes were significantly downregulated in shTET1 cells, suggesting that depletion of total TET1 had a global effect on suppressing AR transcriptional program. On the other hand however, AR-repressed genes were not significantly affected by TET1 total knockdown (**Figure 3.8, right**). This finding is concordant with reduction of AR observed in shTET1 experiments (**Figure 3.7A-B**), and similar results would be expected for shTET1 FL or shTET1 Iso as well, since AR is significantly downregulated in both.

Overall, it can be summarized that TET1 FL and Iso are both required for prostate cancer cell LNCaP's growth and survival, and loss of either leads to reduction in AR expression, manifested at both mRNA and protein levels. However, TET1 FL seems to play a role in suppressing apoptosis. With preliminary evidence, it can be seen that despite its low expression in LNCaP cells, TET1 FL knockdown gives rise to substantial apoptosis in the cells. Therefore, although reduction in AR and its transcriptional activity may be detected in both TET1 FL and Iso knockdown cells, the mechanisms associated with either FL or Iso proteins may not entirely mimic each other, and further in-depth studies are needed to decipher their own unique regulatory roles in prostate cancer.

**Figure 3.8**
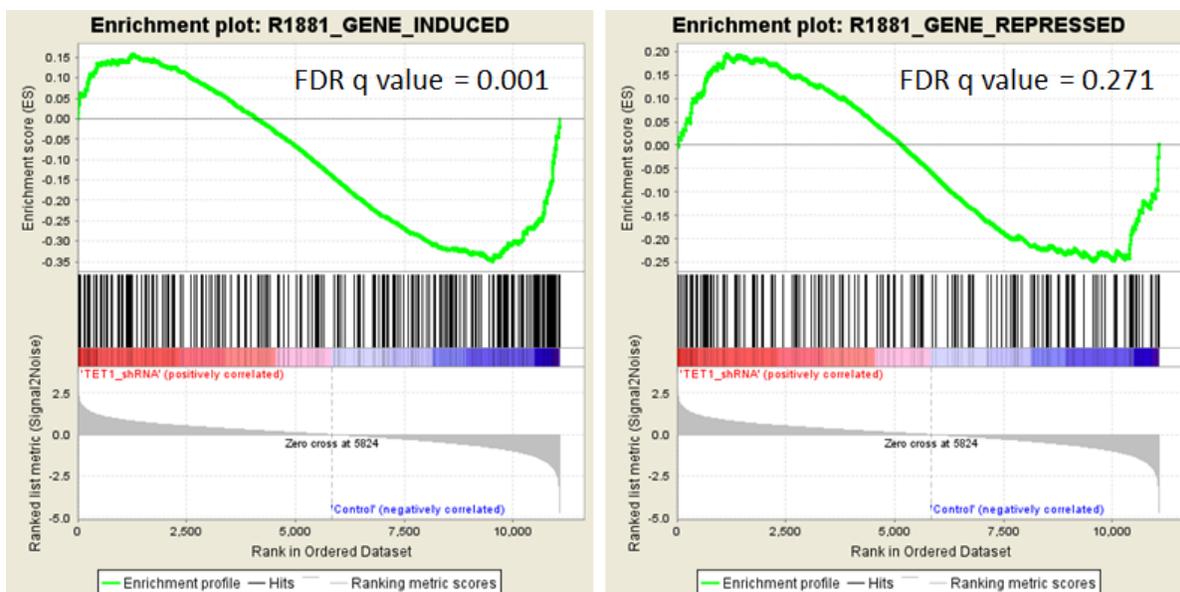
**Figure 3.8 GSEA analyses of TET1-regulated genes in LNCaP.** Using gene expression microarray data of triplicate samples for LNCaP control or shTET1 total, GSEA analysis was performed and showed that while AR-induced genes were significantly downregulated in TET1 knockdown compared to control (left), AR-repressed genes were not statistically differential in the two conditions (right). This figure was generated by Jonathan Zhao.

TET1 expression is higher in patient datasets

In addition to understanding TET1's contribution to prostate cancer development at the molecular level, we were also interested in how it may be manifested in patients. To this end, we utilized publicly available RNA-seq datasets from TCGA and dbGaP (the database of Genotypes and Phenotypes) in conjunction, and compared expression of TET1 in 3 groups of patient samples, normal, primary and metastatic tumor (sample size indicated on x-axis). As depicted in boxplot in **Figure 3.9A**, there is a statistically significant upregulation of TET1 in the metastatic group, but not in the primary tumor group. Thus, it could be understood that higher TET1 level may be associated with more late-stage and aggressive prostate tumors.

Furthermore, since it was shown that TET1 can transcriptionally induce AR, we also assessed correlation between TET1 and AR expressions in TCGA. The scatter plot in **Figure 3.9B** reveals that TET1 and AR mRNA levels are positively correlated, with a statistically significant Pearson coefficient of 0.46. This additional evidence serves to support our findings in molecular studies that TET1 is important for AR expression level and subsequently transcriptional activities, thus rendering TET1 a potential unfavorable factor in context of prostate cancer progression.

**Figure 3.9**

**A.**



**TET1 in TCGA_DBGAP dataset**
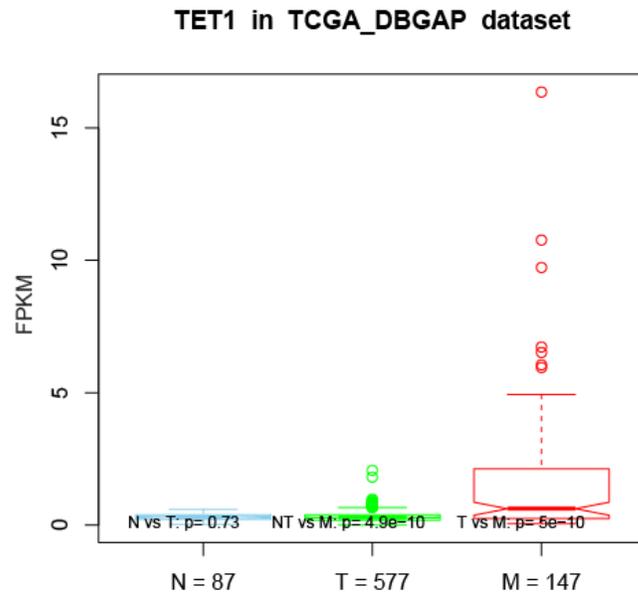
**B.**
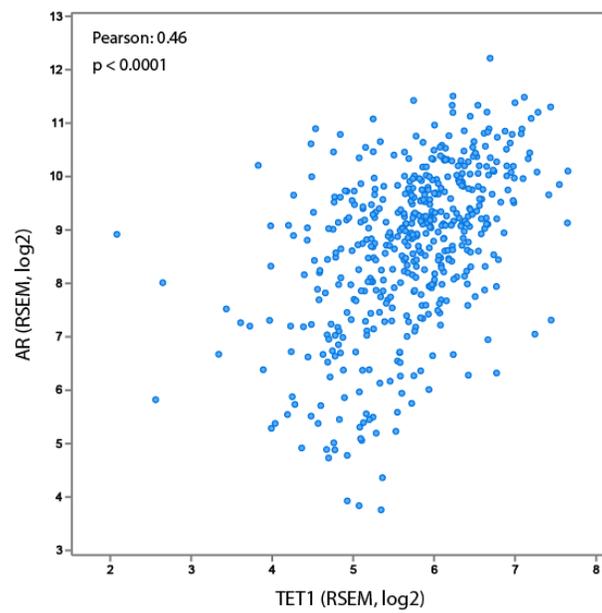
**Figure 3.9 Analyses of prostate cancer patient datasets for TET1 expression.** (**A**) Boxplot of TET1 expression in combined RNA-seq data from TCGA and dbGaP (the database of Genotypes and Phenotypes). Samples were categorized into 3 groups, normal (n=87), primary tumor (n=577), and metastatic tumor (n=147). TET1 is only significantly upregulated in the metastatic group. (**B**) Correlation analysis between TET1 and AR expression in TCGA. Pearson correlation coefficient is 0.46, with p value <0.0001. A was generated by Jonathan Zhao.

**III. Discussion**

Through our endeavor to better understand TET1's biological functions in prostate cancer, we came to discover that unlike in other cell lines, TET1 is manifested as a truncated protein in a number of AR-positive prostate cancer cells. The 150kDa protein detected for TET1, as opposed to its predicted 235kDa, may result from an alternative mRNA transcript driven by a promoter slightly upstream of exon 3 in the original gene. Consequently, this gives rise to an isoform that lacks the first 2 exons but has inclusion of a 300bp cryptic exon before exon 3.

It was only recently reported that in mouse genome an alternative transcript exists for the Tet1 gene, which similarly lacks its first 2 exons[193]. Here, we are providing evidence for the first time that in human genome as well a TET1 isoform is found, and its expression is predominant in a panel of prostate cells, exhibiting 2- to 40-fold higher mRNA level than the full length transcript depending on cell type. In their Molecular Cell paper, Zhang et al. claimed that Tet1 undergoes isoform switching during mouse development, throughout which process the full length gene, high in embryos, embryonic stem cells and primordial germ cells, becomes shut off while the isoform gets expressed when cells start to differentiate into somatic tissues. As the short isoform encodes for a protein without the N-terminus CXXC domain, it has diminished ability to target and bind to DNA, thus has impaired functions in demethylation and imprint erasure.

Correspondingly in our study, we were able to characterize and sequence the variant transcript with 5' RACE, and subsequently assessed its expression level in various cell lines to show that, in consistence with western blot results, the isoform is higher in cancer cells compared to normal or benign cell lines. Moreover, we showed that both the full length and isoform proteins may play a role in transcriptionally regulating AR expression and thus are important for LNCaP cell growth and survival. However, they do slightly differ in that knockdown of FL seems to immediately turn on apoptotic pathways in LNCaP cells, while knockdown of Iso does not affect apoptosis but rather decreases cell viability over a 2-week period. Ultimately, loss of either protein can significantly negatively impact cell survival and colony-forming capabilities in LNCaP. The differential phenotypes associated with specifically depleting

FL or Iso suggest there may be distinct pathways that can be regulated by each, in addition to the fact that both have the ability to regulate AR. Therefore, it would be critical to investigate how these 2 proteins function differently in the context of prostate cancer, and whether the lack of CXXC domain in Iso renders it less effective at demethylating DNA. In addition, gene expression microarray analyses could be performed to examine which genes regulated by FL may be responsible for activation of apoptosis, possibly through caspase 3-mediated mechanism as reflected by accumulation of cleaved PARP, in shTET1 FL cells.

As we are just beginning to learn about the roles of TET1 FL and Iso in prostate cancer, it could also be informative to study how they both regulate AR, with more molecular and mechanistic details. Methylation analyses may be needed to reveal whether decreased AR expression can be attributed to buildup of DNA methylation on AR promoter in the knockdown cells. Moreover, ChIP experiments could help to elucidate whether and where the FL and Iso proteins bind around AR promoter region. Also, importantly, rescue experiments should be conducted to verify the effects FL- or Iso-knockdown by overexpressing the respective protein, in order to demonstrate more compellingly their direct regulation of AR. Along the same line, we could also utilize CRISPR technology to selectively delete FL gene or both genes altogether, to further confirm TET1's regulation of AR, as well as other potential downstream pathways pertaining to cell apoptosis and cell survival.

Furthermore, as seen in **Figure 3.7B**, using single-targeting shRNA to specifically deplete either FL or Iso results in increased protein level of the other, which is potentially indicative of a negative feedback mechanism at the protein level that allows reciprocal regulation between TET1 FL and Iso. A similar autoinhibitory phenomenon has been illustrated for Tet3 in mouse cells, where it was reported that the CXXC domain can interact with CD domain within Tet3 protein and target it for caspase-induced degradation[191]. Although the paper did not identify the same effect on Tet1, but due to the differences in human and mouse TET protein sequences, it remains to be investigated whether a similar regulatory mechanism can be applied to human TET1.

In our analyses of RNA-seq data from TCGA and dbGaP, we observed higher expression of TET1 in patients of metastatic prostate cancer. Since qPCR in cell line panel showed that TET1 Iso is the predominant form in all prostate cells, it is deduced that the upregulation of TET1 seen in patient samples should be attributable to TET1 Iso rather than FL. However, perhaps more advanced algorithms for RNA-seq analysis could be applied in the future to distinguish between FL and Iso in computing their expression levels in a specific manner.

With growing evidence, we are now obtaining clearer pictures of TET1 FL and Iso in prostate cancer development. It appears that both are required for prostate cancer cell growth and survival, therefore conferring oncogenic propensities to the cells and stimulating prostate cancer progression. In future studies, we may focus on elucidating the underlying mechanisms that are employed by either TET1 FL or Iso, in order to comprehensively understand which pathways, in addition to AR signaling, are under regulation by TET1.

# CHAPTER 4: MATERIALS AND METHODS

## Cell lines, plasmids and antibodies

Prostate cancer cell lines LNCaP, VCaP, 22Rv1, BPH1, RWPE-1, DU145 and human embryonic kidney cell line HEK293T cells were obtained from American Type Culture Collection and cultured in either RPMI1640 or Dulbecco's modified Eagle's medium with 10% fetal bovine serum (FBS). For FOXA1 and TET1 FL and domain constructs, human FOXA1 and TET1 cDNA were amplified by reverse transcription polymerase chain reaction (PCR) from LNCaP cells and pENTR223 TET1 (Harvard Plasmid), respectively, and cloned into the entry vector pCR8/GW/TOPO (Invitrogen). Adenoviral construct expressing FOXA1 was generated by recombining pCR8-FOXA1 with pAD/CMV/V5 using LR Clonase II (Invitrogen). Overexpression constructs for TET1 were generated by recombination of pCR8-TET1 with NTSFB destination vector or pLenti CMV/TO Puro DEST (Addgene plasmid 17 293). The pGIPZ lentiviral control and FOXA1 shRNAs were purchased from Open Biosystems. Sequences for scramble (5′-GCGCGCTTTGTAGGATTCG-3′) and total TET1 (5′-GTGGAGAAGTGGACACAAA-3′) shRNA were kindly provided by Dr Debabrata Chakravarti (Northwestern University), and cloned into pLKO lentiviral vector. Target sequences for shTET1 FL and shTET1 Iso are 5'-GCTACGAAGCACCTCTCTTAG-3' and 5'-GAGTTGCTAAGCTCTGCATTG-3', respectively.

The antibodies used in this study include anti-FOXA1 (ab23738) and anti-GAPDH (ab9385) from Abcam, anti-TET1 (GTX627420 and GTX124207) from GeneTex, anti-FLAG (F1804 and F7425) from Sigma, anti-c-Myc (sc-789x) from Santa Cruz, anti-HA (ab9110) from Abcam, anti-alpha Tubulin (sc-32293) from Santa Cruz, anti-Lamin A/C (2032S) from Cell Signaling, anti-5mC (BI-MECY-0100) from Eurogentec, anti-5hmC (39769) from Active Motif, anti-H3K4me2 (07-030) from Millipore, anti-H3K27ac (ab4279) from Abcam.

## Luciferase reporter assay

TET1 promoter and enhancer luciferase reporter assays were conducted according to the manual of Dual-Luciferase Reporter Assay System from Promega. Briefly, LNCaP cells were seeded in a 24-well

plate and co-transfected with the Renilla expression plasmid pRL-TK and the reporter constructs for TET1 promoter and enhancer in pGL4 vector. Cells were infected with LacZ (control) or FOXA1 adenovirus for 48 h to assess the effect of FOXA1 overexpression on luciferase activity. Conversely, to look at FOXA1 depletion effect, lentiviral-transduced shCtrl and shFOXA1 LNCaP cells were used for co-transfection of reporter constructs. Luciferase activities were determined 48–72 h post-transfection and normalized against Renilla internal control values.

## Immunofluorescence staining

Cells were fixed with 4% formaldehyde for 15 min at RT and then permeabilized in 0.1% Triton X-100 for 15 min at RT. Cells were then washed by phosphate buffered saline (PBS) for three times, followed by incubation with 5% normal goat serum for 30 min at RT. Subsequently, cells were incubated with primary antibody, the anti-FLAG antibody (Sigma) and anti-TET1 (Genetex), for 2 h at RT. After washing three times with PBS, cells were incubated with secondary antibody, Alexa Fluor 488 and 594 goat anti-rabbit IgG (Invitrogen), for 1 h at RT. Finally, cells were washed three times with PBS and mounted using Prolong Gold Antifade Reagent (Invitrogen).

## Co-immunoprecipitation

Nuclear proteins were extracted from 293T or LNCaP cells. For S protein pulldown, nuclear extracts were incubated with 30 µl S-protein agarose beads (Millipore) for 3 h at 4°C. The beads/protein complex was then washed four times, and eluted with 30 µl 2× sodium dodecyl sulphate (SDS) sample buffer and subjected to western blot analysis. For LNCaP endogenous co-immunoprecipitation (co-IP), nuclear extracts were incubated with 2 µg antibodies, anti-FOXA1 (Abcam), anti-TET1 (Genetex) and anti-rabbit IgG (Santa Cruz) overnight at 4°C. Dynabeads Protein A (Life Technologies), 25 µl per immunoprecipitation (IP), were added the next day and incubated for 1 h at 4°C. Similarly, the beads/protein complex was washed four times, and eluted with 30 µl 2× SDS sample buffer and subjected to western blot analysis.

## Nuclear protein extraction

For cell pellets with a packed cell volume of 100ul, 1ml of Lysis Buffer A (10mM HEPES pH 7.9, 10mM KCl, 1.5mM MgCl2, 10% Glycerol, 1mM EDTA pH 8.0) with freshly added protease inhibitors (10% PMSF, cOmplete Protease Inhibitor Cocktail from Roche, 1mM NaVO4, 10mM NaF), was used to resuspend. After 10min of incubation on ice, 0.5% TritonX-100 was added, and sample was vortexed for 15sec to lyse the cells. The homogenate was then centrifuged at 2,000g for 2min at 4oC. The supernatant contained cytoplasmic proteins and was discarded. The nucleus-containing pellet was resuspended with 300ul Lysis Buffer B (20mM Tris-HCl,pH 7.5, 420mM NaCl, 1.5mM MgCl2, 1% TritonX-100, 10% Glycerol, 1mM EDTA pH 8.0) with freshly added protease inhibitors (10% PMSF, Complete Protease Inhibitor Cocktail from Roche, 1mM NaVO4, 10mM NaF). The sample was incubated on ice for 60min, with periodic vortexing every 10min. The final nuclear lysate was obtained by centrifuging for 10min at 14,000g at 4oC, after which the supernatant was transferred to a new tube and diluted with Lysis Buffer D (20mM Tris-HCl pH 7.5, 1.5mM MgCl2, 1% TritonX-100, 10% Glycerol, 1mM EDTA pH 8.0), with freshly added protease inhibitors (same as above), to adjust the salt concentration.

In Vitro pulldown assay

To purify bacterial expressed protein, GST-TET1-CXXC, GST-TET1-CXXC A and GST-TET1-CXXC B were first constructed by sub-cloning PCR fragments into pGEX4T1 vector, while MBP-FOXA1-FH-MYC was constructed by sub-cloning into gateway compatible maltose-binding protein (MBP) destination vector, the plasmids were then transformed into Escherichia coli BL21-CodonPlus (DE3)-RIPL competent cells (Agilent technology) and induced with 0.2mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37 °C for 3 h or 18°C ovenight. The GST fusion proteins were purified by binding to glutathione Sepharose (GE healthcare), the MBP fusion proteins by amylose resin (NEB) according to manufacturer instruction. To test direct interaction of TET1 and FOXA1 in vitro, purified GST-TET1-CXXC, GST-TET1-CXXC A and GST-TET1-CXXC B were incubated with MBP-FOXA1-FH-MYC in co-IP lysis buffer containing 2 mg/ml bovine serum albumin, respectively. In a control, GST-GFP was used instead of the GST-TET1-CXXC. After the incubation and extensive washing, the protein complexes bound to glutathione-Sepharose were eluted with SDS sample buffer and examined by immunoblotting.

<u>Chromatin immunoprecipitation (ChIP)</u>

LNCaP cells were first crosslinked with 1% formaldehyde for 10 min at room temperature, following which 0.125M glycine was added for 5 min at RT to terminate crosslinking. Cells were then rinsed with cold 1X PBS twice. All subsequent steps were carried out at 4°C. After scraping off the cells and spinning down, pellets were resuspended and incubated in cell lysis buffer + 10ul/ml PMSF and protease inhibitor cocktail (Roche) for 10 min. The solution was then centrifuged at 5,000 rpm for 5 min and supernatant containing cytoplasmic fractions was discarded. Next, the nuclei pellets were resuspended in nuclear lysis buffer, and then incubated for another 10 min. After complete lysis, the chromatin was subjected to shearing with a probe sonicator to achieve an average fragment length of 500bp. Cellular debris was removed by centrifuging the sheared chromatin at 14,000rpm for 10 min. Then, the chromatin-containing supernatants were transferred to new tubes and incubated with Salmon sperm agarose protein A or G beads (Millipore) for 15 min to pre-clear the chromatin and eliminate non-specific binding to beads. After spinning down at 5,000 rpm, the pre-cleared chromatin was incubated with 3-5ug of antibody at 4°C overnight, 1% or 5% chromatin was saved as input. The next day 50ul agarose protein A or G beads were added for 2 h. After incubation, the beads were washed twice with 1X dialysis buffer and four times with IP wash buffer. The antibody/protein/DNA complexes were then eluted with 150ul IP elution buffer twice. Next, to reverse crosslinking, the eluates were incubated with 10ug RNase A and 0.3M NaCl at 67°C overnight. The following day, DNA/protein complexes were precipitated with ethanol, and the pellets were air-dried and dissolved in 100ul TE. Proteins were then digested by proteinase K treatment at 45°C for 1 h, and DNA was purified with Qiagen PCR purification kit and eluted in 30ul water.

All primers (**Table 1**) were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/), synthesized by Integrated DNA Technologies and used for SYBR Green based real-time PCR. ChIP-quantitative PCR enrichment of target loci was normalized to input DNA and reported as % of 1% input ± SEM. ChIP DNA was prepared into libraries according to standard protocols using Bioo Scientific's DNA Sample Kit (cat. no. 514101). Libraries were sequenced using Illumina Hi-Seq platforms. Sequence reads were aligned to

the Human Reference Genome (assembly hg19) using Burrows–Wheeler alignment tool (bwa) version 0.6.1. New high-throughput data generated in this study have been deposited in GEO database under accession number GSE73363.

Methylated DNA immunoprecipitation (MeDIP)

Total genomic DNA was extracted using QIAamp DNA Mini Kit (Qiagen) and sonicated to obtain fragments between 300 and 1000 bp. Dynabeads M-280 Sheep anti-Mouse IgG (Invitrogen) were incubated with an anti-5-methylcytidine antibody (BI-MECY_0100, Eurogentec, Fremont, CA, USA) overnight at 4°C. The following day, 4 µg of sheared DNA was denatured by boiling at 95°C for 10 min followed by rapid cooling on ice, and subsequently added to the beads/antibody complex. On day 3, the beads were washed three times with PBS + 0.05% Triton X-100 and eluted from beads by incubation at 65°C for 5 min in 150 µl elution buffer (TE + 1% SDS). Elution was repeated for a total of two times. Total eluates were treated with proteinase K and incubated at 50°C for 2 h. QIAquick PCR purification kit (Qiagen) was used to purify the eluted DNA, and lastly qPCR was used to determine the enrichment of target genomic regions using gene-specific primers (listed in Supplementary Data). Enrichment of target loci was normalized to input DNA and reported as Enrichment over input ±SEM.

5hmC chemical labeling (hMe-Seal)

Genomic DNA was fragmented to an average of 400 bp and was incubated with 50 mM HEPES buffer (pH 7.9), 25 mM MgCl2, 100 mM UDP-6-N3-Glc and 2 mM βGT for 1 h at 37°C. The labeled DNA was purified by the QIAquick Nucleotide Removal kit (Qiagen) and eluted in H2O. The click chemistry was performed with the addition of 150 mM of disulfide-biotin, and the mixture was incubated for 2 h at 37°C. The labeled DNA fragments were then purified by the QIAquick Nucleotide Removal kit (Qiagen) and enriched by Dynabeads Streptavidin C1 (Invitrogen), and subsequently released by dithiothreitol (DTT) treatment. The enriched DNA fragments were first purified by Micro Bio-Spin 6 spin columns (Bio-Rad) followed by MinElute PCR Purification Kit (Qiagen).

Quantitative reverse transcription PCR (qRT-PCR)

Cells were harvested and RNA was isolated with Trizol reagent (Life Tech) according to manufacturer's instructions. Then reverse transcription was carried out to synthesize cDNA, in which 500ng RNA was added with RT buffer, dNTP, random primers, RTase and RNase inhibitor (Applied Biosystems) and incubated at $37^{\circ}C$ for 2 h. The resulting cDNA was subjected to qPCR with SYBR reagent (Midwest Scientific) and specific primers (listed in **Table 2**), placed in StepOne Plus real time thermocycler (Applied Biosystems). Gene expression was analyzed by the $2^{-(\Delta(\Delta Ct))}$ method and normalized to GAPDH.

**Table 1. Primers used for ChIP-, MeDIP-, and hMeSeal-PCR**

| Name | Sequence |
|------|----------|
| GPR137B F | CCCTACTGGGGCACTGTCTA |
| GPR137B R | TTGCAGGGTACAGCCTCTCT |
| MME F | TCCTTGAGCTGTGGTGGACT |
| MME R | CTACGCCCACGGAATCTC |
| FAT1 F | GGTTCCAAGCAAGACAATCC |
| FAT1 R | TAGCAGCTGAAGGGTGTGTG |
| CXADR F | CGCAACCTAGATGCACACAG |
| CXADR R | AGACAGGGTTTCACCACATTG |
| NPC1 F | GGATAGGGAAGCTTCTTTCAA |
| NPC1 R | TTAGGCAGGATGGTCTCGAT |
| SNAIL F | GGGTTACACCCGTGAACAAG |
| SNAIL R | CTGGCACCCTTTCATTCTGT |
| FN1 F | CGCATCTCTTTCCTGTCCAT |
| FN1 R | GAGGCACCACGAGAAGTGAC |
| PNLIP F | TGATGTTCCCACAACAATGA |
| PNLIP R | CATGCACATTGGAAGGTGAG |
| CNTNAP2 F | GGCAGGATTTCCTCAAAGAC |
| CNTNAP2 R | GACATCAGCTATCCCCAGGA |
| TMPRSS2 F | TGGAGCTAGTGCTGCATGTC |
| TMPRSS2 R | CTGCCTTGCTGTGTGAAAAA |
| FKBP5 F | GGTTCCTGGGCAGGAGTAAG |
| FKBP5 R | AACGTGGATCCCACACTCTC |
| PSA F | GCCTGGATCTGAGAGAGATATCATC |
| PSA R | ACACCTTTTTTTTTCTGGATTGTTG |
| TET1 enhancer F | CTCAAGCAATCCTCTTGTCTAGG |
| TET1 enhancer R | TACACACTGAGTTCAGAGCAAGC |
| TET1 promoter F | GAACACAGCCCTCATCTGGT |
| TET1 promoter R | AGAAGGTGCCAGGTCAGAGA |
| TET1 iso promoter F | GCAAAGAGGTGTGGTTCCTG |
| TET1 iso promoter R | GCAGGGAGGTTATGTGAGGA |
| KIAA0066 F | CTAGGAGGGTGGAGGTAGGG |
| KIAA0066 R | GCCCCAAACAGGAGTAATGA |

**Table 2. Primers used for qRT-PCR**

| Name | Sequence |
|---|---|
| GAPDH F | TGCACCACCAACTGCTTAGC |
| GAPDH R | GGCATGGACTGTGGTCATGAG |
| FOXA1 F | GAAGATGGAAGGGCATGAAA |
| FOXA1 R | GCCTGAGTTCATGTTGCTGA |
| TET1 exon 1 and 2 F (TET1 FL F) | CTGCCCTGGGAATGTGAC |
| TET1 exon 1 and 2 R (TET1 FL R) | CAGAGTCATTGGTCCTTTGG |
| TET1 exon 11 and 12 F | CCGAATCAAGCGGAAGAATA |
| TET1 exon 11 and 12 R | ACTTCAGGTTGCACGGTCTC |
| TET1 Iso F | GCAAGCAAGATGGCTACCTC |
| TET1 Iso R | TTTGGGCTTCTTTTCCCTCT |
| AR F | CAGTGGATGGGCTGAAAAAT |
| AR R | GGAGCTTGGTGAGCTGGTAG |
| PSA F | ACGCTGGACAGGGGGCAAAAG |
| PSA R | GGGCAGGGCACATGGTTCACT |

# CHAPTER 5: REFERENCES

1. Koonin EV, Wolf YI. Is evolution Darwinian or/and Lamarckian? *Biol Direct.* 2009;4:42.

2. Natt D, Lindqvist N, Stranneheim H, Lundeberg J, Torjesen PA, Jensen P. Inheritance of Acquired Behaviour Adaptations and Brain Gene Expression in Chickens. *Plos One.* 2009;4(7).

3. Weinhold B. Epigenetics: the science of change. *Environ Health Perspect.* 2006;114(3):A160-167.

4. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet.* 2009;10(5):295-304.

5. Ptak C, Petronis A. Epigenetics and complex disease: from etiology to new therapeutics. *Annu Rev Pharmacol Toxicol.* 2008;48:257-276.

6. Schumacher A, Kapranov P, Kaminsky Z, et al. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* 2006;34(2):528-542.

7. Gronbaek K, Hother C, Jones PA. Epigenetic changes in cancer. *APMIS.* 2007;115(10):1039-1059.

8. Barros SP, Offenbacher S. Epigenetics: connecting environment and genotype to phenotype and disease. *J Dent Res.* 2009;88(5):400-408.

9. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002;16(1):6-21.

10. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science.* 2009;324(5929):930-935.

11. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet.* 2011;13(1):7-13.

12. Globisch D, Munzel M, Muller M, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *Plos One.* 2010;5(12):e15367.

13. Pfaffeneder T, Hackner B, Truss M, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl.* 2011;50(31):7008-7012.

14. Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science.* 2011;333(6047):1300-1303.

15. Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol.* 2013;14(6):341-356.

16. Hu L, Lu J, Cheng J, et al. Structural insight into substrate preference for TET-mediated oxidation. *Nature.* 2015;527(7576):118-122.

17. Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle.* 2009;8(11):1698-1710.

18. Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science.* 2007;317(5845):1760-1764.

19. Hashimoto H, Liu Y, Upadhyay AK, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 2012;40(11):4841-4849.

20. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* 2007;67(3):946-950.

21. Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem.* 2011;286(41):35334-35338.

22. He YF, Li BZ, Li Z, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science.* 2011;333(6047):1303-1307.

23. Zhang L, Lu X, Lu J, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol.* 2012;8(4):328-330.

24. Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A.* 1992;89(5):1827-1831.

25. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet.* 2005;37(8):853-862.

26. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13(7):484-492.

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

28. Pastor WA, Pape UJ, Huang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 2011;473(7347):394-397.

29. Matarese F, Carrillo-de Santa Pau E, Stunnenberg HG. 5-Hydroxymethylcytosine: a new kid on the epigenetic block? *Mol Syst Biol.* 2011;7:562.

30. Song CX, Szulwach KE, Fu Y, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol.* 2011;29(1):68-72.

31. Szulwach KE, Li X, Li Y, et al. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 2011;7(6):e1002154.

32. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* 2011;12(6):R54.

33. Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* 2011;25(23):2436-2452.

34. Wu H, Zhang Y. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. *Cell Cycle.* 2011;10(15):2428-2436.

35. Hahn MA, Qiu R, Wu X, et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep.* 2013;3(2):291-300.

36. Szulwach KE, Li X, Li Y, et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci.* 2011;14(12):1607-1616.

37. Hackett JA, Sengupta R, Zylicz JJ, et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science.* 2013;339(6118):448-452.

38. Thomson JP, Lempiainen H, Hackett JA, et al. Non-genotoxic carcinogen exposure induces defined changes in the 5-hydroxymethylome. *Genome Biol.* 2012;13(10):R93.

39.     Lian CG, Xu Y, Ceol C, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell.* 2012;150(6):1135-1146.

40.     Stadler MB, Murr R, Burger L, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature.* 2011;480(7378):490-495.

41.     Yu M, Hon GC, Szulwach KE, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell.* 2012;149(6):1368-1380.

42.     Serandour AA, Avner S, Oger F, et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res.* 2012;40(17):8255-8265.

43.     Whitelaw NC, Whitelaw E. Transgenerational epigenetic inheritance in health and disease. *Curr Opin Genet Dev.* 2008;18(3):273-279.

44.     Schulz WA, Hoffmann MJ. Epigenetic mechanisms in the biology of prostate cancer. *Semin Cancer Biol.* 2009;19(3):172-180.

45.     Dworkin AM, Huang TH, Toland AE. Epigenetic alterations in the breast: Implications for breast cancer detection, prognosis and treatment. *Semin Cancer Biol.* 2009;19(3):165-171.

46.     Ideraabdullah FY, Vigneau S, Bartolomei MS. Genomic imprinting mechanisms in mammals. *Mutat Res.* 2008;647(1-2):77-85.

47.     Oh G, Petronis A. Environmental studies of schizophrenia through the prism of epigenetics. *Schizophr Bull.* 2008;34(6):1122-1129.

48.     Petronis A. Epigenetics and bipolar disorder: new opportunities and challenges. *Am J Med Genet C Semin Med Genet.* 2003;123C(1):65-75.

49.     Mill J, Petronis A. Molecular studies of major depressive disorder: the epigenetic perspective. *Mol Psychiatry.* 2007;12(9):799-814.

50.     Mill J, Tang T, Kaminsky Z, et al. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet.* 2008;82(3):696-711.

51.     Gray JA, Roth BL. Molecular targets for treating cognitive dysfunction in schizophrenia. *Schizophr Bull.* 2007;33(5):1100-1119.

52.     Nelson WG, De Marzo AM, Yegnasubramanian S. Epigenetic alterations in human prostate cancers. *Endocrinology.* 2009;150(9):3991-4002.

53.     Yegnasubramanian S, Kowalski J, Gonzalgo ML, et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. *Cancer Res.* 2004;64(6):1975-1986.

54.     Li LC, Okino ST, Dahiya R. DNA methylation in prostate cancer. *Biochim Biophys Acta.* 2004;1704(2):87-102.

55.     Enokida H, Shiina H, Urakami S, et al. Multigene methylation analysis for detection and staging of prostate cancer. *Clin Cancer Res.* 2005;11(18):6582-6588.

56.     Huang Y, Rao A. Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet.* 2014;30(10):464-474.

57.     Ko M, An J, Pastor WA, Koralov SB, Rajewsky K, Rao A. TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol Rev.* 2015;263(1):6-21.

58.     Coutinho DF, Monte-Mor BC, Vianna DT, et al. TET2 expression level and 5-hydroxymethylcytosine are decreased in refractory cytopenia of childhood. *Leuk Res.* 2015;39(10):1103-1108.

59.     Haffner MC, Chaux A, Meeker AK, et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget.* 2011;2(8):627-637.

60.     Murata A, Baba Y, Ishimoto T, et al. TET family proteins and 5-hydroxymethylcytosine in esophageal squamous cell carcinoma. *Oncotarget.* 2015;6(27):23372-23382.

61.     Udali S, Guarini P, Moruzzi S, et al. Global DNA methylation and hydroxymethylation differ in hepatocellular carcinoma and cholangiocarcinoma and relate to survival rate. *Hepatology.* 2015;62(2):496-504.

62.     Huang H, Jiang X, Li Z, et al. TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc Natl Acad Sci U S A.* 2013;110(29):11994-11999.

63.     Navarro A, Yin P, Ono M, et al. 5-Hydroxymethylcytosine promotes proliferation of human uterine leiomyoma: a biological link to a new epigenetic modification in benign tumors. *J Clin Endocrinol Metab.* 2014;99(11):E2437-2445.

64.     Dehan P, Kustermans G, Guenin S, Horion J, Boniver J, Delvenne P. DNA methylation and cancer diagnosis: new methods and applications. *Expert Rev Mol Diagn.* 2009;9(7):651-657.

65.     Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature.* 2004;429(6990):457-463.

66.     Rodenhiser D, Mann M. Epigenetics and human disease: translating basic biology into clinical applications. *CMAJ.* 2006;174(3):341-348.

67.     Sonpavde G, Aparicio AM, Zhan F, et al. Azacitidine favorably modulates PSA kinetics correlating with plasma DNA LINE-1 hypomethylation in men with chemonaive castration-resistant prostate cancer. *Urol Oncol.* 2011;29(6):682-689.

68.     Singal R, Ramachandran K, Gordian E, Quintero C, Zhao W, Reis IM. Phase I/II study of azacitidine, docetaxel, and prednisone in patients with metastatic castration-resistant prostate cancer previously treated with docetaxel-based therapy. *Clin Genitourin Cancer.* 2015;13(1):22-31.

69.     Azad N, Zahnow CA, Rudin CM, Baylin SB. The future of epigenetic therapy in solid tumours--lessons from the past. *Nat Rev Clin Oncol.* 2013;10(5):256-266.

70.     Knezetic JA, Luse DS. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell.* 1986;45(1):95-104.

71.     Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell.* 2007;128(4):707-719.

72.     Kimura H, Cook PR. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol.* 2001;153(7):1341-1353.

73.     Workman JL. Nucleosome displacement in transcription. *Genes Dev.* 2006;20(15):2009-2017.

74. Owen-Hughes T, Utley RT, Cote J, Peterson CL, Workman JL. Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex. *Science.* 1996;273(5274):513-516.

75. Bruno M, Flaus A, Stockdale C, Rencurel C, Ferreira H, Owen-Hughes T. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Mol Cell.* 2003;12(6):1599-1606.

76. Phelan ML, Schnitzler GR, Kingston RE. Octamer transfer and creation of stably remodeled nucleosomes by human SWI-SNF and its isolated ATPases. *Mol Cell Biol.* 2000;20(17):6380-6389.

77. Lorch Y, Zhang M, Kornberg RD. RSC unravels the nucleosome. *Mol Cell.* 2001;7(1):89-95.

78. Lorch Y, Maier-Davis B, Kornberg RD. Chromatin remodeling by nucleosome disassembly in vitro. *Proc Natl Acad Sci U S A.* 2006;103(9):3090-3093.

79. Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M, Studitsky VM. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell.* 2002;9(3):541-552.

80. Whitehouse I, Flaus A, Cairns BR, White MF, Workman JL, Owen-Hughes T. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature.* 1999;400(6746):784-787.

81. Lorch Y, Zhang M, Kornberg RD. Histone octamer transfer by a chromatin-remodeling complex. *Cell.* 1999;96(3):389-392.

82. Adkins MW, Howar SR, Tyler JK. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. *Mol Cell.* 2004;14(5):657-666.

83. Schwabish MA, Struhl K. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol Cell.* 2006;22(3):415-422.

84. Walter PP, Owen-Hughes TA, Cote J, Workman JL. Stimulation of transcription factor binding and histone displacement by nucleosome assembly protein 1 and nucleoplasmin requires disruption of the histone octamer. *Mol Cell Biol.* 1995;15(11):6178-6187.

85.    Swaminathan V, Kishore AH, Febitha KK, Kundu TK. Human histone chaperone nucleophosmin enhances acetylation-dependent chromatin transcription. *Mol Cell Biol.* 2005;25(17):7534-7545.

86.    Kamakaka RT, Biggins S. Histone variants: deviants? *Genes Dev.* 2005;19(3):295-310.

87.    Zhang H, Roberts DN, Cairns BR. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell.* 2005;123(2):219-231.

88.    Raisner RM, Hartley PD, Meneghini MD, et al. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell.* 2005;123(2):233-248.

89.    Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science.* 2004;303(5656):343-348.

90.    Park YJ, Chodaparambil JV, Bao Y, McBryant SJ, Luger K. Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. *J Biol Chem.* 2005;280(3):1817-1825.

91.    Henikoff S. Labile H3.3+H2A.Z nucleosomes mark 'nucleosome-free regions'. *Nat Genet.* 2009;41(8):865-866.

92.    Yuen BT, Knoepfler PS. Histone H3.3 mutations: a variant path to cancer. *Cancer Cell.* 2013;24(5):567-574.

93.    Pehrson JR, Fried VA. MacroH2A, a core histone containing a large nonhistone region. *Science.* 1992;257(5075):1398-1400.

94.    Doyen CM, An W, Angelov D, et al. Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol Cell Biol.* 2006;26(3):1156-1164.

95.    Berger SL. The complex language of chromatin regulation during transcription. *Nature.* 2007;447(7143):407-412.

96.    Kouzarides T. Chromatin modifications and their function. *Cell.* 2007;128(4):693-705.

97.    Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485(7398):376-380.

98.    Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 2012;485(7398):381-385.

99. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature.* 2012;489(7414):109-113.

100. Jin F, Li Y, Dixon JR, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature.* 2013;503(7475):290-294.

101. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev.* 2005;15(2):136-144.

102. Gondor A. Dynamic chromatin loops bridge health and disease in the nuclear landscape. *Semin Cancer Biol.* 2013;23(2):90-98.

103. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst.* 1960;25:85-109.

104. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature.* 1973;243(5405):290-293.

105. Lugo TG, Pendergast AM, Muller AJ, Witte ON. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science.* 1990;247(4946):1079-1082.

106. Tomlins SA, Rhodes DR, Perner S, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science.* 2005;310(5748):644-648.

107. Demichelis F, Fall K, Perner S, et al. TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. *Oncogene.* 2007;26(31):4596-4599.

108. Clark JP, Cooper CS. ETS gene fusions in prostate cancer. *Nat Rev Urol.* 2009;6(8):429-439.

109. Tomlins SA, Laxman B, Dhanasekaran SM, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature.* 2007;448(7153):595-599.

110. Hermans KG, Bressers AA, van der Korput HA, Dits NF, Jenster G, Trapman J. Two unique novel prostate-specific and androgen-regulated fusion partners of ETV4 in prostate cancer. *Cancer Res.* 2008;68(9):3094-3098.

111. Pflueger D, Rickman DS, Sboner A, et al. N-myc downstream regulated gene 1 (NDRG1) is fused to ERG in prostate cancer. *Neoplasia.* 2009;11(8):804-811.

112. Palanisamy N, Ateeq B, Kalyana-Sundaram S, et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med.* 2010;16(7):793-798.

113. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature.* 2011;470(7333):214-220.

114. Yu J, Cao Q, Mehra R, et al. Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. *Cancer Cell.* 2007;12(5):419-431.

115. Yu J, Cao Q, Yu J, et al. The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer. *Oncogene.* 2010;29(39):5370-5380.

116. Chen H, Tu SW, Hsieh JT. Down-regulation of human DAB2IP gene expression mediated by polycomb Ezh2 complex and histone deacetylase in prostate cancer. *J Biol Chem.* 2005;280(23):22437-22444.

117. Min J, Zaslavsky A, Fedele G, et al. An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor-kappaB. *Nat Med.* 2010;16(3):286-294.

118. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science.* 2001;292(5514):110-113.

119. Metzger E, Wissmann M, Yin N, et al. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature.* 2005;437(7057):436-439.

120. Wang Q, Li W, Zhang Y, et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell.* 2009;138(2):245-256.

121. He HH, Meyer CA, Shin H, et al. Nucleosome dynamics define transcriptional enhancers. *Nat Genet.* 2010;42(4):343-347.

122. Dryhurst D, McMullen B, Fazli L, Rennie PS, Ausio J. Histone H2A.Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated transcription and is upregulated in a model of prostate cancer progression. *Cancer Lett.* 2012;315(1):38-47.

123. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell.* 2011;144(3):327-339.

124. Yu J, Yu J, Mani RS, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell.* 2010;17(5):443-454.

125. Wang Q, Carroll JS, Brown M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol Cell.* 2005;19(5):631-642.

126. Wang Q, Li W, Liu XS, et al. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell.* 2007;27(3):380-392.

127. Chen Z, Zhang C, Wu D, et al. Phospho-MED1-enhanced UBE2C locus looping drives castration-resistant prostate cancer growth. *EMBO J.* 2011;30(12):2405-2419.

128. Carlsson P, Mahlapuu M. Forkhead transcription factors: key players in development and metabolism. *Developmental biology.* 2002;250(1):1-23.

129. Benayoun BA, Caburet S, Veitia RA. Forkhead transcription factors: key players in health and disease. *Trends in genetics : TIG.* 2011;27(6):224-232.

130. Weigel D, Jurgens G, Kuttner F, Seifert E, Jackle H. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell.* 1989;57(4):645-658.

131. Augello MA, Hickey TE, Knudsen KE. FOXA1: master of steroid receptor function in cancer. *EMBO J.* 2011;30(19):3885-3894.

132. Carroll JS, Liu XS, Brodsky AS, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell.* 2005;122(1):33-43.

133. Gao N, Zhang J, Rao MA, et al. The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol Endocrinol.* 2003;17(8):1484-1507.

134. Sahu B, Laakso M, Pihlajamaa P, et al. FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. *Cancer Res.* 2013;73(5):1570-1580.

135. Costa RH, Grayson DR, Darnell JE, Jr. Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes. *Molecular and cellular biology.* 1989;9(4):1415-1425.

136. Besnard V, Wert SE, Hull WM, Whitsett JA. Immunohistochemical localization of Foxa1 and Foxa2 in mouse embryos and adult tissues. *Gene expression patterns : GEP.* 2004;5(2):193-208.

137. Monaghan AP, Kaestner KH, Grau E, Schutz G. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development.* 1993;119(3):567-578.

138. Ang SL, Wierda A, Wong D, et al. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development.* 1993;119(4):1301-1315.

139. Sasaki H, Hogan BL. Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development.* 1993;118(1):47-59.

140. Kaestner KH, Katz J, Liu Y, Drucker DJ, Schutz G. Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. *Genes Dev.* 1999;13(4):495-504.

141. Shih DQ, Navas MA, Kuwajima S, Duncan SA, Stoffel M. Impaired glucose homeostasis and neonatal mortality in hepatocyte nuclear factor 3alpha-deficient mice. *Proc Natl Acad Sci U S A.* 1999;96(18):10152-10157.

142. Gao N, LeLay J, Vatamaniuk MZ, Rieck S, Friedman JR, Kaestner KH. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* 2008;22(24):3435-3448.

143. Besnard V, Wert SE, Kaestner KH, Whitsett JA. Stage-specific regulation of respiratory epithelial cell differentiation by Foxa1. *Am J Physiol Lung Cell Mol Physiol.* 2005;289(5):L750-759.

144. Wan H, Dingle S, Xu Y, et al. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem.* 2005;280(14):13809-13816.

145.    Lee CS, Friedman JR, Fulmer JT, Kaestner KH. The initiation of liver development is dependent on Foxa transcription factors. *Nature.* 2005;435(7044):944-947.

146.    Domanskyi A, Alter H, Vogt MA, Gass P, Vinnikov IA. Transcription factors Foxa1 and Foxa2 are required for adult dopamine neurons maintenance. *Front Cell Neurosci.* 2014;8:275.

147.    Bernardo GM, Lozada KL, Miedler JD, et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development.* 2010;137(12):2045-2054.

148.    Gao N, Ishii K, Mirosevich J, et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development.* 2005;132(15):3431-3443.

149.    Mirosevich J, Gao N, Matusik RJ. Expression of Foxa transcription factors in the developing and adult murine prostate. *Prostate.* 2005;62(4):339-352.

150.    Marker PC, Donjacour AA, Dahiya R, Cunha GR. Hormonal, cellular, and molecular control of prostatic development. *Dev Biol.* 2003;253(2):165-174.

151.    Neben K, Schnittger S, Brors B, et al. Distinct gene expression patterns associated with FLT3- and NRAS-activating mutations in acute myeloid leukemia with normal karyotype. *Oncogene.* 2005;24(9):1580-1588.

152.    Lin L, Miller CT, Contreras JI, et al. The hepatocyte nuclear factor 3 alpha gene, HNF3alpha (FOXA1), on chromosome band 14q13 is amplified and overexpressed in esophageal and lung adenocarcinomas. *Cancer Res.* 2002;62(18):5273-5279.

153.    Sano M, Aoyagi K, Takahashi H, et al. Forkhead box A1 transcriptional pathway in KRT7-expressing esophageal squamous cell carcinomas with extensive lymph node metastasis. *Int J Oncol.* 2010;36(2):321-330.

154.    Nucera C, Eeckhoute J, Finn S, et al. FOXA1 is a potential oncogene in anaplastic thyroid carcinoma. *Clin Cancer Res.* 2009;15(11):3680-3689.

155.    Hu X, Stern HM, Ge L, et al. Genetic alterations and oncogenic pathways associated with breast cancer subtypes. *Mol Cancer Res.* 2009;7(4):511-522.

156.    Robbins CM, Tembe WA, Baker A, et al. Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors. *Genome Res.* 2011;21(1):47-55.

157. Sahu B, Laakso M, Ovaska K, et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J.* 2011;30(19):3962-3976.

158. Gerhardt J, Montani M, Wild P, et al. FOXA1 promotes tumor progression in prostate cancer and represents a novel hallmark of castration-resistant prostate cancer. *Am J Pathol.* 2012;180(2):848-861.

159. Jin HJ, Zhao JC, Ogden I, Bergan RC, Yu J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res.* 2013;73(12):3725-3736.

160. Wang D, Garcia-Bassets I, Benner C, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature.* 2011;474(7351):390-394.

161. Robinson JL, Holmes KA, Carroll JS. FOXA1 mutations in hormone-dependent cancers. *Front Oncol.* 2013;3:20.

162. Grasso CS, Wu YM, Robinson DR, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature.* 2012;487(7406):239-243.

163. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet.* 2012;44(6):685-689.

164. Gao D, Vela I, Sboner A, et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell.* 2014;159(1):176-187.

165. Hu DG, Gardner-Stephen D, Severi G, et al. A novel polymorphism in a forkhead box A1 (FOXA1) binding site of the human UDP glucuronosyltransferase 2B17 gene modulates promoter activity and is associated with altered levels of circulating androstane-3alpha,17beta-diol glucuronide. *Mol Pharmacol.* 2010;78(4):714-722.

166. Hazelett DJ, Coetzee SG, Coetzee GA. A rare variant, which destroys a FoxA1 site at 8q24, is associated with prostate cancer risk. *Cell Cycle.* 2013;12(2):379-380.

167. Xu Y, Chen SY, Ross KN, Balk SP. Androgens induce prostate cancer cell proliferation through mammalian target of rapamycin activation and post-transcriptional increases in cyclin D proteins. *Cancer Res.* 2006;66(15):7783-7792.

168. Jin HJ, Zhao JC, Wu L, Kim J, Yu J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat Commun.* 2014;5:3972.

169. DeGraff DJ, Grabowska MM, Case TC, et al. FOXA1 deletion in luminal epithelium causes prostatic hyperplasia and alteration of differentiated phenotype. *Lab Invest.* 2014;94(7):726-739.

170. Wang H, Meyer CA, Fei T, Wang G, Zhang F, Liu XS. A systematic approach identifies FOXA1 as a key factor in the loss of epithelial traits during the epithelial-to-mesenchymal transition in lung cancer. *BMC Genomics.* 2013;14:680.

171. Song Y, Washington MK, Crawford HC. Loss of FOXA1/2 is essential for the epithelial-to-mesenchymal transition in pancreatic cancer. *Cancer Res.* 2010;70(5):2115-2125.

172. Jain RK, Mehta RJ, Nakshatri H, Idrees MT, Badve SS. High-level expression of forkhead-box protein A1 in metastatic prostate cancer. *Histopathology.* 2011;58(5):766-772.

173. Hannenhalli S, Kaestner KH. The evolution of Fox genes and their role in development and disease. *Nat Rev Genet.* 2009;10(4):233-240.

174. Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature.* 1993;364(6436):412-420.

175. Cirillo LA, McPherson CE, Bossard P, et al. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* 1998;17(1):244-254.

176. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell.* 2002;9(2):279-289.

177. Lupien M, Eeckhoute J, Meyer CA, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell.* 2008;132(6):958-970.

178. Lupien M, Brown M. Cistromics of hormone-dependent cancer. *Endocr Relat Cancer.* 2009;16(2):381-389.

179. Serandour AA, Avner S, Percevault F, et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* 2011;21(4):555-565.

180. Jozwik KM, Carroll JS. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer.* 2012;12(6):381-385.

181. Wang J, Zhuang J, Iyer S, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 2012;22(9):1798-1812.

182. Gifford CA, Meissner A. Epigenetic obstacles encountered by transcription factors: reprogramming against all odds. *Curr Opin Genet Dev.* 2012;22(5):409-415.

183. Maurano MT, Wang H, John S, et al. Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep.* 2015;12(7):1184-1195.

184. Feldmann A, Ivanek R, Murr R, Gaidatzis D, Burger L, Schubeler D. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* 2013;9(12):e1003994.

185. Hon GC, Song CX, Du T, et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell.* 2014;56(2):286-297.

186. Song CX, Clark TA, Lu XY, et al. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods.* 2011;9(1):75-77.

187. Fong KW, Leung JW, Li Y, et al. MTR120/KIAA1383, a novel microtubule-associated protein, promotes microtubule stability and ensures cytokinesis. *J Cell Sci.* 2013;126(Pt 3):825-837.

188. Xu Y, Xu C, Kato A, et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell.* 2012;151(6):1200-1213.

189. Jin C, Lu Y, Jelinek J, et al. TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* 2014;42(11):6956-6971.

190. Wang Y, Zhang Y. Regulation of TET protein stability by calpains. *Cell Rep.* 2014;6(2):278-284.

191. Ko M, An J, Bandukwala HS, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature.* 2013;497(7447):122-126.

192. Iyer LM, Abhiman S, Aravind L. Natural history of eukaryotic DNA methylation systems. *Prog Mol Biol Transl Sci.* 2011;101:25-104.

193. Zhang W, Xia W, Wang Q, et al. Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Mol Cell.* 2016;64(6):1062-1073.

194.    Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet.* 2015;47(3):199-208.