Targeting Histone Chaperones Retinoblastoma Binding Proteins 4 and 7 for the Treatment of Triple Negative Breast Cancer (TNBC)

By

Rebecca Ann Reed

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2016

Doctoral Committee:

Professor Duxin Sun, Chair
Associate Professor Yali Dou
Professor Anna Mapp
Assistant Professor Daniel Southworth
Professor Shaomeng Wang

To my lovely parents, Pat and Mark Reed, whose love and support shaped who I am today. Further to my brother, Matt Reed, who is always there to cheer me up and encourage me to pursue my dreams. And lastly to my fiancé, Michael Moody, who is constantly able to make me laugh and smile.

"We have to continually be jumping off cliffs and developing our wings on the way down." –Kurt Vonnegut

Acknowledgements

First, I would like to thank my advisor, Dr. Duxin Sun, for his insight and guidance throughout my graduate career. His continued optimism and passion for research helped me grow as a scientist and strive for the best. I can't recall a single time I didn't see Dr. Sun with a smile and his positive attitude was contagious. I will be forever grateful for his fantastic mentorship.

I would also like to thank my committee members: Dr. Shaomeng Wang, Dr. Yali Dou, Dr. Dan Southworth and Dr. Anna Mapp. Their advice throughout my dissertation has been immensely beneficial. Their diverse backgrounds also helped shaped my career as a chemical biologist as I was required to examine my project from several angles. Thank you for the constant support and insightful discussions.

I would like to acknowledge fellow graduate students and coworkers who have made my experience in graduate school not only positive, but also enjoyable. To Dr. Hacer Karatas and Dr. Xu Ran, thank you for your help with peptide synthesis and to Dr. Liu Liu, I appreciate your support in assay development. To all of the members of the Sun lab, past and present, you made the lab a great place to work. I have to specially thank Dr. Miao-Chia, Albert Lin, Nicholas Stevers and Samantha Tinsley for their help in my research efforts and suggestions to improve my work. And to Kate Sansanaphongpricha and Nate Truchan, thanks for sharing an office with me and laughing at my awful jokes.

Finally, I would like to thank my friends and family for their love and support throughout my dissertation. When times were tough, I looked to you for help and I couldn't have made it this far without you all.

Coming to Michigan to pursue my doctorate may have been the best decision I have ever made; I not only made great friends and met my fiancé, but I also was immersed in a wonderfully collaborative environment that cultivated my interest in cancer research. Again, I can't thank you all enough.

Table of Contents

List of Figures

List of Tables

Abstract

Retinoblastoma-binding proteins 4 and 7 (RBBP4/7) are ubiquitously expressed histone chaperones intimately tied to epigenetic regulation. They are subunits in several complexes that are involved in transcriptional repression including the Polycomb Repressive Complex 2 (PRC2) and the Nucleosome Remodeling and Deacetylase Complex (NuRD). Both of these complexes have been implicated in the progression of triple negative breast cancer (TNBC), the only subtype for which there is no targeted therapy. Efforts to target these complexes have demonstrated little success in TNBC. Novel strategies to inhibit these complexes are therefore required.

Targeting protein-protein interactions is one of the potential routes for inhibition of epigenetic complexes. RBBP4/7 are thought to be important in recognizing histone H3 and coregulatory factors for optimal complex activity. However, the functional relevance of RBBP4/7 in TNBC is not well understood. In this report, I demonstrate that together RBBP4/7 are critical to proliferation of TNBC and maintenance of the cancer stem cell population. I further show that depletion of RBBP4 and RBBP7 is associated with a switch of gene expression from basal to luminal. The aforementioned changes were due in part to dysregulation of

PRC2 and the decrease in its' associated H3K27me3 mark. These results reveal RBBP4/7 to be a potential therapeutic target for the treatment of TNBC.

As RBBP4/7 promote nucleosome association of their respective epigenetic complexes, we investigated whether interfering with the interaction of RBBP4 and histone H3 would represent a novel therapeutic approach. Examination of this binding interface on RBBP4 led to the discovery that this site is also occupied by the oncogenic transcription factor BCL11A. Treating a TNBC cell line with a BCL11A-derived peptide resulted in a 50% decrease in the activity of aldehyde dehydrogenase (ALDH), an enzyme linked to aggressiveness and poor clinical outcome in breast cancer. Together these results indicate that a decoy peptide, or a small molecule inhibitor, against RBBP4 could be an innovative approach in the treatment of TNBC. An optimized fluorescence polarization (FP) assay was developed to help discover inhibitors of this interaction and better understand its biological consequences and implications in TNBC.

Chapter 1

Breast Cancer, Breast Cancer Stem Cells and the Need for Novel Targeted Therapies

1.1 Introduction.

The history of chemotherapy dates back to World War I with the induction of chemical warfare by Germany[1,2]. Dr. Stewart Francis Alexander made observations that agents such as mustard gas triggered significant lymphoid and myeloid suppression. These observations led Dr. Alexander to the conclusion that these chemical warfare agents could be repurposed to halt the growth of cancerous cells. Supporting studies by Yale pharmacologists, Louis Goodman and Alfred Gilman, resulted in the development of the first ever prototype chemotherapy drug termed a "synthetic lymphocidal chemical", which was nitrogen mustard, the compound used to make mustard gas[1-3]. With these remarkable advancements, nitrogen mustards became the precursor to the chemical war on cancer officially transitioning from "battleground-to-bedside."

Cancer is a disease united by the shared traits of uncontrollable cell growth and ability to invade surrounding tissues. Nitrogen mustards attacked this unifying trait by binding to the cell's DNA, preventing proper DNA replication, and thus halting the cell cycle (Figure 1.1)[3]. However, it is becoming more apparent that

cancer is not a single disease, but rather a group of diseases. Uniform treatments such as chemotherapy and radiation, which target rapidly dividing cells in the human body, cause rampant toxicity as normal healthy cells in the blood, skin, stomach and hair also fall victim to the treatment. These debilitating side effects beget a critical need for the discovery of new cancer therapeutics that can distinguish normal cells from their diseased counterpart, i.e. separate friend from foe.

A new wave of cancer research has focused on "targeted therapies" as a means by which to accomplish this goal[4,5]. With this approach, specific molecular targets of cancerous cells are pursued with small molecule inhibitors or antibodies in hopes of finding an Achilles' heel of the disease. While cancers do share commonalities, their response to treatment can differ extensively and therefore, research efforts have been divided to reflect the reality that cancer is not a single disease. In order to develop successful targeted therapies, it is necessary to understand the molecular underpinnings of each type of cancer, be it breast cancer, liver cancer, skin cancer, .et al. Herein this thesis we focus on breast cancer, as there are close to a quarter of a million new diagnoses each year in the United States alone. Furthermore, breast cancer is the second leading cause of cancer-related deaths amongst women.

1.2 Breast Cancer: Development, Subtypes and the Search for a Cure.

Breast cancer can originate in either the glands that produce milk (lobular) or the ducts that carry the milk (ductal). The majority of breast cancer diagnoses are of the ductal variety with neoplastic transformation from normal to invasive

ductal carcinoma (IDC) occurring through three precursory stages: flat epithelial atypia (FEA), atypical ductal hyperplasia (ADH), and ductal carcinoma in situ (DCIS)[6,7]. Of note, it is estimated that only 30-50% of the DCIS cases identified will develop into the potentially lethal IDC[6,8]. Biopsied samples are graded from 1 to 3 based on how similar the cells appear to normal tissue with lower grades indicating a more-differentiated, normal-like cancer that could be slow to progress and have a better prognosis, while higher grades suggest abnormal cells with a greater chance for invasion.

Breast cancer is further classified by receptor status, i.e. the expression of the estrogen receptor (ER+/-), the progesterone receptor (PR+/-), or the human epidermal growth factor receptor 2 (HER2+/-) (Figure 1.2)[9,10]. Positive receptor status renders the breast cancer susceptible to hormone therapy drugs or targeted treatments. Tamoxifen, a selective estrogen-receptor modulator (SERM), is currently used for treatment of ER+ breast cancer and prevention of breast cancer in those of high-risk[11]. Aromatase inhibitors (AI) are also used for treatment of ER+ breast cancers in post-menopausal women where estrogen levels are dependent on the activity of aromatase enzymes rather than the ovaries[12]. When HER2 overexpression exists, trastuzumab, a monoclonal antibody, is prescribed to prevent ligand-independent HER2-HER3 dimerization and thereby, downstream signaling of the proto-oncogene, specifically the PI3K pathway and cell cycle progression[13-15]. Additionally, trastuzumab mediates antibody-dependent cellular cytotoxicity (ADCC) by serving as a honing device for the immune system[16]. While a significant number of patients benefit from these therapies, breast cancer that does not exhibit

expression of these receptors, also termed triple negative (TNBC), is restricted to radiotherapy and/or traditional cytotoxic chemotherapy.

Molecular analysis and significant efforts in basic research have found that breast cancer can be distinguished by characteristic gene expression patterns[17-19]. These signatures further divide BC into five different intrinsic subtypes: luminal A, luminal B, HER2-enriched, normal-like, and basal-like[10]. Expression profiling not only provides a prognostic value for clinical outcome, but also helps determine the best course for treatment and identify new therapeutic targets[5,20]. Insight into the molecular mechanisms and signaling pathways that fuel breast cancer progression has prompted development of several targeted therapies that are being investigated in clinical trials. For example, the epidermal growth factor receptor (EGFR), a receptor tyrosine kinase (RTK) is often overexpressed in TNBC and its presence is correlated with a poor prognosis. Therefore, research efforts have focused on uncovering EGFR inhibitors such as lapatinib and dacomitinib. Lapatinib is a dual-tyrosine kinase inhibitor of HER2 and EGFR, whereas dacomitinib is a pan-HER irreversible inhibitor[21,22]. While their success for the treatment of TNBC has been limited due to acquired resistance, dacomitinib effectively decreases proliferation of HER-2 enriched breast cancers that are resistant to trastuzumab[22]. Meanwhile, downstream effector pathways of EGFR, including the AKT and MAPK signaling cascades, have lead to the development of mTOR and PI3K inhibitors to supersede acquired resistance. In fact, the PI3K-AKT-mTOR pathway regulates proliferation, apoptosis, and survival with overactivation witnessed in 60% of TNBC cases[5,23]. MLN0128 (against mTOR) and BYL719 (against PI3K) are in clinical trials to target

the hyperactivation of these pathways in cancers[24,25]. Nevertheless, as cancer cells are masters of survival, it is no surprise that inhibition of one pathway leads to the upregulation of different oncogenes and induction of different oncogenic cascades. Evidence of the significant crosstalk between the MAPK and PI3K pathways has proven it will be important to rationally choose combination therapies in order to prevent this phenomenon[26].

Furthermore, the heterogeneous nature of breast cancer, especially in the basal subtype, makes it difficult to effectively target all cells within the tumor. Drug-resistant cells could then repopulate the tumor with a more aggressive and invasive phenotype. A relatively old theory that distinct cell lineages could originate from a single precursor has been applied to cancer to help explain the heterogeneity of the disease. Designated as cancer stem cells (CSCs), these cells are more resistant to chemotherapy and have the ability to give rise to a diverse and plastic population that make up the bulk of the tumor[27-29]. Recognition of CSCs offers a new avenue for cancer treatment that could aid other therapies in preventing recurrence and providing a cure.

1.3 The Existence and Implications of Breast Cancer Stem Cells (BCSCs).

Despite significant advancements in treatment options for breast cancer, the fact remains that over 90% of cancer fatalities are due to metastatic spreading of the disease. Chemoresistance, and thereby tumor relapse and metastasis, remains a critical issue in the war on cancer. Paramount to the development of therapy resistance is the existence of cancer stem cells (CSC). CSCs reside within the tumor

and contribute to its heterogeneous nature; they are characterized by their ability to (1) self-renew and (2) differentiate into cells that form the bulk of the tumor[28].

First discovered in acute myeloid leukemia, the relevance of CSCs was identified in breast cancer in 2003 where it was demonstrated that a biomarker sorted cell population, CD44+/CD24-, had a higher rate of tumor initiation than bulk tumor cells[27,29]. It was since shown in both breast cancer cell lines and patient-derived samples that CSCs survived traditional chemotherapies and exhibited radioresistance[30-32]. Naturally, the presence of cancer stem cells has broader implications for the treatment of cancer. As CSCs drive tumor growth, targeted therapies against these cells will be vital to prevent not only recurrence, but also metastasis[32].

Aside from semantic issues, the theory of CSC, which asserts that a subpopulation of cells has higher tumor-initiating capacity, has been widely accepted given recent research[33]. There are three primary pathways that are known to perpetuate the CSC phenotype and tumorigenicity: WNT, Notch and Hedgehog (Hh) and they are discussed below (Figure 1.3)[33-35].

WNT signaling is intimately tied to development as it controls stem cell pluripotency and cell fate decisions. In relation to breast development, WNT proteins were found to maintain mammary stem cells (MaSC) in cell culture as well as facilitate their ability to reconstitute a functional mammary gland in transplantation assays[36]. Overexpression of WNT signaling proteins has been observed in several breast cancers and breast cancer cell lines[37]. In particular, WNT members, like β-catenin and the WNT receptor frizzled 7 (FZD7), are enriched in

basal-like breast cancers, which demonstrate a less-differentiated phenotype[37,38]. Furthermore, hyperactivation of this pathway is associated with worse overall survival and seems to be an important step in metastatic colonization[39,40].

The signaling cascade is initiated by binding of WNT ligands to the FZD receptor, which is often coupled to low-density lipoprotein receptor-related (LRP) proteins. Upon binding, a signal is then relayed to the multisubunit destruction complex releasing β-catenin and allowing for accumulation in the cytoplasm. Translocation of β-catenin into the nucleus results in its the displacement of HDAC co-repressors from TCF/LEF transcription factors. β-catenin serves as a co-activator alongside TCF/LEF and can recruit epigenetic proteins, such as histone acetyltransferases CREB-binding protein (CBP) and p300, to positively regulate transcription of WNT-targeted genes. Cyclin D1 and c-Myc are two important WNT regulated genes that have been shown to promote cancer cell growth and stemness[41]. Blockade of WNT signaling may provide a new therapeutic avenue for improving patient survival and impeding metastasis[39,42].

Similar to WNT signaling, the Hedgehog (Hh) pathway is critical for embryonic development as well as breast tissue homeostasis. Improper activation of Hh signaling is a common observation in many cancers. In fact, the Hh receptor, smoothened homologue (SMO), and the transcription factor GLI1 are frequently upregulated in TNBC[43-45]. High levels of the secreted ligand, sonic hedgehog (Shh), as assessed by immunohistochemical and RNAseq studies, have also shown to be a predictor of worse overall survival in breast cancer[46]. Targets of Hh include anti-apoptotic protein Bcl-2, oncoprotein c-Myc and the transcription factor SNAIL,

which is important in promoting the epithelial-to-mesenchymal transition (EMT)[47]. EMT is believed to promote invasiveness and is often associated with CSCs[48]. Further studies have demonstrated that Hh signaling is correlated with the CD44+/CD24- BCSC population and that it encourages not only chemoresistance, but also osteolytic metastasis[43]. Non-canonical signaling in the Hh pathway involve the KRAS-MAPK and PI3K-AKT-mTOR kinase cascades to maintain stem-like properties in breast cancer cells and stimulate proliferation[44].

Lastly, another developmental pathway that goes awry in tumorigenesis and supports CSCs is the Notch pathway. Evolutionarily conserved, the Notch pathway is pivotal in maintaining stem-cell renewal via inhibition of differentiation[49]. In breast cancer, the Notch pathway regulates cell metabolism and promotes expression of aldehyde dehydrogenase (ALDH), an enzyme whose activity serves as a CSC marker[50-52]. A Notch reporter system developed by D'Angelo et al. confirmed that increased Notch activity in breast cancer cells resulted in higher tumor initiation capacity in mouse xenografts[53]. Importantly, cells negative for Notch expression failed to generate tumors[53]. Constitutive Notch activation results in increased Nf-kb signaling and the expression of pro-survival, anti-apoptotic proteins[51]. In addition, pro-inflammatory cytokines, such as IL-6 and IL-8, produced from Nf-kb signaling, promote stemness of CSC through activation of the JAK-STAT pathway and subsequent transcription of known stem cell factors: SOX2, OCT4, NOTCH, and WNT[54-57].

It is becoming more evident that these pathways do not work in isolation, but rather together in an interlocked web, with constant signaling crosstalk to maintain

CSC self-renewal and growth[35]. This supports the proposal for combined therapies to prevent chemoresistance by blocking compensatory pathways utilized by cancer cells and CSCs. While there are several inhibitors for WNT, Hh, and Notch, a relatively new approach to eradicate CSCs is to target the mediators of these signaling pathways, which are the epigenetic machineries, recruited via respective CSC transcription factors[58-60]. Targeting epigenetic enzymes and cofactors may have a far greater impact as they elicit control over the transcription of several genes and therefore, many pathways could be affected.

1.4 Epigenetics and Cancer Stem Cells

Epigenetics is defined as the molecular mechanisms by which cells induce heritable change without altering the underlying genetic information, or DNA. Through DNA methylation, histone modifications, chromatin remodeling and regulation of small non-coding RNA, the all-encompassing epigenetic state, or epigenome, of a cell defines a complicated language that translates a specific phenotype from the universal genotype (Figure 1.4). That is to say, epigenetics explains the phenomena where cells with the same underlying DNA sequence can differentiate into several different cell types.

Dysregulation in the epigenome and epigenetic signaling is becoming more apparent in tumorigenesis; and epigenetics is now recognized for its' central role in promoting oncogenic transformation as well as acquisition of cancer hallmarks[61]. In particular, within the last 15 years, the cancer epigenome has garnered a lot of interest for not only the marked alterations in epigenetic regulators, but also for its link to the dysregulation of self-renewal and differentiation in cancer stem cells[61-63].

In fact, a prominent theory is that epigenetic dysfunction primes a cell for carcinogenesis, providing these cells with the selective advantage for proliferation and evasion of apotosis, and also making them receptive to oncogenic drivers[64]. Many cancers also possess an embryonic stem cell signature supporting a possible stem cell origin in tumorigenesis[65].

The dynamic relationship between the genome and the epigenome in cancer promises alternative strategies for targeting cancer stem cell self-renewal and pluripotency. Whereas genetic mutations are permanent, epigenetic modifications maintain a degree of plasticity and are thus reversible. Therefore, epigenetic-targeted therapies may provide an opportunity to reprogram cancer cells or even sensitize the tumor to conventional chemotherapy[63]. In addition, as epigenetic regulators often influence the expression of multiple genes, therapeutic intervention could affect several pathways necessary for cancer initiation, maintenance, and metastasis. Here, we will discuss epigenetic-related therapeutic prospects for abolishing cancer stem cells in relation to the traditional breast cancer stem cell pathways reviewed earlier as well as managing other aspects of cancer stemness such as evading the immune system.

1.5 Targeting Epigenetic Mechanisms for Cancer Stem Cells Therapeutics

1.5.1 DNA Methyltransferases

DNA methylation was the first epigenetic modification to be characterized, and it is now widely appreciated for its role in enabling tumorigenesis[63,66,67]. As it was the first to be identified, it is the most comprehensively studied and it is understood to be relatively stable and associated with long-term gene repression.

DNA methyltransferases (DNMTs) catalyze the methylation of cytosines at the 5C position. DNMT1 maintains DNA methylation patterns during cell division, whereas DNMT3A and DNMT3B are *de novo* methyltransferases that establish new patters of DNA methylation and are important during development. Within cancer cells, there exists a shift in DNA methylation patterns with CpG islands prevailing in a hypermethylated state while there is a concomitant global loss of DNA methylation, or hypomethylation. In cancer, these focal gains in hypermethylation at CpG islands primarily target common tumor suppressor genes such as BRCA1, p16, and Nf-κB inhibitor PDLIM2[68-70]. In regards to CSCs, DNA methylation is also important for maintaining self-renewal by repressing genes critical for differentiation[71,72].

Decitabine (DAC) and azacitadine (AZA) are nucleoside-based DNA methyltransferase inhibitors that have gained momentum for the treatment of acute myeloid leukemia (AML)[73,74]. Their use for the treatment of solid tumors has been hindered by trials that use high doses, which exhibit extreme toxicity and confound the true potential of these inhibitors in solid malignancies. Recently, however, transient low doses of DAC and AZA have proven beneficial in reducing BCSC populations and sustaining long-term gene expression of previously hypermethylated genes[75].

In contrast, DNA hypomethylation is associated with promoting tumor growth via upregulation of oncogenes[67]. Therefore, future studies will need to examine the long-term side effects of treatment with DNMT inhibitors. Development of more specific inhibitors that target only certain promoters'

hypermethylation or particular DNMT isoforms may eliminate some of the off-target effects and toxicities associated with DAC and AZA. Nevertheless, DNMT research highlights the importance of epigenetic pathways in perpetuating genomic "driving" events and tumorigenicity.

1.5.2 Histone Writers, Readers and Erasers

Writer, readers and erasers are names given to epigenetic proteins responsible for placing, sensing or removing histone marks and establishing the "histone code." While histone methylation can either activate or repress gene transcription, histone acetylation is only known to enhance gene transcription. With the success of DNMTi, many efforts have been made to optimize other epigenetic inhibitors for the treatment of cancer, starting with histone deacetylase inhibitors (HDACi) and branching out with other writers, readers and erasers as targets.

Vorinostat, or suberoylanilide hydroxamic acid (SAHA), was the first in a class of histone deacetylase inhibitors approved by the FDA in 2006 for the treatment of cutaneous T cell lymphoma. In inflammatory breast cancer cell lines SUM149 and SUM190, treatment with SAHA results in the reduced expression of stem cell transcription factors Oct4, Nanog and Sox-2 and induces differentiation of BCSC[76,77]. Since the discovery of SAHA, there has been a dramatic increase in the number of HDAC inhibitors, many of which have shown clinical efficacy especially in combination with other drugs[73]. With HDACi treatment, breast cancer cells are not only resensitized to traditional therapies such as trastuzumab, but also show a decrease in metastatic potential and partial reversal of EMT[78].

As HDACs exist in multi-protein complexes, future studies could examine cofactor inhibition as a more specific alternative for targeting BCSC than current pan-deacetylase inhibitors. Bansal et al. recently took to this approach and developed an inhibitor against a PPI within the SIN3A HDAC complex[79,80]. A decoy peptide that mimicked the SIN3 interaction domain (SID) blocked recruitment of adaptor proteins to the paired α-helix (PAH2) domain of SIN3 resulting in the decreased expression of TNBC stem cell genes and those relating to the epithelial-to-mesenchymal transition (EMT)[79]. *In vivo*, this SID decoy led to a reduction in primary tumor growth and metastatic burden[79]. Pursuit of small molecule inhibitors of this interation led to the discovery of the avermectin macrocyclic lactone derivatives selamectin and ivermectin[81]. These avermectin compounds not only decreased the cancer stem cell phenotype in TNBC cell lines, but also sensitized breast cancer cells to other drugs[81]. Epigenetic therapies are therefore able to block CSC signaling, but also promote partial differentiation making once resistant cells susceptible to traditional treatments such as tamoxifen.

Nevertheless, the promiscuity of DNMTs and HDACs has raised concern over their use as a therapeutic target. Therefore, research efforts have shifted towards the more specific histone methyltransferases, demethylases and targeting interactions between oncoproteins and their epigenetic effectors. Two well-studied methyltransferases that are being exploited for CSC therapies include DOT1L and MLL. DOT1L, or disruptor of telomeric silencing 1-like protein, is the only known methyltransferase responsible for the active histone modifications H3K79 mono-, di- and trimethylation. As its name suggests, DOT1L is involved in telomeric

silencing, but also several other important cell processes including cell cycle progression, genomic stability, differentiation and pluripotency[82,83]. DOT1L was recently shown to play a major role in connecting the tumor microenvironment with the propagation of colon cancer stem cells[84]. Via Interleukin-22 mediated STAT3 activation, DOT1L expression is upregulated and H3K79 methylation is increased at the promoters of SOX2, NANOG, and OCT4[84]. Naturally, the increase in these stem cell transcription factors also increased the cancer stemness and tumorigenic potential. As STAT3 is an important factor in BCSC, a similar mechanism could be at play in breast cancer[55-57]. Inhibition of DOT1L with DZNep, EPZ-004777 or SYC-522 decreased mammospheres, EMT and migration of several breast cancer cell lines[83]. One of the therapeutic implications of these inhibitors is that one can target epigenetic enzymes to prevent immune-activated cancer stem cell expansion.

Mixed lineage leukemia (MLL) proteins have a conserved SET domain and catalyze the trimethylation of lysine 4 of histone H3, an activating mark. In total there are more than 60 fusion partners of MLL proteins that can contribute to leukemia by blocking differentiation and enhancing self-renewal capabilities. Importantly, our understanding of the hierarchy that exists within tumors is principally based on studies of leukemia[85,86]. Recently, MLL2 knockdown was shown to epigenetically regulate c-Myc levels in breast cancer and sensitize HER2+ breast cancers to Lapatinib[87]. Furthermore, MLL2 associates with Pygo2 to promote WNT signaling and BCSC expansion[88]. Efforts to inhibit MLL have included targeting the PPI between MLL and WDR5, a cofactor of MLL-protein complexes[89,90]. In

leukemic cell lines, this inhibitor induces differentiation. Whether these inhibitors will have any success in breast cancer is yet to be determined.

Lastly, an encouraging histone methyltransferase target, Enhancer of Zeste 2 (EZH2), resides as the enzymatic component of the Polycomb Repressive Complex 2 (PRC2). PRC2 is implicated in the maintenance of stem cells and has an important role in regulating gene expression during development and cell differentiation[91-96]. Consisting of four core subunits, EZH2, SUZ12, EED and RBBP4 (or RBBP7), PRC2 is responsible for the trimethylation of H3K27 and resultant gene repression via recruitment of other epigenetic complexes to alter chromatin structure. Several of these recruited machineries are also involved in CSC and tumor progression. For example, the H3K27me3 mark serves as a recruiter of the Polycomb Repressive Complex 1 (PRC1), which ubiquitinates H2K119 and further compacts chromatin[97]. One member of PRC1, BMI-1, is a well-known oncogene and implicated in the maintenance of stem cell self-renewal[98,99]. Specifically in breast cancer, hedgehog signaling increases BMI-1 expression and expands the CD44+/CD24- stem cell population[99]. PTC-209, a small-molecule inhibitor of BMI-1, was recently discovered and therapeutically validated in colon cancer stem cells where BMI-1 is also overexpressed[100]. The similarities between BMI-1 expression and its signaling pathways in breast and colon cancer stem cells warrant further investigation of PTC-209 in BCSC. Targeting PRC2 would be another option for inhibition of PRC1, or BMI-1, as it would block the first step in chromatin compaction.

In fact, EZH2, the enzymatic component of PRC2, is also overexpressed in many cancers and its overexpression promotes anchorage-independent growth and

cell invasiveness. Targets of PRC2 include repressors of the WNT pathway, DACT3 and DKK1, and DAB21P, which inhibits both Ras and Nf-kb (Figure 1.5A)[101-104]. EZH2-mediated repression of E-cadherin is also of importance in promoting the epithelial-mesenchymal transition and metastatic potential of breast cancers[105]. A PRC2 gene repression signature is correlated with poor clinical outcome in prostate, breast and small cell lung cancers[101,104]. Moreover, PRC2 target genes have been shown to be downregulated in TNBC and a recent study reported that an upstream regulator of PRC2 controls self-renewal of TNBC stem cells[106,107].

EZH2 also links PRC2 to the noncoding genome via interaction with long noncoding RNA (lncRNA) and regulation of microRNA (miRNA). HOTAIR, a well-studied oncogenic lncRNA, binds to EZH2 and directs PRC2 silencing to metastatic suppressor genes[108,109]. In particular, HOTAIR indirectly inhibits miR-7 by guiding PRC2 to the HOXD10 locus, which regulates miR-7[110]. MiR-7 prevents BCSC formation by targeting the STAT3 pathway and preventing transcription of c-Myc and TWIST[110]. Furthermore, loss of the miR-200 family is often associated with an increase in polycomb group proteins and a transition to a stem-like state[111,112]. Together, these studies suggest that PRC2 connects numerous epigenetic machineries to promote stemness in TNBC.

Several inhibitors of EZH2 have been reported: DZNep, GSK126 and EPZ-6438 (Figure 1.5B)[113-116]. The use of these inhibitors in lymphomas, which bear an EZH2 activating mutation, has shown clinical promise but studies in solid tumors have not been as successful. The failure of these inhibitors to recapitulate EZH2 knockdown studies in solid tumors indicates that in these cases, which often

overexpress EZH2, the tumor is dependent on catalytic and non-catalytic functions of PRC2 or EZH2. As mentioned previously, PRC2 has important functions in linking several epigenetic machineries to dictate control over chromatin structure. EZH2 catalytic inhibitors fail to inhibit the scaffolding role of PRC2. EZH2 itself may also have PRC2-independent roles that would not be affected by the small molecules currently available. A recent study in SWI/SNF-mutant cancers, which are dependent on PRC2 function, demonstrated that enzymatic inhibitors of EZH2 were not fully capable of suppressing EZH2 oncogenic activity[117]. The authors of the study suggest that only inhibitors that disrupt the integrity of the PRC2 complex will translate into successful therapies. In support of this, small molecule inhibitors against the EZH2-EED interaction such as the stabilized α-helix of an EZH2 peptide (SAH-EZH2) and Astemizole were shown to destabilize the PRC2 complex[118,119]. SAH-EZH2 and Astemizole inhibited the methyltransferase activity of PRC2, but more importantly blocked cancer cell proliferation in over-expressing EZH2 cells better than GSK126, an enzymatic inhibitor of EZH2[118,119].

Future studies aimed at the development of inhibitors targeting protein-protein interactions within PRC2 may provide better efficacy against breast cancer stem cells. Herein this thesis, we focus our efforts on the validation of RBBP4 and 7 (RBBP4/7), two PRC2 components important for nucleosome association, as novel targets for the treatment of TNBC and BCSC.

1.6 Targeting PRC2 components RBBP4/7 in TNBC

RBBP4/7 are two histone chaperones that are members of the WD40 repeat protein family and share the characteristic β-propeller. Similar to other WD40

proteins, RBBP4/7 have several interaction surfaces that make it an important member of many epigenetic complexes. In PRC2, for example, RBBP4/7 are required for nucleosome association and integrating the histone code. If the active H3K4me3 mark is present, RBBP4/7 are unable to bind the PRC2 substrate, histone H3, and thereby prevent efficient catalysis of H3K27 trimethylation[120]. Moreover, while RBBP4/7 are not essential for PRC2 function, they are required for optimal activity[121]. Since RBBP4/7 promote nucleosome association, we hypothesized that blocking the interaction between RBBP4/7 and histone H3 with a small molecule would inhibit PRC2 activity. These inhibitors would lead to destabilization of the complex and inhibit the oncogenic activity of PRC2 better than those solely targeting the EZH2 enzymatic domain. Small molecule inhibitors of PRC2 would be therapeutically beneficial in triple-negative breast cancer, where PRC2 targets are downregulated and EZH2 contributes to stemness as well as invasive characteristics of this disease[106]. However, the role of RBBP4/7 in tumorigenesis and PRC2 remains controversial. Therefore, we proposed two aims to study our hypothesis, with the first focused on the biology of RBBP4/7 and target validation and the second on assay development:

Aim 1: Validation of RBBP4/7 as a therapeutic target in TNBC.

After RNAi-mediated knockdown of RBBP4/7, we will demonstrate the effect of these two proteins on TNBC growth, stemness and tumorigenicity *in vitro* and *in vivo*. As RBBP4/7 are implicated in multiple epigenetic complexes (with histone methyltransferase and histone deacetylase activities), we will look at global mRNA

expression analysis after knockdown to see if gene signature patterns match PRC2 and/or other complexes.

Aim 2: Development of a Fluorescence Polarization assay to gauge the chemical tractability of the RBBP4-histone H3 interaction.

RBBP4/7 have several binding surfaces, but the canonical binding site where RBBP4/7 interacts with histone H3 was shown to be important in regulation of complex activity (e.g. when RBBP4/7 are unable to bind histone H3, PRC2 catalytic activity dramatically decreases). Therefore, we chose this pocket to examine for potential small-molecule inhibitors. We will also examine other binding partners of this site, including oncogenic transcription factors SALL4 and BCL11A. Through multiple drug screening techniques (HTS, thermal shift and octet red), we hope to find an inhibitor that can be used to probe the importance of this PPI in TNBC and epigenetic regulation.

At the end of this project, we expect to elucidate the functions of RBBP4/7 in TNBC, with special regards to PRC2 where RBBP4/7's role is not well established. Furthermore, the FP assay will help with the discovery of a small-molecule inhibitor of RBBP4/7. We anticipate this small-molecule would inhibit RBBP4/7's interaction with histone H3 and the oncogenic transcription factor BCL11A. An inhibitor against this PPI will also clarify the importance of RBBP4/7 in recruitment of epigenetic complexes to specific genetic loci and in stabilizing these complexes association with the nucleosome.

References

1 DeVita, V. T., Jr. & Chu, E. A history of cancer chemotherapy. *Cancer Res* **68**, 8643-8653, doi:10.1158/0008-5472.CAN-07-6611 (2008).

2 Morrison, W. B. Cancer chemotherapy: an annotated history. *J Vet Intern Med* **24**, 1249-1262, doi:10.1111/j.1939-1676.2010.0590.x (2010).

3 Masta, A., Gray, P. J. & Phillips, D. R. Nitrogen mustard inhibits transcription and translation in a cell free system. *Nucleic Acids Res* **23**, 3508-3515 (1995).

4 Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353-357, doi:10.1038/nature04296 (2006).

5 Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**, 2750-2767, doi:10.1172/JCI45014 (2011).

6 Bombonati, A. & Sgroi, D. C. The molecular pathology of breast cancer progression. *J Pathol* **223**, 307-317, doi:10.1002/path.2808 (2011).

7 McGee, S. F., Lanigan, F., Gilligan, E. & Groner, B. Mammary gland biology and breast cancer. Conference on Common Molecular Mechanisms of Mammary Gland Development and Breast Cancer Progression. *EMBO Rep* **7**, 1084-1088, doi:10.1038/sj.embor.7400839 (2006).

8 National Comprehensive Cancer Network. NCCN Clinical practices guidelines in oncology: Breast cancer. V.3.2012. http://www.nccn.org.

9 Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70, doi:10.1038/nature11412 (2012).

10 Andrew, H. S., Anthony, H., Sacha, J. H. & Robert, B. C. Origins of breast cancer subtypes and therapeutic implications. *Nature clinical practice. Oncology* **4**, 516-525, doi:10.1038/ncponc0908 (2007).

11 Wiggans, R. G. *et al.* Phase-II trial of tamoxifen in advanced breat cancer. *Cancer chemotherapy and pharmacology* **3**, 45-48 (1979).

12 Geisler, J. *et al.* In vivo inhibition of aromatization by exemestane, a novel irreversible aromatase inhibitor, in postmenopausal breast cancer patients. *Clin Cancer Res* **4**, 2089-2093 (1998).

13 Anindita, C. *et al.* Trastuzumab-resistant cells rely on a HER2-PI3K-FoxO-survivin axis and are sensitive to PI3K inhibitors. *Cancer research* **73**, 1190-1200, doi:10.1158/0008-5472.can-12-2440 (2013).

14 Bartsch, R., Wenzel, C. & Steger, G. G. Trastuzumab in the management of early and advanced stage breast cancer. *Biologics : Targets & Therapy* **1**, 19-31 (2007).

15 Nahta, R. & Esteva, F. J. Trastuzumab: triumphs and tribulations. *Oncogene* **26**, 3637-3643 (2007).

16 Collins, D. M. *et al.* Trastuzumab induces antibody-dependent cell-mediated cytotoxicity (ADCC) in HER-2-non-amplified breast cancer cell lines. *Ann Oncol* **23**, 1788-1795, doi:10.1093/annonc/mdr484 (2012).

17 van 't Veer, L. J. *et al.* Expression profiling predicts outcome in breast cancer. *Breast Cancer Res* **5**, 57-58 (2003).

18 van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009, doi:10.1056/NEJMoa021967 (2002).
19 Weigelt, B. *et al.* Refinement of breast cancer classification by molecular characterization of histological special types. *The Journal of pathology* **216**, 141-150, doi:10.1002/path.2407 (2008).
20 Esserman, L. J. *et al.* Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res Treat* **132**, 1049-1062, doi:10.1007/s10549-011-1895-2 (2012).
21 Wan, X. *et al.* Lapatinib-loaded human serum albumin nanoparticles for the prevention and treatment of triple-negative breast cancer metastasis to the brain. *Oncotarget*, doi:10.18632/oncotarget.8697 (2016).
22 Kalous, O. *et al.* Dacomitinib (PF-00299804), an irreversible Pan-HER inhibitor, inhibits proliferation of HER2-amplified breast cancer cell lines resistant to trastuzumab and lapatinib. *Mol Cancer Ther* **11**, 1978-1987, doi:10.1158/1535-7163.MCT-11-0730 (2012).
23 Majumder, M. *et al.* COX-2 Induces Breast Cancer Stem Cells via EP4/PI3K/AKT/NOTCH/WNT Axis. *Stem Cells*, doi:10.1002/stem.2426 (2016).
24 Gokmen-Polar, Y. *et al.* Investigational drug MLN0128, a novel TORC1/2 inhibitor, demonstrates potent oral antitumor activity in human breast cancer xenograft models. *Breast Cancer Res Treat* **136**, 673-682, doi:10.1007/s10549-012-2298-8 (2012).
25 Mayer, I. A. *et al.* A Phase Ib Study of Alpelisib (BYL719), a PI3Kalpha-specific Inhibitor, with Letrozole in ER+/HER2-Negative Metastatic Breast Cancer. *Clin Cancer Res*, doi:10.1158/1078-0432.CCR-16-0134 (2016).
26 Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci* **36**, 320-328, doi:10.1016/j.tibs.2011.03.006 (2011).
27 Charafe-Jauffret, E. *et al.* Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer research* **69**, 1302-1313, doi:10.1158/0008-5472.CAN-08-2741 (2009).
28 Gangemi, R. *et al.* Cancer stem cells: a new paradigm for understanding tumor growth and progression and drug resistance. *Current medicinal chemistry* **16**, 1688-1703 (2009).
29 Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-3988, doi:10.1073/pnas.0530291100 (2003).
30 Olivier, G., Christoph, A. K. & Patrick, A. B. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature reviews. Cancer* **8**, 755-768, doi:10.1038/nrc2499 (2008).
31 Vidal, S. J., Rodriguez-Bravo, V., Galsky, M., Cordon-Cardo, C. & Domingo-Domenech, J. Targeting cancer stem cells to suppress acquired chemotherapy resistance. *Oncogene* **33**, 4451-4463, doi:10.1038/onc.2013.411 (2014).

32 Dawood, S., Austin, L. & Cristofanilli, M. Cancer stem cells: implications for cancer therapy. *Oncology (Williston Park)* **28**, 1101-1107, 1110 (2014).
33 Pattabiraman, D. R. & Weinberg, R. A. Tackling the cancer stem cells - what challenges do they pose? *Nat Rev Drug Discov* **13**, 497-512, doi:10.1038/nrd4253 (2014).
34 Mani, S. A. *et al.* The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell* **133**, 704-715 (2008).
35 Takebe, N. *et al.* Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat Rev Clin Oncol* **12**, 445-464, doi:10.1038/nrclinonc.2015.61 (2015).
36 Zeng, Y. A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**, 568-577, doi:10.1016/j.stem.2010.03.020 (2010).
37 Khramtsov, A. I. *et al.* Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. *Am J Pathol* **176**, 2911-2920, doi:10.2353/ajpath.2010.091125 (2010).
38 Anastas, J. N. & Moon, R. T. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* **13**, 11-26, doi:10.1038/nrc3419 (2013).
39 Jang, G. B. *et al.* Blockade of Wnt/beta-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. *Sci Rep* **5**, 12465, doi:10.1038/srep12465 (2015).
40 Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85-89, doi:10.1038/nature10694 (2012).
41 Schlange, T., Matsuda, Y., Lienhard, S., Huber, A. & Hynes, N. E. Autocrine WNT signaling contributes to breast cancer cell proliferation via the canonical WNT pathway and EGFR transactivation. *Breast Cancer Res* **9**, R63, doi:10.1186/bcr1769 (2007).
42 Eterno, V. *et al.* AurkA controls self-renewal of breast cancer-initiating cells promoting wnt3a stabilization through suppression of miR-128. *Sci Rep* **6**, 28436, doi:10.1038/srep28436 (2016).
43 Flemban, A. & Qualtrough, D. The Potential Role of Hedgehog Signaling in the Luminal/Basal Phenotype of Breast Epithelia and in Breast Cancer Invasion and Metastasis. *Cancers (Basel)* **7**, 1863-1884, doi:10.3390/cancers7030866 (2015).
44 Han, B. *et al.* FOXC1 Activates Smoothened-Independent Hedgehog Signaling in Basal-like Breast Cancer. *Cell Rep* **13**, 1046-1058, doi:10.1016/j.celrep.2015.09.063 (2015).
45 Cheng, J., Gao, J., Tao, K. & Yu, P. Prognostic role of Gli1 expression in solid malignancies: a meta-analysis. *Sci Rep* **6**, 22184, doi:10.1038/srep22184 (2016).
46 Noman, A. S. *et al.* Overexpression of sonic hedgehog in the triple negative breast cancer: clinicopathological characteristics of high burden breast cancer patients from Bangladesh. *Sci Rep* **6**, 18830, doi:10.1038/srep18830 (2016).

47 Wang, F. *et al.* Hedgehog Signaling Regulates Epithelial-Mesenchymal Transition in Pancreatic Cancer Stem-Like Cells. *J Cancer* **7**, 408-417, doi:10.7150/jca.13305 (2016).
48 Sendurai, A. M. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704-715, doi:10.1016/j.cell.2008.03.027 (2008).
49 Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770-776 (1999).
50 Izrailit, J. & Reedijk, M. Developmental pathways in breast cancer and breast tumor-initiating cells: therapeutic implications. *Cancer letters* **317**, 115-126, doi:10.1016/j.canlet.2011.11.028 (2012).
51 Li, L. *et al.* Notch-1 signaling activates NF-kappaB in human breast carcinoma MDA-MB-231 cells via PP2A-dependent AKT pathway. *Med Oncol* **33**, 33, doi:10.1007/s12032-016-0747-7 (2016).
52 Zhao, D. *et al.* NOTCH-induced aldehyde dehydrogenase 1A1 deacetylation promotes breast cancer stem cells. *J Clin Invest* **124**, 5453-5465, doi:10.1172/JCI76611 (2014).
53 D'Angelo, R. C. *et al.* Notch reporter activity in breast cancer cell lines identifies a subset of cells with stem cell activity. *Mol Cancer Ther* **14**, 779-787, doi:10.1158/1535-7163.MCT-14-0228 (2015).
54 Karin, M. NF-kappaB as a critical link between inflammation and cancer. *Cold Spring Harb Perspect Biol* **1**, a000141, doi:10.1101/cshperspect.a000141 (2009).
55 Iliopoulos, D., Jaeger, S. A., Hirsch, H. A., Bulyk, M. L. & Struhl, K. STAT3 Activation of miR-21 and miR-181b-1 via PTEN and CYLD Are Part of the Epigenetic Switch Linking Inflammation to Cancer. *Molecular Cell* **39**, 493-506, doi:http://dx.doi.org/10.1016/j.molcel.2010.07.023 (2010).
56 Kim, S.-Y. *et al.* Role of the IL-6-JAK1-STAT3-Oct-4 pathway in the conversion of non-stem cancer cells into cancer stem-like cells. *Cellular Signalling* **25**, 961-969, doi:http://dx.doi.org/10.1016/j.cellsig.2013.01.007 (2013).
57 Li, L. *et al.* Evaluation of STAT3 Signaling in ALDH+ and ALDH+/CD44+/CD24− Subpopulations of Breast Cancer Cells. *PLoS ONE* **8**, doi:10.1371/journal.pone.0082821 (2013).
58 Anne, F. S. *et al.* Preclinical and clinical studies of gamma secretase inhibitors with docetaxel on human breast tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 1512-1524, doi:10.1158/1078-0432.CCR-11-3326 (2013).
59 Krop, I. *et al.* Phase I pharmacologic and pharmacodynamic study of the gamma secretase (Notch) inhibitor MK-0752 in adult patients with advanced solid tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 2307-2313, doi:10.1200/JCO.2011.39.1540 (2012).
60 Jang, G. B. *et al.* Wnt/beta-Catenin Small-Molecule Inhibitor CWP232228 Preferentially Inhibits the Growth of Breast Cancer Stem-like Cells. *Cancer Res*, doi:10.1158/0008-5472.CAN-14-2041 (2015).

61 Timp, W. & Feinberg, A. P. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nature reviews. Cancer* **13**, 497-510, doi:10.1038/nrc3486 (2013).
62 Kagara, N. *et al.* Epigenetic regulation of cancer stem cell genes in triple-negative breast cancer. *Am J Pathol* **181**, 257-267, doi:10.1016/j.ajpath.2012.03.019 (2012).
63 Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-734, doi:10.1038/nrc3130 (2011).
64 Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat Rev Genet* **13**, 795-806, doi:10.1038/nrg3317 (2012).
65 Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499-507, doi:10.1038/ng.127 (2008).
66 Agoston, A. T. *et al.* Increased protein stability causes DNA methyltransferase 1 dysregulation in breast cancer. *The Journal of biological chemistry* **280**, 18302-18310, doi:10.1074/jbc.M501675200 (2005).
67 Hernandez-Vargas, H. *et al.* Methylome analysis reveals Jak-STAT pathway deregulation in putative breast cancer stem cells. *Epigenetics : official journal of the DNA Methylation Society* **6**, 428-439 (2011).
68 Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**, 5427-5440, doi:10.1038/sj.onc.1205600 (2002).
69 Qu, Z. *et al.* Epigenetic repression of PDZ-LIM domain-containing protein 2: implications for the biology and treatment of breast cancer. *The Journal of biological chemistry* **285**, 11786-11792, doi:10.1074/jbc.M109.086561 (2010).
70 Qu, Z. *et al.* DNA methylation-dependent repression of PDZ-LIM domain-containing protein 2 in colon cancer and its role as a potential therapeutic target. *Cancer Res* **70**, 1766-1772, doi:10.1158/0008-5472.CAN-09-3263 (2010).
71 Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21-33, doi:10.1038/nrg1748 (2006).
72 Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692, doi:10.1016/j.cell.2007.01.029 (2007).
73 Pathania, R. *et al.* Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth. *Cancer Res* **76**, 3224-3235, doi:10.1158/0008-5472.CAN-15-2249 (2016).
74 Hollenbach, P. W. *et al.* A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines. *PLoS One* **5**, e9001, doi:10.1371/journal.pone.0009001 (2010).
75 Tsai, H. C. *et al.* Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell* **21**, 430-446, doi:10.1016/j.ccr.2011.12.029 (2012).

76 Munster, P. N. *et al.* The histone deacetylase inhibitor suberoylanilide hydroxamic acid induces differentiation of human breast cancer cells. *Cancer Res* **61**, 8492-8497 (2001).
77 Robertson, F. M. *et al.* Suberoylanilide hydroxamic acid blocks self-renewal and homotypic aggregation of inflammatory breast cancer spheroids. *Cancer* **116**, 2760-2767, doi:10.1002/cncr.25176 (2010).
78 Shah, P., Gau, Y. & Sabnis, G. Histone deacetylase inhibitor entinostat reverses epithelial to mesenchymal transition of breast cancer cells by reversing the repression of E-cadherin. *Breast Cancer Res Treat* **143**, 99-111, doi:10.1007/s10549-013-2784-7 (2014).
79 Bansal, N. *et al.* Targeting the SIN3A-PF1 interaction inhibits epithelial to mesenchymal transition and maintenance of a stem cell phenotype in triple negative breast cancer. *Oncotarget* **6**, 34087-34105, doi:10.18632/oncotarget.6048 (2015).
80 Farias, E. F. *et al.* Interference with Sin3 function induces epigenetic reprogramming and differentiation in breast cancer cells. *Proc Natl Acad Sci U S A* **107**, 11811-11816, doi:10.1073/pnas.1006737107 (2010).
81 Kwon, Y. J. *et al.* Selective Inhibition of SIN3 Corepressor with Avermectins as a Novel Therapeutic Strategy in Triple-Negative Breast Cancer. *Mol Cancer Ther* **14**, 1824-1836, doi:10.1158/1535-7163.MCT-14-0980-T (2015).
82 Robert, A. C. Molecular pathways: protein methyltransferases in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 6344-6350, doi:10.1158/1078-0432.ccr-13-0223 (2013).
83 Zhang, L. *et al.* Inhibition of histone H3K79 methylation selectively inhibits proliferation, self-renewal and metastatic potential of breast cancer. *Oncotarget* **5**, 10665-10677 (2014).
84 Kryczek, I. *et al.* IL-22(+)CD4(+) T cells promote colorectal cancer stemness via STAT3 transcription factor activation and induction of the methyltransferase DOT1L. *Immunity* **40**, 772-784, doi:10.1016/j.immuni.2014.03.010 (2014).
85 Cao, F. *et al.* Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Molecular cell* **53**, 247-261, doi:10.1016/j.molcel.2013.12.001 (2014).
86 Latysheva, N. S. & Babu, M. M. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res* **44**, 4487-4503, doi:10.1093/nar/gkw282 (2016).
87 Matkar, S. *et al.* An Epigenetic Pathway Regulates Sensitivity of Breast Cancer Cells to HER2 Inhibition via FOXO/c-Myc Axis. *Cancer Cell* **28**, 472-485, doi:10.1016/j.ccell.2015.09.005 (2015).
88 Chen, J. *et al.* Pygo2 associates with MLL2 histone methyltransferase and GCN5 histone acetyltransferase complexes to augment Wnt target gene expression and breast cancer stem-like cell expansion. *Mol Cell Biol* **30**, 5621-5635, doi:10.1128/MCB.00465-10 (2010).

89 Grebien, F. *et al.* Pharmacological targeting of the Wdr5-MLL interaction in C/EBPalpha N-terminal leukemia. *Nat Chem Biol* **11**, 571-578, doi:10.1038/nchembio.1859 (2015).
90 Cao, F. *et al.* Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell* **53**, 247-261, doi:10.1016/j.molcel.2013.12.001 (2014).
91 Abou El Hassan, M. *et al.* Cancer Cells Hijack PRC2 to Modify Multiple Cytokine Pathways. *PloS one* **10**, doi:10.1371/journal.pone.0126466 (2015).
92 Andrew, C. & Nicholas, C. C. Aberrations of EZH2 in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 2613-2618, doi:10.1158/1078-0432.ccr-10-2156 (2011).
93 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349, doi:10.1038/nature09784 (2011).
94 Teruyuki, S. *et al.* PRC2 overexpression and PRC2-target gene repression relating to poorer prognosis in small cell lung cancer. *Scientific Reports* **3**, doi:10.1038/srep01911 (2013).
95 Vinod, N., Chao, X. & Jinrong, M. Composition, recruitment and regulation of the PRC2 complex. *Nucleus (Austin, Tex.)* **2**, 277-282, doi:10.4161/nucl.2.4.16266 (2010).
96 Vizan, P., Beringer, M., Ballare, C. & Di Croce, L. Role of PRC2-associated factors in stem cells and disease. *FEBS J* **282**, 1723-1735, doi:10.1111/febs.13083 (2015).
97 Jeffrey, A. S. & Robert, E. K. Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nature Reviews Molecular Cell Biology*, doi:10.1038/nrm2763 (2009).
98 Adrian, P. B., Nikolaj, D., Diego, P., Klaus, H. H. & Kristian, H. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development* **20**, 1123-1136, doi:10.1101/gad.381706 (2006).
99 Liu, S. *et al.* Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res* **66**, 6063-6071, doi:10.1158/0008-5472.CAN-06-0054 (2006).
100 Kreso, A. *et al.* Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36, doi:10.1038/nm.3418 (2014).
101 Alba, J.-S. *et al.* Expression of Polycomb Targets Predicts Breast Cancer Prognosis. *Molecular and Cellular Biology* **33**, 3951-3961, doi:10.1128/mcb.00426-13 (2013).
102 Antonis, K. *et al.* Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes & Development* **18**, 1592-1605, doi:10.1101/gad.1200204 (2004).
103 Cheng, A. S. *et al.* EZH2-mediated concordant repression of Wnt antagonists promotes beta-catenin-dependent hepatocarcinogenesis. *Cancer Res* **71**, 4028-4039, doi:10.1158/0008-5472.CAN-10-3342 (2011).
104 Jindan, Y. *et al.* A Polycomb Repression Signature in Metastatic Prostate Cancer Predicts Cancer Outcome. *Cancer Research* **67**, 10657-10663, doi:10.1158/0008-5472.CAN-07-2498 (2007).

105 Fujii, S. *et al.* MEK-ERK pathway regulates EZH2 overexpression in association with aggressive breast cancer subtypes. *Oncogene* **30**, 4118-4128, doi:10.1038/onc.2011.118 (2011).
106 Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100**, 11606-11611, doi:10.1073/pnas.1933744100 (2003).
107 Filippos, K. *et al.* NDY1/KDM2B functions as a master regulator of polycomb complexes and controls self-renewal of breast cancer stem cells. *Cancer research* **74**, 3935-3946, doi:10.1158/0008-5472.CAN-13-2733 (2014).
108 Yuan-Liang, L. W. *et al.* Combined inhibition of EGFR and c-ABL suppresses the growth of triple-negative breast cancer growth through inhibition of HOTAIR. *Oncotarget* (2015).
109 Zhang, K. *et al.* Long non-coding RNA HOTAIR promotes glioblastoma cell cycle progression in an EZH2 dependent manner. *Oncotarget* **6**, 537-546 (2014).
110 Zhang, H. *et al.* MiR-7, inhibited indirectly by lincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by downregulating the STAT3 pathway. *Stem Cells* **32**, 2858-2868, doi:10.1002/stem.1795 (2014).
111 Bai, W.-D. *et al.* MiR-200c suppresses TGF-β signaling and counteracts trastuzumab resistance and metastasis by targeting ZNF217 and ZEB1 in breast cancer. *International Journal of Cancer* **135**, 1356-1368, doi:10.1002/ijc.28782 (2014).
112 Iliopoulos, D. *et al.* Loss of miR-200 inhibition of Suz12 leads to polycomb-mediated repression required for the formation and maintenance of cancer stem cells. *Mol Cell* **39**, 761-772, doi:10.1016/j.molcel.2010.08.013 (2010).
113 Campbell, J. E. *et al.* EPZ011989, A Potent, Orally-Available EZH2 Inhibitor with Robust in Vivo Activity. *ACS medicinal chemistry letters* **6**, 491-495, doi:10.1021/acsmedchemlett.5b00037 (2015).
114 Garapaty-Rao, S. *et al.* Identification of EZH2 and EZH1 small molecule inhibitors with selective impact on diffuse large B cell lymphoma cell growth. *Chemistry & biology* **20**, 1329-1339, doi:10.1016/j.chembiol.2013.09.013 (2013).
115 Sarah, K. K. *et al.* A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nature Chemical Biology* **8**, 890-896, doi:10.1038/nchembio.1084 (2012).
116 Verma, S. K. *et al.* Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS medicinal chemistry letters* **3**, 1091-1096, doi:10.1021/ml3003346 (2012).
117 Kim, K. H. *et al.* SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat Med* **21**, 1491-1496, doi:10.1038/nm.3968 (2015).
118 Kim, W. *et al.* Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. *Nature chemical biology* **9**, 643-650, doi:10.1038/nchembio.1331 (2013).

119 Kong, X. *et al.* Astemizole arrests the proliferation of cancer cells by disrupting the EZH2-EED interaction of polycomb repressive complex 2. *Journal of medicinal chemistry* **57**, 9512-9521, doi:10.1021/jm501230c (2014).
120 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
121 Ketel, C. S. *et al.* Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. *Mol Cell Biol* **25**, 6857-6868, doi:10.1128/MCB.25.16.6857-6868.2005 (2005).
122 Zaidi, S. K. *et al.* Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. *Molecular and cellular biology* **30**, 4758-4766, doi:10.1128/MCB.00646-10 (2010).
123 Prat, A. & Perou, C. M. Mammary development meets cancer genomics. *Nature medicine* **15**, 842-844, doi:10.1038/nm0809-842 (2009).

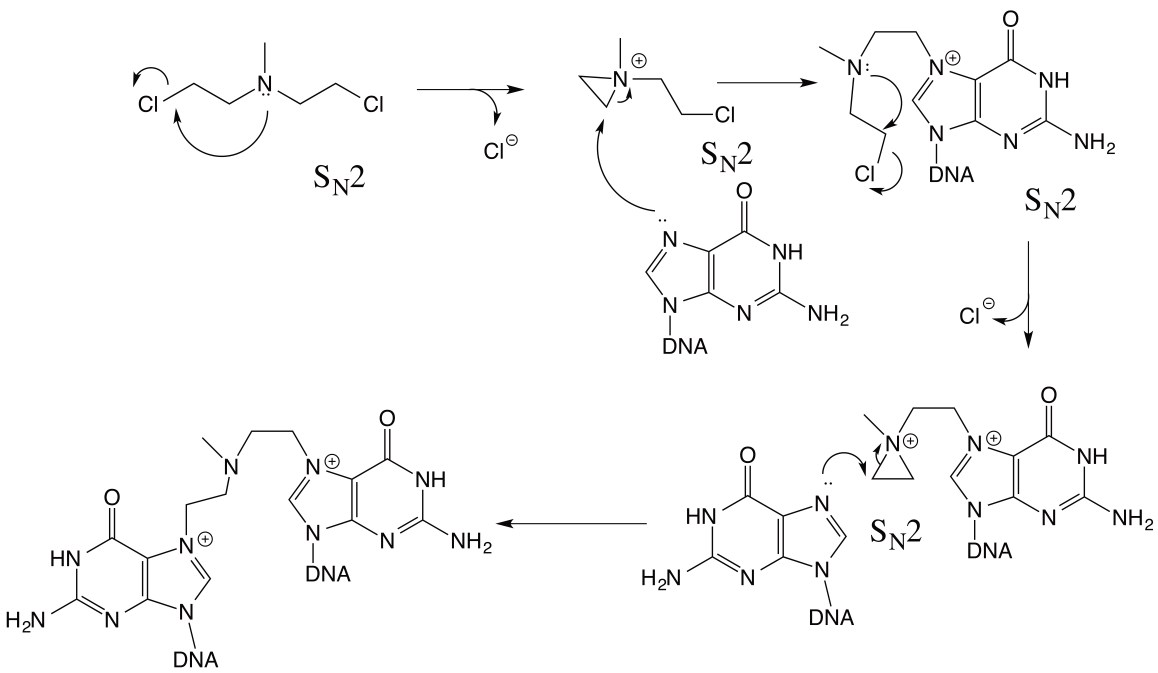


Figure 1.1 DNA Alkylation by Nitrogen Mustard.
The initial step is the formation of the aziridinium cation, which occurs when electrons from the nitrogen atom displaces the chloride. The resulting cation is susceptible to nucleophilic attack. As shown above the N-7 position of guanine is strongly nucleophilic and can be easily alkylated by the aziridinium ion.

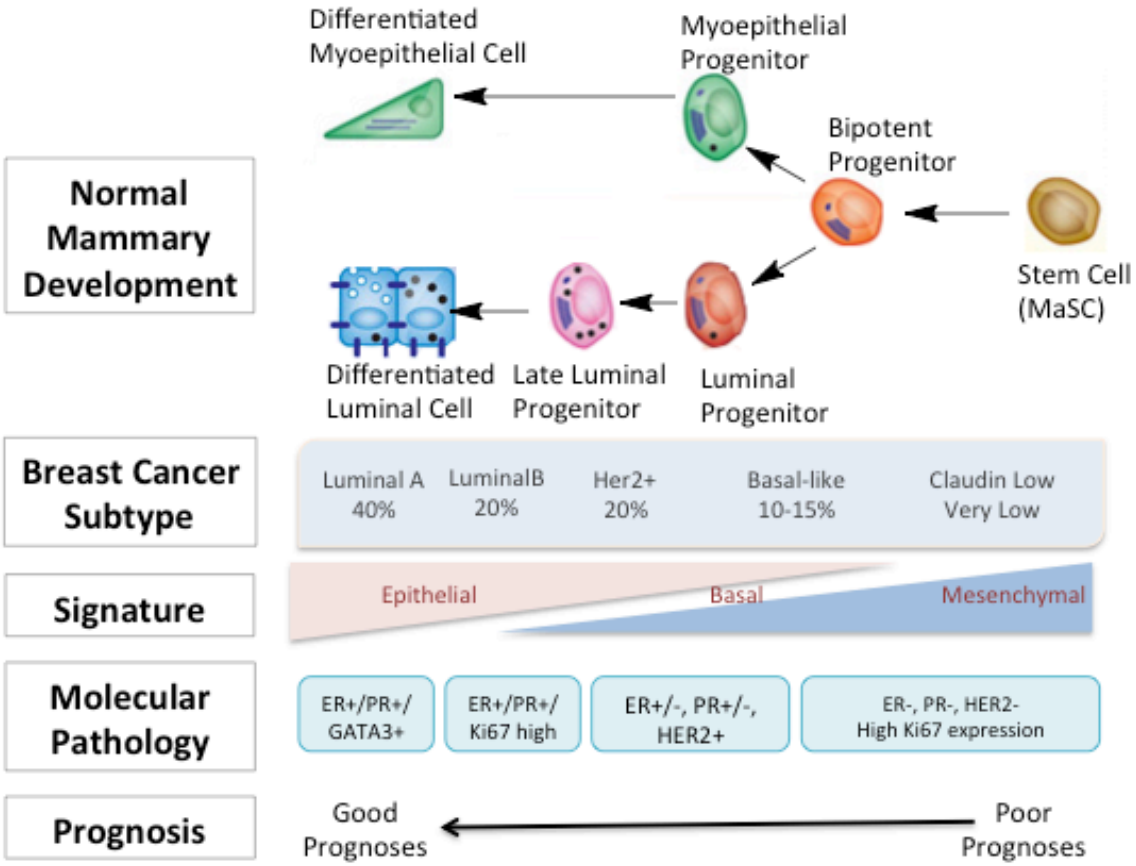


Figure 1.2 Breast Cancer Subtype.
The above figure demonstrates the normal mammary development in breast tissues and the characteristics of the various breast cancer subtypes. Luminal A is the most common form of breast cancer and it usually has low histological grade tumors with good prognosis. Luminal B, on the other hand, has higher grade tumors with a more aggressive phenotype and higher recurrence than Luminal A. The main difference between Luminal A and B tumors are the expression of proliferation genes. HER2 tumors are characterized by the overexpression of HER2 and also have a more aggressive nature with poor prognosis if left untreated. However, HER2+ patients can be treated with trastuzumab and may be more sensitive to other traditional chemotherapies. Basal-like breast cancers are triple negative, ER-/PR-/HER-, and have an aggressive nature with high rates for metastasis to the lung or the brain. TNBC does not have any current targeted treatment. Adapted from ref. 123 and 43.

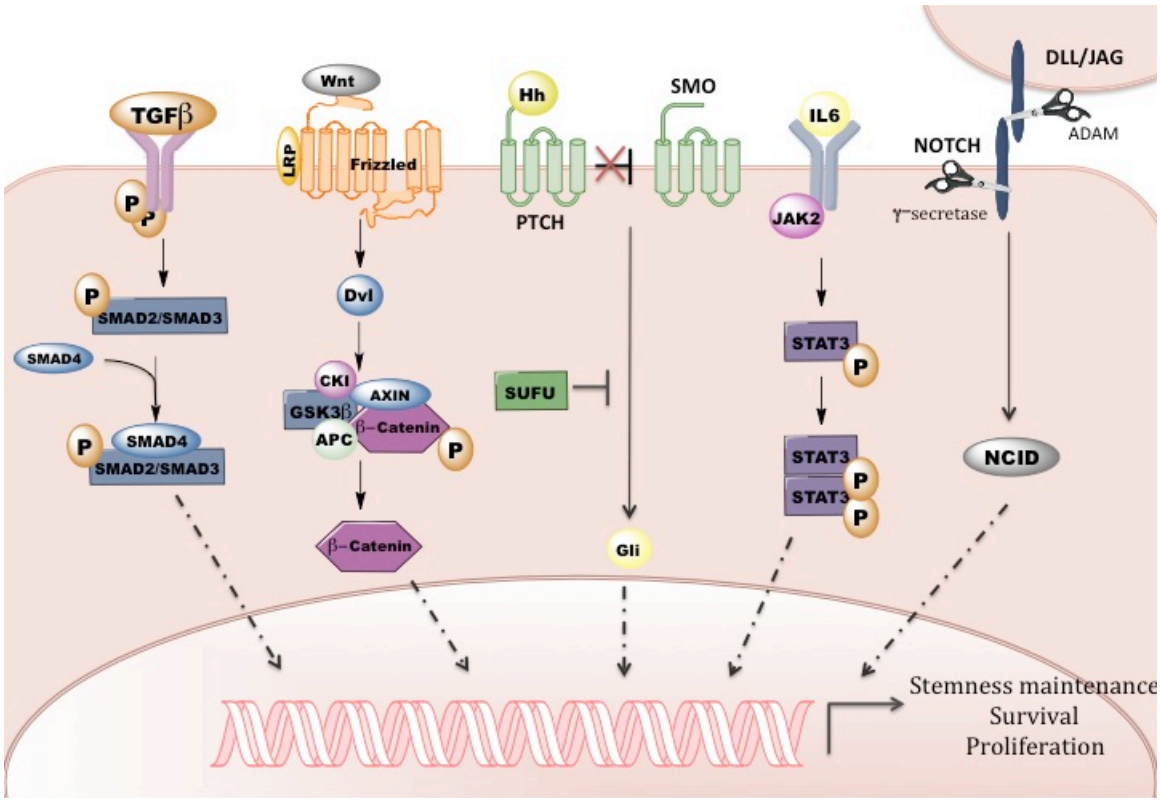


Figure 1.3 Signaling Pathways in Cancer Stem Cells.
The figure above focuses on the minimum necessities of the WNT, Hedgehog (Hh) and Notch pathways in breast cancer. All of these pathways are thought to maintain stemness, proliferation and survival of breast cancer stem cells. There is accumulating evidence that these pathways work in concert with each other to promote tumorigenesis. Figure adapted from ref. 33.

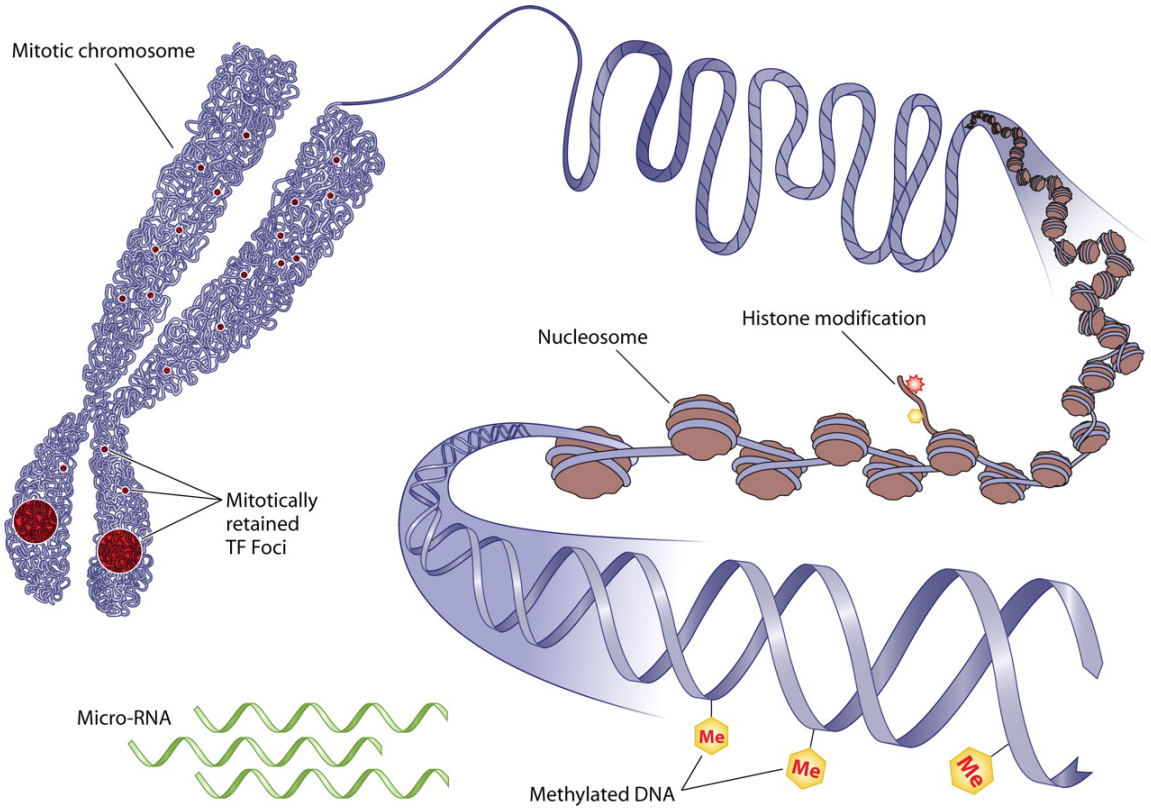


Figure 1.4 The Basis for Epigenetic Control.
Diagram from ref. 122. First and foremost, there must be an understanding of the way in which epigenetic states are created, maintained and altered in tumorigenesis. Epigenetics is intimately tied to the regulation of all DNA-based processes (transcription, DNA repair, and replication). It orchestrates this regulation through influence over the nucleosome and chromatin architecture in the cell and tight control over the access to transcriptionally available DNA. By modification to the histone tails and DNA methylation, nucleosomes are organized into either open euchromatin (the "beads on a string" model) or condensed heterochromatin. Writer, readers and erasers are names given to epigenetic proteins responsible for placing, sensing or removing these marks. Histone acetyltransferases, for example, remove the positive charge of lysine residues, thereby weakening the association between negatively charged DNA and histones; this allows access to RNA polymerases and transcription factors.
Deconstructing epigenetic pathways important for cancer cell stemness will have a far-reaching impact on the treatment of cancer. As chromatin remodelers and epigenetic enzymes dictate the chromatin landscape, they are crucial mediators of cell-fate decisions: self-renewal vs. differentiation, apoptosis vs. proliferation. Epigenetic-targeted therapies could reprogram the cancer cell, promote differentiation and abrogate self-renewal more efficiently rendering the cancer sensitive to conventional chemotherapies. Ref. 122 Zaidi, S. K. *et al.* Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. *Molecular and cellular biology* **30**, 4758-4766, doi:10.1128/MCB.00646-10 (2010).

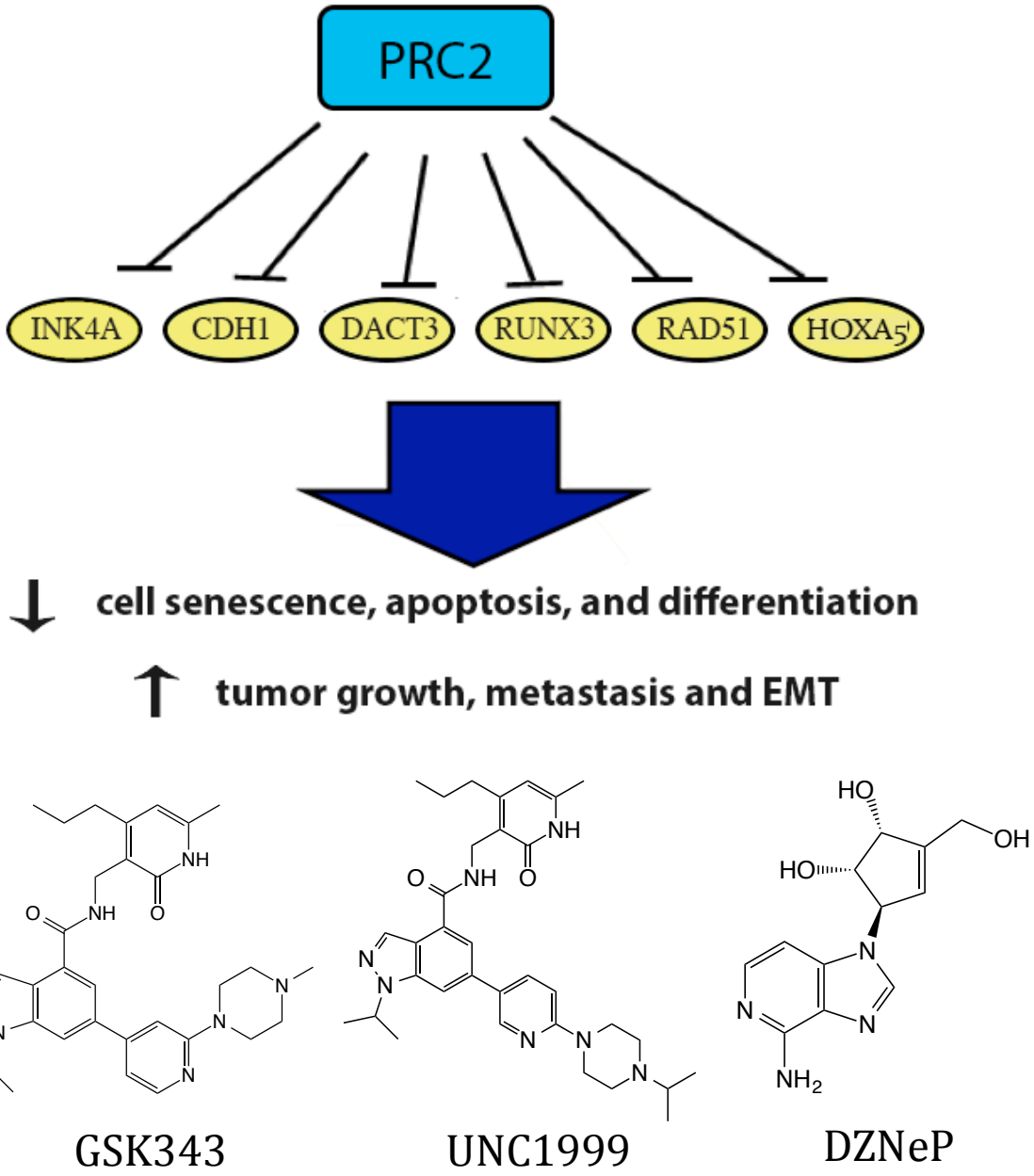


Figure 1.5 PRC2 Targets and Current Inhibitors of EZH2

Chapter 2

RBBP4 and RBBP7 regulate multiple pathways to maintain breast cancer stem cells in triple negative breast cancer

Abstract.

Around 20% of all breast cancers are classified as triple-negative (TNBC). While TNBC is initially responsive to traditional chemotherapies, it can be particularly aggressive and has a higher rate of recurrence than other breast cancer subtypes. It is therefore critical to understand the biological characteristics of TNBC in order to develop successful targeted therapies, like those that exist for receptor-positive breast cancer. In this study, we found the expression of two histone chaperones, retinoblastoma-binding proteins 4 and 7 (RBBP4/7), to be higher in breast cancer and in particular, TNBC, compared to normal breast tissue. Their functional role, however, remains unclear. To evaluate their importance in TNBC, we performed knockdown studies via RNA interference and found RBBP4/7 to be critical in the maintenance of bulk cancer cell proliferation and cancer stem cells (CSC), a population attributed to the heterogeneity of TNBC as well as metastasis and chemoresistance. As a central component to several epigenetic co-repressors, such as NuRD, SIN3A and PRC2, RBBP4/7 knockdown dramatically changed the

epigenetic landscape. RNA sequencing revealed the genetic ramifications of these epigenetic changes with 3319 genes up- and 2605 genes down-regulated by at least 1.2 fold in RBBP4/7 knockdown cells compared to negative control. Gene Set Enrichment Analysis (GSEA) demonstrated that reduced expression of RBBP4/7 was correlated with a decrease in the basal, or TNBC, phenotype as well as a decrease in the stem cell signature. In addition, siRNA against RBBP4/7 significantly reduced tumor growth of SUM149 cells *in vivo*. Our data suggests that RBBP4/7 may provide a novel therapeutic target against TNBC and CSC.

2.1 Introduction.

Triple negative breast cancer (TNBC) is currently the only subtype of breast cancer for which there is no targeted therapy. The inherent heterogeneity of the disease renders it challenging to target and particularly aggressive with a poor prognosis. This difficulty is due in part to the presence of tumor-initiating cells, otherwise referred to as cancer stem cells (CSCs)[1-4] CSCs are characterized by their ability to (1) self-renew and (2) differentiate into cells that form the bulk of the tumor.[5,6] In addition, CSCs are associated with chemoresistance, and thereby tumor relapse and metastasis.[5,7-9] With over 90% of cancer fatalities due to metastatic spreading of the disease, it is imperative to better understand the molecular mechanisms behind cancer stem cells for the development of successful cancer therapeutics. While the focus of cancer research has traditionally been on genetic mechanisms driving the disease, e.g. gene mutations, it is now shifting toward a paradigm that includes epigenetic mechanisms as a medium for stable oncogenic transformation and acquisition of cancer hallmarks. Much of this transition stems

from recent studies indicating the widespread occurrence of mutations, amplifications, or deletions of epigenetic writers, readers and erasers in cancer.[10] In fact, recent research has suggested epigenetic mechanisms can not only contribute to the cancer stem cell phenotype, but also act as drivers for intratumoral heterogeneity and acquired resistance to chemotherapy.[11]

Retinoblastoma binding protein 4 and 7 (RBBP4/7) are two histone chaperones that work in tandem with epigenetic machinery in the cell to facilitate access to the genome and determine cellular identity. As members of the WD40 repeat protein family, they consist of a seven-bladed β-propeller and therefore have several interaction surfaces that help RBBP4/7 act as a keystone for many protein complexes.[12] Their role as integral parts of multisubunit epigenetic complexes is to bind to histones H3 and/or H4, allow for nucleosomal association, and expose the histones to post-translational modifications.[13-15]

RBBP4/7 help to repress gene transcription through the nucleosome remodeling and deacetylation (NuRD) complex, the polycomb repressive complex 2 (PRC2) and the SIN3A complex. All three of these complexes have been implicated in the maintenance of stem cells and the progression of TNBC.[16-23] Specifically, PRC2 represses inhibitors of both the WNT and NF-κB pathways, important signaling pathways utilized by CSCs.[16-18] Moreover, Enhancer of Zeste 2 (EZH2), the catalytic component of PRC2, is linked to aggressiveness and poor prognosis in breast cancer.[19] Whereas, the NuRD complex is believed to promote the epithelial-mesenchymal transition (EMT), a process that promotes metastasis, via its interactions with the oncogenic transcription factor TWIST1.[20,21] Lastly, the SIN3A

complex is a master regulator of STAT3 signaling and thus supports cancer stem cell self-renewal and tumorigenicity.[22,23]

Despite the aforementioned observations and the fact that RBBP4/7 are central to these three epigenetic complexes, very little is known about the contribution of RBBP4/7 to TNBC. In this study, we aim to better understand if RBBP4/7 could influence epigenetic marks, gene regulation and stemness of TNBC. Towards these questions, we performed knockdown studies in TNBC cell lines SUM149, MDA-MB-231 and MDA-MB-436 and evaluated the effect on the epigenetic environment and the cancer stem cell phenotype. We discovered that RBBP4/7 could dramatically alter global histone modifications, TNBC growth, and the frequency of the CD44+/CD24- cancer stem cell population. RBBP4 and RBBP7 exert control over multiple TNBC pathways and disruption of these two histone chaperones results in significantly decreased tumor growth. Taken together, our results indicate that RBBP4/7 might offer a unique therapeutic target for the treatment of TNBC.

2.2 Results

2.2.1 RBBP4 and RBBP7 are highly expressed in triple-negative breast cancer.

In order to better understand the importance of RBBP4 and RBBP7 in breast cancer, we first examined their expression levels in publicly available microarray data through Oncomine. We found that RBBP4 and RBBP7 are significantly overexpressed in breast cancer compared to normal breast in multiple data sets (Figure. 2.1). Importantly, some data sets showed that among breast cancer patients with different molecular subtypes, RBBP4 and RBBP7 expression levels

were higher in TNBC compared to non-TNBC (Figure 2.2). Furthermore, analysis of a panel of breast cancer cell lines confirmed that RBBP4 levels were higher in TNBC (SUM149, SUM159, MC1) versus non-TNBC (HCC1954, MCF7), normal breast tissue, and non-tumorigenic MCF10A (Figure 2.3). For clinical relevance, we sought to determine the significance of their high expression on metastasis and recurrence. Some studies showed that higher expression levels of RBBP4 and RBBP7 were correlated with metastasis and/or poor prognosis in invasive breast carcinoma (Fig. 2.4). These data suggest RBBP4 and RBBP7 could play a significant role in TNBC and could be potential therapeutic targets.

To test this hypothesis, we chose to interrupt RBBP4/7 via siRNA knockdown and assess the effects on various cellular processes including proliferation, cell cycle, and apoptosis and the changes, if any, in the cancer stem cell phenotype. As RBBP4 and RBBP7 are highly homologous, they can act through compensatory and redundant mechanisms.[12,24] Knockdown of one protein can often increase the expression of the other. For that reason, we used siRNA to disrupt both RBBP4 and RBBP7 as well as individually.

2.2.2 Knockdown of RBBP4 and RBBP7 dramatically alters the epigenetic landscape.

Once we confirmed an efficient silencing of RBBP4 and RBBP7 (Figure 2.5A), we first wanted to examine the epigenetic ramifications via changes in the histone post-translational modifications. We evaluated the level of histone marks 6 days post siRNA transfection owing to the prolonged half-life of histone methylation.[25] Dual knockdown resulted in dramatic alterations in global histone marks in the

invasive triple negative breast carcinoma line, SUM149 (Figure 2.5B) with increases in H3K9, H3K36 and H3K79 methylation. Of note, there was a significant reduction in the PRC2-associated repressive mark, H3K27me3. This is in contrast to other studies, which have seen little changes in global H3K27me3 levels after knockdown of p55, a homolog of RBBP4 or RBBP7 in neurospora, and rather only see decreases at specific subtelomeric domains.[26] Our results, however, support the notion that RBBP4/7 are important for nucleosomal association of PRC2 and its function. *In vitro* studies have shown that RBBP4/7 are not required for the histone methyltransferase (HMT) activity of PRC2, but may help promote the catalytic efficiency.[15,27] Knockdown of Nurf55 results in a threefold reduction in the complex's enzymatic activity, which is consistent with our data.[15] The resultant reduction in H3K27me3 in our knockdown studies is met with an increase in H3K27me2 (Figure 2.5B) indicating a potential stall in the conversion of dimethylation to trimethylation.

In contrast to EZH2 inhibition, either by siRNA/shRNA knockdown or small-molecule inhibitors, RBBP4/7 knockdown didn't lead to the reciprocal increase in the activating H3K27Ac mark (Figure 2.5B)[28,29] In addition, both SIN3A and NuRD have been linked to H3K27Ac with a slight increase in the mark upon knockdown of any substituents in either complex.[22,30] Even though we do not see this typical upregulation, it does not rule out changes of this mark at specific loci.[31] In place of H3K27Ac, we see a substantial increase in H3K36me3 and H3K79me3 (Figure 2.5B), two other activating marks. Interestingly, biochemical studies of H3K36me3 have revealed it to be inhibitory to the H3K27me3 mark.[32-34] These results potentially

demonstrate a different mode of action for the reactivation of genes upon knockdown of RBBP4/7 as opposed to EZH2. The global increases in H3K36 and H3K79 trimethylation could also point to ageing chromatin and reduction in proliferation capacity as H3K36me3 is thought to prevent the nucleosome turnover and H3K79me3 accumulates on ageing histones.[71,72]

On the other hand, higher levels of the repressive mark H3K9me3 were observed in our knockdown studies. H3K9me3 is associated with constitutive heterochromatin and is a hallmark of senescence.[24,35] Together RBBP4 and RBBP7 can positively and negatively regulate histone modifications, which may lead to changes in chromatin structure and access to genes.

2.2.3 Knockdown of RBBP4 and RBBP7 inhibits cell proliferation of TNBC cells.

To understand how the changes in the epigenetic environment brought on by RBBP4/7 silencing affect the phenotype of TNBC, we analyzed proliferation, cell cycle and apoptosis. Individually RBBP4 or RBBP7 inactivation only slightly decreased cell growth in comparison to negative control; however, dual knockdown dramatically reduced proliferation by 70% (Figure 2.6A). This inhibition is partially due to an increase in apoptosis from 12.4% to 27.6% as seen in Annexin V staining (Figure 2.6B). Cell cycle analysis also revealed a stall in progression at the G2/M phase after RBBP4/7 inactivation with a slight increase in multiploidy (Figure 2.6C, Table 2.1). Reduction in cell growth was also observed in two other TNBC cell lines MDA-MB-231 and MDA-MB-436 (Figure 2.6D and 2.6E).

SUM149 cells typically exist in two distinct populations with one adopting a mesenchymal spindle-like shape and the other presenting more epithelial

trademarks much like luminal cells. Bright field microscopy of hematoxylin-stained RBBP4/7 knockdown cells showed a dramatic morphological change (Figure 2.6F). Whereas control SUM149 cells showed the characteristic mixed populations, inactivation of RBBP4 and RBBP7 led to a complete shift in the equilibrium away from the mesenchymal and toward the epithelial phenotype. Furthermore, RBBP4/7 knockdown cells were significantly larger than their parental counterparts. These results suggest that RBBP4 and RBBP7 are required for TNBC growth and survival and that knockdown of both proteins may drive cell differentiation (also see below). Analysis of the CD44+/CD24- basal cell marker supported this observation with a substantial decrease in RBBP4/7 knockdown cells (0.74%) verse control (5.6%) (Figure 2.7A).

2.3.4 RBBP4 and RBBP7 are required for the maintenance of the cancer stem cell population in TNBC.

The expression of cell surface markers CD44 and CD24 has been linked to the cancer stem cell phenotype, with the CD44+/CD24- population having higher tumorigenic potential.[36,37] In human breast cancers, CD44+/CD24- cancer stem cells are located primarily at the invasive edge of the tumor and are enriched for genes associated with the epithelial-to-mesenchymal transition (EMT) .[4,38,39] To assess RBBP4/7's function in maintaining the breast CSC (BCSC) population, we examined the percentages of CD44+/CD24- cells. RBBP4/7 knockdown significantly decreased this population (5.6% vs 0.74%) (Figure 2.7A). In addition, we performed mammosphere assays. Sphere-forming assays are a standard *in vitro* method for assessing stem cell self-renewal, as only CSCs are able to survive the non-adherent

conditions. Knockdown of RBBP4/7 decreased the mammosphere formation capacity in SUM149 cells by almost 4 fold (3% vs 0.8%) confirming that these histone chaperones are important for maintaining the BCSC population (Figure 2.7B). Again this reduction in self-renewal capacity was verified in the MDA-MB-436 cell line (Figure 2.7C). Interestingly, when we ectopically expressed either RBBP4 or RBBP7 in SUM149 (Figure 2.7D), we observed an increase in both the number and the size of the mammospheres. These data suggest that RBBP4/7 not only help to maintain BCSCs, but also enhance their capacity for self-renewal.

2.3.5 Knockdown of RBBP4 and RBBP7 inhibits tumor growth of SUM149 cells in vivo.

Following the effect on BCSCs *in vitro*, we next wanted to determine the tumorigenic potential of cancer cells lacking RBBP4 and RBBP7 *in vivo*. We established stable shRNA knockdown SUM149 cell lines using the pGIPZ lentiviral vector. RBBP4/7 knockdown or negative control cells ($5x10^3$) were injected orthotopically into the fourth mammary fat pad of NOD/SCID mice and tumor development was monitored weekly. RBBP4/7 knockdown inhibited SUM149 tumor growth significantly by 11 weeks (Figure 2.8A, $p = 0.0336$). When a similar experiment was performed using siRNA-mediated knockdown of RBBP4 and RBBP7, we witnessed an even more pronounced tumor growth inhibitory effect (Figure 2.8B, $p = 0.0105$ for week 16) probably due to better knockdown efficiency achieved by siRNA than shRNA (data not shown). At the endpoint of the study (16 weeks) the resected tumors from control mice weighed on average 3.5-fold more than those from knockdown mice (Figure 2.8C, $p = 0.0106$). Importantly, when

these tumors were dissociated and the viable cells were subjected to flow cytometric analysis for the CD44+/CD24- BCSC marker, we found a dramatic decrease (11.6% vs 3.1%) in this population in RBBP4/7 knockdown tumors (Figure 2.8D). This finding recapitulates our results from *in vitro* studies and suggests the effects from RBBP4/7 knockdown is lasting.

2.3.6 RBBP4 and RBBP7 regulate multiple TNBC-related pathways.

In order to gain further insight into how RBBP4 and RBBP7 affect TNBC, we performed RNA sequencing (RNA-seq) to investigate changes in signaling pathways and gene expression patterns. To our knowledge this is the first study that examines changes in the transcriptome profile resulting from RBBP4 and RBBP7 knockdown in TNBC. We knocked down RBBP4 and RBBP7 simultaneously as well as individually by siRNA in SUM149 cells (Figure 2.9A). Consistent with the alterations to the epigenetic environment upon RBBP4/7 knockdown, we observed a similarly dramatic change in transcription with 3319 genes up- and 2605 genes down-regulated by at least 1.2 fold (Figure 2.9B). While RBBP4 and RBBP7 do have their distinct downstream-regulated genes, in some instances they work in concert with each other to upregulate (364 genes) or downregulate (447 genes) expression (Figure 2.9B). Comparing the RNA-seq profile of our control SUM149 cells (siControl), which were treated with negative siRNA, to our dual knockdown cells (siRBBP4+7), we found that RBBP4/7 are associated with several important TNBC pathways. Specifically, Gene Set Enrichment Analysis (GSEA) signified that genes normally suppressed in basal breast cancer were upregulated upon RBBP4/7

knockdown (Figure 2.9C).[40-42] On the other hand, genes expressed preferentially in this breast cancer subtype were downregulated (data not shown).

Furthermore, as expected we found that RBBP4/7 knockdown led to de-repression of potential PRC2 target genes. GSEA revealed that genes marked by H3K27me3 in embryonic stem cells are enriched in the RBBP4/7 knockdown phenotype (Figure 2.9D).[43] Among the 1118 genes with the H3K27me3 mark, 142 were upregulated by at least 1.2 fold upon RBBP4/7 knockdown (data not shown). The majority of these genes (104) are also bound by SUZ12 and/or EED in their genomic regions (data not shown).[43] As PRC2 is also known to promote tumorigenesis through the repression of tumor suppressors such as p16INK4a and p14ARF, we examined changes in the expression of mitotic genes and those that prevent aberrant cell growth.[44,45] Consistent with our *in vitro* data of cell proliferation, expression of cell cycle genes was negatively correlated with RBBP4/7 knockdown (Figure 2.9E).[46,47] In addition, targets of p53 were upregulated in cells treated with siRBBP4+7 (Figure 2.10A).[48,49] While p53 mRNA and protein levels did not increase in siRBBP4+7 (Figure 2.10B and C), MDM2, a p53 antagonist, was decreased in knockdown cells compared to control at both mRNA (Figure 2.10B) and protein (Figure 2.10C) levels. Interestingly, Weinken et al. recently reported that MDM2 is a supportive cofactor of PRC2 acting to maintain pluripotency and suppress differentiation.[50]

Reinforcing our hypothesis that RBBP4 and RBBP7 help maintain stemness of TNBC, we found that not only do they control the expression of not only PRC2 targets and MDM2, but also other oncogenic transcription factors including c-MYC

and BCL11A (Figure 2.11A) that are important for BCSCs. While BCL11A was recently found to be an important factor in the regulation of TNBC breast cancer stem cells, c-MYC is well established as a CSC effector.[51] c-MYC was found to repress the WNT inhibitor, DKK1, and we found simultaneous knockdown of RBBP4/7 resulted in a more than 4-fold increase in DKK1 expression (Figure 2.11B, $p < 0.001$).[52] Interestingly, DKK1 was also found to be a target of PRC2 so the effects on DKK1 expression seen in RBBP4/7 knockdown could be due to multiple factors.[53] We also witnessed about a 2-fold decrease in NOTCH1 expression (Figure 2.11C). From our analysis, RBBP4 and RBBP7 could contribute to BCSCs through two prominent cancer stem cell pathways: WNT and NOTCH signaling. Furthermore, GSEA demonstrated a decrease in stem-cell signature from dual knockdown cells in relation to siControl SUM149 cells. With the decrease in CSC genes and signaling, we observed a subsequent increase in genes that are associated with the more differentiated luminal phenotype (Figure 2.11D).[54,55] GATA3 ($p < 0.01$), KRT19 ($p < 0.01$) and CD24 ($p < 0.001$) expression increased upon knockdown of RBBP4/7. Together these results indicate that RBBP4/7 can promote stemness and block differentiation; therefore, they might provide a potential therapeutic target for more aggressive basal-like breast cancer.

2.4 Discussion.

RBBP4 and RBBP7 are two histone chaperones that are involved in nucleosome assembly and are shared amongst three crucial developmental epigenetic complexes: NuRD, SIN3A and PRC2. While each of these complexes has been implicated in the progression of TNBC, little is known about the role of RBBP4

or RBBP7. Operating within these machineries, RBBP4/7 are able to promote association with nucleosomes and help survey the surrounding environment for proper interpretation of the histone code.[13-15] Within PRC2, for instance, when the activating mark H3K4me3 is present, RBBP4/7 is unable to interact with histone H3 and PRC2 methyltransferase activity, as measured by the level of the PRC2 histone mark H3K27me3, is greatly diminished.[15] Accordingly, we hypothesized that RBBP4/7 are important proteins in the promotion of TNBC as they help dictate access to the genome and promote stemness through epigenetic modulation.

From the analysis of publically available microarray data, we found the expression of RBBP4/7 to be higher in breast cancer and specifically in TNBC. In order to further characterize the role of RBBP4 and RBBP7 in TNBC, protein expression was decreased using siRNA and knockdown cells were subjected to several cellular assays. Individual knockdown of either RBBP4 or RBBP7 resulted in the concomitant increase of the other, which suggests that for phenotypic changes to occur both proteins must be targeted. As they share 92% sequence identity, it is not surprising that these proteins also share functional overlap. When both RBBP4 and RBBP7 were depleted, cell proliferation as well as the mammosphere-forming capacity was significantly decreased. These results imply that RBBP4/7 may not only regulate CSCs, but also the bulk tumor population.

With RBBP4 and RBBP7 being central components of epigenetic complexes, we also chose to investigate the changes in histone modifications after RNAi-mediated knockdown. Most of the changes seen in dual knockdown cells were also seen in just RBBP7 depleted cells. This similarity could be due to two reasons: our

RBBP7 siRNA slightly decreased RBBP4 mRNA therefore affecting both proteins and/or the efficiency of RBBP7 knockdown was better than RBBP4 knockdown. The western blot of the global histone marks show both increases and decreases in methylation marks. This aligns with the current knowledge of epigenetic marks in the cell whereby they exist in a dynamic state of checks and balances, i.e. increases in one will cause decreases in the other. In our studies, levels of the activating histone marks H3K36me3 and H3K79me3 dramatically increased suggesting a shift from a more repressed transcriptional network to one that is more open and accessible to RNA polymerases. On the other hand, the repressive mark H3K9me3 was also notably increased in dual knockdown (siRBBP4 and siRBBP7) cells. As trimethylation of Lys9 of histone H3 is associated with closed chromatin and senescence, these results were consistent with the cell proliferation assay where decreased expression of RBBP4/7 significantly hindered TNBC growth.[24] The H3K36me3 and H3K79me3 marks have also recently been reported in ageing histones suggesting that without RBBP4 and RBBP7 the cell is unable to keep chromatin young. These changes in methylation marks may not be directly related to changes in enzymatic components or epigenetic complexes, but rather a natural function of ageing cells. RBBP4/7 knockdown, according to our RNAseq data, did not affect the mRNA levels of DOT1L, G9a, or SET-domain proteins that are responsible for the above marks. We did see a decrease in PHF19, a protein responsible for the demethylation of H3K36me3. PHF19 was shown to recruit PRC2, so disruption of the complex by RBBP4/7 depletion could also explain these observations.[73]

Interestingly, in contrast with other studies, the level of the PRC2-associated mark, H3K27me3, was decreased. While other studies have questioned the role of RBBP4/7 in PRC2 and only cite decreases in H3K27me3 levels at specific genomic locations upon knockdown of these proteins, our results indicate that together RBBP4/7 may be important in global maintenance of H3K27me3.[26,27] Furthermore, we witnessed an increase in H3K27me2 levels in RBBP4/7-silenced cells, which may signify that RBBP4/7 are critical for the conversion from dimethylation to trimethylation. When EZH2 activity is disrupted in breast cancer cells via small-molecules or RNAi, there is a decrease in H3K27me3 levels and a subsequent increase in H3K27 acetylation (H3K27Ac).[19] An unfortunate consequence of the increased H3K27Ac levels, as recently uncovered by Baude et al, is the amplification of Ras-driven transcription.[56] While knockdown of RBBP4/7 may share biological similarities with EZH2 silencing, we do not observe changes in H3K27Ac. This observation may distinguish targeting RBBP4/7 from canonical PRC2 inhibitors.

With the considerable changes to the epigenetic landscape upon RBBP4/7 knockdown, we investigated the genomic consequences in siRBBP4/7 treated cells versus negative control using RNA sequencing. Gene Set Enrichment Analysis (GSEA) confirmed a decrease in expression of mitotic genes as well as genes relating to stemness. In addition, embryonic stem cell genes normally marked with H3K27me3 were upregulated in siRBBP4/7 cells providing further verification of the dysregulation of PRC2 activity without RBBP4/7. In regard to specific PRC2 targets, we found that the expression of a WNT pathway inhibitor, DKK1, was increased upon RBBP4/7 knockdown.[53] Further evidence includes c-MYC, a well-

known oncoprotein and stem cell transcription factor which is upregulated by the WNT signaling cascade, was likewise decreased in siRBBP4/7 treated cells indicating this well-known CSC pathway was inhibited.[52,57,58] c-MYC also functions in a positive feedback loop with NOTCH1, a member of another prominent CSC pathway.[59] RBBP4/7 depletion also decreased NOTCH1 protein levels potentially through this feedback loop. Together RBBP4/7 could promote the CSC phenotype through cooperative pathways of WNT and NOTCH, potentially serving as a nexus for crosstalk between the two.

Functional assays validated the decrease in cancer stem cell properties amongst dual knockdown cells with both the mammosphere formation and the CD44+/CD24- CSC population reduced significantly. These results were translated *in vivo* as siRBBP4/7 tumors were much smaller than their control counterpart. Notably, the changes were lasting, as the CD44+/CD24- population remained much lower in siRBBP4/7 than control cells even after 16 weeks. Therefore, we hypothesized that inhibition of RBBP4 and RBBP7 could drive differentiation. In support of this, we found expression of basal cytokeratins downregulated with a resultant increase in luminal transcription factors, such as GATA3. Several studies have shown that exogenous expression of GATA3 in TNBC cell lines can initiate reprogramming with a decreased incidence of metastasis and tumorigenesis in xenograft models.[60,61] In addition, TNBC with higher expression of GATA3 also shower more favorable prognosis.[62] Together these data suggest that RBBP4/7 are intimately tied to maintenance of the CSC phenotype and may serve as a barrier for differentiation. However, since the effects of RBBP4/7 knockdown may not be

attributed to just one epigenetic complex, it will be necessary to perform chromatin immunoprecipitation (ChIP) assays in the future to further elucidate the interplay between the genetic and epigenetic changes we observed.

Our study is the first to demonstrate that these two proteins may serve as a potential therapeutic target for the treatment of TNBC and BCSCs. Discovering inhibitors of these proteins could interfere with CSC signaling and induce differentiation. Several studies have also demonstrated crosstalk amongst epigenetic pathways and redundant mechanisms that may prevent a single epigenetic therapy from being successful in independent administration. For example, combination therapies of HDAC and DNA methyltransferase (DNMT) inhibitors or inhibitors of EZH2 together with those against the H3K9me3 epigenetic modifier, G9a, are more effective in triggering a positive biological response.[63,64] With RBBP4 and RBBP7 associating with HDAC and histone methyltransferase (HMT) complexes, a single targeted agent against these proteins may act as an "epigenetic sink;" and thus, render once chemoresistant cancer cells susceptible to treatment. The inhibitor may even prevent the rewiring of normal cells into their carcinogenic counterparts through associated oncogenic proteins.

Future studies will be required to identify the critical binding pockets of RBBP4/7 that mediate its' oncogenic function and determine whether these sites are druggable. As RBBP4/7 are thought to be important for nucleosome association, RNAi rescue experiments designed with mutant RBBP4/7 constructs, that render them unable to bind to histone H3, may help delineate the significance of the RBBP4/7:histone H3 PPI in regulating BCSCs.

Methods

Cell culture and reagents. SUM149 was cultured in Ham's F-12 (Invitrogen) supplemented with 5% fetal bovine serum (FBS, Fisher Scientific), 5 μg/ml insulin (Sigma-Aldrich), 1 μg/ml hydrocortisone (Sigma-Aldrich), and 1% antibiotic-antimycotic (AA, Invitrogen). MDA-MB-436 and MDA-MB-231 cells were cultured in DMEM (Invitrogen) with 10% FBS and 1% AA. All cells were maintained in an environment of 5% CO_2. The following antibodies were used in western blotting: Anti-RbAp48 (RBBP4) rabbit polyclonal (ab47456, Abcam; WB 1:2000), Anti-RbAp46 (RBBP7) rabbit polyclonal (ab3535, Abcam; WB 1:2000), Anti-Histone H3 rabbit monoclonal (#4499, Cell Signaling; WB 1:2000), Anti-H3K4me3 rabbit polyclonal (#9727, Cell Signaling; WB 1:1000), Anti-H3K9me3 rabbit polyclonal (39162, Active Motif; WB 1:1000), Anti-H3K27me1 rabbit monoclonal (#7693, Cell Signaling; WB 1:1000), Anti-H3K27me2 rabbit monoclonal (#9728, Cell Signaling; WB 1:1000), Anti-H3K27me3 rabbit monoclonal (#9733, Cell Signaling; WB 1:1000), Anti-H3K27Ac rabbit polyclonal (#4353, Cell Signaling; WB 1:1000), Anti-H3K36me3 rabbit monoclonal (#4909, Cell Signaling; WB 1:1000), Anti-H3K79me3 rabbit polyclonal (#4260, Cell Signaling; WB 1:1000), Anti-cMyc rabbit polyclonal (#9402, Cell Signaling; WB 1:1000), Anti-MDM2 mouse monoclonal (sc-965, Santa Cruz; WB 1:1000), Anti-phospho-MDM2 rabbit polyclonal (#3521, Cell Signaling; WB 1:1000), Anti-p53 mouse monoclonal (#2524, Cell Signaling; WB 1:1000), Anti-DKK1 rabbit monoclonal (#48367, Cell Signaling; WB 1:1000), goat Anti-rabbit IgG, HRP-linked (#7074, Cell Signaling; WB 1:2000), Anti-β-Actin mouse monoclonal (sc-47778), goat Anti-mouse IgG, HRP-linked (sc-2005, Santa Cruz).

ONCOMINE data. Primary sources for the tumor data and for breast cancer cell lines from ONCOMINE were from Stickeler[65], Esserman[66], Chang[67], Desmede[68,69], Liu[9]

Knockdown by siRNA. siGENOME SMART pool small interfering RNA (siRNA) oligonucleotides targeting RBBP4 and RBBP7 were purchased from Dharmacon (GE Life Sciences). A nonsilencing siRNA, AllStars Negative Control siRNA (Qiagen), was utilized as a control. Transfection of cells was carried out using Lipofectamine® RNAiMAX (Thermo Fisher Scientific) according to the manufacturers instructions. Total concentration of siRNA was 50 nM per treatment group: Negative (50 nM siNeg), RBBP4 (25 nM siRBBP4 + 25 nM siNeg), RBBP7 (25 nM siRBBP7 + 25 nM siNeg), and RBBP4+RBBP7 (25 nM siRBBP4 + 25 nM siRBBP7). Cells were treated with siRNA for 72 hours and then assayed for knockdown efficiency via quantitative reverse transcription polymerase chain reaction (qRT-PCR) and western blot.

Lysate preparation and histone extraction. For whole cell lysate, cell pellets were lysed using RIPA buffer with protease inhibitors and phosphatase inhibitors added prior to use. After 30 minutes incubation on ice, lysate was collected via centrifugation. Samples were boiled for 5 min in 1X SDS-PAGE Laemmli loading buffer. To enrich for histones, the Histone Extraction kit (Abcam cat. no. ab113476) was used according to manufacturers instructions. Protein concentration was measured using either a Bradford protein assay (Bio-Rad) or the Pierce™ BCA protein assay (Thermo Scientific).

Western blot. All proteins were resolved on pre-cast gradient 4-15% Mini-PROTEAN® TGX™ (Bio-Rad) gels. Proteins were transferred onto nitrocellulose membranes (0.2 mm) using wet blotting techniques under constant voltage (30V for 70 min for histones or 100 V for 60 min for higher molecular weight proteins). After transfer, membranes were blocked for 1 hour at room temperature with 5% BSA in 0.1% TBST (50 mM Tris-HCl, pH 7.5, 150mM NaCl, 0.1% Tween 20). Primary antibodies were diluted according to manufacturer's recommendations and membranes were incubated with primary antibody overnight on a rotating platform at 4°C. The following morning, membranes were washed 3-4 times with 0.1% TBST for at least 5 min each time. The corresponding secondary antibody was added to the membranes for 1 hour at room temperature. After washing the membrane, enhanced chemiluminescent (ECL) substrate was added and western blot images were captured at various exposures using x-ray film.

CyQUANT proliferation assay. Cell lines were seeded at a density of 3,000 cells per well in 96-well plates and allowed to grow for 0, 24, 48 and 72 hours before being subjected to the CyQUANT (Invitrogen cat. no. C35012) assay. To 100 µL cell suspension was asdded 100 µL 2X detection reagent, which consisted of CyQUANT® Direct nucleic acid stain and CyQUANT® Direct background suppressor I diluted in cell culture medium. Cells were incubated with detection reagent for 60 minutes at 37°C. Fluorescence was read at 480/535 nm using a BioTek Cytation 5 plate reader.

Annexin V assay. siRBBP4, siRBBP7, siRBBP4+7, and siControl cells were harvested using trypsin and washed in cold phosphate-buffered saline (PBS). After

centrifugation and removal of the supernatant, cells were resuspended in annexin-binding buffer (10 mM HEPES, 140 mM NaCl, and 2.5 mM $CaCl_2$, pH 7.4) at a density of $1x10^6$ cells/mL. 5 µL each of Phycoerythrin (PE)-conjugated Annexin V and 7-AAD (BD Biosciences 559763) were added to each 100 µL cell suspension. The resulting mixture was incubated for 15 minutes at room temperature after which 400 µL annexin-binding buffer was added to each sample and tubes were placed on ice. Stained cells were analyzed by flow cytometry.

Propidium Iodide (PI) staining. For cell-cycle analysis, cells were subjected to propidium iodide staining. First, siRNA-treated cells were harvested using trypsin and washed in PBS. Cells were then fixed in 70% ethanol and incubated at 4°C for at least 30 min, or -20°C overnight. After fixation, cells were washed twice with PBS. Cells were resuspended in 200-400 µL of PI stain solution (PBS containing 500 µg/mL heat-inactivated RNase A and 50 µg/mL PI) and incubated at 37°C for 10 min. Stained cells were analyzed by flow cytometry.

Mammosphere formation assay. Treated cells were plated at a density of 3,000 cells per well, unless otherwise specified, in an ultra-low attachment 6-well plate (Corning) in serum-free (SF) medium. SF medium was prepared using MammoCult (STEMCELL technologies) with MammoCult proliferation supplement, hydrocortisone (final concentration 0.48 µg/mL), heparin (final concentration 4 µg/mL), and 1% antibiotic-antimycotic added before use. Mammospheres were cultured in suspension for 5-7 days and those greater than 50 µM were counted using the GelCount™ (Oxford Optronix). Individual mammospheres were visualized

on a Nikon Eclipse TE2000-S microscope and pictures were acquired with MetaMorph 7.6.0.0.

CD44+/CD24- flow cytometry analysis. CD44+/CD24- analysis of cell lines and resected xenograft tumors was performed using APC-conjugated mouse anti-human CD44 (BD Biosciences 559942) and PE-Cy7-conjugated mouse anti-human CD24 (BioLegend 311120) antibodies. PE-conjugated mouse anti-mouse H-2K[d] (BD Biosciences 553566) antibody was included when analyzing xenograft tumor cells to exclude mouse cells. Data acquisition was done by using a BD FACSCanto II flow cytometer with the FACSDiva software. FlowJo was used for data analysis.

Mouse Xenograft Models. *In vivo* studies were performed as described previously. Briefly, female non-obese diabetic/severe combined immunodeficient (NOD/SCID) mice, at 5 weeks old, were purchased from Jackson Laboratories. siRNA or shRNA treated cells (5,000), resuspended in matrigel (BD Biosciences), were injected into the exposed no. 4 inguinal mammary fat pad to generate xenograft models. Tumors were monitored weekly and mice were euthanized with CO_2 inhalation at the endpoint of the experiment or when they became moribund. Resected tumors were analyzed by flow cytometry.

RNA extraction and qRT-PCR. RNA was collected from samples using the RNeasy Mini kit (Qiagen). Isolated RNA was converted to cDNA (QuantiTect Reverse Transcription Kit, Qiagen). Real-time PCR was carried out on a QuantStudio 3 Real Time PCR System (Thermo Fisher Scientific). YWHAZ was used as the internal control.

RNA-seq and data analysis. RNA-seq was performed as previously described[70]. Briefly, total RNA was extracted using the RNeasy Mini kit and the mRNA libraries were prepared (Illumina TruSeq) and sequenced on Illumina Hi-Seq 2000 with 50 cycle single ended reads. RNA-seq reads were annotated using Bowtie. Gene expression is expressed as reads/kilobase/million mapped reads (RPKM) and differences in gene expression were estimated using rSeq. Gene Set Enrichment Analysis (Broad Institute) was used to correlate gene functions and signaling pathways that are significantly affected in RBBP4/7 knockdown cells.

Plasmids. GIPZ control (scrambled), RBBP4 (V2LHS_247612 and V2LHS_57090) and RBBP7 (V2LHS_171278) shRNA lentiviral constructs were obtained from University of Michigan Vector Core.

Lentiviral transduction. The packaging vectors psPAX2 and pMD2.G were co-transfected with GIPZ lentiviral constructs into 293T cells using the PEI method. Lentiviral supernatant was collected 48 hours after transfection and added to cells with 4 μg/mL polybrene (Sigma). Forty-eight hours after infection, cells were selected in puromycin (2 μg/mL) for 3 days. Cells were collected the next day for mouse xenograft implantation.

Statistical analyses. Two-tailed unpaired t-test (Student's t-test) was used for all statistical analyses except for RBBP4 and RBBP7 mRNA levels in breast cancer patients and cell lines in published microarray data, where one-tailed unpaired t-test was used (GraphPad Prism software version 5.0). A p-value of less than 0.05 was considered statistically significant.

Contributions. Rebecca Reed performed siRNA knockdown studies, western blots, helped with *in vivo* experiments and analyzed results. Miao-Chia Lo performed mouse xenografts and constructed shRNA and overexpression constructs. Miao-Chia Lo also completed GSEA analysis. Nicolas Stevers and Samantha Tinsley ran qPCR validation experiments. Albert Lin did flow cytometry experiments in RNAi-mediated knockdown cells.

References

1 Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nat Rev Cancer* **5**, 275-284, doi:10.1038/nrc1590 (2005).

2 Li, X. *et al.* Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J Natl Cancer Inst* **100**, 672-679, doi:10.1093/jnci/djn123 (2008).

3 Gangemi, R. *et al.* Cancer stem cells: a new paradigm for understanding tumor growth and progression and drug resistance. *Curr Med Chem* **16**, 1688-1703 (2009).

4 Charafe-Jauffret, E. *et al.* Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer Res* **69**, 1302-1313, doi:0008-5472.CAN-08-2741 [pii]10.1158/0008-5472.CAN-08-2741 (2009).

5 Charafe-Jauffret, E. *et al.* Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer research* **69**, 1302-1313, doi:10.1158/0008-5472.CAN-08-2741 (2009).

6 Gangemi, R. *et al.* Cancer stem cells: a new paradigm for understanding tumor growth and progression and drug resistance. *Current medicinal chemistry* **16**, 1688-1703 (2009).

7 Charafe-Jauffret, E. *et al.* Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**, 45-55, doi:10.1158/1078-0432.ccr-09-1630 (2009).

8 Hasan, K. *et al.* Activation of an IL6 inflammatory loop mediates trastuzumab resistance in HER2+ breast cancer by expanding the cancer stem cell population. *Molecular cell* **47**, 570-584, doi:10.1016/j.molcel.2012.06.014 (2012).

9 Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78-91, doi:10.1016/j.stemcr.2013.11.009 (2014).

10 Andrew, C. & Nicholas, C. C. Aberrations of EZH2 in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **17**, 2613-2618, doi:10.1158/1078-0432.ccr-10-2156 (2011).

11 Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised epigenetic states and acquired drug resistance in cancer. *Nat Rev Cancer* **14**, 747-753, doi:10.1038/nrc3819 (2014).

12 Xu, C. & Min, J. Structure and function of WD40 domain proteins. *Protein & cell* **2**, 202-214, doi:10.1007/s13238-011-1018-1 (2011).

13 Ahmed, Z. B., Paula, S., Richard, M. S. & Karen, S. RBBP4 Regulates Histone Deacetylation and Bipolar Spindle Assembly During Oocyte Maturation in the Mouse. *Biology of reproduction* **92**, 105, doi:10.1095/biolreprod.115.128298 (2015).

14 Kloet, S. L. *et al.* Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *The FEBS journal*, doi:10.1111/febs.12972 (2014).
15 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
16 Abou El Hassan, M. *et al.* Cancer Cells Hijack PRC2 to Modify Multiple Cytokine Pathways. *PloS one* **10**, doi:10.1371/journal.pone.0126466 (2015).
17 Mirzamohammadi, F. *et al.* Polycomb repressive complex 2 regulates skeletal growth by suppressing Wnt and TGF-beta signalling. *Nat Commun* **7**, 12047, doi:10.1038/ncomms12047 (2016).
18 Ying, Y. & Tao, Q. Epigenetic disruption of the WNT/beta-catenin signaling pathway in human cancers. *Epigenetics* **4**, 307-312 (2009).
19 Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100**, 11606-11611, doi:10.1073/pnas.1933744100 (2003).
20 Fu, J. *et al.* The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Res* **21**, 275-289, doi:10.1038/cr.2010.118 (2011).
21 Gabriel, G. M. *et al.* Architecture of epigenetic reprogramming following Twist1-mediated epithelial-mesenchymal transition. *Genome Biology* **14**, doi:10.1186/gb-2013-14-12-r144 (2013).
22 Icardi, L. *et al.* The Sin3a repressor complex is a master regulator of STAT transcriptional activity. *Proc Natl Acad Sci U S A* **109**, 12058-12063, doi:10.1073/pnas.1206458109 (2012).
23 Jiancong, L. *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nature cell biology* **10**, 731-739, doi:10.1038/ncb1736 (2008).
24 Pegoraro, G. *et al.* Ageing-related chromatin defects through loss of the NURD complex. *Nat Cell Biol* **11**, 1261-1267, doi:10.1038/ncb1971 (2009).
25 Byvoet, P., Shepherd, G. R., Hardin, J. M. & Noland, B. J. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch Biochem Biophys* **148**, 558-567 (1972).
26 Jamieson, K., Rountree, M. R., Lewis, Z. A., Stajich, J. E. & Selker, E. U. Regional control of histone H3 lysine 27 methylation in Neurospora. *Proc Natl Acad Sci U S A* **110**, 6027-6032, doi:10.1073/pnas.1303750110 (2013).
27 Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell* **15**, 57-67, doi:10.1016/j.molcel.2004.06.020 (2004).
28 Pasini, D. *et al.* Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* **38**, 4958-4969, doi:10.1093/nar/gkq244 (2010).
29 Tie, F. *et al.* CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* **136**, 3131-3141, doi:10.1242/dev.037127 (2009).

30 Reynolds, N. *et al.* NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J* **31**, 593-605, doi:10.1038/emboj.2011.431 (2012).
31 Yamada, T. *et al.* Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron* **83**, 122-134, doi:10.1016/j.neuron.2014.05.039 (2014).
32 Filippos, K. *et al.* NDY1/KDM2B functions as a master regulator of polycomb complexes and controls self-renewal of breast cancer stem cells. *Cancer research* **74**, 3935-3946, doi:10.1158/0008-5472.CAN-13-2733 (2014).
33 Frank, W. S. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Molecular Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
34 Ezponda, T. & Licht, J. D. Molecular pathways: deregulation of histone h3 lysine 27 methylation in cancer-different paths, same destination. *Clin Cancer Res* **20**, 5001-5008, doi:10.1158/1078-0432.CCR-13-2499 (2014).
35 Dorr, J. R. *et al.* Synthetic lethal metabolic targeting of cellular senescence in cancer therapy. *Nature* **501**, 421-425, doi:10.1038/nature12437 (2013).
36 Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-3988, doi:10.1073/pnas.0530291100 (2003).
37 Sheridan, C. *et al.* CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res* **8**, R59, doi:10.1186/bcr1610 (2006).
38 Fillmore, C. M. & Kuperwasser, C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast cancer research : BCR* **10**, doi:10.1186/bcr1982 (2007).
39 Suling, L. *et al.* Breast Cancer Stem Cells Transition between Epithelial and Mesenchymal States Reflective of their Normal Counterparts. *Stem Cell Reports* **2**, doi:10.1016/j.stemcr.2013.11.009 (2014).
40 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
41 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).
42 Smid, M. *et al.* Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68**, 3108-3114, doi:10.1158/0008-5472.CAN-07-5644 (2008).
43 Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499-507, doi:10.1038/ng.127 (2008).
44 Ren, G. *et al.* Polycomb protein EZH2 regulates tumor invasion via the transcriptional repression of the metastasis suppressor RKIP in breast and prostate cancer. *Cancer research* **72**, 3091-3104, doi:10.1158/0008-5472.CAN-11-3546 (2012).

45 Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**, 525-530, doi:10.1101/gad.415507 (2007).
46 Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**, 1977-2000, doi:10.1091/mbc.02-02-0030. (2002).
47 Reichert, N. *et al.* Lin9, a subunit of the mammalian DREAM complex, is essential for embryonic development, for survival of adult mice, and for tumor suppression. *Mol Cell Biol* **30**, 2896-2908, doi:10.1128/MCB.00028-10 (2010).
48 Perez, C. A., Ott, J., Mays, D. J. & Pietenpol, J. A. p63 consensus DNA-binding site: identification, analysis and application into a p63MH algorithm. *Oncogene* **26**, 7363-7370, doi:10.1038/sj.onc.1210561 (2007).
49 Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* **37**, D674-679, doi:10.1093/nar/gkn653 (2009).
50 Wienken, M. *et al.* MDM2 Associates with Polycomb Repressor Complex 2 and Enhances Stemness-Promoting Chromatin Modifications Independent of p53. *Mol Cell* **61**, 68-83, doi:10.1016/j.molcel.2015.12.008 (2016).
51 Walid, T. K. *et al.* BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nature Communications* **6**, doi:10.1038/ncomms6987 (2015).
52 Cowling, V. H., D'Cruz, C. M., Chodosh, L. A. & Cole, M. D. c-Myc transforms human mammary epithelial cells through repression of the Wnt inhibitors DKK1 and SFRP1. *Mol Cell Biol* **27**, 5135-5146, doi:10.1128/MCB.02282-06 (2007).
53 Cheng, A. S. *et al.* EZH2-mediated concordant repression of Wnt antagonists promotes beta-catenin-dependent hepatocarcinogenesis. *Cancer Res* **71**, 4028-4039, doi:10.1158/0008-5472.CAN-10-3342 (2011).
54 Weigelt, B. *et al.* Refinement of breast cancer classification by molecular characterization of histological special types. *The Journal of pathology* **216**, 141-150, doi:10.1002/path.2407 (2008).
55 Abd El-Rehim, D. M. *et al.* Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* **203**, 661-671, doi:10.1002/path.1559 (2004).
56 Baude, A., Lindroth, A. M. & Plass, C. PRC2 loss amplifies Ras signaling in cancer. *Nat Genet* **46**, 1154-1155, doi:10.1038/ng.3124 (2014).
57 He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509-1512 (1998).
58 Ziegler, S. *et al.* Novel target genes of the Wnt pathway and statistical insights into Wnt target promoter regulation. *FEBS J* **272**, 1600-1615, doi:10.1111/j.1742-4658.2005.04581.x (2005).
59 Palomero, T. *et al.* NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc Natl Acad Sci U S A* **103**, 18261-18266, doi:10.1073/pnas.0606108103 (2006).

60 Dydensborg, A. B. *et al.* GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. *Oncogene* **28**, 2634-2642, doi:10.1038/onc.2009.126 (2009).
61 Yan, W., Cao, Q. J., Arenas, R. B., Bentley, B. & Shao, R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J Biol Chem* **285**, 14042-14051, doi:10.1074/jbc.M110.105262 (2010).
62 Yu, K. D. *et al.* Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin Cancer Res* **19**, 2723-2733, doi:10.1158/1078-0432.CCR-12-2986 (2013).
63 Curry, E. *et al.* Dual EZH2 and EHMT2 histone methyltransferase inhibition increases biological efficacy in breast cancer cells. *Clin Epigenetics* **7**, 84, doi:10.1186/s13148-015-0118-9 (2015).
64 Pathania, R. *et al.* Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth. *Cancer Res* **76**, 3224-3235, doi:10.1158/0008-5472.CAN-15-2249 (2016).
65 Stickeler, E. *et al.* Basal-like molecular subtype and HER4 up-regulation and response to neoadjuvant chemotherapy in breast cancer. *Oncol Rep* **26**, 1037-1045, doi:10.3892/or.2011.1392 (2011).
66 Esserman, L. J. *et al.* Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res Treat* **132**, 1049-1062, doi:10.1007/s10549-011-1895-2 (2012).
67 Chang, J. C. *et al.* Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol* **23**, 1169-1177, doi:10.1200/JCO.2005.03.156 (2005).
68 Desmedt, C. *et al.* Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* **13**, 3207-3214, doi:10.1158/1078-0432.CCR-06-2765 (2007).
69 Patil, P., Bachant-Winner, P. O., Haibe-Kains, B. & Leek, J. T. Test set bias affects reproducibility of gene signatures. *Bioinformatics* **31**, 2318-2323, doi:10.1093/bioinformatics/btv157 (2015).
70 Sun, L. *et al.* Novel cancer stem cell targets during epithelial to mesenchymal transition in PTEN-deficient trastuzumab-resistant breast cancer. *Oncotarget*, doi:10.18632/oncotarget.9839 (2016).
71 Venkatesh, S. *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* **489**, 452-455, doi:10.1038/nature11326 (2012).
72 De Vos, D. *et al.* Progressive methylation of ageing histones by Dot1 functions as a timer. *EMBO Rep* **12**, 956-962, doi:10.1038/embor.2011.131 (2011).
73 Brien, G. L. *et al.* Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* **19**, 1273-1281, doi:10.1038/nsmb.2449 (2012).

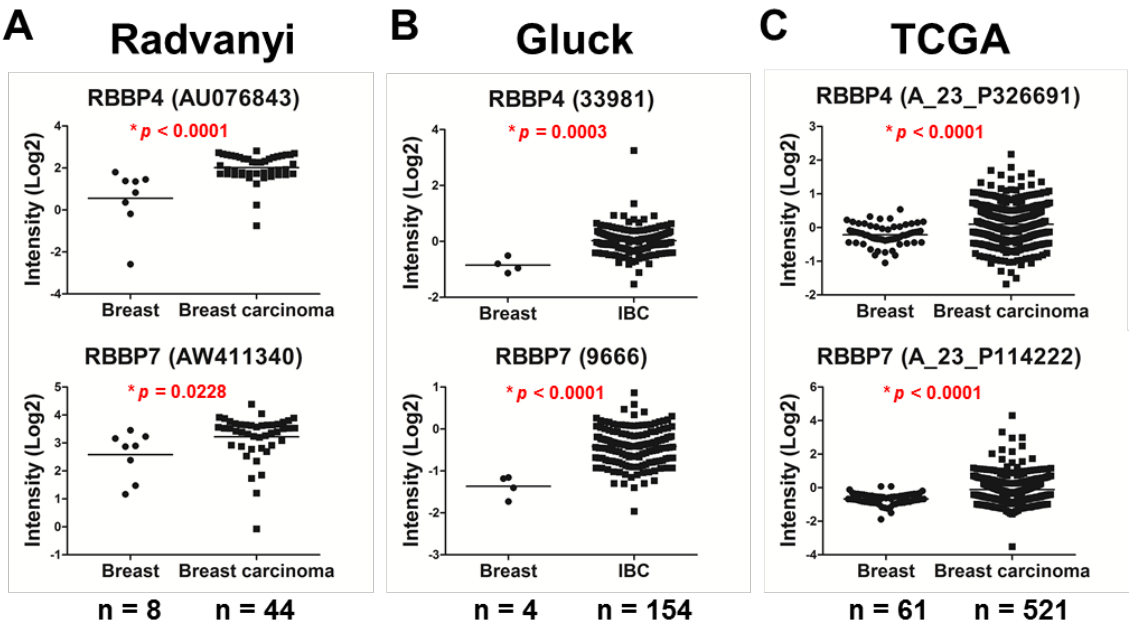


Figure 2.1 The mRNA levels of RBBP4 and RBBP7 in breast cancer versus normal breast tissues.

Dot plots showing the mRNA levels of RBBP4 and RBBP7 in breast cancer versus normal breast tissues in Radvanyi (A), Gluck (B) and TCGA (C) microarray studies. The y-axis represents the log2 median-centered ratios of the reporters indicated in the parentheses, which were obtained from Oncomine (www.oncomine.org). The horizontal bars are the mean of each patient group. IBC: invasive breast carcinoma. ns: not significant. *p < 0.05

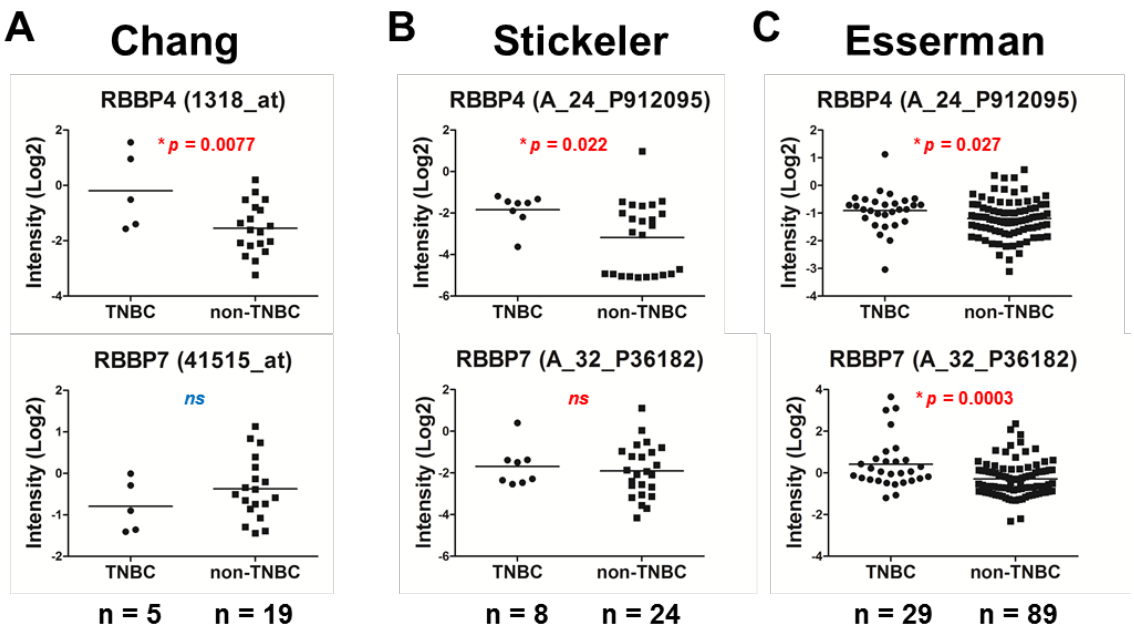


Figure 2.2 The mRNA levels of RBBP4 and RBBP7 in TNBC versus non-TNBC. Dot plots showing the mRNA levels of RBBP4 and RBBP7 in TNBC vs. non-TNBC patients in Chang (A), Stickeler (B), and Esserman (C) microarray studies. The y-axis represents the log2 median-centered ratios of the reporters indicated in the parentheses, which were obtained from Oncomine (www.oncomine.org). The horizontal bars are the mean of each patient group. Red and blue indicate upregulation and downregulation, respectively, in TNBC. ns: not significant. *p < 0.05

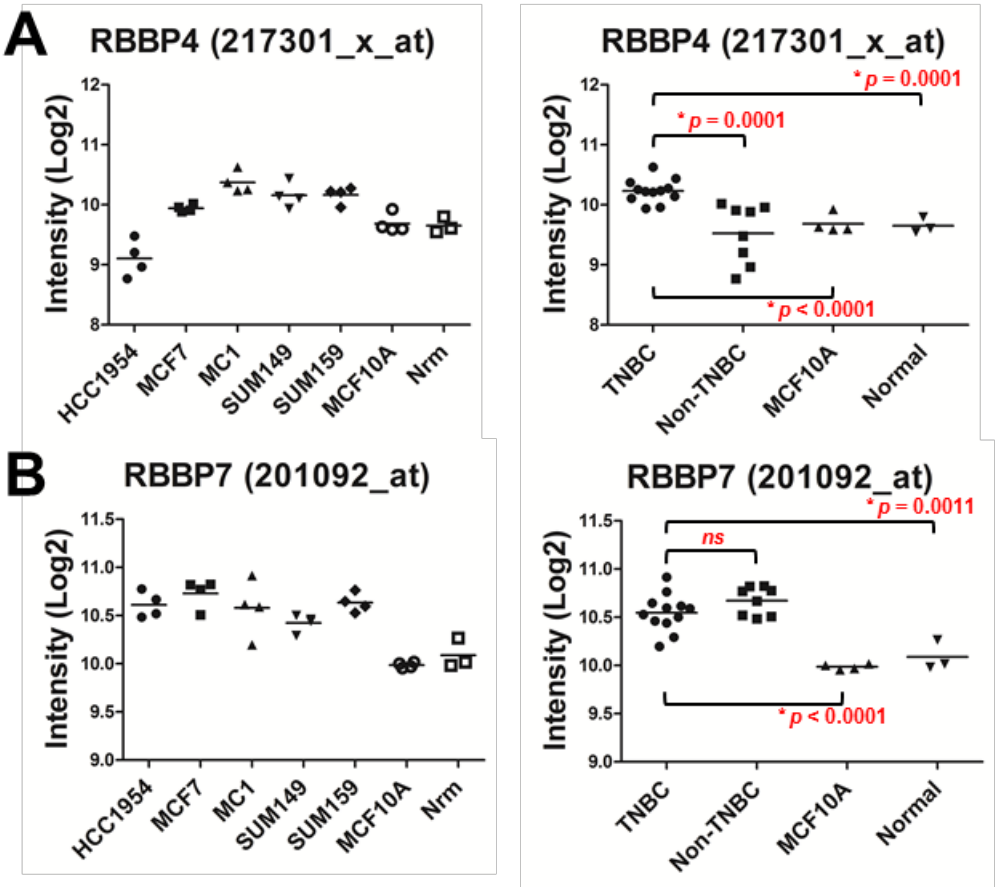


Figure 2.3 The mRNA levels of RBBP4 and RBBP7 in TNBC versus non-TNBC cell lines.

Dot plots showing the mRNA levels of RBBP4 (A) and RBBP7 (B) in two non-TNBC (HCC1954, MCF7) and three TNBC (MC1, SUM149, SUM159) cell lines, MCF10A and normal breast cells from Liu et al. microarray study obtained from GEO database (GSE52262). The y-axis represents the log2 signal intensity of the reporters indicated in the parentheses. The horizontal bars are the mean of each cell line or group. ns: not significant. *p < 0.05

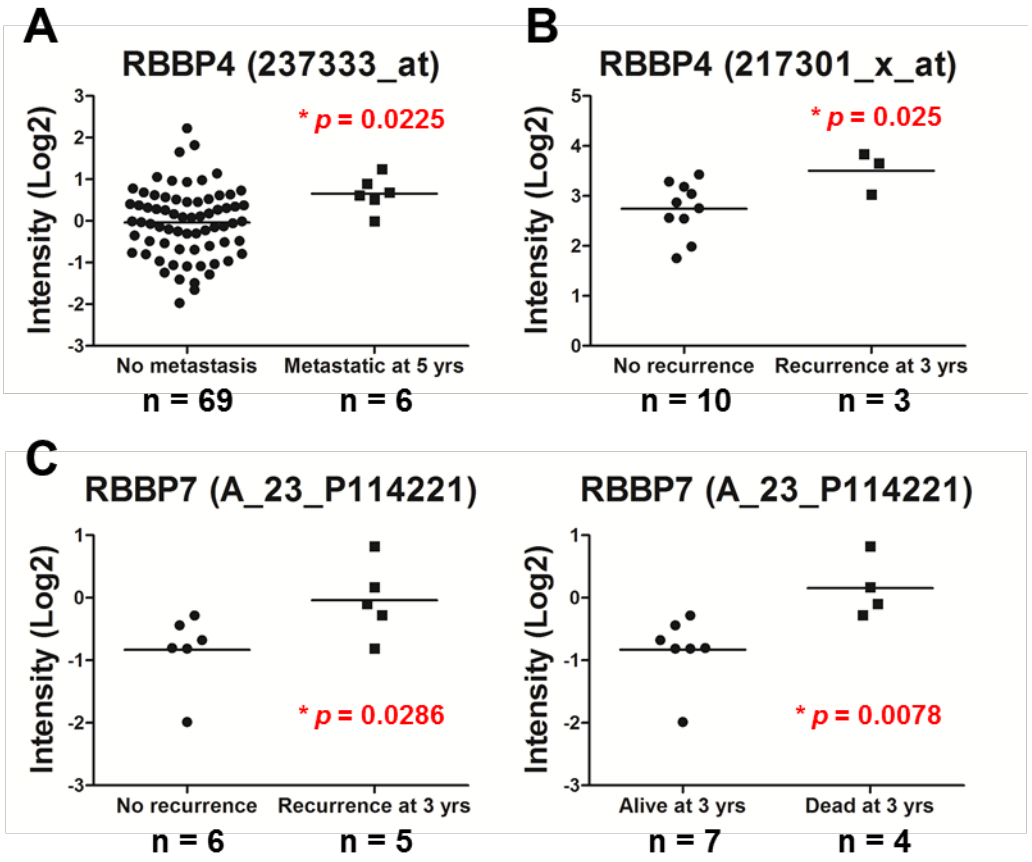


Figure 2.4 The mRNA levels of RBBP4 and RBBP7 in breast cancer patients stratified based on metastasis or recurrence status.
(A) RBBP4 mRNA levels in breast carcinoma patients from Loi microarray study stratified based on metastasis status at 5 years. (B) RBBP4 mRNA levels in invasive lobular breast carcinoma patients from Desmedt microarray study stratified based on recurrence status at 3 years. (C) RBBP7 mRNA levels in invasive breast carcinoma patients from Esserman microarray study stratified based on recurrence (left) or survival (right) status at 3 years. The y-axis represents the log2 median-centered ratios [obtained from Oncomine (www.oncomine.org)] of the reporters indicated in the parentheses. The horizontal bars are the mean of each patient group. *p < 0.05

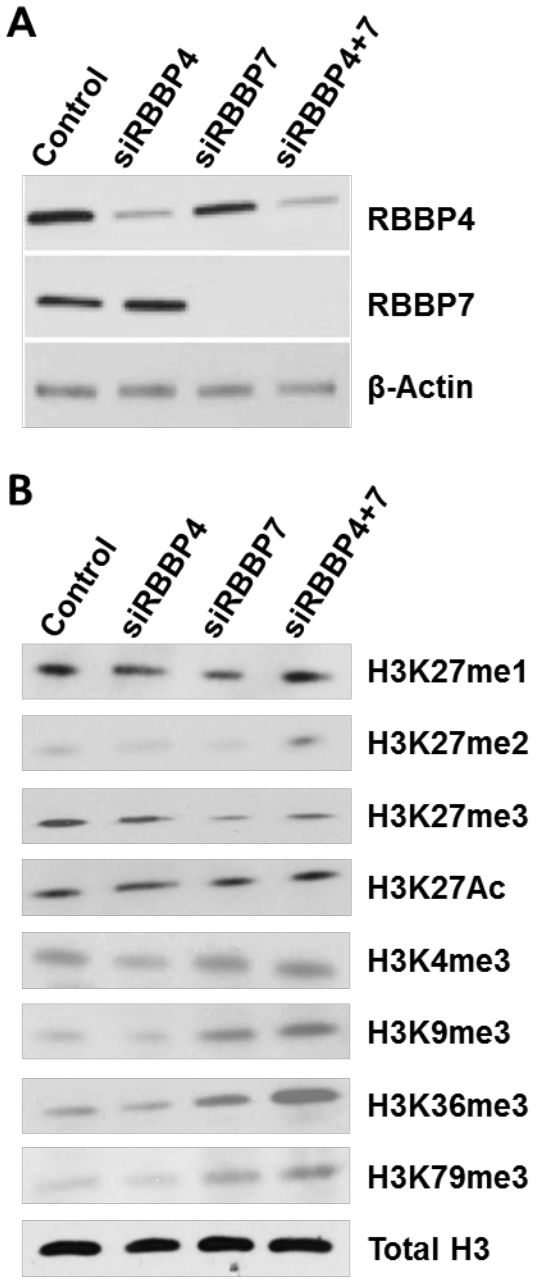


Figure 2.5 RBBP4 and RBBP7 knockdown affects global histone modification. SUM149 cells were transfected with the indicated siRNA. After 72 hr, RBBP4 and RBBP7 protein levels (A) and various histone modifications (B) were examined by western blot.

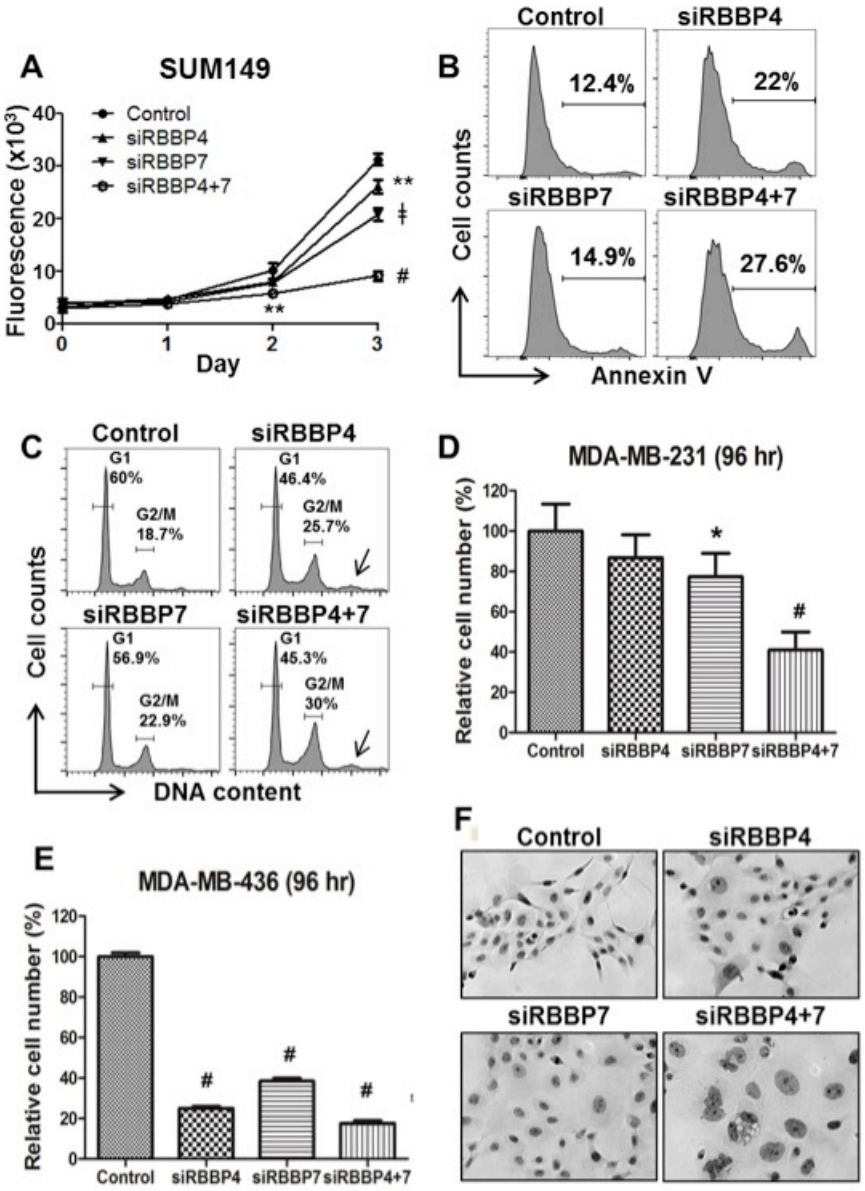


Figure 2.6 RBBP4 and RBBP7 knockdown inhibits growth of TNBC cell lines. SUM149 cells were transfected with the indicated siRNA. After 72 hr, cells were plated for growth curve using the CyQUANT assay (A). RBBP4 or RBBP7 knockdown significantly decreased cell numbers on day 3 when compared with control siRNA treated cells. Knockdown of both RBBP4 and RBBP7 showed a better growth inhibition ($p < 0.01$ on day 2) than knockdown of either RBBP protein. Each time point represents the average with s.d. of triplicate samples. Representative results from at least three independent experiments are shown. (B) Apoptosis assay by flow cytometry at 72 hr after siRNA transfection. Knockdown of RBBP4 and RBBP7 showed an increase in apoptosis. (C) Cell cycle analysis by flow cytometry at 72 hr after siRNA transfection. Cells were fixed and stained with propidium iodide. The

percentage of cells with 2N (G1) or 4N (G2/M) DNA content are indicated. RBBP4 and RBBP7 knockdown caused an increase in G2/M phase at the expense of G1 phase. The arrows indicate the increase of multiploid cells upon siRBBP4 and siRBBP4+7 knockdowns. Representative results from two independent experiments are shown in (B) and (C). (D) and (E) MDA-MB-231 and MDA-MB-436 cells were treated with the indicated siRNA for 72 hr. Equal number of cells were then plated and the relative cell numbers were determined by CyQUANT after 96 hr. The averages with s.d. of at least four samples are shown. (F) Hematoxylin staining of SUM149 cells 10 days after siRNA transfection. Images were taken using Nikon TE 2000S microscope with the 10× objective. *$p < 0.05$, **$p < 0.01$, ǂ$p < 0.001$, #$p < 0.0001$

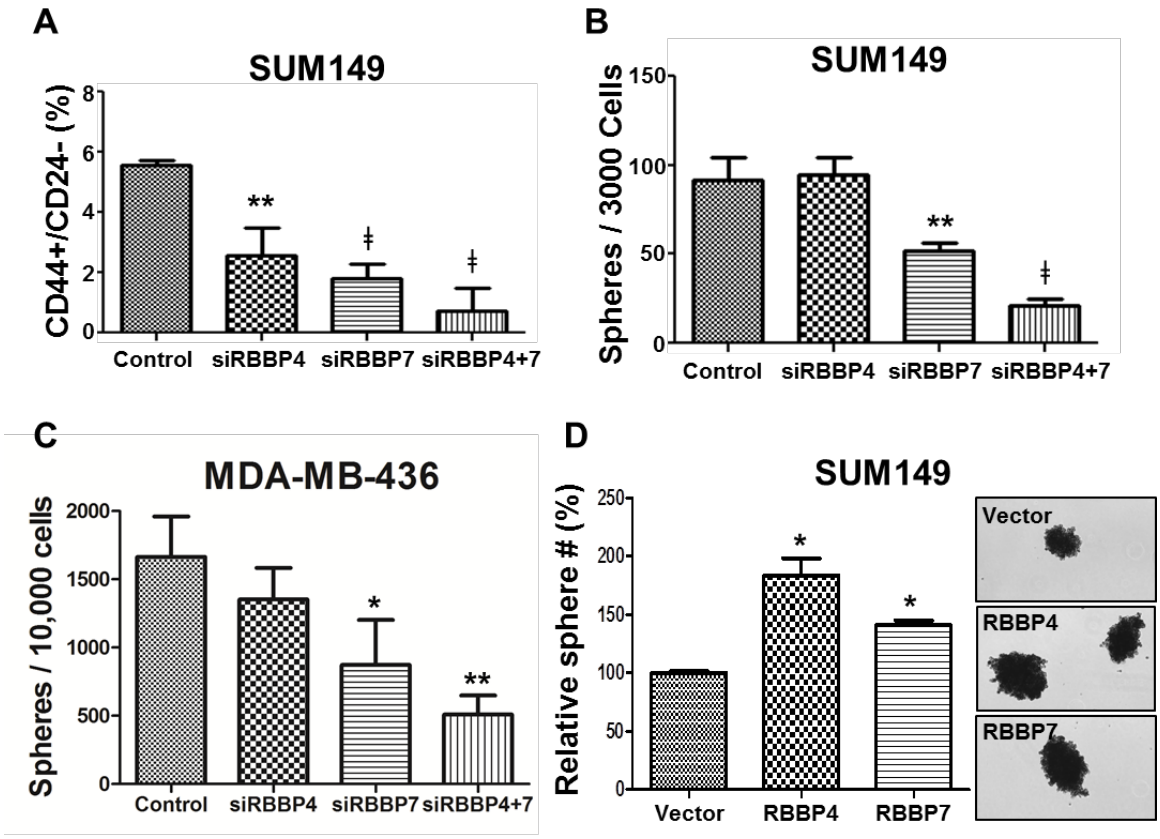


Figure 2.7 RBBP4 and RBBP7 knockdown decreases the cancer stem cell population in TNBC cells.

(A) Analysis of the CD44+/CD24- CSC population by flow cytometry. (B) After 72 hours, knockdown cells were trypsinized, collected and resuspended in serum-free MammoCult medium at a density of 3,000 cells/mL. 1 mL of cell suspension was added per well to an ultra low attachment plate and mammospheres were cultured for 7 days before counting. (C) Sphere counts of siRNA treated MDA-MB-436 cells. (D) Relative mammosphere counts of SUM149 cells exogenously expressing RBBP4, RBBP7 and the control vector, which is presented as 100% (left). The representative spheres are shown (right). *$p < 0.05$, **$p < 0.01$, ǂ$p < 0.001$

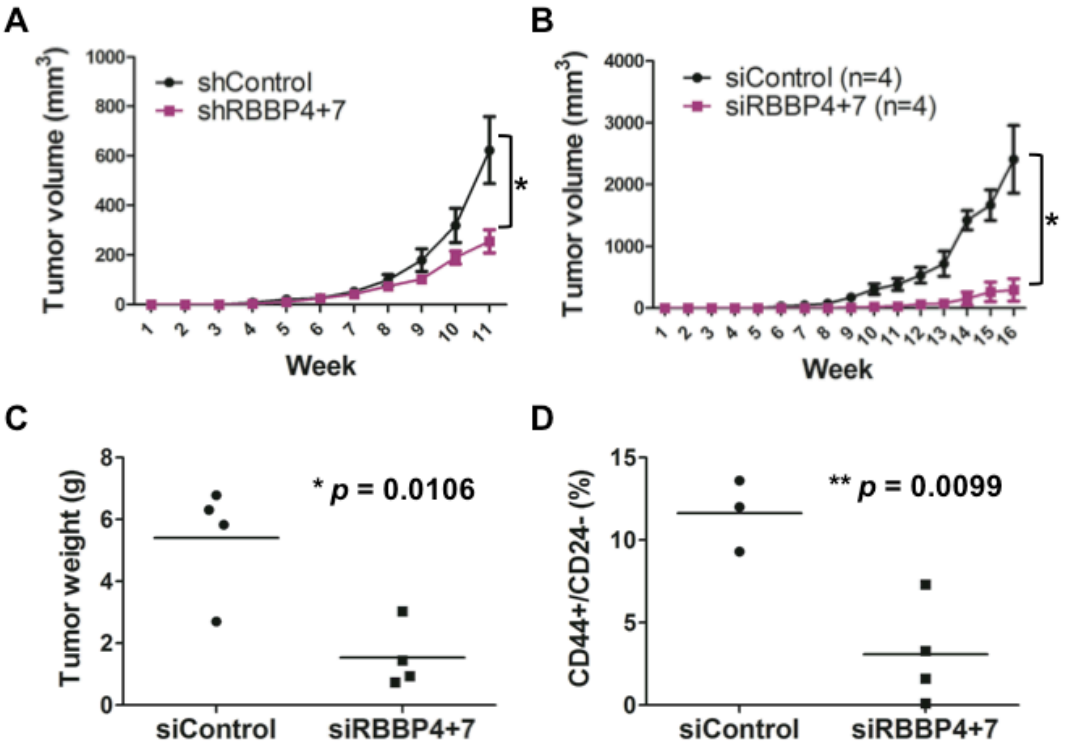


Figure 2.8 Knockdown of both RBBP4 and RBBP7 inhibits tumor growth of SUM149 cells in vivo.
(A) 5x10^3 shControl- or shRBBP4+7- and (B) 5x10^3 siControl- or siRBBP4+7-transduced SUM149 cells were orthotopically injected into the fourth mammary fat pad of 6 to 8-week-old female NOD-SCID mice. Tumor volume was measured weekly. (C) Weight of the resected tumors from mice in (B) at the endpoint of the study. (D) Cells isolated from siControl and siRBBP4+7 tumors were assessed for CD44+/CD24- percentages by flow cytometry. Horizontal bars indicate the mean of each treatment group. *p < 0.05, **p < 0.01

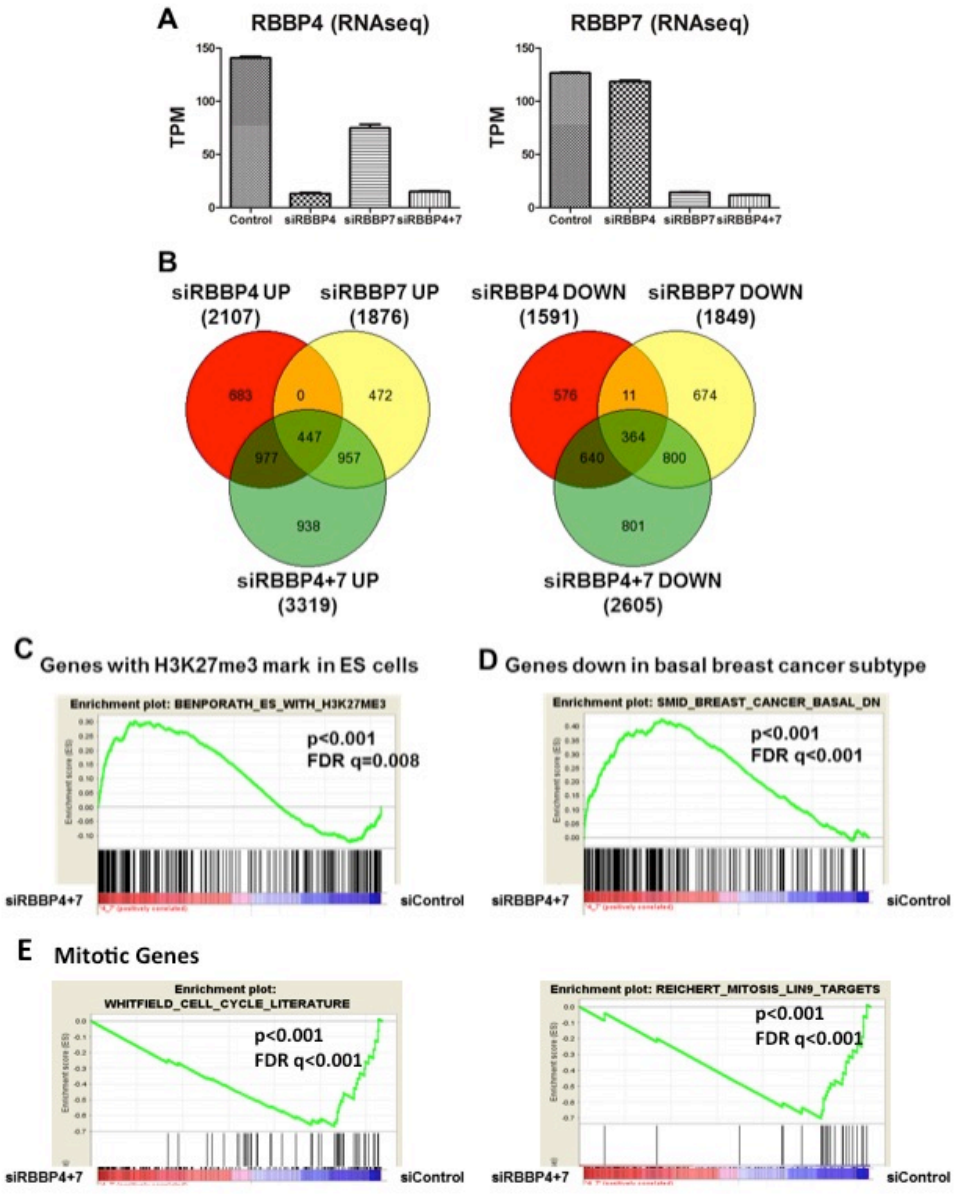


Figure 2.9 Knockdown of both RBBP4 and RBBP7 in SUM149 cells upregulates genes with H3K27me3 marks and results in reversion of basal subtype breast cancer gene signature.

72 hours after siRNA transfection, RNA-seq analysis was performed. The RNA-seq data was then subjected to GSEA for biological interpretations. (A) RNA-seq data showing knockdown of RBBP4 and/or RBBP7 mRNA in SUM149 cells. TPM: transcripts per million. (B) Venn diagrams showing number of genes up (left) and down (right) by at least 1.2 folds in SUM149 cells treated with siRBBP4, siRBBP7 or both. (C) and (D) GSEA showing genes with H3K27me3 marks in embryonic stem cells and genes under-expressed in basal subtype breast cancer, respectively, were significantly enriched in siRBBP4+7 cells compared to siControl cells. (E) GSEA showing expression of mitotic genes was negatively correlated with RBBP4/7 knockdown.

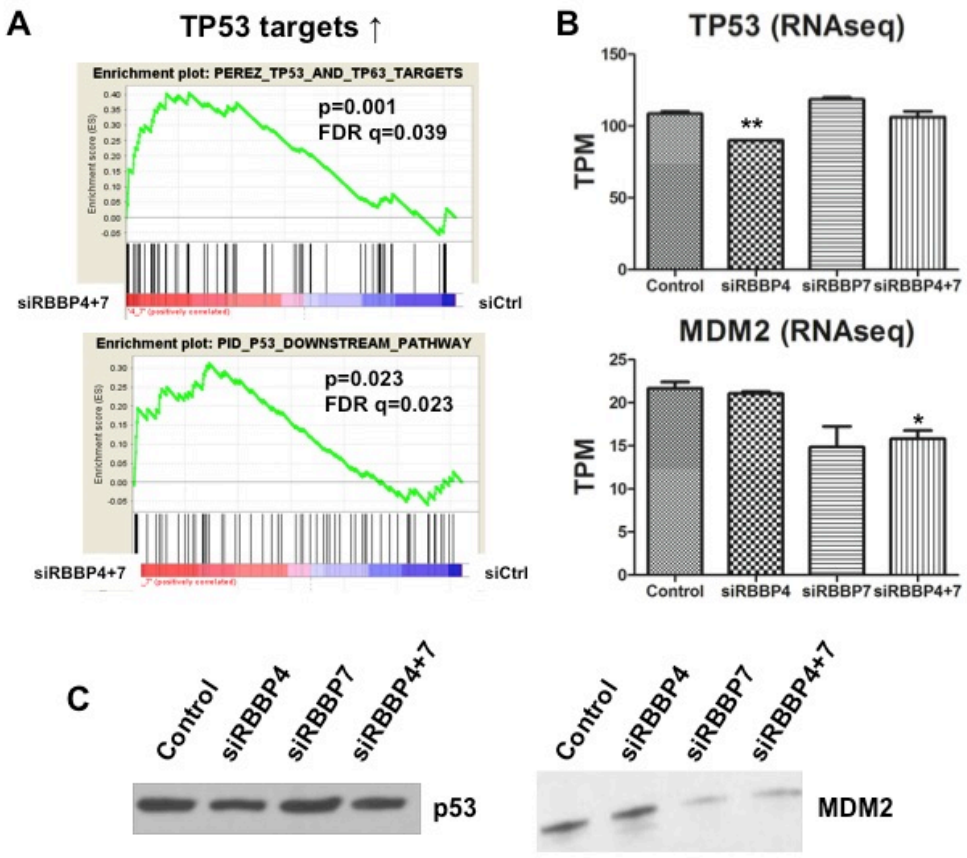


Figure 2.10 Knockdown of both RBBP4 and RBBP7 in SUM149 cells upregulates p53 target genes.
(A) GSEA analysis of siRBBP4+7 samples in comparison to control (siCtrl) indicating upregulation of p53 targets. (B) RNA-seq demonstrating no increase in p53 mRNA levels, but a decrease in the p53 antagonist, MDM2, potentially contributing to the upregulation of p53 targets. (C) Immunoblots of p53 and MDM2 in siRNA treated SUM149 cells reiterating the decrease in MDM2, but little to no change in p53 levels. *$p < 0.05$, **$p < 0.01$

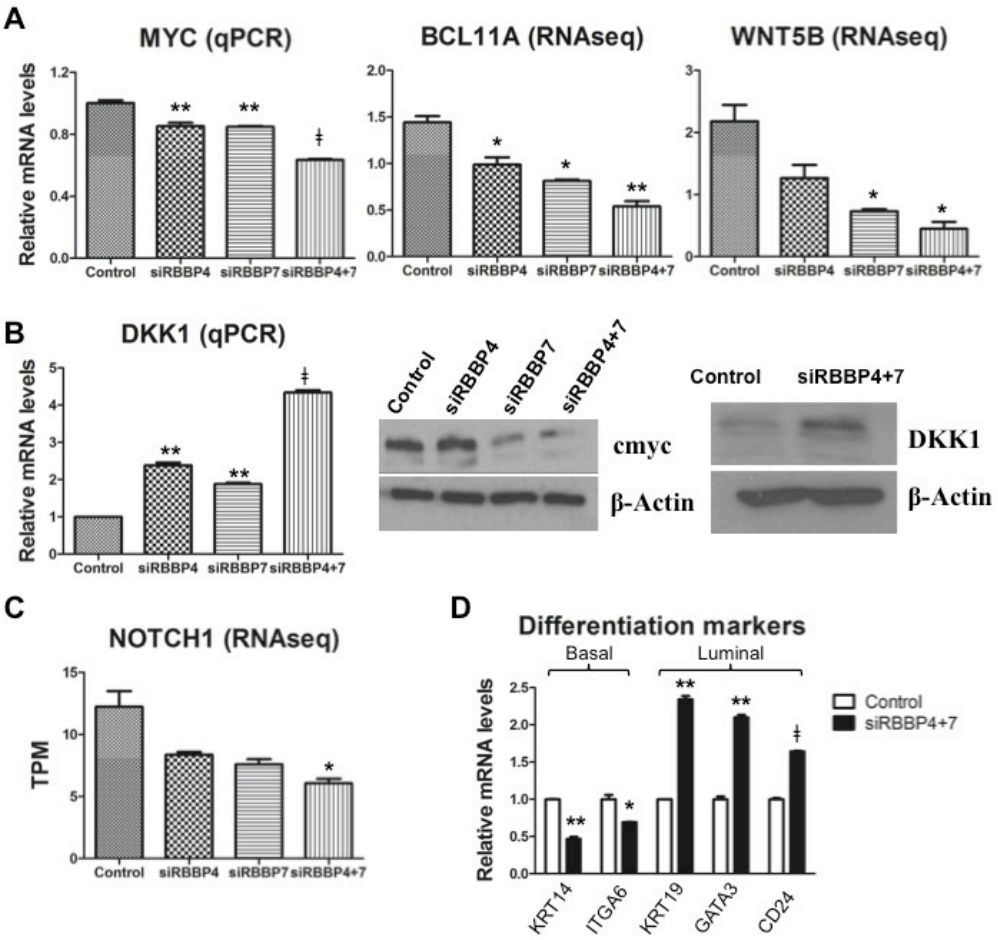


Figure 2.11 Knockdown of both RBBP4 and RBBP7 in SUM149 cells downregulates multiple cancer stem cell pathways and pushes basal cell differentiation.

(A) mRNA analysis of common stem cell proteins including oncogenic transcription factors c-MYC and BCL11A and WNT signaling protein WNT5B in siRNA treated SUM149 cells. (B) (left panel) RT-qPCR showing DKK1, an inhibitor of WNT signaling, which is usually repressed by c-MYC, is upregulated in siRBBP4+7 treated cells. (middle and right panels) Immunoblots of DKK1 and c-MYC suggest upregulation of DKK1 could be the result of decreased c-MYC expression. (C) Another critical CSC protein, NOTCH1, is decreased in dual knockdown cells. (D) RNA-seq data showing the decreased expression of basal markers KRT14 and ITGA6 (CD49f) and the increased expression of luminal markers KRT19, GATA3 and CD24 following the knockdown of RBBP4 and RBBP7. *p < 0.05, **p < 0.01, ǂp < 0.001

Table 2.1 Percentages of cell cycle stages in SUM149 cells treated with siRNA.

Treatment	G1 (%)	S (%)	G2/M (%)	Multi-ploid (%)
Control siRNA	60	13.3	18.7	7.93
RBBP4 siRNA	46.4	15.2	25.7	12.7
RBBP7 siRNA	56.9	10.7	22.9	9.45
RBBP4+RBBP7 siRNA	45.3	13	30	11.7

Table 2.2 Extreme limiting dilution analysis for SUM149-Control and RBBP4+7 shRNA mouse xenograft.

Group	Limiting dilutions Tumors/Implantations				CSC frequency (1 in/..)	*P*
	50000	5000	500	50		
Control shRNA	5/5	5/5	3/4	0/5	455	0.0877
RBBP4+7 shRNA	4/4	5/5	0/4	0/6	2022	

Chapter 3

Understanding the interaction between RBBP4/7 and the oncogenic transcription factor BCL11A

Abstract

Currently triple negative breast cancer (TNBC) is the only subtype of breast cancer for which there is no targeted therapy. Accumulating studies have focused on rectifying this unmet need by concentrating on the molecular foundation of this disease. B-cell lymphoma/leukemia 11A (BCL11A), a hematopoietic transcription factor, was recently discovered to be a driving force in TNBC and contributes to the maintenance of the chemoresistant cancer stem cell (CSC) population. BCL11A was found to interact with several co-repressor complexes to impart transcriptional control over select genomic loci; however, how this protein interacts with these complexes remains unknown. In this study, we connect BCL11A with histone methyltransferases (PRC2) and histone deacetylases (NuRD and SIN3A) through its interaction with the common core subunit, RBBP4. We reveal the crystal structure of RBBP4 in complex with an amino terminal-derived peptide (residues 2-16) of BCL11A. The BCL11A peptide binds to the top face of RBBP4's β-propeller; an interaction surface shared by histone H3 as well as other transcription factors

FOG-1 and SALL4. The structure confirms the importance of C-terminal residues of the BCL11A peptide as they anchor the peptide to the side of RBBP4. We further show, through pulldown studies, that this BCL11A peptide is able to associate with PRC2, SIN3A, and NuRD. However, the epigenetic complex CoREST, which is devoid of RBBP4, is not able to associate with the N-terminal peptide derived from BCL11A. Lastly, we transfected the TNBC cell line, SUM149, with the BCL11A peptide and witnessed a 50% decrease in ALDH+ activity, a marker of CSC. Together our studies suggest that the BCL11A-RBBP4 interface could be a novel drug target for TNBC. A small molecule against this protein-protein interaction could prevent oncogenic BCL11A from redirecting epigenetic complexes, thus inhibiting the rewiring of the epigenetic landscape and CSC-mediated tumorigenesis.

3.1 Introduction.

B-cell lymphoma/leukemia 11A (BCL11A) is a zinc-finger transcription factor that is necessary for hematopoiesis and regulates developmental globin switching as well as silencing[1,2]. Like its name suggests, BCL11A is also viewed as a proto-oncogene important in the development of leukemia[3]. In lymphoid malignancies BCL11A expression is often deregulated through chromosomal translocations or amplifications. It is closely tied with important signaling pathways that control cell growth and apoptosis[4-7]. In particular, it negatively regulates tumor suppressor p21 in addition to p53 via upregulation of MDM2[6,7].

Similar to many other hematopoietic transcription factors, like GATA3 and MYB, BCL11A was recently found to have a profound role in breast cancer, specifically in triple negative breast cancer (TNBC)[8-11]. While high expression of this

protein is usually restricted to fetal brain and germinal center B-cells, Khaled and colleagues demonstrated that BCL11A is upregulated in TNBC through copy number gains and hypomethylation of the BCL11A locus[11]. Furthermore, exogenous overexpression of BCL11A increased clonogenicity of non-tumorigenic mammary epithelial cells[11]. Knockdown of BCL11A, on the other hand, resulted in reduced tumor size and frequency of tumor formation indicating a decrease in tumor-initiating cells[11]. These data suggest BCL11A is an ideal target for the treatment of TNBC. Understanding how BCL11A operates within the cell, and specifically cancer, is paramount to the development of new and more effective therapies. This is especially true of TNBC for which there are currently no targeted treatments.

Co-immunoprecipitation studies suggest that BCL11A interacts with several histone modifying and chromatin remodeling complexes, but how this transcription factor interacts and recruits these epigenetic complexes remains to be determined[1]. BCL11A has been shown to associate with the nucleosome remodeling and deacetylase complex (NuRD)[12]. NuRD consists of six core components and uniquely ties together two enzymatic functions through the chromodomain-helicase-DNA binding protein 3/4 (CDH 3/4) and histone deacetylase 1/2 (HDAC 1/2). A non-enzymatic core component of NuRD is the histone chaperones retinoblastoma binding proteins 4 and 7 (RBBP4/7). RBBP4/7, which are known to bind to histone H3 and H4, are thought to promote nucleosomal association of their respective epigenetic complexes. Recent evidence indicates many transcription factors including FOG1, SALL4, and PHF6 connect with NuRD through its RBBP4/7 subunit.

Therefore, it is highly likely that BCL11A recruits NuRD through a similar mechanism, yet the biological consequences are unknown.

Several studies have shown that NuRD interacts with oncogenes in turn promoting tumorigenesis through repression of downstream targets, such as the tumor suppressor PTEN or E-cadherin. However, with the promiscuity of HDAC inhibitors there is a greater effort to find more specific inhibitors through targeting protein-protein interactions within the complex. To better characterize the interaction between BCL11A and NuRD we combined a panel of cell-based assays with crystallization studies. Using a fluorescence polarization (FP) assay we are able to show that a 12 amino acid peptide derived from the N-terminus of BCL11A is sufficient to bind RBBP4/7. It contains a conserved binding motif shared by FOG-1, SALL4, and others. The peptides derived from BCL11A compete with Histone H3 for binding to the top face of RBBP4's β-propeller. Crystal structures of BCL11A (aa 2-16) in complex with RBBP4 depict interactions that extend beyond the top face of RBBP4 and expose new and potentially druggable pockets on the histone chaperone. We also confirmed that BCL11A (aa 2-16) is able to pull down RBBP4 and RBBP7 as well as the NuRD components HDAC1 and MTA1 from cell lysate of the TNBC cell line SUM149. Furthermore, using an 11-amino acid peptide from BCL11A (aa 2-12), we treated SUM149 cells and found a significant decrease in the aldehyde dehydrogenase (ALDH)-positive tumor-initiating cell population. Together our results suggest that the RBBP4-BCL11A interaction is a potential therapeutic target for the treatment of TNBC.

3.2 Results

3.2.1 BCL11A competes with Histone H3 for binding to RBBP4.

Previous mass spectrometry and pulldown studies have indicated that BCL11A interacts with the NuRD epigenetic complex as well as the Polycomb Repressive Complex 2 (PRC2) and SIN3 deacetylase complex (SIN3A)[1]. RBBP4/7 are shared amongst all three complexes and therefore, we hypothesized that BCL11A may recruit the aforementioned epigenetic machineries through their common core subunit, RBBP4/7. In fact, BCL11A shares a similar N-terminal sequence as the transcription factors FOG-1 and SALL4 which have been shown to interact with NuRD and specifically RBBP4 (Figure 3.1A)[13,14]. It is highly likely this shared sequence motif is a conserved mechanism by which these transcription factors enlist epigenetic complexes to control transcription.

In order to test our hypothesis we developed a fluorescence polarization (FP) assay with histone H3, a known RBBP4/7 binding partner, as the tracer. Histone H3 (aa 1-21) was labeled with 5-FAM at the C-terminus and was incubated at 20 nM with increasing concentrations of full-length RBBP4. The binding affinity of the H3 tracer for RBBP4 was determined to be 0.84 ± 0.08 μM, stable for over 24 hours and consistent with published literature (Figure 3.1B and C)[15]. Increasing concentrations of salt and/or decreasing the pH of the system resulted in a decreased binding affinity. This indicates the importance of electrostatics in maintaining the interaction between the top face of RBBP4 and its binding partners (Figure 3.1D and E).

Competitive binding curves were designed with synthetic peptides of BCL11A, SALL4, and Histone H3. A scrambled peptide was also tested to determine the specificity of the assay. While SALL4 (aa 2-12) and BCL11A (aa 2-12) had similar inhibitory constants (K_i) of 90 nM vs 268 nM, respectively (Figure 3.2A), it is interesting to note that a shorter peptide corresponding to SALL4 amino acids 2-10 resulted in a 160-fold reduced inhibitory potential (Ki = 15 μM) (Figure 3.2B). This suggests that there may be important contacts between the 11^{th} and 12^{th} amino acids of BCL11A or SALL4, which consist of the basic histidine residue and the hydrophobic residue isoleucine or leucine, with RBBP4. In this platform, a longer peptide of BCL11A (aa 2-16) did not greatly change the K_i (268 nM vs 110 nM). A scrambled version of the BCL11A (aa 2-12) had an IC50 value > 1 mM indicating that total positive charge of the peptide is not enough to maintain the interaction with RBBP4.

3.2.2 Crystal Structure of RBBP4 with the BCL11A peptide.

To understand the molecular basis for the binding of BCL11A to RBBP4, we solved the crystal structure of RBBP4 bound to BCL11A (2-16) to 2.4 Å (Figure 3.3). Crystallography data and refinement statistics are listed in Table I. The structure was solved by molecular replacement with the apo RBBP4 protein and the initial electron density map showed clear density for the peptide bound to the surface of the WD40 propeller in the same binding pocket as H3 (Figure 3.4). The structure of RBBP4 bound to BCL11A is identical to the apo structure with an RMSD of 0.4152 A between the core atoms. The first residue of the BCL11A peptide, Ser2, was disordered, as was the sidechain of Arg3, but we saw clear density for the remainder

of the BCL11A peptide through residue 16. The sidechain of Arg4 of BCL11A occupies a similar pocket to Arg3 of H3 and makes a hydrogen bond between NH1 and Glu231 OE1 of RBBP4. There is also a well-ordered glycerol molecule in the binding pocket, which interacts with the Arg4 sidechain. Similar to Lys4 of H3, the major interaction between RBBP4 and BCL11A is Lys5, which lies in a highly negatively charged pocket on RBBP4 lined by Glu179 and Glu126 and the OD1 atom of Asn128. Gln6 OE1 makes a hydrogen bond to the backbone carbonyl of Glu395. Gly7 and Lys8 are slightly pushed out and make no interactions with the RPPB4 protein. The sidechain of Lys8 is pointed toward the solvent and disordered. Pro9 stacks with His71 of RBBP4, which pulls this residue in and starts to turn the peptide down the side of the propeller. The sidechain NE2 of Gln10 is hydrogen bonded to the backbone carbonyl of Pro9 on the peptide. His11 NE2 is hydrogen bonded to the OE2 of Glu41. The sidechain of Leu12 is pointing toward the solvent, but an important hydrogen bond between the backbone carbonyl of Leu12 and the backbone amide of Lys14 forms a tight turn of the peptide. Ser12 is pointed toward solvent. The amide of Lys14 of BCL11A is hydrogen bonded to the backbone carbonyl of Ala30 of RBBP4. Arg15 is bound to Glu41, effectively sandwiching this residue between His11 and Arg15 of the peptide. The final residue of the peptide Glu16 is hydrogen bound to NE2 of His38. Hydrogen bonds and hydrophobic interactions between RBBP4 and BCL11A, as well as intra-molecular interactions within the BCL11A peptide, are listed in Tables 3.2-3.4.

3.2.3 BCL11A peptide capable of interacting with epigenetic complexes.

To assess if, and which, epigenetic complexes interacted with BCL11A we performed a pulldown experiment with a biotin-labeled BCL11A peptide. As BCL11A is implied to be a triple-negative breast cancer gene, we chose the TNBC cell line SUM149 for these experiments. Confirming our biophysical studies, immobilized BCL11A (2-16) was able to bind RBBP4 to a significant degree over scramble control in SUM149 lysate (Figure 3.5B). In addition, RBBP7 was pulled down by BCL11A (2-16) implying that BCL11A may bind to both histone chaperones in a similar manner.

With RBBP4/7 being critical to many multi-subunit epigenetic complexes, we examined the level at which enzymatic components were pulled down by BCL11A wildtype verse scramble peptides. HDAC1 and HDAC2 were strongly associated with the BCL11A wt peptide and were indiscernible in the scramble pulldown (Figure 3.5B). These complexes pulled down by BCL11A also exhibited deacetylase activity, which is consistent with other similar studies using SALL4 as the bait (Figure 3.5D). Notably, the complexes that associated with the wt peptide had 12.8-fold greater HDAC activity (nmol/min/mL) compared to scramble BCL11A (2.04 vs. 0.16, respectively).

To distinguish between the several HDAC-containing complexes, we evaluated the presence of the NuRD-specific subunit metastasis-associated protein 1(MTA1). SIN3a and CoREST, two proteins whose names are shared with their respective complexes and associate with HDAC activity, were also analyzed. As seen in Figure 3.5, MTA1 was strongly detected in the pulldowns of wt BCL11A over

scramble, while SIN3a was modestly identified in the wt pulldown. CoREST was only witnessed in lysate loading control. As CoREST complexes are lacking RBBP4 or RBBP7, our results imply that the aminoterminus (aa 2-12) of BCL11A only interacts with epigenetic machineries that contain these two histone chaperones.

We also found the PRC2 proteins enhancer of zeste 2 (EZH2) and SUZ12 in the pulldown lysates (Figure 3.5B). While weaker washing conditions (0.1% triton) did not show any noticeable difference between scramble and wt pulldowns, harsher conditions (1% triton) clearly indicate that the interaction is more specific to wt BCL11A. These results demonstrate that BCL11A may interact with both histone deacetylase complexes and histone methyltransferases to dictate transcriptional control.

3.2.4 Treatment of SUM149 with BCL11A peptide decreases ALDH+ cancer stem cell population.

In order to better understand the functional and biological significance of the BCL11A-RBBP4 interaction, we treated SUM149 cells with the wt BCL11A (2-12) peptide for 72 hours. We hypothesized that the peptide would be able to outcompete BCL11A protein within the cells and allow us to study the impact of blocking this protein-protein interaction. First, we evaluated the delivery efficiency of a 5-FAM labeled BCL11A peptide using the Pep-1 carrier. Using flow cytometry, we found that over 98% were successfully transfected (Figure 5A) when cells were transfected at a confluency of 25%. Treatment at lower cell densities decreased transfection efficiency (data not shown).

Due to the fact that BCL11A is known to support the growth and maintenance of cancer stem cells (CSC), the treated cells were exposed to an aldefluor-activity assay[11]. This assay measures the levels of the enzyme aldehyde dehydrogenase (ALDH), which is a well-established marker for cancer stem cell activity. Treatment of the SUM149 cells with wt peptide decreased the ALDH+ population by 50%, as measured by flow cytometry, in comparison to cells treated with a BCL11A scramble peptide (Figure 5B and C). These results support the notion that blocking the interaction between BCL11A and RBBP4 could provide a novel method for the treatment of CSCs.

3.3 Discussion.

As breast cancer is the second-most leading cause of death in women, there are growing demands for the development of new and more effective therapeutics. This is especially true of TNBC for which there are currently no targeted treatments. BCL11A was recently discovered to be a critical TNBC gene, sparking a lot of interest in better understanding the molecular mechanisms by which it operates. Furthermore, as BCL11A is a transcription factor it is notoriously difficult to inhibit the DNA-protein interface with a small molecule. Uncovering binding partners of BCL11A and studying how this transcription factor recruits epigenetic complexes is therefore crucial for the development of inhibitors. These inhibitors can be developed to abrogate BCL11A-mediated repression by targeting downstream effectors or BCL11A-associated proteins that may be more susceptible to chemical intervention.

In this study, we found that BCL11A resides in a binding site on the top face of RBBP4; one that is shared by histone H3 as well as other transcription factors, including SALL4, FOG1, and PHF6[13-17]. As it shares the binding pocket, these interactions are likely mutually exclusive. Likewise, since RBBP4 and RBBP7 share 92% sequence homology, BCL11A may interact with either histone chaperone[18]. BCL11A may recruit epigenetic complexes to specific chromatin regions through its' interaction with RBBP4/7 and thereby alter transcription of certain genes. The other known binding pocket on RBBP4/7, existing on the side of the β-propeller between three distinguishing structural elements (PP loop, C-helix, and N-helix), is occupied by either MTA1 of the NuRD complex, SUZ12 of PRC2, or histone H4[15,19,20]. Compiling these data and crystal structures supports the understanding of the architecture and recruitment of multi-subunit complexes to their targets. While RBBP4/7 is able to interact with both histones H3 and H4 during nucleosome assembly, when present in PRC2 or NuRD, RBBP4/7 may be limited to interactions with H3 or transcription factors.

Recently, studies have suggested that multiple copies of RBBP4 or RBBP7 could be recruited to the NuRD complex through its' MTA1 subunit[21,22]. Similarly, PRC2 may exist in a dimer state within the cell[23]. When taken together, RBBP4/7 may be capable of surveying the surrounding epigenetic landscape and thereby coordinate signals from histone tails and transcription factors. This could be a mechanism by which RBBP4/7 fine-tune the recruitment of its' associated co-repressor complexes. It may also provide a threshold whereby multiple copies of a transcription factor must be present to elicit a certain response. The

aforementioned hypothesis was demonstrated for FOG1, which interacts with RBBP4/7 and MTA1; it suggests that multiple points of contact must be made for the optimal recruitment of NuRD[14]. As BCL11A shares a highly similar sequence to FOG1, it is possible that it may also interact with MTA1.

Our crystallography studies further our understanding of the interactions between BCL11A and RBBP4. Similar to other transcription factors that bind to the top face of RBBP4, Arg-4 of BCL11A was inserted into the core of the β-propeller while Lys-5 was anchored to the surface of RBBP4 via hydrogen bonds. Modifications to either of these residues hinder binding to RBBP4[15]. Consistent with our structure, Arg-3 is not an important mediator for this PPI or others[14,16]. Interestingly, the crystal structure of BCL11A in complex with RBBP4 exposes new interactions on the side face of the histone chaperone that are not seen with NURF55/RBBP4 and histone H3 or RBBP4 and FOG1. Whereas no discernable interactions occur outside residues aa 11-14 of either the H3 peptide or FOG1, in our crystal structure we observe H-bonding that anchors His-11, Arg-15 and Glu-16 of BCL11A to the side of RBBP4[14]. Glu-16 occupies a small trough-like cavity that may be susceptible to inhibition via small-molecules. In addition, it was recently shown that truncating FOG1 below the first 15 amino acids drastically impacted the ability of the peptide to pulldown the NuRD complex[24]. Our results help explain the necessity of the C-terminal residues of the transcription factors, FOG1, SALL4 or BCL11A, in efficiently interacting with RBBP4 and endogenous epigenetic complexes.

From our pulldown experiments, we were able to demonstrate that the BCL11A (aa 2-16) peptide was able to interact strongly with RBBP4 and RBBP7 over the scramble control in lysate from SUM149. As RBBP4/7 are common to several epigenetic complexes, we investigated the pulldown lysate for existing proteins that could help discriminate between NuRD, PRC2, SIN3A and CoREST. We were able to demonstrate that BCL11A strongly recruits NuRD, but also interacts with PRC2 and SIN3A to a lesser extent. Indeed, the isolated proteins exhibit notable HDAC activity. These data suggest BCL11A may impose transcriptional repression through either histone deacetylase complexes or histone methyltrasnferases. Future studies will focus on an endogenous co-IP with wt BCL11A verses an N-terminal truncated (aa 1-12) form of BCL11A to identify if these residues are essential to the recruitment of epigenetic complexes and function of BCL11A in TNBC.

From a therapeutic perspective we found that incubation of an aminoterminal (aa 2-12) peptide of BCL11A was able to decrease the ALDH+ CSC population in SUM149. This demonstrates that the BCL11A (2-12) peptide was capable of blocking the interaction of endogenous BCL11A with RBBP4 and thereby hinders the recruitment of epigenetic complexes. Reducing the ALDH+ cell population is imperative for preventing recurrence and improving survival in breast cancer as it is associated with poor clinical outcome and metastasis [25-28]. Therefore, our data warrants further investigation into the discovery of small-molecule inhibitors of this PPI. Blocking the interaction between BCL11A and RBBP4/7 will not only provide a route to eliminate breast cancer stem cells, but it will also shed light on the genes targeted by BCL11A and how it regulates tumorigenesis (Figure

3.7). Insight into the regulation of transcriptional control via multiprotein complexes, like NuRD and PRC2, is of high importance as these mechanisms are often awry in disease-states.

Methods

Cell Culture and reagents. SUM149 cells were cultured in Ham's F-12 medium supplemented with 5% FBS, hydrocortisone (1 μg/mL), insulin (5 μg/mL) and 5 mL Antibiotic/Antimycotic (Thermo Fisher). The following antibodies were used in this study: Anti-RbAp48 (RBBP4) rabbit polyclonal (ab47456, Abcam; WB 1:2000), Anti-RbAp46 (RBBP7) rabbit polyclonal (ab3535, Abcam; WB 1:2000), Anti-SUZ12 rabbit monoclonal (#3737, Cell Signaling; WB 1:1000), Anti-EZH2 rabbit monoclonal (#5426, Cell Signaling; WB 1:1000), Anti-HDAC1 mouse monoclonal (#5356, Cell Signaling; WB 1:1000), Anti-HDAC2 mouse monoclonal (#5113, Cell Signaling; WB 1:1000), Anti-CoREST rabbit monoclonal (#14567, Cell Signaling; WB 1:1000), Anti-SIN3A rabbit monoclonal (#8056, Cell Signaling; WB 1:1000), Anti-MTA1 rabbit monoclonal (#5647, Cell Signaling; WB 1:1000), goat Anti-rabbit IgG, HRP-linked (#7074, Cell Signaling; WB 1:2000), goat Anti-mouse IgG, HRP-linked (sc-2005, Santa Cruz).

Protein Purification. Cloning, expression and purification of RBBP4 was performed similarly to previous reports[14]. Briefly, full length RBBP4 (aa 1-425) was cloned into a pFastBac HT-A vector with both an N-terminal His_6 tag and a tobacco etch virus (TEV) cleavage site. Generation of recombinant bacmid and virus was performed using the Bac-to-Bac (Invitrogen) expression system and according to manufacturer's recommendations. Protein was expressed in Tn5 (High Five™; Invitrogen) cells using P3 virus. Infected cells were collected via centrifugation, washed with ice cold PBS twice and the pellet was lysed immediately or flash frozen in liquid nitrogen for later use. All purification steps were performed at 4°C unless

otherwise noted. Infected cells were lysed using lysis buffer (50 mM Tris (ph 8), 500 mM NaCl , 0.1% NP40, 5 mM imidazole,10% glycerol) with protease inhibitors and nuclease (Pierce™ Thermo Scientific) added right before use. Lysis was further aided by sonication. The lysate was clarified by centrifugation for 40 min at 12,000 rpm and supernatant was added to nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen) for 2 hours on a shaker. Resin was collected and washed with high-salt buffer (50 mM Tris (ph 8), 500 mM NaCl , 10 mM imidazole,10% glycerol) three times. Resin was then washed with low-salt buffer (50 mM Tris (ph 8), 150 mM NaCl , 10 mM imidazole,10% glycerol) to equilibrate before elution. RBBP4 was eluted using the low-salt buffer with 250 mM imidazole. Fractions containing RBBP4 were collected, pooled, and concentrated using an Amicon® Ultra-4 Centrifugal Filter (EMD Millipore).

Peptide Synthesis. All peptides were synthesized manually or with an ABI 433 peptide synthesizer using Fmoc chemistry with Rink Amide resin as the solid support with the exception of H3 (1-21), FAM labeled probe that was purchased from Anaspec (cat. No. AS-63824). Either DIC/HOAt or HOBt/HBTU was used as the coupling reagent. Following completion of the peptide, a cleavage cocktail composed of TFA:TIS:H_2O:Thioanisole (19 mL: 0.5 mL: 1 mL: 1 mL) removed the peptide from the resin as well as any side-chain protecting groups. The resulting solution was evaporated and the crude peptide was precipitated with diethyl ether. Peptides were purified via RP-HPLC (Waters, Sunfire Prep C18, 19 mm x 150 mm, 5 μm) and confirmed by electrospray ionization mass spectroscopy (ESI-MS). For biotin-labeled peptides, Fmoc-PEG Biotin NovaTag™ Resin (EMD Millipore) was used as

solid support. To generate a scramble control, permutations of the original sequence were made[29].

Fluorescence Polarization Assay. All fluorescence polarization experiments were conducted in 384-well, black, low volume, round-bottom plates (Corning) using a BioTek Synergy 2 plate reader (Winooski, VT). To each well, was added increasing amounts of protein and the 5-carboxyfluorescein (5-Fam) labeled Histone H3 N-terminal probe/tracer (20 nM) to a final volume of 20 uL in the assay buffer (50 mM Tris-HCl, pH=7.4, 100 mM NaCl, 0.1% glycerol unless otherwise indicated). The plate was allowed to incubate at room temperature for 5 min to reach equilibrium. The polarization values (mP) were measured at an excitation wavelength at 485 nm and an emission wavelength at 528 nm. An equilibrium-binding isotherm was constructed by plotting the mP values as a function of the protein concentration at a fixed concentration of tracer (20 nM). All experimental data were analyzed using Prism 7.0 software (Graphpad Software, San Diego, CA) and WinNonlin (7.3).

Competitive Binding Assay. All IC50 values discussed were identified using the competitive binding assay. To a 384-well, black, low volume, round-bottom plate (Corning) was added 10 μL RBBP4 at 2 μM and 4 μL of 100 nM H3 (1-21) 5-FAM probe. Dilutions of test compound (in assay buffer) or peptide were added to the wells to give a final volume of 20 μL and final concentrations of 1 μM RBBP4 and 20 nM probe. All experiments were run in 50 mM tris-hcl pH 7.5, 100 mM NaCl and 0.1% glycerol unless otherwise noted. In addition, there were three controls on each plate: blank (buffer only), 100% inhibition (probe only) and 0% inhibition

(probe and protein only). K_i values were determined using the previously derived equation in Nikolovska-Coleska, et al[30].

Protein Crystallization and Structure Determination. Prior to crystallization, the N-terminal His_6-tag was removed from RbBP4 via overnight dialysis against 50 mM Tris, pH 7.5, 150 mM NaCl and X mgs of TEV protease. The cleaved protein was then purified by gel filtration on a Superdex 75 column in buffer containing 50 mM Tris, pH 7.5 and 150 mM NaCl. The protein was concentrated to 7.7 mg/mL and incubated with a 1.1-fold molar excess of Bcl11a (2-16) overnight at 4 °C. The complex crystallized from drops containing equal volumes of complex and well solution (25% polyethylene glycol monomethyl ether (MW 2000) and 0.1 M potassium thiocynate). Prior to data collection, crystals were cryoprotected in well solution containing 25% glycerol. All data were collected at LS-CAT at the Advanced Photon Source at Argonne National Lab on line 21-ID-G equipped with a Mar300 detector. Data were processed and scaled with HKL2000[31]. The structure was solved by molecular replacement with Phaser (CCP4 suite[32]) using a previously solved structure of RPPB4 (PDB 4R7A) as the starting model. The structure was refined using Buster[33] with iterative rounds of fitting in COOT[34]. Structures were validated with Molprobity[35]. Data refinement and statistics are given in Table 3.1.

Pulldown. Cellular lysate was prepared from SUM149 cells using RIPA buffer (with protease and phosphatase inhibitors added prior to use). Cells were incubated for 30 min on a shaker at 4°C and lysate was cleared by centrifugation. Protein concentration was measured using the Bradford assay (Bio-Rad). Pierce

Streptavidin Agarose Resin (Thermo Scientific) were mixed by inversion and 30 µL was pipetted into two eppendorf tubes. Resin was washed two times with 500 µL buffer A (PBS + 0.1% Triton X-100) and centrifuged at 3000 rpm for 5 min to remove the supernatant. To prepare the resin for pulldown, 1 µL (1 mg/mL) biotinylated peptide (either wt 2-16 BCL11A or scramble) was added to the washed resin in 500 µL buffer A and incubated for 2 hours at 4°C on a rocking shaker. The beads were washed three times with buffer A and used immediately. 600 µg of SUM149 cellular lysate was added to each tube and the volume was increased to 500 µL with buffer A (plus protease and phosphatase inhibitors). The mixture was incubated overnight at 4°C on a rocking shaker. The following morning the beads were centrifuged for 5 min at 3000 rpm and the supernatant was removed. Beads were washed three times with variations of buffer A (increasing detergent and/or salt, plus protease and phosphatase inhibitors). After the final wash, centrifugation and collection of the beads, 40 µL 2X SDS was added to each tube and heated for 5 min at 95 °C. Samples were separated and analyzed via SDS page.

HDAC Activity Assay. HDAC activity was measured using the HDAC Fluorometric Activity Assay Kit (Cat. No. 10011563, Cayman Chemical). Briefly, after performing pull down as described above, beads were washed 3 times and then resuspended in HDAC assay buffer (25 mM Tris-HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, and 1 mM $MgCl_2$). 10 µL of each sample was added to a well of black 96-well plate in triplicate. The samples were then diluted further in 140 µL assay buffer. 10 µL Trichostatin A (final concentration 1 µM) was added to a control well for each sample group to remove background signal. HDAC substrate (10 µL) was added last to all wells and

the plate was incubated at 37°C on an orbital shaker for 30 minutes. Developer was added to each well following incubation. After 15 minutes at room temperature, fluorescence was read (ex350/em450) on a BioTek Synergy HT plate reader. Fluorescence of the standard wells was plotted as a function of deacetylated substrate concentration to give a standard curve. The deacetylated concentration of the samples was measured using this equation:

Deacetylated compound (μM) =

$$\frac{((\text{Average fluorescence of samples} - \text{average fluorescence of TSA treated samples}) - (\text{y} - \text{int}))}{\text{Slope}}$$

HDAC activity (nmol/min/mL) was calculated taking the deacetylated substrate concentration, dividing by the incubation time (30 minutes), and then multiplying by sample dilution.

Western Blot. Collected samples (15 μL) from the pulldown were run on a 4-15% gradient gel for 5 minutes at 50V, then 120V for 1-1.5 hour(s). Proteins were transferred to a nitrocellulose membrane (0.2 mm) under constant current of 220 mA. The membrane was blocked with 5% Milk in PBST (PBS + 0.1% Tween 20). Primary antibodies were added according to manufacturer's recommendations in 1% Milk in PBST. The following morning, the membrane was washed 3 times with PBST and then the corresponding secondary antibody was added to the membrane for 2 hours at room temperature. The membrane was washed again using PBST for three times before being exposed to enhanced chemiluminescence (ECL) reagent (Pierce Thermo Scientific). Proteins were detected using X-ray development.

Transient Transfection. BCL11A (aa 2-12) or scramble peptides were delivered to SUM149 cells using Chariot™ Protein Delivery Reagent (Active Motif) according to manufacture's instructions. SUM149 cells were seeded into a 6 well plate 24 hours prior to treatment. Peptide stocks (2 mM) were diluted in 100 µL PBS so that the final treatment concentration would be 20 µM. Chariot reagent, diluted 1:10, was further diluted in 100 µL sterile water. Peptide and Chariot solutions were combined and incubated for 30 min, after which the mixture was overlaid onto cells. 400 µL serum-free media was added and cells were incubated for 2 hours at 37°C before complete media was added. Treated cells were assayed for ALDH activity 72 hours post-transfection. Transfection efficiency and conditions were optimized using a 5-FAM labeled BCL11A (aa 2-12) peptide.

Aldehyde dehydrogenase (ALDH+) assay. Aldehyde dehydrogenase (ALDH) activity was measured by the ALDEFLUOR kit (StemCell Technologies). Briefly, cells were incubated with ALDEFLUOR reagent for 30 minutes at 37 °C. As negative control, cells were incubated with the reagent and the ALDH inhibitor, diethylaminobenzaldehyde (DEAB). After incubation, cells were washed by PBS containing 2% FBS and then re-suspended with the Assay buffer for FACS analysis. ALDEFLUOR-positive gates were based on the DEAB treated negative control samples.

Acknowledgements: This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817).

Contributions: Rebecca Reed expressed and purified RBBP4/7 protein, synthesized all peptides, performed all FP experiments, prepared cell lysates and designed pulldown experiment. Nicholas Stevers and Samantha Tinsley ran western blot of pulldown experiments. Jennifer Meagher solved crystal structure of RBBP4 in complex with the BCL11A peptide. Albert Lin did the ALDH+ activity assay on cells prepared by Rebecca Reed.

References

1 Xu, J. *et al.* Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci U S A* **110**, 6518-6523, doi:10.1073/pnas.1303976110 (2013).

2 Liu, H. *et al.* Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. *Mol Cancer* **5**, 18, doi:10.1186/1476-4598-5-18 (2006).

3 Satterwhite, E. *et al.* The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. *Blood* **98**, 3413-3420 (2001).

4 Jiang, B. Y. *et al.* BCL11A overexpression predicts survival and relapse in non-small cell lung cancer and is modulated by microRNA-30a and gene amplification. *Mol Cancer* **12**, 61, doi:10.1186/1476-4598-12-61 (2013).

5 Senawong, T., Peterson, V. J. & Leid, M. BCL11A-dependent recruitment of SIRT1 to a promoter template in mammalian cells results in histone deacetylation and transcriptional repression. *Arch Biochem Biophys* **434**, 316-325, doi:10.1016/j.abb.2004.10.028 (2005).

6 Yin, B. *et al.* A retroviral mutagenesis screen reveals strong cooperation between Bcl11a overexpression and loss of the Nf1 tumor suppressor gene. *Blood* **113**, 1075-1085, doi:10.1182/blood-2008-03-144436 (2009).

7 Yu, Y. *et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med* **209**, 2467-2483, doi:10.1084/jem.20121846 (2012).

8 Chou, J. *et al.* GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nat Cell Biol* **15**, 201-213, doi:10.1038/ncb2672 (2013).

9 Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. & Werb, Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**, 1041-1055, doi:10.1016/j.cell.2006.09.048 (2006).

10 Drabsch, Y., Robert, R. G. & Gonda, T. J. MYB suppresses differentiation and apoptosis of human breast cancer cells. *Breast Cancer Res* **12**, R55, doi:10.1186/bcr2614 (2010).

11 Khaled, W. T. *et al.* BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nat Commun* **6**, 5987, doi:10.1038/ncomms6987 (2015).

12 Topark-Ngarm, A. *et al.* CTIP2 associates with the NuRD complex on the promoter of p57KIP2, a newly identified CTIP2 target gene. *J Biol Chem* **281**, 32272-32283, doi:10.1074/jbc.M602776200 (2006).

13 Gao, C. *et al.* Targeting transcription factor SALL4 in acute myeloid leukemia by interrupting its interaction with an epigenetic complex. *Blood* **121**, 1413-1421, doi:10.1182/blood-2012-04-424275 (2013).

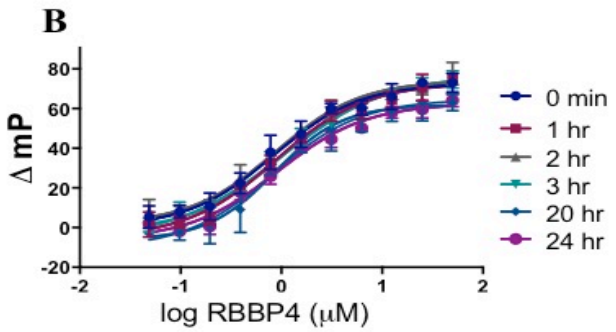
14 Lejon, S. *et al.* Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48.FOG-1 complex. *J Biol Chem* **286**, 1196-1203, doi:10.1074/jbc.M110.195842 (2011).

15 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
16 Lauberth, S. M. & Rauchman, M. A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J Biol Chem* **281**, 23922-23931, doi:10.1074/jbc.M513461200 (2006).
17 Liu, Z. *et al.* Structural basis of plant homeodomain finger 6 (PHF6) recognition by the retinoblastoma binding protein 4 (RBBP4) component of the nucleosome remodeling and deacetylase (NuRD) complex. *J Biol Chem* **290**, 6630-6638, doi:10.1074/jbc.M114.610196 (2015).
18 Xu, C. & Min, J. Structure and function of WD40 domain proteins. *Protein Cell* **2**, 202-214, doi:10.1007/s13238-011-1018-1 (2011).
19 Alqarni, S. S. *et al.* Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. *J Biol Chem* **289**, 21844-21855, doi:10.1074/jbc.M114.558940 (2014).
20 Murzina, N. V. *et al.* Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure* **16**, 1077-1085, doi:10.1016/j.str.2008.05.006 (2008).
21 Schmidberger, J. W. *et al.* The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7. *Protein Sci* **25**, 1472-1482, doi:10.1002/pro.2943 (2016).
22 Millard, C. J. *et al.* The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *Elife* **5**, e13941, doi:10.7554/eLife.13941 (2016).
23 Davidovich, C., Goodrich, K. J., Gooding, A. R. & Cech, T. R. A dimeric state for PRC2. *Nucleic Acids Res* **42**, 9236-9248, doi:10.1093/nar/gku540 (2014).
24 Saathoff, H. *et al.* A peptide affinity reagent for isolating an intact and catalytically active multi-protein complex from mammalian cells. *Bioorg Med Chem* **23**, 960-965, doi:10.1016/j.bmc.2015.01.023 (2015).
25 Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78-91, doi:10.1016/j.stemcr.2013.11.009 (2014).
26 Ginestier, C. *et al.* ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**, 555-567, doi:10.1016/j.stem.2007.08.014 (2007).
27 Marcato, P. *et al.* Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. *Stem Cells* **29**, 32-45, doi:10.1002/stem.563 (2011).
28 Charafe-Jauffret, E. *et al.* Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clin Cancer Res* **16**, 45-55, doi:10.1158/1078-0432.CCR-09-1630 (2010).
29 Connarn, J. N. *et al.* The molecular chaperone Hsp70 activates protein phosphatase 5 (PP5) by binding the tetratricopeptide repeat (TPR) domain. *J Biol Chem* **289**, 2908-2917, doi:10.1074/jbc.M113.519421 (2014).

30 Nikolovska-Coleska, Z. *et al.* Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem* **332**, 261-273, doi:10.1016/j.ab.2004.05.055 (2004).
31 Otwinowski, Z. a. W. M. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **Volume 276: Macromolecular Crystallography, part A**, p.307-326 (1997).
32 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
33 BUSTER (Global Phasing Ltd, Cambridge, United Kingdom, 2011).
34 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).
35 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).

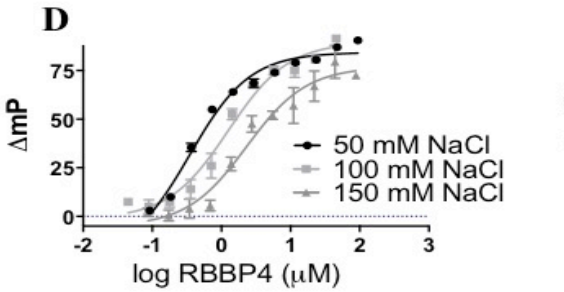
A

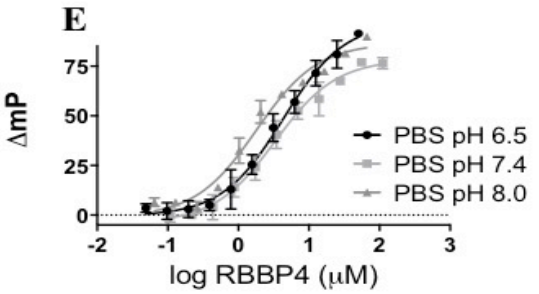
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
FOG1	M	S	R	R	K	Q	S	N	P	R	Q	I	K	R	S	L
SALL4A	M	S	R	R	K	Q	A	K	P	Q	H	I	N	S	E	E
Histone H3	M	A	R	T	K	Q	T	A	R	K	S	T	G	G	K	A
BCL11A	M	S	R	R	K	Q	G	K	P	Q	H	L	S	K	R	E



C

	Kd (μM)
RBBP4	0.84 ± 0.08
RBBP7	1.20 ± 0.10



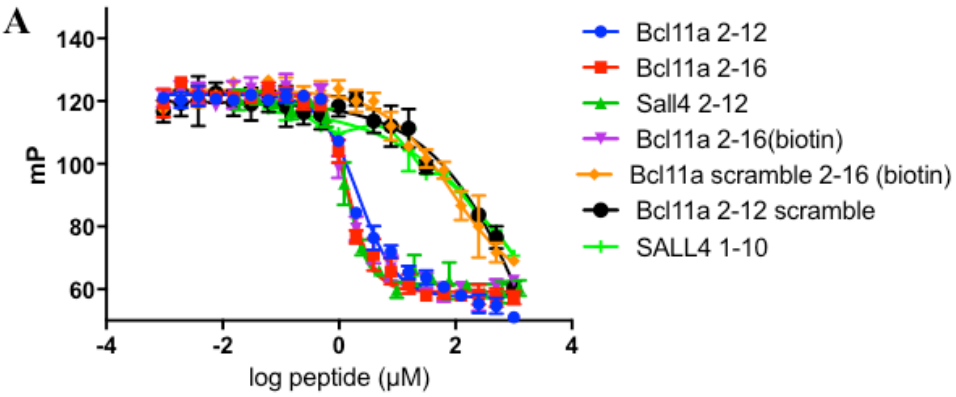


	Kd (μM)
50 mM NaCl	0.348
100 mM NaCl	1.26
150 mM NaCl	2.31

	Kd (μM)
pH 6.5	4.3
pH 7.4	3.3
pH 8.0	1.8

Figure 3.1 Full Binding Curves for RBBP4 and RBBP7 to 5-FAM H3 peptide. (A) Sequences of BCL11A, FOG1, SALL4, and Histone H3 peptides along with numbering. All peptides share RKQ motif thought to be most important for binding to RBBP4 (B) Full binding isotherm of a H3 aa 1-21 peptide labeled at C-terminus with 5-FAM to RBBP4. Labeled peptide was kept constant at 20 nM and incubated with increasing concentrations of purified full length RBBP4. mP values were measured at the times indicated. Interaction was stable over 24 hours. (C) K_d values of H3 1-21 (5FAM) to either RBBP4 or RBBP7. The effects of salt concentration (D) and pH (E) on binding affinity were also examined. Data presented as mean ± S.D. (n=3).

A



B

	IC50 (μM)	K_i (μM)
BCL11A 2-12	2.16	0.27
BCL11A 2-16	1.45	0.11
SALL4A 2-12	1.37	0.09
BCL11A 2-16(biotin)	1.34	0.08
BCL11A scramble2-16(biotin)	81.3	18.1
BCL11A scramble 2-12	>1000	N/A
SALL4A 1-10	65.5	14.6

Figure 3.2 Competitive Binding Curves.
(A) Representative competitive binding curves of N-terminal peptides of BCL11A, SALL4A and scramble peptides to RBBP4 in optimized buffer conditions (50 mM Tris, 50 mM NaCl, 0.1% glycerol, pH 7.5). Peptides used in competitive binding experiments had a free amino-terminal end with a capped amide at C-terminus. As unlabeled peptide concentration increased, tracer was subsequently displaced from binding to RBBP4. Control scramble peptide of BCL11A demonstrates interaction is specific as it has diminished capacity to displace probe. Data presented as mean ± S.D. (n=3). (B) IC50 and K_i values (calculated using the equation in ref.30) of the peptides tested.

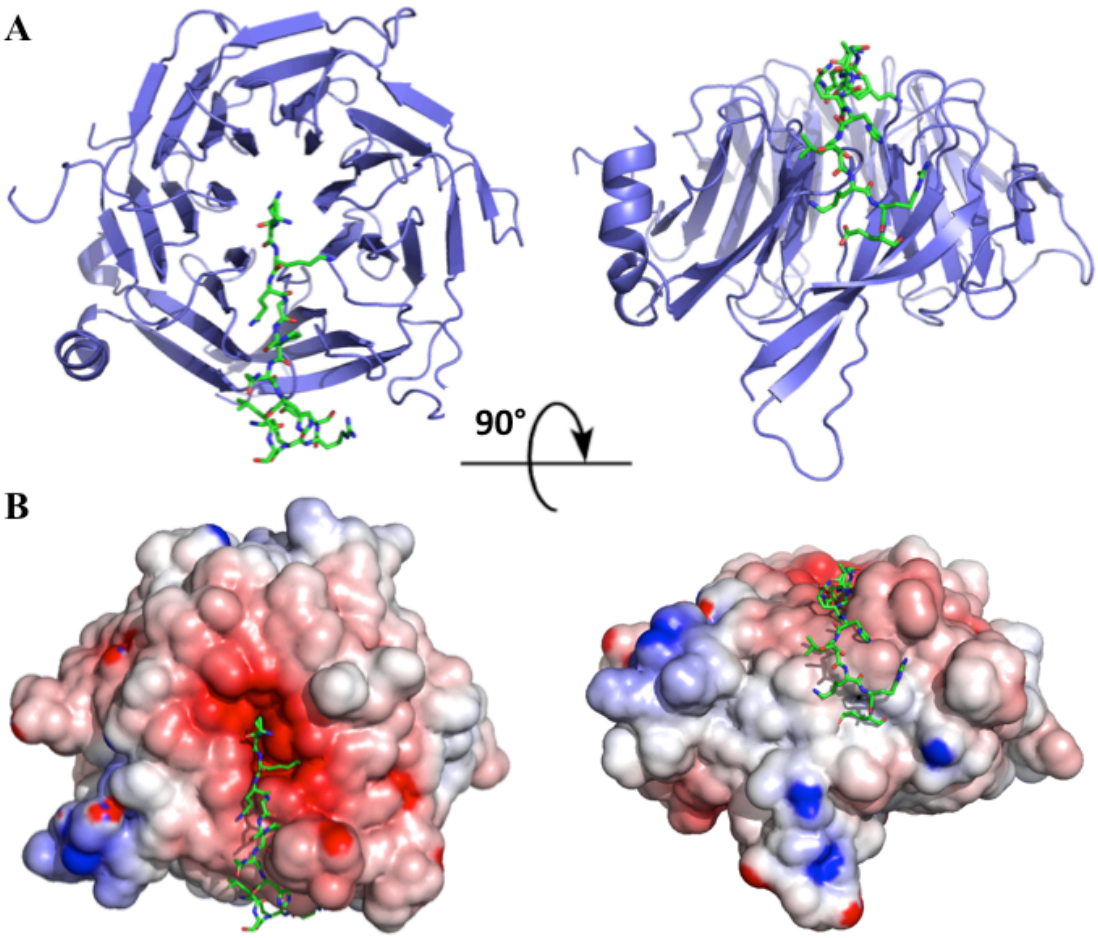


Figure 3.3 Complex of RBBP4 and BCL11A-derived peptide.
Crystal Structure of BCL11A peptide bound to RBBP4. (A) BCL11A (2-16) shown in green bound to the top face of the RBBP4 β-propeller. (B) The same figures shown with representative surface potential of the RBBP4 binding pocket.

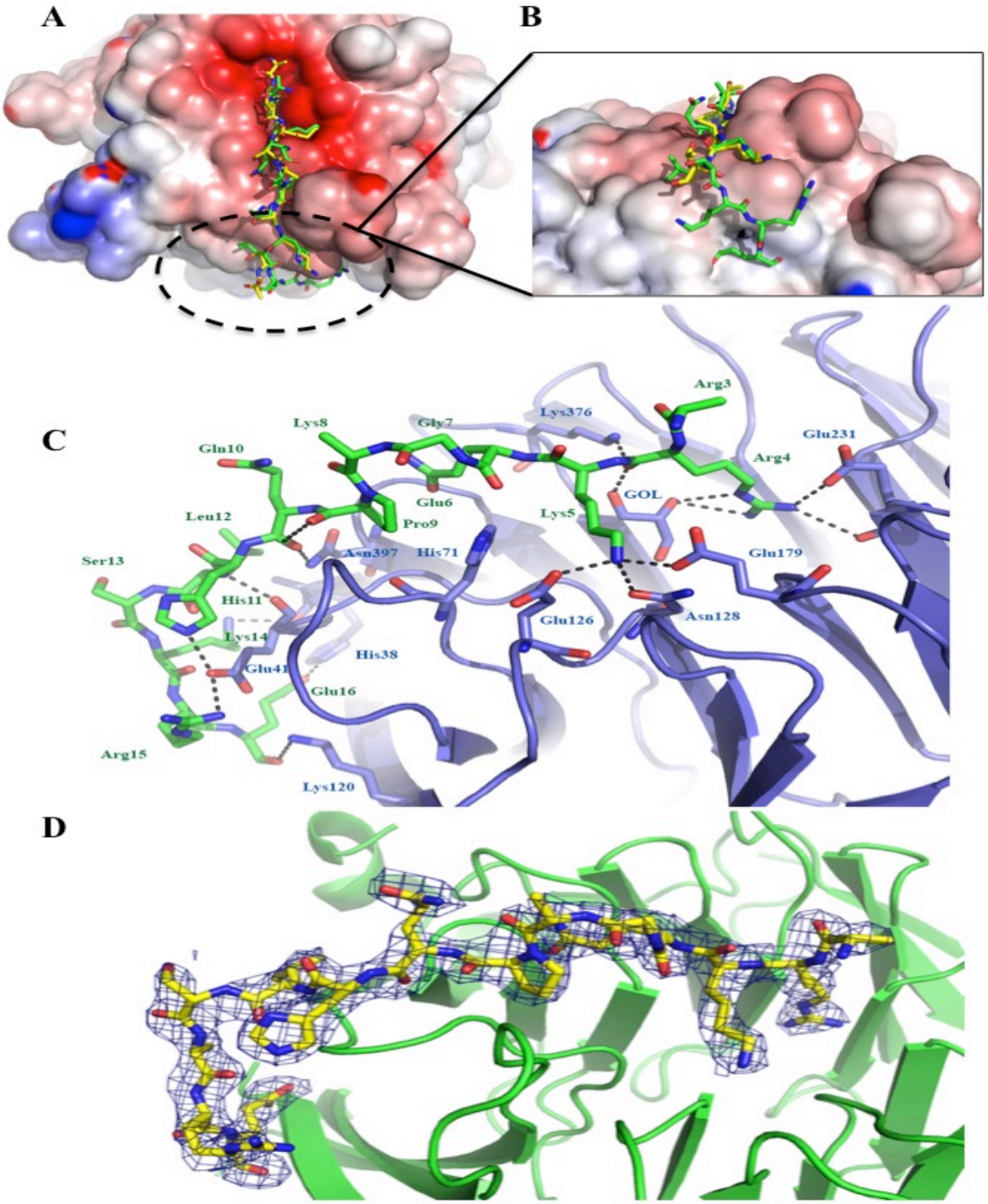


Figure 3.4 Comparison of histone H3 and BCL11A peptides bound to RBBP4. (A) Overlay of the H3 peptide (shown in yellow)[15] (PDB: 2YBA) and the BCL11A peptide (shown in green) bound to the top face of the RBBP4 β-propeller indicating the interactions are likely mutually exclusive. (B) A view of the novel interactions of the BCL11A peptide to the side of RBBP4 and (C) a detailed map of the hydrogen bonds between RBBP4 and BCL11A amino acids. (D) 2fo-fc Electron density map of BCL11A Peptide

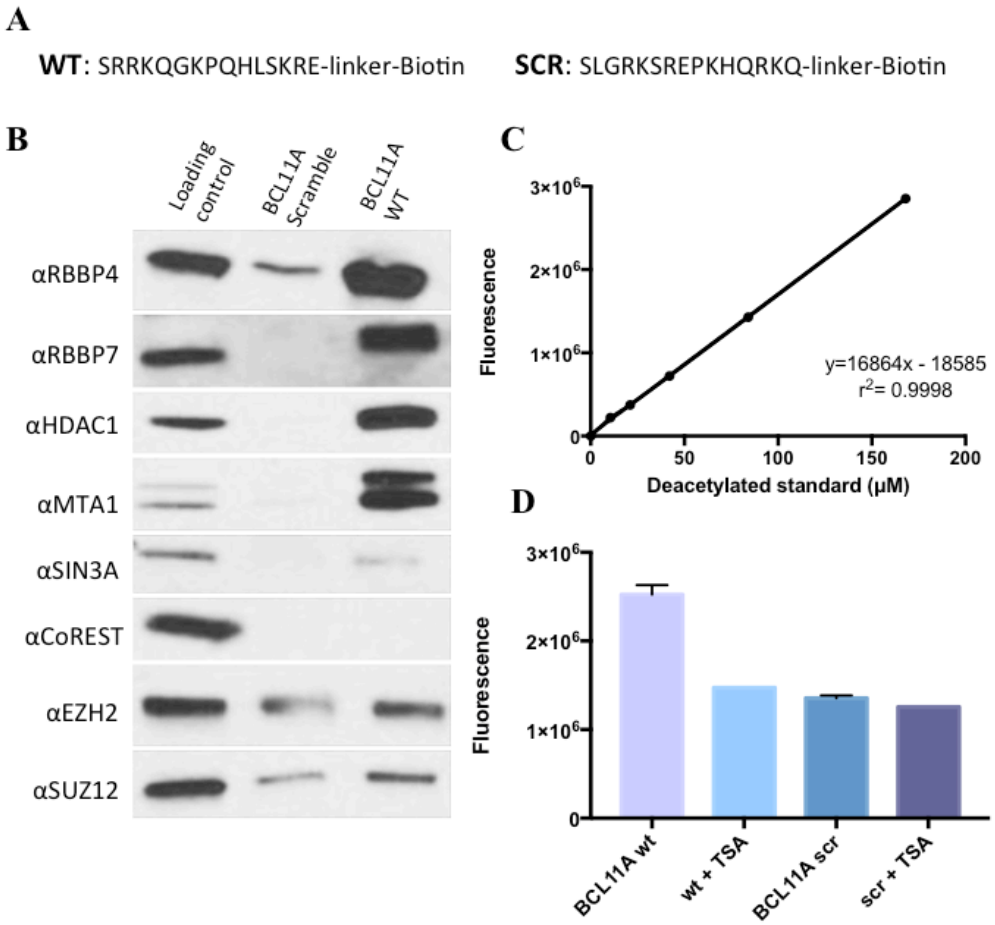


Figure 3.5 A peptide derived from the aminoterminus of BCL11A is able to interact with several epigenetic complexes.
(A) The top panel shows the amino acid sequences of the biotin-labeled BCL11A and scramble peptides used in our pulldown studies. (B). The wt peptide interacts strongly with RBBP4 and RBBP7 over scramble control. Biotin-labeled peptides where pre-incubated with streptavidin agarose beads and then exposed to lysate from SUM149 cells. Proteins bound to peptide were eluted using 2X SDS and collected via centrifugation, followed by western blotting using anti-RBBP4 and anti-RBBP7. BCL11A, through the first 16 amino acids, interacts only with epigenetic complexes that contain RBBP4 or RBBP7. Western blots were performed with anti-HDAC1, anti-HDAC2, anti-MTA1, anti-EZH2, anti-SUZ12, anti-SIN3A and anti-CoREST to distinguish between potential co-effector epigenetic complexes. CoREST, which does not associate with RBBP4/RBBP7, was only seen in lysate-loaded control. (C) Standard curve of deacetylated substrate used to calculate HDAC activity of pulldown supernatants. (D) BCL11A peptide interacts with active HDAC complexes. SUM149 lysate was incubated with biotin-labeled BCL11A or scramble peptide and then immunoprecipitated with streptavidin agarose beads. Complexes that were pulled down were tested in an HDAC activity assay. A 10% loading input, or lysate that was not subjected to the pulldown, was used as a control.

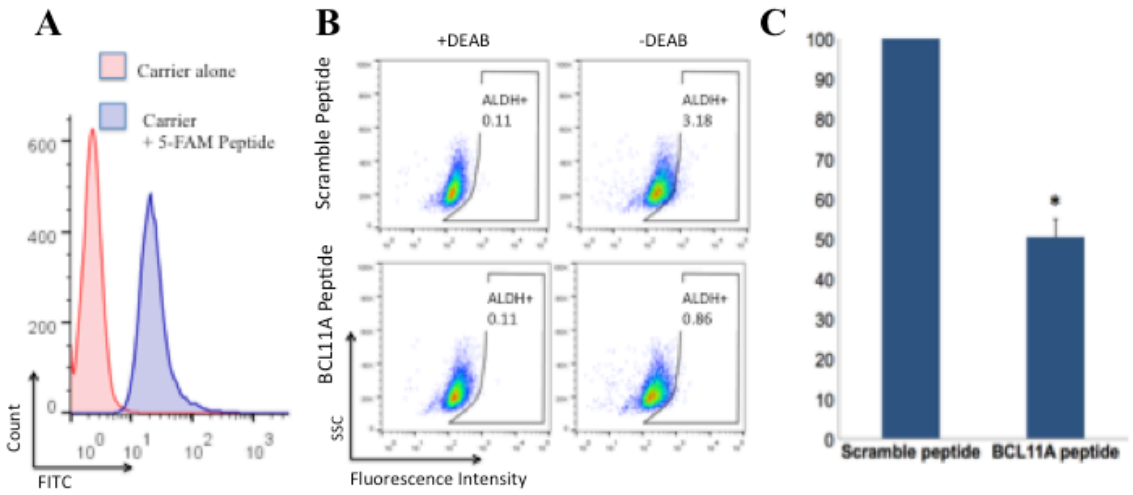


Figure 3.6 An 11-mer BCL11A peptide is able to decrease the ALDH+ Cancer Stem Cell (CSC) population.

(A) Optimization of the transfection of SUM149 cells with a 5-FAM labeled control peptide using Chariot, Pep-1, as a carrier. Treatment of SUM149 cells at 25% confluency with peptide and carrier resulted in a successful transfection of over 95% over the control, non-treated cells. (B) SUM149 cells treated were treated with BCL11A (2-12) wt peptide or scramble for 72 hours, collected, and then subjected to the ALDEFLUOR assay as described in the Methods. Top panel shows representative dot plots of scramble-treated cells with and without DEAB, while the bottom panel shows representative dot plots of wt-treated cells. (C) BCL11A wt peptide decreased ALDH+ activity by 50% over scramble in SUM149 cells. Data presented as mean ± S.D. of three independent experiments (n=3). * $p < 0.05$

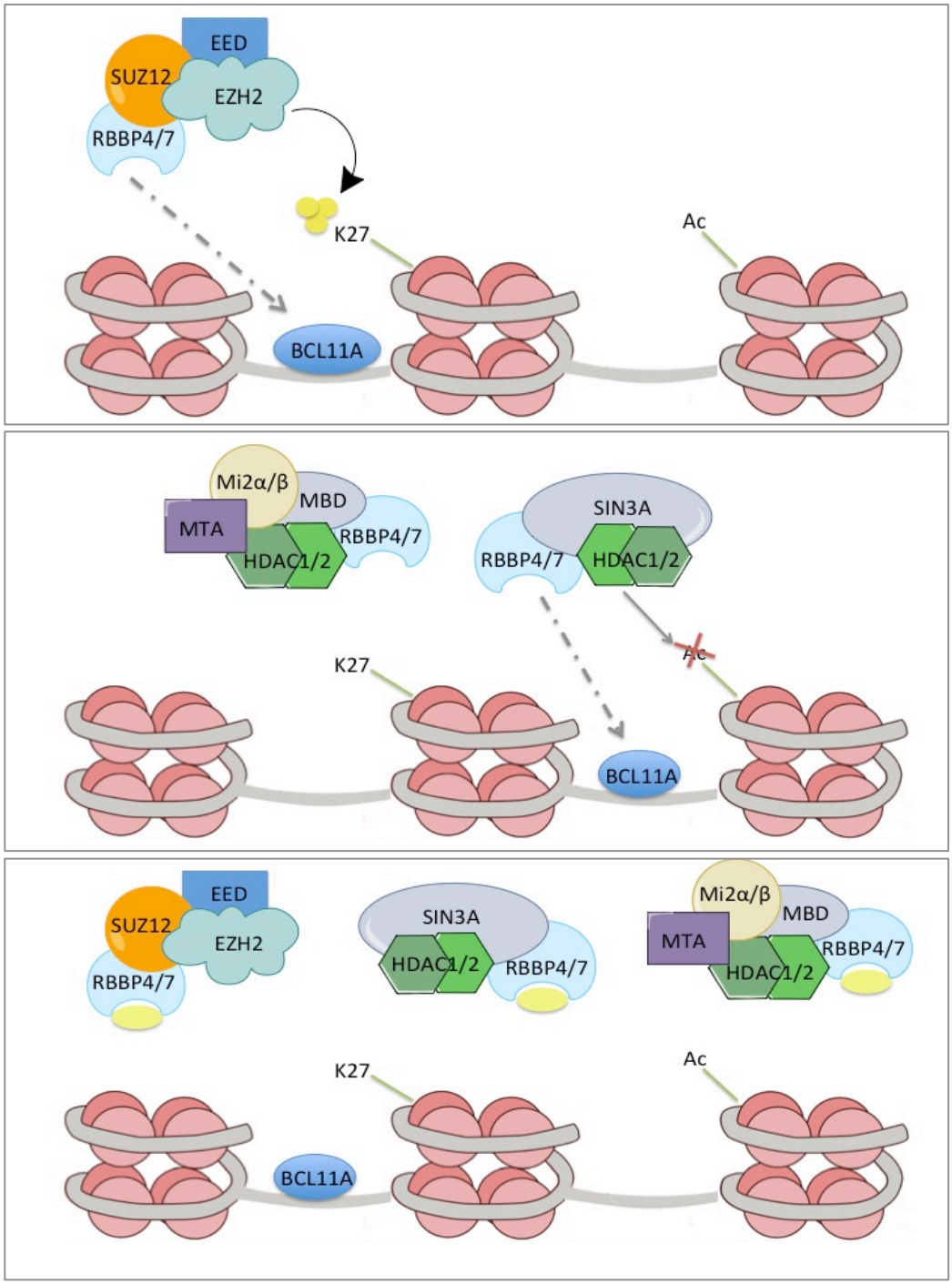


Figure 3.7 Diagram of the potential use for small-molecule inhibitors of the BCL11A-RBBP4 protein-protein interaction.
In the top two panels, BCL11A recruits PRC2 or HDAC complexes (SIN3A and NuRD) to repress gene transcription. With an inhibitor, as depicted in bottom panel, BCL11A will fail to direct these epigenetic complexes to its' target genes.

Table 3.1: Crystallography Data Collection and Refinement Statistics

Data Collection	RBBP4-Bcl11a (2-16)
SpaceGroup	$P2_12_12_1$
Unit Cell a, b, c (Å)	
Wavelength (Å)	0.97786
Resolution (Å)[1]	2.40 (2.44-2.40)
Rsym](%)[2]	11.5 (56.6)
<I/sI>[3]	10 (3)
Completeness (%)[4]	99.7 (98.7)
Redundancy	8.2 (8.0)
Refinement	
Resolution (Å)	2.40
R-Factor [5]	18.6
Rfree [6]	24.0
Protein atoms	2902
Water Molecules	162
Unique Reflections	14915
R.m.s.d.[7]	
Bonds	0.010
Angles	1.18
MolProbity Score[8]	
Clash Score[8]	

[1]Statistics for highest resolution bin of reflections in parentheses.
[2]$R_{sym} = \Sigma_h\Sigma_j \mid I_{hj} - \langle I_h \rangle \mid / \Sigma_h\Sigma_j I_{hj}$, where I_{hj} is the intensity of observation j of reflection h and $\langle I_h \rangle$ is the mean intensity for multiply recorded reflections.
[3]Intensity signal-to-noise ratio.
[4]Completeness of the unique diffraction data.
[5]R-factor = $\Sigma_h \mid |F_o| - |F_c| \mid / \Sigma_h |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes for reflection h.
[6]R_{free} is calculated against a 10% random sampling of the reflections that were removed before structure refinement.
[7]Root mean square deviation of bond lengths and bond angles.
[8] ref 35

Table 3.2 Hydrogen Bonds between RBBP4 and BCL11A

Bcl11A	Atom	RBBP4	Atom	Distance
Arg 4	O	Lys 376	NZ	2.94
Arg 4	NH1	Glu 231	OE1	2.92
Lys 5	NZ	Glu126	OE2	2.75
		Asn 128	OD1	2.72
		Glu 179	OE1	2.75
Gln10	O	Asn397	ND2	3.05
His11	NE2	Glu41	OE2	2.69
Leu12	N	Glu41	O	2.72
Lys14	NZ	ALA39	O	2.97
Arg15	NH1	Glu41	OE2	2.82
Glu16	OE2	HIS38	NE2	2.82

Table 3.3 Intra-Bcl11a Interactions

Bcl11A	Atom	BCL11A	Atom
Gln10	NE2	PEPTIDE Lys7	O
Lys14	NH	PEPTIDE Leu11	O

Table 3.4 Hydrophobic Interactions between RBBP4 and BCL11A

Bcl11A	RBBP4 Residues
Arg4	Tyr 181 and Phe 323
Pro9	His 71

Chapter 4

Screening for Inhibitors of the RBBP4-Histone H3 Protein-Protein Interaction (PPI).

Abstract

Retinoblastoma binding protein 4 and 7 (RBBP4/7) are two histone chaperones intimately tied to epigenetic regulation within the cell. As such, RBBP4/7 function can be easily corrupted by oncogenes to promote tumorigenesis. RBBP4/7 contain WD40 repeats that fold into the characteristic seven-bladed β-propeller providing many surfaces to mediate interactions in its' associated complexes including the Polycomb Repressive Complex 2 (PRC2) and the Nucleosome Remodeling and Deacetylase complex (NuRD). The top face, or the canonical binding site, interacts with histone H3 and several oncogenic transcription factors, such as SALL4 and BCL11A, recruiting epigenetic complexes to their target genes. Given the significance of epigenetics in the development of diseases and cancer, targeting the canonical binding site of RBBP4/7 will help probe and interrogate the functional importance of this interface of PPIs in transcriptional regulation. This is true especially considering the newly defined role of BCL11A and the established role of PRC2 in triple-negative breast cancer, both of which interact with RBBP4. Herein chapter 4, we apply orthogonal screening techniques to identify potential inhibitors

of RBBP4/7. Through a high-throughput fluorescence polarization assay and a fragment-based thermal shift screen, we found several chemical scaffolds that are in the process of being validated and optimized for use either as a chemical probe or a drug-like molecule.

4.1 Introduction to High-Throughput Screening and Targeting Protein-Protein Interactions.

High-throughput screening (HTS), a technique vastly utilized by pharmaceutical companies for lead discovery in drug development, has started branching into academia in recent years. This is due in part to the increased interest in discovering new molecular targets of disease-states and the expansion in chemical compounds available. The rise in HTS begot development of robust assays as well as improvements in automation and miniaturization (i.e. decreases in the amount of reagents required and increases in sampling size per assay from 96-well plates to 384-well and even 1536-well).[1,2] Currently, HTS is one of the most productive and universally applied techniques in drug discovery.

Nevertheless, HTS is being scrutinized by many for its lack of perceived success and is even being held responsible for slowing productivity in the pharmaceutical industry.[3] Drawbacks of HTS include the high attrition rates, the demanding validation process, and the fact that hit compounds are not immediately impactful as drugs, but require a great deal of optimization.[3] These problems stem from the fact that lead compounds that come from HTS are only as good as the library screened. While initial HTS libraries focused on quantity rather than the quality of small molecules screened, that is not necessarily the case today.[1]

Improvements to libraries are dedicated to balancing novel chemotypes with ideal drug-like properties (i.e. those that adhere to Lipinski's rule of five) and thus, improve the starting point for drug design and development.[4,5]

Pharmaceutical companies concentrated HTS efforts on easily tractable, or druggable, targets with emphasis on those attributed to prominent diseases. Academic HTS endeavors, therefore, have little competition in their efforts against rare diseases or more challenging drug targets like protein-protein interactions (PPIs). PPIs are important for all biochemical pathways, many of which are dysregulated in cancer. As such, PPIs provide many opportunities for chemical intervention as there estimated to be over 300,000 PPIs within a cell.[6,7] PPIs were previously viewed as undruggable because of their shallow surfaces and interactions occurring over large areas.[7] In addition, PPIs were thought to rely on multiple weak interactions that collectively attributed to a higher binding affinity, which would understandably be difficult to interrupt with small molecule inhibitors.[7] For those reasons, targeting PPIs remains to be an underrepresented area in drug discovery.[6,7] However, recent studies have identified "hotspots" of PPIs, or the residues responsible for the majority of binding affinity. These "hotspots" are enriched for tyrosine, tryptophan and arginine residues that participate in a combination of hydrophobic, hydrogen bonding, π-π stacking, electrostatic, and cation-π interactions with its' natural ligand.[6,8] Analysis of "hotspots" and depth of pockets enhances viable target selection. Many "success stories" from HTS efforts against PPIs are due to these improvements in our understanding of potential targets and their interaction surfaces.

A few well-known examples of PPI inhibitors include those targeted against p53-MDM2 and the B-cell lymphoma-2 (Bcl-2) family of proteins. MDM2 is often overexpressed in cancers and impairs tumor-suppressor p53's function. HTS identified a class of cis-imidazoline analogs, thereafter referred to as Nutlins, capable of inhibiting this PPI, which resulted in anti-proliferative and apoptotic effects[9]. Subsequently, several of these inhibitors have entered early clinical trials.[10] Inhibitors against Bcl-xL-Bak PPI were discovered in a similar manner and are successful in prompting the release of pro-apoptotic Bak and committing the cell to death. These inhibitors (ABT-737, ABT-199, and WEHI-539) mimic the BH3 domain of Bak and prevent its' sequestration by Bcl-2 family members.[11-14] As Bcl-2 proteins have an important role in hematological malignancies, the inhibitors are being enthusiastically studied in these cases. Several other PPI inhibitors are gaining traction in both the laboratory and in clinical trials, spurring the development of this area of drug-discovery.

Furthermore, targeting PPIs offers hope for more specific inhibitors as traditional targets, such as enzymes, often share similar binding pockets to members of the same class. They may also provide different biological consequences from enzymatic inhibitors. For example, a hydrocarbon-stapled peptide that was able to structurally mimic the embryonic ectoderm development (EED) binding domain of enhancer of zeste homolog 2 (EZH2), referred to as SAH-EZH2, produced drastically different pharmacological results than an EZH2 enzymatic inhibitor, GSK126.[15] EZH2, the catalytic component of the polycomb repressive complex 2 (PRC2), is overexpressed and over-activated in many cancers

and thought to be an oncogenic driver of tumorigenesis.[16-18] Although GSK126 was more effective in reducing the PRC2 H3K27me3 mark, SAH-EZH2 had greater cytotoxic effect in cell lines characterized by the aberrant activity of EZH2.[15] These differences were attributed to the disruption of the EZH2-EED interaction, which resulted in the destabilization of EZH2 and a decrease in EZH2 protein levels. PRC2 and EZH2 may, therefore, have important functions outside of its' histone methyltransferase activity.[19] These results emphasize the importance of finding novel inhibitors for cancer therapeutics beyond the traditional enzymatic targets.

4.2 Background of RBBP4/7: its' known interactions and implications in disease.

Retinoblastoma binding proteins 4/7 (RBBP4/7) are two other non-enzymatic components of PRC2 that have yet to be explored as potential therapeutic targets against this epigenetic complex. Importantly, RBBP4/7 are not limited to roles in only PRC2 but are in fact subunits in histone deacetylase (HDAC) complexes as well. This links these proteins to two critical epigenetic mechanisms for regulating gene expression. Furthermore, previous studies have demonstrated that RBBP4 is important for maintaining pluripotency in stem cells, perpetuates k-Ras signaling, and increases the sensitivity of glioblastoma cells towards traditional chemotherapies.[20-22] RBBP7, however, may help regulate the epithelial to mesenchymal transition (EMT) supporting metastasis.[23,24] Studying these proteins, therefore, may provide novel cancer therapeutics.

RBBP4/7 consist of WD40-repeats that fold into a characteristic seven-bladed β-propeller. As such, they have several interaction surfaces and thereby

function as a protein-protein platform. Using their side surface, RBBP4/7 can interact with the first alpha helix of histone H4.[25] In epigenetic complexes, this site is reserved for SUZ12 of PRC2 or metastasis-associated protein 1 (MTA1) of the nucleosome remodeling and deacetylation (NuRD) co-repressor.[26-28] Histone H3, on the other hand, binds to the top face of the β-propeller, referred to as the canonical binding site of many WD40 family proteins. Recently, studies revealed that RBBP4/7 were unable to bind histone H3 in the presence of the activating H3K4me3 mark causing a decrease in PRC2 activity.[26] With PRC2 having significant functions in breast tumorigenesis and coupled with our results discussed in Chapter 2, we chose to pursue screening techniques for inhibitors of RBBP4/7. Small-molecule inhibitors of this interaction will help discern the importance of this PPI for RBBP4/7's function and whether RBBP4/7 are viable targets for the treatment of TNBC.[29]

While the goal of the work described in this chapter is to develop a drug-like molecule for the PPI between RBBP4/7 and Histone H3, the lead compounds identified may be able to serve as biochemical probes that would help answer the questions surrounding RBBP4/7's role in epigenetic complexes and disease-states. Specifically, whether RBBP4/7's interaction with histone H3 is necessary for nucleosome association and activity of PRC2, whether it is a valid drug target, or if inhibiting this PPI is too toxic.

4.3 Results and Discussion

4.3.1 Purification of RBBP4.

Notably, when we first expressed RBBP4 protein in bacterial cells using various tags including His-tag, maltose binding protein (MBP) or small ubiquitin-related modifier (SUMO), there was no change detected in the polarization of the histone H3 probe. Consequently, we switched to a higher eukaryotic system in insect cells using the Bac-to-Bac baculovirus expression system. In contrast to bacteria, insect cells offer various protein chaperones for proper folding as well as the opportunity for post-translational modifications, which may be necessary for certain proteins (a description of the process is in Figure 4.1). Once we had 'active' protein we were able to precede with the development and optimization of the competitive assay.

4.3.2 Development and Optimization of Fluorescence-Polarization (FP) based Competitive Binding Assay.

In order to gauge the chemical tractability of the RBBP4-Histone H3 PPI, we used the FP assay that was described previously in Chapter 3. Briefly, the histone H3 peptide (aa 1-21) was labeled at the C-terminus with fluorescein and served as the probe. The equilibrium dissociation constant (K_D) was determined by titrating RBBP4 protein at a fixed concentration of the 5-FAM H3 probe (20 nM). The K_D of the RBBP4-Histone H3 PPI was found to be 0.84 ± 0.08 μM and did not vary significantly from the K_D of the RBBP7-Histone H3 PPI, which was 1.20 ± 0.10 μM (Figure 4.2). This is consistent with the notion that inhibition of either histone chaperone, RBBP4 or RBBP7, will likely inhibit the other. As these proteins are highly similar, there is quite a bit of functional overlap between the two and

inhibition of either RBBP4 or RBBP7 alone may not provide a significant therapeutic response (as discussed in Chapter 2).

The principle of the FP assay for use in hit-lead identification is described in Figure 4.3. To improve the cost efficiency of HTS, it is essential to 1) determine the probe with the tightest binding affinity and 2) minimize the necessary amount of protein to achieve a sensitive assay with a reasonable dynamic range (ΔmP). The development of a successful competitive binding assay will aid in the investigation of small molecule inhibitors and natural product extracts in HTS. Not only will improving the binding affinity of the probe decrease the amount of protein required, but it will also improve the range of inhibitor potency that can be resolved. This is due to the fact that the K_D of the fluorescent probe serves as the limiting factor for the maximum K_i that can be determined by the assay.[30]

Therefore, we first sought to identify the minimal binding motif necessary for maintaining high-affinity binding and investigate if we could improve K_D. Subsequent truncations of the peptide from the C-terminus of the H3 peptide resulted in an increase in the competitive inhibition constants (K_i) (Table 4.1), with longer peptides, such as H3 aa 1-10, having better capability for displacement of the 5-FAM H3 probe. Further deletion of Lys-9 from the peptide, giving H3 1-8, decreased binding affinity by 30 times. Interestingly, modifications to the N-terminus of either the H3 1-10 or H3 1-9 peptide with an acetyl group diminished binding affinity to the extent where accurate K_i values could not be determined under our assay conditions. From the crystal structure of Nurf55 (a homolog of RBBP4/7) in complex with an H3 peptide, we can see that the α-amino group of H3

Ala1 forms a hydrogen bond with Nurf55 Asp252 (Figure 4.4). This could indicate that a free amine is necessary to help anchor the H3 peptide to the top surface of RBBP4. According to the aforementioned crystal structure, the side-chain of H3 Lys9 of is solvent exposed, but maintains hydrophobic contacts with on the surface of Nurf55 via Trp46. These results suggest that the minimal binding motif for RBBP4 recognition of H3 is ARTKQTARK.

Alanine scans from previous studies have shown that the critical residues that mediate the interaction of transcription factors with RBBP4 are RRKQXXP.[31,32] Histone H3 shares a similar sequence with R-KQXXK. Consistent with this data, mutations to the H3 sequence involving Arg3 or Lys4 impaired binding affinity (Table 4.2). If we were able to block these important interactions using a small molecule, it may prevent recruitment of epigenetic co-repressors to their targets and thus alter transcription. While the H3 peptide lies in an extended confirmation along the acidic top face of RBBP4, there is a small pocket where H3 Arg2 sits, that may be amenable to interference with an inhibitor (Figure 4.4).

Once we determined the optimal length of the H3 probe to be aa 1-21, as it demonstrated the highest K_D, we next focused on the minimal concentration of the fluorescent ligand that could provide a stable signal. This is based on previous literature findings whereby higher concentrations of probe, one that is above the true K_D value, could result in higher apparent binding affinities. To determine the minimal level of the fluorophore required, we added increasing concentrations of the ligand (from 0.5 nM to 1 μM) to the assay buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1% glycerol) at a final volume of 20 μL and FP was analyzed. FP signal

was stable at concentrations higher than 20 nM (Figure 4.5) and was, therefore, chosen as the optimal concentration for the H3 5-FAM probe. Importantly, higher concentrations of the fluorescent probe can potentially overcome interference of compounds with weak fluorescent capability.

Next, we sought to determine the ideal amount of protein that would maximize the sensitivity of the assay and improve the dynamic range, but also remain in the linear range of the saturation-binding curve (Figure 3.2). As mentioned in Chapter 3, decreasing the salt concentration from 100 mM to 50 mM improved the binding affinity from 1.26 μM to 0.35 μM. For HTS, we chose lower salt conditions to minimize total RBBP4 protein required. Under these conditions, using RBBP4 at about three times the K_D (1.2 μM), we were able to reach a desirable dynamic range of about 65 mP with a Z' of > 0.75, which is well above the value necessary for a quality HTS (Figure 4.6).

Finally, as most compounds in a high-throughput screening library are stored in DMSO, it was imperative to determine the assay's tolerance to DMSO. Increasing concentrations of DMSO (from 0.1% to 50% final assay volume) were incubated in solutions containing 50 μM RBBP4 and 20 nM H3 5-FAM probe (Figure 4.5). Stable mP signals were seen up to 12.5% DMSO indicating the traditional 4-8% DMSO used in HTS would be well tolerated.

4.3.3 HTS Screen.

To begin screening, we ran one plate under the conditions previously mentioned with half of the plate dictating positive controls (only probe and DMSO) and the other half as negative controls (probe, protein, and DMSO) (Figure 4.6). The

plate was run at various time points to assess the stability and we found that after 2 hours the signal decreased. Thus, all plates assayed were run after 30 minutes of incubation, but before the 120-minute mark. The resultant mP signals from the control plate were used to determine percent inhibition of compounds screened.

Initial screens were performed against the LOPAC and Prestwick pilot libraries in which 0.2 μL of the small molecule in DMSO was pin tooled into a corresponding well of the FP plate which contained RBBP4 in 10 μL assay buffer. The last two columns on each plate were reserved for positive controls. All compounds tested were first screened at one concentration (20 μM) and full dose-response curves would later be performed for confirmed "hits." After addition of the small molecules, 10 μL of tracer was added to each well of the plate and fluorescence polarization was measured on a PHERAstar plate reader.

While hit rate was relatively low (<1%), we proceeded to screen against natural product extracts (NPE) collected by David Sherman's laboratory from sediments, sponges, and cyanobacteria from around the world. Natural products (NP) offer a diverse source of chemical and structural entities that lend themselves favorably to medicinal purposes. Around a third of drugs approved by the US Food and Drug Administration (FDA) in the past thirty years are natural products or NP-derived.[33] In particular, NPs are useful in disrupting protein-protein interactions (PPIs), like that between RBBP4 and Histone H3. As they usually have larger and more complex structures, natural products may be better suited to inhibit PPIs as they occur over larger surfaces with little surface topology. A review of several efficacious anti-cancer natural products can be seen in Figure 4.7.

However, despite their success, NPs are often difficult to synthesize, thereby leading to low yields and precluding structural modifications (and structure-activity relationship studies). In addition, with our screen against extracts, which contain multiple NPs, it will take time to deconvolute and characterize the compound responsible for the efficacy. Thus, we also chose to screen small molecules from the Maybridge library as well as half of the ChemDiv library. All of the data from the pilot and primary screens is in Figure 4.6B.

Compounds were defined as "hits" if they were greater than 3 standard deviations from the mean (of the negative control) and if fluorescence was less than 2 times the total fluorescence of controls. Moreover, any compounds that were flagged for reactive structures or PAINS were removed. All "hits" were confirmed in triplicate (and those that had greater than 25% inhibition are shown in Figure 4.8).

We are currently in the process of selecting which compounds to move forward with for dose-response curves, counter-screens (to identify hits selective against the RBBP4-Histone H3 PPI), and finally, lead optimization. The first step in this process will be to order fresh sample of the compounds as those in the HTS library have been stored for a long period of time and may be subject to degradation and/or modification. Interestingly, dipyridamole and several naphthalimide analogs are among the confirmed hits. Both of these compounds are in the process of being (or have been) used as anti-cancer therapies.[34,35] 1,8-Naphthalimide derivatives are investigated as potential poly(ADP-ribose) polymerase (PARP) inhibitors, where they prevent DNA repair and force cancer cells towards apoptosis. Dipyridamole, on the other hand, is an FDA-approved medication to prevent clot

formation. It has been used historically in combination therapies for cancer treatment and has been implicated in preventing TNBC progression.[36] Nevertheless, structural studies and secondary screening will be necessary to validate these compounds.

4.3.4 Parallel Screening with Fragment-based Libraries.

To maximize our chances of finding an inhibitor of this PPI, we performed a parallel screen with a thermal-shift assay against a fragment-based library. Differences between HTS and fragment-based drug discovery (FBDD) are summarized in Figure 4.9.[37,38] The thermal-shift assay measures changes in protein stability, as gauged by a change in melting temperature of the protein (ΔTm). As temperature increases, the protein unfolds and exposes a hydrophobic core that can be bound by a fluorescent marker. If a fragment binds to RBBP4, it would be expected to increase the stability of RBBP4 and thus, the Tm.

To begin our screen, we first needed to determine optimal conditions of dye and protein concentration as well as whether DMSO and/or detergent were tolerated. Various protein and dye concentrations were tested in replicates of eight, and we found that the best conditions were 2 μM RBBP4 and 2x dye in phosphate buffered saline (PBS). While DMSO was tolerated up to 5% under these conditions, Triton and Tween detergents decreased the signal-to-noise ratio and were therefore excluded.

Once optimized, we screened around 1400 fragments from ChemDiv and Life Science libraries. All compounds were tested in quadruplicates at a concentration of 500 μM and fluorescence was measured at regular intervals. Those that resulted in

Tm shifts that were greater than 2 times the standard deviation of the controls (RBBP4 and DMSO in PBS) were selected for confirmation. Of the 100 compounds selected from the initial screen, 29 fragments were confirmed to stabilize RBBP4 (Figure 4.10 and 4.11).

With thermal shift assays, the Tm shift indicates binding and stabilization, but not where the interactions occur. To distinguish between fragments that bind to the top face of RBBP4, we ordered dry powders of the 29 hits, resuspended in DMSO to make fresh stocks, and tested all in the FP assay. Only one of the fragments was capable of competing with the H3 tracer for binding to RBBP4, and it had a K_i of 94 μM (Figure 4.12). Using that core structure, we performed a substructure search through ChemDiv and Chembridge for "catalog-to-analog" SAR. All analogs were tested in FP at a single concentration (1 mM) and those that exhibited inhibitory activity were selected for full dose response curves (results summarized in Table 4.3).

Three of the aforementioned analogs were able to bind RBBP4, but only one reduced the K_i by almost 5-fold to 21 μM. This fragment (cat. no. 14133809, Chembridge), referred to as compound 58, had a pyrazolopyrimidine core in comparison to the triazolopyridine of the initial hit. In addition, 58 had a propyl group at position 5 and a triazole ring connected via an amine to position 7 (Figure 4.13). To determine the binding mode for compound 58, we turned to in silico methods. Using the crystal structure of RBBP4 in complex with FOG1 (PDB: 2XU7), we removed the native ligand and performed molecular simulation and docking studies to assess possible binding conformations of compound 58 (Figure 4.14).

Overlay of the histone H3 peptide shows that the heterocyclic ring of 58 occupies the same space as Arg2 of H3, i.e. 58 is sandwiched between tyrosine and phenylalanine residues within RBBP4. Furthermore, the triazole moiety appears to hold the same buried pocket as Ala1 and may help anchor the fragment to RBBP4.

Comparing the structures of the analogs garners some insight into the druggability of the pocket and potential SAR. Bulky substituents off of the heterocyclic ring, for instance, cannot fit into the small pocket on the top face of RBBP4's β-propeller. Changing the electrostatics of the compound also removes potential for binding. To better understand and probe the binding pocket, x-ray structures of RBBP4 in complex with compound 58 will be necessary and will also promote the structure-based design of more potent inhibitors.

4.3.5 Pursuit of other RBBP4 Inhibitors.

From the initial 29 compounds discovered from our thermal-shift screen, only one of those was able to compete with the histone H3 peptide for binding to RBBP4. This leaves 28 other compounds that may be able to bind to RBBP4 and inhibit its' function in another way. To determine which of these fragments are involved in true binding events, we developed a label-free BioLayer Interferometry (BLI) assay using the Octet RED platform. This assay provides valuable information on the specificity, affinity, and kinetics of biomolecular complex formation in real-time. BLI is an optical analytical technique that measures the interference patterns between the biocompatible surface layer and the internal reference layer of a biosensor. When a fragment or small molecule binds to RBBP4, which is immobilized on the surface of the biosensor, this creates a change in the

interference patterns which is represented as a change in wavelength, or spectral shift (Δnm).

RBBP4 was biotinylated and then immobilized on Super Streptavidin (SSA) biosensors. We optimized loading concentrations of RBBP4 to achieve maximal saturation response of 11-12 nm (Figure 4.15A). Biotinylated streptavidin was also loaded onto SSA biosensors as an inactive negative control. Following immobilization, a stable base line was established and then association-dissociation curves cycles were performed with the fragments starting at a concentration of 31.25 μM to 500 μM (Figure 4.15B).

Initial testing indicated 5 compounds that bind to RBBP4 with weak affinity (K_D > 100 μM), whereas one of the fragments, fragment 18, had moderate affinity (K_D < 100 μM) (Figure 4.15C). The potential binding could occur on the bottom face of RBBP4's β-propeller, the site opposite of the histone H3 interaction (Figure 4.16). This pocket is much larger, but little is known about the interactions that occur at this face. Future thermal shift screens coupled with NMR assays could help uncover compounds that bind to this pocket. These potential inhibitors might also disrupt RBBP4's function by interrupting the proper formation of epigenetic complexes.

4.4 Conclusions. Herein this Chapter, we described orthogonal techniques to find potential inhibitors of RBBP4/7. As histone chaperones, RBBP4/7 are thought to be crucial to the recruitment of multi-subunit complexes to nucleosomes and potentially their target genes. As such, they may easily be corrupted in cancer cells where its' binding partners, such as the oncogenic transcription factor BCL11A, are overexpressed leading to alterations in the epigenetic landscape. In addition,

RBBP4/7 have important roles in maintaining the integrity of the epigenetic complexes and integrating signals, such as the histone code.

Impending validation studies will separate actual hits from compounds that are false positives. With any HTS there is potential for false positives that could come from pan-assay interference compounds (PAINS) or compounding effects from differing purification schemes.[39] While the canonical binding site of RBBP4/7 is relatively small, it does not necessarily preclude small-molecule inhibitors. It is to be expected that small changes to our initial hit from fragment screening could result in no significant binding, as the affinity was not high to begin with. Once a better inhibitor with tighter affinity is discovered analog screening will be more successful. Moreover, deconvolution of the hits from natural product extracts will expand the chemical scaffolds that we can pursue as lead compounds.

If we find an inhibitor of the PPI between RBBP4 and histone H3, it will help answer several biological questions surrounding RBBP4/7's function in PRC2 and potentially other epigenetic complexes. Studies resulting from treatment of TNBC cell lines with a potential inhibitor are expected to differ from our knockdown studies as depletion of RBBP4/7 may have indirect effects on many epigenetic complexes. Theoretically the inhibitor will only affect those machineries that are dependent on the canonical binding site of RBBP4/7. Through the small-molecule intervention of this PPI, we may also study its' effect on the recruitment of epigenetic complexes to particular genomic loci.

Perhaps an interesting outcome of these studies will be compounds that can bind to RBBP4/7 on the bottom face of its' β-propeller. Currently, there are no

known interaction partners at this site. Recently, the oncogenic factor c-Myc was found to bind to a similar side of the WD40 protein, WD repeat-containing protein 5 (WDR5). In the studied capacity WDR5 was acting as an essential cofactor in Myc-driven tumorigenesis.[40,41] Targeting, or rather exploring, this site of RBBP4/7 will provide not only a better understanding of the architecture and recruitment of co-repressor complexes but also a potential alternative for therapeutic intervention.

4.5 Materials and Methods

Materials. All compounds were purchased in powder form from ChemDiv, ChemBridge, or eMolecules, Inc.

Protein Expression and Purification. Initially, full length RBBP4 was expressed in Escherichia coli BL21 (DE3) cells from pMCSG9 plasmids. Selected colonies were grown in terrific broth (TB) medium supplemented with ampicillin (50 mg/L) at 37°C with shaking until OD_{600} was between 0.6 and 0.8. Protein expression was induced at 18°C overnight by the addition of IPTG (final concentration 0.2 mM). The following morning cells were harvested by centrifugation (4000xg, 10 min, 4°C). Collected cell pellets were suspended in lysis buffer and subject to sonication. All proteins were purified by His-tag on Ni-NTA resin (Qiagen). To improve yields, RBBP4 was expressed from pET28A-SUMO as a small ubiquitin-related modifier (SUMO) fusion. This His-SUMO fusion of RBBP4 was expressed in BL21 DE3 pLysS codon (+) and purified in a similar manner as listed above.

Cloning, expression and purification of RBBP4 and RBBP7 were performed similarly to previous reports. Briefly, full length RBBP4 (aa 1-425) and RBBP7 (aa 1-425) were cloned into a pFastBac HT-A or pFastBac HT-B vector with both an N-terminal His_6 tag and a tobacco etch virus (TEV) cleavage site. Generation of recombinant bacmid (in DH10Bac cells) and virus (using Cellfectin® II in Sf9 cells) was performed using the Bac-to-Bac (Invitrogen) expression system and according to manufacturer's recommendations. Protein was expressed in Tn5 (High Five™; Invitrogen) cells using P3 virus. Infected cells were collected via centrifugation,

washed with ice cold PBS twice and the pellet was lysed immediately or flash frozen in liquid nitrogen for later use. All purification steps were performed at 4°C unless otherwise noted. Infected cells were lysed using lysis buffer (50 mM Tris (ph 8), 500 mM NaCl , 0.1% NP40, 5 mM imidazole,10% glycerol) with protease inhibitors and nuclease (Pierce™ Thermo Scientific) added right before use. Lysis was further aided by sonication. The lysate was clarified by centrifugation for 40 min at 12,000 rpm and supernatant was added to nickel-nitrilotriacetic acid (Ni-NTA) resin (Qiagen) for 2 hours on a shaker. Resin was collected and washed with high-salt buffer (50 mM Tris (ph 8), 500 mM NaCl , 10 mM imidazole,10% glycerol) three times. Resin was then washed with low-salt buffer (50 mM Tris (ph 8), 150 mM NaCl , 10 mM imidazole,10% glycerol) to equilibrate before elution. RBBP4 was eluted using the low-salt buffer with 250 mM imidazole. Fractions containing RBBP4 were collected, pooled, and concentrated using an Amicon® Ultra-4 Centrifugal Filter (EMD Millipore).

Peptide purification. All peptides were synthesized manually using Fmoc chemistry and Rink Amide resin as the solid support. The exception is the 5-FAM H3 (1-21) probe which was purchased from Anaspec (cat. No. AS-63824). DIC/HOAt were used as the coupling reagent. After the final amino acid was added, the peptide was cleaved from the resin using a cleavage cocktail of trifluoroacetic acid, triisopropylsilane, H_2O and thioanisole. The resulting solution was evaporated and the crude peptide was precipitated in diethyl ether. Peptides were purified via RP-HPLC (Waters, Sunfire Prep C18, 19 mm x 150 mm, 5 μm) and confirmed by electrospray ionization mass spectroscopy (ESI-MS).

Fluorescence Polarization Assay. All fluorescence polarization experiments were conducted in 384-well, black, low volume, round-bottom plates (Corning) using a BioTek Synergy 2 plate reader (Winooski, VT). To each well, was added increasing amounts of protein and the 5-carboxyfluorescein (5-Fam) labeled Histone H3 N-terminal probe/tracer (20 nM) to a final volume of 20 uL in the assay buffer (50 mM Tris-HCl, pH=7.4, 100 mM NaCl, 0.1% glycerol unless otherwise indicated). The plate was allowed to incubate at room temperature for 5 min to reach equilibrium. The polarization values (mP) were measured at an excitation wavelength at 485 nm and an emission wavelength at 528 nm. An equilibrium-binding isotherm was constructed by plotting the mP values as a function of the protein concentration at a fixed concentration of tracer (20 nM). For competitive binding experiments, RBBP4 was added to the well at a concentration around three times the K_D with 5FAM probe (20 nM) and increasing concentrations of nonlabeled peptide (in 50 mM Tris-HCl, pH=7.4, 100 mM NaCl, 0.1% glycerol unless otherwise indicated). Inhibitory constants (K_i) were calculated using a previously determined equation[42]. All experimental data were analyzed using Prism 7.0 software (Graphpad Software, San Diego, CA) and WinNonlin (version 5.3).

High Throughput Screening. Screening was performed at the Center of Chemical Genomics (CCG) at the University of Michigan. Libraries screened included the Prestwick and LOPAC pilot libraries, as well as natural product extracts, MayBridge 24K, and part of the ChemDiv library for a total of around 70,000 compounds. Compounds screened were pintooled (200 nL) into a 384-well plate using a Biomek FX dual head (Beckman) for a final concentration of 20 μM. An initial plate was run

with 50/50 positive (probe only) and negative controls (probe, protein and DMSO) to calculate Z' factor (equation below)

$$Z' = 1 - \frac{3\left(\sigma_p + \sigma_n\right)}{\left|\mu_p - \mu_n\right|}$$

σ = Standard deviation of the positive or negative controls

μ = Mean of the positive or negative control

Z' = 1 is an ideal assay, whereas Z' > 0.5 is deemed an excellent assay. After primary screening (n=1), compounds or natural products that were at least 3 standard deviations above the negative control were selected for confirmation screening in triplicate.

Thermal Shift Assay. RBBP4 (2 μM) and Protein Thermal Shift™ (Applied Biosystems) dye (2X final concentration) were mixed in PBS in a MicroAmp Fast Optical 384-Well plate. 1 μL of fragment from ChemDiv or Life Science libraries was added to the respective wells bringing the final volume to 20 μL. Following addition of the fragment, the microplates were sealed with an adhesive optical clear seal (MicroAmp Optical Adhesive Film) and centrifuged. A melt curve experiment was measured on a QuantStudio™ 7 Flex Real-Time PCR System (Thermo Scientific) and fluorescence was measured at regular intervals over a temperature range from 25 to 99°C. Controls included no-protein controls (PBS, dye and DMSO) and negative reference (RBBP4, dye, DMSO in PBS). The temperature at which 50% of the protein is unfolded represents the apparent melting temperature (Tm). Raw data was analyzed using Protein Thermal Shift™ software, which produced Boltzmann-

derived Tm values. Fragments that were greater than 2 standard deviations from the reference wells were selected for further evaluation.

BLI/Octet RED. Purified recombinant RBBP4 was biotinylated using the Thermo EZ-Link long-chain biotinylation reagent. To avoid over-biotinylation, protein and biotinylation reagent were mixed at a 1:1 molar ratio in PBS at 4°C for 2 hours. Following the completion of the reaction, the mixture was dialyzed using FisherSci 10,000 MWCO dialysis cassettes to remove unreacted biotin. All assays were run in 50 mM Tris, pH 7.5, 50 mM NaCl, 0.005% Tween-20 and 5% DMSO on an OctetRED 96 instrument (PALL/ForteBio). Biotinylated RBBP4 (50 μg/mL) was then immobilized on Super Streptavidin (SSA) biosensors (ForteBio) and saturation (11-12 nm) was reached within 15 minutes. As an inactive control, biotinylated blocked streptavidin sensors were prepared according to manufacturers' instructions. Following loading of the sensors, the probes were incubated in buffer for 10 min to establish a stable baseline and wash away unbound protein. Association-dissociation cycles were then performed with the increasing concentrations of compounds. Default times for the association-dissociation cycles were 1 min for each. Data was analyzed using the software provided by the manufacturer with both buffer only and inactive protein references subtracted. 1:1 binding models were applied to generate kinetic data.

Docking Studies. Docking of initial fragment hit from FP screening, T482-1661, was performed using GOLD[43,44] with goldscore and chemscore scoring functions. Crystal structure of RbBP4 (PDB ID: 2XU7) was downloaded and prepared for

docking using MOE. The carbon atom of the guanidinium group of Arg2 from the cocrystalized H3 peptide was used as the origin for docking, with default settings for docking runs with both the scoring functions. The top 10 poses were saved from each run for further analysis. Each pose obtained was rescored using additional scoring functions; MScore[45], XScore[46] and DrugScore[47]. Similar process was followed for docking of other hits identified, and analogs. Analogs after docking and rescoring were clustered based on compounds that were ranked among the top 50 for two or more scoring functions, to purchase for testing.

Acknowledgements: This research used resources of University of Michigan's Center for Chemical Genomics (CCG). Screening reported in this chapter was supported by the National Cancer Institute of the National Institutes of Health under award number P30CA046592. Thermal shift screening and Octet Red was performed using equipment from Shaomeng Wang's Laboratory at the University of Michigan with help from Dr. Liu Liu and Dr. Xu Ran.

Contributions: Rebecca A. Reed expressed and purified RBBP4/7. SUMO-RBBP4 construct was obtained from Shirley Lee from Yali Dou's Laboratory. Rebecca Reed synthesized, purified and characterized peptides, and performed all screening and analysis. Garrett Johnson assisted with OctetRed assay development. Dr. Denzil Bernard performed all docking studies.

References:

1 Mayr, L. M. & Fuerst, P. The future of high-throughput screening. *J Biomol Screen* **13**, 443-448, doi:10.1177/1087057108319644 (2008).

2 Szymanski, P., Markowicz, M. & Mikiciuk-Olasik, E. Adaptation of high-throughput screening in drug discovery-toxicological screening tests. *Int J Mol Sci* **13**, 427-452, doi:10.3390/ijms13010427 (2012).

3 Macarron, R. *et al.* Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **10**, 188-195, doi:10.1038/nrd3368 (2011).

4 Rees, S., Gribbon, P., Birmingham, K., Janzen, W. P. & Pairaudeau, G. Towards a hit for every target. *Nat Rev Drug Discov* **15**, 1-2, doi:10.1038/nrd.2015.19 (2016).

5 Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* **1**, 337-341, doi:10.1016/j.ddtec.2004.11.007 (2004).

6 Sheng, C., Dong, G., Miao, Z., Zhang, W. & Wang, W. State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. *Chem Soc Rev* **44**, 8238-8259, doi:10.1039/c5cs00252d (2015).

7 Nero, T. L., Morton, C. J., Holien, J. K., Wielens, J. & Parker, M. W. Oncogenic protein interfaces: small molecules, big challenges. *Nat Rev Cancer* **14**, 248-262, doi:10.1038/nrc3690 (2014).

8 Guo, W., Wisniewski, J. A. & Ji, H. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett* **24**, 2546-2554, doi:10.1016/j.bmcl.2014.03.095 (2014).

9 Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844-848, doi:10.1126/science.1092472 (2004).

10 Burgess, A. *et al.* Clinical Overview of MDM2/X-Targeted Therapies. *Front Oncol* **6**, 7, doi:10.3389/fonc.2016.00007 (2016).

11 Cang, S., Iragavarapu, C., Savooji, J., Song, Y. & Liu, D. ABT-199 (venetoclax) and BCL-2 inhibitors in clinical development. *J Hematol Oncol* **8**, 129, doi:10.1186/s13045-015-0224-3 (2015).

12 Chen, L. *et al.* Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Mol Cell* **17**, 393-403, doi:10.1016/j.molcel.2004.12.030 (2005).

13 Degterev, A. *et al.* Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat Cell Biol* **3**, 173-182 (2001).

14 Lessene, G. *et al.* Structure-guided design of a selective BCL-X(L) inhibitor. *Nat Chem Biol* **9**, 390-397, doi:10.1038/nchembio.1246 (2013).

15 Kim, W. *et al.* Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. *Nat Chem Biol* **9**, 643-650, doi:10.1038/nchembio.1331 (2013).

16 Gonzalez, M. E. *et al.* Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. *Oncogene* **28**, 843-853, doi:10.1038/onc.2008.433 (2009).

17 Kleer, C. G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A* **100**, 11606-11611, doi:10.1073/pnas.1933744100 (2003).
18 Mu, Z. *et al.* EZH2 knockdown suppresses the growth and invasion of human inflammatory breast cancer cells. *J Exp Clin Cancer Res* **32**, 70, doi:10.1186/1756-9966-32-70 (2013).
19 Kim, K. H. *et al.* SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat Med* **21**, 1491-1496, doi:10.1038/nm.3968 (2015).
20 Kitange, G. J. *et al.* Retinoblastoma Binding Protein 4 Modulates Temozolomide Sensitivity in Glioblastoma by Regulating DNA Repair Proteins. *Cell Rep* **14**, 2587-2598, doi:10.1016/j.celrep.2016.02.045 (2016).
21 O'Connor, M. D. *et al.* Retinoblastoma-binding proteins 4 and 9 are important for human pluripotent stem cell maintenance. *Exp Hematol* **39**, 866-879 e861, doi:10.1016/j.exphem.2011.05.008 (2011).
22 Scuto, A. *et al.* RbAp48 regulates cytoskeletal organization and morphology by increasing K-Ras activity and signaling through mitogen-activated protein kinase. *Cancer Res* **67**, 10317-10324, doi:10.1158/0008-5472.CAN-06-3313 (2007).
23 Fu, J. *et al.* The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Res* **21**, 275-289, doi:10.1038/cr.2010.118 (2011).
24 Li, G. C. & Wang, Z. Y. Constitutive expression of RbAp46 induces epithelial-mesenchymal transition in mammary epithelial cells. *Anticancer Res* **26**, 3555-3560 (2006).
25 Murzina, N. V. *et al.* Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. *Structure* **16**, 1077-1085, doi:10.1016/j.str.2008.05.006 (2008).
26 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
27 Alqarni, S. S. *et al.* Insight into the architecture of the NuRD complex: structure of the RbAp48-MTA1 subcomplex. *J Biol Chem* **289**, 21844-21855, doi:10.1074/jbc.M114.558940 (2014).
28 Schmidberger, J. W. *et al.* The MTA1 subunit of the nucleosome remodeling and deacetylase complex can recruit two copies of RBBP4/7. *Protein Sci* **25**, 1472-1482, doi:10.1002/pro.2943 (2016).
29 Jene-Sanz, A. *et al.* Expression of polycomb targets predicts breast cancer prognosis. *Mol Cell Biol* **33**, 3951-3961, doi:10.1128/MCB.00426-13 (2013).
30 Huang, X. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J Biomol Screen* **8**, 34-38, doi:10.1177/1087057102239666 (2003).
31 Liu, Z. *et al.* Structural basis of plant homeodomain finger 6 (PHF6) recognition by the retinoblastoma binding protein 4 (RBBP4) component of the nucleosome remodeling and deacetylase (NuRD) complex. *J Biol Chem* **290**, 6630-6638, doi:10.1074/jbc.M114.610196 (2015).

32 Lauberth, S. M. & Rauchman, M. A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex. *J Biol Chem* **281**, 23922-23931, doi:10.1074/jbc.M513461200 (2006).
33 Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* **14**, 111-129, doi:10.1038/nrd4510 (2015).
34 Munoz-Gamez, J. A. *et al.* PARP inhibition sensitizes p53-deficient breast cancer cells to doxorubicin-induced apoptosis. *Biochem J* **386**, 119-125, doi:10.1042/BJ20040776 (2005).
35 Wang, C., Schwab, L. P., Fan, M., Seagroves, T. N. & Buolamwini, J. K. Chemoprevention activity of dipyridamole in the MMTV-PyMT transgenic mouse model of breast cancer. *Cancer Prev Res (Phila)* **6**, 437-447, doi:10.1158/1940-6207.CAPR-12-0345 (2013).
36 Spano, D. *et al.* Dipyridamole prevents triple-negative breast-cancer progression. *Clin Exp Metastasis* **30**, 47-68, doi:10.1007/s10585-012-9506-0 (2013).
37 Mashalidis, E. H., Sledz, P., Lang, S. & Abell, C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* **8**, 2309-2324, doi:10.1038/nprot.2013.130 (2013).
38 Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat Chem* **1**, 187-192, doi:10.1038/nchem.217 (2009).
39 Baell, J. & Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* **513**, 481-483, doi:10.1038/513481a (2014).
40 Sun, Y. *et al.* WDR5 Supports an N-Myc Transcriptional Complex That Drives a Protumorigenic Gene Expression Signature in Neuroblastoma. *Cancer Res* **75**, 5143-5154, doi:10.1158/0008-5472.CAN-15-0423 (2015).
41 Thomas, L. R., Foshage, A. M., Weissmiller, A. M. & Tansey, W. P. The MYC-WDR5 Nexus and Cancer. *Cancer Res* **75**, 4012-4015, doi:10.1158/0008-5472.CAN-15-1216 (2015).
42 Nikolovska-Coleska, Z. *et al.* Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem* **332**, 261-273, doi:10.1016/j.ab.2004.05.055 (2004).
43 Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **267**, 727-748, doi:10.1006/jmbi.1996.0897 (1997).
44 Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609-623, doi:10.1002/prot.10465 (2003).
45 Yang, C. Y., Wang, R. & Wang, S. M-score: a knowledge-based potential scoring function accounting for protein atom mobility. *J Med Chem* **49**, 5903-5911, doi:10.1021/jm050043w (2006).
46 Wang, R., Lai, L. & Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **16**, 11-26 (2002).

47 Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **295**, 337-356, doi:10.1006/jmbi.1999.3371 (2000).
48 Migliori, V. *et al.* Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol* **19**, 136-144, doi:10.1038/nsmb.2209 (2012).
49 Yanyan, L. *et al.* Sulforaphane, a dietary component of broccoli/broccoli sprouts, inhibits breast cancer stem cells. *Clinical cancer research : an official journal of the American Association for Cancer Research* **16**, 2580-2590, doi:10.1158/1078-0432.ccr-09-2937 (2010).
50 Slichenmyer, W. J. & Von Hoff, D. D. Taxol: a new and effective anti-cancer drug. *Anticancer Drugs* **2**, 519-530 (1991).
51 Yonekura, S. *et al.* Effects of the HIF1 inhibitor, echinomycin, on growth and NOTCH signalling in leukaemia cells. *Anticancer Res* **33**, 3099-3103 (2013).
52 Liu, Y. Q. *et al.* Perspectives on biologically active camptothecin derivatives. *Med Res Rev* **35**, 753-789, doi:10.1002/med.21342 (2015).

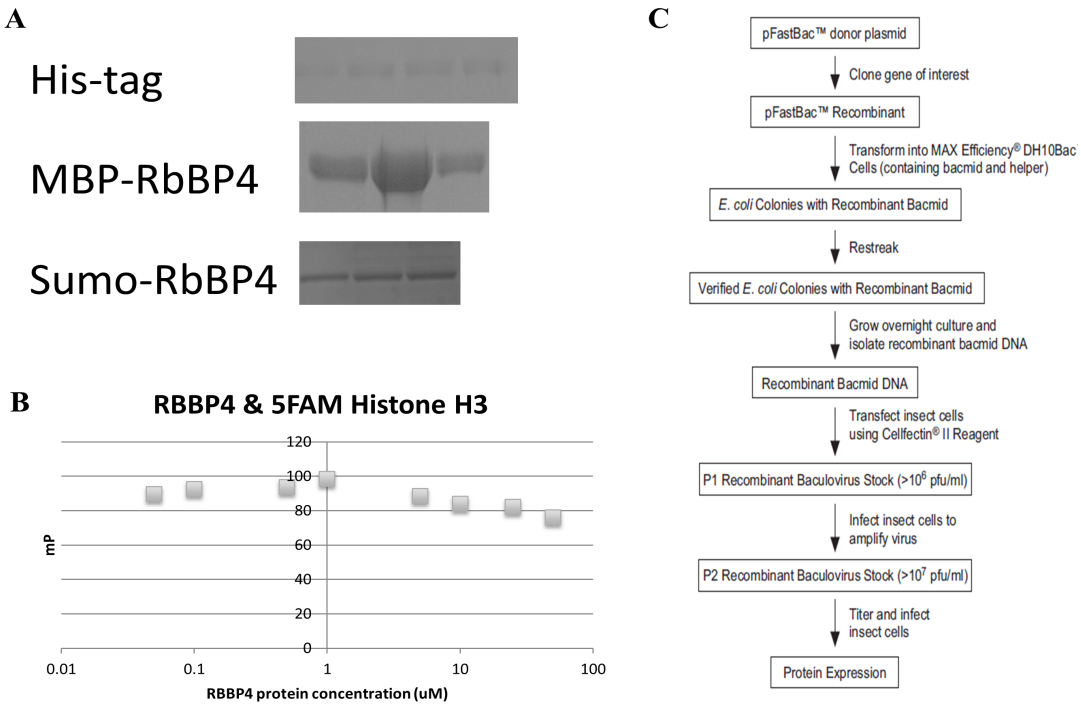
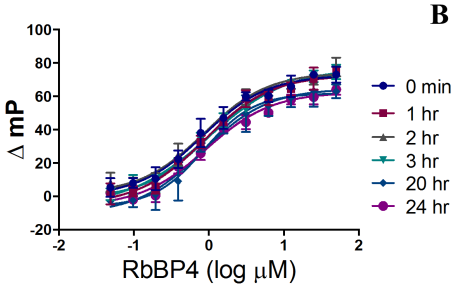


Figure 4.1 Schematic of Protein Purification.
(A) RBBP4 fusion constructs were constructed with various purification and solubility tags and gel pictures show purified RBBP4 expression levels. (B) All RBBP4 protein purified from bacteria did not produce any change in mP signal when incubated with a 5-FAM H3 peptide. (C) Schematic of protein expression and expression using the Bac-to-Bac expression system taken from Invitrogen Life Technologies.

A



B

	Kd (μM)
RBBP4	0.84 ± 0.08
RBBP7	1.20 ± 0.10

Figure 4.2 Binding curve of 5FAM-Histone H3 to RBBP4 and RBBP7.
(A) Full binding isotherm of RBBP4 bound to the H3 derived 21 amino acid 5FAM probe indicating the interaction is stable over 24 hours. (B) Kd values for the association between RBBP4 and RBBP7 to the H3 probe.

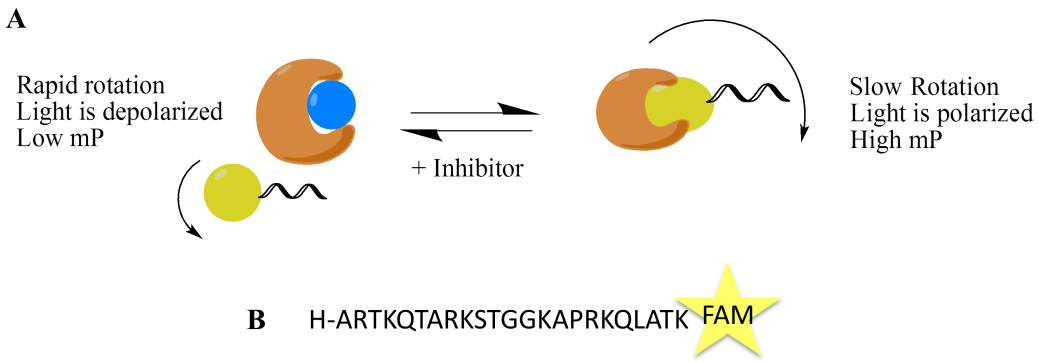
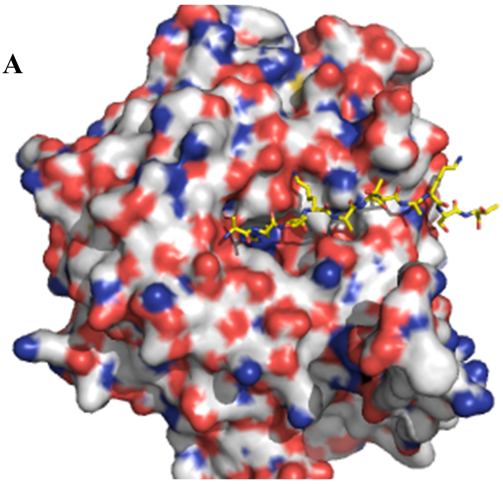
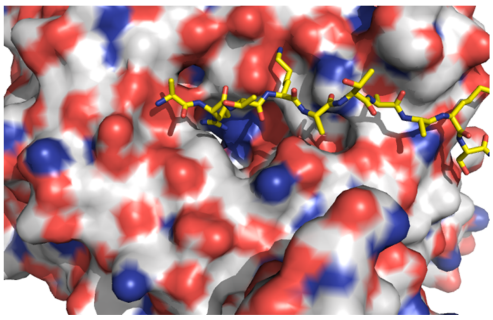


Figure 4.3 Principles of the Fluorescence Polarization (FP) assay.
(A)When small molecular weight probe is unbound to protein, it rotates rapidly and light remains depolarized. However, when probe interacts with a larger molecular weight protein, rotation will be slowed resulting in polarization of the light and a higher mP signal. If an inhibitor competitively blocks association of the probe and protein then we would expect a low mP signal resulting from rapid rotation of the free dye/probe. (B) Sequence of the histone H3 probe used in HTS screening.

A





B

Histone H3	RBBP4 (Nurf55)
Ala-1	Asp-252, Asn-281,Glu-323, H_2O
Arg-2	Phe-325, Glu-235, Tyr-185
Lys-4	Glu-183, Asn-132, Glu-130
Lys-9	Trp-46

C

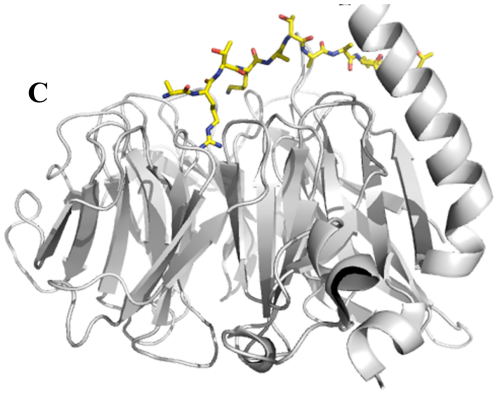


Figure 4.4 Crystal Structure of H3 peptide bound to RBBP4.
Ref. 26. (A) H3 (1-12) shown in yellow bound to the top face of the RBBP4 β-propeller. (B) A table of the important interactions between the H3 peptide and RBBP4. (C) Side view of the H3 peptide bound to RBBP4.

Table 4.1: Truncation study.

Peptide	K_i (μM)
Ac H3 1-10	> 200
H3 1-10	2.2
Ac H3 1-9	> 200
H3 1-9	2.6
H3 1-8	63
H3 1-7	193
H3 1-6	176
H3 1-4	143
RTK (H3 2-4)	> 200

Table 4.2: Modifications to Lys4 or Arg5 of the H3 peptide and the effect on Ki.

ART**X**QTARKST

Peptide	K_i (μM)
K	2.2
R	38.6
M	>200
Kme1, me2, me3, Ac	>200

A**X**TKQTARKST

Peptide	K_i (μM)
R	2.2
Cit	10.4
F(Nh2)	>200
F(guanidino)	17.5
αme2	4.8
me2	>200

*me1, me2, me3, Ac of Lys4 and αme2 of Arg5 from ref. 26 **me2 of Arg5 from ref. 48

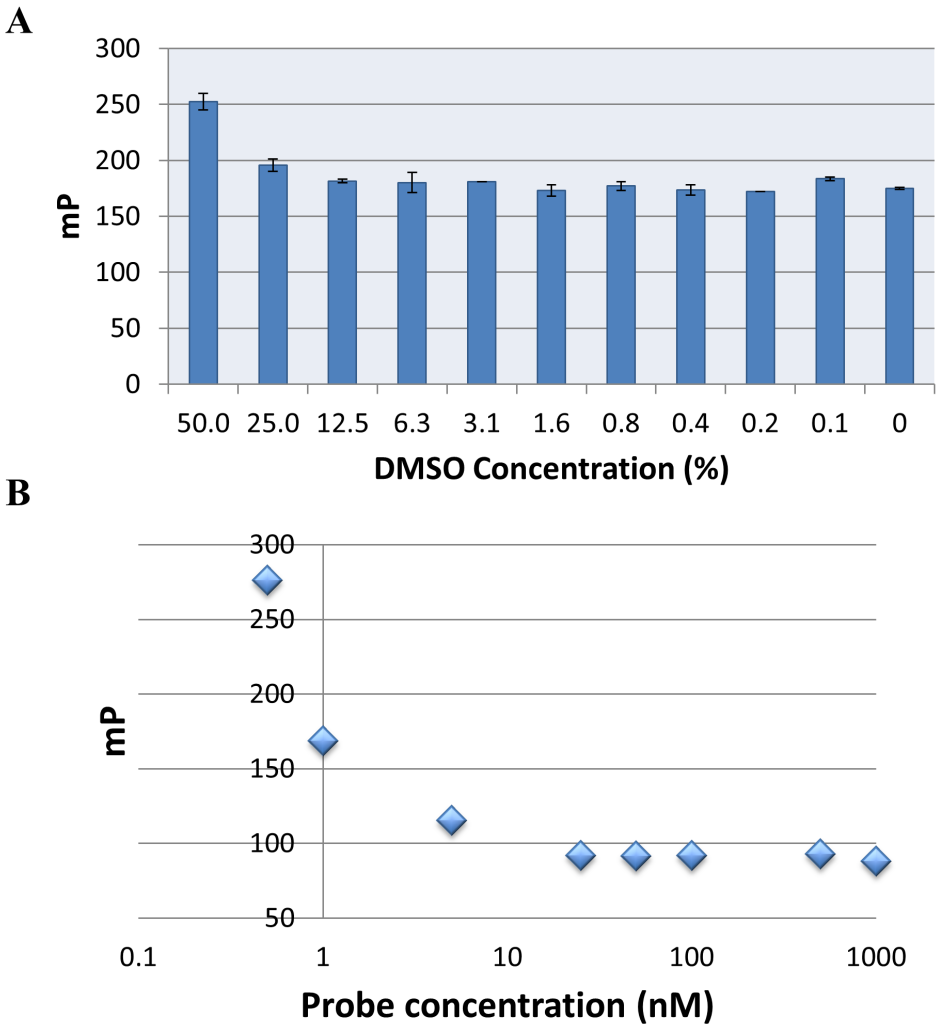
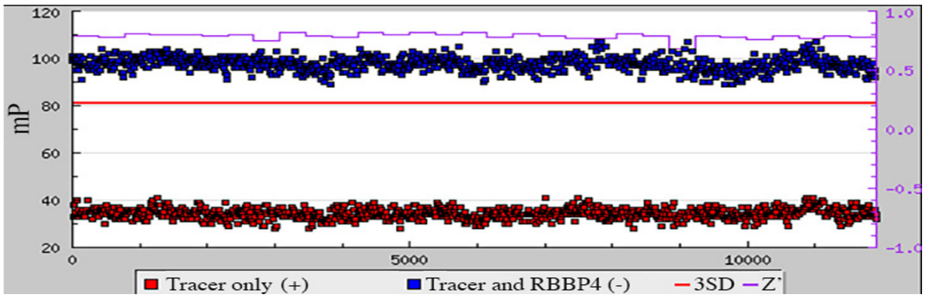
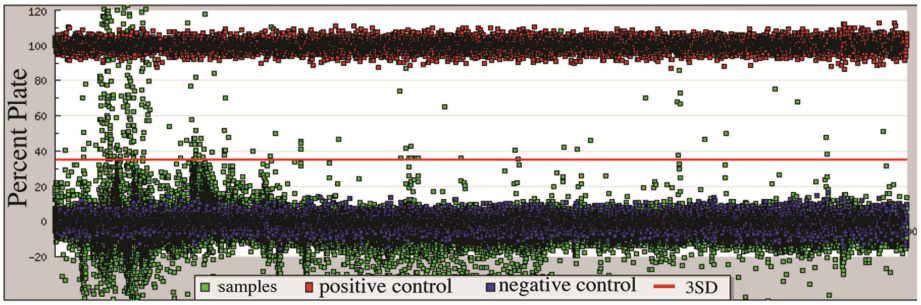


Figure 4.5 Optimization of the DMSO and probe concentration.
(A) 50 µM of RBBP4 and 20 nM of the H3 tracer were incubated in assay buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 0.1% glycerol) with increasing concentrations of DMSO. mP signal was stable around 175 until DMSO reached a concentration of 25% at which time it started to rise. This indicates the assay is stable in concentrations of DMSO up to 12.5%. (B) Increasing concentrations of the H3 tracer was incubated in assay buffer at concentrations ranging from 0.5 to 1000 nM. mP signal was stable at concentrations above 15 nM.

A



B



1875 Actives on 78104 screened

Library	Hits	Total Screened	Hit Ratio
Maybridge_24K	62	23552	0.26 %
Natural Extracts	1687	19839	8.5 %
ChemDiv_100K	89	32640	0.27 %
Prestwick	30	1280	2.34 %
LOPAC	11	1280	0.86 %

Figure 4.6 Results from HTS.
(A) Control plate run at CCG. Half of a 384 well plate consisted of Negative Controls (Tracer and RBBP4) and the other half Positive Controls (tracer only). Z' factor was greater than 0.75 indicating assay was suitable for HTS. (B) Initial screen against compounds (n=1) from natural products, ChemDiv and Maybridge libraries. (C) Hit rate amongst libraries screened.

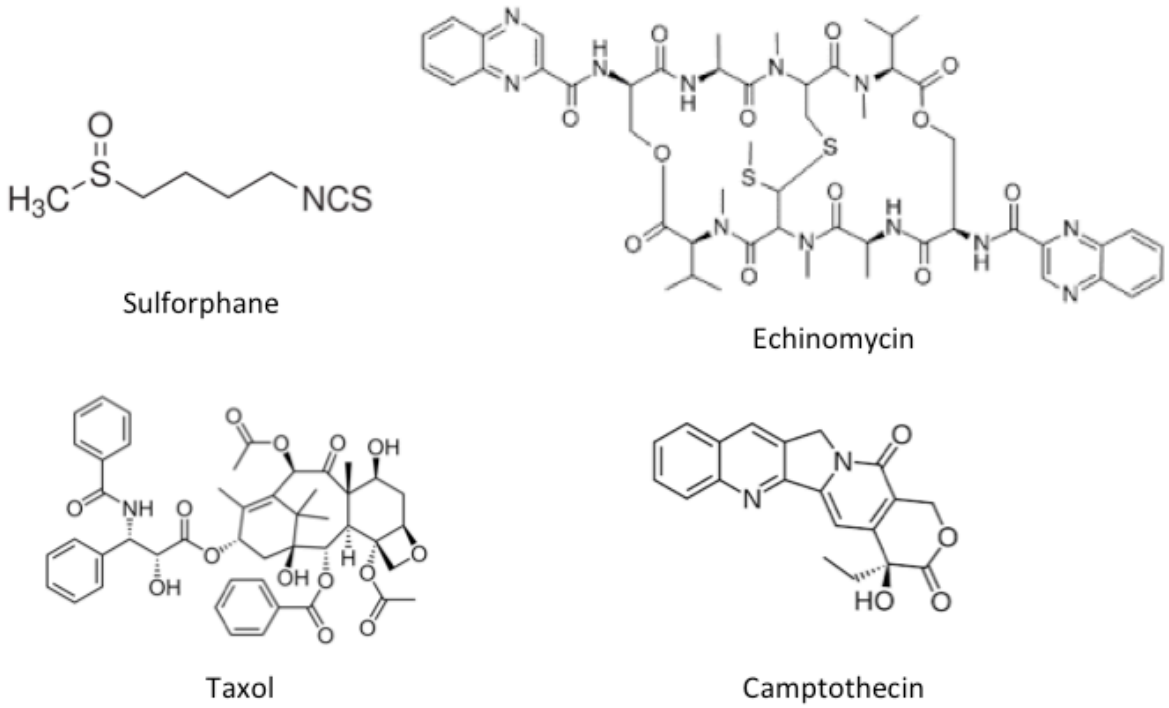


Figure 4.7 Anti-cancerous Natural Products.
Several anti-cancerous Natural Products have been discovered and optimized for use in cancer therapy. For example, sulforaphane, an isothiocyanate discovered in broccoli, block cancer proliferation via upregulation of p53 and downregulation of antiapoptotic proteins. Moreover it was discovered to CSC signaling in breast cancer (ref. 49). Echinomycin inhibits hypoxia-inducible factor (HIF) and suppresses leukemia via reduced NOTCH expression (ref. 51). Taxol which was derived from the bark of the Pacific Yew tree interferes with the breakdown of microtubules during cell division and is used in the treatment of breast, ovarian and lung cancers among others (ref. 50). Lastly, camptothecin derived from the bark and stem of the *Camptotheca acuminate* tree was discovered in 1966 and found to be an inhibitor of DNA topoisomerase I (ref. 52).

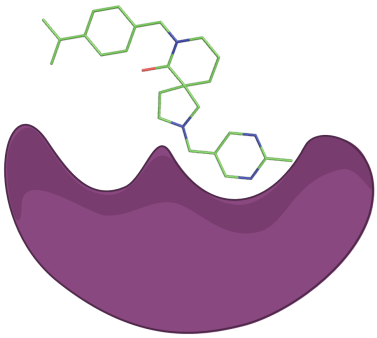
A

CCG_Number	Value (mP)	% Inhibition	ACTIVE_SCREENS	TOTAL_SCREENS
CCG-253841	45	82.8	1	2
CCG-111023	46	81.4	14	51
CCG-114392	46	81.4	6	49
CCG-114393	48	78.6	9	49
CCG-58238	53	71.6	1	46
CCG-41266	55	68.8	6	68
CCG-16379	57	66	6	95
CCG-220792	57	66	14	37
CCG-116287	59	63.2	7	47
CCG-247750	61	60.4	1	2
CCG-236023	64	56.2	1	2
CCG-39016	64	56.2	49	160
CCG-104775	65	54.8	4	57
CCG-123005	65	54.8	8	49
CCG-114390	67	52	9	48
CCG-236014	67	52	2	4
CCG-235966	74	42.1	1	4
CCG-113449	75	40.7	5	49
CCG-133529	75	40.7	10	48
CCG-39478	75	40.7	23	163
CCG-40190	75	40.7	9	152
CCG-103351	77	37.9	2	51
CCG-103407	77	37.9	17	58
CCG-39326	79	35.1	10	157
CCG-112058	80	33.7	13	52
CCG-221129	80	33.7	11	36
CCG-111026	82	30.9	8	50
CCG-220456	82	30.9	17	39
CCG-39135	83	29.5	10	158
CCG-103447	84	28.1	7	52
CCG-109643	84	28.1	5	49

Figure 4.8A HTS small molecules with > 25% inhibition.
(A) CCG compounds identified as hits (n=4) in FP assay with greater than 25% inhibition.

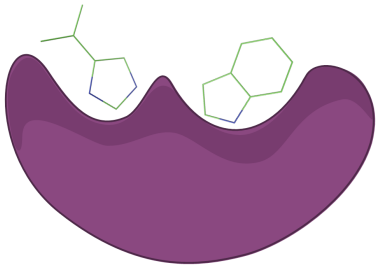
CCG_Number	Value (mP)	% Inhibition	ACTIVE_SCREENS	TOTAL_SCREENS
SID-23650	43	85.6	4	29
SID-171929	60	61.8	8	13
SID-174116	64	56.2	4	12
SID-173887	66	53.4	3	13
SID-23490	67	52	5	30
SID-33212	68	50.6	13	21
SID-28196	75	40.7	18	30
SID-33211	75	40.7	11	21
SID-24620	77	37.9	15	28
SID-26098	77	37.9	5	27
SID-33185	79	35.1	7	20
SID-142490	80	33.7	8	17
SID-32729	80	33.7	14	22
SID-68119	80	33.7	8	20
SID-142659	81	32.3	10	18
SID-23411	81	32.3	6	29
SID-33184	81	32.3	9	20
SID-68993	81	32.3	10	21
SID-143599	82	30.9	6	15
SID-174152	82	30.9	3	12
SID-26759	82	30.9	18	31
SID-172962	83	29.5	3	9
SID-26318	83	29.5	11	30
SID-67414	83	29.5	9	22
SID-174155	84	28.1	5	12
SID-31267	84	28.1	7	20
SID-33194	84	28.1	9	19
SID-67998	84	28.1	6	18
SID-23640	85	26.7	4	29
SID-69626	85	26.7	9	20

Figure 4.8B Positive Natural Product Extracts after triplicate confirmation. Table of the NPEs that had greater than 25% inhibition (n=4) as well as their percent inhibition and the number of screens this extract was deemed active.



High-throughput Screening (HTS)

- $>10^5$ compounds (>300 kDa)
- Poor coverage of chemical space
- Broader range of targets (ex: whole cell screening)
- Higher affinity hits (K_d in low µM range)
- Difficult to optimize hits as they may be very complex
- Biochemical assays (less protein and performed in higher throughput)



Fragment-based drug discovery(FBDD)

- Compounds of smaller size (<300 kDa)
- Greater coverage of chemical space
- Requires well-characterized targets
- Low affinity hits (K_d ~100µM-1 mM)
- Step-by-step optimization increasing the size of the molecule
- Biophysical Screening methods which are low-to-medium throughput and require large amounts of protein

Figure 4.9 Comparison between High-throughput Screening and Fragment-based drug discovery.
Adapted from ref. 37 and 38.

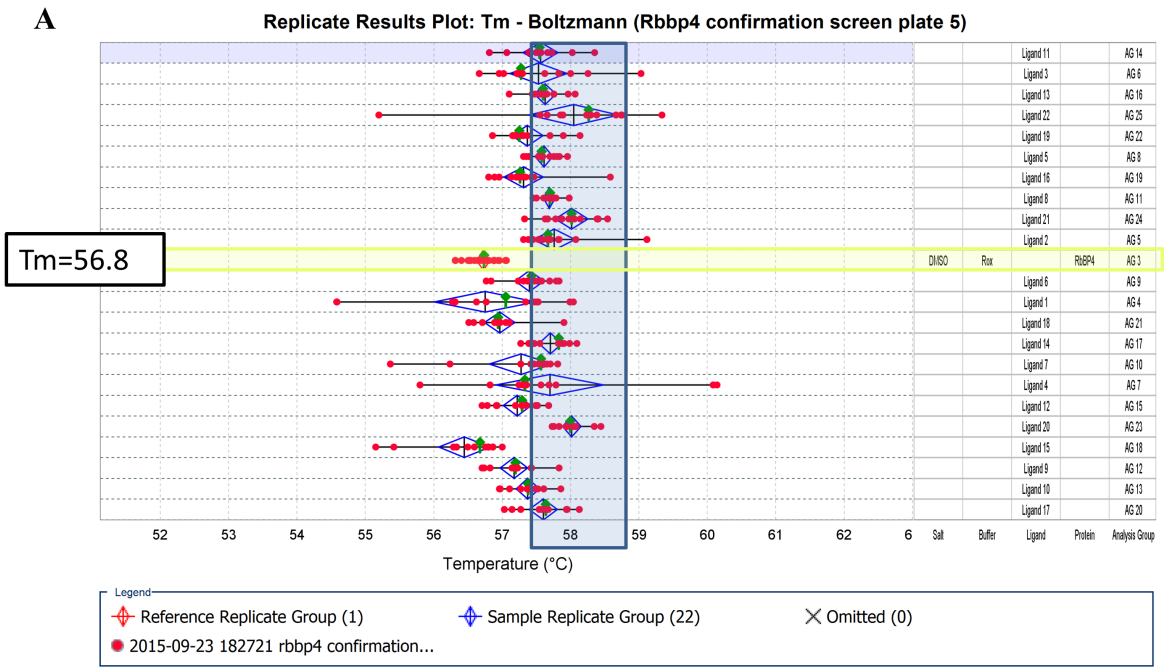


Figure 4.10 Analysis of Boltzmann Tm shifts.
(A) Example screening results from a confirmation plate run with 12 replicates of each fragment. Tm of RBBP4 and DMSO (highlighted in yellow) was 56.8°C. The blue window indicates ligands that shifted the melting temperature by greater than 2 S.D. (in this case 1 degree).

ID	Renumbered	Tm Shift
2047-9680	1	2
8018-8696	2	1
8020-0036	3	0.8
D205-0138	4	0.8
G349-2146	5	1.5
K780-1342	6	1
M893-0132	7	1
M893-0795	8	1
P277-0135	9	1.2
R052-2310	10	1
R315-0015	11	1
T482-1662	12	1
T786-0038	13	1.2
T992-0173	14	1
Y041-4077	15	1.2
Y200-0520	16	1.2
Y020-4431	17	1.5
8020-2436	18	3
F0191-4581	19	1
F1912-0069	20	1.5
F1957-0067	21	1
F2124-0078	22	1
F2145-0303	23	2
F2145-0815	24	1.2
F2147-1064	25	1
F2513-0171	26	1.5
F3284-8093	27	2
F5857-0380	28	0.9

Figure 4.11 Confirmed Fragments that bind RBBP via Thermal Shift Assay. Table of the compound ID and the Tm Shift it produced.

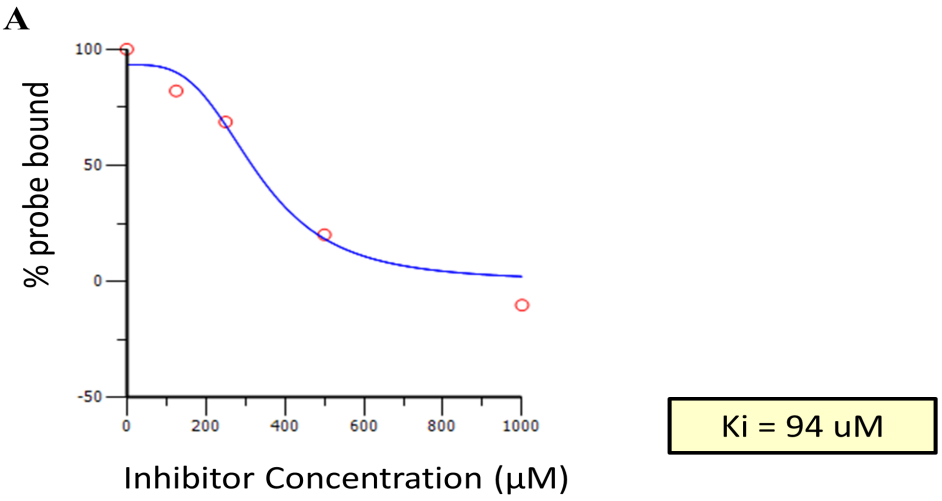


Figure 4.12. Compound 12 fragment, T482-1662, competed with H3 tracer to bind RBBP4.
(A) Initial competitive binding curve of compound 12 (T482-1662), run at 5 concentrations, against H3 tracer.

Table 4.3 Compound 12 Analogs purchased

Compounds that exhibited some inhibition activity against H3 peptide are highlighted in yellow.

A

Compound ID	Number	Ki (μM)
T482-0925	1	
T482-1005	2	130
T482-1008	3	139
T482-1012	4	
T482-1022	5	
T482-1111	6	
T482-1335	7	
T482-1659	8	
T482-2325	9	
T482-2403	10	
T482-2758	11	
T482-3112	12	
T482-3115	13	

B

Compound ID	Number	Ki (μM)
9334608	14	
18511202	15	
9193268	16	
9307278	17	
9116453	18	
90037826	19	
9020130	20	
9127030	21	
9333777	22	
99498404	23	
9117090	24	
9322675	25	
9335021	26	
45486440	27	
41381074	28	
39522521	29	
15434536	30	
31579039	31	
47699300	32	
11216153	33	
91170205	34	
32823916	35	
11342657	36	
5274452	37	

Compound ID	Number	Ki (μM)
9117560	38	
9123258	39	
55828978	40	
9343400	41	
9293424	42	
9295393	43	
9308579	44	
9303558	45	
4101370	46	
5270579	47	
9339196	48	
4101369	49	
9293020	50	
9263671	51	
9193270	52	
93678577	53	
9121582	54	
9126709	55	
9026525	56	
9030574	57	
14133809	58	21
9022278	59	
19645138	60	

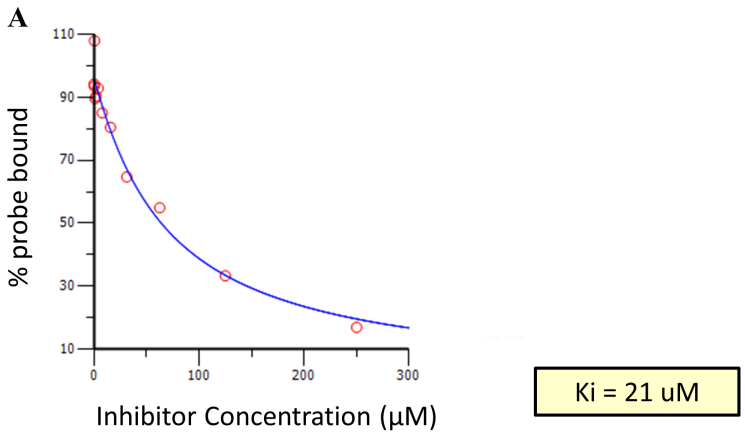


Figure 4.13 Improvement in inhibitory ability by compound 12 analog.
(A) 60 analogs of compound 12 were purchased from ChemDiv and Chembridge and run in a competition assay with RBBP4 and H3 tracer. (B) Full competition curve of compound 58 (14133809) indicated a Ki value of 21 µM (Ki calculated using ref. 42). The structure differed from the original hit in (1) a pyrazolopyrimidine core in place of the triazolopyridine, (2) an amine to replace the urea, (3) a triazole ring, and (4) an aliphatic propyl group.

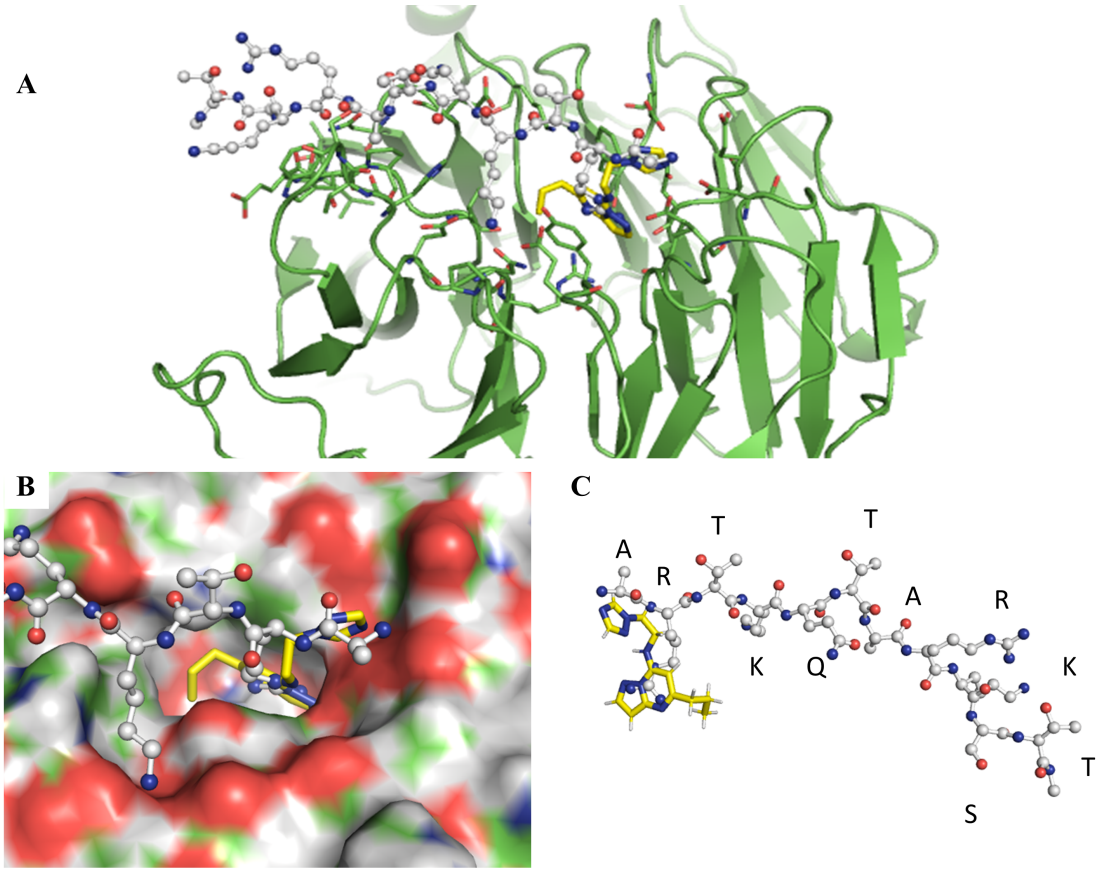


Figure 4.14 Docking of Compound 58 to RBBP4.
(A) Side profile view of compound 58, shown in yellow, overlaid against the histone H3 peptide, shown in white). (B) Compound 58 inserts into the top face of RBBP4 occupying the canonical binding site. (C) Compound 58 occupies the same binding site as the Arg2 residue of the H3 peptide and the triazole ring rests in the shallow surface pocket similar to Ala1 of H3.

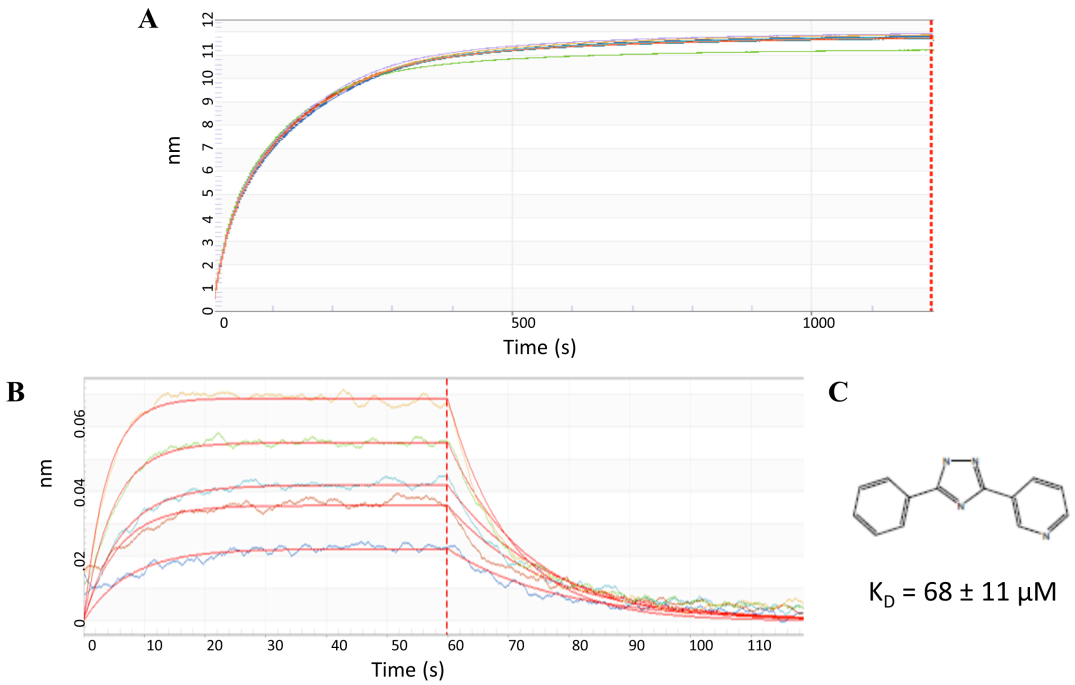


Figure 4.15 OctetRED assay development.
(A) Representative loading curve of biotin-labeled RBBP4 to Super Streptavidin (SSA) biosensors showing a maximal shift around 11-12 nm and saturation reached by 15 minutes. (B) Association-Dissociation cycles of compound 18 at various concentrations and 1:1 binding model fit curves (in read). (C) Structure of potential lead fragment 18 and K_D (68 ± 11 µM) as determined by ForteBio software.

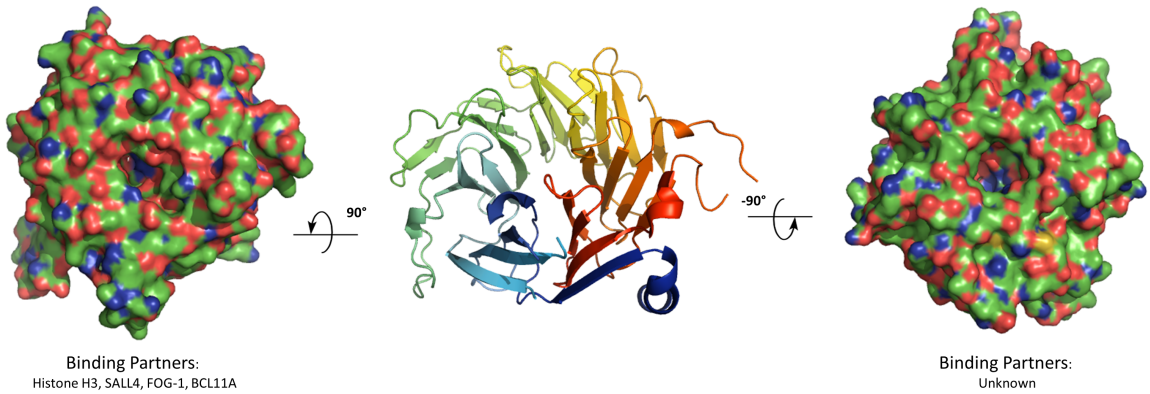


Figure 4.16 Binding Pockets of RBBP4.

Chapter 5

Conclusions & Future Directions

While breast cancer research has uncovered several successful treatments and novel targeted therapies, there still exist the critical issues of recurrence and metastasis. Among the four subtypes of breast cancer, TNBC (ER-/PR-/HER2-) has high rates of recurrence and metastasis and thereby has a generally poor prognosis.[1,2] The recent identification of breast cancer stem cells (BCSCs), which are attributed to tumor initiation, chemoresistance, and metastasis, helps explain the aggressive nature of TNBC.[3-5] Therefore, there has been growing interest in understanding pathways and proteins important for maintaining the BCSC population so as to develop effective therapies for the prevention and treatment of metastatic breast cancer.

Accumulating evidence has indicated that epigenetic mechanisms are important for the progression of cancer as well as the maintenance of the two BCSC defining characteristics: self-renewal and multipotency.[6-10] Epigenetic complexes may alter the dynamic state of chromatin and regulate access to the genome for all DNA-related processes such as repair, transcription and replication. In fact global epigenetic alterations have been shown to precede genetic mutations in cancer.[11,12] In addition, changes in histone modifications such as loss of acetylation at Lys 16

and trimethylation of Lys20 of histone H4 are now viewed as hallmarks of cancer.[13] In BCSCs, epigenetic complexes that have important roles in blocking differentiation, such as NuRD and PRC2, may be exploited by oncogenic transcription factors to promote stemness of cancer cells. Aberrant recruitment of these epigenetic machineries can result in drastic transcriptional changes and gene expression leading to tumorigenesis.

Within this thesis, we demonstrate that RBBP4 and RBBP7, which are involved in several epigenetic complexes, are important proteins in maintaining TNBC proliferation and BCSC regulation. In Chapter 2, we show that knockdown of RBBP4/7 not only decreases TNBC growth *in vitro*, but also *in vivo* with a decrease in the CD44+/CD24- CSC population. We further show that there is a decrease in the basal markers cytokeratin CK14 (or KRT14) and integrin ITGA6 (or CD49f) and a concomitant increase in luminal markers (CD24, CK19, and GATA3). Both ITGA6 and CK14 have been associated with invasive carcinoma, drug resistance, and cancer stem cell activity.[14,15] Expression of GATA3 and CK19, on the other hand, are correlated with longer relapse-free and overall survival.[15-17] The shift towards luminal differentiation in RBBP4/7 knockdown cells could make them more susceptible to traditional endocrine therapies. Inhibition of RBBP4/7 may therefore transform aggressive basal-like cancer into a cancer with a far better prognosis. Gene set enrichment analysis (GSEA) confirmed the decrease in basal breast cancer signaling. GSEA also suggested RBBP4/7 knockdown leads to inhibition of PRC2 function, as genes normally marked with the H3K27me3 mark in the undifferentiated embryonic stem cells, were upregulated in dual knockdown cells,

which may contribute to the observed phenotypes. Global levels of H3K27 trimethylation were decreased confirming reduced PRC2 activity. In addition, we observed downregulation in two pathways that are critical to development and maintenance of self-renewal in breast cancer stem cells: WNT and Notch. The Notch receptor, NOTCH1, and the WNT receptor, WNT5B, were downregulated in RBBP4/7 KD cells. The Notch target gene, c-Myc was also significantly decreased. Furthermore, the PRC2 target gene, DKK1, which is an antagonist of the WNT pathway, was upregulated in RBBP4/7 KD. These results suggest that RBBP4/7 are important in maintaining these two signaling pathways in breast cancer stem cells, but the mechanism remains to be found. Current experiments in our lab are looking at the differences between siEZH2 and siRBBP4/7 treated cells in efforts to discriminate between PRC2-associated functions of RBBP4/7 and others. Interestingly, we found that expression of DKK1 was upregulated to a greater extent in siRBBP4/7 cells compared to siEZH2 (data not shown). It is well established that combination therapies against epigenetic targets mediate greater changes in gene expression than single epigenetic inhibitors. Therefore, a potential explanation for better DKK1 induction could be the fact that RBBP4/7 work in concert with multiple epigenetic machineries to control transcription.

The multifaceted functions of RBBP4/7, with its involvement in many epigenetic complexes, may complicate the interpretation of our data. Ultimately, additional studies will be required to determine the direct versus indirect effects of RBBP4/7 depletion in TNBC. Chromatin immunoprecipitation (ChIP) assays coupled with sequencing (ChIP-seq) will allow us to globally identify genes bound

by RBBP4 and RBBP7. For particular genes of interest, further ChIP-qPCR assays will help assess which RBBP4/7 associated epigenetic complex, or complexes, is responsible for the target gene regulation. Due to the fact that RBBP4/7 are mainly associated with repression, it logically follows that genes upregulated in the knockdown cells may be directly affected by RBBP4/7. That does not mean, however, that genes downregulated in the knockdown studies are not directly bound by RBBP4/7 and we cannot rule out other functions of RBBP4/7 outside of repression.

While further studies are needed to better understand the biological consequences of RBBP4/7 in TNBC, we did establish that these histone chaperones could serve as a novel therapeutic target to inhibit TNBC proliferation and CSC activity. Therefore, we sought to develop an assay to gauge the chemical tractability of RBBP4 and aid in the discovery of small-molecule inhibitors of RBBP4/7. Assessment of the RBBP4 structure, a WD40 protein with the characteristic β-propeller, directed our attention to the binding interface between RBBP4 and histone H3. This interaction, which occurs at the top face of the β-propeller, was previously reported to be important in the regulation of PRC2 activity.[18] When an activating H3K4me3 mark was present on the histone tail, it precluded RBBP4 binding and reduced catalytic efficiency of EZH2, the enzymatic component of PRC2.[18] For this reason and because RBBP4/7 promote nucleosome association of epigenetic complexes, we hypothesized that by blocking this site we would prevent recruitment of PRC2, and potentially NuRD, to their histone substrate.

In our studies we found that the canonical binding site of RBBP4 is also occupied by the transcription factor BCL11A, which was recently implicated in the maintenance of BCSCs and specifically in TNBC. We determined the crystal structure of RBBP4 in complex with an aminoterminus derived peptide of BCL11A (aa 2-16). The structure points to important contacts at the C-terminus of the BCL11A peptide. While these amino acids (aa 12-16) did not seem to be necessary in our competitive FP-based assay, other studies have indicated that these residues might be important for optimal interaction with epigenetic complexes. In one instance a FOG1 peptide, which has a similar sequence to BCL11A, was far less effective at pulling down epigenetic complexes if the peptide was truncated below the first 15 amino acids.[19] Initially, it was believed that the 'RRKQ' motif shared amongst these transcription factors and histone H3 was the only important contact that mediated binding to RBBP4/7. Our study indicates that this is not the case and reveals the molecular basis for interactions at the C-terminus of BCL11A. Potential routes to block recruitment of RBBP4 and its associated complexes could therefore focus on the canonical binding site or the newly found interactions exposed in our crystal structure. As a proof-of-concept study we used a BCL11A derived peptide to treat SUM149, a TNBC cell line. We believe this peptide could compete with endogenous BCL11A for binding to RBBP4 and prevent BCL11A-directed transcriptional regulation. We found that treatment with this peptide led to a 50% reduction in a CSC population marked by aberrant aldehyde dehydrogenase (ALDH) activity. Inhibition of RBBP4/7 will therefore interfere with recruitment and activity epigenetic complexes leading to a decrease in CSCs. Coupled with our data

from Chapter 2, we are challenging the notion that these proteins are just histone chaperones, in fact we believe the primary function of RBBP4/7 could be in their association with transcription factors and epigenetic complexes to bridge histones with histone-modifying enzymes.

To find inhibitors of the top face of RBBP4/7, we performed high-throughput screening with our optimized FP assay. We chose to screen against small molecule libraries as well as natural product extracts to improve our sampling of the chemical space and chances for finding an inhibitor. We are currently involved in the triage and validation process for small molecules, but the deconvolution of hits from NP extracts will take a longer time. Nevertheless, our studies pave the way for a novel therapeutic approach in the treatment of TNBC: targeting RBBP4/7 to block recruitment of epigenetic complexes to their target loci and inhibit BCSCs. Identified inhibitors will also serve as a biological probe to provide insights into the architecture and recruitment of RBBP4/7's associated epigenetic complexes and the implications in disease.

References

1 Cetin, I. & Topcul, M. Triple negative breast cancer. *Asian Pacific journal of cancer prevention : APJCP* **15**, 2427-2431 (2013).

2 Lehmann, B. D. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* **121**, 2750-2767, doi:10.1172/JCI45014 (2011).

3 Gangemi, R. *et al.* Cancer stem cells: a new paradigm for understanding tumor growth and progression and drug resistance. *Current medicinal chemistry* **16**, 1688-1703 (2009).

4 Liu, S. *et al.* Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports* **2**, 78-91, doi:10.1016/j.stemcr.2013.11.009 (2014).

5 Olivier, G., Christoph, A. K. & Patrick, A. B. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nature reviews. Cancer* **8**, 755-768, doi:10.1038/nrc2499 (2008).

6 Lu, D. *et al.* MicroRNA100 inhibits self-renewal of breast cancer stem-like cells and breast tumor development. *Cancer research* **74**, 6648-6660, doi:10.1158/0008-5472.CAN-13-3710 (2014).

7 Munoz, P., Iliou, M. S. & Esteller, M. Epigenetic alterations involved in cancer stem cell reprogramming. *Mol Oncol* **6**, 620-636, doi:10.1016/j.molonc.2012.10.006 (2012).

8 Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised epigenetic states and acquired drug resistance in cancer. *Nat Rev Cancer* **14**, 747-753, doi:10.1038/nrc3819 (2014).

9 Kagara, N. *et al.* Epigenetic regulation of cancer stem cell genes in triple-negative breast cancer. *Am J Pathol* **181**, 257-267, doi:10.1016/j.ajpath.2012.03.019 (2012).

10 Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat Genet* **39**, 157-158, doi:10.1038/ng1941 (2007).

11 You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22**, 9-20, doi:10.1016/j.ccr.2012.06.008 (2012).

12 Hitchins, M. P. *et al.* Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. *Cancer Cell* **20**, 200-213, doi:10.1016/j.ccr.2011.07.003 (2011).

13 Fraga, M. F. *et al.* Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* **37**, 391-400, doi:10.1038/ng1531 (2005).

14 Brooks, D. L. *et al.* ITGA6 is directly regulated by hypoxia-inducible factors and enriches for cancer stem cell activity and invasion in metastatic breast cancer models. *Molecular cancer* **15**, 26, doi:10.1186/s12943-016-0510-x (2016).

15 Abd El-Rehim, D. M. *et al.* Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* **203**, 661-671, doi:10.1002/path.1559 (2004).

16 Yan, W., Cao, Q. J., Arenas, R. B., Bentley, B. & Shao, R. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J Biol Chem* **285**, 14042-14051, doi:10.1074/jbc.M110.105262 (2010).
17 Yu, K. D. *et al.* Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin Cancer Res* **19**, 2723-2733, doi:10.1158/1078-0432.CCR-12-2986 (2013).
18 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol Cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
19 Saathoff, H. *et al.* A peptide affinity reagent for isolating an intact and catalytically active multi-protein complex from mammalian cells. *Bioorg Med Chem* **23**, 960-965, doi:10.1016/j.bmc.2015.01.023 (2015).