KMT2C DIRECTS ESTROGEN RECEPTOR ALPHA ACTIVITY IN NORMAL AND TRANSFORMED MAMMARY CELLS

by

Kinisha Pankaj Gala

A dissertation presented to the faculty of the

Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences

Memorial Sloan Kettering Cancer Center

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

New York, NY

June, 2017

_____ _____

Sarat Chandarlapaty, M.D./Ph.D. Date
Dissertation Mentor

DEDICATION

This thesis is dedicated to my parents, Asha and Pankaj Gala, my brother, Bhavin Gala, and my husband, Neil Mendhiratta. My stars, my moon and my sun. You are the great loves of my life. Thank you, thank you, thank you.

ABSTRACT

Estrogen receptor alpha (ERα) is a ligand-activated nuclear receptor that promotes proliferation and differentiation in mammary derived-epithelial cells and other estrogen-regulated cell types. These functions of ERα largely require its genomic transcriptional activity and, as such are often dependent on trans-acting factors such as pioneer factors and chromatin modifiers to establish a genomic landscape conducive to ERα transcriptional activity. Here, we identify the H3K4 methyltransferase KMT2C as necessary for hormone-driven ERα activity in ER+ breast cancer proliferation and pubertal mammary gland development KMT2C knockdown suppresses estrogen-dependent gene expression in multiple ER+ models and causes H3K4me1 and H3K27ac loss at ERα enhancers. Correspondingly, KMT2C loss impairs estrogen-driven breast cell proliferation and mammary ductal branching. While loss of KMT2C is not seen to dramatically affect global ERα binding to the DNA, it does change the ERα cofactor binding profile suggesting that the role of KMT2C is ERα cofactor recruitment. Downstream ERα signaling is known to be a key driver of proliferation in ER+ breast cancer. The most widely used treatment against ER+ breast cancer are drugs that reduce the amounts of estrogen in the body. When mimicking these hormone-deprivation therapies we found that KMT2C loss promotes outgrowth under hormone-depleted conditions. In accordance, *KMT2C* is one of the most frequently mutated genes in breast cancer with *KMT2C* mutation frequency increasing specifically in hormone resistant tumors. In addition, *KMT2C* mutation and gene expression signatures of KMT2C loss are associated with poorer

clinical outcomes in Luminal A/B breast cancer. From a therapeutic standpoint, KMT2C depleted cells that develop hormone independence retain their dependence on ERα, displaying ongoing sensitivity to direct ERα antagonists. Moreover, we find that ERα has a completely reprogrammed binding profile in these resistant cells, that may be partially dependent on AP-1. We conclude that KMT2C is a key regulator of ERα enhancer activity whose loss uncouples breast cell proliferation from hormone sensitivity.

BIOGRAPHICAL SKETCH

Kinisha Gala was born to Pankaj and Asha Gala on November 15th, 1989 in Yardley, Pennsylvania. Her father, Pankaj, would eventually give her his love for science, his love for Star Wars and his love for being outside. Unfortunately, Pankaj was unable to pass on his unparalleled patience and forever calm demeanor, but Kinisha is still a work in progress. From her mother, Asha, Kinisha inherited her love for reading, her love for dancing, and her need for everything to be perfect. Asha showered Kinisha's childhood with lots of hugs, kisses and the best birthday gifts ever. At seven months of age, Kinisha and her parents moved to Branchburg, New Jersey where they welcomed her brother, Bhavin Gala. Bhavin would go on to become Kinisha's best friend, travel buddy, and the kind of person that sees two circles when Kinisha can only see one. When Kinisha was eight years old, her family moved to Puerto Rico where Kinisha discovered the wonders of coquis, bioluminescent bays and snorkeling. After four happy years, she eventually moved back to New Jersey. Kinisha graduated salutatorian from Voorhees High School in 2008 and then began college at The College of New Jersey (TCNJ). She graduated from TCNJ in 2012 with a Bachelor of Science in Biology. While at TCNJ, Kinisha was inducted into the Phi Beta Kappa Society and competed her undergraduate honors thesis under the mentorship of Dr. Tracy Kress, who showed Kinisha that scientists could be happy, fun and unbelievably kind. Also, while in college, Kinisha met Neil Mendhiratta, the boy with whom she would go on to share innumerable best days ever. She soon got to know Neil's parents, Arjun and Anurita Mendhiratta, whose generosity and

grace truly put everyone else to shame, and Neil's brother, Vikas Mendhiratta, whose vast kindness is evidenced by his willingness to laugh at all of Kinisha's horrible jokes. In July 2012, Kinisha fulfilled her lifelong dream of living in New York City. She began her PhD training at the Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences at Memorial Sloan Kettering Cancer Center (GSK) after being very impressed by the caliber of research conducted by the students. Five years at GSK has only strengthened this initial impression. While at GSK, she joined the laboratory of Dr. Sarat Chandarlapaty in the Human Oncology and Pathogenesis Program, a decision that she has never and will never regret. She spent many painstaking hours working on KMT2C, and while her love for the protein diminished rapidly over time, her love for the lab mates, with which she spent those many hours, only deepened. In May 2017, Kinisha married Neil, so grateful to be able to call this wonderful man her husband forever.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my mentor, Dr. Sarat Chandarlapaty, for his generosity, his patience, his support, and his wisdom. Sarat, having you as my mentor was the single best decision I made during my time in graduate school. I would like to sincerely thank Dr. Weiyi Toy, Dr. Zhiqiang Li, Trusha Bhatt and Alison Smith their friendship, encouragement and readiness to help. I hope you all know that you are a major reason I stayed sane and happy over the past couple years. I would also like to thank Dr. Cheng Cheng Yang, Sharmeen Rahman, Mayur Gadiya, Dr. Bo Liu, Dr. Antonio Strillacci, Jacqueline Giovanniello, and Dr. Marie Will for making our workplace an open, welcoming and exciting place. I would like to sincerely thank my committee members, Drs. James Fagin and Ping Chi for all their support and sound advice over the course of my Ph.D. And thank you to Dr. Steffi Oesterreich and Dr. David Solit for agreeing to participate on my examining committee. I would also like to thank our many wonderful collaborators. Thank you to Dr. Amit Sinha for analyzing the numerous sequencing experiments conducted over the course of the project. Thank you to Young Rock Chung for his patience as I tepidly learned how to take care of mouse colonies. Thank you to Drs. Francisco Sanchez-Vega, Alessandro Pastore, and Nikolous Schultz for their dedication to helping us better understand the effects of *KMT2C* mutation across patient populations. Thank you to Dr. Ronald Hendrickson for analyzing our proteomics studies. Thank you to Drs. Omar Abdul-Wahab, Ping Chi, Neal Rosen, Charles Sawyers, Weiyi Toy and to Alison Smith for all their help in editing our manuscript. A very big thank you to

the GSK office for all your hard work and commitment to the student body. To our dean, Dr. Ken Marians, the GSK students are unbelievably lucky to have your vision, generosity and dedication. Thank you. To our benefactor, Louis V. Gerstner, thank you for making this education possible. I, of course, need to thank my incredible friends and family, made of the strongest, smartest and most loving superheroes in the world. They lift me up and give me purpose. Finally, I would be amiss to leave out all the teachers and mentors that have taught me and trained me over the past 27 years. Thank you for all your patience, dedication and care. You made a world of difference.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AF	Activating function
AR	Androgen Receptor
ASCOM	ASC-2 containing complex
bp	basepairs
E2	Estradiol
ER+	Estrogen receptor positive
ERE	Estrogen response element
ERa	Estrogen Reception alpha
FXR	Farnesoid X receptor
GR	Glucocorticoid receptor
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
H3K27	Histone H3 Lysine 27
H3K27ac	Histone H3 Lysine 27 acetylation
H3K4	Histone H3 Lysine 4
H3K4me2	Histone H3 Lysine 4 dimethylation
H3K4me1	Histone H3 Lysine 4 monomethylation
H3K4me3	Histone H3 Lysine 4 trimethylation
HMG	High mobility group
HMT	Histone methyltransferase
K	Lysine
LTED	Long-term estrogen deprivation

LXR	Liver X receptor
MMTV	Mouse mammary tumor virus
NCoR	nuclear receptor corepressor
PDAC	Pancreatic ductal adenocarcinoma
PHD	Plant homology domain
PolII	RNA Polymerase II
PPARy	Peroxisome proliferator-activated receptor gamma
PR	Progesterone receptor
PRC2	Polycomb repressive complex 2
qRTPCR	Quantitative Reverse Transcriptase Polyermase Chain Reaction
R	Arginine
RIME	Rapid immunoprecipitation and mass spectrometry of endogenous protein complexes
SERD	Selective ERa Degrader
SERM	Selective estrogen receptor modulator
shRNA	Short Hairpin RNA
SILAC	Stable isotope labeling with amino acids in cell culture
TCGA	The Cancer Genome Atlas
Trr	Trithorax related
Trx	Trithorax
WAP	Whey Acidic Protein
WT	Wildtype

CHAPTER 1: INTRODUCTION

Introduction to epigenetic regulation

Organization of chromatinized DNA

Chromatin occurs in two basic states, tightly closed heterochromatin or open euchromatin(1). Nucleosomes are the basic units of chromatin. They consist of approximately 146 basepairs (bp) of DNA organized into two superhelical turns wrapped around an octamer of core histone proteins. The core histone proteins include an H3-H4 tetramer and two H2A-H2B dimers(2). Chromatinized DNA also includes a linker histone protein, H1, which resides in between nucleosomes, binding to the DNA entry and exit sites(3, 4).

Nucleosomes form the primary sites for protein-DNA interactions. For upstream signaling to be able to exert an effect on gene expression, it must be able to interact with only the appropriate target DNA sites. Therefore, nucleosomes associated with actively transcribed genes need to be more open while nucleosomes associated with inactive genes need to be tightly bound(5). This level of regulation is in large part achieved through modification of the chromatin structure.

The four core histone proteins, H2A, H2B, H3 and H4 are made up of two domains. The histone fold domain, which encompasses the middle to the C-terminal end of the protein, mediates histone-histone and histone-DNA interactions to maintain the nucleosome as a unit(6-8). The N-terminal end consists of a flexible tail that functions to mediate higher order interactions with neighboring nucleosomes(8-10). These N-terminal histone tails, especially the H3 and H4 tails, are subject to post-translational modifications including

methylation, acetylation, phosphorylation, ubiquitination, ADP ribosylation, deimination, biotinylation, butyrylation, N-formylation and proline isomerization (11). These modifications modulate histone tails interactions with other nucleosome and also with other chromatin-bound proteins, thereby influencing the DNA accessibility of transcription factors and regulatory proteins(12, 13).

Histone modifications and their effect on gene expression

In general, acetylation is considered to demarcate regions of active transcription as the acetyl moiety, placed exclusively on lysine (K) residues, is negatively charged and can decrease the affinity for neighboring proteins and DNA, serving to open up the surrounding chromatin(14-16). Methylation occurs on arginine (R) and K residues and, unlike acetylation, can either demarcate regions of active, poised or inactive transcription based on which residue is modified (17). The ability for residues to be mono-, di- or tri-methylated also adds further complexity to methylation-dependent regulation(18-20). While methyl residues do not influence the histone charge, these modifications are important for protein recruitment. Specifically, plant homology domains (PHDs) recognize H3K4me3, tudor domains recognize H3K4me3 and H4K20me, chromodomains recognize H3K9me and WD-40 domains recognize H3K27me3(21-26). The proteins that bind to lysine methylation marks have very diverse functions with some acting as positive regulators of transcription and others acting as negative regulators of transcription(27). In terms of other histone modifications, there is some data suggesting that phosphorylation can serve as a positive regulator of

mitogen-stimulated early response genes(28, 29) and can mediate cell cycle regulation(30, 31) and DNA-damage response(32). The effect of other histone modifications, including ubiquitination, sumoylation, and others has yet to be fully understood, but the presence of these marks suggest that chromatin regulation is a highly dynamic process. Together, the various chromatin modifications make up the “epigenetic code” (12).

Histone modification enzymes

The enzymes partially responsible for placing the epigenetic code, the histone modification enzymes, provide a dynamic layer for gene expression regulation. While there are a host of kinases, ubiquitin ligases and other modification-specific enzymes, the enzymes responsible for histone methylation and acetylation are the most widely studied. GCN5, a histone acetyltransferase (HAT), was the first histone modification enzyme to be identified(33). HATs, in general, fall into multiple families that are thought to be responsible for acetylating specific sites and generating distinct patterns of acetylation(34). The counterbalance to HAT activity is maintained within histone deacetyltransferases (HDACs), which generally negatively regulate gene expression(35). HATs and HDACs often function within larger, multiprotein complexes, which helps determine substrate specificity(36, 37).

Histone methyltransferases (HMTs) are responsible for the transfer of a methyl group from S-adenosylmethionine (SAM) on either R or L residues(38). Protein R methyltransferases (PRMTs) can either monomethylate,

asymmetrically dimethylate or symmetrically dimethylate residues depending on if they are classified as Type I, Type II or Type III PRMTs(39). Lysine-specific histone methyltransferases all have a highly conserved SET domain. The most widely studied lysine HMT is EZH2, which is specific for di- or trimethylating H3K27 residues. EZH2 is generally found in the Polycomb repressive complex 2 (PRC2), known for silencing the HOX gene cluster, and can affect a wide range of cellular processes including apoptosis, proliferation, differentiation and cell cycle regulation(40). HMTs also include the family of proteins responsible for H3K9 methylation, which is known to demarcate regions of constitutive heterochromatin and be important in promoting genome stability(41). Finally, H3K4 HMTs are positive regulators of transcription, found at both enhancer and promoter regions of the genome(20).

The KMT2 family of H3K4 histone methyl transferases

Members of the KMT2 family

The KMT2 family includes KMT2A, KMT2B, KMT2C, KMT2D, KMT2F and KMT2G. These family members are highly conserved with paralogs found in unicellular eukaryotes(42, 43). In Drosophila, the KMT2A and KMT2B homolog, Trithorax (Trx), the KMT2C and KMT2D homolog, Trithorax related (Trr) and the KMT2F and KMT2G homolog, Set1, are known H3K4 methyltransferases(44). Trr was implicated as the major Drosophila H3K4 monomethyltrasferase, thought to specifically be active at enhancer regions(45). KMT2E is another KMT2 family member related to the Set3 family of proteins; however, this protein has no

known methyltransferase activity (46) and will therefore be excluded from this introduction.

The KMT2 family has differential substrate specificity. KMT2A and KMT2B have mono- and dimethyltransferase activity but very low trimethyltransferase activity and KMT2C and KMT2D having predominantly monomethyltransferase activity(47-49). By contrast, KMT2F and KMT2G are able to mono-, di- and trimethylate H3K4 residues, similar to their yeast homolog Set1(43). The reason behind the differential substrate specificity is poorly understood and may require additional structural studies.

KMT2F and KMT2G are predominantly found at promoter regions of the genome(50) and KMT2C and KMT2D are predominantly found at enhancer regions of the genome(51, 52). KMT2A and KMT2B are found at both enhancer and promoter regions(52, 53). Therefore, their substrate specificity correlates with their genome-wide binding profile as promoter regions are abundant in trimethylated H3K4 and enhancer regions are abundant in monomethylated H3K4(17).

KMT2C family coregulatory proteins

KMT2 family coregulatory proteins are thought to be especially important for their full KMT2 enzymatic activity. Indeed, while purified KMT2C-mediated HMT activity was weak, it was soon confirmed that KMT2C, as well as other KMT2 family members, interacted with several core proteins that promoted KMT2 enzymatic activity(54, 55). These core components include ASHL2, WDR5,

RbBP5, and hDPY-30. Functional characterization of these core components has been primarily carried out in the context of KMT2A-mediated methyltransferase activity with the assumption that they play similar roles for the other KMT2 proteins. WDR5 is thought to make direct contact with histone H3(55, 56) and bridge the KMT2 protein to its substrate. WDR5 is also thought to directly stimulate KMT2A methyltransferase activity along with RbBP5(57). ASHL2 is able to stimulate KMT2A and KMT2F methyltransferase activity in response to H2B ubiquitination marks. Interestingly, KMT2C does not seem to respond to H2B ubiquitination the same way that KMT2A does, suggesting differential roles for the core proteins in mediating the activity of the different KMT2 proteins(48). The final component, hDPY-30 has yet to be studied in mammalian cells; however, its homolog in yeast is necessary for yeast Set1-mediated HMT activity(58).

Physiologic functions of KMT2 family members

In terms of the marks regulated by KMT2 family members, H3K4me3 at promoters regions has been shown to be integral for the binding of transcriptional regulators. Specifically, KMT2-family members have been shown to influence RNA polymerase II (PolII) binding to promoters(51, 52). H3K4me1 is known to demarcate both “active” and “poised” distal enhancer regions and is the most widely used histone modification to delineate lineage- and developmental-specific enhancers. Differentiating between “active” and “poised” enhancers requires analysis of H3K27ac, which is enriched in the former(17, 59). The

presence of H3K4me1 is thought to prime enhancers for the binding of other transcriptional activators, explaining its importance in coordinating lineage-specific transcription factors. In fact, it has been shown that the activity of KMT2A and KMT2D is also important for non-coding transcription at enhancers(52, 60).

Despite this family of proteins being the major regulators of H3K4 methylation, deletion of individual KMT2 genes had little effect on global gene expression in mouse embryonic fibroblasts (MEFs) suggesting some level of redundancy(60). However, deletion of *Kmt2c* or *Kmt2d* in the whole mouse results in large-scale gene expression changes during adipogenesis and during the trans-differentiation of pre-adipocytes into myocytes. Indeed, KMT2C and KMT2D were recently shown to be the major HMTs responsible for methylating adipocyte- and myocyte-specific enhancers(51). KMT2-deletion, in general, has also been shown to be important for embryonic stem cell differentiation and ligand-dependent cell stimulation(61-63).

Aside from their fundamental importance in transcriptional regulation, KMT2 family members are also important for a number of other physiological functions. For example, KMT2A binding to CpG islands was shown to protect these islands from DNA methylation(64-66). In fact, DNA methylation is thought to be largely mutually exclusive with H3K4me(67, 68). The expression of KMT2A is also tightly coupled to the cell cycle. Similar to the KMT2A homolog, Trx, KMT2A remains tightly bound to DNA during S phase and mitosis to ensure proper DNA replication. Then, KMT2A is subject to ubiquitin-mediated degradation during the S-G1 transition and the M-G1 transition(69-71). Such

regulation of KMT2 activity would ensure proper spatiotemporal regulation of H3K4me.

KMT2 family members in cancer

KMT2 mutations have been reported across multiple cancers suggesting a key role for this family of proteins in cancer progression. *KMT2A* is frequently involved in rearrangements, tandem duplications and amplifications in the context of mixed lineage leukemia (MLL) where somatic mutations in *KMT2A* are rare(49). In contrast, *KMT2A* is quite frequently mutated in solid tumors including lung, bladder, endometrial, breast and colon(72-75) and ~40% of these mutations result in truncated protein with loss of enzymatic activity. While zygosity is not known for every case of *KMT2A* mutation, for the ~60% of cases where zygosity is known, wildtype (WT) alleles of *KMT2A* remain intact(49). *KMT2B* is not as frequently mutated as *KTM2A*; however, *KMT2B* mutations also mostly consist of truncation and missense mutations that are disproportionately enriched in the SET domain. Over 90% of *KMT2B* mutations are heterozygous(49).

Modeling Kmt2a loss in mice suggests that *KTM2A* heterozygous deletion is not sufficient to induce leukemogenesis or tumirogenesis(76). However, modeling of gain-of-function Kmt2a rearrangements with fusion partners did lead to malignant MLL transformation. Surprisingly a variety of cancer cell line models are sensitive to KMT2A knockdown and cells expressing *KMT2A*-fusion genes are sensitive to loss of the WT *KTM2A* allele(77). Similarly, in vitro embryonic stem cell and in vivo knockout of *KMT2B* resulted in reduced proliferation and

embryonic lethality(78, 79). Given that both KMT2A and KMT2B are known to bind to *MYC* super enhancers and positively regulate cell cycle regulators(80), targeting WT KMT2A or KMT2B might result in cancer cell inhibition.

KMT2F and *KMT2G* also have multiple frameshift and missense mutations that span across the length of the protein. *KMT2F*/*G* missesnse mutations tend to cluster around the SET and N-SET domains with fewer missense mutations found in the N-terminal RNA binding domain. Similar to the other KMT2 family members, very few KMT2F/G mutations were found to be homozygous.(49) These mutations have yet to be modeled in vitro so we do not yet know how they directly contribute to cancer progression.

KMT2C and *KMT2D* are among the most frequently mutated genes across all cancer. As is seen for other KMT2 family members, *KMT2C*/*D* mutations tend to be loss of function mutations with a disproportionate amount of missense mutations (16% for *KMT2C* and 6% for *KMT2D*) found in the PHD domain. KMT2C has been suggested to be a tumor suppressor in multiple experimental studies for leukemia and bladder cancer(81, 82). KMT2D has been suggested to act as a coactivator of p53(83) but other studies show that KMT2D loss impairs proliferation of breast cancer xenografts in mice and in multiple cancer cell lines(84-86). *KMT2C* is most frequently mutated in lung, large intestine, breast, endometrial and bladder carcinomas(73, 87-89) but the selective advantage of its mutation is not very well understood across these multiple cancers and may be very context specific.

KMT2C

KMT2C as a member of ASCOM

KMT2C was initially discovered in 2001 and was thought to play a role in transcriptional activation due to its protein domain structure consisting of a C-terminal SET domain, multiple PHD-fingers, an high mobility group (HMG) box, several nuclear localization signals and several nuclear receptor binding motifs(90). Soon after, KMT2C was identified in a study aimed to isolate a steady-state complex for the nuclear receptor co-activator protein, ASC-2. Immunoprecipitation of the ASC-2-containing complex (ASCOM) resulted in the identification of both KTM2C and KMT2D as ASCOM members(91). In the same study, it was verified that ASCOM specifically interacted with and methylated histone H3, residue K4. ASCOM contains all the core regulatory components known to be necessary for complete HMT activity. Aside from the core components, ASCOM also has specific components that do not bind KMT2A or KMT2B. These include ASC-2, PTIP, PA1 and UTX. ASC-2 is thought to act as a coactivator for nuclear receptors(92). PTIP is responsible for thethering PA1 to the rest of the complex(54) and is also capable to interacting with the transcription factor PAX2, an important developmental regulator(93). UTX is another histone modification enzyme that demethylates the repressive H3K27me2/3 marks(94). Therefore, ASCOM, in pairing the H3K4 HMT activity of KMT2C or KMT2D with the H3K27 demethylase activity of UTX, functions to make chromatin more accessible to transcription factors and promotes transcription.

KMT2C monomethyltransferase activity

In mammalian cells and drosophila cells, it was suggested that KMT2C and KMT2D, and their homology Trr, predominantly monomethylate H3K4 residues at enhancer regions(95). As such, the dominant theory was that KMT2C/D along with the ASCOM complex facilitate H3K4me1 at enhancer regions while other SET domain containing proteins, such as Set1a/b, facilitate H3K4me3 at promoter regions(51, 52) (Fig. 1.1). It should be noted, however, that another study showed that MEFs expressing catalytically inactive KMT2C lost H3K4me3 primarily at promoter regions(96). Therefore, there is chance that the precise role of KMT2C, as well as the ASCOM complex, may vary with tissue type. It is also possible that the overall scheme of methylating H3K4 residues may involve multiple SET-domain containing proteins and demethylating proteins with KMT2C being an early enzyme capable of H3K4me1. In terms of KMT2C expression, northern blot analysis of KMT2C mRNA in different adult and fetal tissues showed KMT2C expression in placenta, peripheral blood, testes, fetal thymus, heart, brain, lung, liver and kidney tissues(97), suggesting that KMT2C is ubiquitously expressed.

Physiological functions of KMT2C

KMT2C as a nuclear receptor coactivator

The KMT2 family of methyltransferases has been repeatedly implicated as mediators of nuclear receptor activity, especially since ASC-2 and the KMT2A/B

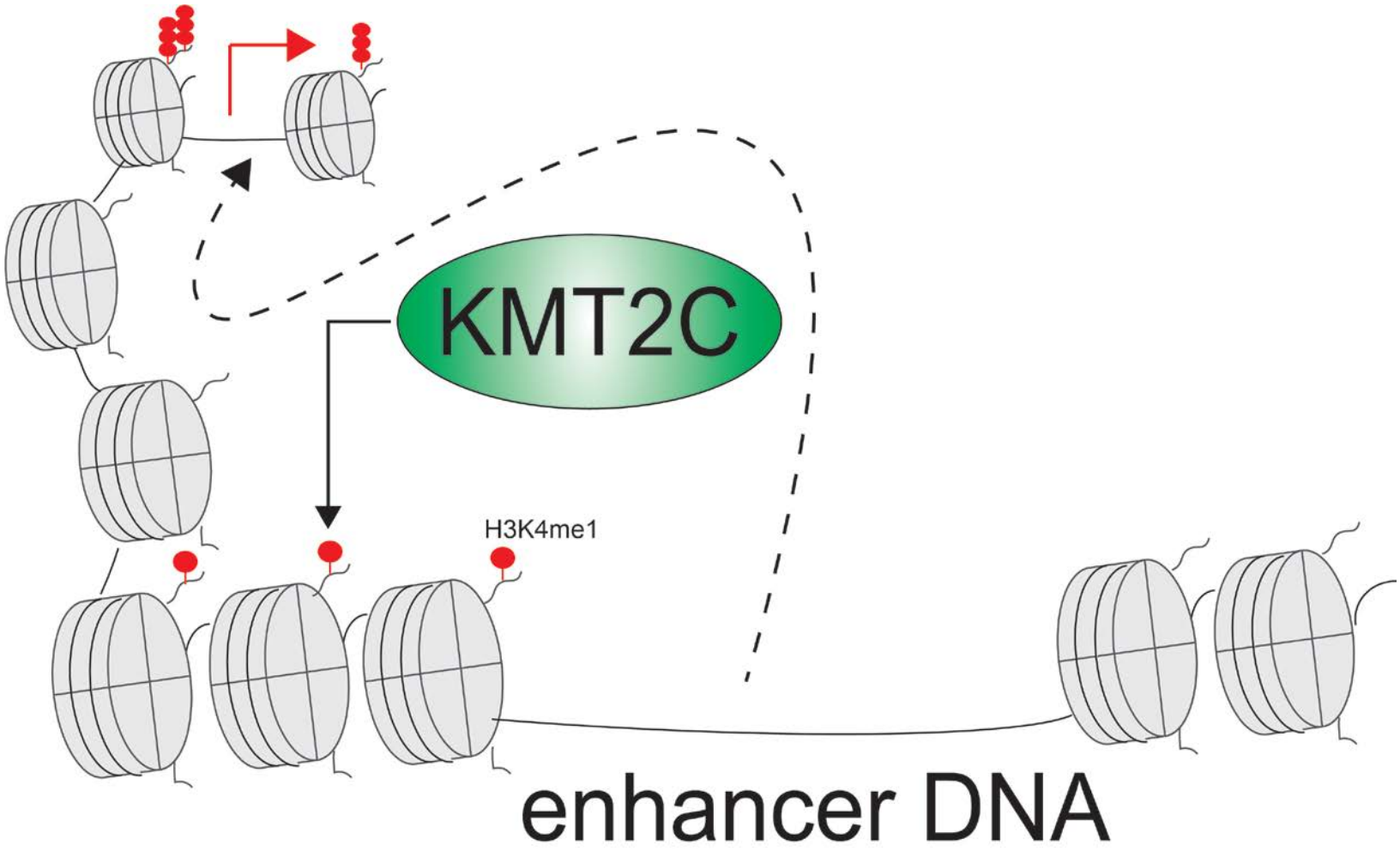

Figure 1.1 KMT2C is an H3K4 monomethyltransferase that is active predominantly at enhancer regions of the DNA.

regulator, Menin, are candidate nuclear receptor binding proteins. In Drosophila, Trr coimmunoprecipitated and genetically interacted with the ecdosyne-receptor, which is homologous to nuclear receptors found in human cells(98). In mammalian cells, chromatin immunoprecipitation experiments showed the recruitment of ASCOM to retinoic acid receptor (RAR) target genes in the presence of retinoic acid(91). Since ASC-2 was already well regarded as a nuclear receptor coactivator, these early studies sparked interest in characterizing ASCOM, as a whole, as a transcriptional coactivator for multiple nuclear receptors.

Aside from RAR, ASCOM was also shown to coactivate liver X receptors (LXRs) in mouse liver tissue and KMT2C/D methyltransferase activity was shown to be crucial for LXR-mediated gene transcription(83). Similarly, ASCOM was shown to be a coactivator for the farnesoid X receptor (FXR) and glucocorticoid receptor (GR) during the activation of bile acid transporter genes in hepatocellular carcinoma cells(99, 100) and for the progesterone receptor during activation of a stably integrated MMTV-luciferase transgene in T47D cells(4). ASCOM can also influence the activity of the peroxisome proliferator-activated receptor gamma (PPARγ**).** Through these various studies, it also became apparent that siRNA-mediated knockdown of both KMT2C and KMT2D resulted in more significant decreases in transcriptional activation than knockdown of either one alone, suggesting their redundancy in serving as nuclear receptors coactivators.

Interestingly, KMT2C has also been implicated as a coactivator of the estrogen receptor (ERα), specifically in the context of HOX gene transcriptional regulation. HOX genes are evolutionarily conserved genes that play key roles in development(101). ERα, KMT2C and KMT2B were shown to be recruited to the *HOXC6* promoter in the presence of estrogen. Knockdown of KMT2B or KMT2C resulted in reduced *HOXC6* transcription in the presence of estrogen(102). KMT2C was also seen to play a role in the estrogen-mediated transcription of *HOXB9*(103), *HOXC13*(104), *HOXC10*(84), and *HOTAIR*(105), a long noncoding RNA that is transcribed with the HOXC gene cluster. Most of these studies have been done in JAR placental choriocarcinoma cells(106), so it remains unclear if KMT2C serves as a regulator of HOX gene transcriptions in estrogen-regulated tissue such as the mammary gland. More recently, KMT2C was found to bind FOXA1, an ERα coregulator, KMT2C loss, in this study, was shown to effect a number of non-HOX ERα target genes(107). KMT2C was, however, not seen in among ERα binding partners shown to coimmunprecipitate with ERα suggesting that these two proteins do not physically interact(108, 109). However, whether KMT2C acts as a global regulator of ERα-dependent gene expression has yet to be understood.

KMT2C as a tumor suppressor

KMT2C is among the most commonly mutated genes across all cancer with the majority being truncation mutations that are upstream of the SET domain. Therefore, all truncation mutations seen would result in loss of

enzymatic activity. *KMT2C* missense mutations are very rare in the SET domain (<1%) and, compared to other KMT2 family members, are disproportionately high in the PHD domains (~16%). The majority of *KMT2C* mutations are heterozygous with approximately only 15% being homozygous.

Multiple lines of evidence suggest that KMT2C acts as a tumor suppressor. Heterozygous knockout of the *KMT2C* Drosophila homolog, *trr*, in eye imaginal disks results in a growth advantage over cells with WT *trr* cells(45). Moreover, *KMT2C* whole body homozygous knockout in mice resulted in urothelial carcinomas in approximately 50% of the mice. This penetrance reached 100% when KMT2C was knocked out in a p53 null background(82). However, the results presented with whole body *KMT2C* knockout do not recapitulate the haploinsufficiency seen in human urothelial cancer and other *KMT2C* mutated cancers. Instead, this haploinsuffiency was recapitulated when KMT2C knockdown and knockout was attempted in hematopoietic stem cells that were p53 and NF1 deficient(81). In this model, only approximately 50% KMT2C knockdown and only heterozygous *KMT2C* knockout was achieved implying that KMT2C is haploinsufficient. Additionally, in this model, KMT2C loss resulted in increased AML progression upon adoptive transfer of these cells into host mice(81). In the context of mammary stem cells, it was shown that *KMT2C* knockout, in the context of p53 deletion and PI3K pathway activation, resulted in a more aggressive mammary carcinoma(110). However, the mammary carcinoma described in this study did not recapitulate any known breast cancer subtype and so it remains unclear if this model presents the true contribution of

KMT2C loss in breast cancer. Breast cancer is an especially interesting setting to study KMT2C loss as *KMT2C* is among the ten most frequently mutated genes in this disease.

Estrogen receptor positive (ER+) breast cancer

Breast cancer disease

In 2016 there were over 230,000 new cases of invasive breast cancer resulting in over 40,000 deaths, making breast cancer the second largest cause of cancer related deaths in American women(111). Prognosis and therapy for breast cancer patients is largely guided by the expression of nuclear receptors, including ERα and the progesterone receptor (PR), and the extracellular membrane-bound receptor tyrosine kinase, HER2. Luminal A and Luminal B breast cancer tend to express ERα with Luminal A being largely HER2- and Luminal B being either HER2- or HER2+. These subtypes are generally referred to as being ER+ and dependent on the ERα signaling pathway for proliferation. Luminal A and Luminal B breast cancer comprise over two thirds of all breast cancer and are generally associated with favorable prognosis, with Luminal A breast cancer having higher survival rates than Luminal B. HER2+ breast cancer tends to overexpress HER2, making it sensitive to direct antagonism of the HER2 signaling pathway. Anywhere between 5-15% of breast cancer is HER2+. Finally, triple negative breast cancer is named so for lack of expression of ERα, PR and HER2. These cancers are largely basal in cell type and have the worst prognosis of all breast cancer subtypes, in part due to the lack of targeted therapeutics.

Triple negative breast cancer makes up about 15-20% of all breast cancer cases(111).

ER+ breast cancer therapeutics and resistance

As ER+ breast cancer is highly dependent on estrogen-mediated mitogenic signals, therapies targeting ER+ breast cancer have predominantly focused on reducing the amount of estrogen present in the body or directly antagonizing ERα. Ovary ablation in premenopausal women as well as aromatase inhibition in postmenopausal women reduces estrogen production. Alternatively, selective estrogen receptor modulators (SERMs, e.g. Tamoxifen) and selective estrogen receptor degradors (SERDs, e.g. Fulvestrant) directly antagonize ERα activity(112). Despite the high success rate of these drugs, many patients eventually develop resistance with the reoccurrence of tumors that are no longer sensitive to these hormone therapies.

Until recently, the most accepted mechanism of resistance to hormone therapy was overexpression of HER2 and/or activation of the PI3K pathway. Activation of these alternative growth-promoting pathways can serve to bypass the dependence on ERα. Additionally, there are multiple lines of evidence indicating crosstalk between the PI3K-AKT-mTOR and ERα signaling pathways that may be contributing to the ability of PI3K activation to contribute to endocrine resistance(113). Specifically, studies have indicated that treatment of ER+ breast cancer cells with fulvestrant or the aromatase inhibitor, letrozole, results in increased phospho-AKT(114, 115) suggesting that ERα inhibition induces PI3K-

AKT pathway activity. Similarly, it has also been shown that PI3K-AKT pathway inhibitors increase ERα transcriptional activity by inducing increased ERα binding to the promoters of target genes such as *PGR* and *GREB1*(114). This may be partially mediated by the release of AKT-mediated inhibition of KMT2D, whose activity may contribute to ERα-dependent transcription (116). Given the cross talk between the ERα signaling pathway and the PI3K pathway and the frequency of PI3K mutations among ER+ patients, the BOLERO-2 trial found that combined treatment of everolimus, an mTORC1 allosteric inhibitor, and letrozole improved progression-free survival(117). Other proposed mechanisms of resistance include increased expression of cell cycle regulators such as MYC, Cyclin D1 and Cyclin E1, decreased expression or activity of Rb, p21 and p27, increased expression of ERα coactivators, such as AIB1, or loss of ERα expression(118).

More recently, across multiple efforts to sequence the tumors of patients with hormone-resistant ER+ metastatic breast cancer, *ESR1* mutations have been found in approximately 25-40% of this cohort(119-122). These *ESR1* mutations are rarely found in primary, hormone-sensitive tumors, where they occur at a frequency of less that 1%(119). Most of these mutations occur within the ERα ligand-binding domain with residues Y537 and D538 being the most frequently altered(123). Patients with the Y537S or the D538G mutation have significantly worse prognosis than patients with WT *ESR1*(124). Structural analysis revealed that many of these mutations promote a constitutively active conformation even in the absence of ligand(125). Functional characterization of

these mutant ERα proteins revealed that many of them confer ligand-independent activation and were resistant to doses of tamoxifen and fulvestrant that would be sufficient to inhibit WT ERα(119, 122, 123). However, xenograft models of these mutations seem to remain sensitive to second generation SERDs, such as AZD2549, that have better bioavailability than fulvestrant(126).

Estrogen receptor signaling pathway

Estrogen receptor alpha

ERα is a hormone-dependent member of the nuclear receptor superfamily of ligand-dependent transcription factors. The ERα protein includes an activating function (AF)-1 domain, a DNA binding domain, a dimerization domain and an AF2 domain. AF1 is poorly conserved and is thought to have tissue- or promoter-specific activities(127). The DNA binding domain is highly conserved and contains two zinc-finger binding motifs that are thought to mediate direct contact with the DNA(128, 129). Finally, the AF2 domain mediates ligand binding and ligand-dependent transcriptional activity(130).

The ERα protein is subject to multiple levels of post-translational modification including phosphorylation, methylation, acetylation, sumoylation and ubiquitination. These post-translation modifications are thought to influence ERα protein stability, localization, dimerization, interaction to coregulatory proteins, and genomic activity(131). Phosphorylation at S104/106, S118 and S167 are among the better studied sites of ERα post-translation modification. Phosphorylation at S104/106 is induced by estradiol (E2), but not growth factor

stimulation(131). While the direct kinase responsible for phosphorylating this mark is unknown, it has been speculated that MAPK(132), GSK3(133) or CDK2(134) may be involved. Phosphorylation of S118 is thought to be downstream of growth factor signaling, specifically downstream of EGF or IGF-1 mediated RAS-MAPK pathway activation(135, 136). S118 is the most widely studied ERα modified site and is thought to be important for ERα dimerization(137). S118 phosphorylation is also thought to be important for direct binding to ERα target genes(138) and coactivator binding(139, 140) suggesting that it plays a key role in the genomic activities of ERα. Similarly, phosphorylation of S167 is also considered to be important for ERα transcriptional activity(141) but this site is thought to be phosphorylated by AKT(142) or p90RSK(141). Overall, post-translation modification of ERα is one way that the ERα signaling pathway can converge with other known signaling pathways such as the MAPK pathway and the PI3K-AKT pathway.

Ligand-dependent ERα activity

In the absence of its ligand, estrogen, ERα is thought to be relatively unstable and subject to ubiquitin-mediated degradation. Unliganded, cytoplasmic ERα is therefore often found in complex with chaperone proteins, such as HSP90 (143-146). Ligand-dependent ERα signaling begins upon estrogen binding which induces a conformational change that stabilizes the receptor and mediates its dissociation from chaperone proteins(147). Aside from receptor stabilization, ligand-binding simultaneously unmasks the domains necessary for ERα

dimerization (148), nuclear localization, DNA-binding, and coactivator/ corepressor binding(149). Therefore, ligand-bound ERα can then homodimerize and move into the nucleus where it can bind to estrogen response elements (EREs) in the DNA, bind to coactivators, and recruit the necessary enzymatic components necessary for the transcription of ERα-target genes (Fig. 1.2).

Ligand-bound ERα is also thought to have nongenomic functions where it activates growth factor signaling by acting predominantly as a scaffolding protein in the cytoplasm(150). Specifically, ligand-bound ERα can associate with Src and Shc and promote MAPK pathway activation through the Shc/ Src/ Ras/ ERK node(151).

Physiologic role of ERα

ERα plays key roles in the development of reproductive organs including the ovaries, the uterus and the mammary gland, among others. Whole body knockout of *Esr1* in mice resulted in defects in these very organs. Specifically, the ovaries had large cysts and diminished ovulation. While prepubertal uteruses from *Esr1* knockout mice look similar to their control counterparts, they were not able to respond to estrogen-stimulation during puberty and displayed inhibited epithelial cell proliferation and reduced size. Finally, mammary glands from *Esr1* knockout mice were completely devoid of epithelial ducts branching(152). During puberty, epithelial ducts branch outward across the entire mammary gland. These ducts remain relatively dormant following puberty and then differentiate

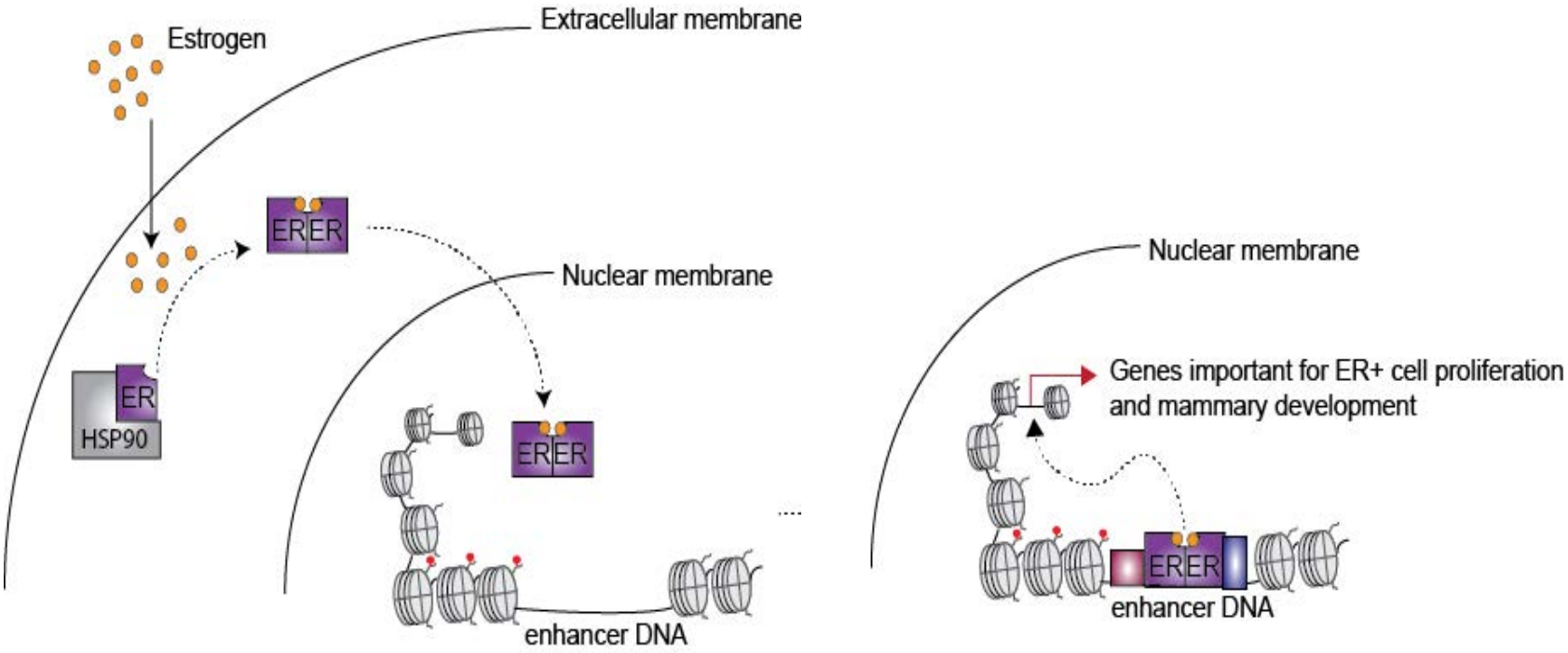

Figure 1.2 ERα signaling pathway. In the absence of ligand, ERα is relatively unstable and is thought to complex with chaperone proteins, such as HSP90. In the presence of estrogens, ERα can bind the estrogen, stabilize, homodimerize and move into the nucleus. In the nucleus, ERα binds to EREs predominantly at enhancer regions of the genome where it then recruits coactivators. ERα can then mediate the transcription of target genes, many of which are important divers of proliferation and of mammary gland development.

again into alveolar cells during pregnancy forming alveolar ducts that produce milk used to nurse pups(153, 154).

Floxed *Esr1* mice were used to study the effects of mammary-specific *Esr1* deletion on the mammary gland development. Use of the mouse mammary tumor virus (MMTV) promoter allowed for the assessment of puberty-specific mammary development and use of the whey acidic protein (WAP) promoter allowed for the assessment of pregnancy-related mammary development(155). MMTVcre-mediated *Esr1* deletion resulted in mammary glands that had significantly diminished epithelial duct branching. Specifically, the ERα depleted mammary glands lost secondary branching(156). It should be noted that the MMTV-mediated phenotype was not as severe as that seen in the mammary glands of whole body *Esr1* knockout mice. This suggests that perhaps ERα mediates some degree of mammary gland development that is prepubertal. Alternatively, it could also suggest that use of the MMTV promoter results in incomplete *Esr1* deletion or does not result in deletion immediately at the onset of puberty allowing for some pubertal ERα-mediated development to occur before it is deleted.

WAPcre-mediated deletion of *Esr1* resulted in dilated alveolar ducts with reduced milk production in the mammary glands of nursing mothers. This subsequently resulted in reduced weight of pups that were nursed by mothers with ERα deficient mammary glands as compared to pups nursed by control mothers(156). Together, these data indicate that ERα is a key regulator of multiple reproductive organs.

ERα coregulatory proteins

The presence of coactivator and corepressor proteins plays a major role in ERα-mediated transcriptional activity. These coactivators generally bind ERα through a highly conserved LXXLL domain called the "nuclear receptor box(157)." It is thought that these nuclear receptor box-containing proteins may act as scaffolding platforms for the recruitment of complex regulatory proteins and chromatin modification enzymes. Examples of such ERα coactivators include TIFα, Tip60, AIB1, SRC-1. SRC-2, and TRAP220(158). ERα is also known to regulate transcription at non-ERE sites by tethering to other transcription factors such as AP-1(159-161), Sp1(162, 163) and NF-κB(158). In fact, knockdown of cFOS, a component of the AP-1 transcription complex, reduces expression of over one-third of all estrogen-regulated target genes(164). While ERα is generally considered a transcriptional activator, it is also important to note that for many genes ERα acts as a transcriptional repressor. ERα is known to interact with corepressors such as the nuclear receptor corepressor (NCoR) and multiple HDACs(165).

In addition to coactivator or corepressor proteins, ERα is also known to be heavily reliant on the activity of pioneer factors, such as FOXA1. FOXA1 silencing dramatically reduces the ability of ERα to bind to a large majority of estrogen-regulated genes(166). FOXA1 was found to be localized to enhancers of ERα target genes in the absence of E2, suggesting that it primes the chromatin for ERα binding(167).

Interestingly, enhancers of ERα target genes that were bound by FOXA1 were also enriched with H3K4me1/2 residues. Overexpression of the H3K4me1- and H3K4me2-specific demethylase KDM1 resulted in loss of H3K4me and diminished FOXA1 recruitment at those enhancer marks(167). FOXA1 has been shown to directly interact with KMT2C(107). Whether FOXA1 binding promotes the addition of H3K4me or whether H3K4me promotes FOXA1 binding is still unclear. In MCF7 cells, FOXA1 silencing did not reduce the presence of the H3K4me marks at enhancer regions(167). However, when FOXA1 was ectopically expressed in the ER- cell line, MDA-MB-231, there was an increase in H3K4 methylation at FOXA1 binding sites. Regardless, H3K4 methylation and FOXA1 act to prime DNA for ERα binding. The importance of H3K4 methylation marks in ERα activation is further evidenced as TIP60, a histone acetyltransferase required for the expression of certain ERα targets, is recruited by H3K4me1 marks(168).

KMT2C, was shown to effect the ERα-mediated transcription of certain HOX genes. However, in the case of TIP60 recruitment or E2- mediated *TFF1* transcription, it seems as though SET1a and KMT2A/B are the primary HMTs(168-170). Another study determined that, in MCF7 cells, there is a different set of coregulators necessary for each of six ERα target genes and that certain coregulators, such as the histone methyltransferase G9a, can act as a coactivator of one gene while serving as a corepressor for another(171). These studies convey the importance of coactivator proteins and chromatin modification enzymes, including the KMT2 family of histone methyltransferases, in regulating

ERα transcriptional activity and show that different ERα target genes are differentially regulated by the coactivators/ corepressors.

Introduction to dissertation

KMT2C is an H3K4 monomethyltransferase and is a member of the KMT2 family of histone methyltransferases. Multiple lines of evidence implicate many members of this family as regulators of nuclear receptor transcriptional activity. In particular, KMT2C has been shown to regulate the activity of LXR, PPARγ, ERα, and GR for candidate target genes as part of a larger complex called ASCOM. However, these studies have not always been conducted in tissues where these nuclear receptors are relevant. In addition, whether or not KMT2 family members act as global regulators of nuclear receptor gene expression has yet to be understood.

Given the potential for KMT2C, and other KMT2 family, to act as coregulators for nuclear receptors and other transcription factors, they may serve as key mediators of differentiation and development in tissues types where these transcription factors are relevant. One such tissue type is mammary tissue, which is highly dependent upon the actions of ERα. Interestingly, KMT2C ranks among the top ten most frequently mutated genes in mammary carcinomas, of which 70% express ERα and depend upon ERα downstream signaling. Despite this frequent mutation, we have yet to understand the role of KMT2C in normal mammary development, in mammary homeostasis, and the selective advantage of *KMT2C* mutation in breast cancers.

Here, we show that KTM2C serves as a regulator of a subset of ERα target genes. Loss of KMT2C impairs ERα-dependent mammary gland development and inhibits ER+ breast cancer cell proliferation. In our efforts to better understand -the selective advantage of KMT2C mutation, we find that loss of KMT2C promotes resistance to hormone deprivation, which is commonly used to treat ER+ breast cancer. As such, gene expression signatures of KMT2C loss predict for poorer survival in Luminal A/B breast cancer patients. Overall, we nominate KMT2C as a candidate regulator of ERα activity whose loss allows cells to escape hormone sensitivity.

CHAPTER 2: KMT2C IS A KEY REGULATOR OF ESTROGEN RECEPTOR ALPHA TRANSCRIPTIONAL ACTIVITY

Introduction

KMT2C is a member of the family of H3K4 histone methyltransferases. Members of this family have an enzymatically active SET domain that is responsible for mono-, di- or tri- methylating H3K4 residues (172). H3K4 methylation, in general, is known to demarcate regions of active transcription with H3K4me1 being more abundant at enhancers regions of the genome and H3K4me3 being more abundant at promoter regions of the genome (173-175). KMT2C, in particular is known to specifically be an H3K4 monomethylatransferase and is therefore thought to be more active an enhancer regions of the genome (51, 52).

Within the KMT2 family, KMT2A and KMT2B are thought have some functionally redundancies as they can both interact with the nuclear receptor coactivator, Menin (176). Similarly, KMT2C and KMT2D are also thought to be somewhat functional redundant as they can both interact with ASC-2 (91), another nuclear receptor coactivator (177), and participate in the ASCOM complex. The ASCOM complex also includes known regulators of H3K4 methyltransferase activity including ASH2L and WDR5, as well as the H3K27 demthylase, UTX(178). As part of the ASCOM complex, it has been suggested that KMT2C and KMT2D can influence the activity of multiple nuclear receptors including the LXR, FXR, GR, PPARY, PR and ERα (83, 84, 99, 178, 179) by promoting a transcriptionally active chromatin state. However, the effects of KMT2C on global nuclear receptor activity have yet to be characterized.

In this report, we show that KMT2C is among the most frequently mutated genes in breast cancer, a disease in which the nuclear receptor, ERα, is known to play a major role. We go on to show that KMT2C serves as a regulator of global ERα activity with KMT2C loss specifically affecting known ERα-regulated phenotypes including proliferation and mammary-gland development.

Results

KMT2C is among the most frequently mutated genes in breast cancer

Upon assessing the most frequently mutated genes in over 700 metastatic breast tumors sequenced using the MSKCC-IMPACT platform, we found that *KMT2C* was among the top 10 most frequently mutated genes with a mutation frequency of approximately 9.8% (Fig. 2.1A). This mutation frequency is relatively consistent with the 7% *KMT2C* mutation frequency reported by the Cancer Genome Atlas (TCGA) (180) for primary breast cancer.

We then wanted to compare *KMT2C* mutation frequency to that of its most closely related homologs to determine if frequent mutation of H3K4 histone methyltransferases was common among multiple family members. We found that *KMT2C* was by far the most frequently mutated H3K4 histone methylatransferase within its family with the second most frequently mutated family member, *KMT2D*, having a mutation frequency less than 3% (Fig. 2.1B).

The KMT2C mutation spectrum mostly consists of truncation and missense mutations that are spread out along the length of the protein (Fig. 2.1C). It should also be noted that the KMT2C enzymatically active SET domain

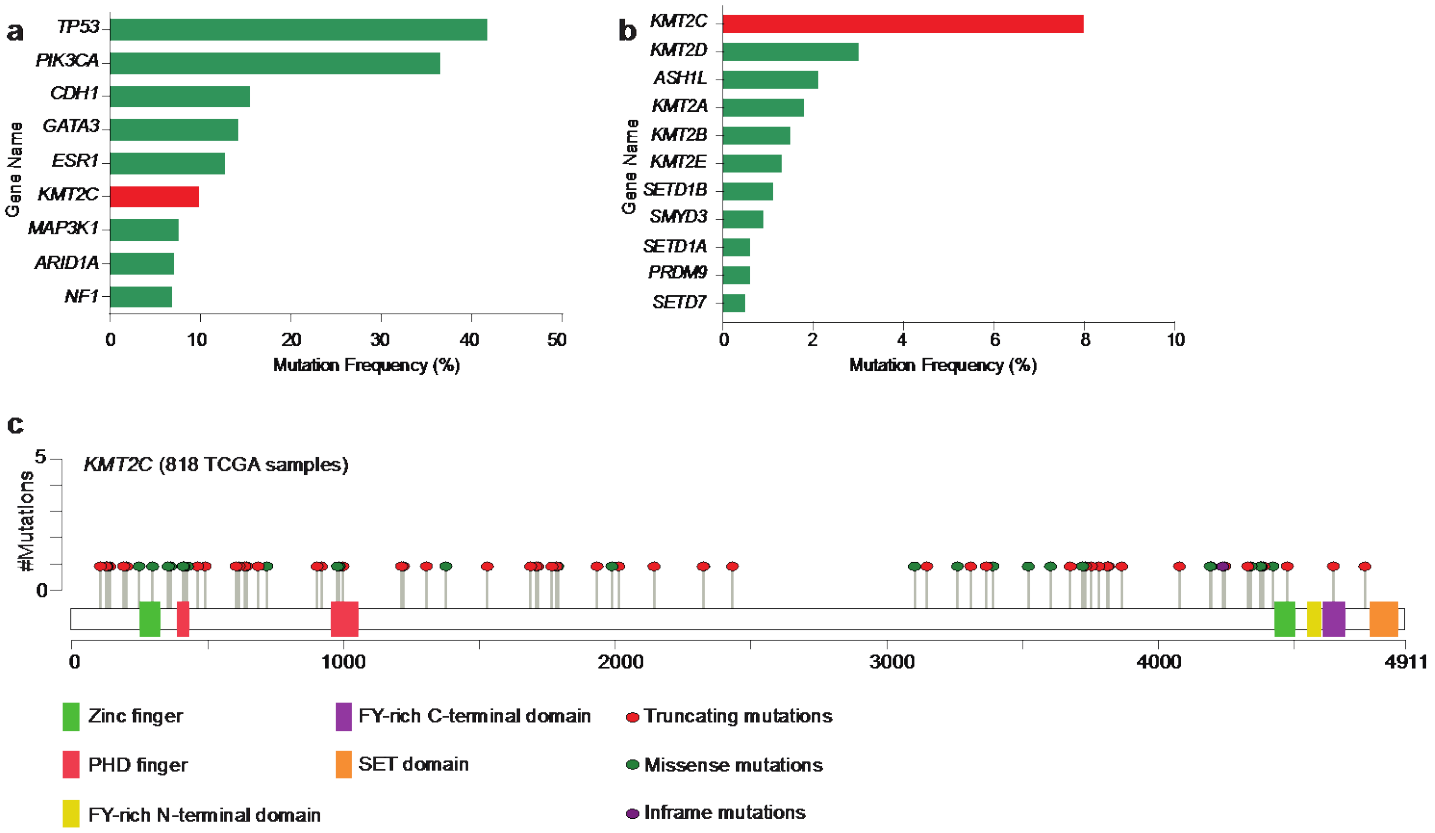

Figure 2.1 KMT2C ranks among the most frequently mutated genes in breast cancer. (A) Percentage of cases with gene mutations detected in the MSKCC-IMPACT metastatic breast cancer patient cohort(181) (n= 746). (B) Mutation frequencies of H3K4 histone methyltransferases in breast cancer tissue samples from the TCGA dataset(180) (n=818). (C) Diagram of KMT2C domains with the locations of the identified TCGA mutations. y-axis corresponds to the number of cases with indicated mutation (n=818).

is located at the C-terminal end of the protein. Therefore, even if any of the truncation mutations resulted in a truncated protein that was expressed, it would likely have no enzymatic activity. This mutation spectrum suggested that KMT2C is a tumor suppressor whose loss of function is somehow selectively advantageous to breast cancer cells.

Given that different breast cancer subtypes have different proliferative drivers, we wanted to see if KMT2C mutation was more or less frequent in any particular subtype. However, we found that *KMT2C* mutation frequency was relatively evenly distributed across breast cancer subtype with a range of 5-12% (Fig. 2.2), consistent with the frequency found in total breast cancer. Given the *KMT2C* mutation spectrum and the even distribution of *KMT2C* mutation across breast cancer subtype, we decided to model KMT2C loss across a variety of breast cancer cell lines that are representative of the different subtypes.

KMTC2 knockdown in breast cancer cell lines

In order to model KMT2C loss in breast cancer, we cloned two short hairpin RNAs (shRNAs) against *KMT2C* (shKMT2C) (182). We found that both short hairpins, shKMT2C#1 and shKMT2C#2, specifically reduced expression of KMT2C by 60-70% without affecting expression of its most closely related homologs, KMT2A, B, or D (Fig. 2.3A). We stably expressed shKMT2C#1 and shKMT2C#2 in a panel of cell lines representative of the clinical subtypes of breast cancer including ER+, ER-HER2+, ER+HER2+ and triple negative and found a similar degree of knockdown across the cell lines (Fig. 2.3B). The breast

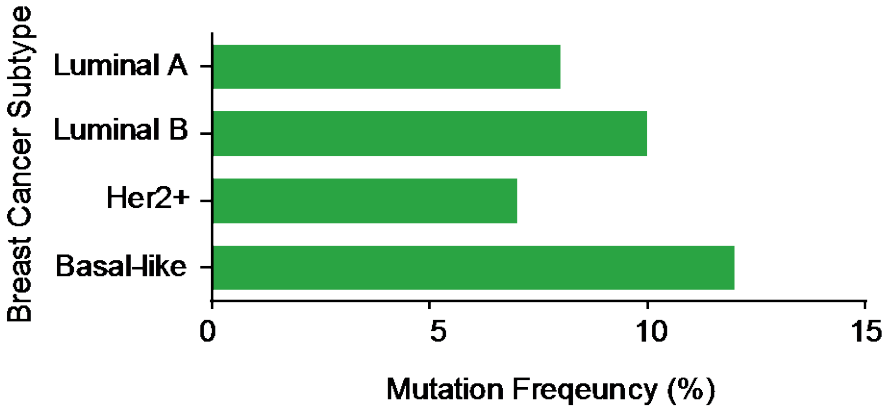

Figure 2.2 KMT2C mutations are evenly distributed across breast cancer subtype. Percentage of KMT2C mutations across different breast cancer subtypes from the TCGA dataset(180) (n=818).

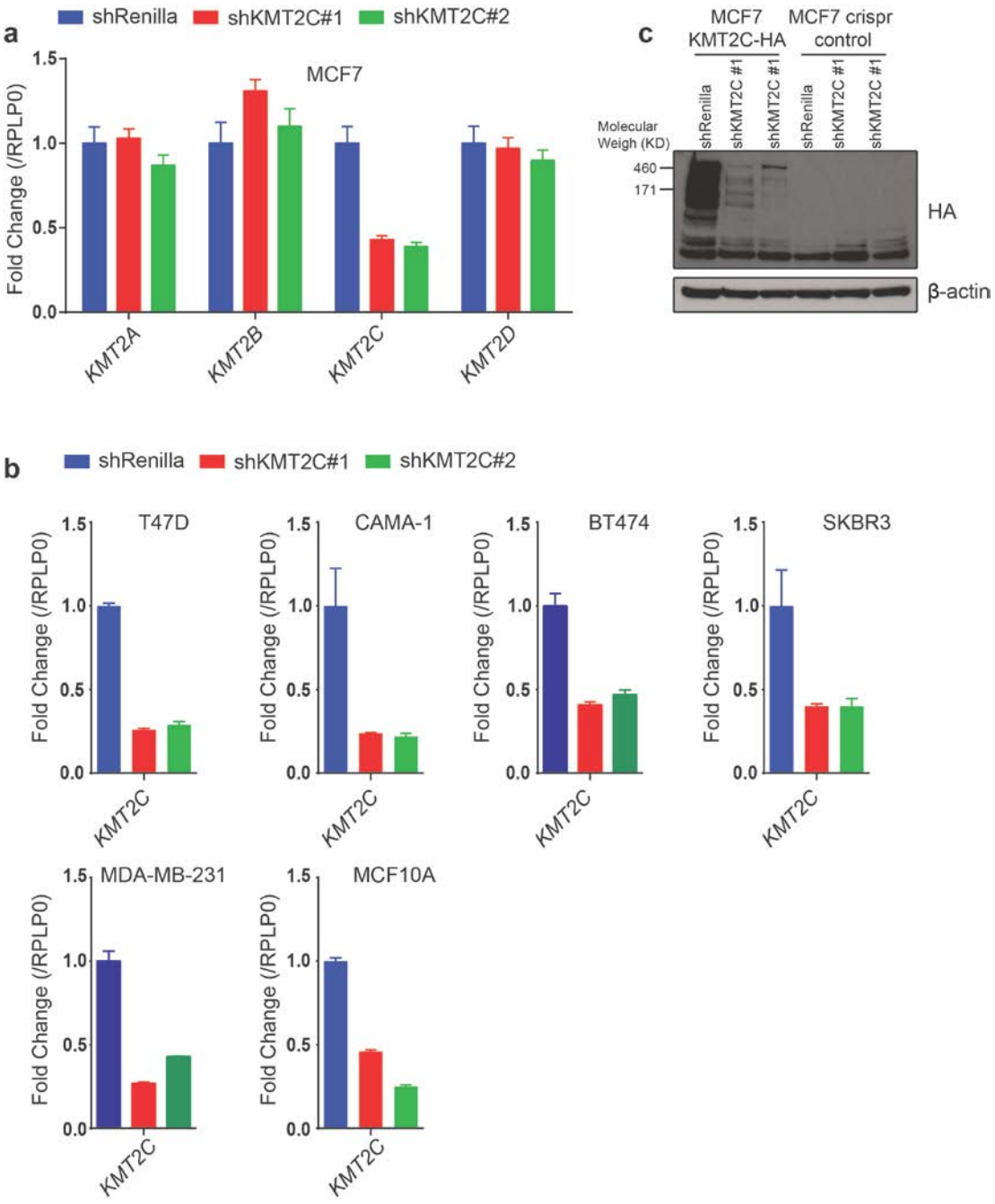

Figure 2.3 KMT2C knockdown in breast cancer cell lines. (A) mRNA levels of indicated genes, as measured by qRT-PCR, in shRenilla, shKMT2C#1 and shKMT2C#2 MCF7 cells. Values correspond to the mean of three replicates ± s.e.m. (B) mRNA levels of KMT2C, as measured by qRT-PCR, in shRenilla, shKMT2C#1 and shKMT2C#2 T47D, CAMA-1, SKBR3, MDA-MB-231 and MCF10A cells. Values correspond to the mean of three replicates ± s.e.m. Data correspond to one representative assay from a total of two or three independent assays. (C) Immunoblot of the indicated proteins in shRenilla or shKMT2C KMT2C-HA or Crispr control MCF7 cells cells. β-actin was used as a loading control.

cancer cell lines used all underwent next generation sequencing using the MSKCC-IMPACT platform and showed no clear deleterious mutation in KMT2C (Table 2.1).

To confirm sufficient knockdown of KMT2C protein levels, we used MCF7 cells engineered to express HA at the C-terminal end of an endogenous *KMT2C* allele (KMT2C-HA cells, Fig. 2.4). The generation of this cell line allowed us to probe for KMT2C using antibodies against HA. We confirmed expression of KMT2C-HA at both mRNA and protein levels and also confirmed that mRNA levels of total KMT2C were unchanged (Fig. 2.4A-C). Finally, we confirmed that the KMT2C-HA cell line maintained nuclear expression of KMT2C-HA (Fig. 2.4D) suggesting proper localization.

Expression of shKMT2C #1 and shKMT2C #2 in these cells resulted in knockdown of KMT2C-HA by immunoblotting (Fig. 2.3C). It should be noted that we were, at best, able to achieve close to 75% knockdown of KMT2C. This is in keeping with previously published data suggesting that cancer cells may select for haploinsufficiency (81).

KMT2C loss inhibits proliferation specifically in breast cancer cell lines dependent on ERα signaling

We subsequently used the shKMT2C expressing breast cancer cell lines to assay whether KMT2C loss had any effects on cell proliferation. We found that KMT2C knockdown resulted in a 40-70% reduction in proliferation selectively in the three ER+HER2- cell lines examined, MCF7, T47D and Cama-1 (Fig. 2.5A).

Table 2.1. *KMT2C* mutation status in breast cell lines.

Cell Line	Chromosome	Nucleotide	Ref	Alt
Cama-1	***KMT2C* WT**			
MDA-MB-361	***KMT2C* WT**			
MCF7	***KMT2C* WT**			
MDA-MB-231	***KMT2C* WT**			
MCF10A	***KMT2C* WT**			
T47D	**7**	**151932996**	**C**	**T**
SKBR3	**7**	**151878785**	**G**	**C**

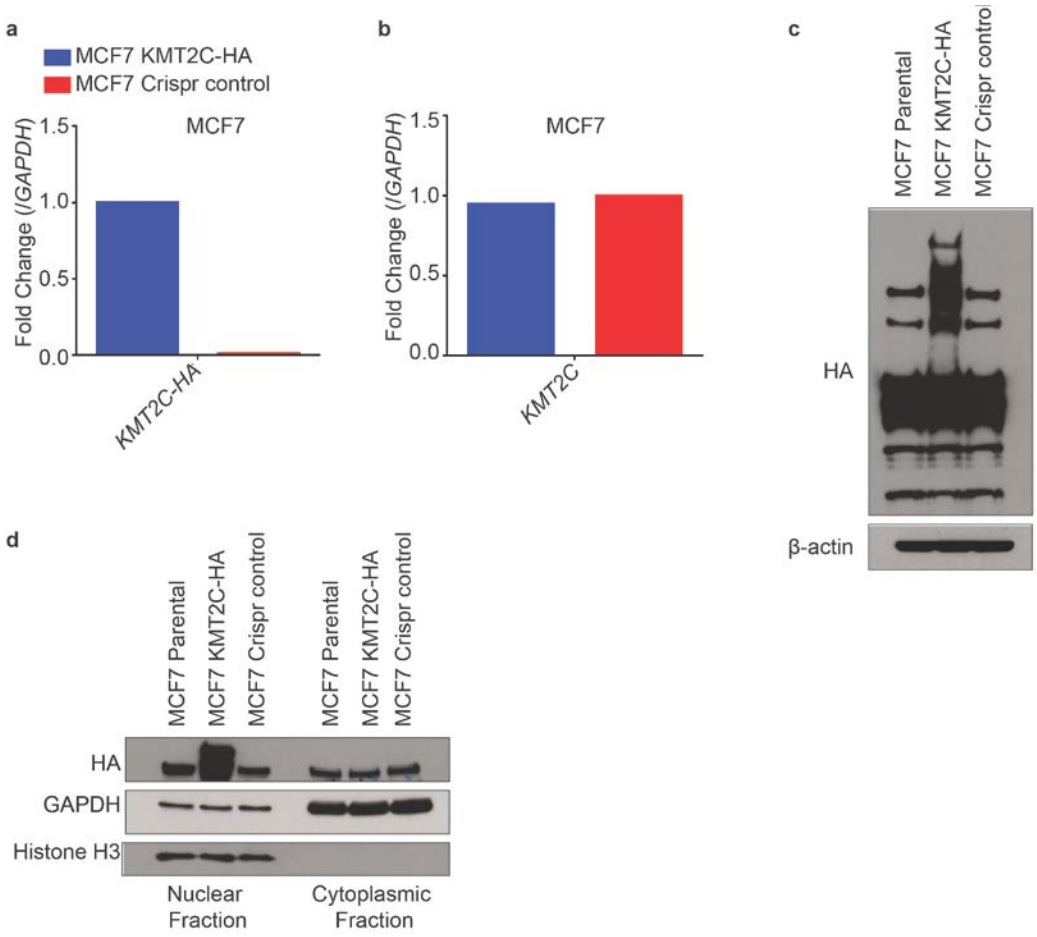

Figure 2.4 MCF7 KMT2C-HA cells have proper expression and localization of KMT2C-HA. (A) mRNA levels of KMT2C-HA, as measured by qRT-PCR, in MCF7 KTM2C-HA cells and MCF7 Crispr control cells. (B) mRNA levels of KMT2C, as measured by qRT-PCR, in MCF7 KTM2C-HA cells and MCF7 Crispr control cells. (C) Immunoblot of indicated proteins in shRenilla or shKMT2C-HA cells. β-actin was used as a loading control. (D) Immunoblot of indicated proteins in shRenilla or shKMT2C-HA cells following nuclear fractionation. GAPDH was used as a cytoplasm-specific control, Histone H3 was used as a nucleus-specific control.

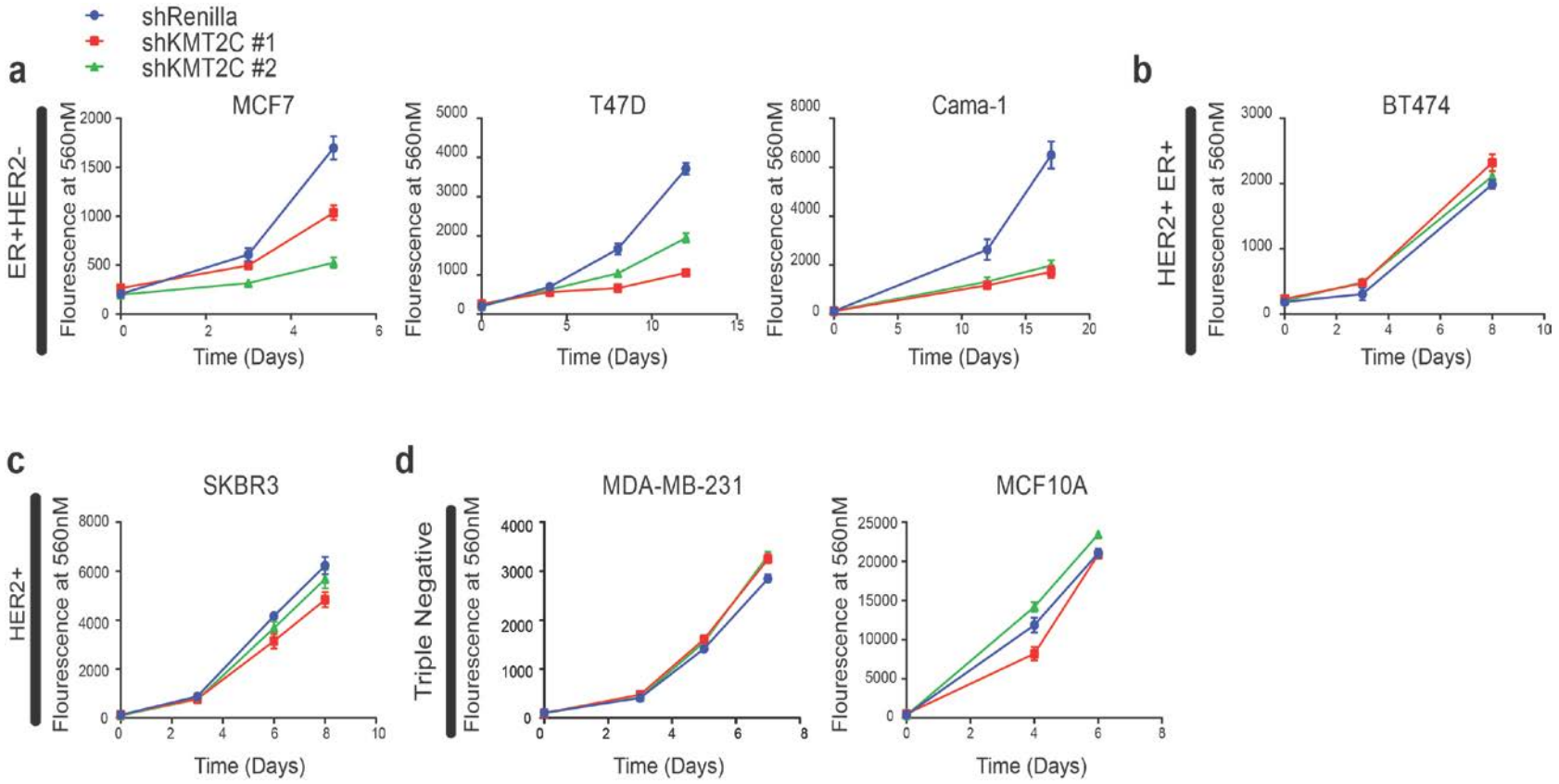

Figure 2.5 KMT2C loss inhibits proliferation specifically in breast cancer cell lines dependent on ERα signaling. (A-D) Indicated breast cancer cells stably expressing either shRenilla or shKMT2C were assayed for proliferation using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m.

In contrast to the effects seen in the ER+HER2- cells lines, KMT2C knockdown had no effect on the proliferation of the ER+HER2+ cell line, BT-474, the ER-HER2+ cell line, SKBR3 and the triple negative cell lines, MDA-MB-231 and MCF10A (Fig. 2.5B-D).

We wanted to confirm that the cell lines in which we saw the inhibited proliferation were in fact dependent on ERα signaling. Therefore, we assessed the sensitivity of MCF7, T47D and Cama-1 cells to E2 stimulation and fulvestrant inhibition and found that all three cell lines were responsive to E2 and sensitive to fulvestrant (Fig. 2.6). Together, we found that KMT2C selectively inhibits the proliferation of ER+HER2- cells that are dependent on ERα signaling.

CRISPR/CAS9nickase mediated deletion of KMT2C inhibits proliferation of MCF7 cells

We next wanted to confirm that knockout of *KMT2C* alleles would have the same effect as shRNA-mediated knockdown. To this end, we used the CRISPR/Cas9nickase system (183) to specifically target exon 6 of *KMT2C* (Fig. 2.7). In keeping with our knockdown results, we were unable to recover any cells in which all alleles of *KMT2C* could be knocked out (Fig. 2.7A). Again, this was consistent with previously published data where *KMT2C* was unable to be homozygously knocked out in leukemia cells (81). In addition, MCF7 cells are known to be aneuploid making homozygous knockout more difficult. Nevertheless, we were still able to show that reduced KMT2C expression

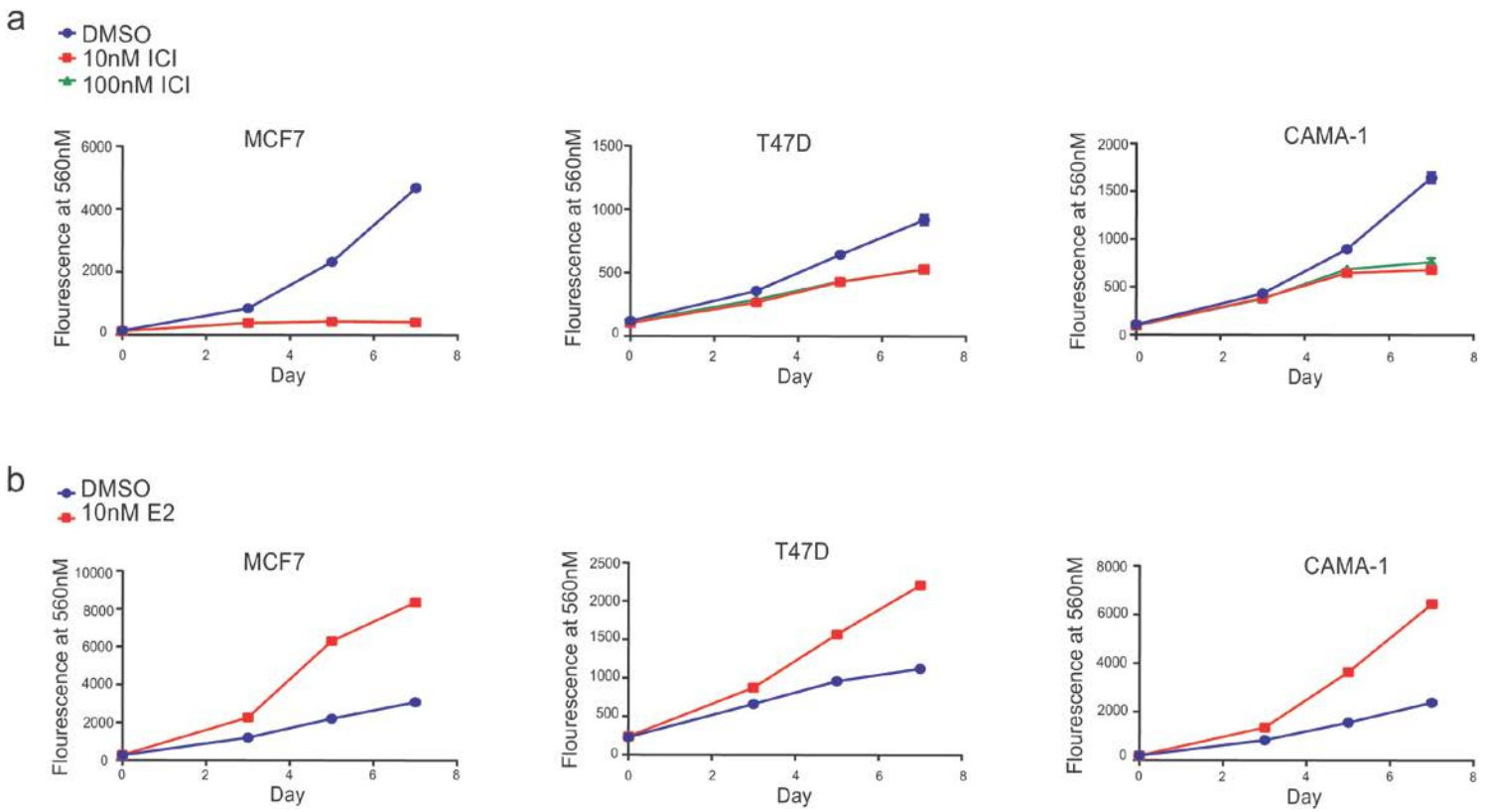

Figure 2.6 ER+HER2- cells are dependent on estrogen and ERα for proliferation (A) Indicated breast cancer cells were assayed for proliferation in full serum in the presence of either DMSO or fulvestrant using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m. (B) Indicated breast cancer cells were assayed for proliferation in phenol-red free media with charcoal stripped serum in the presence of either DMSO or estradiol (E2) using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m. Data correspond to one representative assay from a total of two or three independent assays.

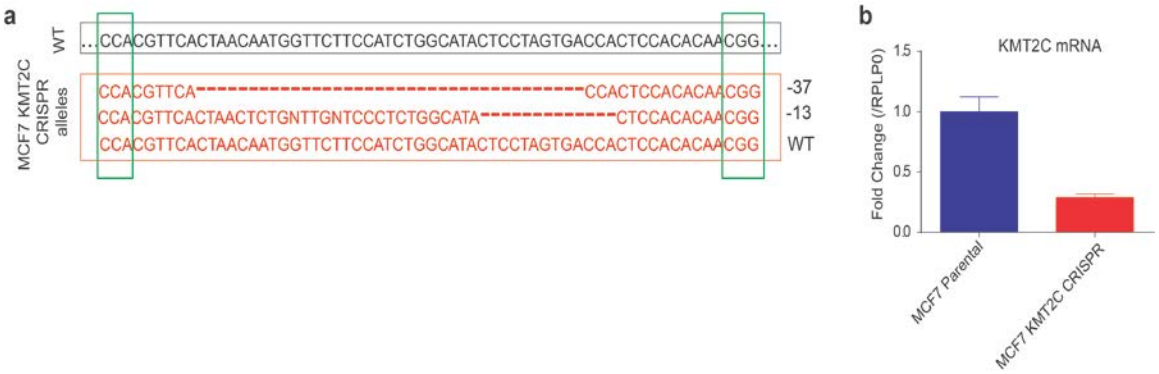

Figure 2.7 CRISPR/CAS9nickase mediated deletion of KMT2C in MCF7 cells. (A) Mutant DNA sequences induced by *KMT2C* exon 6 sgRNAs in MCF7 (resulting cell line referred to as MCF7 KMT2C CRISPR cells). The PAM sequences are shown in the green outline and deleted nucleotides shown as dashes. WT, wild-type. MCF7 KMT2C CRISPR cells retained one WT copy as indicated. (B) mRNA levels, as measured by qRT-PCR, in parental and MCF7 KMT2C CRISPR cells. Values correspond to the mean of three technical replicates ± s.e.m. Data correspond to one representative assay from a total of three independent assays.

mediated by Cas9nickase (Fig. 2.7B) resulted in inhibited proliferative ability compared to our control cells (Fig. 2.8) consistent with the knockdown results.

Differential sensitivity of breast cancer cell lines to loss of various KMT2 family members

Given the close homology between KMT2C and its remaining KMT2 family members, especially KMT2D, we wanted to examine whether loss of remaining other KMT2 family members would have similar effects on the proliferation of breast cancer cell lines. Therefore, we expressed shKMT2A, shKMT2B and shKTM2D in a representative panel of breast cancer cell lines resulting in approximately 60-75% knockdown of KMT2A mRNA, 30-60% knockdown of KMT2B mRNA and 50-70% knockdown of KMT2D mRNA (Fig. 2.9).

We observed that while the effects of KMT2C loss on proliferation were specific to the ER+HER2- cell line, MCF7, loss of KMT2A had milder effects on MCF7 and also modestly inhibited proliferation of BT474 and MCF10A cells. Loss of KMT2B did inhibit the proliferation of MCF7 by approximately 70% but also inhibited the proliferation of SKBR3 by approximately 50% and BT474 by approximately 40%. Strikingly loss of KMT2D drastically inhibited the proliferation of MCF7, BT474 and SKBR3 cells by 70-90% and also inhibited the proliferation of MCF10A cells by approximately 50% (Fig. 2.10). Together this suggests that different members of the KMT2 family have differential effects on the proliferation of breast cancer cell lines and may suggest that they regulate different pathways. Moreover, our data shows that the effects of KMT2C on ER+/HER2- breast

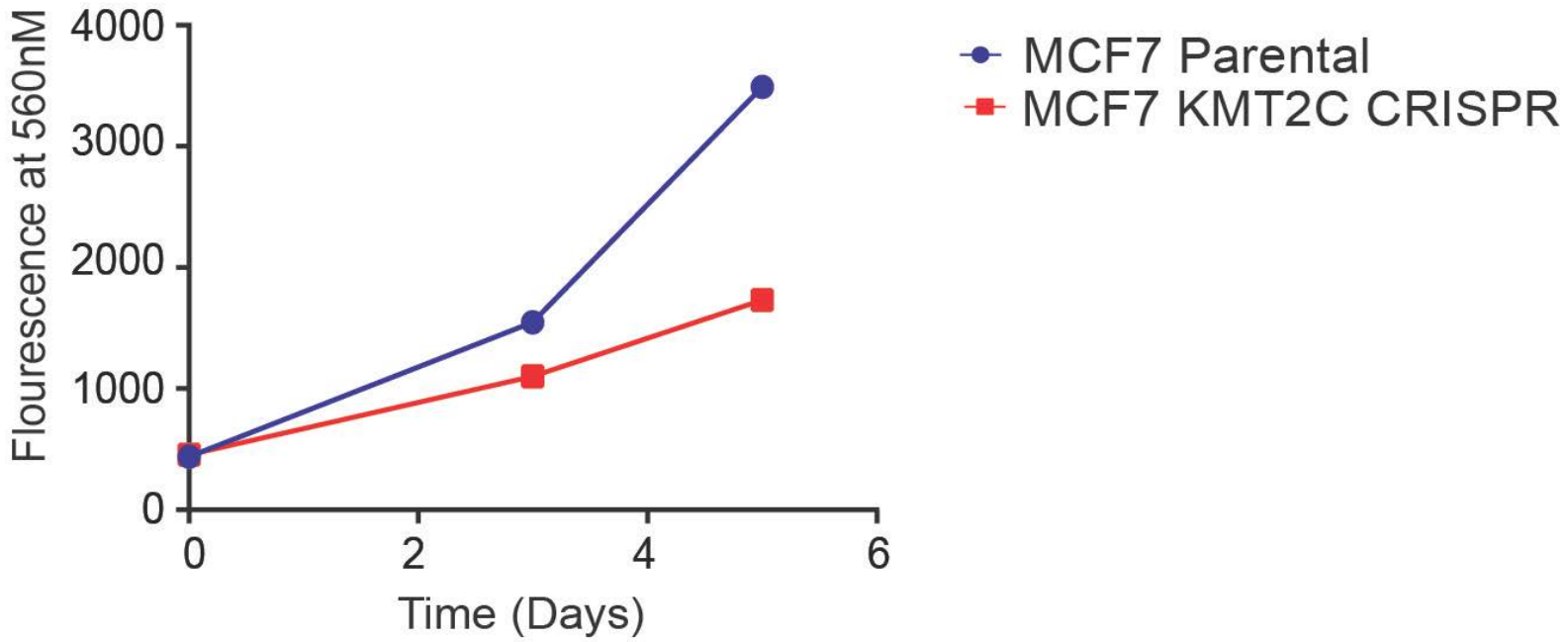

Figure 2.8 CRISPR/CAS9nickase mediated deletion of KMT2C inhibits proliferation of MCF7 cells. Parental and MCF7 KMT2C CRISPR cells were assayed for proliferation using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m. Data correspond to one representative assay from a total of two or three independent assays.

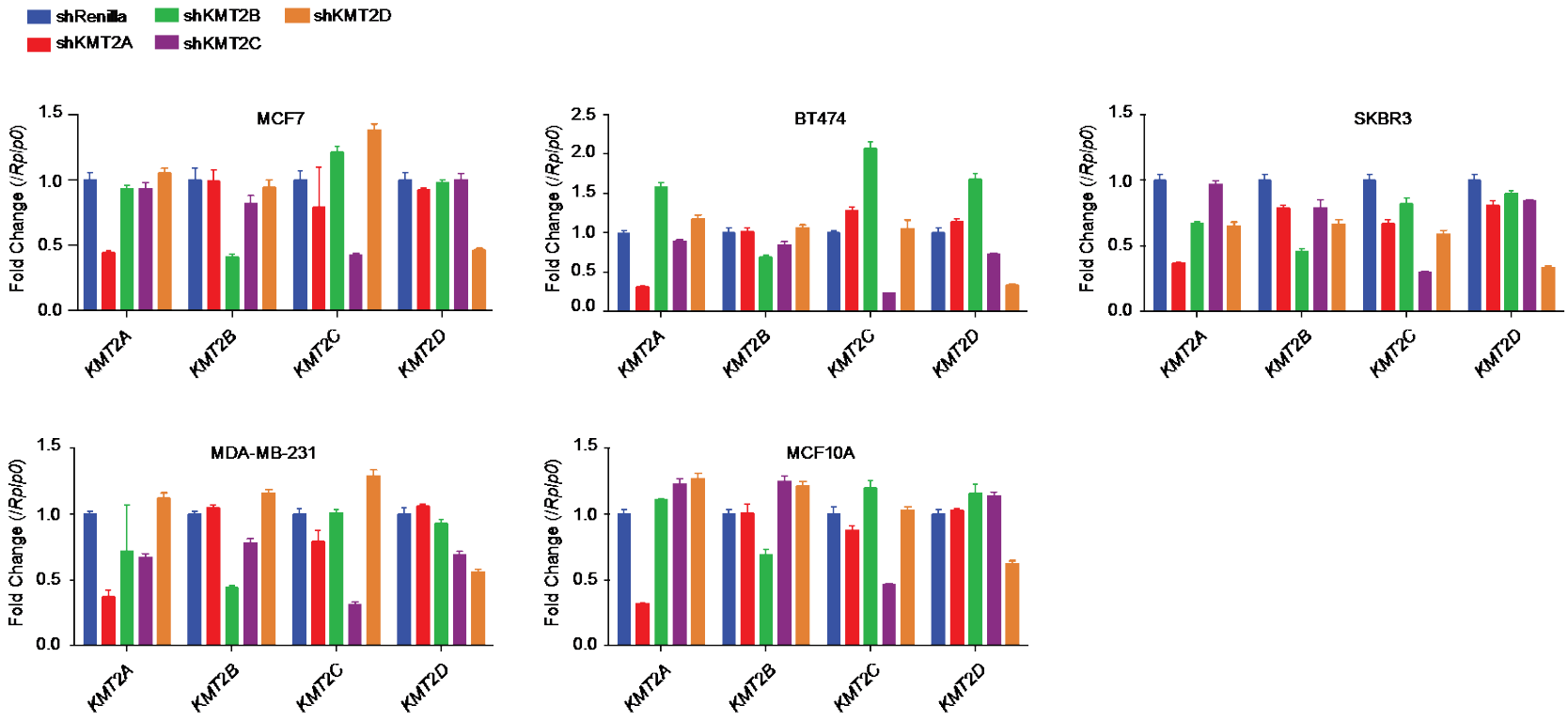

Figure 2.9 Knockdown of KMT2A, B, C and D in breast cancer cell lines. mRNA levels of indicated genes, as measured by qRT-PCR, in shRenilla, shKMT2A#1, shKMT2B#1, shKMT2C#1, and shKMT2D#1 indicated breast cancer cell lines. Values correspond to the mean of three replicates ± s.e.m.

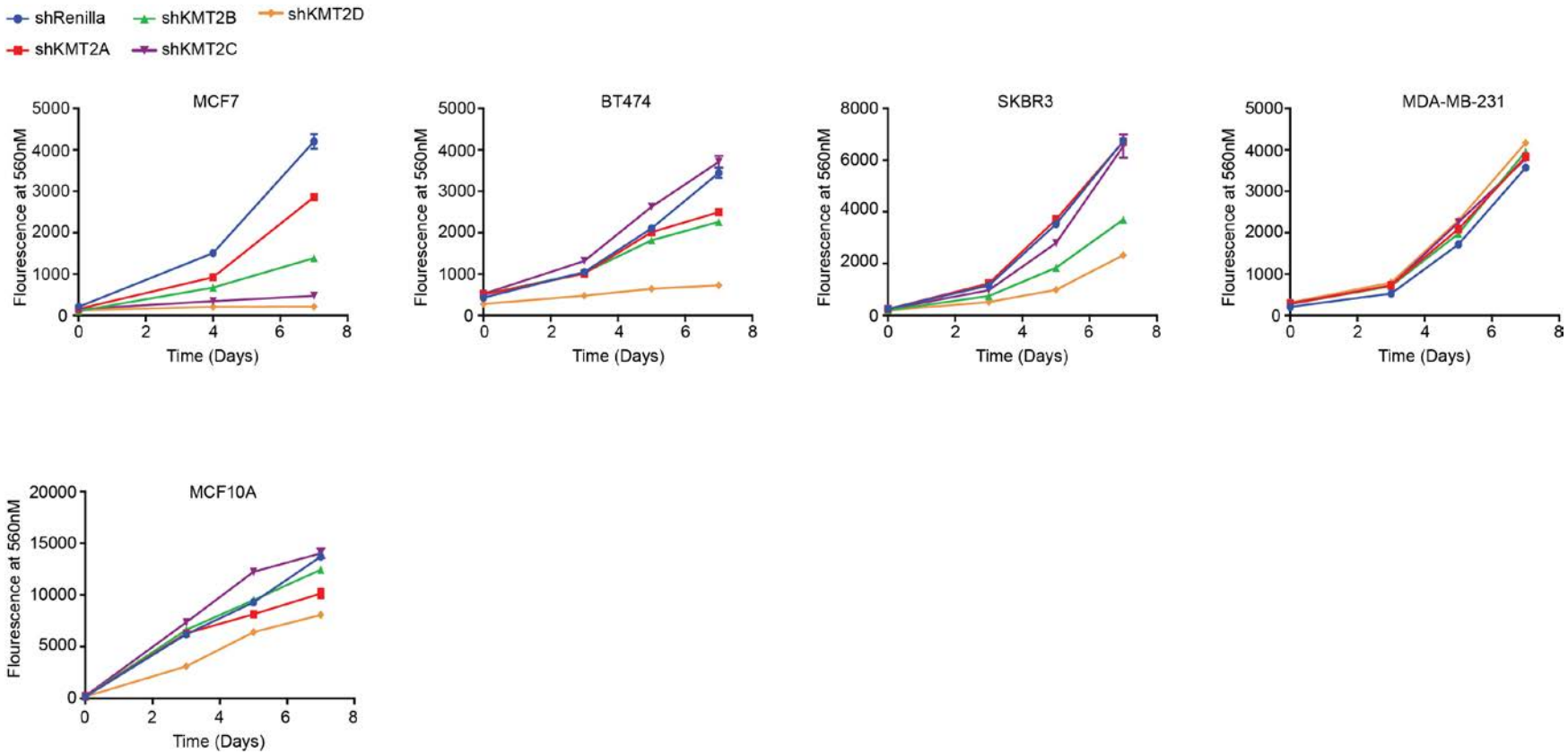

Figure 2.10 Differential sensitivity of breast cancer cell lines to loss of various KMT2 family members. Indicated breast cancer cells stably expressing either shRenilla, shKMT2A#1, shKMT2B#1, shKMT2C#1, and shKMT2D#1 were assayed for proliferation using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m.

cancer proliferation may be unique and not shared by all H3K4 methyltransferases giving credence to further investigation of the role KMT2C plays in estrogen-dependent signaling.

KMT2C regulation of pubertal mammary gland development

After establishing the effects of KMT2C on cell lines dependent on the ERα signaling pathway, we investigated whether KMT2C would have an effect on pubertal mammary gland development, which represents a physiologic setting of ERα activity. To this end, we generated mice with floxed alleles of *Kmt2c* (Fig. 2.11A-C) and crossed them with mice expressing Cre recombinase under the control of the mammary specific MMTV promoter (Fig. 2.11D-F). The MMTV promoter is known to become active at the onset of puberty and can therefore be used to study pubertal mammary development(155). Mammary specific knockout of *Kmt2c* was verified through PCR from DNA extracted from the mammary gland and through qRT-PCR from RNA extracted from the mammary gland. When examining mammary fat pads from fully mature 12-week old females, we found that homozygous MMTV-driven excision of *Kmt2c* resulted in diminished side branching and ductal elongation (Fig. 2.12A-D). We quantified the area occupied by the epithelial ducts and found a significant reduction as compared to control mice (Fig. 2.12E). Additionally, to assess the effects of Kmt2c loss specifically at the end of puberty, we also examined the mammary fat pads of 8-week old females and observed similar loss of side branching in the MMTVCre^{+}*Kmt2c*$^{F/F}$ mice (Fig. 2.12F, G).

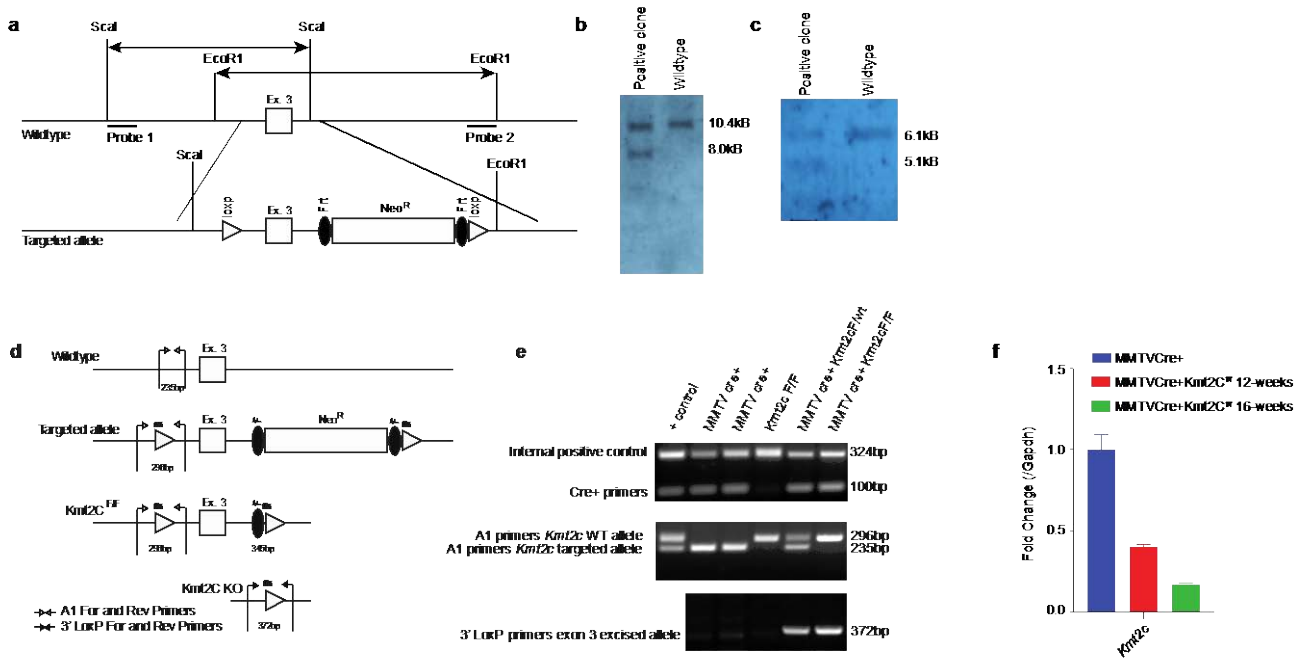

Figure 2.11 Generation of MMTVCre^{+}KMT2C$^{F/F}$ knockout mice. (A) Diagram indicating components of the *Kmt2c* wildtype and targeted allele and the location of ScaI and EcoRI restriction enzyme sites and Southern blot probe recognition sites. (B) Southern blot with probe 1 following digestion with ScaI. Wildtype band is 10.4kB and targeted allele is 8.0kB (C) Southern blot with probe 2 following digestion with EcoR1. Wildtype band is 6.1kB and targeted allele is 5.1kB. (D) Targeted and conditional disruption of the mouse *Kmt2c* gene, using the Cre-loxP recombination system. Components of the *Kmt2c* exon 3 WT allele, the targeted allele after homologous recombination in ES cells, the floxed allele (*Kmt2c*$^{F/F}$) after deletion of the neomycin resistance gene (pGK-Neo) in ES cells, and the deleted *Kmt2c* allele in Cre-recombinase transgenic mice (Kmt2cKO). (E) Genotyping mammary glands of virgin mice with the indicated genotypes with the positive control (+ control) being toe clips from an MMTVCre^{+}*Kmt2c*$^{F/F}$ mouse. (F) mRNA levels of indicated genes, as measured by qRT-PCR, from mammary glands of female mice of the indicated genotypes. Values correspond to the mean of three replicates ± s.e.m

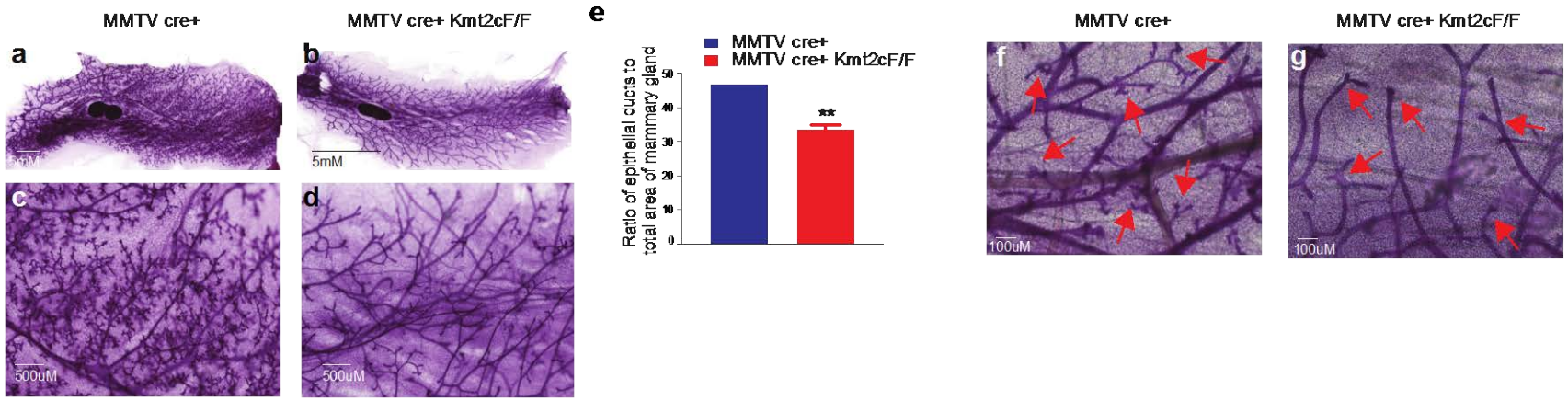

Figure 2.12 MMTVCre^{+}KMT2C$^{F/F}$ Knockout mice have defects in secondary epithelial duct branching. (A-D) Representative whole mounts of mammary glands from 12-week old, mature, virgin MMTVCre^{+} mouse (A, C), and MMTVCre^{+}*Kmt2c*$^{F/F}$ mouse (B, D). Images correspond to one representative mammary gland from a total of four MMTVCre^{+}*Kmt2c*$^{F/F}$ mice, and six MMTVCre+ control mice. (E) Quantification of surface area occupied by mammary epithelial ducts as compared to total surface area of the whole mount mammary epithelial gland. Values correspond to the mean from six MMTVCre+ control mice and four MMTVCre^{+}*Kmt2c*$^{F/F}$. (F,G) Representative whole mounts of mammary glands from 8-week old, mature, virgin MMTVCre^{+} mouse (F) and MMTVCre^{+}*Kmt2c*$^{f/f}$ mouse (G). Images correspond to one representative mammary gland from a total of two MMTVCre^{+}*Kmt2c*$^{f/f}$ mice and two MMTVCre+ control mice. Red arrows indicate terminal end buds.

KMT2C has no effects on pregnancy-related mammary gland development

In addition to being active during pubertal mammary development, ERα is also important for regulating pregnancy-related mammary development. In order to assess whether KMT2C has an effect on pregnancy-related mammary development, we next crossed our *Kmt2c* floxed mice to mice expressing Cre under the control of the WAP promoter. The WAP promoter is only expressed in alveolar epithelial cells during lactation (155) and can therefore be used to study pregnancy-related mammary gland development during the second pregnancy and beyond. We assessed the mammary glands on lactation day 4 from 4-5 month old females that had been pregnant at least once before and found no change in their morphology (Fig. 2.13A). Additionally, we weighed pups from WAPCre^{+}*Kmt2c*$^{F/F}$ and control mothers that had been pregnant at least once before and found no change in pup weight suggesting that lactation was occurring normally in the *Kmt2c* knockout mice (Fig. 2.13B). Together, this data suggests that Kmt2c does not regulate pregnancy related mammary gland development.

KMT2C does not affect ERα expression or stability

Given the effects of KMT2C loss on ERα-driven breast cancer proliferation and on pubertal mammary gland development, we hypothesized that KMT2C may be regulating ERα activity. In addition, our findings that KMT2C does not effect pregnancy-related mammary gland development suggested that perhaps KMT2C only regulates certain, specific aspects of ERα signaling. To assess

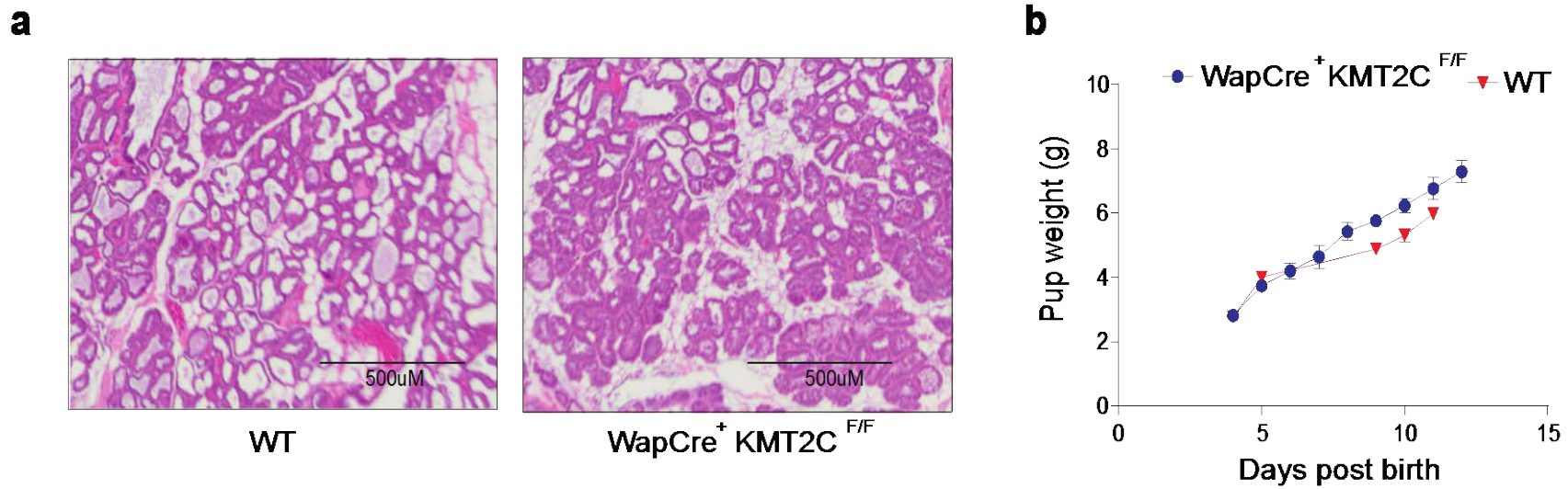

Figure 2.13 WapCre^{+}*Kmt2c*$^{f/f}$ knockout mice do not display defects in pregnancy-related mammary gland development. (A) H&E staining were used to analyze the morphology of the mammary glands at second lactation day 4 of the 4-5 month old WT controls and WAPCre+*Kmt2c*$^{F/F}$ females. Images correspond to one representative mammary gland from a total of 5 WT and 5 WAPCre+*Kmt2c*$^{f/f}$ mice. (B) Pups nursed by their WAPCre+*Kmt2c*$^{f/f}$ mothers (n = 5) were of similar size to WT pups nursed by their WT mothers (n = 2). Data are presented as the mean ± s.e.m.

whether and how KMT2C might be regulating ERα signaling, we first evaluated ERα expression, phosphorylation and conformational stability in the presence of KMT2C loss. We found no changes in the levels of total ERα or pERα (S118 or S104/106) or in the HSP90 dependence of ERα, as measured by sensitivity to the HSP90 inhibitor SNX-2112. (Fig. 2.14). Together, this suggests that KMT2C does not affect ERα expression or stability.

KMT2C loss suppresses ERα target gene expression in ER+HER2- cell lines

Seeing no effects on ERα protein expression or conformation, we next sought to determine whether there might be changes to KMT2C-dependent ERα-regulated gene expression. To this end, we performed RNA sequencing on shRenilla and shKMT2C MCF7 cell lines (Fig. 2.15A). We then compared the gene expression changes following KMT2C knockdown to those following five-day hormone deprivation of shRenilla cells. We verified that after 5 days in hormone-deprived media, well-known ERα target genes, such as *PGR*, *TFF1*, *GREB1*, and *XBP1* were downregulated suggesting that 5 days is sufficient to inhibit estrogen-dependent gene expression (Fig. 2.16). When comparing KMT2C loss to hormone deprivation, we found that genes that were upregulated following KMT2C knockdown were strongly enriched among genes that were upregulated following hormone deprivation (Fig 2.15B). A similar degree of enrichment was seen for genes that were downregulated following KMT2C knockdown among genes downregulated following hormone deprivation (Fig. 2.15C). Together this suggests that loss of KMT2C partially phenocopies loss of

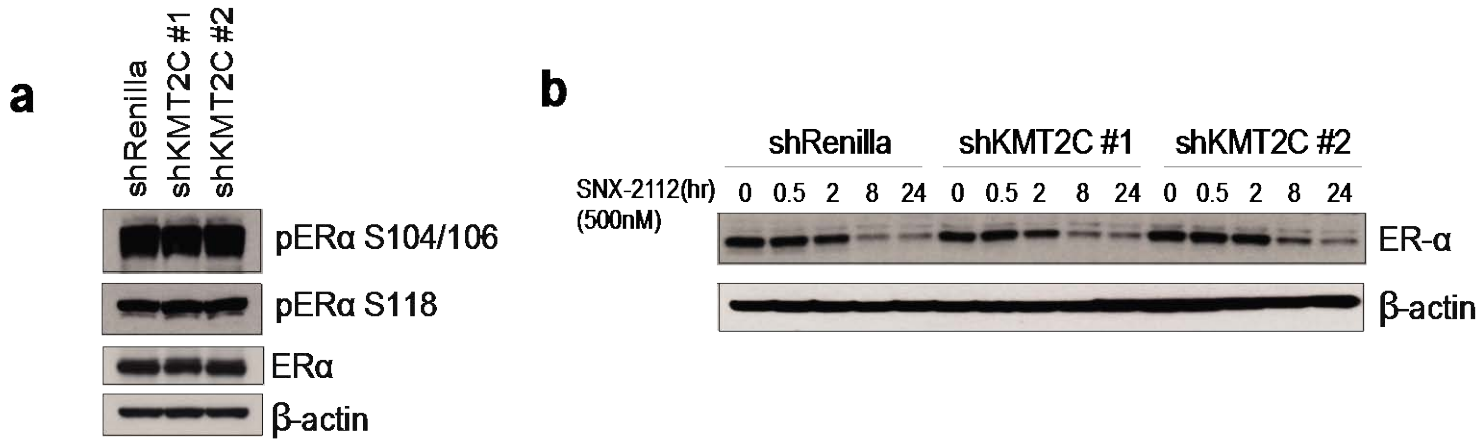

Figure 2.14 KMT2C loss does not affect ERα expression levels, phosphorylation or stability. (A) Immunoblot of the indicated proteins in shRenilla or shKMT2C MCF7 cells. β-actin was used as a loading control. (B) shRenilla, shKMT2C#1 or shKMT2C#2 MCF7 cells were treated with HSP90 inhibitor, SNX2112 (500nmol/L), and collected at indicated times. Immunoblots of the indicated proteins are shown. β-actin was used as a loading control.

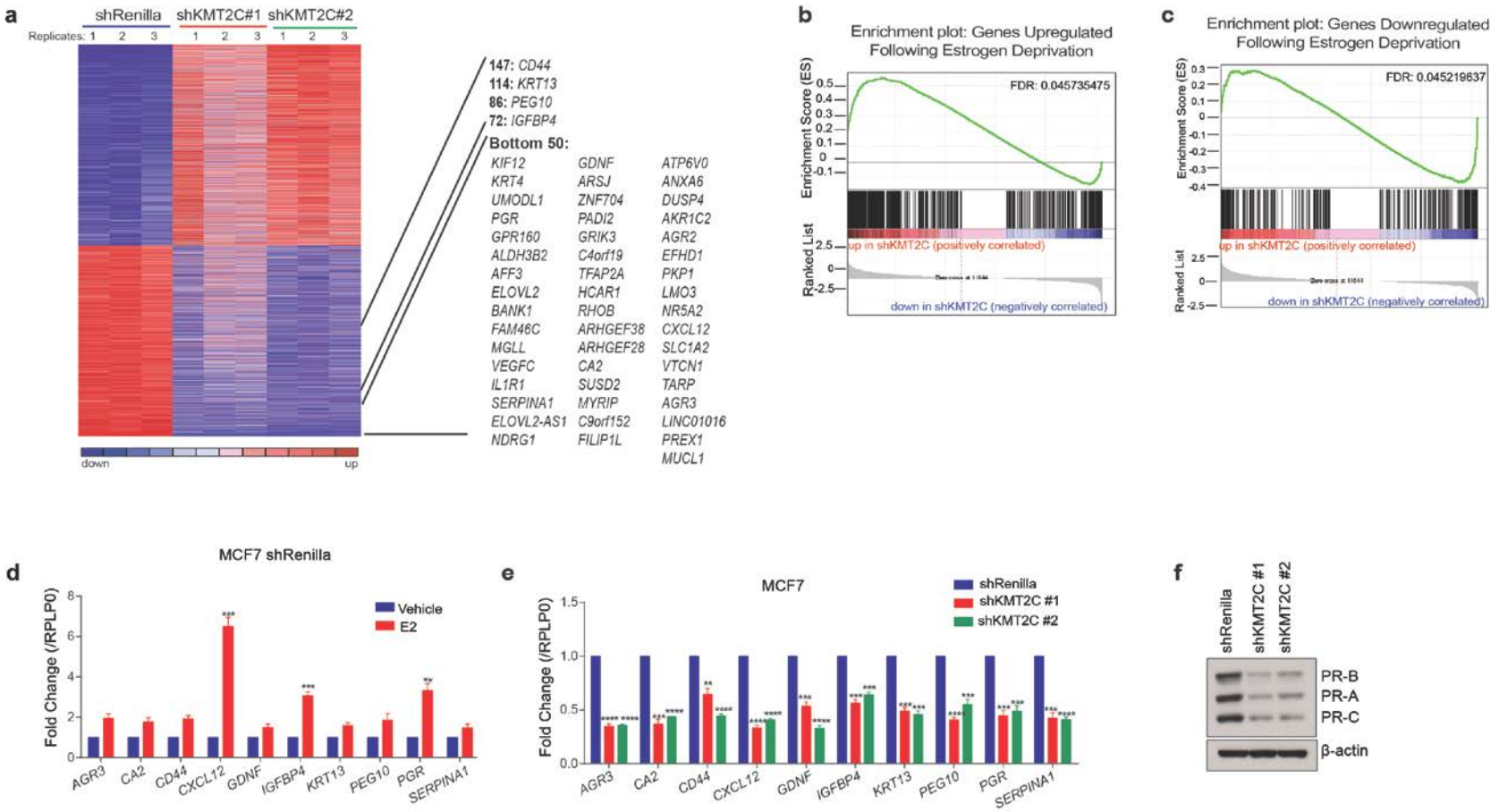

Figure 2.15 KMT2C loss suppresses estrogen-dependent gene expression in MCF7 cells. (A) Supervised analysis of the 7938 differentially expressed genes between MCF7 shRenilla and MCF7 shKMT2C (shKMT2C#1 and shKMT2C#5 samples combined). Columns represent individual replicates for shRenilla and shKMT2C samples, rows correspond to the

different genes. Color reflects the normalized expression count (number of reads mapping to each gene); scaled by row. (B) GSEA showing 4081 genes upregulated in shKMT2C cells are enriched among genes upregulated following 5 day estrogen deprivation of shRenilla cells (≥ 3-fold). Estrogen deprivation entailed culturing the cells in phenol-red free DMEM F12 media supplemented with 10% charcoal stripped serum. (C) GSEA showing 3857 genes downregulated in shKMT2C cells are enriched among genes downregulated following 5 day estrogen deprivation of shRenilla cells (≥ 3-fold). (D) mRNA levels, as measured by qRT-PCR, in MCF7 shRenilla cells in charcoal stripped media or in the presence of estradiol (E2). Values correspond to the mean of three replicates ± s.e.m. (E) mRNA levels, as measured by qRT-PCR, in MCF7 cells constitutively expressing either shRenilla, shKMT2C#1 or shKMT2C#2. Values correspond to the average of three replicates ± s.e.m.; two-tailed Student's *t*-test with a desired FDR = 1% was used to determine statistical significance; ***P* < 0.01, ****P* <0.001, *****P* <0.0001. Data correspond to one representative assay from a total of two or three independent assays. (F) Immunoblot of the indicated proteins in shRenilla or shKMT2C MCF7 cells. β-actin was used as a loading control.

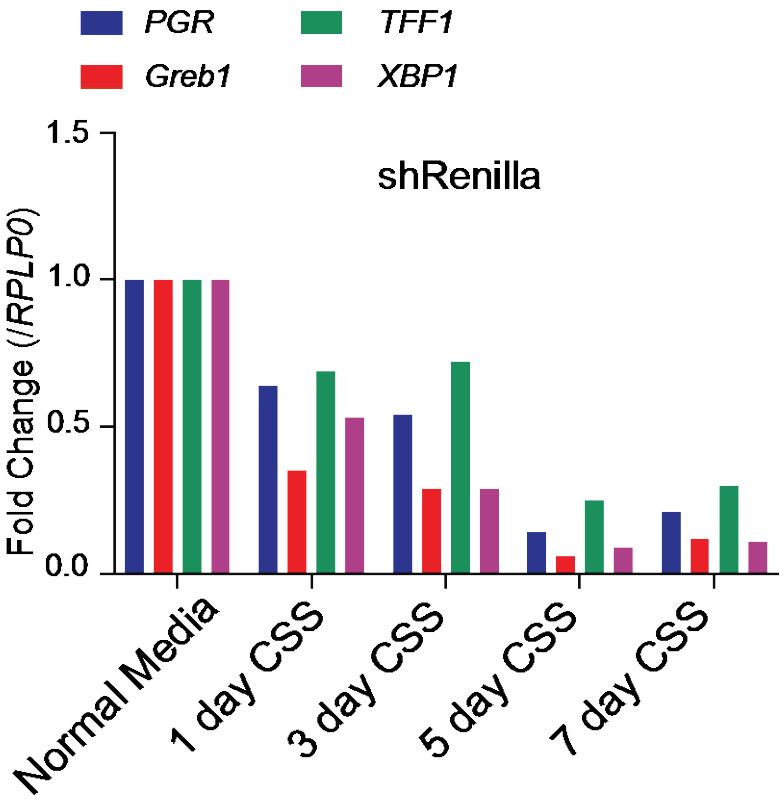

Figure 2.16 shRenilla MCF7 cells significantly reduce estrogen-dependent signaling after 5 days in charcoal-stripped serum. mRNA levels, as measured by qRT-PCR, in MCF7 cells constitutively expressing shRenilla, cultured in phenol-red free DMEMF12 media with 10% charcoal stripped fetal bovine serum (CSS) for indicated times. Values correspond to the average of three replicates ± s.e.m.

estrogen. We validated the differential gene expression data using qRT-PCR on a panel of ERα target genes whose expression was stimulated with E2 (Fig. 2.15D). Expression of all of these transcripts was reduced by 40-50% following KMT2C knockdown in MCF7 cells (Fig. 2.15E). To further confirm these findings, we examined protein expression of progesterone receptor (PR), a well-established ERα target, and found it to be decreased by approximately 70% following KMT2C loss (Fig. 2.15F).

In support of KMT2C loss affecting ERα target genes, we also verified that KMT2C was affecting the same target genes and pathway in other ER+HER2- cells. We performed RNA sequencing of T47D cells following KMT2C knockdown and again found genes downregulated by KMT2C loss to be highly enriched among ERα target gene sets (Fig. 2.17A-C). We then assessed the genes that were affected by KMT2C knockdown in MCF7 for KMT2C dependence in Cama-1 and T47D cells. We found that the majority of these genes were also downregulated by 10-60% in the absence of KMT2C in these other ER+HER2- models (Fig. 2.17D, E).

Kmt2c loss suppresses ERα target gene expression in the mammary gland

Next, we assayed for expression of ERα target genes in both control and *Kmt2c* knockout mammary glands to determine if KMT2C similarly affected ERα target gene expression in vivo. We found reduced expression of multiple known ERα target genes in the *Kmt2c* knockout mice to levels that were approximately 40-90% below levels seen in control mice (Fig. 2.18A). In addition, we assayed

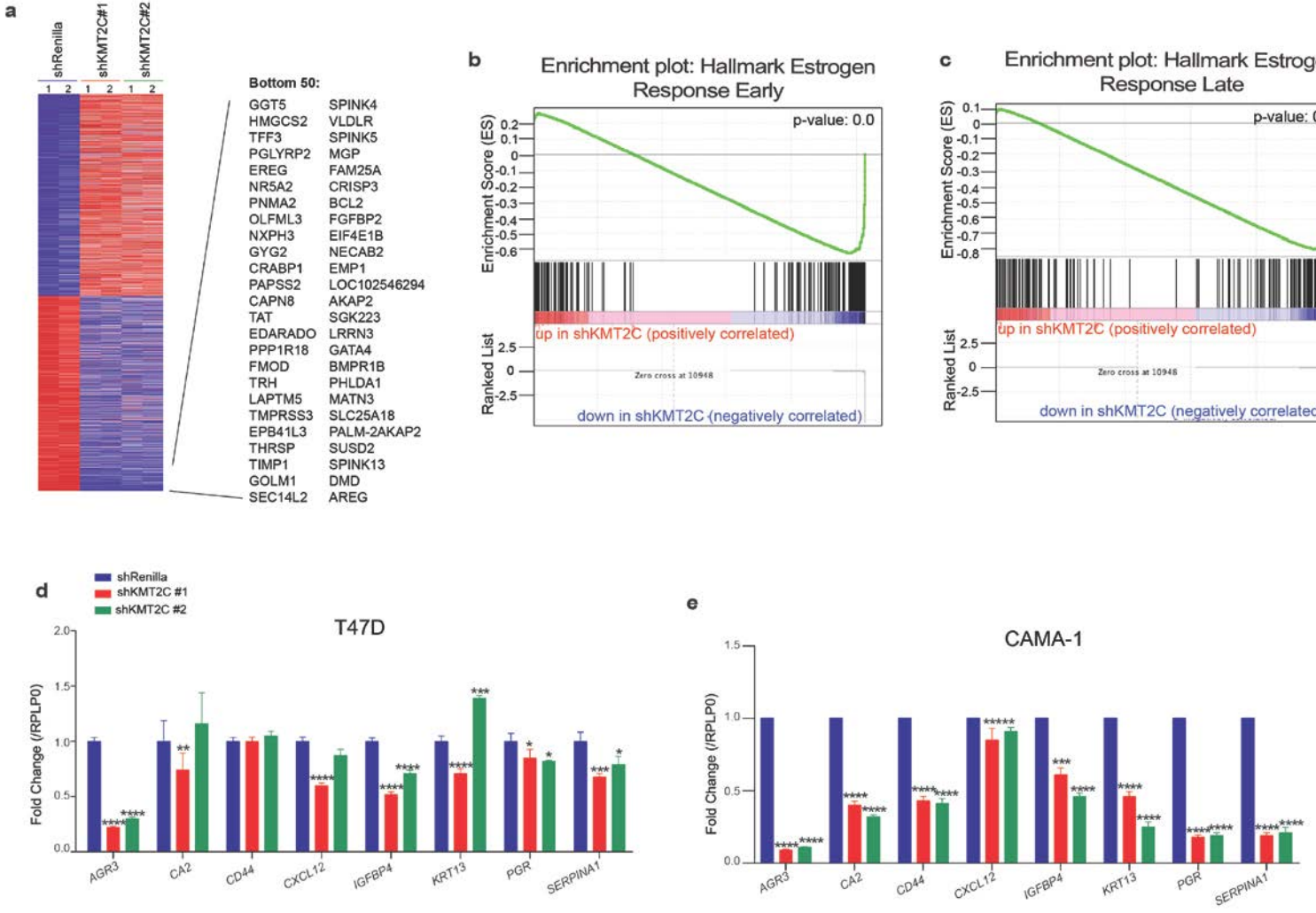
a
shRenilla
shKMT2C#1
shKMT2C#2
1 2 1 2 1 2
Bottom 50:
GGT5 SPINK4
HMGCS2 VLDLR
TFF3 SPINK5
PGLYRP2 MGP
EREG FAM25A
NR5A2 CRISP3
PNMA2 BCL2
OLFML3 FGFBP2
NXPH3 EIF4E1B
GYG2 NECAB2
CRABP1 EMP1
PAPSS2 LOC102546294
CAPN8 AKAP2
TAT SGK223
EDARADO LRRN3
PPP1R18 GATA4
FMOD BMPR1B
TRH PHLDA1
LAPTM5 MATN3
TMPRSS3 SLC25A18
EPB41L3 PALM-2AKAP2
THRSP SUSD2
TIMP1 SPINK13
GOLM1 DMD
SEC14L2 AREG
b
Enrichment plot: Hallmark Estrogen Response Early
p-value: 0.0
Enrichment Score (ES)
up in shKMT2C (positively correlated)
Ranked List
Zero cross at 10948
down in shKMT2C (negatively correlated)
c
Enrichment plot: Hallmark Estrogen Response Late
p-value: 0.0
Enrichment Score (ES)
up in shKMT2C (positively correlated)
Ranked List
Zero cross at 10948
down in shKMT2C (negatively correlated)
d
shRenilla
shKMT2C #1
shKMT2C #2
T47D
Fold Change (/RPLP0)
AGR3 CA2 CD44 CXCL12 IGFBP4 KRT13 PGR SERPINA1
e
CAMA-1
Fold Change (/RPLP0)
AGR3 CA2 CD44 CXCL12 IGFBP4 KRT13 PGR SERPINA1

Figure 2.17 KMT2C loss suppresses estrogen-dependent gene expression in T47D and Cama-1 cells. (A) Supervised analysis of the 4349 differentially expressed genes between T47D shRenilla and T47D shKMT2C (shKMT2C#1 and shKMT2C#5 samples combined). Columns represent individual replicates for shRenilla and shKMT2C samples, rows correspond to the different genes. Color reflects the normalized expression count (number of reads mapping to each gene); scaled by row. (B) GSEA of 2137 genes downregulated in T47D shKMT2C vs. shRenilla as compared to ranked genes from the Hallmark Estrogen Response Early Geneset (from the Broad Institute Molecular Signature Database) (C) GSEA of 2137 genes downregulated in T47D shKMT2C vs shRenilla as compared to ranked genes from the Hallmark Estrogen Response Late Geneset (from the Broad Institute Molecular Signature Database). (D) mRNA levels, as measured by qRT-PCR, in T47D cells constitutively expressing either shRenilla, shKMT2C#1 or shKMT2C#2. Values correspond to the average of three replicates ± s.e.m.; two-tailed Student's *t*-test with a desired FDR = 1% was used to determine statistical significance; **P* < 0.05, ***P* < 0.01, ****P* <0.001, *****P* <0.0001. (E) mRNA levels, as measured by qRT-PCR, in Cama-1 cells constitutively expressing either shRenilla, shKMT2C#1 or shKMT2C#2. Values correspond to the average of three replicates ± s.e.m.; two-tailed Student's *t*-test with a desired FDR = 1% was used to determine statistical significance; ***P* < 0.01, ****P* <0.001, *****P* <0.0001. Data correspond to one representative assay from a total of two or three independent assays.

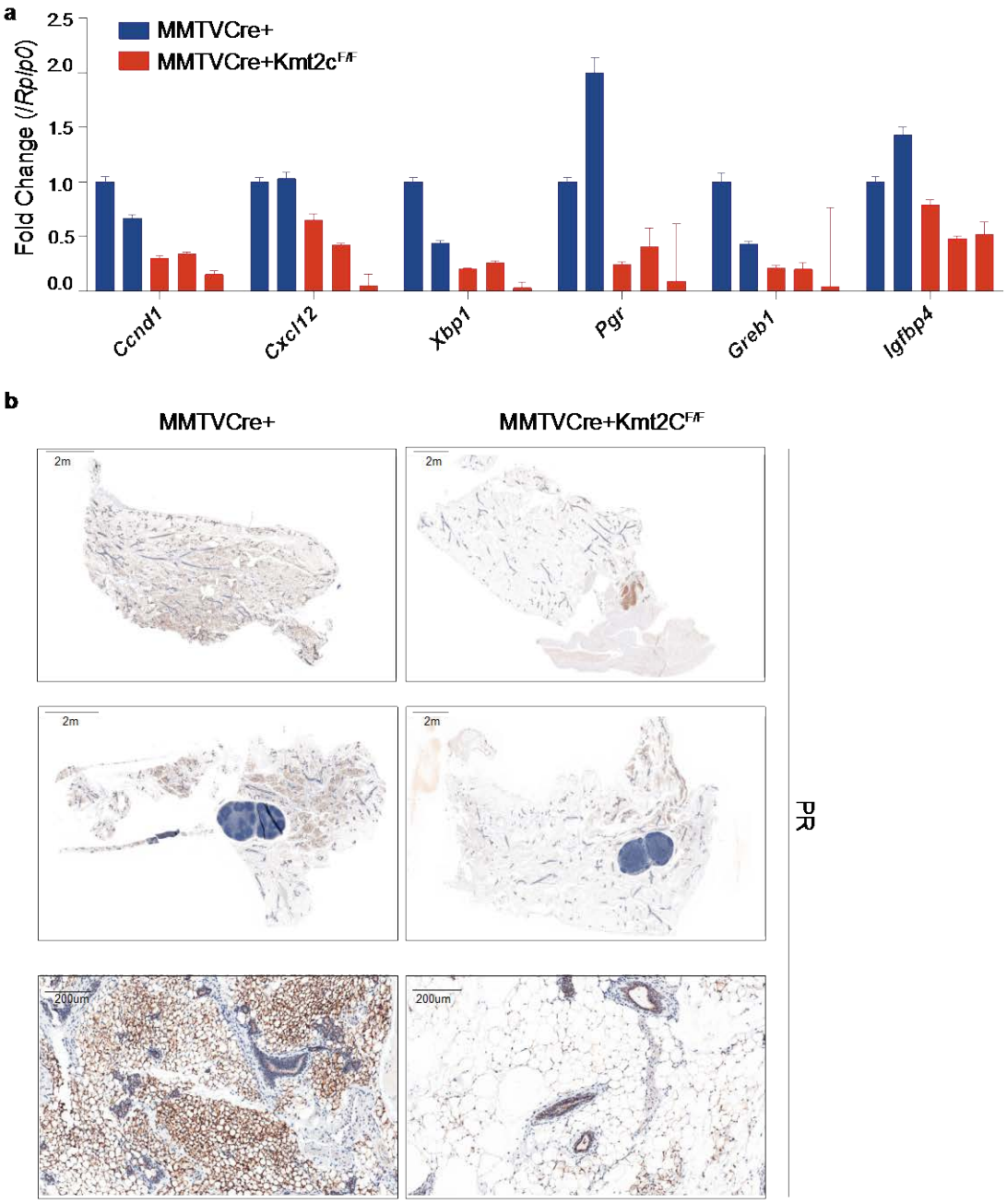

Figure 2.18 Kmt2c loss suppresses ERα target gene expression in the mammary gland. (A) mRNA levels, as measured by qRT-PCR, from RNA taken from the mammary glands of 12-week old virgin females. Values correspond to the mean of three replicates ± s.e.m. (B) Representative slides of PR expression in mammary glands from 12-wk old virgin females.

for expression of PR by IHC and found significant reductions in PR protein expression across the entire mammary gland (Fig. 2.18B). Together, our in vitro and in vivo data confirm that KMT2C regulates key aspects of the estrogen response program while not globally controlling ERα expression or stability.

KMT2C may more generally regulate other hormone-dependent nuclear receptors

Given the effects of KMT2C loss on ERα signaling, we wanted to assess whether KMT2C similarly affects other hormone-driven nuclear receptors. The androgen receptor (AR) is another such nuclear receptor that is expressed in MCF7 cells. Regulation of AR has been shown to have many parallels to regulation of ERα, especially in terms of their dependence on pioneer factors and H3K4me(184) We tested the expression of AR target genes in our KMT2C knockdown cells and found that the expression of *KLK2*, *TMPRSS2*, *FKBP5*, *PRLR* were reduced by approximately 15-60% in MCF7 cells (Fig. 2.19). This suggests that perhaps KMT2C can regulate the expression downstream of other nuclear receptors in breast cancer cells and beyond, which has been suggested by other groups in the context of LXR, GR, FXR and PPARϒ (54, 83, 84, 100, 103, 179).

KMT2C does not regulate candidate HOX gene expression in MCF7 cells

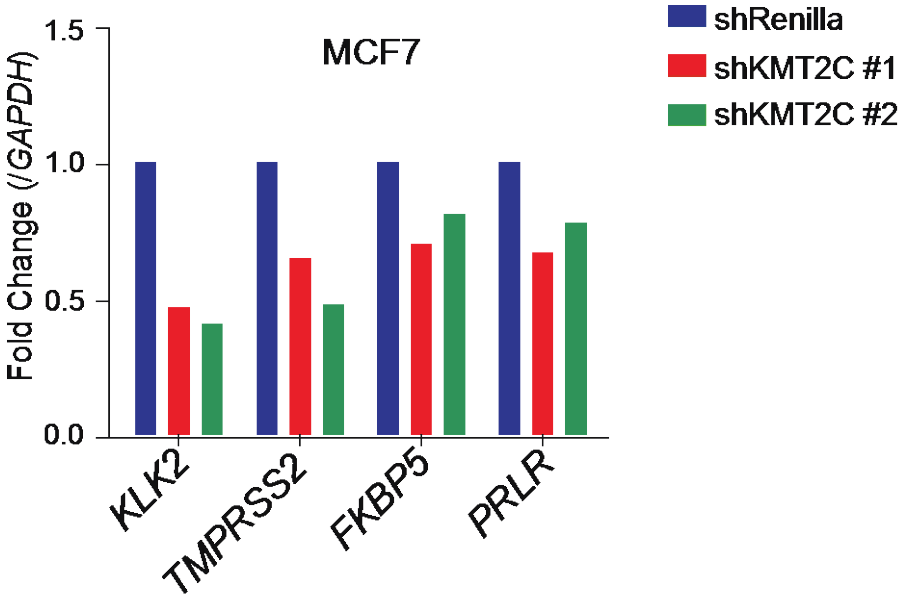

Figure 2.19 KMT2C loss moderately suppresses select AR target genes expressed in breast cancer cells. mRNA levels, as measured by qRT-PCR, in MCF7 cells constitutively expressing either shRenilla, shKMT2C#1 or shKMT2C#2. Values correspond to the average of three replicates ± s.e.m.

Previous studies have suggested that KMT2C can regulate the expression of certain HOX genes through ERα and other nuclear receptors(84, 103). Therefore, we wanted to specifically assess if candidate HOX gene expression was affected in our KMT2C knockdown cells. In particular, *HOXA4*, *HOXB7*, *HOXB8* and *HOXC10* have previously been implicated to be regulated by KMT2C(84, 102, 103) and are also expressed in MCF7 cells. Upon measuring their gene expression levels, we found no significant change in our shKMT2C cells as compared to our shRenilla cells (Fig. 2.20). This suggests that in the context of ER+HER2- breast cancer, KMT2C may not be a regulator of HOX gene expression and may more specifically be a regulator of hormone-driven gene expression such as that downstream of AR or ERα.

KMT2 family members do not regulate the same genes as KMT2C

Given the close homology of KMT2C with the other KMT2 family members and given that all KMT2 family members inhibited MCF7 proliferation to varying degrees, we next asked if the ability of KTM2C to regulate candidate ERα target gene expression was shared among the other KMT2 family members. Upon testing our panel of E2-stimulated target genes in shKMT2A, shKMT2B and shKMT2D cell lines, we found that KMT2A loss downregulated the expression of many KMT2C target genes with very modest effects on *AGR3*, *CD44* and *IGFBP4*; however, it also upregulated the expression of *SERPINA1*, *KRT13* and *GDNF*. The effects of KMT2B loss were the least similar to that of KMT2C. Loss of KMT2B resulted in upregulation of the majority of KMT2C target genes, with

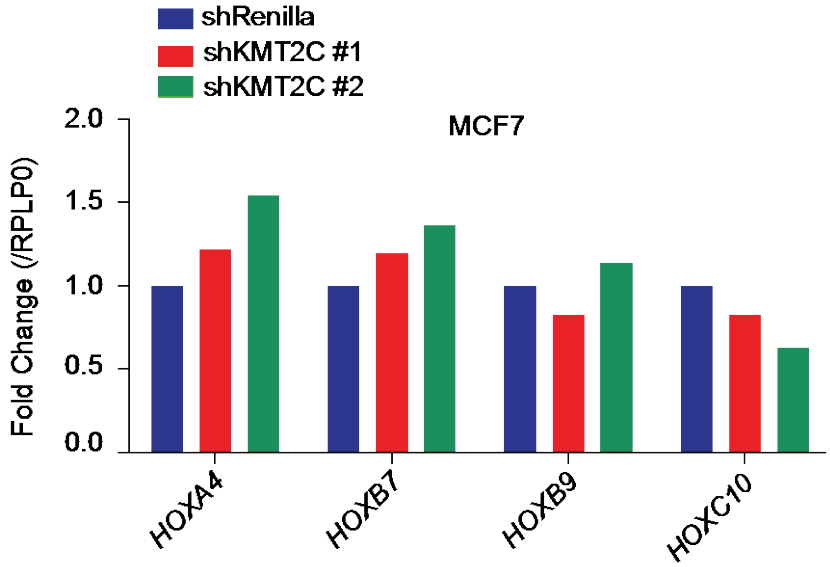

Figure 2.20 KMT2C loss does not affect candidate HOX gene expression. mRNA levels, as measured by qRT-PCR, in MCF7 cells constitutively expressing either shRenilla, shKMT2C#1 or shKMT2C#2. Values correspond to the average of three replicates ± s.e.m.

only modest reductions in *CA2*, *CD44* and *CXCL12*. Strikingly, the effects of KMT2D loss were the most similar to that of KMT2C loss resulting in significant downregulation of most of the same target genes with over 70% reduction in *AGR3*, *CA2*, *IGFBP4* and *PEG10*. However, KMT2D also upregulated *GDNF* and *CD44* but only by 10-20% (Fig. 2.21). This similarity in gene expression regulation by KMT2C and KMT2D is especially intriguing as KMT2D is thought to have certain functional redundancies with KTM2C (83).

KMT2C loss suppresses ERα function at enhancer sites

It has been speculated that H3K4me may help demarcate regions for ERα binding and activity (166). Given the known enzymatic activity of KMT2C in generating H3K4me(96), we hypothesized that KMT2C may regulate ERα activity by placing these marks at ERα binding sites. This would provide a mechanistic explanation for how KMT2C is able to effect ERα-dependent gene expression without directly effecting ERα expression or stability. We found no change in total H3K4me1 and H3K4me3 levels following KMT2C loss, implying that KMT2C is not the sole methyltransferase responsible for these marks in MCF7 cells (Fig. 2.22A). We next examined the pattern of genome-wide H3K4me1 and H3K4me3 following KMT2C loss using ChIP-sequencing. Whereas KMT2C loss showed no meaningful effect upon H3K4me3 across the genome, it appeared to selectively alter H3K4me1 at 869 affected loci, or 1% of the total 84,484 sites of H3K4me1 occupancy in control cells (Fig. 2.22B). This is in accordance with KMT2C being predominantly an H3K4 monomethyltransferase.

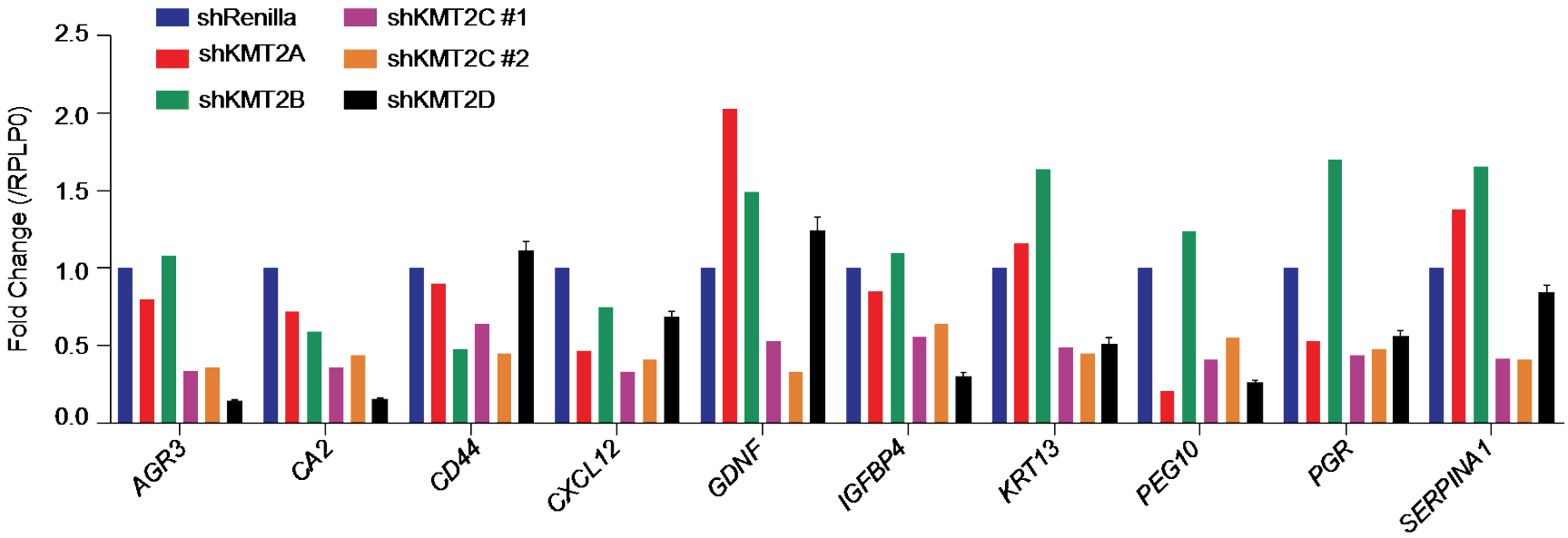

Figure 2.21 KMT2 family members have differential effects on KMT2C-regulated target genes. mRNA levels, as measured by qRT-PCR, in MCF7 cells constitutively expressing either shRenilla, shKMT2A, shKMT2B, shKMT2C or shKMT2D. Values correspond to the average of three replicates ± s.e.m.

To analyze the functional significance of H3K4me1 loss, we integrated the H3K4me1 ChIP-sequencing data with the RNA sequencing data. We found a strong enrichment for sites with H3K4me1 loss among sites whose nearby genes showed diminished expression following KMT2C knockdown (Fig. 2.22C). These genes have previously been shown to be predominantly ERα target genes. Together this suggests that H3K4me1 is predominantly decreased at enhancers of ERα target genes (Fig. 2.15). To confirm these effects, we examined H3K4me1 at known ERα enhancer sites and found decreases in H3K4me1 from 15% at the *PGR* enhancer to 80% at the *CA2* enhancer with both ChIP-sequencing as well as ChIP-PCR (Fig. 2.22D, E).

We then went on to confirm that these sites of decreased H3K4me1 occupancy were direct targets of KMT2C by comparing our H3K4me1 ChIP-sequencing data to recently published KMT2C ChIP sequencing data (107). We found that sites with decreased H3K4me1 occupancy directly overlapped with KMT2C binding sites (Fig. 2.22D), suggesting that they were all indeed direct KMT2C targets. Given the prior data that ERα regulates gene expression through its actions at promoter and enhancer sites, our integrated H3K4 methylation and RNA sequencing data points to KMT2C selectively altering ERα-dependent gene expression via its effects upon ERα enhancer sites.

Sites of H3K4me1 loss in KMT2C knockdown cells are additionally characterized by loss of H3K27ac

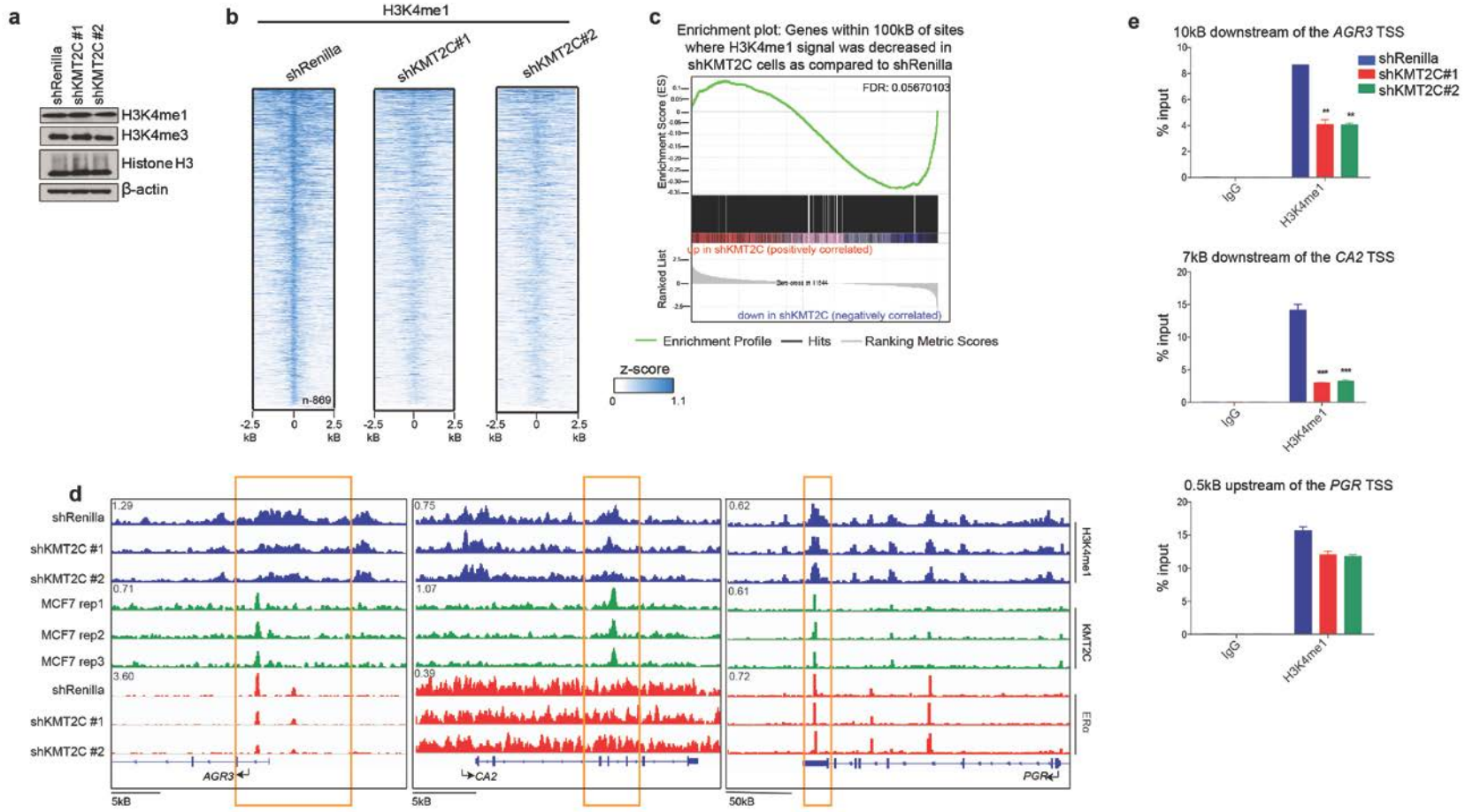

Figure 2.22 KMT2C loss results in reduced H3K4me1 occupancy at ERα target gene enhancers. (A) Immunoblot of the indicated proteins in shRenilla or shKMT2C MCF7 cells. β-actin was used as a loading control. (B) Normalized heatmaps for H3K4me1 occupancy in shRenilla, shKMT2C#1 and shKMT2C#2 cells among the 869 differential sites. Heatmaps are centered at the peak summit. (C) GSEA of 3857 genes significantly downregulated in MCF7 shKMT2C

cells (p-val < 0.05) as compared to ranked genes with a ≥ 25% reduction in H3K4me1 marks at enhancers in shKMT2C MCF7 cells. (D) IGV browser views for H3K4me1 in shRenilla, shKMT2C#1 and shKMT2C#2 MCF7 cells, KMT2C in MCF7 cells(107), and ERα in shRenilla, shKMT2C#1 and shKMT2C#2 MCF7 cells at three representative loci. The y-axis corresponds to the ChIP-seq signal intensity. Outlined in orange are sites of H3K4me1 loss. (E) qChIP analysis for H3K4me1 occupancy loss in enhancer regions of specified ERα target genes after KMT2C knockdown in MCF7 cells. IgG was used as a negative control. Values correspond to mean percentage of input enrichment ± s.e.m. of triplicate qPCR reactions of a single replicate. Two-tailed Student's *t*-test with a desired FDR = 1% was used to determine statistical significance; ***P* < 0.01, ****P* <0.001, *****P* <0.0001. Data correspond to one representative assay from a total of two or three independent assays.

As the H3K4me1 mark is known to demarcate both regions of active transcription and regions 'poised' for transcription, histone H3K27 acetylation (H3K27ac) can be used to more accurately delineate sites that are actively engaged in transcription (185). To ensure that KMT2C was regulating active enhancers, we assayed for H3K27ac at the *AGR3*, *CA2*, and *PGR* enhancer sites where we saw loss of H3K4me1 following KMT2C knockdown. We found significant reduction of H3K27ac at these sites (Fig. 2.23A). We went on to perform H3K27ac ChIP-sequencing on KMT2C knockdown MCF7 cells and found that H3K27ac was coordinately reduced across sites of H3K4me1 loss (Fig. 2.23B, C). These sites marked by loss of both H3K4me1 and H3K27ac subsequently correlated with reduced gene expression of their surrounding genes (Fig. 2.23D). Together, this establishes a model where KMT2C specifically functions at ERα enhancer regions that are actively engaged in transcription.

KMT2C loss results in decreased RNA Polymerase II occupancy at the PGR promoter

After having established that KMT2C loss results in inactivation of ERα enhancers, we wanted to see if there was any direct result on the promoters of the genes whose enhancer regions were affected. We assayed for PolII occupancy as a surrogate for promoter activity at the *PGR* promoter and found reduced PolII occupancy KMT2C knockdown cells (Fig. 2.24). This supports a hypothesis where loss of H3K4me1 and H3K27ac at *PGR* enhancer regions can directly affect the activity at the *PGR* promoter.

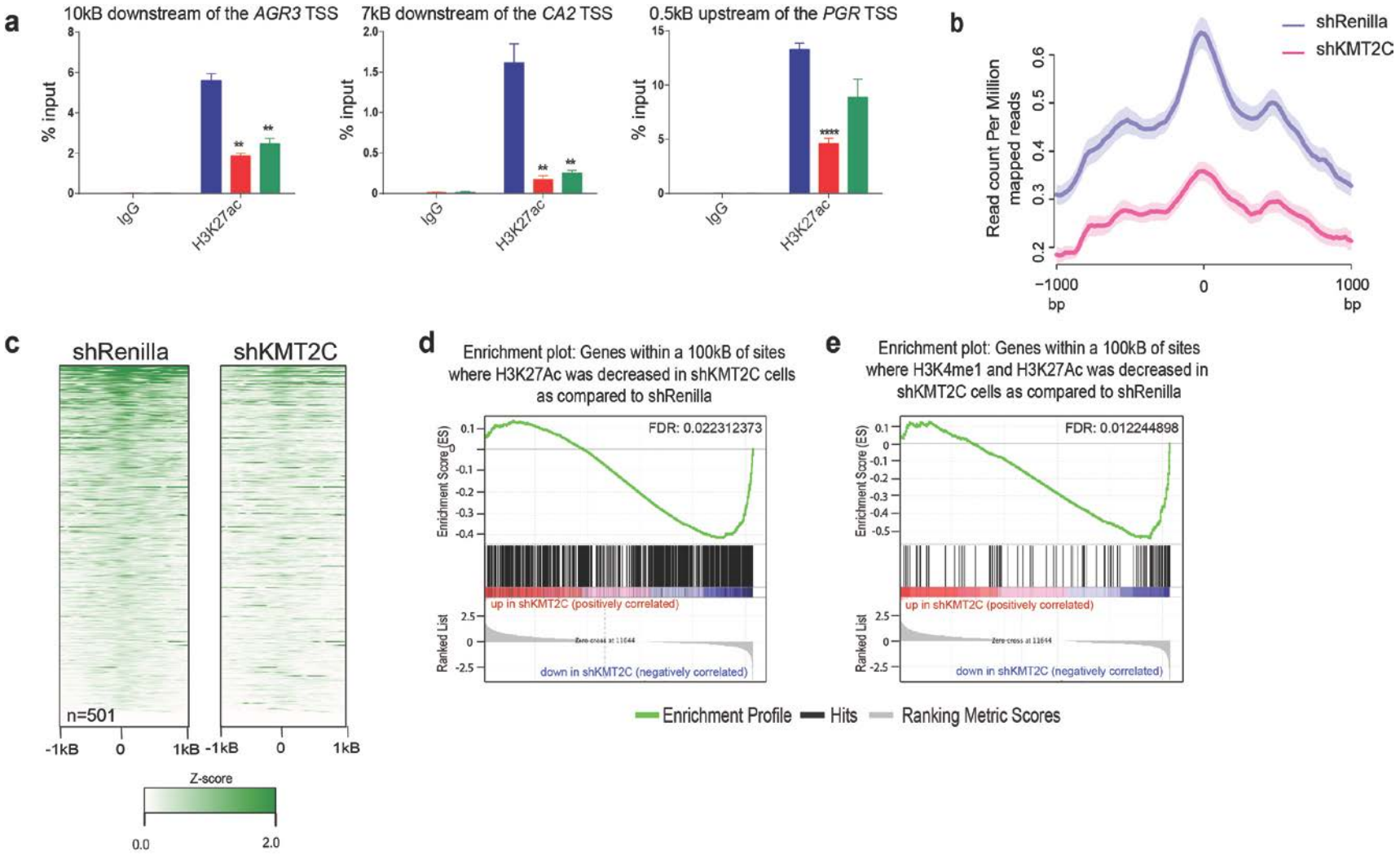
a
10kB downstream of the AGR3 TSS
7kB downstream of the CA2 TSS
0.5kB upstream of the PGR TSS
% input
IgG
H3K27ac
b
shRenilla
shKMT2C
Read count Per Million mapped reads
−1000 bp
0
1000 bp
c
shRenilla
shKMT2C
n=501
-1kB
0
1kB
Z-score
0.0
2.0
d
Enrichment plot: Genes within a 100kB of sites where H3K27Ac was decreased in shKMT2C cells as compared to shRenilla
FDR: 0.022312373
Enrichment Score (ES)
up in shKMT2C (positively correlated)
Ranked List
down in shKMT2C (negatively correlated)
e
Enrichment plot: Genes within a 100kB of sites where H3K4me1 and H3K27Ac was decreased in shKMT2C cells as compared to shRenilla
FDR: 0.012244898
Enrichment Profile
Hits
Ranking Metric Scores

Figure 2.23 Sites of H3K4me1 loss in KMT2C knockdown cells are additionally characterized by loss of H3K27ac. (A) qChIP analysis for H3K27ac occupancy loss in enhancer regions of specified ERα target genes after KMT2C knockdown in MCF7 cells. IgG was used as a negative control. Values correspond to mean percentage of input enrichment ± s.e.m. of triplicate qPCR reactions of a single replicate. Two-tailed Student's *t*-test with a desired FDR = 1% was used to determine statistical significance; ***P* < 0.01, ****P* <0.001, *****P* <0.0001. Data correspond to one representative assay from a total of two or three independent assays. (B) The average binding of H3K27ac in shRenilla and shKMT2C MCF7 cells at sites of H3K4me1 loss at a region centered at the peaks and showing ±1kB. shKMT2C#1 and shKMT2C#2 H3K27ac peaks were averaged to generate the shKMT2C profile. (C) Normalized heatmaps for H3K27ac occupancy at sites of H3K4me1 loss in shRenilla, shKMT2C#1 and shKMT2C#2 cells among 501 sites. Heatmaps are centered at the H3K27ac peak summit. (D) GSEA of 3857 genes significantly downregulated in MCF7 shKMT2C cells (p-val < 0.05) as compared to ranked genes with a ≥ 25% reduction in H3K27ac marks at enhancers in shKMT2C MCF7 cells. (E) GSEA of 3857 genes significantly downregulated in MCF7 shKMT2C cells (p-val < 0.05) as compared to ranked genes with a ≥ 25% reduction in both H3K27ac and H3K4me1 marks at enhancers in shKMT2C MCF7 cells.

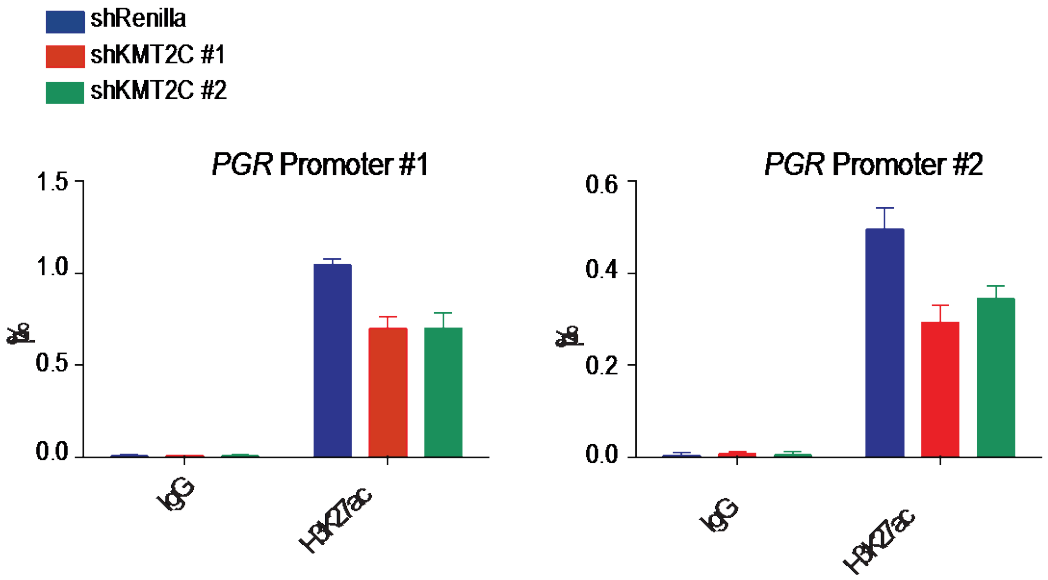

Figure 2.24 KMT2C loss results in decreased RNA Polymerase II occupancy at the *PGR* promoter. qChIP analysis for RNA PolII occupancy loss in promoter regions of *PGR* after KMT2C knockdown in MCF7 cells. IgG was used as a negative control. Values correspond to mean percentage of input enrichment ± s.e.m. of triplicate qPCR reactions of a single replicate.

KMT2C loss does not alter ERα DNA binding

H3K4me1 may facilitate enhancer function by promoting binding of any of a number of components necessary for site-specific transcription. It has been speculated that DNA-binding of ERα itself may require H3K4 methylation and pioneer factor binding (166, 167, 186). Therefore, we wanted to test the hypothesis that loss of H3K4me would results in loss of ERα binding to DNA explaining the decrease in ERα transcriptional activity. We examined ERα ChIP-sequencing in KMT2C knockdown cells. Unlike the case with H3K27ac, we did not observe major loss of ERα binding at sites accompanied by changes in gene expression (Fig. 2.25A). Not surprisingly, we did find a handful of sites with decreased ERα binding (Fig 2.23D, left) and concomitant decreases in proximal gene expression (Fig. 2.25B). However, the 12 genes accounted for by loss of ERα binding were only 0.3% of the total gene expression changes observed. Moreover, we observed no change in levels of ERα at >75% of ERα binding sites proximal to sites where we found loss of H3K4me1 in the KMT2C knockdown cells (Fig. S2.25A, Fig 2.23D middle, right).

KMT2C loss may prevent recruitment of ERα coactivators

The data thus far points to KMT2C regulating ERα enhancer function perhaps via its facilitation of essential ERα coactivators, but not via recruitment of ERα itself. In order to nominate potential coactivators involved in mediating the effects of KMT2C upon ERα, we first performed *de novo* motif analysis on a 1kB radius around sites where H3K4me1 had decreased in shKMT2C cells. We found

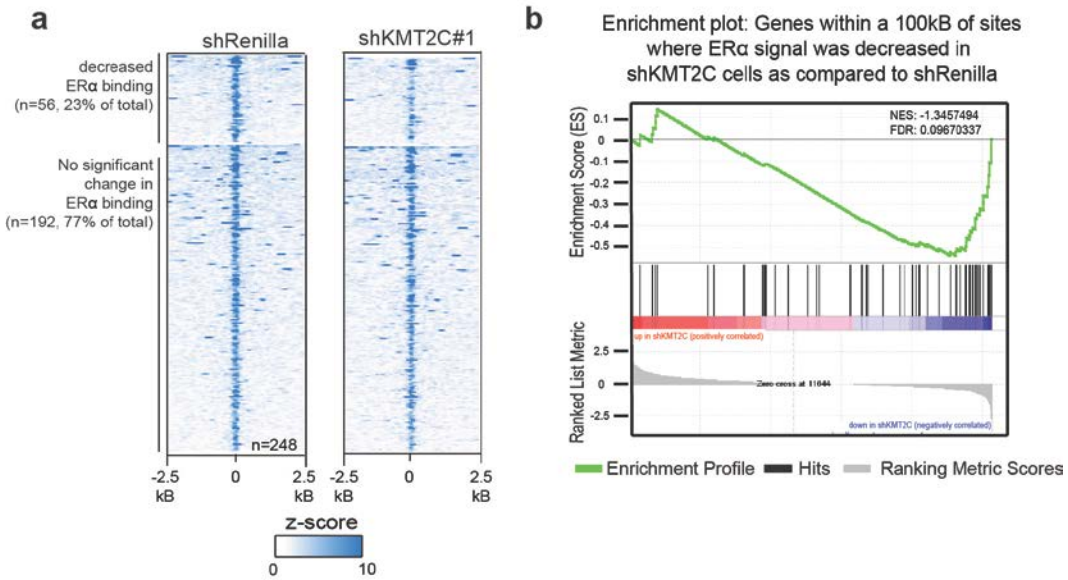

Figure 2.25 KMT2C loss does not alter ERα DNA binding. (A) ERα ChIP Sequencing was performed on MCF7 cells cultured in full serum. Normalized heatmaps for ERα occupancy at sites with decreased H3K4me1 in shKMT2C#1 cells as compared to shRenilla cells. Heatmaps are centered at the peak summit. (B) GSEA of 3857 genes significantly downregulated in MCF7 shKMT2C cells (p-val < 0.05) as compared to ranked genes with a ≥ 25% reduction in ERα marks at enhancers in shKMT2C MCF7 cells.

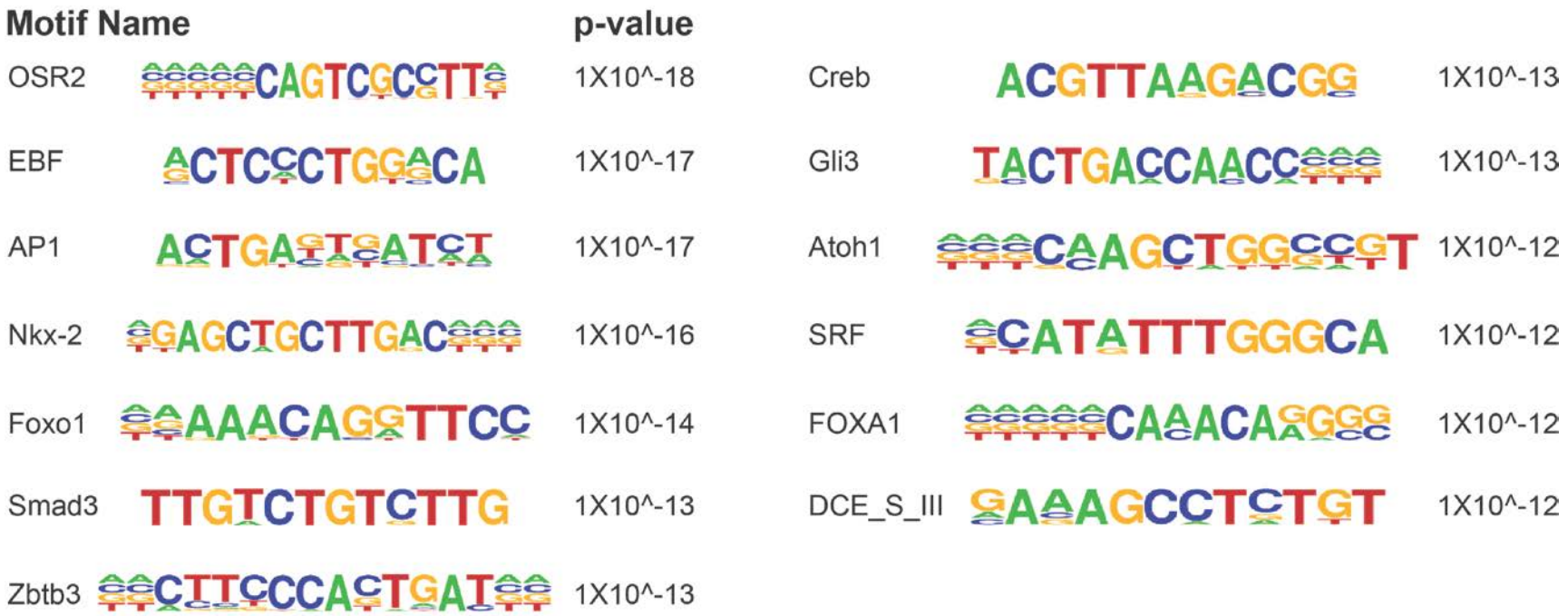

Figure 2.26 Sites of H3K4me1 loss are enriched for ERα coactivator motifs. Motif analysis, using HOMER, of a 1kB radius around the 869 loci where H3K4me1 enrichment was reduced in shKMT2C#1 and shKMT2C#2 MCF7 cells.

that these regions were significantly enriched for the well-known ERα pioneer factor, FOXA1(166) as well as the ERα coactivator, AP-1(187) (Fig. 2.26). While a number of motifs were considered to be significant in this analysis, it is notable that these two known factors necessary for ERα activity.

To further screen for potential intermediaries between KMT2C and ERα activity, we looked for changes in the ERα interactome following KMT2C loss. Given the large number of potential interactions, we sought to restrict this search to interactions occurring on DNA. To this end, we employed Rapid Immunoprecipitation Mass Spectrometry of Endogenous Proteins (RIME)(109, 188) with shKMT2C#2 MCF7 cells cultured in heavy SILAC (stable isotope labeling with amino acids in cell culture) media and shRenilla MCF7 cells cultured in light SILAC media. We compared our list of peptides differentially bound to ERα with previously published ERα binding partners(108). These data revealed only modest changes with most interactions. However, they did reveal reduced binding of ERα to MCM4, FOXA1, and TRPS1 in shKMT2C cells (Table 2.2). These data begin to establish a model where ERα coactivators mediate the functional interaction between KMT2C and ERα at enhancer sites. Furthermore, the motif and interactome analyses both point to a possible role for regulation of FOXA1, a well-established pioneer factor essential for ERα activity.

Conclusions

Table 2.2 Differential ERα binding partners in the absence of KMT2C

Protein Name	H/L Ratio	p-value
MCM4	0.51522	0
FOXA1	0.57893	0
TRPS1	0.62315	0

* H/L refers to the ratio of peptides found in the heavy (shKMT2C) vs the light (shRenilla) labeled lysates. List was restricted to previously established ERα binding partners(108) with a H/L ratio of 0.666. Assay done in triplicate

Here, we uncovered that the H3K4 methyltransferase KMT2C is a critical regulator of hormone-dependent ERα activity (Fig. 2.27). Although it is unlikely that KMT2C will prove to be the only histone modifier that regulates ERα function, the findings in this report suggest it to be a key regulator of the estrogen responsive phenotypes such as proliferation and mammary gland development.

We first suspected that KMT2C may have a role in estrogen dependence from its frequent mutation in primary breast cancer. As most mutations appear to be loss of function mutations, we modeled KMT2C loss in breast cancer cell lines that were representative of the variety of breast cancer subtypes. Given that the KMT2C mutation spectrum strongly suggests KMT2C to be a tumor suppressor, we were initially surprised to find that KMT2C loss led to a selective proliferative defect in ER+ driven breast cancers while having no effect on the proliferation of other models tested. As previous studies have reported interactions between H3K4me and nuclear receptor activation (82-84, 91, 96, 99, 100, 102, 103, 105, 179, 189), we hypothesized that KMT2C might be serving as a regulator of ERα activity. Moreover, we find that while KMT2C seems to selectively effect the proliferation of ER+HER2- cells, other members of the KMT2 family differentially effect the proliferation across multiple breast cancer subtypes. This suggests that the other members of the KMT2 family may have effects on proliferation that are not restricted to effects on the ERα signaling pathway.

To examine the mechanisms of KMT2C-mediated regulation of ER+ breast cancer proliferation, we analyzed the effect of its loss upon established gene expression signatures of ligand-activated ERα. We found that KMT2C loss

downregulated estrogen responsive gene expression. Additionally, mammary-specific inactivation of Kmt2c resulted in diminished side branching during mammary development. Interestingly, while MMTV-driven deletion of Kmt2c dramatically affected secondary epithelial duct branching, WAP-driven deletion of Kmt2c did not seem to have any effects on pregnancy-related mammary development. Upon comparing these phenotypes to MMTV- and WAP-driven deletion of *Esr1* we find that MMTV-mediated deletion of *Esr1* results in diminished secondary epithelial duct branching, much like what is seen after loss of Kmt2c. However, WAPcre mediated deletion of *Esr1* resulted in dilated alveolar cells, and pups that were born from WAPcre+*Esr1*$^{F/F}$ mothers had significantly reduced weight. The phenotypes we see from our mammary-specific Kmt2c deletion models suggest that Kmt2c might affect Esr1-regulated pubertal mammary development but not Esr1-regulated pregnancy-related mammary development. The idea that KMT2C may only regulate a subset of ERα-driven phenotypes is in accordance with the gene expression data where we see that loss of KMT2C downregulates a specific subset of ERα-target genes. It will now be imperative to determine when KMT2C is expressed along different stages of mammary development to see if that correlates with the stages where is seems to affect development.

As we were interested in finding the mechanism for KTM2C regulation of ERα activity, we noted that early analyses of KMT2C biochemical activities and KMT2C knockout cells pointed to an essential role for this protein in generating the H3K4me1 mark associated more frequently with enhancer sites(96) In

accordance, we found H3K4me1, but not H3K4me3, to be lost specifically at ERα enhancer sites in KMT2C knockdown cells. Moreover, ChIP-sequencing of H3K27ac demonstrated that H3K4me1 loss coincided with sites of H3K27ac loss, implying that the sites affected by KMT2C loss were actually sites of active ERα driven transcription. For *PGR*, we were specifically able to connect the modest loss of H3K4me1 at the *PGR* enhancer with significant loss of PolII binding at the promoter. These data begin to specify the mechanism of KMT2C regulation of ERα function – through enabling transcriptional activity at ERα enhancer sites. These data are significant in the context of better delineating regulation of ERα activity especially because it is known that approximately 96% of ERα binding sites are at enhancer regions of the genome(187). However, much of the analyses on regulation of ERα activity have been limited to better understanding its role at gene promoters. Here we present KMT2C as the first known enhancer-specific regulator of ERα activity.

To understand more specifically how H3K4me1 might influence ERα activity, we conducted ERα ChIP-sequencing and surprisingly found no significant loss of ERα binding despite loss of proximal ERα target gene expression. We then noted that the regions encompassing sites of H3K4me1 loss were enriched for motifs that predicted for ERα binding partners, such as FOXA1 and AP-1. In addition, we observed that KMT2C loss led to decreases in ERα interactions with multiple known binding factors including MCM4, FOXA1 and TRPS1. Finally, as an independent proof of potential interactions of KMT2C and FOXA1, Jozwik et al recently uncovered a direct interaction between KMT2C

and FOXA1(107). How FOXA1, AP-1, and other specific factors ultimately coordinate KMT2C-dependent ERα activity will be an area for further investigation.

CHAPTER 3: KMT2C LOSS IS A MEANS FOR BREAST CANCER CELLS TO ESCAPE HORMONE DEPENDENCE

Introduction

KMT2C is an H3K4 histone methyltransferase that is among the most frequently mutated genes in breast cancer. The *KMT2C* mutation spectrum strongly suggests that it is a tumor suppressor as there is an abundance of truncation and missense mutations that span along the length of the protein. Aside, from breast cancer, *KMT2C* is also very frequently mutated in melanoma, uterine carcinoma, bladder cancer, lung adenocarcinoma and others(89). Across these cancer types, KMT2C loss of function is repeatedly selected for whether it is through deletion of chromosome 7, as in the case of acute myeloid leukemia, or through truncation mutations, as in bladder and breast cancer.

Despite its frequent mutation in breast cancer, we do not understand the selective advantage of KMT2C loss. In other cancer types including acute myeloid leukemia and bladder cancer, it has been suggested that loss of KMT2C may promote a more stem-like gene expression profile(81) or that KMT2C may act as a coactivator of p53(82). In breast cancer stem cells, KMT2C loss has been suggested to promote a more stem-like phenotype through activation of the HIF signaling pathway(110). However, in the context of breast cancer epithelial cells, our previous studies have indicated that KMT2C serves as a key regulator of ERα transcriptional activity, suggesting that its frequent mutation is paradoxical. Here, we show that KMT2C loss allows for hormone-independent proliferation of the ER+HER2- MCF7 breast cancer cell line. KMT2C deficient MCF7 cells that are hormone-independent retain their dependence on unliganded ERα as ERα acquires approximately 10,000 novel binding sites. In

accordance, KMT2C mutation frequency increases selectively in breast cancer patients with ER+ hormone refractory disease.

Results

KMT2C mutation and gene expression signatures of KMT2C loss predict for poor survival outcomes

Our data on KMT2C loss in ER+ breast cancer cell lines suggest that it functions as a coactivator of ERα signaling. Given the integral role of ERα signaling in ER+ breast cancer, the described roles of KMT2C would suggest it to have oncogenic functions instead of tumor suppressive roles. However, this contradicts the presence of numerous truncating mutations that span along the length of the *KMT2C* gene. Therefore, we wanted to confirm if the presence of *KMT2C* mutation correlated with patient prognosis, which might imply that these mutations were contributing to a more or less aggressive tumor. To this end, we looked at disease-free survival and overall survival in patients with *KMT2C* mutations from the TCGA database. We found that mutation in *KMT2C* was associated with significantly worse disease-free survival and even more significantly associated with worse overall survival (Fig. 3.1A, B).

Given that *KMT2C* mutations include missense mutations, which may not always result in loss of KMT2C activity, and given that there are other mechanisms for loss of KMT2C function including reducing KMT2C expression, we wanted to more carefully assess the effects of KMT2C loss on patient prognosis. Therefore, we derived a gene expression signature from our MCF7

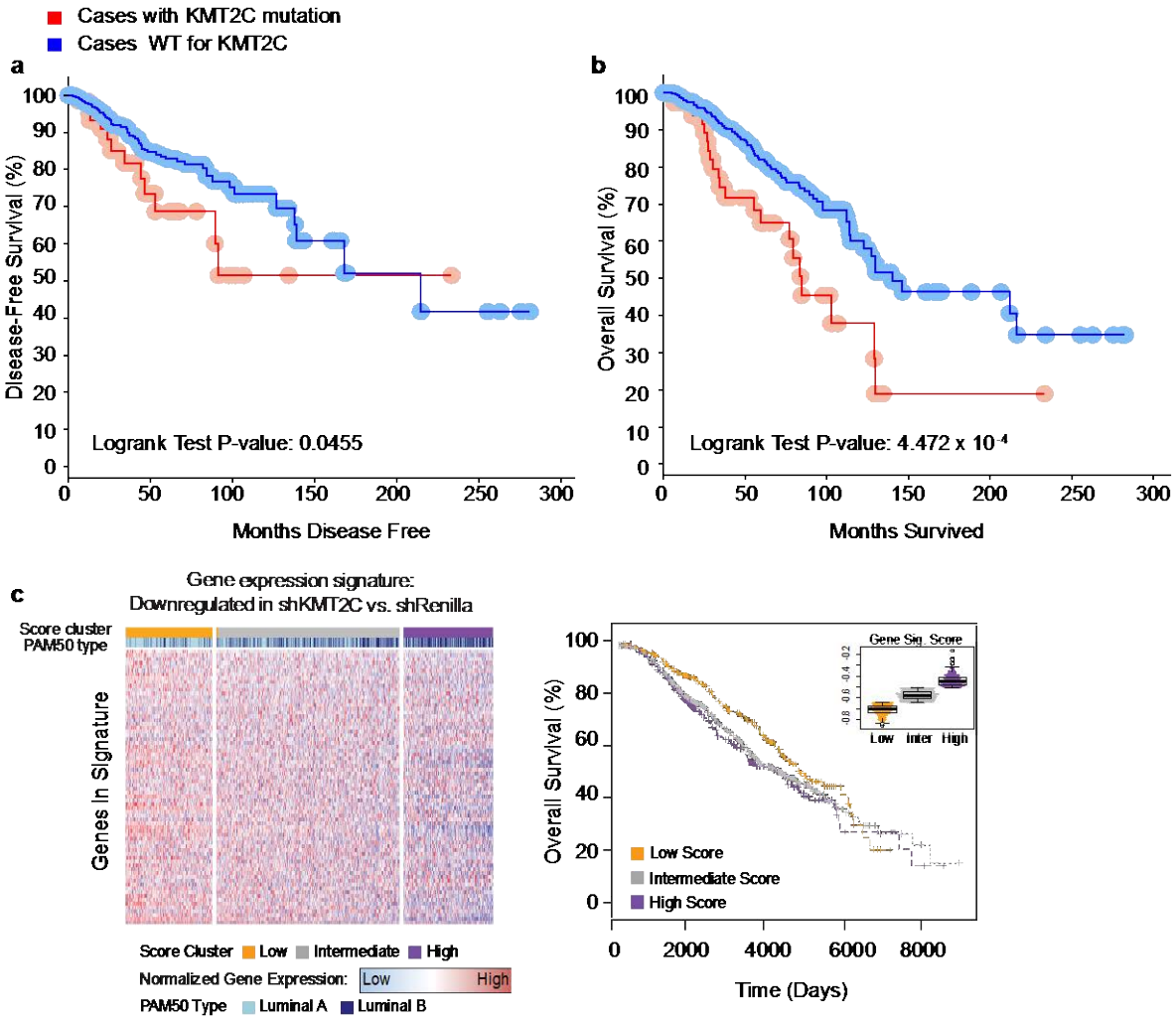

Figure 3.1 KMT2C mutation and gene expression signature is correlated with poorer outcomes. (A,B). Overall survival and disease-free survival curves were generated through cBioPortal webtool using the TCGA(180) breast cancer cohort (n= 818) (C) Molecular stratification of Luminal A and Luminal B patients in the METABRIC cohort using gene expression signatures derived from KMT2C knockdown experiments. Results are shown for an experimentally derived gene signatures of genes downregulated in shKMT2C vs. shRenilla. Kaplan-Meier curve shows overall survival for patients with high signature scores (1st quartile, purple), low signature scores (4th quartile, yellow) and intermediate signature scores (2nd and 3rd quartiles, gray). Heat maps show normalized gene expression for the genes included in each signature across the 1209 patients, ranked from left to right in increasing order of signature score. Luminal A or B PAM50 subtyping indicated by light or dark blue lines. Columns represent individual patients and rows represent probes. Three-class comparison log-rank p-value for survival curve= 0.0277. Two-class comparison log-rank p-value between high and low scoring patients= 0.0093.

KMT2C knockdown cells. We then compared this derived gene expression signature to derived KMT2C loss gene expression signatures of 1209 Luminal A and Luminal B breast cancer patients from the METABRIC database. We were able to separate the patients into three groups (high, mid, low) based on the similarity of their gene expression data to the derived gene expression signatures (Fig 3.1C, top panel). Those with a high score had a gene expression profile most similar to that of KMT2C knockdown while those with a low score had a gene expression profile least similar to that of KMT2C. Intriguingly, despite the fact that this gene expression signature was derived from slower growing shKMT2C MCF7 cells, we found that patients with a gene expression signature similar to that of KMT2C loss has worse overall survival than patients with a gene expression signature less similar to that of KMT2C loss (Fig. 3.1C, bottom panel). Consistent with these outcomes, we found that patients with high scores were more likely to be Luminal B in subtype than Luminal A (Fig. 3.1C, top panel). Together with the *KMT2C* mutation spectrum, this strongly suggests that KMT2C is a tumor suppressor whose loss contributes to more aggressive tumors.

KMT2C loss does not result in increased transformation of MCF10A cells

Given that KMT2C loss in patients correlates with poorer survival, we wanted to determine the contribution that KMT2C loss was having on tumor aggressiveness. Having already established that loss of KMT2C does not result in increased proliferation across multiple breast cancer cell lines, we set out to assay for others aspects of cancer progression. One such aspect is

transformation of normal mammary epithelial cells, which we tested by performing a colony formation assay using MCF10A cells. MCF10A in an immortalized mammary epithelial cell line used to model normal breast tissue. Previous studies have shown that overexpression of a *NEU* transgene in these cells result in a more spindle-like morphology and increased colony formation in matrigel(190). Therefore, we used overexpression of the *NEU* transgene as a positive control and tested if CAS9nickase mediated loss of KMT2C would result in a similar phenotype. Upon performing a 10-day colony formation assay on matrigel we found that our *NEU*- expressing controls cells quickly formed large colonies and morphed into more spindle-shaped cells. In contrast the *KMT2C* knockout cells lines looked very similar to both parental MCF10A and crispr control MCF10A cells (Fig. 3.2). This suggested that loss of KMT2C alone is not enough to transform these cells.

KMT2C does not increase metastatic potential of breast cancer cell lines

In addition to transformation, we also tested whether loss of KMT2C could promote metastases. We rationalized that certain pro-metastatic factors, such as TGFβ, are known to suppress proliferation but also promote metastasis(191). Given that loss of KMT2C does suppress the proliferation of ER+HER2- breast cancer cells, we tested whether the selective advantage for its mutation stems from its ability to promote metastasis. We injected luciferase-expressing shRenilla and shKMT2C MCF7 cells into the tail vein of nude female mice and

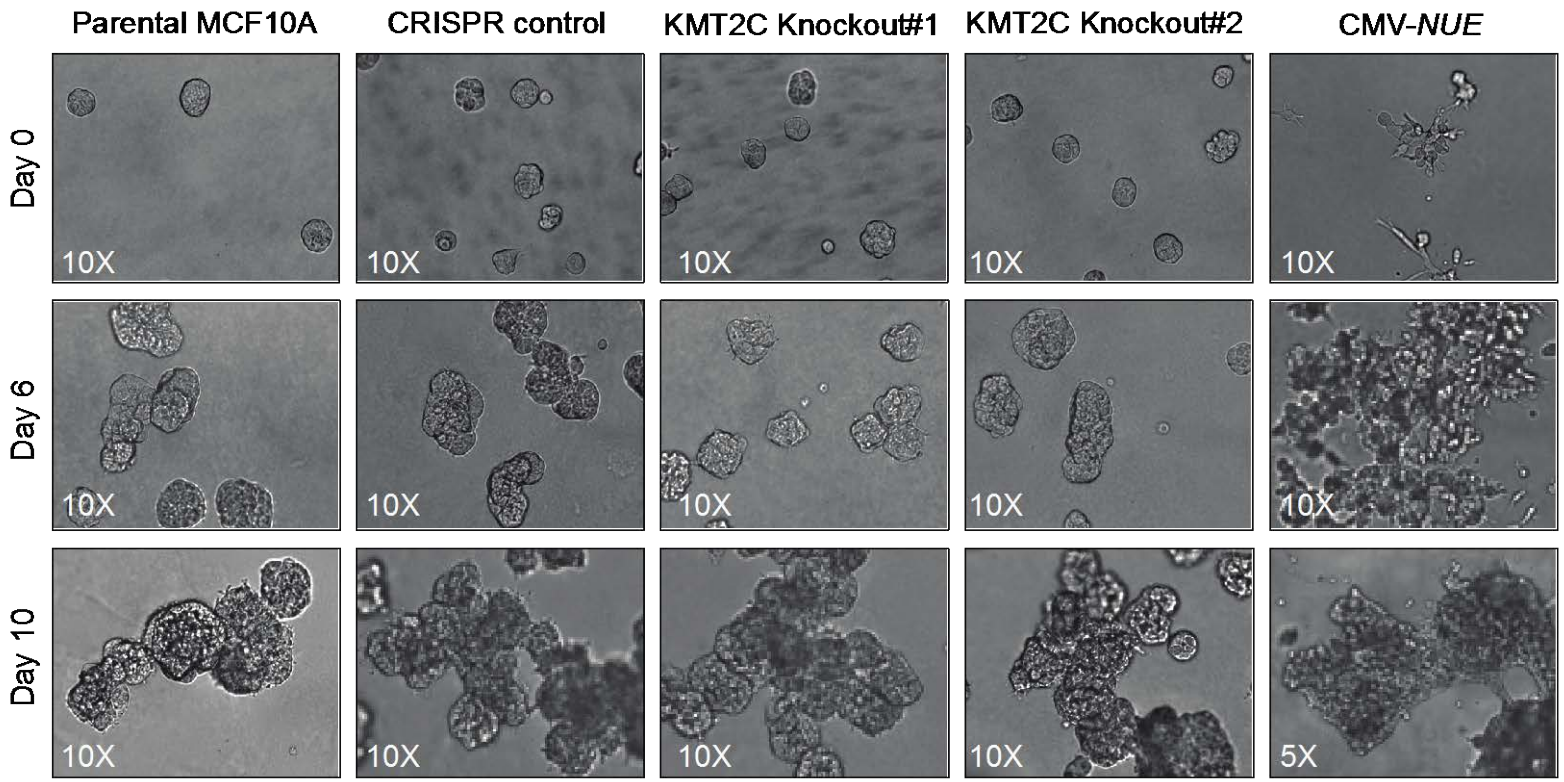

Figure 3.2 KMT2C loss does not result in increased colony formation in MCF10A cells. Indicated MCF10A cells were cultured on matrigel for up to 10 days. CMV-*NEU* was used as a positive control for MCF10A transformation.

monitored for colonization of cells by weekly bioluminescence imaging. Our results show modest colonization of both shRenilla and shKTM2C cells with no significant increases in metastatic potential mediated by KMT2C loss (Table 3.1).

KMT2C loss promotes hormone independent ER+ breast cancer cell proliferation

Given the role of KMT2C in estrogen-driven growth, we wanted to determine how loss of KMT2C affected ER+ cell response to estrogen deprivation. MCF7 cells are known to arrest their growth in the absence of estrogen mimicking clinical responses to estrogen deprivation therapies including ovariectomies and aromatase inhibition. However, after 3-4 months MCF7 cells are known to be able to outgrow these hormone-deprived conditions (192, 193). To assess the role of KMT2C in this context, we cultured KMT2C deficient cells in hormone-deprived media and observed that KMT2C deficient MCF7 cells displayed a much more rapid time to growth with cells emerging in as little as two weeks (Fig. 3.3). As expected, shRenilla cells continued to be growth arrested for 10-12 weeks beyond the time KMT2C knockdown cells grew.

After establishing hormone resistance in vitro, we wanted to test the response of KMT2C deprived cells to hormone deprivation in vivo. To this end, we implanted both control and KMT2C knockdown cells into the mammary fat pads of nude mice that had been implanted with estrogen pellets. Once the tumors reached approximately $100mm^3$ in size, we removed the estrogen pellets and continued to measure tumor size. We found that both control and KMT2C

Table 3.1 KMT2C loss does not enhance metastases

MCF7 xenograft model	Mice with mets/ total mice	Fluorescence (p/sec/cm^3/sr)
shRenilla no E2 pellet	1/15	73400
shKMT2C no E2 pellet	1/15	14200, 13000, 84300
shRenilla with E2 pellet	3/15	74400
shKTM2C with E2 pellet	1/15	41500

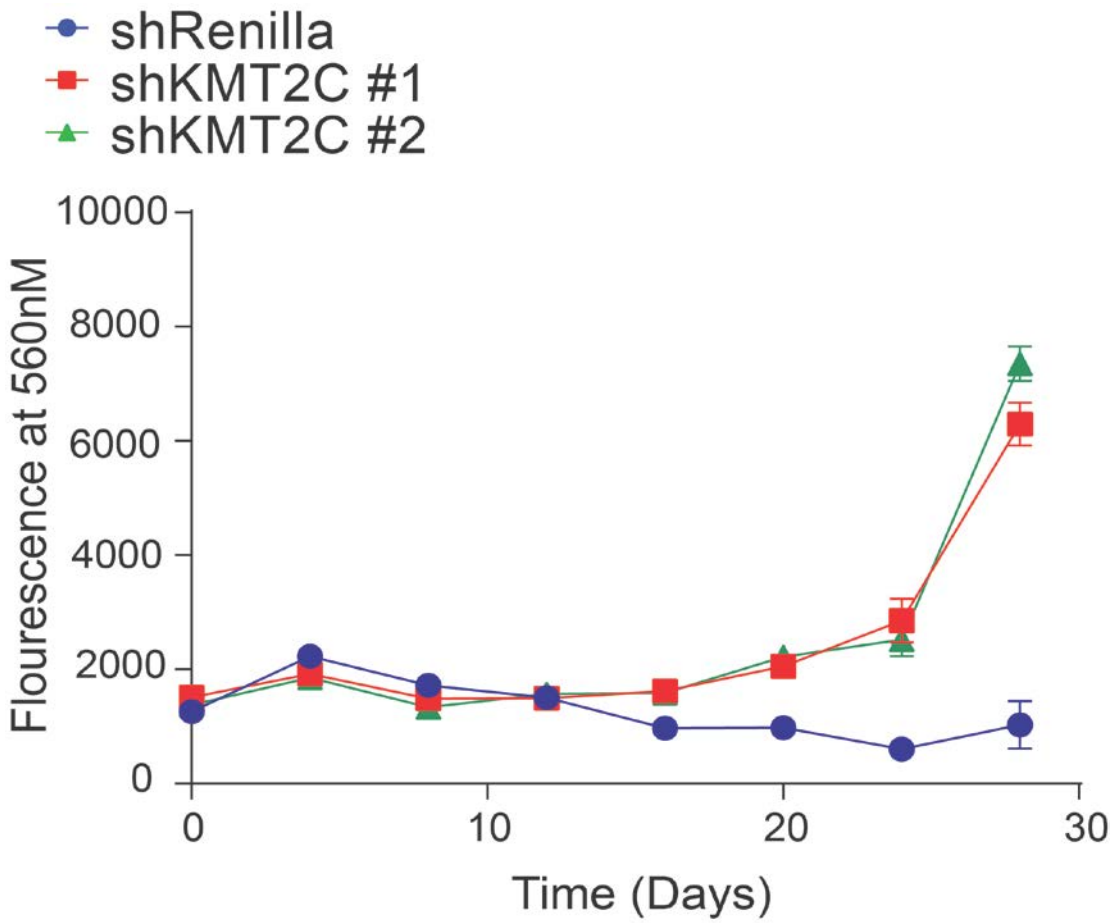

Figure 3.3 KMT2C loss results in rapid outgrowth of ER+ breast cancer cells in hormone deprived media. MCF7 cells stably expressing shRenilla or shKMT2C were assayed for proliferation in phenol-red free DMEMF12 with 10% charcoal stripped serum (CSS) medium using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m. Data correspond to one representative assay from a total of two independent assays.

knockdown cells did eventually grow out under hormone-deprived conditions but KMT2C knockdown cells did so at a faster rate than control cells, reaching 500mm^3 approximately four weeks before control cells (Fig. 3.4).

KMT2C mutation frequency goes up in hormone resistant patients

We next wanted to assess whether our finding that KMT2C loss promotes rapid hormone-independent outgrowth was supported by clinical data. We hypothesized that perhaps *KMT2C* mutation frequency would increase among hormone-refractory patients if the ability to promote hormone-independent proliferation were indeed a selective advantage of its mutation. We specifically analyzed sequencing data from patients with metastatic, hormone-refractory ER+ breast cancer and found that in this patient population, *KMT2C* mutation frequency was approximately 33% (Fig. 3.5). This was a dramatic increase from the ~7% mutation frequency seen in primary, untreated breast cancer. In both cases, mutations appear to be associated with loss of function and occur 5' to the SET domain (as shown in Fig. 2.1C). This suggested that even in the clinical setting, *KMT2C* mutation may promote hormone resistance.

Knockdown of KMT2C in other ER+ cell lines has differential effects on hormone-deprived proliferation

We then wanted to assess the effects of KMT2C loss in other ER+ cells lines to determine if its ability to promote hormone-independent growth was universal across multiple breast cancer cell lines. We found that the other ER+

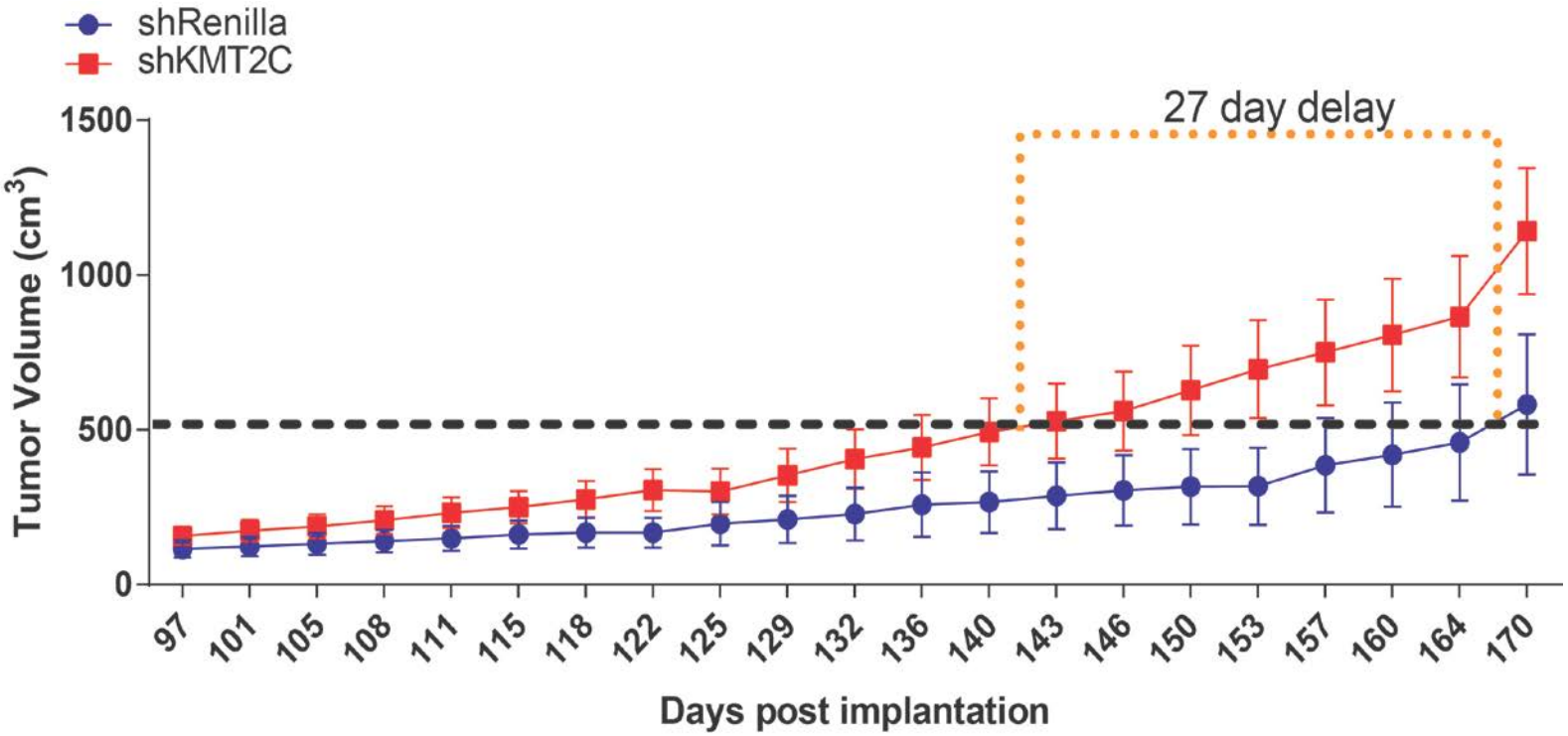

Figure 3.4 KMT2C loss results in outgrowth of ER+ breast cancer cells in xenograft models of estrogen deprivation. Mice bearing tumors derived from MCF7 cells with shRenilla or shKMT2C short hairpins had their E2 pellets removed when the tumors reached an approximate volume of 100 mm3. Values correspond to the average of 10 mice per group ± s.e.m.

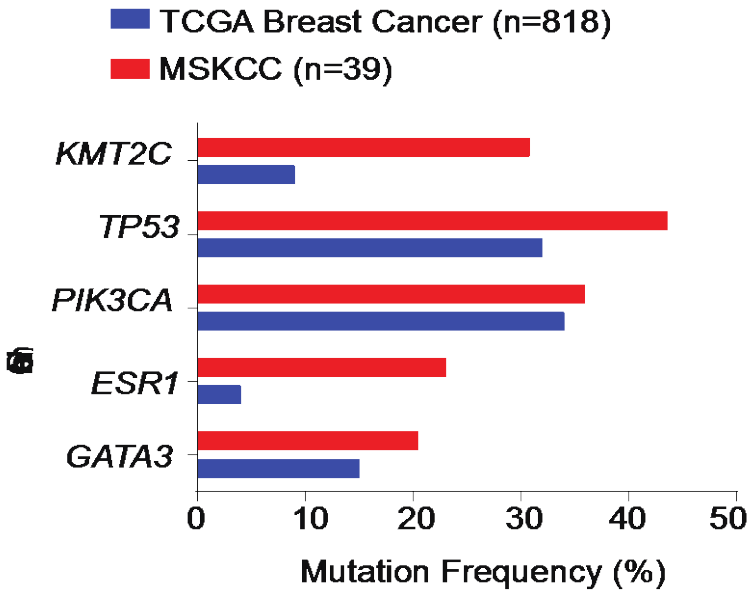

Figure 3.5 KMT2C mutation frequency goes up in hormone resistant patients. Percentage of cases with gene mutations detected in patient samples (MSKCC(119)) were compared to those of TCGA(180).

cells lines were largely insensitive to hormone-deprivation making them unsuitable for studying the effects of long-term estrogen deprivation. In fact, we found that Cama-1 was the only ER+ cell line that remained sensitive to hormone-deprivation and in this cell line, we see cell death in the presence of hormone-deprivation following ~12 days in charcoal stripped serum (Fig. 3.6). Also, these cells in normal media are the most sensitive to KMT2C loss (as shown in Fig. 2.5) suggesting that they might not be able to overcome the detrimental effects of KMT2C loss on their proliferation to allow for growth in hormone-deprived media.

Knockdown of KMT2A, B and D does not result in outgrowth in hormone-deprived media

Given that KMTA, B and D are also mutated in breast cancer and knockdown of these proteins in MCF7 cells had similar effects on proliferation as did knocking down KMT2C, we wanted to assess whether loss of the other KMT2 family members would also promote hormone-independent outgrowth. We cultured cells deficient in KMT2A, B, C and D along with control cells and found that loss of KMT2C was the only one that promoted rapid outgrowth in hormone-deprived media (Fig. 3.7). Consistent with this, we do not see increased mutation frequency of the other members of the KMT2 family in our hormone-refractory patients (data not shown).

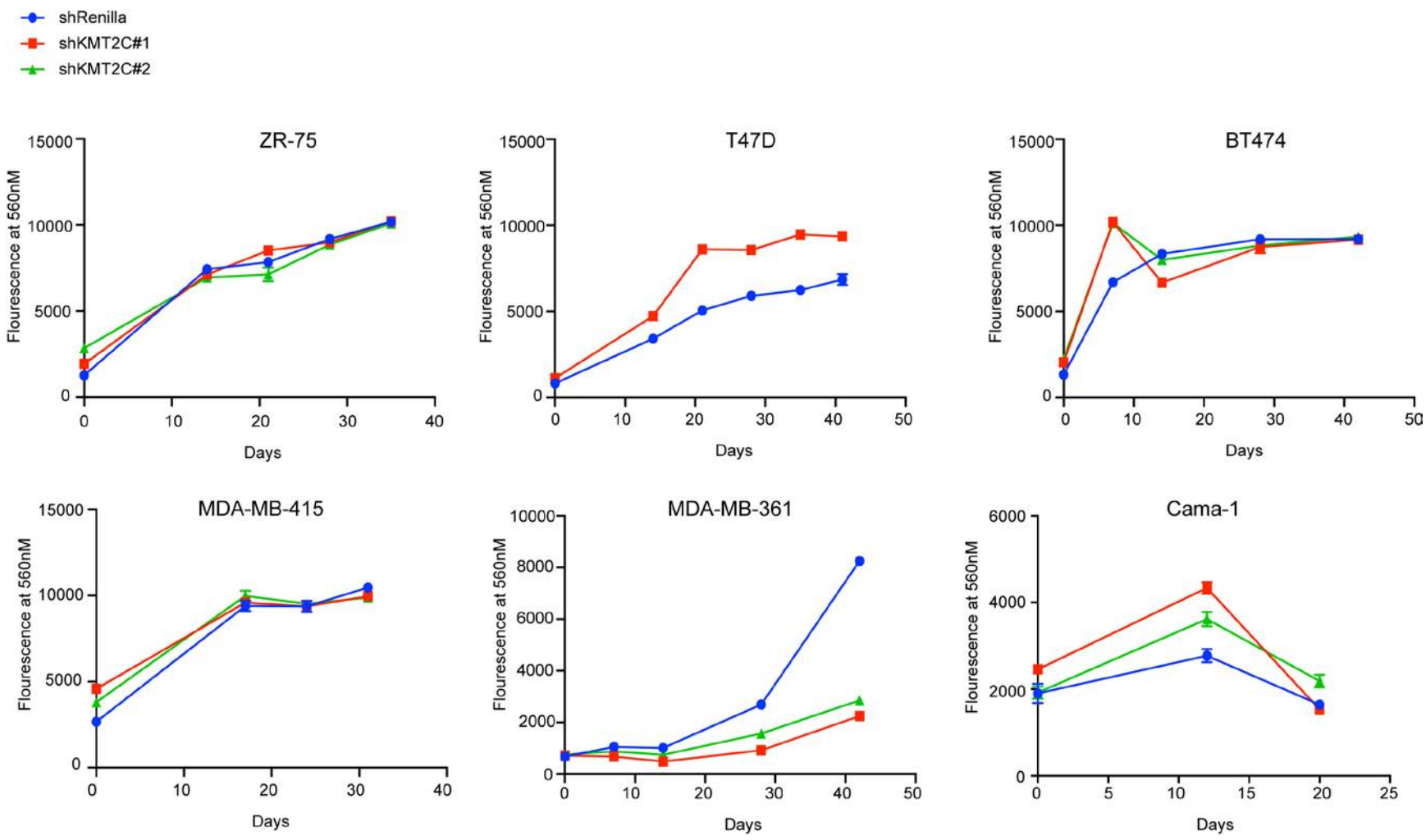
shRenilla
shKMT2C#1
shKMT2C#2
ZR-75
T47D
BT474
MDA-MB-415
MDA-MB-361
Cama-1
Flourescence at 560nM
Days

Figure 3.6 Knockdown of KMT2C in other ER+ cell lines has differential effects on hormone-deprived proliferation. Indicated cells stably expressing shRenilla or shKMT2C were assayed for proliferation in phenol-red free DMEMF12 with 10% charcoal stripped serum (CSS) medium using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m. Data correspond to one representative assay from a total of two independent assays.

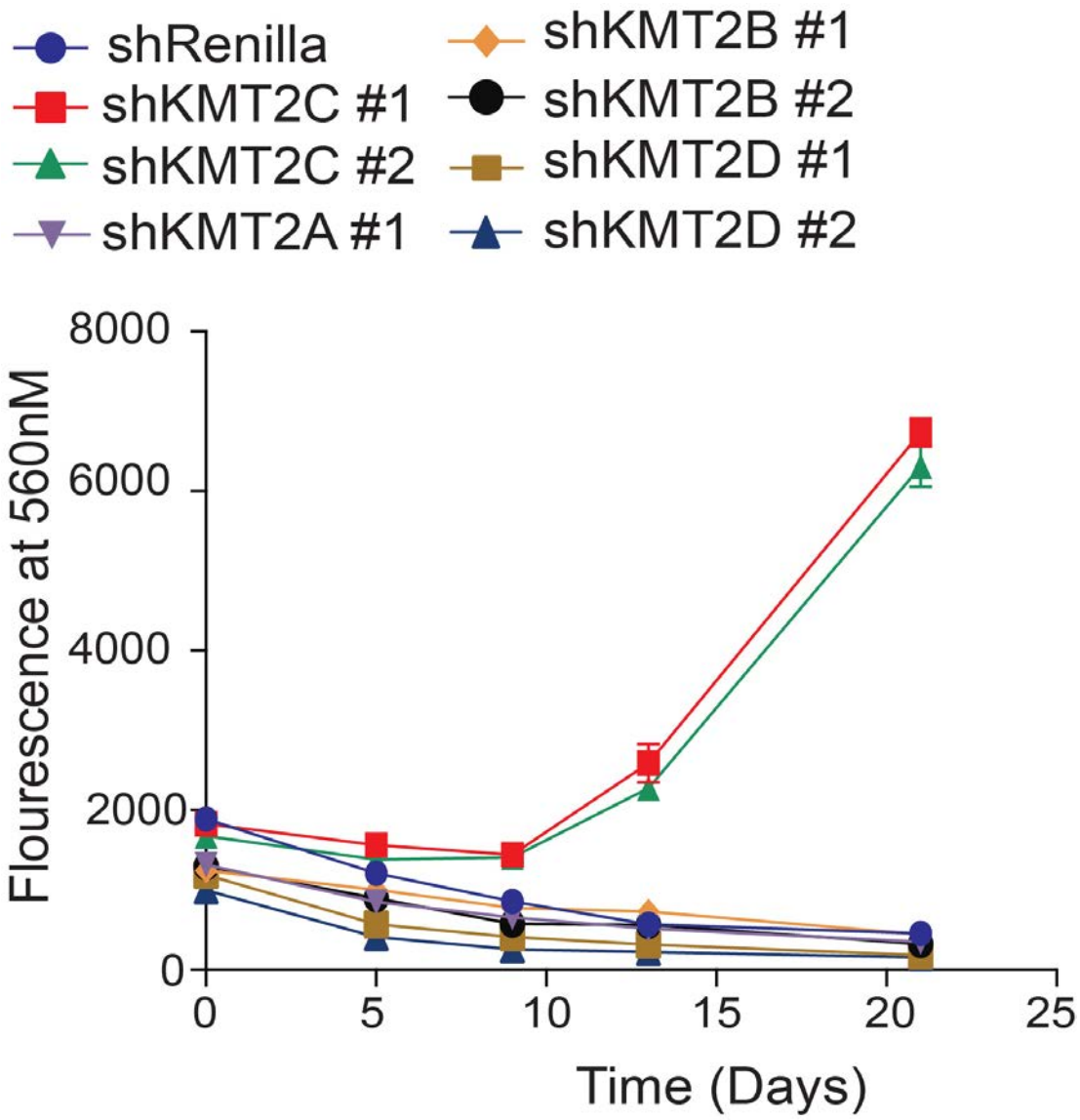

Figure 3.7 KMT2C knockdown but not that of its family members results in rapid outgrowth in hormone deprived media. MCF7 cells stably expressing shRenilla, shKMT2A, shKMT2B, shKMT2C or shKMT2D were assayed for proliferation in phenol-red free DMEMF12 with 10% charcoal stripped serum (CSS) medium using the alamarBlue cell viability assay. Values correspond to the mean of six experimental replicates ± s.e.m. Data correspond to one representative assay from a total of two independent assays.

shKTM2C-R cells maintain KMT2C expression levels and an H3K4me1 occupancy signature consistent with maintained KMT2C knockdown

To investigate the basis of hormone-independent growth in the context of KMT2C knockdown, we established a hormone-independent MCF7 shKMT2C#2 cell line referred to as shKMT2C-R. Next generation DNA sequencing using the MSKCC-IMPACT platform revealed no obvious acquired mutations compared to parental cells that could explain the rapid outgrowth in hormone-deprived media (Table 3.2).

We assayed for KMT2C mRNA levels and found that KMT2C remained knocked down whether the cells were cultured in full serum or hormone-deprived charcoal stripped serum (Fig. 3.8A, B). Additionally, we also performed H3K4me1 ChIP-sequencing in these cells and found that the original 869 sites with reduced H3K4me1 continued to have reduced levels of H3K4me1 (Fig. 3.8C, D). These data suggested to us that the development of hormone-independent growth was not through replacement of KMT2C activity upon ERα enhancers by upregulating a different H3K4 methyltransferase or suppressing the demethylase.

Gene expression signatures of shKMT2C-R cells correlate with poorer prognosis

Next we wanted to do a similar assessment on the shKMT2C-R gene expression profile and determine if it also correlated with patient prognosis as did the gene expression profile of shKMT2C cells cultured in normal media. Again,

Table 3.2 Mutations in shKMT2C-R cells

Gene	Protein Change	Allele Freq. shKMT2C#2 parental	Allele Freq. shKMT2C#2-R	COSMIC
ABL1	**P900L**	0.06	not present	
BRD4	**P956Tfs*137**	0.15	not present	
PIK3C2G	**Q404E**	0.16	not present	
RPTOR	**R330Q**	not present	0.19	2
KEAP1	**H451Y**	not present	0.05	
PIK3CA	**E545K**	0.30	0.23	1431
GATA3	**D335Gfs*17**	0.46	0.50	1
ATRX	**D2136V**	0.07	0.07	2
ERBB4	**Y1242C**	0.22	0.24	2
PTPRD	**G61E**	0.33	0.29	2
FGF3	**R120W**	0.22	0.24	1
IRS1	**R327C**	0.32	0.33	1
MAP3K13	**D380N**	0.23	0.24	1
CHEK2	**R145L**	0.18	0.23	
HIST1H1C	**A24T**	0.45	0.47	
NBN	**P325H**	0.09	0.10	
NBN	**R43***	0.68	0.67	
PHOX2B	**Y83C**	0.41	0.13	
SF3B1	**E870K**	0.24	0.21	
SPOP	**S330L**	0.19	0.22	
TBX3	**P602S**	0.21	0.18	
ZFHX3	**Q1740_Q1741del**	0.25	0.24	

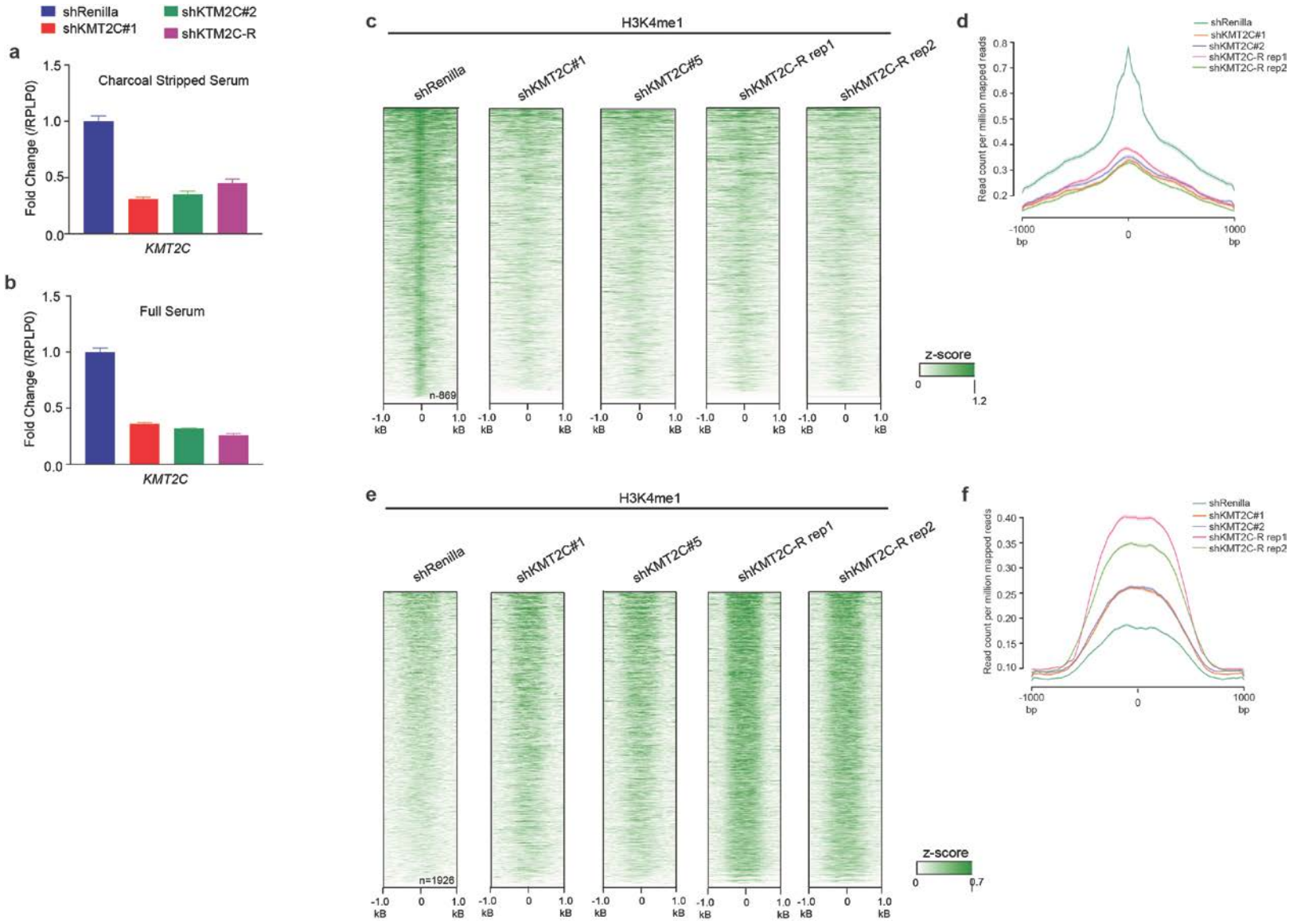
shRenilla
shKMT2C#1
shKTM2C#2
shKTM2C-R
a
Charcoal Stripped Serum
Fold Change (/RPLP0)
KMT2C
b
Full Serum
Fold Change (/RPLP0)
KMT2C
c
H3K4me1
shRenilla
shKMT2C#1
shKMT2C#5
shKMT2C-R rep1
shKMT2C-R rep2
n-869
z-score
d
Read count per million mapped reads
bp
e
H3K4me1
n=1926
f

Figure 3.8 shKMT2C-R cells maintain reduced KMT2C expression and an H3K4me1 occupancy signature of KMT2C knockdown. (A) mRNA levels of *KMT2C*, as measured by qRT-PCR, in shRenilla, shKMT2C#1, shKMT2C#2 and shKMT2C-R MCF7 cells. Values correspond to the average of three replicates ± s.e.m. shRenilla, shKMT2C#1 and shKMT2C#2 were cultured in charcoal stripped serum for 48 hours prior to collection. shKMT2C-R cells were maintained and collected in charcoal stripped serum. (B) mRNA levels of *KMT2C*, as measured by qRT-PCR, in shRenilla, shKMT2C#1, shKMT2C#2 and shKMT2C-R MCF7 cells. Values correspond to the average of three replicates ± s.e.m. shKMT2C-R was cultured in full serum for 48 hours prior to collection. shRenilla, shKMT2C#1 and shKMT2C#2 cells were maintained and collected in full serum. (C) Normalized heatmaps for H3K4me1 occupancy in shRenilla, shKMT2C#1, shKMT2C#2 and shKMT2C-R MCF7 cells among 869 differential sites where H3K4me1 was reduced in shKMT2C#1, shKMT2C#2 as compared to shRenilla (same sites as depicted in Figure 4). Heatmaps are centered at the peak summit. (D) The average binding of H3K4me1 in shRenilla, shKMT2C#1, shKMT2C#2 and shKMT2C-R MCF7 cells at 869 sites of H3K4me1 loss at a region centered at the peaks and showing ±1kB. (E) Motif analysis, using HOMER, at the 1,749 H3K4me1 loci in shKMT2C-R (maintained in charcoal stripped serum) as compared to shKMT2C#2 MCF7 cells (maintained in full serum).

upon comparing Luminal A and B breast cancer patient derived gene expression signatures to shKMT2C-R derived gene expression signatures, we found that patients that were more similar to shKMT2C-R had worse overall survival (Fig. 3.9). Additionally, patients that were more similar to shKMT2C-R were more likely to be Luminal B in subtype than Luminal A (Fig. 3.9 top panels). Overall, this data suggested to us that the shKMT2C-R cells describe a more aggressive cancer.

shKMT2C-R cells further downregulated ERα target genes

To interrogate what pathways might be allowing for hormone-independent proliferation of the shKMT2C-R cells, we performed RNA sequencing on these cells and compared their gene expression to that of parental shKMT2C#2 or shRenilla MCF7 cells maintained in hormone-deprived media for five days (Fig. 3.10A, B). We found that genes downregulated in shKMT2C-R cells were enriched for ERα target gene signatures (Fig 3.10C, D) and for growth factor stimulation signatures among others (Fig. 3.10E, F). We also examined genes upregulated in the shKMT2C-R cells and found enrichment for genes involved in metabolic pathways including fatty acid, inositol phosphate, glucose and amino acid metabolic pathways. However, no individual pathway appeared significant when we corrected for multiple hypothesis testing (data not shown). This data suggested that there was ongoing suppression of ERα target gene expression, consistent with the culturing of these cells in hormone-deprived conditions. This also suggested to us that it was not a re-expression of ERα-target genes that was promoting their outgrowth under conditions of hormone deprivation.

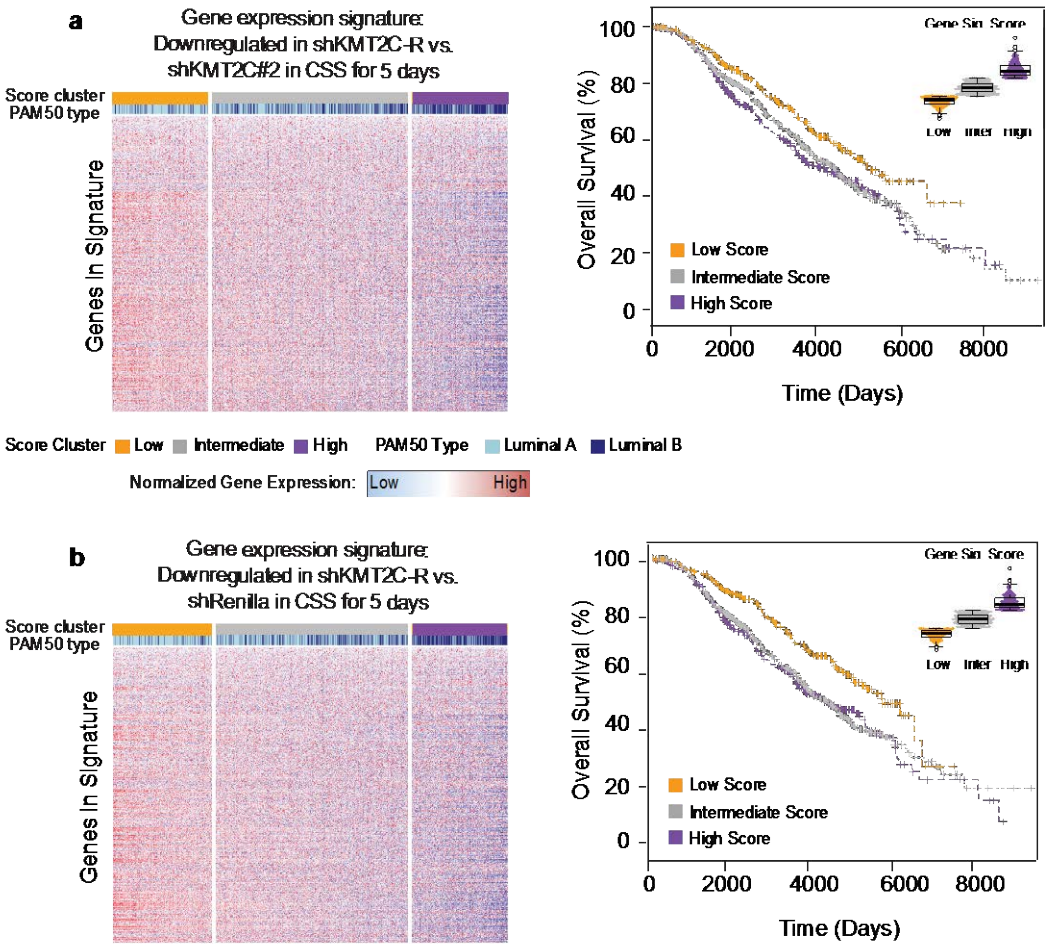

Figure 3.9 Gene expression profiles of shKMT2C-R cells correlated with Luminal B breast cancer and poorer outcomes. (A,B) Molecular stratification of Luminal A and Luminal B patients in the METABRIC cohort using gene expression signatures derived from KMT2C knockdown experiments. Results are shown for three experimentally derived gene signatures that include genes downregulated in shKMT2C-R vs. shKMT2C#2 in CSS (A) and genes downregulated in shKMT2C-R vs. shRenilla in CSS (B). Kaplan-Meier curves show overall survival for patients with high signature scores (1st quartile, purple), low signature scores (4th quartile, yellow) and intermediate signature scores (2nd and 3rd quartiles, gray). Heat maps show normalized gene expression for the genes included in each signature across the 1209 patients, ranked from left to right in increasing order of signature score. Luminal A or B PAM50 subtyping indicated by light or dark blue lines. Columns represent individual patients and rows represent probes. Three-class comparison log-rank p-values for survival curves= 0.0086, 0.0002 for the left, middle and right graphs respectively. Two-class comparison log-rank p-values between high and low scoring patients= 0.0031, 0.00024 for left, middle and right graphs, respectively.

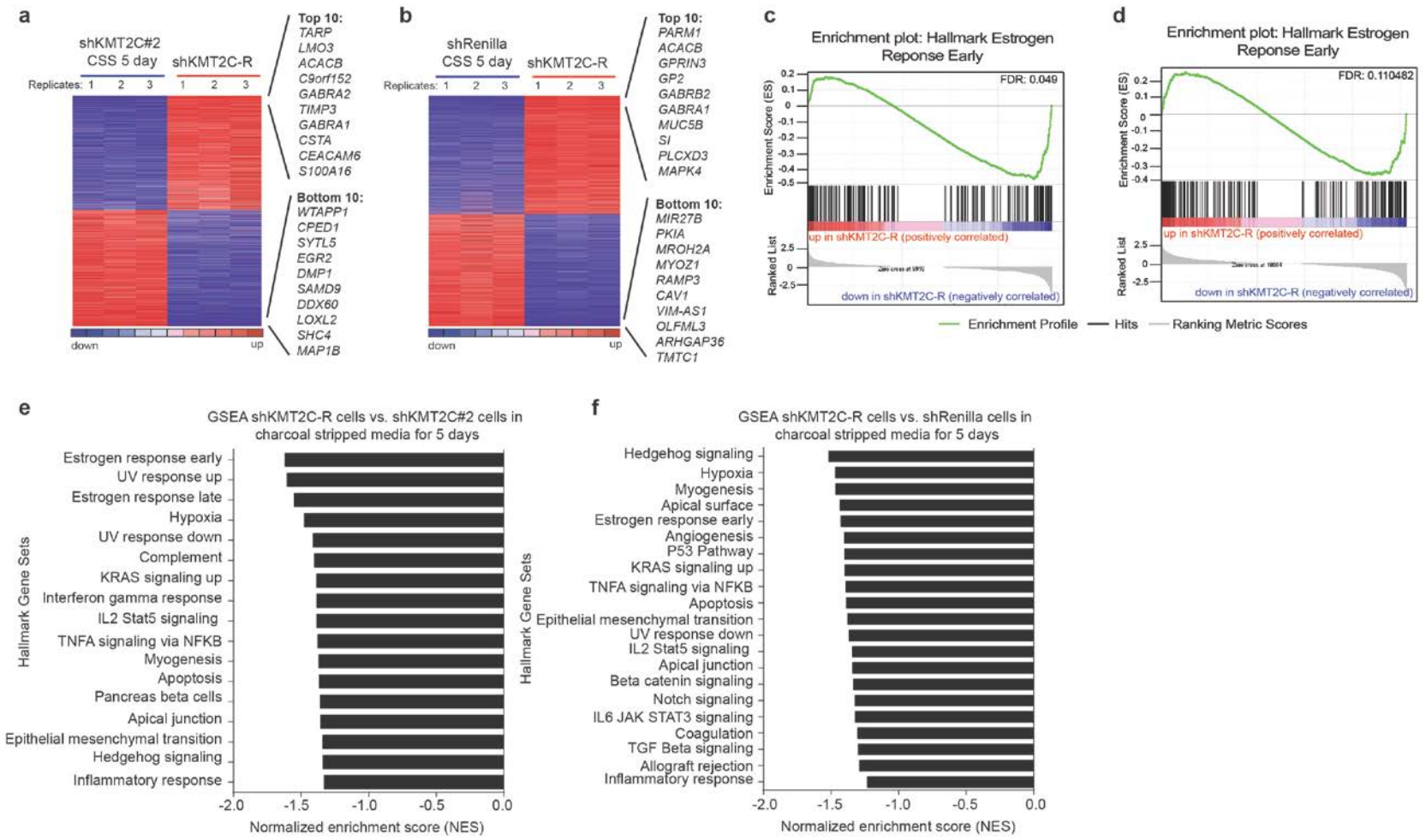

Figure 3.10 GSEA reveals that shKTM2C-R cells further downregulate ERα target genes. (A) Supervised analysis of the 8622 differentially expressed genes between MCF7 shKMT2C #2 cells in charcoal stripped media for 5 days and

MCF7 shKMT2-R. Columns represent individual replicates for MCF7 shKMT2C #2 cells and shKMT2C-R samples, rows correspond to the different genes. Color reflects the normalized expression count (number of reads mapping to each gene); scaled by row. CSS, charcoal stripped media (B) Supervised analysis of the 9398 differentially expressed genes between MCF7 shRenilla cells in charcoal stripped media for 5 days and MCF7 shKMT2-R. Columns represent individual replicates for MCF7 shRenilla cells and shKMT2C-R samples, rows correspond to the different genes. Color reflects the normalized expression count (number of reads mapping to each gene); scaled by row (C) GSEA of 4347 genes significantly downregulated in MCF7 shKMT2C-R cells against MCF7 shKMT2C#2 cells in charcoal stripped media for 5 days (p-val < 0.05) as compared to ranked genes from the Hallmark Estrogen Response Early Geneset (from the Broad Institute Molecular Signature Database) (D) GSEA of 4559 genes significantly downregulated in MCF7 shKMT2C-R cells against MCF7 shRenilla cells in charcoal stripped media for 5 days (p-val < 0.05) as compared to ranked genes from the Hallmark Estrogen Response Early Geneset (from the Broad Institute Molecular Signature Database). (E) GSEA of 4347 genes significantly downregulated in MCF7 shKMT2C-R cells as compared to MCF7 shKMT2C cells in charcoal stripped media for 5 days (p-val < 0.05) against Hallmark panel of gene sets (50 gene sets, downloaded from the Broad Institute Molecular Signature Database). Gene sets shown all had an FDR q-value of at least 0.25 (F) GSEA of 4559 genes significantly downregulated in MCF7 shKMT2C-R cells as compared to MCF7 shRenilla cells in charcoal stripped media for 5 days (p-val < 0.05) against Hallmark panel of gene sets (50 gene sets, downloaded from the Broad Institute Molecular Signature Database). Gene sets shown all had an FDR q-value of at least 0.25.

shKMT2C-R cells maintain responsiveness to estrogen and continue to be sensitive to ERα antagonists

As the gene expression data suggested an ongoing suppression of ERα target genes, we hypothesized that these cells were no longer sensitive to the estrogen stimulation. We tested the response of shKMT2C-R to E2 and surprisingly found that these cells continue to respond to E2 by increasing their proliferative rate (Fig. 3.11). These data suggest an ongoing responsiveness to the estrogen stimulated growth program, but a capability to proliferate in the absence of ERα ligands. To determine if unliganded ERα was still promoting growth in these cells, we examined their response to ERα antagonism. Intriguingly, we found that shKMT2C-R cells were still dependent on ERα as they were very sensitive to inhibition by multiple SERDs including fulvestrant, AZD9496, ARN1917, SRN927, and RU58668 (Fig. 3.11). These data together suggested that the shKMT2C-R cells continue to be responsive to an E2-driven growth program while unliganded ERα remains necessary for the proliferation of these cells.

ERα is not overexpressed in the shKMT2C-R cells

Given the maintained dependence on ERα, we hypothesized that the shKMT2C-R cells may have upregulated ERα expression. Upregulation of ERα has previously been shown to be a characteristic of long-term estrogen deprived (LTED) MCF7 cells(194). We assayed for *ESR1* mRNA levels in our shKMT2C-R cells as compared to parental and control MCF7 cells under conditions of

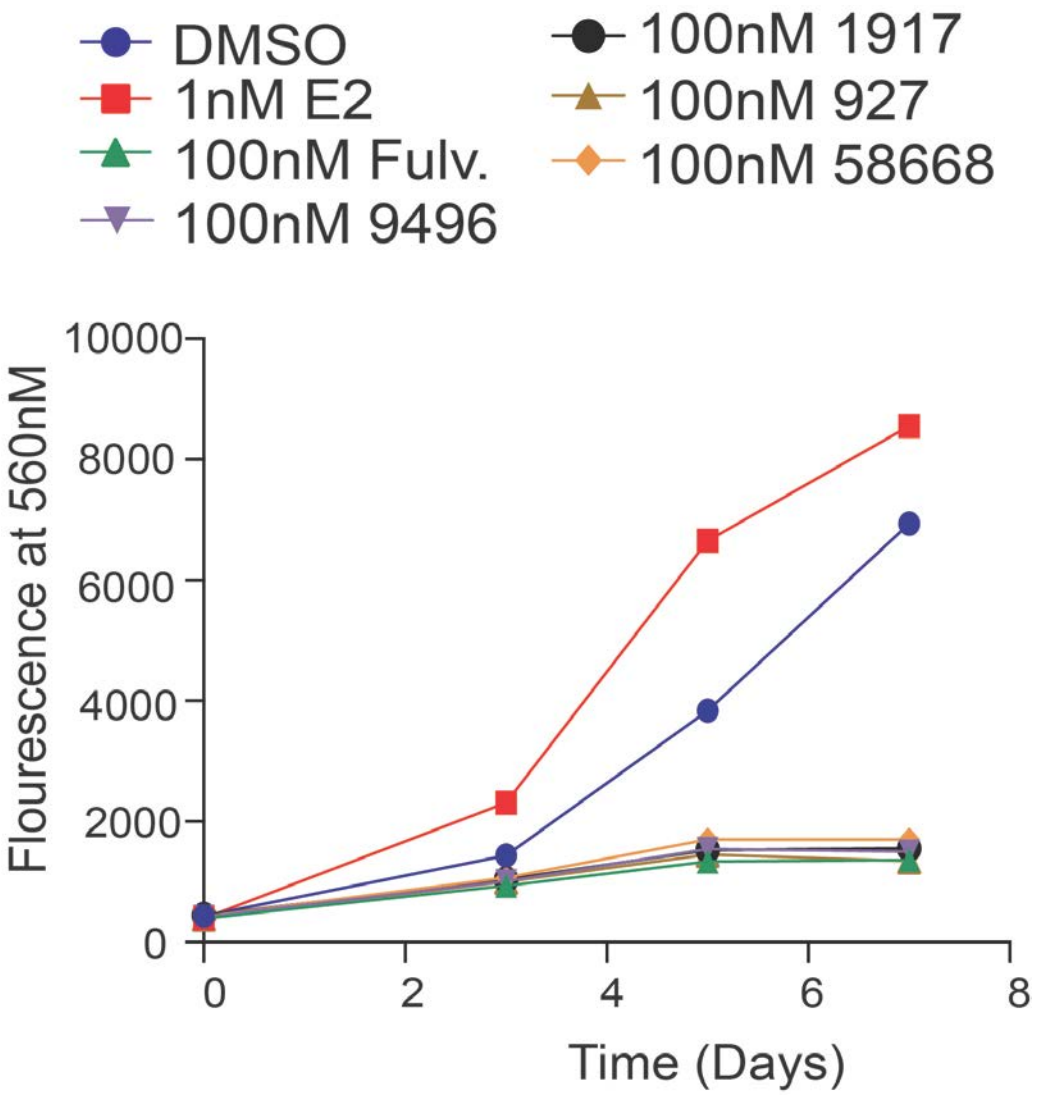

Figure 3.11 shKMT2C-R cells maintain responsiveness to estrogen and continue to be sensitive to ERα antagonists. MCF7 cells were treated with DMSO, estradiol (E2), fulvestrant (Fulv.), AZD9496 (9496), ARN1917 (1917), GDC927 (927), or RU58668 (58668) at indicated doses. Proliferation assayed over the course of 7 days in treatment using the alamarBlue cell viability assay. Values correspond to the average of six experimental replicates ± s.e.m.

hormone deprivation as well as in full serum and found no changes in *ESR1* mRNA levels (Fig. 3.12A). We then immunoblotted for ERα is our shKMTC-R cells as compared with parents shKMT2C cells as well as control cells and again we saw no significant change in ERα protein levels (Fig. 3.12B).

ERα is reprogrammed in the shKMT2C-R cells

Our findings for shKMT2C-R cells demonstrated that they no longer required estrogen to support growth but maintained a dependence of ERα expression without upregulating ERα. To examine the role of the unliganded ERα in this context, we performed ERα ChIP-sequencing of the shKMT2C-R cells. Whereas we continued to identify many of the sites seen in shRenilla cells (n=6598), we also identified over 10,000 novel ERα binding sites (Fig. 3.13). This suggests that ERα has been dramatically reprogrammed in the shKMT2C-R cells. Given the sensitivity to SERD treatment, this reprogrammed ERα may be actively contributing to the proliferation of these cells.

Reprogrammed ERα in shKMT2C-R cells may be partially dependent on AP-1

In order to determine if there were any cofactors or coregulators that might be mediating the reprogramming or affecting the activity of the reprogrammed ERα, we performed motif analysis at these novel binding sites. We found that the motifs of FOXA1 and AP-1, among others, were enriched (Fig. 3.14A). AP-1, in particular, has been previously shown to play a major role in the hormone-independent outgrowth of ER+ breast cancer(161, 195-200). To test the

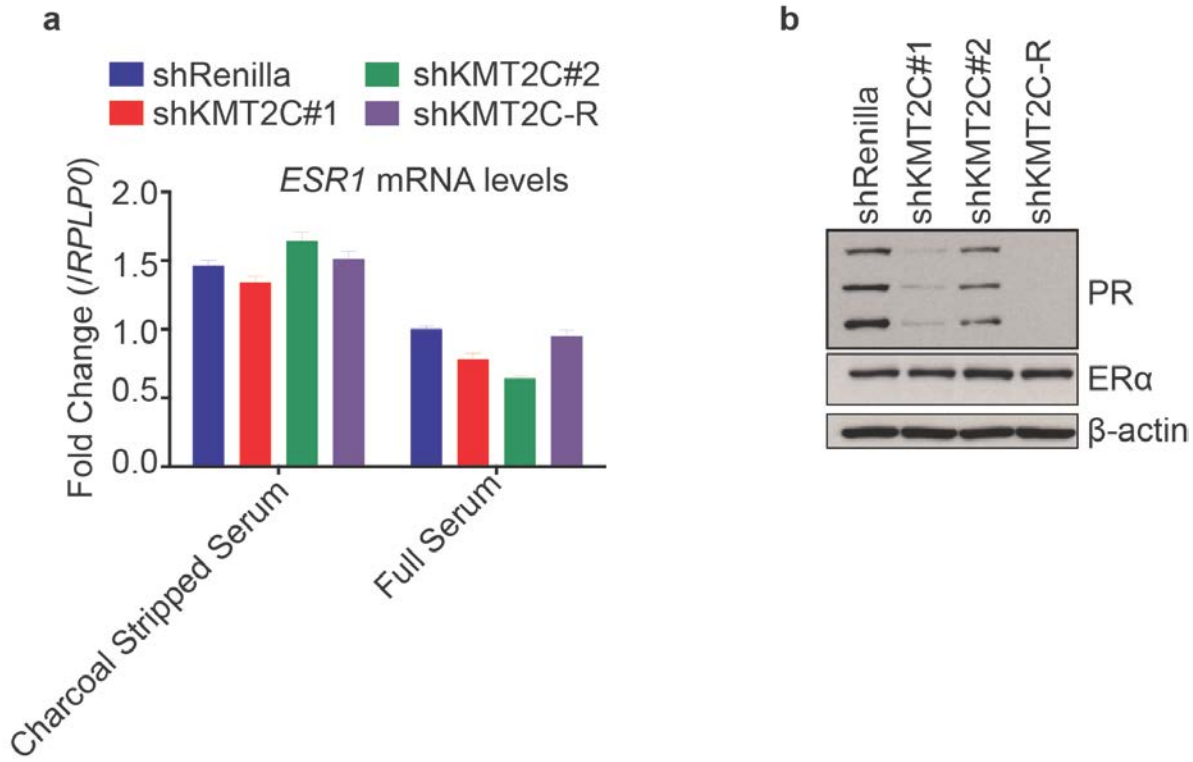

Figure 3.12 ERα expression is unchanged in the shKTM2C-R cells as compared to parental and control MCF7 cells. (A) mRNA level of indicated genes, as measured by qRT-PCR, in shRenilla, shKMT2C#1, shKTM2C#2 and shKMT2C-R cells cultured in either full serum or charcoal stripped serum. (B) Immunoblot of indicated proteins. β-actin was used as a loading control.

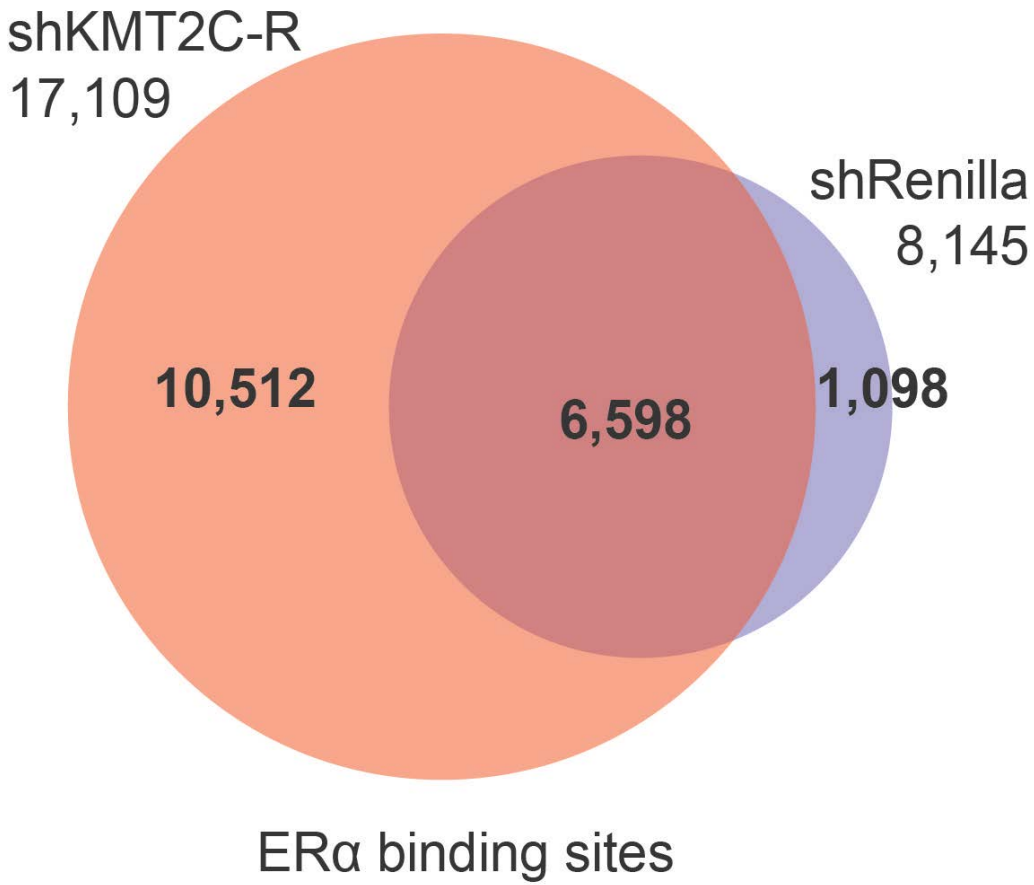

Figure 3.13 ERα is reprogrammed in the shKMT2C-R cells. Venn diagram showing overlap between shKMT2C-R (in CSS, red) and shRenilla (in full serum, purple) ERα binding sites.

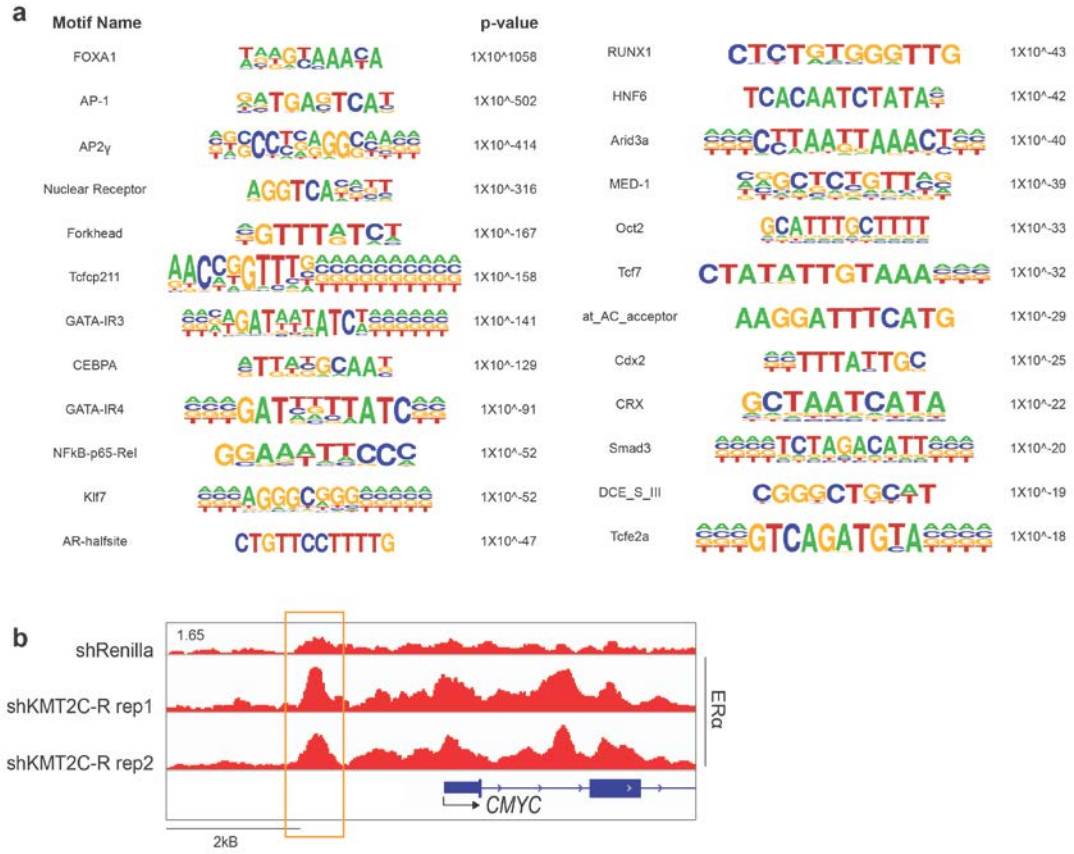

Figure 3.14 Reprogrammed ERα in shKMT2C-R cells may be partially dependent on AP-1. (A) Motif analysis, using HOMER, at the 10,512 loci novel ERα loci in shKMT2C-R (maintained in charcoal stripped serum) as compared to shRenilla MCF7 cells (maintained in full serum). (B) IGV browser views for ERα in shRenilla and shKMT2C-R and MCF7 cells. Outlined in orange is a previously defined AP-1 dependent ERα binding site(201). The y-axis corresponds to the ChIP-seq signal intensity.

relevance of AP-1 for the reprogammed ERα, we looked for ERα binding around a known AP-1-dependent ERα binding site upstream of *c-MYC*, a key driver of hormone-independent growth(201). We found significant increases in ERα binding at this site in shKMT2C-R cells as compared to control cells (Fig. 3.14B). This supports the hypothesis that AP-1 may, in part, be responsible for reprogramming ERα under conditions of KMT2C loss.

Conclusions

One of the major goals of our study was to understand the selective advantage of KMT2C loss. To this end, we queried *KMT2C* mutation frequency at different stages of breast cancer progression and observed a ~7% mutation rate in primary breast cancer, a similar rate (~10%) in metastatic breast cancer and a much higher rate (~30%) in hormone-independent ER+ breast cancer. These data were in some contrast to the proliferative defects we observed upon KMT2C knockdown in ER+ cell lines growing in hormone-containing media. Therefore, we hypothesized that loss of KMT2C may confer a proliferation-independent selective advantage to breast cancer cells. We tested for the ability of KMT2C loss to promote transformation of MCF10A cells and also tested for the ability of KMT2C loss to promote increased metastases. However, we found that loss of KMT2C had no effect on either of those two processes.

Given the established role of KMT2C in ERα transcriptional signaling, we next wanted to investigate how loss of KMT2C affected cell response to estrogen-depleted conditions. We observed that KMT2C loss led to markedly

accelerated outgrowth of cells in the absence of hormone both in vitro and in vivo. These data were further supported upon finding that *KMT2C* mutation in primary breast cancer correlated with poorer outcomes in terms of overall survival and disease-free survival. Also, loss of KMT2C gene expression signatures impacted Luminal A/B breast cancer patient survival in the METABRIC data set and correlated with a more aggressive Luminal B subtype and poor outcomes. This is especially intriguing after considering the fact that the signature is derived from cells that were growing more slowly than control cells. In accordance, reduced KMT2C expression was significantly associated with poor outcome(202). Together, these data suggest that KMT2C loss is likely to have complex effects on the clinical behaviors of breast cancers.

In accordance with our data on KMT2C loss inhibiting proliferation, we do anticipate that under some circumstances KMT2C loss will slow estrogen-driven tumor growth and potentially be associated with favorable outcomes. In other contexts, loss of KMT2C may be permissive for tumors developing an estrogen-independent growth program, perhaps only in selected genotypes where another growth program is sufficiently active. Interestingly however, even in this context, our data would suggest ERα may be targetable as the shKMT2C-R cells were still sensitive to SERD-mediated ERα antagonism. Our analyses of these shKMT2C-R cells suggested that shKMT2C-R cells had large scale reprogramming of ERα such that it acquired thousands of new binding loci. The function of these new binding sites is not yet known, but the sensitivity of the cells to ERα inhibitors suggest there may be key oncogenic functions gained.

Indeed, we do see increased ERα binding to an AP-1 dependent *MYC* enhancer site suggesting that perhaps some of these new ERα binding sites may regulate the expression of known drivers of proliferation. In comparing these shKMT2C-R to other LTED models, we note some key differences. While LTED cells have been shown to upregulate expression ERα and activate the MAPK, PI3K and/or mTOR signaling pathways (194), we do not see ERα upregulation or pathway activation in our shKTM2C-R cells. Therefore, It will thus be imperative to further understand the growth requirements for different *KMT2C* mutated tumors to better utilize this biomarker in selecting the types of endocrine therapy most appropriate for this common clinical entity.

CHAPTER 4: DISCUSSION

In this dissertation, we present KMT2C as a candidate regulator of the ERα signaling pathway with key roles in maintaining the hormone sensitivity of ER+ breast cancer cells (Fig. 4.1).

KMT2C is frequently mutated in breast cancer

KMT2C is among the most frequently mutated genes in breast cancer with a mutation frequency of approximately 7% in primary breast cancer, 10% in metastatic breast cancer and as high as 30% when looking specifically at hormone-resistant ER+ breast cancer. This high mutation frequency places it among the top ten most frequently mutated genes in breast cancer. Moreover, *KMT2C* mutation and reduced KMT2C expression correlate with poorer patient prognosis in breast cancer(202). In this dissertation, we attempted to detail the selective advantage of its loss in breast cancer. We modeled KMT2C loss using both shRNA-mediated knockdown and CRISPR/Cas9-mediated knockout. We then used these models of KMT2C loss to assess for multiple hallmarks of cancer, including increased proliferative capacity, increased transformation of normal breast cells, and increased metastatic potential. We found that KMT2C loss does not overtly contribute to these hallmarks of cancer when modeled in breast epithelial cells suggesting that the selective advantage of its mutation may not be fully captured in these particular assays.

Our results were in some contrast to studies where KMT2C loss was modeled specifically in breast stem cells(110). In the breast stem cell context, KMT2C loss, when combined with PI3K pathway activation and p53 loss,

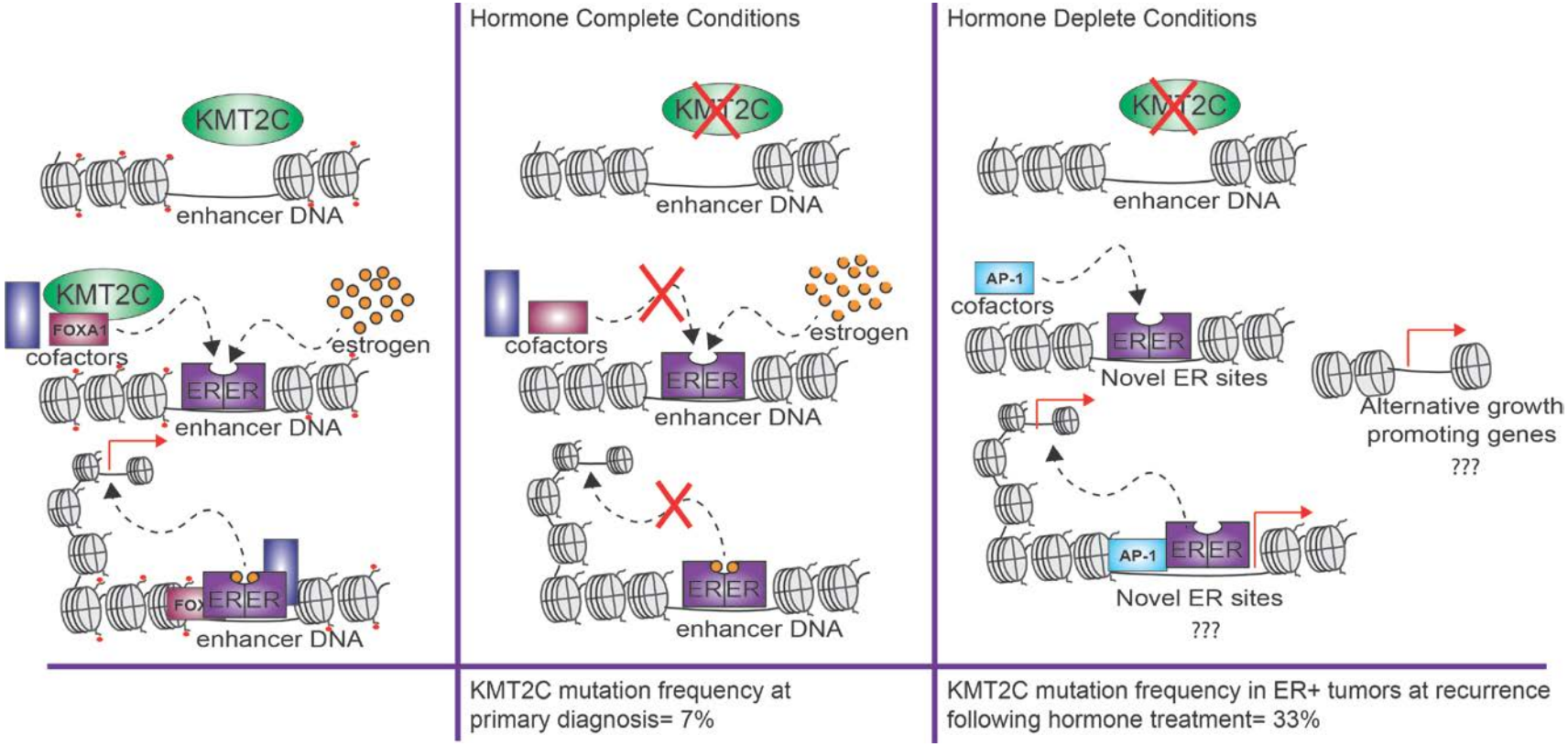

Figure 4.1 Conclusion. (Left) Within the ERα signaling pathway, KMT2C plays a key role in allowing proper cofactor binding to ERα, thereby allowing for full ERα transcriptional output in the presence of the ERα ligand, E2 at enhancers. (Middle) In the absence of KMT2C, E2 bound ERα is still able to bind to DNA but is no longer able to properly bind to cofactors, resulting in diminished expression of target genes. (Right) In hormone- deplete conditions, KMT2C loss may allow for ERα reprogramming that is partially AP-1 dependent. Under these conditions, KMT2C loss may also allow for alternate growth promoting pathways. Together, activation of these various pathways allows cells to grow in the absence of hormone. Correspondingly, *KMT2C* mutation frequency goes up specifically in hormone refractory ER+ breast cance

resulted in a more aggressive cancer and significantly reduced survival. Moreover, this study indicated that KMT2C loss may specifically result in HIF pathway activation as measured through gene expression profiling. It remains difficult, however, to appreciate how well this represents the true contributions of KMT2C loss in breast cancer, as the models do not recapitulate known breast cancer subtypes.

KMT2C as a mediator of specific ERα-dependent phenotypes

Interestingly, while we did not find that KMT2C loss conferred a proliferative advantage to any of the breast cancer cell line models tested, we did find that KMT2C loss resulted in a proliferative defect that was specific to ER+HER2- cells. These cells are all specifically dependent on the ERα signaling pathway. Given that KMT2C has already been implicated in regulating the transcriptional activity of a number of nuclear receptors(54, 83, 100, 103, 178, 179), we hypothesized that KMT2C may in fact be a key regulator of ERα activity. In keeping with this hypothesis, we found that the in vivo phenotypic consequences of MMTV-driven, mammary specific *Kmt2c* deletion were similar, though less severe, to those of MMTV-driven, mammary specific *Esr1* deletion. Indeed, both *Kmt2c* and *Esr1* mammary specific knockout mice both had significantly diminished secondary branching of the mammary epithelial ducts(156). This suggests that knockout of both these genes may be affecting the same signaling pathway.

Upon assessing the effects of KMT2C on gene expression in ER+ breast cancer cell lines, we found that ERα target genes were specifically downregulated in KMT2C knockdown cells. Our results were confimed by another study that published siRNA-mediated knockdown of KMT2C affected candidate ERα target gene expression(107). Similarly, in the mammary gland, deletion of *Kmt2c* resulted in reduced expression of multiple known ERα target genes. It should be noted that some of the genes that were affected in the mammary gland were not differentially expressed in the KMT2C knockdown cell lines. This discrepancy may reflect differences in ERα signaling in untransformed mammary glands versus breast cancer cells or may more specifically reflect differences in the ability of KMT2C to regulate ERα in these different contexts.

In keeping with the idea the Kmt2c may differentially regulate ERα in different contexts, we found that pregnancy specific, WAP-mediated deletion of *Kmt2c* actually did not reflect WAP-mediated deletion of *Esr1*(156). This suggests that the function of ERα during pregnancy-related mammary gland development is independent of Kmt2c. Follow-up on these studies would require a detailed understanding of when Kmt2c is expressed in the mammary glands and also a better understanding of the downstream ERα targets during both pubertal and pregnancy related mammary gland development.

KMT2C as a key regulator of ERα transcriptional activity

We next wanted to better understand the mechanism by which KMT2C directly influenced ERα transcriptional activity. Previous mediators of ERα

transcriptional activity have been shown to regulate its stability, expression, post-translation modifications and ability to bind DNA(153, 158). Therefore, we asked whether loss of KMT2C had any effect on these properties of ERα regulation. More specifically, we assayed for ERα stability in the presence of an HSP90 inhibitor, for ERα steady-state expression and for ERα phosphorylation at two sites that are known to correlate well with ERα activity. We found that none of these nodes were affected by KTM2C loss suggesting that KMT2C may not directly affect the ERα protein. It should be noted that there are multiple other ERα post-translation modifications that have been reported that we did not assay for. Therefore, it remains a possibility that KMT2C loss affects any of these modifications.

Upon initiating our analysis on how KMT2C affects ERα transcriptional activity, our most promising hypothesis was that KMT2C loss affects ERα binding to the DNA. This hypothesis was supported by previous studies suggesting that ERα binding occurs at sites of H3K4 methylation(166, 184, 187). Moreover, previous studies have also indicated that ERα binding to DNA is dependent on binding of FOXA1, and that FOXA1 binding to DNA is dependent on the presence of H3K4me1(107, 166, 167, 184). Together, this suggested that ERα binding to the DNA was preceded by the presence of H3K4me1. Therefore, we were especially surprised to find that ERα DNA binding was largely unaffected by KMT2C loss, even at sites where we had seen loss of H3K4me1. The vast majority of the ERα target gene expression changes could not be explained by a loss of ERα binding.

We then hypothesized that while ERα may still bind to DNA in the absence of KMT2C, it may be unable to recruit the coactivators necessary to mediate its full transcriptional activity. We indeed found significant loss of a number of known ERα cofactors in the through KMT2C knockdown. Most interesting among these binding partners was FOXA1. FOXA1 has, very recently, been implicated as being able to directly bind to KMT2C. Therefore, we speculate that FOXA1 plays a key role in mediating the link between KTM2C methyltransferase activity and ERα transcriptional activity. However, FOXA1 knockdown was suggested to prevent KMT2C recruitment, which would support a model in which the presence of FOXA1 serves to recruit KMT2C. Our data, on the other hand, would support a model where the lack of KMT2C at these target gene enhancers prevents the ability for FOXA1 and ERα to bind to each other, perhaps by preventing recruitment of FOXA1. One potential mechanism that could accommodate both studies would suggest that FOXA1 as a pioneer factor is able to bind to DNA and recruit ERα. However, the presence of KMT2C and perhaps its methyltransferase activity is necessary for the formation of the complete complex in which ERα and FOXA1 would bind to each other. Therefore, a better understanding of the exact interplay between KMT2C, FOXA1, ERα and other ERα cofactors is needed to understand the kinetics and the order for generating ERα-transcriptional complexes. Furthermore, it should be noted that we have yet to rule out the possibility that KMT2C is involved in transcriptional activation in ways that are independent of its methyltransferase activity which would require more directed

deletion of KTM2C methyltransferase activity instead of knockdown or knockout of the entire protein.

Key differences between KMT2 family members in breast cancer

After establishing that KMT2C is a key regulator of ERα transcriptional activity, we wanted to see if the other KMT2 family members had similar effects on ERα activity and ERα dependent phenotypes. IMPACT sequencing on metastatic breast cancer revealed mutations in *KMT2A* and *KMT2D*, in addition to the known mutations in *KMT2C*. *KMT2B*, *KMT2F*, and *KMT2G* are excluded from the IMPACT panel. In addition, the other members of the KMT2 family have also been implicated as regulators of nuclear receptors similar to KMT2C. In particular, it was shown that KMT2A was necessary for ERα-mediated transcription of *TFF1*, which is a very well established ERα target gene. It should be noted that KMT2F and KMT2G were removed from our analysis due to their lack of expression in a number of the breast cancer cell lines.

While the effects of KMT2C loss on proliferation were specific to the ER+HER2- cell types, loss of other KMT2 proteins affected the proliferation of all cell lines tested including those dependent on the HER2 pathway signaling and those that were triple negative. This suggests that while the other KMT2 family members may affect ERα signaling they may also affect other growth signaling pathways or might affect downstream mediators that are common among multiple growth signaling pathways. Also, it may suggest that the effects that their

loss have on ER+HER2- cells may be independent of ERα signaling and may instead affect more general aspects of transcription.

To assay whether the other KMT2 family members did effect ERα target gene signaling, we went on to look at KMT2C target gene expression in the presence of KMT2A/ B /D knockdown. Interestingly, we found that many of the KMT2C target genes were also affected with knockdown of the other family members. The effects of KMT2D were the most similar to those of KMT2C, which is not surprising given that these two proteins have a high degree of sequence similarity and are found to complex with the same coregulatory proteins. However, loss of KTM2D had significantly more severe effects on proliferation across multiple subtypes of breast cancer suggesting that it may have some significant non-redundant roles with KMT2C. These non-redundant functions of KMT2C and KMT2D have also been suggested in other cancer models where KMT2D and KMT2C were suggested to regulate H3K4me1 at different sites of the genome(85). Overall, we conclude that the different members of the KMT2 family do seem to have non-redundant roles and are seen to differentially regulate a panel of genes that are seen to be KMT2C target genes.

KMT2C loss is advantageous upon hormone deprivation

Given our data and the data of others establishing KMT2C as a key component of the ERα signaling pathway, we wanted to better understand how KMT2C loss affected the response of ER+ breast cancer cells to hormone therapy. As ER+ breast cancer is very dependent on estrogen-mediated

signaling, patients are often treated with hormone therapies that serve to reduce the amount of estrogen present in the body(112). Upon mimicking conditions of hormone deprivation in our control and KMT2C knockdown cell lines, we found that KMT2C loss resulted in a proliferative advantage as compared to control cells. KMT2C was the only member of the KMT2 family whose loss promoted insensitivity to hormone deprivation.

Upon further analysis of these KMT2C deficient, hormone-resistant cells, we were surprised to see that ERα acquired over 10,000 novel binding sites in and that these cells were still highly sensitive to ERα degradation by SERDs. Together, this suggests that unliganded ERα plays a key role in driving proliferation in these cells and supports the idea that perhaps *KMT2C* mutant tumors that are resistant to hormone therapy may continue to be sensitive to direct inhibition of ERα.

This KMT2C deficient model of hormone resistance can be used to study hormone resistance in general, and can be compared to other established LTED models. MCF7 LTED cells are known to grow out in hormone-deprived media; however, this typically occurs following 3-4 months of growth arrest. Mechanisms of LTED proliferation include upregulation of ERα expression, downregulation of ERα expression, MAPK pathway activation, and PI3K-mTOR pathway activation. Upregulated ERα has also been shown to participate in non-genomic signaling activity. Upon testing for the importance of these pathways in our KMT2C deficient model of LTED, we did not see evidence for any of these pathways playing a role. There was no notable increase in MAPK or PI3K-mTOR pathway

activation nor was there any change in ERα expression levels. In fact, we see a further downregulation of canonical ERα-target genes. Therefore, we propose that these KMT2C deficient cells utilize an alternate pathway for hormone-resistance that involves reprogramming of unliganded ERα. There, however, remains the possibility that other growth promoting pathways, unrelated to unliganded ERα, are still necessary. Also it remains a possibility that even though this unliganded ERα does bind to novel sites on the genome, it could still be participating in non-genomic activities that serve to promote the survival of these cells. These possibilities still need to be explored.

Another important point to note about these resistant cells is that we have not yet established whether they continue to be dependent upon KMT2C loss or if that simply allowed the cells to rapidly become insensitive to hormone deprivation. More specifically, it is possible that the chromatin state established by loss of KMT2C allowed cell to initially overcome hormone deprivation and after establishing that chromatin state, KMT2C loss is irrelevant. It is somewhat encouraging to note that KMT2C remains knocked down in the resistant cells and that they retain a H3K4me1 signature of KMT2C loss. However, to sufficiently answer this question, we need to re-express shRNA-resistant KMT2C in the resistant cells and see if this resensitizes them to hormone deprivation. Given the large size of KMT2C, this may not immediately be easy to accomplish.

The importance of context when understanding *KMT2C* mutation

Together, our results indicate that KMT2C loss has different effects in hormone-complete and hormone-deplete condition, inhibiting growth in the former while promoting rapid outgrowth in the latter. This highlights the importance of context when understanding the function of KMT2C as well as the selective advantage of KMT2C loss.

The importance of context becomes even more apparent when studying the effects of KMT2C loss across the various cancer types where it is frequently mutated. For instance, when KMT2C loss was modeled in the context of bladder cancer, it was found to form a tumor suppressive complex with p53(82). Furthermore, in the context of acute myeloid leukemia, it was proposed that loss of KMT2C promoted a more stem cell like gene expression signature(81) but there were no mechanistic details as to how that may be achieved. In the context of breast stem cells, loss of KMT2C was also shown to promote a more murine stem cell like gene expression signature possibly through activation of HIF signaling(110). Finally, in the context of pancreatic ductal adenocarcinoma (PDAC), loss of KMT2C was shown to suppress proliferation, similar to what we see in ER+ breast cancer cells in hormone-complete conditions. In accordance, PDAC patients with *KMT2C* mutations saw improved survival(203). Together, these studies show that KMT2C loss can affect multiple signaling pathways depending on the cell type and cancer model.

The importance of context when studying KMT2C loss may also be important in our efforts to better understand the selective advantage of its mutation in primary cancer. Again, *KMT2C* is very frequently mutated in primary

cancer and its mutation frequency is relatively evenly distributed across the different breast cancer subtypes. This could either mean that KMT2C loss has some selective advantage that is consistent across the various breast cancer subtypes or that in each subtype, KMT2C loss is contributing in a different manner. The results presented in this dissertation do not provide a sufficient explanation for the selective advantage of *KMT2C* mutations in primary breast cancer. However, tumors may undergo periods of low estrogen depending on the microenvironment allowing for KMT2C loss to provide a selective advantage. The universality of this finding, however, has yet to be confirmed.

Future directions

In this report we show that KMT2C loss has effects on a large number of ERα target genes. In terms of mechanism, our analysis is largely restricted to H3K4me1, H3K27ac and ERα ChIP sequencing, leaving the layer of intra-chromatin dynamics largely untouched. Therefore, an important avenue of follow-up to the work presented in this dissertation is a better understanding of the three-dimensional chromatin architecture and how that changes in the presence of KMT2C loss. In this dissertation, we used proximity to determine whether we would expect any given enhancer region to have effects on any given promoter. However, it is possible for regions of DNA that are further apart than our limits to interact with one another based on how the chromatin folds. In order to look at the three-dimensional chromatin architecture, we can use sequencing methods such as Hi-C(204) that preserve the chromatin architecture and look for long-

range chromatin interactions. Moreover, in our efforts to better understand KMT2C-mediated regulation of ERα activity, we would want to follow-up on our hypothesis that KTM2C loss prevents ERα cofactor recruitment. This would require large-scale ChIP sequencing of multiple ERα coregulatory proteins, namely MCM4, TRPS1 and FOXA1.

In our continued efforts to understand the more general role that KMT2C may have in transcriptional regulation and beyond, we are interested in better understanding the functions of the numerous other KMT2C domains. Aside from methyltransferase activity, KMT2C is able to recognize and bind to H3K4me1, through its PHD domain, and can potentially directly interact through DNA with its zinc finger motifs. Given that KMT2C was found to be a part of the FOXA1 binding complex(107), KMT2C may have other relevant scaffolding functions.

In terms of KMT2C loss in cancer, we want to begin to shift our focus to KMT2C loss in primary cancer. To this end, we want to cross our mice with mammary-specific deletion of *Kmt2c* to mice to a variety of mammary carcinoma models. These models include mammary specific *Pi3kca* mutation, mammary specific *Neu* overexpression, and DMBA-induced carcinogenesis(205). One major issue with mammary specific mouse models is that they do not completely recapitulate the characteristics of human breast cancer and there is no effective model for hormone-dependent breast cancer. Nonetheless, we may be able to uncover whether KMT2C plays more generalized tumor suppressive roles and may be able to confirm previous studies that indicate that KMT2C can promote a more stem cell like gene expression signature(110).

Another important aspect to understanding the effects of KTM2C loss in cancers is to make sure we are properly modeling the types of mutations we see in patients. *KMT2C* has a number of missense mutations that tend to be clustered within its PHD domains, suggesting a specific function for this particular domain. It will be productive to model these missense mutations to determine whether there are any specific gains of function. Also, despite the fact that the truncated mutations result in loss of enzymatic activity, there are still a number of domains that could be preserved in expressed truncated proteins. Therefore, by strictly modeling knockdown or knockout of KMT2C we may not be capturing the function, if any, of the truncated proteins. It would also be of value to model the exact truncation mutations, determine if the truncated proteins are expressed, and determine whether they carry out any specific functions.

Finally, in terms of following up on KMT2C-mediated resistance to hormone-therapy, we want to more specifically delineate the dependencies of the resistant cells. Firstly, we are most interested in better understanding the function of ERα binding, if any, at its novel binding sites. Beyond the continued ERα dependencies of these cells, we also want to determine if there are other growth-promoting pathways that have since been activated and are contributing to the hormone-independent proliferation.

CHAPTER 5: METHODS AND MATERIALS

IMPACT sequencing of breast cancer cell lines

DNA was extracted from frozen cell pellets of Cama-1, MDA-MB-361, MCF7, MDA-MB-231, MCF10A, T47D, BT474 and SKBR3 cells using the PureLink Genomic DNA Mini Kit (Invitrogen, Cat. No. K1820-01). DNA was quantified using a Nanodrop Spectrophotometer and approximately 7ug were sent for MSKCC-IMPACT sequencing. The MSKCC IMPACT sequencing panel includes 341 genes, including *KMT2C*, which have been identified as being important for cancer cells. MSKCC-IMPACT uses the Illumina HiSeq platform with a coverage of approximately 500X-1000X per sample. This sequencing data was used to determine whether *KMT2C* harbored any deleterious mutations in any of the cell lines used for experiments.

Generation of short hairpin RNAs (shRNAs)

For shRNA knockdown, we used the miR-E backbone, as previously described(182). Target shRNA sequences in the miR-E backbone were ordered as Ultramers from IDT. Hairpin ultramers were then amplified using the following primers: For 5'-TACAATACTCGAGAAGGTATATTGCTGTTGACAGTGAGCG-3', Rev 5'-TTAGATGAATTCTAGCCCCTTGAAGTCCGAGGCAGTAGGCA-3'. Amplified ultramer and lentiviral SPEG vector were digested with EcoR1-HF and Xho1 restriction enzymes and ligated using an insert to vector molar ratio of 5:1. Ligation products were then transformed into OneShot Stbl3 competent E. Coli cells (ThermoFisher Scientific Cat. No. C737303). Proper insertion was verified

by Sanger sequencing using a specific sequencing primer: 5'-TGTTTGAATGAGGCTTCAGTAC-3'.

List of shRNA sequences used

shKTM2C#1- 5'-TTTGTCTTTGAAACACATCTGC-3'

shKMT2C#2- 5'- TCTTTGTCTATCTACCTCCTGC-3'

shRNA Transfection

HEK293CT packaging cells (a kind gift from P. Chi) were plated at 2.2 × 10^7 cells/ tissue culture dish (10cm in diameter) and transfected them with 4.5μg of the SGEP lentiviral vector (encoding shRenilla, or target shRNAs), 4.5 μg of psPAX2 and 1μg of pVSVG with X-tremeGENE HP (Roche) diluted in 500uL of Reduced Serum Opti-MEM (Invitrogen, Cat. No. 31985070) and added to the cells after a 20 minute incubation at room temperature. Conditioned medium containing recombinant lentivirus was collected from the transfected HEK293CT cells 48hrs after transfection, filtered through non-pyrogenic filters with a pore size of 0.45 μm (Millipore) and applied immediately to target cells, which had been plated 18 h before infection at a density of 3 × 10^6 cells/tissue culture dish (10cm in diameter). Polybrene (Sigma, No. AL-118) was added to a final concentration of 8 μg/ml, and supernatants were incubated with cells for 12 h. Conditioned media was reapplied 24 hrs after first application for a second infection. After two rounds of infection, cells were placed in fresh growth medium

and cultured as usual. Selection with 5-7 μg/ml puromycin was initiated 48 h after infection to select for shRNA expressing cells.

Generation of MCF7 KMT2C-HA cells

The following guide RNA sequence targeting KMT2C exon 60 was selected using the Optimized CRISPR Design tool (http://crispr.mit.edu(206)): 5'-AGCCGCCCGCTGAGCTAGCA-3'.

DNA oligonucleotides were purchased from IDT and cloned into the px458-GFP vector(207). For homologous recombination, we purchased a custom IDT Ultramer 200bp repair template with the HA sequence (5'-TACCCATACGATGTTCCAGATTACGCT-3') directly upstream the KMT2C stop codon. This 200bp fragment contained two silent mutations, one to remove the PAM site (TCC>TTC) and another to introduce a HincII restriction enzyme (GTGAAC>GTTAAC) site upstream of the HA tag. The 200bp fragment also contained 87bps of homology to KMT2C exon 60 upstream of the HA tag and 83bp of homology to the KMT2C 3"UTR downstream of the stop codon. 12ug of the targeting construct and 240nM repair template were nucleofected into MCF7 cells using the Lonza Nucleofector V kit and Program P020 on the Nucleofector device. Nucleofected cells were single cell sorted based on GFP positivity 48-hours following nucleofection. Clones were screened for the presence of successful HA insertion by *KMT2C* exon 60 PCR and subsequent restriction enzyme digest with HincII. Positive clones show two digested products while negative clones only had a single undigested band. A single positive clone,

called KMT2C-HA, containing the HA coding sequencing downstream KMT2C was selected to carry out further studies. As a control, we used one of the negative clones from the screening process, called KMT2C crispr control.

Proliferation Assays

500-1000 cells in 200uL of media were seeded per well of a 96-well plate with 6 replicates per sample. If necessary, cells were treated the following day, (Day 0). On days when the plates were measured, 25uL of Resazurin (R&D Systems AR002) was added per well followed by 4 hr incubation at 37°C. Plates were then read using a SpectraMax M5 (Molecular Devices) and results were analyzed with Softmax Pro 6.2.2 software with an Endpoint Readtype (Excitation: 560nM, Emission: 590nM). Results were normalized to blank media with no cells.

For long-term estrogen deprived proliferation assay, 5000 cells were seeded in 200uL of charcoal stripped media. Media was replenished once or twice a week during the course of the proliferation assay.

Generation of CRISPR-CAS9nickase mediated deletion of KMT2C

The following guide RNA sequences targeting KMT2C exon 6 were selected using the Optimized CRISPR Design tool (http://crispr.mit.edu(206)): gRNA1 5'-CTAGTGACCACTCCACACAACGG-3', gRNA2 5'-AGAACCATTGTTAGTGAACGTGG-3'. DNA oligonucleotides were purchased from IDT and cloned into pX335-GFP vector(207) to generate targeting constructs that were subsequently co-transfected in an equimolar ratio into MCF7

cells using Lipofectamine 2000 (Thermo Fischer Scientific). 72 h after transfection, cells were sorted using a MoFlo cell sorter (Beckman Coulter) for cells expressing Cas9 nickase (GFP-positive cells) and left to recover for 1 week before sorting for single cells and allowing colonies to form. Indels within the KMT2C exon 6 were detected using the Surveyor Mutation Detection kit (IDT) and confirmed by subcloning and sequencing of KMT2C exon 6. One MCF7 clone showing knockout of all 2/3 KMT2C alleles and two MCF10A clones showing complete knockout of KMT2C were used for subsequent analysis.

Generation of KMT2C$^{F/F}$ mice

Mice with the *Kmt2c* floxed (F) allele were generated by Biocytogen (Worcester, MA) in a C57BL/6 background. ES cells were verified using Southern blotting.

Breeding scheme to generation MMTVCre^{+}*Kmt2c*$^{F/F}$ knockout mice and WAPCre^{+}*Kmt2c*$^{F/F}$ knockout mice

Kmt2c$^{F/F}$ mice were bred to transgenic mice expressing the Cre enzyme in the mammary tissue under the control of the MMTV long-terminal repeat (mice were a kind gift from S. Lowe) or under the control of the WAP promoter (mice were a kind gift from M. Jasin). Cre-mediated excision of the *Kmt2c* gene was verified by genotyping PCR of DNA extracted from toe clips or mammary fat pads. Genotyping for the MMTV-cre transgenes were done using generic Cre primers, forward (oIMR1084): 5'- GCGGTCTGGCAGTAAAAACTATC-3', reverse (oIMR1085): 5'- GTGAAACAGCATTGCTGTCACTT- 3', as well as internal

positive controls, forward (oIMR7338): 5'-CTAGGCCACAGAATTGAAAGATCT-3, reverse (oIMR7339): 5'-GTAGGTGGAAATTCTAGCATCATCC-3'. Genotyping for the WAP-cre transgene was done using WAP-cre specific primers, forward (oIMR0575): 5'- TAGAGCTGTGCCAGCCTCTTC-3' and reverse (oIMR1085): 5'-GTGAAACAGCATTGCTGTCACTT-3', as well as internal positive controls, forward (oIMR8744) 5'-CAAATGTTGCTTGTCTGGTG-3' and reverse (oIMR8745) 5'-GTCAGTCGAGTGCACAGTTT-3'. *Kmt2c* floxed allele was genotyped using A1 primers, forward: 5'-GCAGAATCAAGGAGCTGTCTG-3', reverse: 5'-TGCCTTGAGGTCAACGTACAATTG-3'. Cre mediated excision of KMT2C exon 3 was verified by the presence of a 3'LoxP band at 372bp using the following primers, forward: 5'-GGAGCTGTCTGTTCAAGTATTTAGC-3', reverse: 5'-GTGTGCCTAGGTATCCACAGAG-3'. For MMTV experiments, 12-week and 8-week old virgin females were used. All mice were housed in the MSKCC animal facility. For WAP experiments, females were sacrificed after at least their second pregnancy.

Mammary gland whole mounts

Mammary glands were dissected, mounted on glass slides, and fixed and stained in carmine alum solution as described for whole-mount analysis(208). Briefly, slides were first immersed in Carnoy's fixative (EtOH: choloform: acetic acid, 6:3:1) for 4 hours at room temperature. Then, slides were immersed into 70% ethanol for 15 minutes. 70% ethanol was gradually changed to distilled water by replacing half the volume with distilled water and incubating for 5 minutes, three

times. Slides were then placed in pure distilled water for 5 minutes. Slides were then immersed in carmine alum solution (1g carmine natural red, 2.5g aluminum potassium sulfate in 500mL H2O) overnight. Slides were then immersed in 70%, 95% and 100% ethanol for 5 minutes each to progressively dehydrate the glands. Mammary fat pads were then cleared by immersing the slides in xylene overnight. Glands were mounted on the slides with Permount (Fisher Scientific) and sldies were scanned using Axio Scan.Z1 (Zeiss). Quantification of the area occupied by the mouse mammary epithelial cells and the area of the whole mouse mammary gland whole mount was done using ImageJ.

Immunohistochemistry of mammary glands

Tissue were fixed in 10% neutral buffered formalin for 48 hours, were processed in xylene and ethanol, embedded in paraffin, sectioned at 4 μm thickness, and stained with hematoxylin and eosin or with an antibody against the progesterone receptor (Santa Cruz sc-538) and then detected with SignalStain Boost IHC Detection Reagent (Cell Signaling, 8114). Slides were examined by a board certified pathologist (S. Monette, MSKCC).

RNA Sequencing

RNA was extracted from MCF7 and T47D cells stably expressing shRenilla, shKMT2C #1 and shKMT2C #2 using the RNeasy Mini kit (Qiagen) according to the manufacturer's protocol. RNA size, concentration and integrity were analyzed using Agilent 2100 Bioanalyzer. Libraries were generated using Illumina's

TruSeq RNA sample Prep Kit v2, following the maufacturer's protocol. Sequencing was done on the HiSeq2500 sequencer as 50-bp paired-read with 30-40 million reads per sample by the NYU Genome Technology Center. The raw data was analyzed using Basepair software (http://www.basepairtech.com/) with pipelines including the following steps: Raw reads were aligned to the transcriptome derived from UCSC genome assembly hg19 using STAR(209) with default parameters. Read counts for each transcript were measured using featureCounts(210). Differentially expressed genes were determined using DESeq2(211) and a cut-off of 0.05 on adjusted p-value (corrected for multiple hypotheses testing) was used for creating lists and heatmaps, unless otherwise stated.

Gene Set Enrichment Analysis (GSEA)

GSEA was performed using the downloaded javaGSEA application (http://software.broadinstitute.org/gsea/downloads.jsp). Files were created as previously described. Hg19 was used as the chip file.

Quantitative Reverse Transcriptase Polymerase Chain Reaction

Cells were washed once with cold PBS and scraped off the plate with a rubber policeman. The cell suspension was briefly centrifuged to pellet cells, PBS was removed, and the cell pellet was stored at −80 °C until lysis. RNA was extracted from cells using the RNeasy Mini kit (Qiagen) according to the manufacturer's protocol. cDNA was synthesized with 1 µg of RNA from each sample using the

qScript cDNA SuperMix (Quanta Biosciences, No. 95048) according to the manufacturer's protocol. Synthesized cDNA was diluted with one volume of DEPC-treated water, and 2 µl of the mixture was added to TaqMan PCR Master Mix (Applied Biosystems) along with primers. Relative quantification for each mRNA was performed using the comparative CT method with a ViiA 7 Real-Time PCR system (Applied Biosystems, No. 4364340). Samples were run in triplicate, and mRNA levels were normalized to those of *RPLP0* or *Gapdh* for each reaction. TaqMan primers were all purchased from Applied Biosystems.

List of Primers used for qRTPCR

AGR3	Hs00411286_m1
BANK1	Hs01009378_m1
CA2	Hs01070108_m1
CACNA1G	Hs00367969_m1
CAV1	Hs00971716_m1
CD44	Hs01075861_m1
CXCL12	Hs03676656_mH
ESR1	Hs00174860_m1
GDNF	Hs01931883_s1
IGFBP4	Hs01057900_m1
KMT2A	Hs00610538_m1
KMT2B	Hs00207065_m1
KMT2C	Hs01005521_m1
KMT2D	Hs00231606_m1
KRT13	Hs02558881_s1
PEG10	Hs00248288_s1
PGR	Hs01556707_m1
RPLP0	Hs99999902_m1
SERPINA1	Hs00165475_m1
SOX5	Hs00753050_s1
TFF3	Hs00902278_m1
TMPRSS3	Hs00917537_m1
Kmt2c	Mm01156942_m1
Pgr	Mm00435628_m1
Gapdh	4331182

Western Blotting

Cells were washed once with cold PBS and scraped off the plate with a rubber policeman. The cell suspension was briefly centrifuged to pellet cells, PBS was removed, and the cell pellet was stored at −80 °C until lysis. For cell lysis, pellets were resuspended in a non-denaturing lysis buffer (Cell Signaling Technology) or RIPA buffer (Pierce, No. 89901) supplemented with Halt protease and phosphatase inhibitors (Thermo Scientific, No. 1861281) and were sonicated briefly for 2 mins at 30sec on/ 30 sec off. Lysates were cleared by centrifugation at 14,000g for 10 min, and protein concentration was determined using the BCA kit (Pierce No. 23223), which measures the reduction of Cu2+ to Cu1+ by protein in an alkaline medium. For each sample, 30 µg of protein lysate was loaded onto 4–12% SDS-PAGE minigels (Invitrogen) for electrophoresis. Protein was then transferred onto nitrocellulose membranes for immunoblotting.

List of antibodies used for Western blotting

against H3K4me1(Cell Signaling 5326), H3K4m3(Cell Signaling 9727), Histone H3(Cell Signaling 9715), ERα(Santa Cruz 7207), PR(Cell Signaling 8757), β-actin(Cell Signaling 4970).

Chromatin Immunoprecipitation followed by DNA sequencing

2×10^7 MCF7 stably expressing either shRenilla, shKMT2C #1 or shKMT2C #2 cells were fixed with 1% formaldehyde for 10 minutes at room temperature and the reaction was subsequently quenched using 125mM glycine. Cells were then lysed using LB1 (50mM HEPES-KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5%NP-40, 0.25% Triton-X100), incubated at 4°C rotating and spun down. The pellet was then lysed with LB2 (10mM Tris-HCl, pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA) incubated at 4°C rotating and spun down. The pellet was then lysed with LB3 (20mM Tris-HCl, pH 8.0, 150mM NaCl, 2mM EDTA, 0,5mM EGTA, 1% Triton-X100) and sonicated leading to a DNA average size of 200bp. 1ug of target-specific antibodies was added to the samples and incubated overnight at 4°C. The complexes were purified using ChIP grade protein-A/G magnetic beads (Thermo Scientific QB210263) followed by elution from the beads, reverse cross-linking and treatment with RNAse and Proteinase K. DNA was purified using PCR purification columns (QIAGEN). ChIP-seq libraries were prepared using 5ng of DNA and KAPA Hyper Prep Kit for Illumina, according to the manufacturer's protocol. Libraries were validated using the Agilent 2200 Tapestation and KAPA library quantification qPCR assay and sequenced on a HiSeq2500 sequencer as 50-bp single-end reads by the NYU Genome Technology Center.The raw data was analyzed using Basepair software (http://www.basepairtech.com/) with pipelines including the following steps: The raw fastq data were trimmed using trim_galore to remove low-quality ends from reads (quality < 15) and adapter sequences. The trimmed data was aligned using Bowtie2(212) to UCSC genome assembly hg19. Duplicate reads were removed

using Picard and bigwig files were created for visualization. Peaks were identified with Macs1.4(213) and transcription factor binding motifs were detected with Homer. Motifs were identified as enriched over 48,246 randomly selected background sequences with matched %GC content. Peaks overlapping with Satellite repeat regions were discarded and remaining filtered peaks were annotated using custom scripts based on UCSC Flat data(214), where peaks between -2500 bp to 2500 bp of a transcription start site were marked as Promoter, the overlapping gene body were marked as Genebody and the rest were marked as Intergenic. For intergenic peaks, a gene was considered a target if it was within 1Mb of the peak.

Chromatin Immunoprecipitation followed by quantitative polymerase chain reaction

Cells were fixed with 1% formaldehyde for 10 minutes at room temperature and the reaction was subsequently quenched using 125mM glycine. Fixed cells were lysed using the Low Salt Wash Buffer (0.1% SDS, 1%Triton-X100, 2mM EDTA, 20mM Tris-HCl pH 8.1, 150mM NaCl). Lysates were sonicated for 30 minutes (High, 30 sec on and 30 sec off), spun down, and cleared lysates were used for immunoprecipitation. Complexes were purified using target-specific antibodies. Complexes were then washed with Low Salt Wash Buffer, High Salt Wash Buffer (0.1% SDS, 1% Triton-X100, 2mM EDTA, 20mM Tris-HCl pH 8.1, 500mM NaCl), LiCl Wash Buffer (1% Deoxycholate, 1% NP-40, 1mM EDTA, 10mM Tris-HCl, pH 8.1, 250mM LiCl), and TE Wash Buffer (1X TE Buffer with 5uM NaCl). Samples

were eluted off beads after 1 hr. at 65°C and then treated with RNAse for 30 min. at 37°C and Proteinase K overnight at 55°C. DNA was then purified using the PCR Purification Kit (Qiagen). 2 µl of the purified DNA was added to SYBR Green PCR Master Mix (Applied Biosystems) along with primers. Relative quantification for each sample was performed using the comparative CT method with a ViiA 7 Real-Time PCR system (Applied Biosystems, No. 4364340). Samples were run in triplicate, and DNA levels were normalized to those of 1-10% input for each sample.

List of antibodies used for ChIP

H3K4me1 (abcam 8895) and H3K27ac (abcam 4729)-

H3K4me1 (abcam 8895), H3K27ac (abcam 4729) and ERα (Santa Cruz HC-20)-

List of primers used for ChIP-qPCR

ChIP primers are as follows *CA2* enhancer forward: 5'-GCAATCCTGAGAAACTGCAA-3', reverse: 5'-GCTCCTGGCCTCAAACTATC-3', *PGR* enhancer forward: 5'- GAGGGCTGTTGAAACAAACA-3', reverse: 5'-TGCTGAGATCACACCTAGACAA-3', *AGR2* enhancer forward: 5'-ACGAAGCCTGCTT CTGAAC-3', reverse: 5'-GTTTACAAGACATCAAACAACATGA-3'.

Rapid immunoprecipitation mass spectrometry of endogenous protein complexes (RIME)

RIME was preformed as previously described(109, 188). Briefly, MCF7 cells were grown in SILAC Protein Quantification Kit DMEM:F12 (Thermo Fischer Scientific 88439) supplemented with dialyzed FBS (Life Technologies 88440) with either L-Lysine-2HCl (Thermo Scientific 88429) and L-Arginine-HCl (Thermo Scientific 88427) for light-labeled media or L-Arginine-HCl 13C6 (Cambridge Isotope Laboratories 13E-430) and L-Lysine-2HCl 13C6, 15N2 (Cambridge Isotope Laboratories 14F-832) for heavy-labeled media. 60 million cells total with 30 million light labeled shRenilla cells and 30 million heavy labeled shKMT2C#2 cells was used for each replicate. Cells were lysed with LB1, LB2 and LB3. 10ug of antibody against ERα (Santa Cruz sc-543) was used for each replicate. There were three replicates done in total. Immunoprecipitated samples were washed with RIPA wash buffer as previously described and digested overnight with trypsin, peptides desalted using C18 zip tips, and then dried by vacuum centrifugation. Each sample was reconstituted in 10 uL 0.1% (vol/vol) formic acid and 4 uL analyzed by microcapillary liquid chromatography with tandem mass spectrometry using the NanoAcquity (Waters) with a 100-μm-inner-diameter × 10-cm- length C18 column (1.7 um BEH130, Waters) configured with a 180-μm × 2-cm trap column coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). Peptides were eluted with a linear gradient of 0-35% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 150 mins with a flow rate of 300 nL/min. The QE Plus was operated in automatic, data-dependent MS/MS acquisition mode with one MS full scan (380–1800 m/z) at 70,000 mass resolution and up to ten concurrent MS/MS scans for the ten most

intense peaks selected from each survey scan. Survey scans were acquired in profile mode and MS/MS scans were acquired in centroid mode at 17500 resolution and isolation window of 1.5 amu and normalized collision energy of 27. AGC was set to 1 x 10(6) for MS1 and 5 x 10(4) and 100ms IT for MS2. Charge exclusion of unassigned and greater than 6 enabled with dynamic exclusion of 15s. All MS/MS samples were analyzed using MaxQuant (Max Planck Institute of Biochemistry, Martinsried, Germany; version 1.5.3.3) at default settings with a few modifications. The default was used for first search tolerance and main search tolerance: 20 and 4.5 ppm, respectively. Labels were set to Arg6 and Lys8. MaxQuant was set up to search the reference human proteome database downloaded from Uniprot on September 6, 2016. Maxquant performed the search assuming trypsin digestion with up to two missed cleavages. Peptide, Site, and Protein FDR were all set to 1%. One unique peptide was required for high-confidence protein identifications and a minimum ratio count of two peptides (one unique and one razor) were required for SILAC ratio determination. The following modifications were used as variable modifications for identifications and included for protein quantification: Oxidation of methionine, acetylation of the protein N-terminus, phosphorylation of serine, threonine and tyrosine residues.

METABRIC patient data acquisition

1209 Luminal A and B breast cancer patients with Illumina HumanHT-12 V3 expression beadchip array and complete clinical annotation from the METABRIC dataset were included in the analysis. Patient data was collected retrospectively.

KMT2C derived gene expression signature

ssGSEA(215) was used to compute a set of different scores based on *in vitro* derived shKMT2C expression signatures. For each sample independently, an ssGSEA score was calculated after gene expression levels were rank normalized and ordered, using the empirical cumulative distribution functions of genes in the signature and the reaming genes as previously described(215-217) using the bioconductor package GSVA(218). Patients were stratified into high, low and intermediate scorers using population quartiles for the numerical ssGSEA score.

Xenograft studies

We obtained 6- to 8-week-old nu/nu athymic BALB/c female mice from Envigo Laboratories and maintained them in pressurized, ventilated caging. All studies were performed in compliance with institutional guidelines under an institutional animal care and use committee (IACUC)-approved protocol (MSKCC 12-10-016). Xenograft tumors of MCF7 cells stably expressing either shRenilla or shKMT2C#2 were established in nude mice by subcutaneously implanting 0.72 mg of sustained-release 17β-estradiol pellets with a 10-g trocar into one flank and injecting 1 × 107 cells suspended 1:1 in reconstituted basement membrane (Matrigel, Collaborative Research) into the opposite flank 3 d later. We removed 17β-estradiol pellets from the mice when tumors reached a size of ~100 mm3. Tumor dimensions were measured with vernier calipers, and tumor volume was calculated (π/6 × larger diameter × (smaller diameter)2). In this study, no blinding

of investigators or randomization of mice was performed. On the basis of our previous work measuring the variability in size and growth of MCF7 xenografts, we estimated that ten mice per group would allow us to detect differences in tumor size of >200 mm3.

MCF10A colony formation assay

MCF10A colony formation assay was performed using 8-well chamber slides. 45uL of matrigel was spread evenly on each slides and allowed to solidify for 30 minutes at 37°C. 5000 MCF10A cells (either parental, CRISPR control or KMT2C knockout) in 400uL of overlay media were used per well. Overlay media consisted of DMEM:F12, 2% Horse Serum, 5ng/mL EGF and 2% matrigel. Wells were left in 37°C for over the course of the experiments. Slide images were taken using a Zeiss Widefield Microscope. Each cell line was tested in triplicate.

Tail Vein Xenografts

1 million MCF7 cells (either stably expressing shRenilla or shKMT2C) were injected into the tail veins of 6-8 week old nu/nu athymic BALB/c female mice from Envigo Laboratories. 0.72mg b-estradiol pellets were planted into the “with estrogen” groups 3 days prior to cell injection. Mice were imaged once a week starting at 3 weeks post injected. For imaging, mice were retro-orbitally injected with 75mg/ kg D-luciferin in PBS prior to acquiring bioluminescence images.

Study Approval

All mice were treated with procedures approved by the Institutional Animal Care and Use Committee. Access to the METABRIC(219) cohort was obtained with consent from the MSKCC institutional review board for anonymized evaluation of genomic associations.

REFERENCES

1. Ball P (2003) Portrait of a molecule. *Nature* 421(6921):421-422.
2. Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* 184(4139):868-871.
3. Wisniewski JR, Zougman A, Kruger S, & Mann M (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol Cell Proteomics* 6(1):72-87.
4. Vicent GP*, et al.* (2011) Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. *Genes Dev* 25(8):845-862.
5. Weintraub H & Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. *Science* 193(4256):848-856.
6. Arents G & Moudrianakis EN (1993) Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. *Proceedings of the National Academy of Sciences of the United States of America* 90(22):10489-10493.
7. Arents G, Burlingame RW, Wang BC, Love WE, & Moudrianakis EN (1991) The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proceedings of the National Academy of Sciences of the United States of America* 88(22):10148-10152.
8. Luger K, Mader AW, Richmond RK, Sargent DF, & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389(6648):251-260.
9. Cheung P, Allis CD, & Sassone-Corsi P (2000) Signaling to chromatin through histone modifications. *Cell* 103(2):263-271.
10. Marmorstein R (2001) Protein modules that manipulate histone tails for chromatin regulation. *Nat Rev Mol Cell Biol* 2(6):422-432.
11. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128(4):693-705.
12. Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* 403(6765):41-45.
13. Turner BM (2002) Cellular memory and the histone code. *Cell* 111(3):285-291.
14. Gregory PD, Wagner K, & Horz W (2001) Histone acetylation and chromatin remodeling. *Exp Cell Res* 265(2):195-202.
15. Morales V & Richard-Foy H (2000) Role of histone N-terminal tails and their acetylation in nucleosome dynamics. *Molecular and cellular biology* 20(19):7230-7237.
16. Wolffe AP & Hayes JJ (1999) Chromatin disruption and modification. *Nucleic acids research* 27(3):711-720.
17. Calo E & Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Molecular cell* 49(5):825-837.
18. Schneider R, Bannister AJ, & Kouzarides T (2002) Unsafe SETs: histone lysine methyltransferases and cancer. *Trends Biochem Sci* 27(8):396-402.

19. Bannister AJ, Schneider R, & Kouzarides T (2002) Histone methylation: dynamic or static? *Cell* 109(7):801-806.
20. Santos-Rosa H*, et al.* (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419(6905):407-411.
21. Shi X*, et al.* (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442(7098):96-99.
22. Li H*, et al.* (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442(7098):91-95.
23. Lee J, Thompson JR, Botuyan MV, & Mer G (2008) Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nature structural & molecular biology* 15(1):109-111.
24. Nielsen PR*, et al.* (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* 416(6876):103-107.
25. Xu C*, et al.* (2010) Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proceedings of the National Academy of Sciences of the United States of America* 107(45):19266-19271.
26. Couture JF, Collazo E, & Trievel RC (2006) Molecular recognition of histone H3 by the WD40 protein WDR5. *Nature structural & molecular biology* 13(8):698-703.
27. Cohen I, Poreba E, Kamieniarz K, & Schneider R (2011) Histone modifiers in cancer: friends or foes? *Genes Cancer* 2(6):631-647.
28. Sassone-Corsi P*, et al.* (1999) Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science* 285(5429):886-891.
29. Soloaga A*, et al.* (2003) MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *EMBO J* 22(11):2788-2797.
30. Langan TA*, et al.* (1989) Mammalian growth-associated H1 histone kinase: a homolog of cdc2+/CDC28 protein kinases controlling mitotic entry in yeast and frog cells. *Molecular and cellular biology* 9(9):3860-3868.
31. Boggs BA, Allis CD, & Chinault AC (2000) Immunofluorescent studies of human chromosomes with antibodies against phosphorylated H1 histone. *Chromosoma* 108(8):485-490.
32. Shroff R*, et al.* (2004) Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Curr Biol* 14(19):1703-1711.
33. Brownell JE*, et al.* (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84(6):843-851.
34. Kuo MH & Allis CD (1998) Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 20(8):615-626.
35. Hildmann C, Riester D, & Schwienhorst A (2007) Histone deacetylases--an important class of cellular regulators with a variety of functions. *Appl Microbiol Biotechnol* 75(3):487-497.
36. Lee KK & Workman JL (2007) Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* 8(4):284-295.

37. Shahbazian MD & Grunstein M (2007) Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem* 76:75-100.
38. Loenen WA (2006) S-adenosylmethionine: jack of all trades and master of everything? *Biochem Soc Trans* 34(Pt 2):330-333.
39. Di Lorenzo A & Bedford MT (2011) Histone arginine methylation. *FEBS Lett* 585(13):2024-2031.
40. Margueron R & Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* 469(7330):343-349.
41. Towbin BD, *et al.* (2012) Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* 150(5):934-947.
42. Schuettengruber B, Martinez AM, Iovino N, & Cavalli G (2011) Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 12(12):799-814.
43. Shilatifard A (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem* 75:243-269.
44. Poux S, Horard B, Sigrist CJ, & Pirrotta V (2002) The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. *Development* 129(10):2483-2493.
45. Herz HM, *et al.* (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & development* 26(23):2604-2620.
46. Zhang X, Novera W, Zhang Y, & Deng LW (2017) MLL5 (KMT2E): structure, function, and clinical relevance. *Cell Mol Life Sci*.
47. Patel A, Vought VE, Dharmarajan V, & Cosgrove MS (2008) A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem* 283(47):32162-32175.
48. Wu L, *et al.* (2013) ASH2L regulates ubiquitylation signaling to MLL: trans-regulation of H3 K4 methylation in higher eukaryotes. *Molecular cell* 49(6):1108-1120.
49. Rao RC & Dou Y (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* 15(6):334-346.
50. Deng C, *et al.* (2013) USF1 and hSET1A mediated epigenetic modifications regulate lineage differentiation and HoxB4 transcription. *PLoS Genet* 9(6):e1003524.
51. Lee JE, *et al.* (2013) H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* 2:e01503.
52. Kaikkonen MU, *et al.* (2013) Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell* 51(3):310-325.
53. Guenther MG, *et al.* (2005) Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* 102(24):8603-8608.

54. Cho YW, *et al.* (2007) PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* 282(28):20395-20406.
55. Dou Y, *et al.* (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature structural & molecular biology* 13(8):713-719.
56. Ruthenburg AJ, *et al.* (2006) Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nature structural & molecular biology* 13(8):704-712.
57. Avdic V, *et al.* (2011) Structural and biochemical insights into MLL1 core complex assembly. *Structure* 19(1):101-108.
58. South PF, Fingerman IM, Mersman DP, Du HN, & Briggs SD (2010) A conserved interaction between the SDI domain of Bre2 and the Dpy-30 domain of Sdc1 is required for histone methylation and gene expression. *The Journal of biological chemistry* 285(1):595-607.
59. Buecker C & Wysocka J (2012) Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet* 28(6):276-284.
60. Wang P, *et al.* (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Molecular and cellular biology* 29(22):6074-6085.
61. Ang YS, *et al.* (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* 145(2):183-197.
62. Austenaa L, *et al.* (2012) The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. *Immunity* 36(4):572-585.
63. Guo C, *et al.* (2012) Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proceedings of the National Academy of Sciences of the United States of America* 109(43):17603-17608.
64. Cierpicki T, *et al.* (2010) Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. *Nature structural & molecular biology* 17(1):62-68.
65. Risner LE, *et al.* (2013) Functional specificity of CpG DNA-binding CXXC domains in mixed lineage leukemia. *J Biol Chem* 288(41):29901-29910.
66. Erfurth FE, *et al.* (2008) MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. *Proceedings of the National Academy of Sciences of the United States of America* 105(21):7517-7522.
67. Ooi SK, *et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448(7154):714-717.
68. Zhang Y, *et al.* (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic acids research* 38(13):4246-4253.
69. Wang J, Muntean AG, Wu L, & Hess JL (2012) A subset of mixed lineage leukemia proteins has plant homeodomain (PHD)-mediated E3 ligase activity. *J Biol Chem* 287(52):43410-43416.

70. Wang J, Muntean AG, & Hess JL (2012) ECSASB2 mediates MLL degradation during hematopoietic differentiation. *Blood* 119(5):1151-1161.
71. Liu H, Cheng EH, & Hsieh JJ (2007) Bimodal degradation of MLL by SCFSkp2 and APCCdc20 assures cell cycle execution: a critical regulatory circuit lost in leukemogenic MLL fusions. *Genes Dev* 21(19):2385-2398.
72. Ding L, *et al.* (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455(7216):1069-1075.
73. Wood LD, *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853):1108-1113.
74. Gui Y, *et al.* (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature genetics* 43(9):875-878.
75. Cancer Genome Atlas Research N, *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature* 497(7447):67-73.
76. Jones WD, *et al.* (2012) De novo mutations in MLL cause Wiedemann-Steiner syndrome. *Am J Hum Genet* 91(2):358-364.
77. Ansari KI, Kasiri S, & Mandal SS (2013) Histone methylase MLL1 has critical roles in tumor growth and angiogenesis and its knockdown suppresses tumor growth in vivo. *Oncogene* 32(28):3359-3370.
78. Glaser S, *et al.* (2006) Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* 133(8):1423-1432.
79. Lubitz S, Glaser S, Schaft J, Stewart AF, & Anastassiadis K (2007) Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Mol Biol Cell* 18(6):2356-2366.
80. Sierra J, Yoshida T, Joazeiro CA, & Jones KA (2006) The APC tumor suppressor counteracts beta-catenin activation and H3K4 methylation at Wnt target genes. *Genes Dev* 20(5):586-600.
81. Chen C, *et al.* (2014) MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer cell* 25(5):652-665.
82. Lee J, *et al.* (2009) A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. *Proceedings of the National Academy of Sciences of the United States of America* 106(21):8513-8518.
83. Lee S, Lee J, Lee SK, & Lee JW (2008) Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. *Mol Endocrinol* 22(6):1312-1319.
84. Ansari KI, Hussain I, Kasiri S, & Mandal SS (2012) HOXC10 is overexpressed in breast cancer and transcriptionally regulated by estrogen via involvement of histone methylases MLL3 and MLL4. *Journal of molecular endocrinology* 48(1):61-75.
85. Guo C, *et al.* (2013) KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. *Oncotarget* 4(11):2144-2153.
86. Kim JH, *et al.* (2014) UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. *Cancer Res* 74(6):1705-1717.

87. Peifer M, *et al.* (2012) Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature genetics* 44(10):1104-1110.
88. Sjoblom T, *et al.* (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314(5797):268-274.
89. Lawrence MS, *et al.* (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495-501.
90. Tan YC & Chow VT (2001) Novel human HALR (MLL3) gene encodes a protein homologous to ALR and to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated with leukemia and developmental defects. *Cancer detection and prevention* 25(5):454-469.
91. Goo YH, *et al.* (2003) Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. *Molecular and cellular biology* 23(1):140-149.
92. Caira F, Antonson P, Pelto-Huikko M, Treuter E, & Gustafsson JA (2000) Cloning and characterization of RAP250, a novel nuclear receptor coactivator. *Journal of Biological Chemistry* 275(8):5308-5317.
93. Patel SR, Kim D, Levitan I, & Dressler GR (2007) The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Developmental cell* 13(4):580-592.
94. Agger K, *et al.* (2007) UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449(7163):731-734.
95. Hu D, *et al.* (2013) The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology* 33(23):4745-4754.
96. Valekunja UK, *et al.* (2013) Histone methyltransferase MLL3 contributes to genome-scale circadian transcription. *Proceedings of the National Academy of Sciences of the United States of America* 110(4):1554-1559.
97. Ruault M, Brun ME, Ventura M, Roizes G, & De Sario A (2002) MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. *Gene* 284(1-2):73-81.
98. Sedkov Y, *et al.* (2003) Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. *Nature* 426(6962):78-83.
99. Kim DH, Lee J, Lee B, & Lee JW (2009) ASCOM controls farnesoid X receptor transactivation through its associated histone H3 lysine 4 methyltransferase activity. *Mol Endocrinol* 23(10):1556-1562.
100. Ananthanarayanan M, *et al.* (2011) Histone H3K4 trimethylation by MLL3 as part of ASCOM complex is critical for NR activation of bile acid transporter genes and is downregulated in cholestasis. *American journal of physiology. Gastrointestinal and liver physiology* 300(5):G771-781.
101. Lappin TR, Grier DG, Thompson A, & Halliday HL (2006) HOX genes: seductive science, mysterious mechanisms. *The Ulster medical journal* 75(1):23-31.
102. Ansari KI, Hussain I, Shrestha B, Kasiri S, & Mandal SS (2011) HOXC6 Is transcriptionally regulated via coordination of MLL histone methylase and

estrogen receptor in an estrogen environment. *Journal of molecular biology* 411(2):334-349.
103. Ansari KI, Shrestha B, Hussain I, Kasiri S, & Mandal SS (2011) Histone methylases MLL1 and MLL3 coordinate with estrogen receptors in estrogen-mediated HOXB9 expression. *Biochemistry* 50(17):3517-3527.
104. Ansari KI, Kasiri S, Hussain I, & Mandal SS (2009) Mixed lineage leukemia histone methylases play critical roles in estrogen-mediated regulation of HOXC13. *The FEBS journal* 276(24):7400-7411.
105. Bhan A*, et al.* (2013) Antisense transcript long noncoding RNA (lncRNA) HOTAIR is transcriptionally induced by estradiol. *Journal of molecular biology* 425(19):3707-3722.
106. Lewis MT (2000) Homeobox genes in mammary gland development and neoplasia. *Breast Cancer Res* 2(3):158-169.
107. Jozwik KM, Chernukhin I, Serandour AA, Nagarajan S, & Carroll JS (2016) FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell reports* 17(10):2715-2723.
108. Mohammed H*, et al.* (2013) Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell reports* 3(2):342-349.
109. Mohammed H*, et al.* (2015) Progesterone receptor modulates ERalpha action in breast cancer. *Nature* 523(7560):313-317.
110. Zhang Z*, et al.* (2016) Mammary-Stem-Cell-Based Somatic Mouse Models Reveal Breast Cancer Drivers Causing Cell Fate Dysregulation. *Cell reports* 16(12):3146-3156.
111. Society AC (2013) Breast Cancer Facts and Figures 2013-2014. *Atlanta: American Cancer Society, Inc.*
112. Lumachi F, Brunello A, Maruzzo M, Basso U, & Basso SM (2013) Treatment of estrogen receptor-positive breast cancer. *Curr Med Chem* 20(5):596-604.
113. Miller TW, Balko JM, & Arteaga CL (2011) Phosphatidylinositol 3-kinase and antiestrogen resistance in breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29(33):4452-4461.
114. Bosch A*, et al.* (2015) PI3K inhibition results in enhanced estrogen receptor function and dependence in hormone receptor-positive breast cancer. *Sci Transl Med* 7(283):283ra251.
115. Hoeflich KP*, et al.* (2016) The PI3K inhibitor taselisib overcomes letrozole resistance in a breast cancer model expressing aromatase. *Genes Cancer* 7(3-4):73-85.
116. Toska E*, et al.* (2017) PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. *Science* 355(6331):1324-1330.
117. Baselga J*, et al.* (2012) Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med* 366(6):520-529.
118. Osborne CK & Schiff R (2011) Mechanisms of endocrine resistance in breast cancer. *Annual review of medicine* 62:233-247.
119. Toy W*, et al.* (2013) ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nature genetics* 45(12):1439-1445.

120. Li S, *et al.* (2013) Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell reports* 4(6):1116-1130.
121. Jeselsohn R, *et al.* (2014) Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* 20(7):1757-1767.
122. Robinson DR, *et al.* (2013) Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nature genetics* 45(12):1446-1451.
123. Toy W, *et al.* (2017) Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists. *Cancer Discov* 7(3):277-287.
124. Chandarlapaty S, *et al.* (2016) Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. *JAMA Oncol* 2(10):1310-1315.
125. Fanning SW, *et al.* (2016) Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *Elife* 5.
126. Weir HM, *et al.* (2016) AZD9496: An Oral Estrogen Receptor Inhibitor That Blocks the Growth of ER-Positive and ESR1-Mutant Breast Tumors in Preclinical Models. *Cancer Res* 76(11):3307-3318.
127. Tzukerman MT, *et al.* (1994) Human estrogen receptor transactivational capacity is determined by both cellular and promoter context and mediated by two functionally distinct intramolecular regions. *Mol Endocrinol* 8(1):21-30.
128. Zilliacus J, Wright AP, Carlstedt-Duke J, Nilsson L, & Gustafsson JA (1995) Modulation of DNA-binding specificity within the nuclear receptor family by substitutions at a single amino acid position. *Proteins* 21(1):57-67.
129. Zilliacus J, Wright AP, Carlstedt-Duke J, & Gustafsson JA (1995) Structural determinants of DNA-binding specificity by steroid receptors. *Mol Endocrinol* 9(4):389-400.
130. Gronemeyer H (1991) Transcription activation by estrogen and progesterone receptors. *Annu Rev Genet* 25:89-123.
131. Le Romancer M, *et al.* (2011) Cracking the estrogen receptor's posttranslational code in breast tumors. *Endocr Rev* 32(5):597-622.
132. Thomas RS, Sarwar N, Phoenix F, Coombes RC, & Ali S (2008) Phosphorylation at serines 104 and 106 by Erk1/2 MAPK is important for estrogen receptor-alpha activity. *Journal of molecular endocrinology* 40(4):173-184.
133. Medunjanin S, *et al.* (2005) Glycogen synthase kinase-3 interacts with and phosphorylates estrogen receptor alpha and is involved in the regulation of receptor activity. *J Biol Chem* 280(38):33006-33014.
134. Rogatsky I, Trowbridge JM, & Garabedian MJ (1999) Potentiation of human estrogen receptor alpha transcriptional activation through phosphorylation of serines 104 and 106 by the cyclin A-CDK2 complex. *J Biol Chem* 274(32):22296-22302.

135. Chen D, *et al.* (2002) Phosphorylation of human estrogen receptor alpha at serine 118 by two distinct signal transduction pathways revealed by phosphorylation-specific antisera. *Oncogene* 21(32):4921-4931.
136. Kato S, *et al.* (1995) Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* 270(5241):1491-1494.
137. Masuhiro Y, *et al.* (2005) Splicing potentiation by growth factor signals via estrogen receptor phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 102(23):8126-8131.
138. Chambliss KL, Yuhanna IS, Anderson RG, Mendelsohn ME, & Shaul PW (2002) ERbeta has nongenomic action in caveolae. *Mol Endocrinol* 16(5):938-946.
139. Dutertre M & Smith CL (2003) Ligand-independent interactions of p160/steroid receptor coactivators and CREB-binding protein (CBP) with estrogen receptor-alpha: regulation by phosphorylation sites in the A/B region depends on other receptor domains. *Mol Endocrinol* 17(7):1296-1314.
140. Shah YM & Rowan BG (2005) The Src kinase pathway promotes tamoxifen agonist action in Ishikawa endometrial cells through phosphorylation-dependent stabilization of estrogen receptor (alpha) promoter interaction and elevated steroid receptor coactivator 1 activity. *Mol Endocrinol* 19(3):732-748.
141. Joel PB, *et al.* (1998) pp90rsk1 regulates estrogen receptor-mediated transcription through phosphorylation of Ser-167. *Molecular and cellular biology* 18(4):1978-1984.
142. Campbell RA, *et al.* (2001) Phosphatidylinositol 3-kinase/AKT-mediated activation of estrogen receptor alpha: a new model for anti-estrogen resistance. *J Biol Chem* 276(13):9817-9824.
143. Inano K, Curtis SW, Korach KS, Omata S, & Horigome T (1994) Heat shock protein 90 strongly stimulates the binding of purified estrogen receptor to its responsive element. *J Biochem* 116(4):759-766.
144. Inano K, Ishida T, Omata S, & Horigome T (1992) In vitro formation of estrogen receptor-heat shock protein 90 complexes. *J Biochem* 112(4):535-540.
145. Inano K, *et al.* (1990) Reconstitution of the 9 S estrogen receptor with heat shock protein 90. *FEBS Lett* 267(1):157-159.
146. Echeverria PC & Picard D (2010) Molecular chaperones, essential partners of steroid hormone receptors for activity and mobility. *Biochim Biophys Acta* 1803(6):641-649.
147. Pratt WB & Toft DO (1997) Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr Rev* 18(3):306-360.
148. Pettersson K, Grandien K, Kuiper GG, & Gustafsson JA (1997) Mouse estrogen receptor beta forms estrogen response element-binding heterodimers with estrogen receptor alpha. *Mol Endocrinol* 11(10):1486-1496.
149. Maggi A (2011) Liganded and unliganded activation of estrogen receptor and hormone replacement therapies. *Biochim Biophys Acta* 1812(8):1054-1060.
150. Titolo D, Mayer CM, Dhillon SS, Cai F, & Belsham DD (2008) Estrogen facilitates both phosphatidylinositol 3-kinase/Akt and ERK1/2 mitogen-

activated protein kinase membrane signaling required for long-term neuropeptide Y transcriptional regulation in clonal, immortalized neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28(25):6473-6482.
151. Migliaccio A*, et al.* (1996) Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. *EMBO J* 15(6):1292-1300.
152. Lee HR, Kim TH, & Choi KC (2012) Functions and physiological roles of two types of estrogen receptors, ERalpha and ERbeta, identified by estrogen receptor knockout mouse. *Lab Anim Res* 28(2):71-76.
153. Tanos T, Rojo L, Echeverria P, & Brisken C (2012) ER and PR signaling nodes during mammary gland development. *Breast cancer research : BCR* 14(4):210.
154. Brisken C & O'Malley B (2010) Hormone action in the mammary gland. *Cold Spring Harbor perspectives in biology* 2(12):a003178.
155. Wagner KU*, et al.* (1997) Cre-mediated gene deletion in the mammary gland. *Nucleic acids research* 25(21):4323-4330.
156. Feng Y, Manka D, Wagner KU, & Khan SA (2007) Estrogen receptor-alpha expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America* 104(37):14718-14723.
157. Heery DM, Kalkhoven E, Hoare S, & Parker MG (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387(6634):733-736.
158. Klinge CM (2000) Estrogen receptor interaction with co-activators and co-repressors. *Steroids* 65(5):227-251.
159. Uht RM, Anderson CM, Webb P, & Kushner PJ (1997) Transcriptional activities of estrogen and glucocorticoid receptors are functionally integrated at the AP-1 response element. *Endocrinology* 138(7):2900-2908.
160. Webb P, Lopez GN, Uht RM, & Kushner PJ (1995) Tamoxifen activation of the estrogen receptor/AP-1 pathway: potential origin for the cell-specific estrogen-like effects of antiestrogens. *Mol Endocrinol* 9(4):443-456.
161. Paech K*, et al.* (1997) Differential ligand activation of estrogen receptors ERalpha and ERbeta at AP1 sites. *Science* 277(5331):1508-1510.
162. Porter W, Saville B, Hoivik D, & Safe S (1997) Functional synergy between the transcription factor Sp1 and the estrogen receptor. *Mol Endocrinol* 11(11):1569-1580.
163. Porter W, Wang F, Wang W, Duan R, & Safe S (1996) Role of estrogen receptor/Sp1 complexes in estrogen-induced heat shock protein 27 gene expression. *Mol Endocrinol* 10(11):1371-1378.
164. Dahlman-Wright K*, et al.* (2012) Interplay between AP-1 and estrogen receptor alpha in regulating gene expression and proliferation networks in breast cancer cells. *Carcinogenesis* 33(9):1684-1691.
165. Jepsen K & Rosenfeld MG (2002) Biological roles and mechanistic actions of co-repressor complexes. *J Cell Sci* 115(Pt 4):689-698.

166. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, & Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics* 43(1):27-33.
167. Lupien M*, et al.* (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132(6):958-970.
168. Jeong KW*, et al.* (2011) Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. *Nature structural & molecular biology* 18(12):1358-1365.
169. Jeong KW, Andreu-Vieyra C, You JS, Jones PA, & Stallcup MR (2013) Establishment of active chromatin structure at enhancer elements by mixed-lineage leukemia 1 to initiate estrogen-dependent gene expression. *Nucleic acids research*.
170. Dreijerink KM*, et al.* (2006) Menin links estrogen receptor activation to histone H3K4 trimethylation. *Cancer research* 66(9):4929-4935.
171. Won Jeong K, Chodankar R, Purcell DJ, Bittencourt D, & Stallcup MR (2012) Gene-specific patterns of coregulator requirements by estrogen receptor-alpha in breast cancer cells. *Mol Endocrinol* 26(6):955-966.
172. Dillon SC, Zhang X, Trievel RC, & Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology* 6(8):227.
173. Bernstein BE*, et al.* (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America* 99(13):8695-8700.
174. Krogan NJ*, et al.* (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* 277(13):10753-10755.
175. Ng HH, Robert F, Young RA, & Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* 11(3):709-719.
176. Hughes CM*, et al.* (2004) Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Molecular cell* 13(4):587-597.
177. Lee SK*, et al.* (2001) Two distinct nuclear receptor-interaction domains and CREB-binding protein-dependent transactivation function of activating signal cointegrator-2. *Mol Endocrinol* 15(2):241-254.
178. Lee J*, et al.* (2008) Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 105(49):19229-19234.
179. Lee S*, et al.* (2009) Crucial roles for interactions between MLL3/4 and INI1 in nuclear receptor transactivation. *Mol Endocrinol* 23(5):610-619.
180. Ciriello G*, et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 163(2):506-519.
181. Pedram Razavi MTC, Sumit Middha, Dara S. Ross, Ahmet Zehir, Tracy A. Proverbs-Singh, Cyriac Kandoth, Sarat Chandarlapaty, Maura N. Dickler, Jorge S. Reis-Filho, Sujata Patil, Venkatraman Seshan, Lillian Smyth, Neil M. Iyengar, Komal Jhaveri, Shanu Modi, Chau T. Dang, Mark E. Robson, Larry

Norton, Clifford A. Hudis, Marc Ladanyi, Maurizio Scaltriti, Nikolaus Schultz, David Hyman, Michael F. Berger, Barry S. Taylor, David B. Solit and José Baselga (2016) Clinical genomic profiling of 1000 metastatic breast cancer patients: actionable targets, novel alterations, and clinical correlations. *AACR*, (Cancer Research).

182. Fellmann C*, et al.* (2013) An optimized microRNA backbone for effective single-copy RNAi. *Cell reports* 5(6):1704-1713.
183. Shen B*, et al.* (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature methods* 11(4):399-402.
184. Robinson JL & Carroll JS (2012) FoxA1 is a key mediator of hormonal response in breast and prostate cancer. *Front Endocrinol (Lausanne)* 3:68.
185. Creyghton MP*, et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* 107(50):21931-21936.
186. Joseph R*, et al.* (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Molecular systems biology* 6:456.
187. Carroll JS*, et al.* (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* 38(11):1289-1297.
188. Mohammed H*, et al.* (2016) Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc* 11(2):316-326.
189. Kim DH*, et al.* (2015) Crucial roles of mixed-lineage leukemia 3 and 4 as epigenetic switches of the hepatic circadian clock controlling bile acid homeostasis in mice. *Hepatology* 61(3):1012-1023.
190. Giunciuglio D*, et al.* (1995) Invasive phenotype of MCF10A cells overexpressing c-Ha-ras and c-erbB-2 oncogenes. *Int J Cancer* 63(6):815-822.
191. de Caestecker MP, Piek E, & Roberts AB (2000) Role of transforming growth factor-beta signaling in cancer. *J Natl Cancer Inst* 92(17):1388-1402.
192. Masamura S, Santner SJ, Heitjan DF, & Santen RJ (1995) Estrogen deprivation causes estradiol hypersensitivity in human breast cancer cells. *The Journal of clinical endocrinology and metabolism* 80(10):2918-2925.
193. Jeng MH*, et al.* (1998) Estrogen receptor expression and function in long-term estrogen-deprived human breast cancer cells. *Endocrinology* 139(10):4164-4174.
194. Santen RJ*, et al.* (2005) Long-term estradiol deprivation in breast cancer cells up-regulates growth factor signaling and enhances estrogen sensitivity. *Endocr Relat Cancer* 12 Suppl 1:S61-73.
195. Dumont JA*, et al.* (1996) Progression of MCF-7 breast cancer cells to antiestrogen-resistant phenotype is accompanied by elevated levels of AP-1 DNA-binding activity. *Cell Growth Differ* 7(3):351-359.
196. Johnston SR*, et al.* (1999) Increased activator protein-1 DNA binding and c-Jun NH2-terminal kinase activity in human breast tumors with acquired tamoxifen resistance. *Clin Cancer Res* 5(2):251-256.
197. Malorni L*, et al.* (2016) Blockade of AP-1 Potentiates Endocrine Therapy and Overcomes Resistance. *Mol Cancer Res* 14(5):470-481.

198. Schiff R, *et al.* (2000) Oxidative stress and AP-1 activity in tamoxifen-resistant breast tumors in vivo. *J Natl Cancer Inst* 92(23):1926-1934.
199. Zhou Y, *et al.* (2007) Enhanced NF kappa B and AP-1 transcriptional activity associated with antiestrogen resistant breast cancer. *BMC Cancer* 7:59.
200. Gutierrez MC, *et al.* (2005) Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase. *J Clin Oncol* 23(11):2469-2476.
201. Wang C, *et al.* (2011) Estrogen induces c-myc gene expression via an upstream enhancer activated by the estrogen receptor and the AP-1 transcription factor. *Mol Endocrinol* 25(9):1527-1538.
202. Sato K & Akimoto K (2016) Expression Levels of KMT2C and SLC20A1 Identified by Information-theoretical Analysis Are Powerful Prognostic Biomarkers in Estrogen Receptor-positive Breast Cancer. *Clin Breast Cancer*.
203. Dawkins JB, *et al.* (2016) Reduced Expression of Histone Methyltransferases KMT2C and KMT2D Correlates with Improved Outcome in Pancreatic Ductal Adenocarcinoma. *Cancer Res* 76(16):4861-4871.
204. van Berkum NL, *et al.* (2010) Hi-C: a method to study the three-dimensional architecture of genomes. *J Vis Exp* (39).
205. Hutchinson JN & Muller WJ (2000) Transgenic mouse models of human breast cancer. *Oncogene* 19(53):6130-6137.
206. Hsu PD, *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9):827-832.
207. Cong L, *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819-823.
208. Bocchinfuso WP, *et al.* (2000) Induction of mammary gland development in estrogen receptor-alpha knockout mice. *Endocrinology* 141(8):2982-2994.
209. Dobin A, *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15-21.
210. Liao Y, Smyth GK, & Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923-930.
211. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12):550.
212. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* 9(4):357-359.
213. Zhang Y, *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9(9):R137.
214. Heinz S, *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* 38(4):576-589.
215. Kannengiesser C, *et al.* (2008) Gene expression signature associated with BRAF mutations in human primary cutaneous melanomas. *Molecular oncology* 1(4):425-430.
216. Flockhart RJ, *et al.* (2012) BRAFV600E remodels the melanocyte transcriptome and induces BANCR to regulate melanoma cell migration. *Genome research* 22(6):1006-1014.

217. Richart L, Real FX, & Sanchez-Arevalo Lobo VJ (2016) c-MYC partners with BPTF in human cancer. *Molecular & cellular oncology* 3(3):e1152346.
218. Richart L, *et al.* (2016) BPTF is required for c-MYC transcriptional activity and in vivo tumorigenesis. *Nature communications* 7:10153.
219. Curtis C, *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403):346-352.