UNDERSTANDING THE CONTRIBUTIONS OF THE POLYCOMB CBX PARALOGS IN BINDING AND ONCOGENSIS

by

Katelyn Elizabeth Connelly

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Medicinal Chemistry and Molecular Pharmacology

West Lafayette, Indiana

December 2018

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Emily Dykhuizen, Chair

Department of Medicinal Chemistry and Molecular Pharmacology

Dr. Michael Wendt

Department of Medicinal Chemistry and Molecular Pharmacology

Dr. Scott Briggs

Department of Biochemistry

Dr. W. Andy Tao

Department of Biochemistry

Approved by:

Dr. Andy Hudmon

Head of the Graduate Program

To my parents, Scott and Evelyn, and my sisters, Ashley and Emily for their unwavering support and encouragement,

&

To all of my science teachers, especially Mrs. Yasmeen Youngs and Mr. Craig Oldham, for sparking my love for science at a young age.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my PhD advisor and mentor Dr. Emily Dykhuizen for taking me on as her first graduate student and her continual support and encouragement throughout my PhD. I cannot thank you enough for the opportunities to grow as a scientist and person. Additionally, a special thank you to my dissertation committee, Dr. Scott Briggs, Dr. W. Andy Tao, and Dr. Michael Wendt for the fruitful discussions and career guidance.

I would like to thank members of the Dykhuizen lab, past and present for the support, encouragement, guidance, and conversations. My PhD would not be the same without you and I will miss working with everyone. I wish you all the best in your future endeavors. A special thank you to Libby who has become one of my closest friends over the past several years. It has been wonderful to be able to bounce ideas, troubleshoot, analyze data, etc. and have lunch with you daily. It's no wonder people are surprised when they see one of us alone. I wish all of you the best. I know you will do great things and I can't wait to collaborate down the road.

Additionally, many thanks to all of our collaborators. This work would not be possible without you. Thank you to Dr. Uma Aryal and the Purdue Proteomics Facility for working with me to develop a system to study nuclear protein complexes. It taught me more than I ever imagined it would. A special thank you goes to Dr. Catherine Musselman and Tyler Weaver for jumping all in on our chromodomain binding project. It has been a pleasure working together from afar and at the bench. Tyler, thank you for sharing your structural biology perspective with me. I wish you the best with your future endeavors.

To my friends and family, thank you for adding balance to my life and supporting me day in and day out.

Finally, big thank you to Minervo who has been my biggest supporter over the last several years. I cannot imagine grad school without having you in my corner. You've helped me grow as a person and scientist over the years even from 600 miles away. I look forward to what the future holds for us, and I am excited to begin our next journey together.

Thank you to my sources of funding: the Purdue University Ross Assistantship and Bilsland Dissertation Fellowship and the Indiana Clinical and Translational Sciences Institute pre-doctoral training award.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

bps, base pairs

PcG, Polycomb Group

PRC1, Polycomb Repressive Complex 1

cPRC1, canonical PRC1

PRC2, Polycomb Repressive Complex 2

Pc, Polycomb

Ph, Polyhomeotic

Psc, Posterior sex combs

Sce, Sex combs extra

H3K27me3, histone H3 lysine 27 trimethylation

SAM, sterile alpha motif

PCGF, Polycomb Group Finger

PHC, human Polyhomeotic

CBX, chromobox homolog

PREs, Polycomb Response Elements

MLL, mixed lineage leukemia

SWI/SNF, Switch/sucrose non-fermentable

ESC embryonic stem cell

HSC, hematopoietic stem cell

NPCs, neural progenitor cells

ChIP, chromatin immunoprecipitation

MEF, mouse embryonic fibroblast

MES, early cardiac mesoderm precursor

PLFZ, promyelocytic leukemia zinc finger

lncRNA, long non-coding RNA

MS, mass spectrometry

miRNA, miR, microRNA

MAPK, mitogen-activated protein kinases

IHC, immunohistochemistry

CBP, Creb-binding protein

EAP, elongation assisting proteins

GBM, glioblastoma multiforme

TMZ, temozolomide

Dox, doxorubicin

CD, chromodomain

dPc, Drosophila Polycomb

BF, breast fibroblast

SSE, sequential salt extraction;

HEK, human embryonic kidney

ΔCD, chromodomain deletion

WT, wild-type

NMR, nuclear magnetic resonance

HSQC, heteronuclear single quantum coherence

CSP, chemical shift perturbation

Kd, dissociation constant

Δδ, normalized CSP

MLA, methyl-lysine analog

EMSA, electrophoretic mobility shift assay

NCPs, nucleosome core particles

ENCODE, Encyclopedia of DNA Elements

GEO, Gene Expression Omnibus

SEM, standard error of the mean

DEG, differentially expressed genes

ABSTRACT

Author: Connelly, Katelyn, E. PhD
Institution: Purdue University
Degree Received: December 2018
Title: Understanding the Contributions of the Polycomb CBX Paralogs in Binding and Oncogenesis
Committee Chair: Emily Dykhuizen

The transcriptional repressor Polycomb Repressive Complex 1 (PRC1) is critical for stem cell maintenance and proper differentiation and as such is involved in the development and progression of cancer. Canonical PRC1, composed of PCGF, PHC, RING and CBX, binds histone H3 lysine 27 trimethylation (H3K27me3) allowing for ubiquitination, chromatin compaction and subsequently transcriptional silencing. In mammals, each subunit has multiple paralogs creating functional and compositional diversity. The greatest diversity is contributed by the CBX targeting subunit with five mutually exclusive paralogs (CBX2/4/6/7/8). The CBX paralogs contain an N-terminal chromodomain for methyl-lysine binding. There has been interest in the CBX paralogs due to their misregulation in various cancers and the "druggability" of the chromodomain histone interaction. However, the unique biochemical and transcriptional functions of the paralogs are unclear. Expression changes during lineage specification and the context-dependent misregulation of CBX paralogs in cancers suggest the paralogs have paralog-specific functions. However, little has been done to define differences in paralog-mediated chromatin binding and regulation. This work utilizes a variety of approaches to tease apart the biological and biochemical functions of the CBX paralogs in chromatin binding and oncogenesis. In this dissertation, we identify a combinatorial therapeutic strategy using a CBX chromodomain inhibitor to enhance chemotherapeutic response. Further, this work demonstrates a role for CBX8 and its chromodomain in glioblastoma oncogenesis suggesting it may serve as a therapeutic target. Finally, we identify a binding mechanism for the CBX8 chromodomain in which DNA and H3K27me3 binding contribute to full chromatin association.

CHAPTER 1. INTRODUCTION

The following chapter was reproduced and modified with permission from Connelly, K.E., and Dykhuizen, E.C., *Biochim Biophys Acta*. 2017 Feb; 1860(2): 233-245. doi: 10.1016/j.bbagrm.2016.12.006.

1.1 Introduction

1.1.1 Genome Evolution

One of the more surprising findings from sequencing the human genome is the relatively small number of protein coding genes.[1] Because genome size roughly correlates with organismal complexity, it was predicted that the number of protein coding genes would too. In prokaryotes where the genome contains almost exclusively protein coding sequences, genome size is proportional to gene number; however, even in simple eukaryotic systems, such as *S. cerevisae*, a significant jump in genome size (12 million base pairs (bps)) compared to most prokaryotic genomes does not come with a significant jump in gene number (6,600). This is reflective of an increased proportion of the genome dedicated to gene regulation instead of protein coding. *S. cerevisae* organize their genome via chromatin and express a multitude of proteins involved in epigenetic regulation that are not found in prokaryotes. Many of these proteins are conserved in higher organisms, making yeast a tremendously useful system for studying many epigenetic mechanisms involved in gene regulation, including splicing, histone modifications, and chromatin remodeling.

The evolution of multicellularity represents a major jump in organismal complexity, which is represented in Drosophila by an increase in gene number to 14,000 and even bigger increase in genome size to 140 million bps. Genes involved in cellular differentiation and identity emerge, along with epigenetic modifiers to regulate these genes. Interestingly, in the evolution to higher-order organisms, such as vertebrates, a continual increase in genome-size occurs without a significant increase in gene number. In fact, mammals have only small increases in the number of genes compared to Drosophila (21,000 in humans) but huge increases in genome size (3200 million bps). This increase in genome size is primarily non-coding and represents increasingly complex mechanisms of

gene regulation. Many of the complex phenotypes observed in higher-order organisms are not a result of protein-coding sequences but a result of increasingly complex mechanisms of transcriptional regulation, with increased importance for splicing, non-coding RNA, genome organization, and histone modification.

1.1.2 Polycomb as a regulator of cell differentiation

One family of epigenetic regulators important during the advent of multicellularity is the Polycomb group proteins. There are no Polycomb complexes in unicellular yeast such as *S. pombe* and *S. cerevisiae*; however, there is evidence of a Polycomb-like complex in *C. Neoformans*[2] as well as unicellular algae *C. reinhardtii,*[3] indicating a possible co-opting of mechanisms of adaptation for developmental processes. Originally discovered in *Drosophila melanogaster* as regulators of appropriate body segmentation,[4] Polycomb group (PcG) proteins form multiple complexes. The two most well studied complexes are known as Polycomb Repressive Complex 1 and 2 (PRC1/2). PRC2, composed of three core proteins Esc, E(z), and Su(z)12, is a histone methyltransferase catalyzing the trimethylation of histone H3 lysine 27 (H3K27me3) via the SET domain in E(z).[5] PRC1 comprised of four core subunits, Polycomb (Pc), Polyhomeotic (Ph), posterior sex combs (Psc), and sex combs extra (Sce or dRing), binds the H3K27me3 and monoubiquitinates lysine 118 (119 in humans) on H2A to repress gene transcription.[6]

In Drosophila, the PRC1 subunits contribute differently to proper function. A heterodimer of dRing and Psc form the E3 ubiquitin ligase, which is responsible for the monoubiquitination of H2AK118 (119 in humans) and subsequent transcriptional repression.[7,8] The Pc chromodomain is a chromatin-binding domain that recognizes the ARK(me3)S amino acid sequence to bind H3K27me3. This approximately 60 amino acid N-terminal domain is critical for complex localization across the genome.[9] In addition to its role in ubiquitination, the Psc subunit is capable of compacting nucleosomal templates *in vitro.*[10] The function of the Ph subunit is relatively unknown, but it does contain a sterile alpha motif (SAM) domain that is necessary for Ph polymerization and repressive function.[11] Numerous studies have examined the mechanisms in which PRC1 represses gene transcription in Drosophila, although a definitive mechanism has yet to be fully resolved. While it was originally thought that the canonical mechanism of transcriptional

repression involves H2AK119 ubiquitination in order to repress transcription,[6] Kingston and colleagues have demonstrated that the ubiquitination is not essential for nucleosome compaction *in vitro*.[10,12,13] PRC1 binding alone is able to alter the conformation of chromatin, and pre-incubation of PRC1 prevents the RNA polymerase II pre-initiation complexes from forming on DNA[14] and inhibits the chromatin remodeling capabilities of SWI/SNF *in vitro,*[8] leading to transcriptional repression. Therefore, even in Drosophila, there may be multiple mechanisms by which PRC1 regulates transcription.

1.1.3 Mammalian Polycomb group proteins

Polycomb proteins are conserved across all multicellular organisms and universally involved in developmental processes. In plants, the general mechanism of Polycomb-mediated repression is maintained, although the complexes differ significantly from animals[15] and differ significantly among members of the plant kingdom.[16] In vertebrates, however, the homology to Drosophila Polycomb proteins is high, although the number of PcG genes has more than doubled compared to Drosophila.[17] While the only expansion of the core PRC2 subunits is the duplication of E(z) to EZH1 and EZH2, the PcG proteins that comprise canonical PRC1 underwent tremendous expansion (Figure 1.1). Psc is known as Polycomb Group Finger (PCGF) in vertebrates and consists of six paralogs (*PCGF1-6*); Ph expanded to three paralogs (*PHC1-3*); the dRING homologs are RING1A and RING1B, and Pc, known as CBX in mammals, has expanded to five different genes (*CBX2,4,6,7,8*) (Figure 1.2). The various paralogs resulted from two genome duplications that occurred during vertebrate evolution. While many of these duplicated genes were slowly lost over time, the duplicated *HOX* and PRC1 genes remained, implying a shared importance in cell identity and differentiation. Despite being highly homologous, the different paralogs are more closely related among species than they are to the other paralogs within the same species.[18] For example, human CBX2 is more similar to mouse CBX2 than it is to human CBX7, indicating functional specificity that has been evolutionarily maintained. This expansion of canonical PRC1 subunits adds a new level of complexity to deciphering the functional role of PRC1 in regulating transcription and defining cell state.

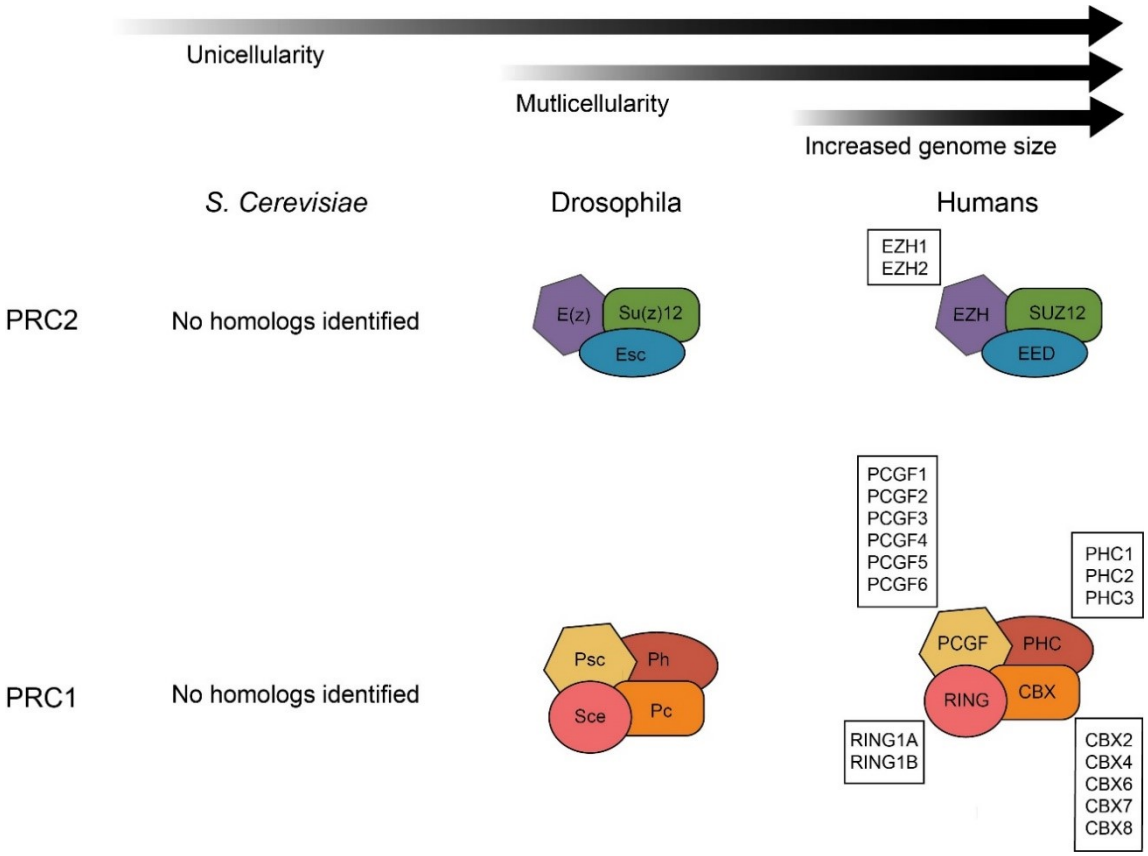


Figure 1.1 Evolution of the PcG Family.

Expansion of the Polycomb genes occurs in vertebrates as the genome size increases.

To add to that complexity, multiple non-canonical PRC1 complexes containing PCGF and RING1A/B but not CBX or PHC, have been identified (reviewed in [19,20]). Non-canonical PRC1 complexes comprise the majority of PRC1 complexes in many cell types, including embryonic stem cells (ESCs), and provide a mechanism by which PRC1 is targeted independent of H3K27me3.[21] In fact, in the absence of DNA-encoded Polycomb Response Elements (PREs) in mammals, non-canonical PRC1 complexes may be one mechanism for the recruitment of PRC2.[22,23] RING1B is the only canonical PcG protein essential for early embryonic development,[24] and the essential nature of it and non-canonical subunits such as RYBP[25] is in line with a role in cellular maintenance, cell-cycle regulation and regulation of metabolism genes.[26] In contrast, canonical PRC1-mediated repression is only critical during lineage commitment and differentiation, and as such, the relative protein expression of canonical PRC1 increases from less than 10% of all PRC1 complexes in embryonic stem cells to over 15% in neural progenitors.[21] Since canonical and non-canonical PRC1 complexes share many of the same subunits and transcriptional targets,[26] non-canonical PRC1 provides an additional layer of regulation for canonical PRC1 complex formation and function; however, this dissertation will focus primarily on the diversity of function within canonical PRC1 in mammals.

Many other chromatin-regulating complexes display similar increases in compositional heterogeneity and functional diversity in higher-order organisms. For example, chromatin remodeling complexes such as MLL and SWI/SNF are similarly heterogeneous, similarly important for cell-type specific transcriptional function, and similarly misregulated in cancer.[19,27,28] The presence of conserved homologs in yeast has greatly advanced our understanding of basic transcriptional functions for these regulators, while our understanding of PcG proteins, particularly PRC1 subunits, lags behind. Genetic deletion of PRC1 subunits in Drosophila confirms a role for PRC1 in cooperating with PRC2 to repress *Hox* genes during development,[29] however, the functional role of PRC1 in mammals has been more difficult to discern, in part due to the high compositional complexity. With the recent successes in targeting aberrant PRC2 methyltransferase

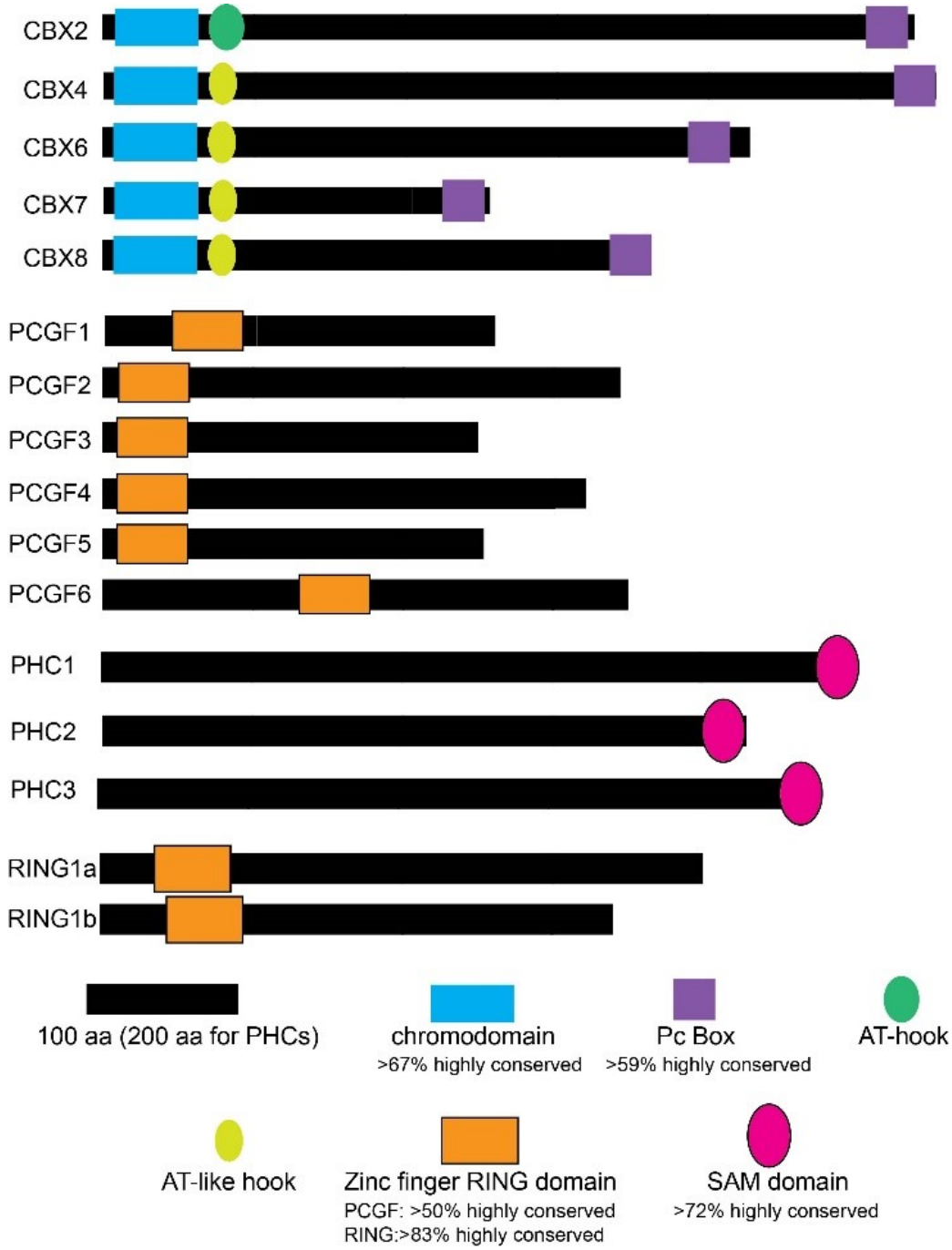


Figure 1.2 Paralogs of canonical PRC1 subunits.

Mammalian PRC1 has at least two paralogs for every subunit. The major conserved domains are depicted to visualize similarities and differences among the paralogs. The highly conserved percentage is denoted for each domain relative to the Drosophila homolog, according to Protein Basic Local Alignment Search Tool (BLAST).

function in cancer,[30,31] increased focus has turned to understanding how aberrant PRC1 function might promote oncogenesis, metastasis and drug resistance.[32–34] PRC1 function in cancer is likely related to its roles in maintaining stem cell states,[35] as such, a greater understanding of how unique PRC1 complexes define cell state will lead to improved cancer treatments. There are three basic mechanisms by which paralogs might impart specific functionality to PRC1 complexes: differing effects on transcriptional repression through ubiquitination, chromatin compaction or other undefined enzymatic/biochemical properties, association with unique protein partners, and differential chromatin targeting. In this dissertation, we attempt to address some of the differing roles the diverse array of canonical PRC1 complexes may play in gene regulation and cell identity.

1.2 Transcriptional Regulation of PRC1 proteins

Many studies have examined the relative expression of PRC1 subunit paralogs in cell lines and tissues in order to begin defining cell-type specific functional roles. Knockdown, knockout, and overexpression studies have been further used to elucidate many of the known developmental functions of PRC1. Knockout of PRC1 subunits, with the exception of *RING1B*, are not embryonic lethal but often result in skeletal deformities (reviewed in [19,36–40]). *CBX2* knockout mice undergo a male to female sex reversal[41,42] while *CBX7* knockout mice are normal but have a greater tendency to develop liver and lung cancer.[43] Since studies in other cancers suggest an oncogenic role for *CBX7,*[44–48] there are likely context dependent roles for some PRC1 subunits. To further relate these phenotypes to transcriptional regulation, many studies have sought to define PRC1 gene targets using RNA sequencing (RNA-seq) and chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq).[26,49–54] While all these studies seek to answer the questions pertaining to PRC1 function, there are still significant gaps in our knowledge of how unique biochemical properties of each paralog relate to transcriptional profiles important for defining cell state. We will attempt to provide a brief synopsis of some of the multitude of studies investigating PRC1 subunit expression, dependencies, and transcriptional output in order to set the stage for more in depth discussion of biochemical function, although we direct readers to the many excellent reviews covering this topic in more detail.[19,20,35,37,40,55]

1.2.1 PRC1 subunit expression

There are many lines of evidence supporting a role for cell-type specific expression of PRC1 subunits. A tissue microarray study in 2006 characterized the expression of PRC1 subunits *RING1A*, *RING1B*, *BMI-1* (*PCGF4*), *MEL18* (*PCGF2*), and *PHC1* in various tissues.[56] *RING1A*, in accordance with The Human Protein Atlas, was identified as being expressed in almost all tissues whereas *RING1B* expression has been identified in only 60% of tissues.[56,57] *BMI-1* and *MEL18* expression was positive in a majority of the normal tissue and often co-expressed, while *PHC1* was only detected in few tissues: parathyroid, pituitary gland, testis, and pancreas.[56] Most importantly, the expression of these proteins varied between cell types and even between the same cell-type in different organ systems.[56] For example, the glial cells in the cerebral cortex had low *BMI-1* expression, but in the brain stem and basal nuclei, *BMI-1* staining was higher.[56]

To pinpoint more precisely how PRC1 subunit expression correlates with cellular identity, several differentiation systems have defined changes in PRC1 subunit transcript and protein expression upon lineage commitment. In every system investigated, lineage specific cell types express new PRC1 complexes with alterations in gene targets.[52,54,58,59] In ESCs, PRC1 is predominantly non-canonical; however, the canonical PRC1 complex in ESCs contains CBX7, MEL18, PHC1, and RING1B (Figure 1.3) while the remaining canonical PRC1 subunits are transcriptionally repressed.[21,52,60] The expression of the other canonical PRC1 subunits are negligible until differentiation to embryoid bodies, upon which new complexes composed of BMI-1, CBX2, and CBX4 are formed.[21,52] Differentiation of ESCs into cardiomyocytes revealed three unique complexes throughout the process (Figure 1.3). Upon induction of differentiation, the predominant PRC1 complex in early cardiac mesoderm precursor cells (MES) consisted of MEL18, CBX2, RING1B, and PHC2; however, upon further development into cardiomyocytes, Morey *et al*. observed a shift in complex formation to CBX4 and PHC3.[54] In hematopoietic stem cells (HSCs), CBX7 is the predominant chromobox homolog, however, upon differentiation into hematopoietic lineages CBX8 replaces CBX7 (Figure 1.3).[58] In HSCs, CBX2, CBX4, and CBX8 overexpression can induce differentiation, whereas CBX7 and BMI-1 overexpression drive self-renewal capabilities.[58,61] Overexpression studies of CBX7 and CBX8 suggest that both paralogs can associate with BMI-1 and MEL18,[58] despite several

proteomic studies suggesting endogenous CBX7 does not associate with BMI-1.[45,52] Neural progenitor cells (NPCs), on the other hand, do not have a single PRC1 complex. Multiple CBX (2, 4, and 8) and PHC (1, 2, and 3) paralogs[52,53] are present with BMI-1, but not MEL18 (Figure 1.3).[54,62] Similarly, fibroblasts express 15 of the 16 PcG genes,[63] although as is the case with many differentiated cell types, the dependency on PRC1 for normal function is unclear. The dysregulation in cancer implies an important function in maintaining a differentiated state, but these roles have not been well-defined in non-transformed cells. Ectopic overexpression of BMI-1 and CBX8 in the fibroblast line TIG3 imparts the ability of cells to bypass senescence[64] but isn't sufficient to transform cells. Though complexes have been defined in select cell types, the PRC1 complexes have yet to be defined in the majority of cell types. In addition, how unique complexes relate to unique biochemical and transcriptional function is still unclear.

1.2.2 PRC1 transcriptional regulation

Knockdown studies and genome-wide localization studies have identified transcriptional autoregulation by PRC1 subunit paralogs to be a common theme.[49,52,65] In ESCs, CBX7 was found localized at the loci of *CBX2, 4, 8*, and *BMI-1*, repressing their expression.[52] In fibroblasts, however, knockdown of *CBX7* does not affect *CBX8* expression level,[52,66] suggesting that CBX7 does not regulate *CBX8* expression in all cell types. An inverse correlation between BMI-1 and MEL18 has been noted for numerous cell systems. As stated above, ESCs exclusively contain MEL18, which is replaced with BMI-1 in progenitor cells, such as hematopoietic and neuronal. The reverse process of BMI-1 upregulation and MEL18 downregulation has been noted in many cancers leading to the hypothesis that BMI-1 is the primary oncogenic driver while MEL18 acts as a tumor suppressor. Overexpression of MEL18 can repress *BMI-1* in different cell lines; however, it is unclear if this is a direct or indirect target of MEL18.[67,68] In human diploid fibroblasts, overexpression of MEL18 decreases the expression of BMI-1 and induces senescence, however, it has been shown to be indirect as MEL18 negatively regulates *MYC* expression, which activates *BMI-1*.[65] While it is clear that differential expression of the PCGF paralog alters gene expression, it is not clear how they alter the biochemical functions of PRC1.

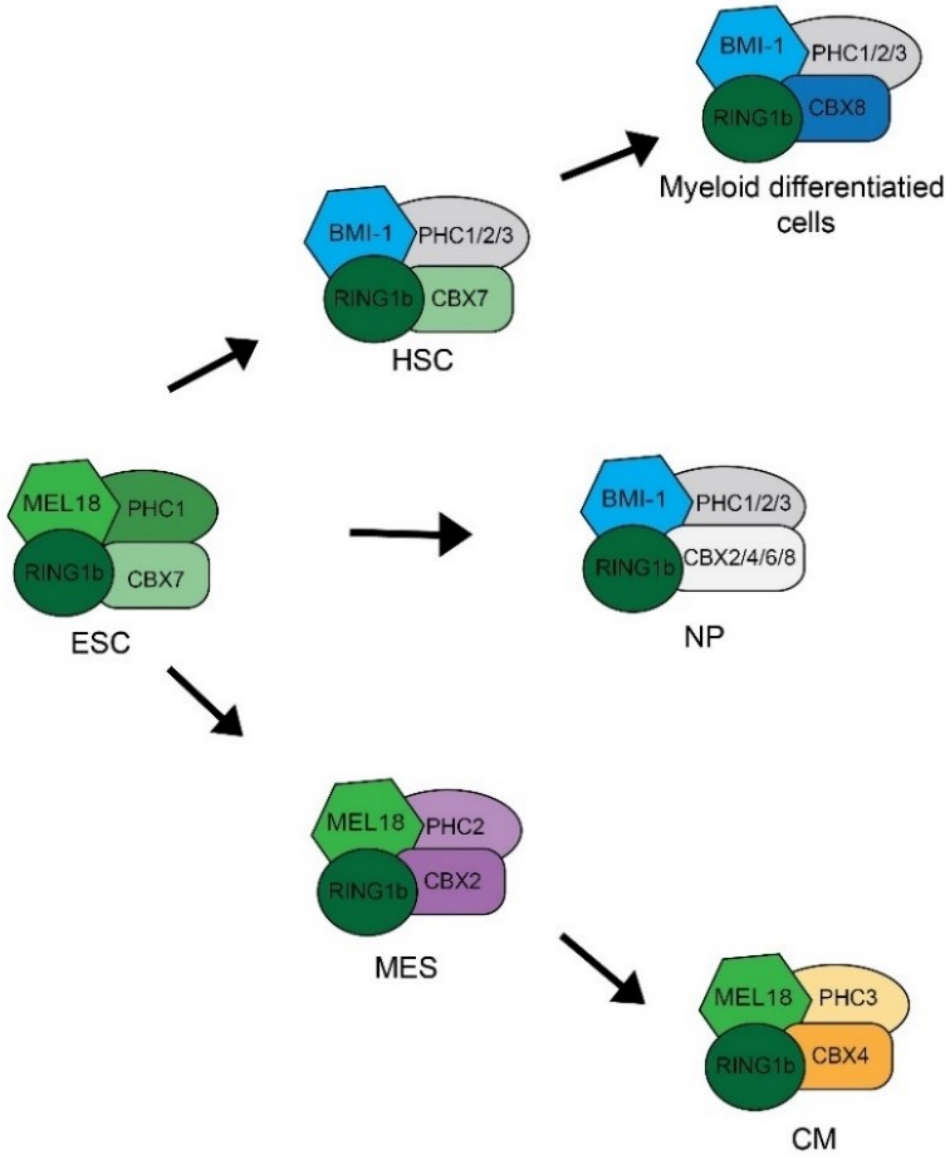


Figure 1.3 Map of compositional changes in PRC1 during lineage specification.

In ESCs, one canonical PRC1 complex has been defined; however, under different differentiation pathways the composition of PRC1 changes depending on the lineage and stage of differentiation.

Micro RNAs (miRNAs) play an important role in coordinating the regulation of PRC1 subunit expression.[69] Several miRNAs that inhibit *RING1B* (miR181a,b, miR200b,c) and *BMI-1* (miR203, miR200b,c) expression were identified.[70] Interestingly, the repression of these miRNAs is controlled by PRC2, suggesting a positive feedback loop.[70] O'Loghlen *et al.* identified two miRNA families, miR181 and miR125, that are important for mediating CBX7 expression during ESC differentiation.[60] While not expressed in ESCs, the miR181 and miR125 are induced upon differentiation, particularly neuronal differentiation.[60] They bind to the 3'UTR of *CBX7* to posttranscriptionally reduce CBX7 expression and allow for upregulation of the other CBX paralogs.[60] An additional miRNA family was identified to mediate CBX7 expression.[71] The miR9 family, whose expression induces senescence, was found to downregulate CBX7 resulting in upregulation of p16 and cellular senescence.[71] Like the miRNA181 and miR200, the miR9 family is part of a negative feedback loop with CBX7.[71] While miR9 downregulates CBX7, CBX7 was found to repress miR9 expression and bound to its loci.[71]

1.3 Unique PRC1 complex formation

Considering all of the possible combinations of the 16 paralogs for canonical PRC1 subunits, there is a potential for 180 possible PRC1 complexes to exist in mammals. An important question is whether all of these possible complexes can actually exist. There are several pressing questions in the field: what are the compositions of PRC1 complexes?; do these complexes have unique function?; and how do they regulate cell type-specific transcriptional programs? A major breakthrough in defining PRC1 complexes comes from a comprehensive proteomic and genomic analysis of PRC1 subunits from HEK293T performed by Reinberg and coworkers.[72] Prior to this comprehensive study, several studies determined many of the same associations but focused on only a few subunits at a time. Proteomic mass spectrometry (MS) analysis identified PCGF as the defining subunit of unique complexes PRC1.1-PRC1.6.[72] MS analysis indicates that while all PCGF subunits

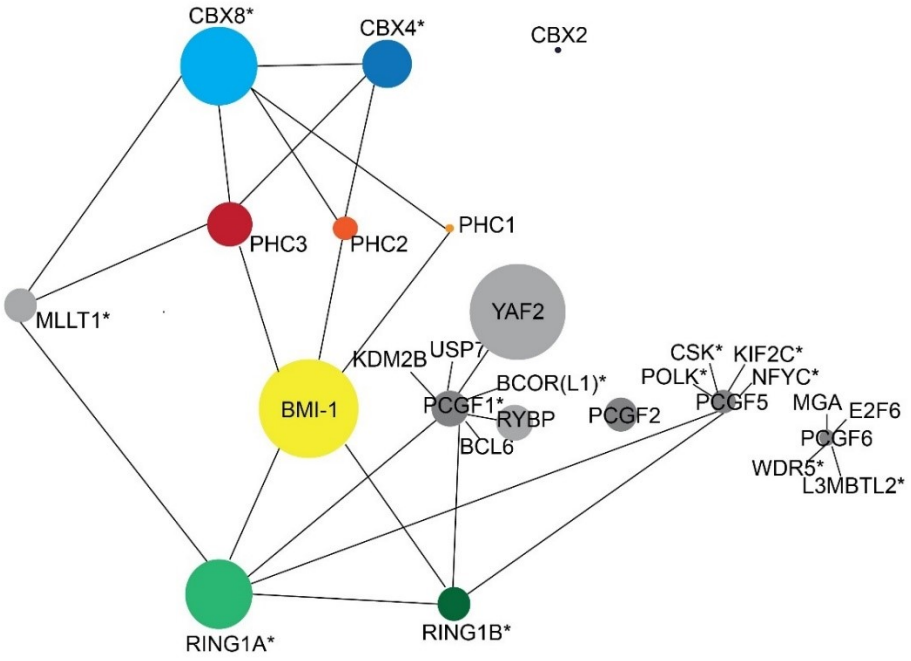


Figure 1.4 PRC1 interactome.

This illustration depicts canonical and non-canonical PRC1 interactions based proteomic studies in HeLa cells[73]. Circle size indicates protein expression level. If a subunit is not represented, the endogenous expression level in HeLa is negligible. Canonical PRC1 members are illustrated in color. The asterisk (*) indicates the protein served as a bait.

associate with RING1A and RING1B, only PCGF2 (MEL18) and PCGF4 (BMI-1) associate with PHC and CBX to produce canonical PRC1. In turn PCGF1 associates with BCOR(L1), KDM2B, and USP7, PCGF3 associates with AUTS2, PCGF5 associates with FBRS(L1) and CKII, and PCGF6 associates with L3MBTL2, E2F6, WDR5, MAX, MGA and HDACs.[72] In addition, all six PCGF paralogs associate with mutually exclusive subunits RYBP or YAF2;[72] however, for MEL18 and BMI-1-containing complexes, the binding of RYBP/YAF2 competes with CBX to bind the same surface of RING1B.[72,74] Because of this mutually exclusive relationship, RYBP/YAF2 inclusion is used to define a diverse set of non-canonical PRC1 complexes with mechanisms of targeting independent of H3K27me3.[72]

Since many complexes are found to be cell-type specific, an important question is whether the various PRC1 complexes defined in HEK293T cells are similar in other cell

types. A recent wide-scale proteomics study identifying not just interactions, but also stoichiometries, confirms many of these interactions in HeLa cells (Figure 1.4).[73] Interestingly, in addition to confirming PRC1.1-PRC1.6 subcomplexes, this study also defines an alternate canonical PRC1 complex containing CBX8 and ENL (MLLT1).[73] Additional proteomics studies utilizing overexpression systems have confirmed these associations but also identified PCGF paralogs 1, 3, and 5 to co-purify with canonical CBX subunits,[73,75,76] however, much work needs to be done to confirm such association. There are some complications and caveats that come with proteomic analyses. First, many different PRC1 complexes closely associate at the same loci making it highly possible for immunoprecipitation to capture associating interactions that may not be true complex members. In such instance, there would be multiple paralogs co-immunoprecipitating that may not be in the same complex but rather associating. Additionally, overexpression of a single PRC1 paralog could alter stoichiometry and produce PRC1 complexes not found endogenously. Kloet *et al.* noted that when CBX4 was overexpressed in ESCs, it did not compete with CBX7 for incorporation into canonical PRC1 but was able to interact independently with RING1B and form additional canonical PRC1 complexes. Whether CBX4 competes with non-canonical subunit for RING1B binding is unknown, but it does indicate that a hierarchy may exist for subunit incorporation into PRC1 complexes.[21]

Several studies have demonstrated that CBX7 is in a complex only with MEL18 and not with BMI-1 in endogenous systems;[45,52,54] however, when CBX7 is overexpressed in HeLa cells, which normally have little CBX7, it readily interacts with BMI-1 whose expression is five-fold higher than MEL18.[73,75] It is possible that under endogenous conditions, CBX7 does not prefer BMI-1 as an interacting partner, but when over-expressed the homeostasis of PRC1 is shifted. This aberrant PRC1 complex formation upon subunit misregulation may in fact explain several of the PRC1 complexes recently observed in cancer that contain both canonical and non-canonical subunits. In breast cancer, CBX8 forms a non-canonical complex with WDR5, a subunit of the MLL1 methyltransferase complex and previously identified member of the PRC1.6 complex.[51,77] In germinal center B cells and the lymphomas derived from them, CBX8 is upregulated and associates with the non-canonical BCOR PRC1.1 complex,[76,78] an association which has been further recapitulated in overexpression systems using HEK293T cells.[78] Further

studies are necessary to understand the balance of PRC1 expression and formation in normal and oncogenic cells to aid in delineating PRC1's function.

This data strongly supports the hypothesis that cell-type specific PRC1 complexes exist and are important for function; however, the PRC1 complexes have not been defined for most differentiated tissues. Many studies have identified cell-type specific functions for particular PRC1 subunits.[40,79] This could be a result of differing cell conditions or could be result of novel PRC1 subunit partners. In other words, differential PRC1 function may not just be a simple addition of the functions of individual subunits but unique functions that result from novel composite surfaces.

1.4 Unique Biochemical properties

It is known that PRC1 can bind, compact, and ubiquitinate chromatin; however, the individual subunits (and paralogs) have unique biochemical functions that potentially affect PRC1's function (Figure 1.5).

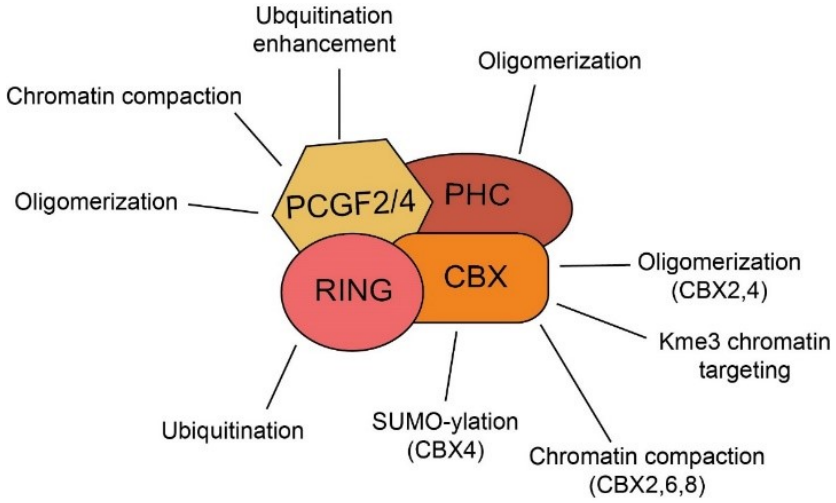


Figure 1.5 Summary of PRC1 biochemical functions by subunit.

The individual subunits of PRC1 have unique biochemical functions, however, not all paralogs have the same functions.

1.4.1 Ubiquitination

Although paralogous PRC1 subunits are highly homologous, many display unique biochemical properties. The RING ubiquitin ligases are responsible for H2AK119

ubiquitination, the role for which is still not entirely clear. H2AK119 ubiquitination is proposed to mediate transcriptional silencing by restraining POLII, preventing the eviction of nucleosomes, and preventing the trimethylation of H3K4, possibly through steric hindrance.[80–82] RING1A can mildly compensate for loss of the RING1B paralog; however, *in vivo* loss of RING1B has a much larger effect on H2AK119 ubiquitination than RING1A and is embryonic lethal.[52,83,84] Interestingly, *in vitro* experiments demonstrate that RING1A and RING1B have similar enzymatic activity alone[85] but that RING1A stimulates RING1B ubiquitin ligase activity when combined.[7] Although the mechanism by which this happens is unknown, it may be due to trans-ubiquitination. In particular, the ubiquitination of multiple lysines in the C terminus of RING1B contributes to its function as a H2A E3 ubiquitin ligase[86] even though the primary site of ubiquitination on lysine 112 appears to be indispensable.[87] RING1A and RING1B are ubiquitously expressed together and are both simultaneously incorporated into all PRC1 complexes examined.[72] Their relative expression levels do vary in different cell types, possibly proving a mechanism for modulating PRC1 ubiquitin ligase activity.

The PCGF subunit enhances ubiquitination by RING,[7] and the PCGF paralogs can have different enhancement capabilities. BMI-1 significantly stimulates RING ubiquitination *in vitro*, whereas MEL18 has little to no impact on ubiquitination;[7] similarly deletion of BMI-1 significantly affects *HOX* gene expression, while MEL18 deletion does not.[7] Part of BMI-1's activity may be related to its ability to stabilize RING1B and mediate its self-ubiquitination by reducing the number of available lysines four-fold.[86] How the other PCGF subunits might influence RING ubiquitin ligase activity has not yet been established; however, the associating factors found with other PCGF subunits may influence their activities. Non-canonical BMI1-containing PRC1 complexes with RYBP have significantly higher ubiquitin ligase activity than canonical BMI-1-containing PRC1 complexes with PHC and CBX.[72] This is similar to data in Drosophila in which the dRAF complex (dRING, PSC, and dKDM2 and occasionally RYBP) is more effective at catalyzing H2AK119ub1 than canonical PRC1.[88] In line with this, ubiquitination in Drosophila requires KDM2 and no requirement for Ph.[88]

In addition to the ubiquitin ligase activity of the RING subunit, CBX4 is an E3 SUMO ligase despite being non-homologous to other E3 ligases. CBX4 enhances the

deposition of the SUMO modification on proteins such as the transcriptional co-repressor CtBP and the DNA methyltransferase DNMT3A among other proteins.[89–91] Additionally, BMI-1 can be SUMOylated by CBX4 during the DNA damage response.[91] Intriguingly, CBX4 itself can also have SUMO-modifications,[89] although the functional consequence of these modifications is unclear.

1.4.2 Chromatin compaction

In addition to ubiquitination**,** an important biochemical function for PRC1 complexes is chromatin compaction, which is frequently measured *in vitro* using nucleosomal templates. Chromatin compaction does not require histone tails and is driven by highly basic regions of Psc in Drosophila.[10] In mammals, however, chromatin compaction is driven by highly basic regions of the CBX proteins.[12] Grau *et al*. demonstrated that particular CBX paralogs (CBX2, 6, and 8) can compact chromatin in the absence of the PRC1 complex due to inclusion of a highly basic region that is necessary to drive chromatin compaction. Interestingly, in contrast, CBX7, which lacks a highly basic region, fails to compact chromatin efficiently,[12] and no information was available for CBX4, although it does not have an obvious basic region in its amino acid sequence. Perhaps as a result of this function for CBX proteins in mammals, canonical PRC1 complexes are slightly more effective at compacting nucleosomes *in vitro* than non-canonical PRC1 complexes.[72] This indicates a possible separation of PRC1 function in mammals such that non-canonical PRC1 is primarily responsible for ubiquitin ligase activity while canonical PRC1 complexes are primarily involved in compaction. While both of these functions for PRC1 are transcriptionally repressive, they are independent functions.[13] This implies that canonical and non-canonical PRC1 complexes may in fact have slightly different mechanisms of transcriptional repression and that the balance of canonical and non-canonical PRC1 may be a factor in determining cell state. As such non-canonical PRC1 predominates in ESCs (and is required for pluripotency) while the fraction of canonical PRC1 increases upon differentiation.[21]

1.4.3 Looping

We often think about gene transcription and regulation in a linear manner; however, chromatin forms higher-order structures important for transcriptional regulation. One potential mechanism of PRC1 repression involves the formation of higher-order chromatin structures (reviewed in [92–95]). Via microscopy, it has been known for some time that Polycomb proteins tend to cluster in discrete nuclear foci termed Polycomb bodies. Using immunohistochemistry (IHC) or live cell imaging, different paralogs, such as the CBX paralogs, have different patterns of Polycomb body distribution.[96,97] In 2004, Kingston and colleagues determined that PRC1 complexes compact chromatin by moving nucleosomes and creating DNA loops.[10] Often, numerous PRC1 complexes form PcG bodies, or "hubs," by looping the DNA and bringing distal genomic regions together.[92] For example, the *HOX* loci are looped to multiple promoter hubs[98] and are a model for how certain genes are regulated by multiple paralogs simultaneously. A similar phenomenon likely explains the regulation of the *INK4a/ARF* locus by multiple paralogs of PRC1.[66] Looping is likely mediated through oligomerization of PRC1 complexes, as subunits from disparate complexes can be identified at the same loci but without direct overlap.[66,72] These structures are multi-looped and dynamic and associated with insulator elements.[92,98] The size of Polycomb bodies can be large, as observed for the *HOX* loci, or small; however, the number of Polycomb bodies is less than the number of genes bound by Polycomb as determined by ChIP-Seq.[63,93] It is unknown whether all of the thousands of ChIP-Seq targets bound by PRC1 are necessarily in Polycomb bodies.

The size and number of Polycomb bodies varies between cell types. ESCs display a few large Polycomb bodies while more differentiated cells have more, smaller Polycomb bodies.[96] This is in line with evidence that a loss of looping occurs during differentiation.[99] A major question is how Polycomb bodies are maintained and why IHC with Drosophila shows that disruption of H3K27me3 by EZH2 inhibition does not dramatically disrupt PRC1 chromatin distribution within Polycomb bodies.[100,101] In addition, while mutation of CBX chromodomains completely abrogates chromatin binding to discrete loci as observed with ChIP,[58] it does not alter gross nuclear localization into Polycomb bodies as determined by live cell fluorescence.[96] Based on these studies as well as the measured affinities of CBX chromodomains to H3K27me3,[102,103] a reasonable hypothesis might be that

distribution of PRC1 into discrete Polycomb bodies serves to increase the local concentration of PcG proteins, which facilitates chromodomain-mediated binding at specific loci. How canonical PRC1 might be localized into these discrete bodies is not clear but may be controlled by self-association.

1.4.4 PRC1 self-association and oligomerization

PRC1 compaction and looping is facilitated through PRC1 self-association and oligomerization. While many genomic sites indicate the co-localization of paralogs, extensive immunoprecipitation and mass spectrometry analyses clearly find no co-association between paralogs in solution. This may be explained by *in vitro* results indicating only self-association for many of the PRC1 subunits. As mentioned in the introduction for the Ph subunit of Drosophila, the PHC subunits can self-associate through their SAM domains.[104] This polymerization is critical in PRC1 clustering, condensation of chromatin, and repression;[104] however, this function is moderated by O-GlcNAcylation, without which over-aggregation impedes Ph's repressive capacity.[11] In addition, similar to Psc in Drosophila,[105] both BMI-1 and MEL18 are able to dimerize in solution.[106] A splice variant of CBX2 without the Pc box required for association in the PRC1 complex was found to form multimeric complexes *in vivo* indicating that self-association of CBX proteins does not require PRC1 complex formation.[107] While CBX self-association may not require PRC1 complex formation, mass spectrometry studies identify the association between two CBX4 splice variants that do contain the Pc box and are incorporated into PRC1,[75] indicating that CBX self-association can also occur in the context of PRC1 formation. It is likely that the self-association of PRC1, driven through multiple subunits to form PcG bodies, enhances the capability of PRC1 to compact and repress chromatin. This mechanism could be analogous to the well-established mechanism for HP1-mediated chromatin compaction, although initial studies to probe the stoichiometry of PRC1 using single molecule fluorescence imaging indicate that PRC1 complexes from ESCs are primarily monomeric and span several nucleosomes.[108] Interestingly, however, upon ESC differentiation, the authors observe a significant increase in the fraction of PRC1 dimers,[108] indicating that complex composition or cellular environment can influence PRC1 oligomerization.

1.5 PRC1 targeting

In Drosophila, PcG proteins are recruited to specific genes via PREs.[109,110] Despite extensive studies, mammalian PREs have not been identified; however, in ESCs, PcG proteins are enriched at a subset of CpG islands recognized by KDM2B in non-canonical PRC1.1 complexes.[111,112] Subsequent recruitment of PRC2[22] to these sites, followed by H3K27me3 sets the stage for canonical PRC1 binding in the developing embryo.[113] While PRC1 binding correlates with sites of H3K27me3, canonical PRC1 typically only binds a subset of H3K27me3 sites across the genome.[63] In addition, these targets change upon the expression of alternate PRC1 subunits during differentiation[52,58] implying a possible role for canonical PRC1 paralogs in differential targeting.

1.5.1 Paralog redundancy in targeting

In Drosophila, H3K27me3 is deposited by PRC2 at genetically defined PREs that recruit a complex array of DNA binding proteins. In turn, the chromodomain of Pc recognizes H3K27me3 and represses gene transcription through direct compaction of nucleosomes, H2AK119 ubiquitination and eviction of SWI/SNF chromatin remodelers.[114] The mammalian CBX chromodomains are highly homologous to the Pc chromodomain,[102,103] and this model of recruitment has been carried over to mammals. Application of this model to a mammalian system suggests that the different paralogs of the PRC1 subunits can act redundantly to repress the same genes marked by H3K27me3. This redundancy in targeting is also observed on the molecular level. The classical mammalian PRC1 target, *INK4a/ARF*, is the perfect example of different PRC1 complexes acting redundantly; CBX4, 7, and 8, MEL18, and BMI-1 have all been identified at the locus in the same cells.[66,115] Further, depletion of CBX7 or BMI-1 disrupts the binding of CBX8 or MEL18 at the *INK4a/ARF* and vice versa, indicating an interdependence of paralogs for silencing at this locus.[66] Additionally, genome-wide association of the CBX paralogs 6, 7, and 8 in two different fibroblast cell lines identified a vast number of overlapping gene targets.[63] This is further substantiated in embryonic stem cells, the chronic myeloid leukemia cell line K562, and hematopoietic stem and progenitor cells (HSPCs), where CBX2/4, CBX2/8, and CBX7/8 have overlapping gene targets,

respectively.[50,52,58] In contrast to the *INK4a/ARF* locus, CBX7 and 8 compete for binding sites in HSPCs, indicating cooperativity in function at some sites and antagonism at others.

1.5.2 CBX and chromodomain-mediated targeting

Without PREs in mammals, it is still unclear how H3K27me3 marks are deposited across the genome. Nevertheless, it has been established that H3K27me3 marks can be cell-type specific and affect cell-type specific changes in transcription.[116,117] Accordingly, fluorescent-tagged PRC1 complexes exhibit different localization patterns in ESCs compared to mouse embryonic fibroblasts (MEFs).[97] Similarly, a change in gross nuclear localization of tagged CBX paralogs is observed upon ESC differentiation into neuronal or adipocyte cells[96] suggesting cell-type specific targeting of PRC1. As the only subunit containing a known chromatin targeting domain, the CBX subunit is proposed to be responsible for PRC1 targeting across the genome, and cell-type specific changes in CBX expression have been proposed to control PRC1 targeting in a cell-type specific manner. In ESCs, the predominant CBX paralog is CBX7, however, upon differentiation to MES cells, CBX2 replaces CBX7.[54] In mesoderm precursor (MES) cells, CBX2 localizes to over 2000 genes and approximately half of those genes are CBX7 targets in ESCs,[54] while the other half were MES-specific gene targets.[54] Genome-wide binding studies in different cell types have observed similar targeting changes. In both ESCs and the leukemia K562 cell line, PRC1 localized to similar gene targets; however, in the K562 cells there are additional PRC1 targets bound by CBX2 or CBX8 that are not bound by CBX7 in ESCs.[50] Not only are CBX targets shared between cell types but genome-wide binding studies in a fibroblast cell line indicate that the majority of the targets of CBX paralogs are overlapping as well.[50,63] For any of these systems, CBX paralogs do have unique gene targets as well, as defined from ChIP and knockdown studies (Figure 1.6).[50,52,54,63] To determine how CBX paralog expression dictates these targets, a recent study examining PRC1 composition and gene targets in ESCs and neural progenitor cells demonstrated that when CBX4 is expressed in ESCs, it targets PRC1 to CBX4-NPC target genes.[21] From these studies, it appears as though the CBX paralogs share many targets but also are able to localize PRC1 to unique targets. How CBX paralogs may mediate this recruitment to unique sites is still an active area of investigation.

The chromodomain, which binds H3K27me3 in Drosophila through Pc, has been identified as the primary chromatin-targeting domain for the PRC1 complex in mammals.[9,102,103,118] CBX does generally co-localize with H3K27me3 marks across the genome,[63] and mutation or deletion of the CBX chromodomain abrogates its ability to bind regions of H3K27me3 and localize PRC1 to specific gene targets.[9,58,64,102] In Drosophila, chromodomain swapping experiments with HP1 (recognizes H3K9me3) and Pc confirmed that the chromodomains were sufficient for the differential targeting observed for these two complexes;[9] however, whether or not this specificity extends to the mammalian CBX paralogs is still under debate. All of the PRC1 CBX paralogs, except CBX4 were found to localize to Xi heterochromatin in mice;[102] swapping the CBX2 and CBX4 chromodomains revealed that the chromodomain is required for this localization.[102] Whether the chromodomains alone could really be responsible for CBX paralog specific targeting seems unlikely, as the chromodomains have high sequence conservation (>80%) while other regions of the protein differ significantly; even the Pc boxes are only 45% similar. In spite of this similarity, soluble CBX chromodomains have varying affinities and specificities for histone peptides and not all the chromodomains prefer H3K27me3 *in vitro*.[102,103] CBX4 and CBX7 chromodomains consistently prefer H3K9me3 but also bind H3K27me3 (CBX4 K_d~ 70, 205 μM; CBX7 K_d~ 55, 110 μM, respectively), whereas CBX6 and CBX8 chromodomains bind weakly to both H3K9me3 and H3K27me3 (K_d > 500 μM).[102,103] *In vivo* studies, however, have identified higher overlap between CBX paralogs and H3K27me3 peaks throughout the genome, even for CBX6 and CBX8.[49,52,63] In fact, Klauke *et al*. observe greater enrichment of exogenous CBX8 compared to CBX7 using chromatin immunoprecipitation, proposing the possibility that CBX8 is actually more tightly associated with chromatin than CBX7 *in vivo*, in contrast to what the *in vitro* data with recombinant chromodomains would suggest.[58] Similarly, a proteomics approach to identify proteins that interact with H3K9me3 and H3K27me3 peptides identified all PRC1 subunits associating with H3K27me3, and CBX2,4,8, PHC2,3, BMI-1, RING1A also associating with H3K9me3,[119] suggesting that *in vitro* studies with soluble chromodomains may not provide a complete picture of PRC1-histone interactions. To support this, several studies have demonstrated that the chromodomain is not necessary for PRC1 association with histones or DNA, suggesting that other regions of CBX, as well as other PRC1 subunits,

may contribute to PRC1 binding affinity and targeting specificity.[97] While data suggest that CBX paralogs dictate specific PRC1 targets and that the highly homologous chromodomains are required for association with H3K27me3, it is still unclear the extent to which chromodomain binding contributes to the paralog specific targeting.

1.5.3 Transcription factor mediated localization

Another proposed mechanism for paralog-specific targeting is the association between PRC1 and transcription factors (Figure 1.6).[120–123] It is important to note that while PRC1 interacts with transcription factors that guide in its localization throughout the genome they only guide PRC1 to a subset of its binding locations.[124,125] Proteomic studies have identified numerous transcription factors that associate with PRC1 though additional studies are necessary to understand the precise role the transcription factors have in PRC1 targeting.[73,75] Transcription factors can recognize specific DNA sequences, and several of these transcription factors have been identified to recruit PRC1 to the DNA motif.[124–127]

An *in vitro* study examining the role of the Drosophila transcription factor Zeste revealed that when purified with PRC1, PRC1 was targeted to Zeste binding sequences and was more effective at inhibiting chromatin remodeling activity.[127] In mammals, however, for specific targeting of PRC1 paralogs, it is necessary to identify paralog/subunit-specific transcription factors. The PCGF paralog, BMI-1, was co-purified with the promyelocytic leukemia zinc finger (PLZF) transcription factor and could only bind *PLZF* sequences.[120] Interestingly, BMI-1 was incapable of binding to and repressing a *HOX* oligo *in vitro* in the absence of PLZF.[120] PLZF has been identified as a negative regulator of the *HOX* genes, but in the absence of BMI-1 it was unable to repress the *HOXD11* reporter.[120]

In ESCs, the CBX proteins as well as other PRC1 subunits were found to interact with the transcription factor REST, which is critical for neuronal development.[125,128,129] REST recognizes RE1 binding sites and was found to co-localize with PRC1 at both proximal and distal RE1 binding sites.[125] REST interaction with the CBX proteins is mediated through the N-terminus of both REST and CBX.[125] PRC1 localization to distal RE1 sites is dependent on its interaction with REST.[125] Interestingly, when REST and PRC1 localization is lost at distal regions, genes are de-repressed, however, the opposite occurs at proximal regions.[125]

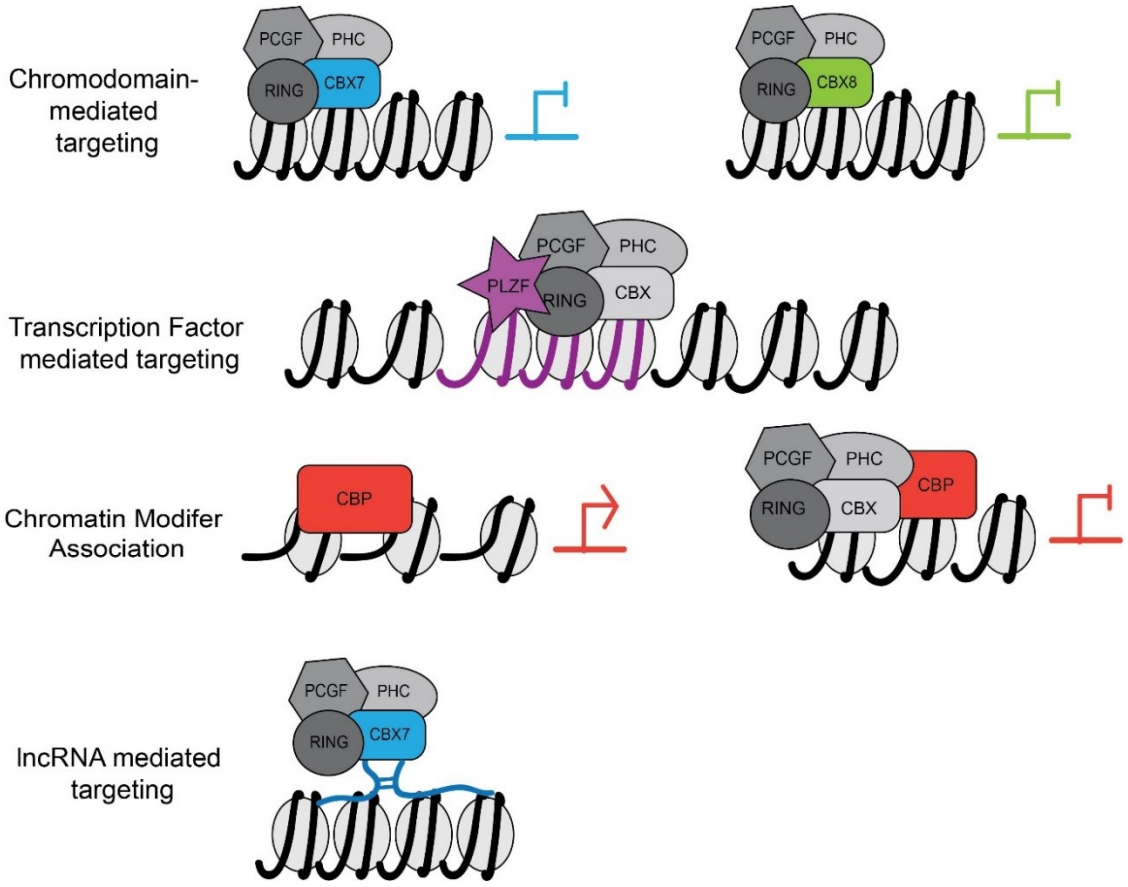


Figure 1.6 PRC1 targeting mechanisms.

PRC1 can be targeted in a paralog specific manner, e.g. CBX7has specific gene targets it recognizes that the other paralogs do not. PRC1 localization can be mediated by transcription factors to DNA with transcription factor binding motifs or through association with other chromatin modifiers. lncRNA can guide PRC1 to a particular gene or subset of genes.

Other proteomic studies in murine megakaryoblastic cells identified the core binding transcription factor RUNX1 as a novel protein interactor with BMI-1 and RING1B.[124] In a PRC2-independent manner, RUNX1 recruits PRC1 to genes allowing for H2AK119 ubiquitination and gene repression.[124] Interestingly, RUNX1 and PRC1 are essential for hematopoiesis.[58,130] Correspondingly, loss of *bmi-1* and *ring1b* in zebrafish phenocopy that of *runx1* deletion.[124]

1.5.4 Association with other chromatin modifiers

PRC1 subunits have been found to interact with a variety of other chromatin regulators to repress transcription.[121–123,131] In Drosophila, PRC1 interacts with Creb-binding Protein (CBP) through the Pc subunit in a histone-independent manner.[131] This association is conserved in mammals where the CBX proteins bind the histone acetyltransferase activity module of CBP to block histone H3K27 acetylation and gene activation.[131] CBX7 interacts with H3K9 methyltransferase SUV39H2 to induce H3K9me3 at sites of PRC1 localization.[132] Additionally, the CBX7 and CBX8 chromodomains bind to a lysine trimethylation site on the histone H3K9 methyltransferase, SETDB1, at its amino acid sequence FALK(me3)S.[103] The biological implications of this interaction are unknown; however, it has been hypothesized that PRC1 can be recruited to sites of gene repression via its interaction with SETDB1 rather than H3K27me3.[103]

CBX8 has also been potentially implicated in several unique interactions that result in gene activation.[122,123,133] ENL (MLLT1), a member of the EAP (elongation assisting proteins) complex, specifically binds CBX8 at a unique C-terminal and inhibits its ability to repress transcription.[122,133] CBX8 also binds the homologous EAP protein AF9 (MLLT3)[123] where it competes with DOT1L for binding. CBX8 sequestration of AF9 prevents H3K79 methylation by DOT1L, resulting in activation of *ENaCα*.[123] CBX8 has also been shown to interact with the MLL methyltransferase binding partner WDR5[51,121–123,133,134] in agreement with evidence that PRC1 is often observed with MLL at bivalent domains. Bivalent domains contain both repressive H3K27me3 as well as activating H3K4me3 and are typically associated with genes that are poised for activation. Surprisingly, this is not the only context in which CBX8 might actually play a transcriptionally activating role. During neuronal differentiation, CBX8 is required for the

activation of neuronal genes[53] but not the maintenance of neuronal gene transcription after differentiation. This unique association of CBX8 with gene activating chromatin-modifying complexes could provide a general mechanism by which PRC1-repressed genes can be marked for activation upon differentiation.

1.5.5 Chromosomal translocations alter PRC1 function

One type of leukemia, known as MLL (mixed lineage leukemia), has a characteristic chromosomal translocation that results in MLL fusion proteins, such as MLL-AF9 and MLL-ENL.[135] These fusion partners of MLL, as discussed above, can interact with CBX8.[121,122,133,134,136] Interestingly, the interaction with CBX8 drives the oncogenesis.[122,136] When CBX8 interacts with MLL-ENL, CBX8-mediated repression is lost resulting in hematopoietic cell transformation.[122] However, the capacity for CBX8 to drive transformation with MLL-ENL is highly dependent on the ratio of the two proteins. If CBX8 exceeds MLL-ENL, it will function normally to repress transcription and prevent transformation.[122] Similar results are observed when CBX8 interacts with the MLL-AF9 fusion protein; the fusion protein alone is not sufficient to induce transformation, CBX8 is necessary.[136]

PRC1 interaction with fusion proteins is not limited to MLL fusions. PLZF, which interacts with BMI-1, can form a fusion protein with RARA.[126] Similarly to MLL-AF9, PLZF-RARA cannot transform cells in the absence of BMI-1.[126] Because PLZF aids in the localization of PRC1 to gene targets, the fusion protein aberrantly recruits PRC1 to the transcription factor DNA binding motifs.[126]

1.5.6 RNA-mediated localization of PRC1

Long non-coding RNA (lncRNAs) assist in targeting a number of chromatin-modifying complexes.[137–139] While numerous lncRNAs have been identified for PRC2,[140–142] only a few have been identified for PRC1.[143–145] The chromodomains of CBX4, 6, 7, and 8 can non-specifically bind to single stranded RNAs.[102] CBX7 binds with the highest affinity and when treated with RNase, CBX7 could no longer bind chromatin.[102] Accordingly, thus far the only CBX-specific lncRNA identified, *ANRIL*, interacts with CBX7. *ANRIL* is an antisense lncRNA from the *INK4a/ARF* locus, a well-studied PRC1

gene target.[143,145] CBX7 specifically binds *ANRIL i*n addition to H3K27me3 to repress transcription of *INK4a/ARF* (Figure 1-6).[145] Loss of *ANRIL* results in a reduction of CBX7 at the locus, suggesting that *ANRIL* assists in CBX7 localization.[145] RNA binding has also been shown to physically alter the conformation of CBX7 chromodomain to permit association with trimethylated lysines,[146] which may contribute to RNA-mediated CBX7 localization. Recently a new lncRNA, *CAT7*, was identified to interact with PRC1 using a RNA immunoprecipitation screen. *CAT7* fine tunes PRC1 gene targeting, particularly during neuronal differentiation[144] and expression of human *CAT7* rescues a *cat7l* morpholino in zebra fish, suggesting that the two are functionally similar.[144] *CAT7* may also be involved in other cell types as Bmi-1 was found to interact with *cat7l* to repress the *ink4a/arf* locus in zebrafish.[144] As research continues, the discovery of new lncRNAs will shed light onto the specificity of PRC1 targeting.

1.6 Phosphorylation regulates PRC1 function

Phosphorylation plays a role in the regulation of PRC1 function and may contribute to cell-type specific paralog function. CBX2 can be phosphorylated at the Ser42 position within the chromodomain,[147] which alters the binding specificity of CBX2 *in vitro*. Phosphorylated CBX2 preferentially binds H3K9me3 whereas unphosphorylated CBX2 prefers H3K27me3.[147] Ser42 is conserved in the other CBX paralogs as Thr, which could possibly be phosphorylated, however, studies have not examined Thr phosphorylation in the chromodomain.[147] Additionally, CBX7 has been found to be phosphorylated near the Pc box (Thr118) via the Mitogen-Activated Protein Kinase pathway.[148] CBX7 phosphorylation improved its ability to be incorporated into the PRC1 complex, resulting in enhanced CBX7-mediated repression of the *INK4a/ARF* locus.[148] Phosphorylation of CBX2 and CBX7 is important in fine-tuning the function of PRC1 by altering their binding preference and affinity to RING1A/B, respectively.[147,148]

The CBX proteins are not the only subunit that can be phosphorylated. Both BMI-1 and MEL18 have been found to be phosphorylated, each with unique roles. BMI-1 phosphorylation occurs in a cell-cycle dependent manner. In G1/S phase, BMI-1 is hypophoshorylated and tightly bound to chromatin; however, in G2/M, BMI-1 is phosphorylated and dissociates from chromatin.[149,150] The 3pk kinase from the MAPK

family interacts with the PHC2 subunit and is responsible for BMI-1 phosphorylation, in addition to other PcG proteins.[150] Further, the PCGF paralog MEL18 can also be phosphorylated by protein kinase C,[106] preventing homodimerization, which can have a large impact on PRC1 stoichiometry and function.[106]

Phosphorylation is a signal from the surrounding environment and is dependent on a multitude of kinases and phosphatases. Being regulated by phosphorylation is one way that PRC1 can adapt to different cellular signals, such as stress. The apparent opposing phenotypes sometimes reported for individual paralogs could reflect these environmental mechanisms in place for activation or inactivation. For example, paralogs that compete for a genomic binding site could display differing transcriptional activities depending on the environment, further contributing to the complexity and diversity of PRC1. Further characterization of these modifications is critical in defining the different functions of PRC1.

1.7 Conclusion

Gene regulation in mammals is an intricate process with numerous mechanisms working together to regulate cell-type transcriptional programs, while also allowing for a certain amount of plasticity in order to regulate cell-type changes in response to stimuli. While many chromatin regulators have homologs in unicellular organisms, PcG proteins emerged upon the advent of multicellularity and further expanded during vertebrate evolution. The expansion of Polycomb genes in vertebrates is more pronounced for PRC1 genes, adding two to six paralogs for each subunit but only one additional paralog for the PRC2 methyltransferase. The advent of PRC1 paralogs with increasing organismal complexity and genome size suggests that PRC1 plays a crucial role in proper development and lineage specification. Unsurprisingly, genetic studies of PRC1 in vertebrates continually support this hypothesis. Without PRC1 proteins, embryos survive to birth but have severe skeletal deformities regardless of which PRC1 subunit is absent.[7,41,151–153] Furthermore, PRC1 subunits are critical for the maintenance and self-renewal capabilities of stem cells.[37,45,52,58,154] Based on these roles PRC1 plays, it is not surprising that PRC1 is implicated in a plethora of cancers (reviewed in [19,33]). Just as PRC1 has context-dependent functions, its misregulation in cancer is also context-dependent, with some subunits acting

as oncogenes and some as tumor suppressors. In the last decade, chromatin regulators, PRC1 included, have been at the forefront in the development of new drugs;[146,155–157] however, the development of these small molecules is reliant on understanding the biochemical functions of PRC1, which is still lacking.

With sixteen different paralogs, over a hundred unique PRC1 complexes can possibly be formed; however, we do not know the exact composition in every cell-type or what combinations actually exist. It is clear that different subunits have different functions in different cell types, which we do not understand how this is related to PRC1 targeting or transcriptional regulation. In addition, we are still working to understand how external factors, including posttranslational modifications, autoregulation, transcription factors, and lncRNAs further regulate PRC1 targeting and function.

This dissertation focuses on understanding CBX paralog diversity and complexities. Specifically, we demonstrate that inhibition of the CBX7 chromodomain enhances chemotherapeutic toxicity, while loss of CBX8 and its chromodomain are necessary for glioblastoma cell viability. Further, this work characterizes a new CBX8 chromodomain binding mechanism where the CBX8 chromodomain engages both DNA and H3K27me3 for full chromatin association.

CHAPTER 2. CBX CHROMODOMAIN INHIBITION ENAHNCES CHEMOTHERAPY RESPONSE IN GLIOBLASTOMA MULTIFORME

The following chapter is reproduced and modified with permission from Connelly, K.E., Martin, E.C., and Dykhuizen, E.C. *Yale J. Biol. Med.* 2016 Dec., 89(4), 431-440. © 2016, *Yale Journal of Biology and Medicine*

2.1 Introduction

Glioblastoma multiforme (GBM) is a malignant brain tumor that comprises the majority of all gliomas.[158] Due to GBM's highly aggressive nature, patients are left with a survival time of approximately twelve months.[159] Current treatments for GBM include surgery, radiation, and chemotherapy; however, numerous challenges exist in eradicating the tumor. Complete resection of the tumor is difficult due to the sticky finger-like morphology. Additionally, small molecules must be capable of permeating the blood brain barrier. Although the development of temozolomide (TMZ) in 2005 initially seemed promising, the five year survival rate for GBM patients has not improved.[160,161] It has been hypothesized that GBM develops resistance through its stem cell-like properties.[162,163]

Polycomb group (PcG) proteins have a fundamental role in the development and maintenance of adult stem-cells.[37,164,165] Known to serve as transcriptional repressors, PcG proteins are classified into two complexes known as Polycomb Repressive Complex 1 and 2 (PRC1/2). PRC2, via its catalytic subunit EZH2, is responsible for the trimethylation of histone H3 lysine 27 (H3K27me3). Consequently, PRC1 is recruited to H3K27me3 via its chromatin-organization modifier domain (chromodomain) where it ubiquitinates H2A lysine 119 resulting in chromatin compaction and gene repression.

The PRC complexes are composed of three subunits (PRC2) and four subunits (PRC1); however, gene duplications have resulted in numerous mutually exclusive paralogs for each subunit.[18] PRC2 is comprised of EZH1 or EZH2, EED, and SUZ12. PRC1, on the other hand, can be much more diverse and is comprised of RING E3 ubiquitin ligase (RING1A/B), Polycomb Group Finger (BMI-1 or MEL18), Polyhomeotic (PHC1-3), and the chromodomain-containing chromobox homolog (CBX2,4,6,7,8).

Numerous PcG proteins have been implicated in GBM progression and maintenance.[166,167] The PRC2 catalytic subunit EZH2 is overexpressed and is important in the development of GBM resistance.[166,168,169] The PRC1 subunit BMI-1 has been demonstrated to play a role in GBM self-renewal and promotes stem cell-like characteristics.[167] The chromobox homolog protein has several misregulated paralogs in GBM;[166] the CBX6 and CBX7 paralogs are downregulated compared to normal tissue, whereas the CBX8 paralog is upregulated. While phenotypical studies have demonstrated the importance of these CBX paralogs,[166] the mechanism in which CBX misregulation impacts GBM progression and maintenance is unknown.

Recently, it has been suggested that epigenetic processes may serve as good therapeutic targets.[170] Over the past decade, there has been an emergence of epigenetic inhibitors, including small molecules against the PcG proteins. These inhibitors function in different ways to derepress gene transcription and alter chromatin structure. For example, EZH2 inhibitors block the catalytic methyltransferase domain preventing H3K27 trimethylation. Current inhibitors against PRC1 include RING inhibitors that block histone ubiquitination, CBX inhibitors that prevent chromodomain binding, and inhibitors that block the transcription or incorporation of BMI-1.[155–157] None of these PcG inhibitors have been tested for efficacy against GBM cell lines.

Because of the limited therapies and low survival time, new therapeutic strategies need to be explored for glioblastoma, particularly to combat chemotherapy resistance. Previous studies have demonstrated that the knockdown of EZH2 and BMI-1 improves response to chemotherapies in a resistant cell population.[168,171] Here, we identify a novel therapeutic strategy to inhibit CBX chromodomain binding to improve GBM response to standard chemotherapy treatment.

2.2 Materials and Methods

2.2.1 Cell culture

U118MG, T98G, A172, and SVGp12 cells from ATCC were maintained in Eagle's Minimum Essential Medium (Corning) with 10% fetal bovine serum (Omega Scientific, Inc.), 1% penicillin-streptomycin (Corning), 1% non-essential amino acids (Corning), and

1% glutamine (Corning) at 37 °C and 5% CO_2. MDA-MB-231 cells were maintained in Dulbecco's Modified Essential Medium (Corning) with 10% fetal bovine serum (Omega Scientific, Inc), 1% penicillin-streptomycin (Corning), 1% sodium pyruvate (Corning), 1% glutamine (Corning) at 37 °C and 5% CO_2. Cells were plated in a 96 well plate at $2x10^3$ ($8.9x10^3$ for MDA-MB-231) cells/well 48 hours prior to treatment.

2.2.2 Drug Screen

Cells were treated 48 hours after plating. Cells were dosed in a grid format for every combination of drug at the designated dose: PRT4165 40 μM (Cayman), PTC209 200 nM (Cayman), DZnep 25 μM (Cayman), GSK343 400 nM (Cayman), MS37452 200 μM (Cayman), Doxorubicin 200 nM, temozolomide 50 μM (Cayman), SAHA 1 μM (Cayman). Control cells were treated with 1% DMSO or a single drug. Cells were treated for a total of five days but redosed with the same treatments after 48 hours. Following five days of treatment, cells were fixed with 50% trichloroacetic acid (TCA) for an hour at 4 °C, washed with water, incubated in sulforhodamine B for 10 minutes, washed with 1% acetic acid and dried overnight. Protein was solubilized with 10 mM tris and 515 nm absorbance readings were taken.

2.2.3 Dose response curves

U118MG, A172, SVGp12, and MDA-MB-231 cells were treated in a 96 well-plate with varying concentrations of MS37452 (0, 15.6, 31.25, 62.5, 125, 250 μM) in combination with doxorubicin (200 nM) or DMSO for five days or and with varying concentrations of doxorubicin with standard dose of MS37452 (250 μM) or DMSO. After five days, protein levels were measured with the sulforhodamine assay described above. Cells were plated at $1.3x10^4$ cells/well in a 24 well plate and treated with MS37452 (0, 31.25, 250 μM) and doxorubicin (100 nM) or DMSO in a similar format. Additional cells were plated at $2x10^3$ cells/well in a 96 well plate and treated with MS37452 (0, 31.25, 250 μM) and doxorubicin (50 nM) or DMSO. Cells were washed with PBS, trypsinized, harvested and counted.

2.2.4 Western Blot Analysis

Cells were treated as indicated above for five days in a 6 well plate. Cells were harvested and lysed with Buffer A (25 mM HEPES, 5 mM KCl, 25 mM $MgCl_2$, 0.05 mM EDTA,

10% glycerol, 0.1% NP-40, protease inhibitors) for 15 minutes on ice. Nuclei were pelleted and resuspended in RIPA buffer (50 mM tris, 150 mM NaCl, 0.1% SDS, 0.5% Na DOC, 1% triton-X, protease inhibitors, benzonase (Sigma) for 15 minutes. Lysates with LDS and BME loading buffer were boiled and ran on an SDS-page 4-12% gel (Invitrogen). Gels were transferred to PVDF membranes (Millipore) and exposed to 5% BSA in PBSt (0.1% tween-20). Membranes were blotted with primary antibodies overnight at 4°C. The blots were washed with PBSt and incubated for an hour at room temperature with goat anti-rabbit or mouse conjugated to IR800CW or IRDye 680 (LI-COR) secondary antibody. Blots were imaged on the LI-COR Odyssey. Antibodies used are rabbit CBX8 (Bethyl, A300-882A), rabbit CBX7 (Abcam, ab21873), rabbit cleaved PARP-1 (Cell Signaling Technology, 9541), mouse H3 (Active Motif, am-61475), and rabbit phospho-histone H2A.X (ser139) (Cell Signaling Technology, 9718).

2.2.5 Peptide pulldown

Nuclear lysate was made with untreated cells lysed with Buffer A for 15 minutes on ice. Nuclei were pelleted and resuspended in IP buffer (25 mM tris, 300 mM NaCl, 1 mM EDTA, 1%NP-40) for 15 minutes. 3 µg of biotinylated [Lys(Me3)27]-Histone H3 (21-44) and Histone H3 (21-44) (Eurogentec) were incubated with 10 µL of equilibrated streptavidin agarose resin (Solulink) (150 mM NaCl, 0.5 mM DTT, 50 mM tris, 1%NP-40) [119] at 4°C for an hour. Nuclear lysate was aliquoted and MS37452 was added to the lysate (0, 31.25, 250 µM). Lysate was divided and added to each peptide saturated resin and incubated at 4°C overnight. Resin was washed with equilibration buffer twice for ten minutes at 4°C. LDS with BME was added to the sample, boiled and ran on an SDS-page 4-12% gel (Invitrogen) as described above.

2.2.6 Flow cytometry

U118 cells were plated in a 6 well plate 24 hours prior to drug treatment. Cells were treated with DMSO or CBX7i (250 µM) every 48 hours for four days. On the fifth day of treatment, doxorubicin (100 nM) was added to the cells for 16 hours. Cells were fixed and permeabilized following the Click iT Plus EdU Alexa Fluor 488 Flow cytometry protocol (Life Technologies). FxCycle PI/RNase staining solution (Life Technologies) was used for

PI staining following the product manual. Cells were pelleted and resuspended in 300 µL of PBS and loaded on a 96-well plate. Unstained and single color controls were used to adjust the laser intensity of the Guava Easy-Cyte flow cytometer (Guava Technologies). 10,000 events were collected and data was processed and analyzed on FlowJo.

2.2.7 Immunofluorescence

U118MG cells were plated on glass coverslips in a 24-well plate 7 days prior to staining. Cells were treated with CBX7i (250 µM) every 48 hours following plating. On day five, cells were treated with 1 µM of doxorubicin for an hour. After an hour, doxorubicin was removed, and cells were allowed to recover for 0, 2, 4, 8, or 12 hours. Cells were fixed with 4% paraformaldehyde for 20 minutes. Fixed cells were stained with mouse phospho-H2A.X antibody (Millipore) overnight at 4 °C followed by secondary staining with a rabbit anti-mouse Alexa Fluor 594 antibody (Molecular Probes) for an hour at room temperature. The nuclei were stained with DAPI at 5µg/mL for 10 minutes. All cells were imaged at the same light intensity, brightness, and contrast at 40x magnification.

2.2.8 Data Analysis

Replicate absorbances (n=3) were averaged and normalized to the DMSO control (n=24) or to a single drug treatment (n=3) to generate a heat map was generated using RStudio. Red indicating low cell viability and blue indicating high cell viability. The mean corrected total cell fluorescence (CTCF) per cell was calculated using ImageJ and the equation CTCF = Integrated Density-(area of the cell x mean fluorescence background). Student's t-tests were performed on the 515 nm absorbance data, cell count, flow cytometry data, and mean CTCF for microscopy. Immunoblots were quantitated with ImageJ software. The intensity of each band was measured, normalized to the loading control (H3). Normalized data was used to calculate fold change compared to the DMSO control. Peptide immunoblots were quantitated in a similar manner and normalized to the H3K27me3 0 µM treatment.

2.3 Results

2.3.1 CBX7 chromodomain inhibition enhances chemotherapy toxicity

2.3.1.1 Identifying drug combinations enhancing cell response

Initially, we performed a drug screen with a wide array of epigenetic inhibitors and chemotherapies to identify therapeutic combinations that decrease GBM cell viability. In a grid format to ensure all possible combinations, we treated GBM U118MG cells with inhibitors against histone deacetylase complexes (SAHA), CBX7 chromodomain (MS3742),[155] BMI-1 (PTC-209),[157] RING1 ubiquitin ligase (PRT4165),[156] and histone methyltransferases (GSK343, specific for EZH2;[172] DZnep, inhibits global histone methylation).[173] Additionally, we included the chemotherapies, doxorubicin, a topoisomerase II inhibitor, and temozolomide, a DNA alkylating agent.[174] After treatment, bulk protein in the screen was stained with sulforhodamine B and the absorbance at 515 nm was taken and correlated to cell viability.

From our screen, we identified several combinations that resulted in consistently decreased cell viability compared to DMSO treated and single drug treatment:

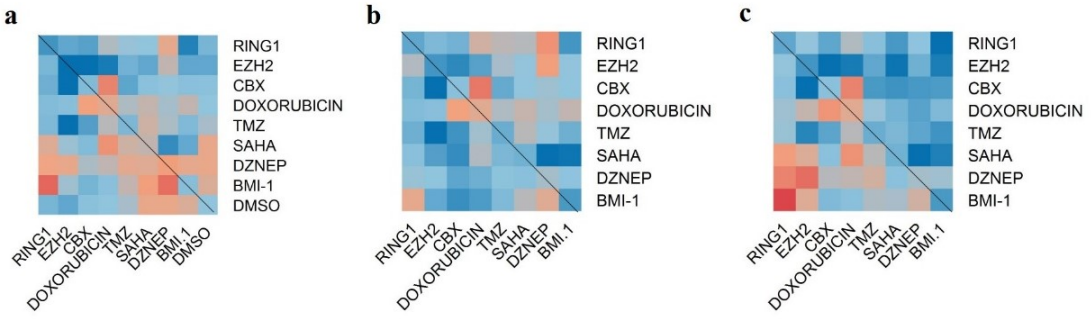


Figure 2.1 Select epigenetic inhibitors sensitize GBM U118MG cells to doxorubicin and temozolomide chemotherapies.

a) A heat map demonstrates the effect of a dual drug treatment on cell viability (n = 3, high viability (> 77 percent), blue; low viability (< 76 percent), red) compared to DMSO treated cells (n = 18). b,c) Heat maps represent the effect of a dual drug treatment compared to cells treated with a single drug. Each dual treatment was normalized to the single drug treatment indicated on the right column (b) or the drug indicated on the bottom (c). Heat maps in a-c were generated from sulforhodamine 515 nm assays to correlate bulk protein to cell viability.

SAHA/TMZ and MS37452/doxorubicin (Figure 2.1 a-c). It was necessary to normalize the absorbance values to both DMSO-treated and a single drug treatment. Normalization to a single drug treatment is important to ensure that the decrease in viability is a result of the combination of drugs, and not just one of the drugs. If it were the result of only one of the two drugs normalized viability would be 100%. The SAHA/TMZ combined treatment has been identified prior to our study, and clinical trials examining the effect SAHA and TMZ have together on GBM progression are in progress;[175] however, MS37452, which is a CBX7 chromodomain inhibitor (CBX7i), in combination with doxorubicin is a novel therapeutic strategy.

2.3.1.2 CBX7 chromodomain inhibition with doxorubicin decreases viability

We subjected U118MG cells to varying concentrations of CBX7i while the doxorubicin concentration remained constant (200 nM) for five days. Following the five-day treatment regimen, we used a sulforhodamine assay to measure bulk protein adhered to the plate, which should correlate to cell viability. Interestingly, we observed a slight increase in cell viability with increasing concentrations of CBX7i; however, at high concentrations of CBX7i with doxorubicin, cell viability was significantly decreased compared to CBX7i or doxorubicin only treatments (Figure 2.2a). Unsurprisingly, when we treated cells with varying concentrations of doxorubicin and kept CBX7i concentration constant (250 μM), we observe a similar response to the combinatorial treatment (Figure 2.2b).

The sulforhodamine assay is limited to measuring bulk protein, which can remain adhered to the plate after cell death, so we also measured viability by counting live cells. Under the same conditions, we observe a similar trend to the sulforhodamine assay, confirming our results (Figure 2.2c). Finally, at minimal doxorubicin (50 nM), we still observe a significant decrease in cell viability (Figure 2.2d). This suggests that dual therapeutic strategy for GBM will require lower doses of doxorubicin, potentially minimizing the toxic side effects of chemotherapy.

GBM tumors, however, are highly heterogeneous, thus it is important to determine if this phenomenon is observed in other GBM derived cells. To answer this question, we performed our sulforhodamine cell viability assay with the A172 GBM cell line. As

expected, the dual treatment increases doxorubicin toxicity in A172 cells (Figure 2.3a). Interestingly, unlike the U118MG cell lines, the A172 cells displayed sensitivity to the CBX7i (Figure 2.3b). Because of the observed sensitivity and decrease in cell viability in the presence of only the CBX7i, it is difficult to interpret if the combined effect is a result of increased toxicity to doxorubicin.

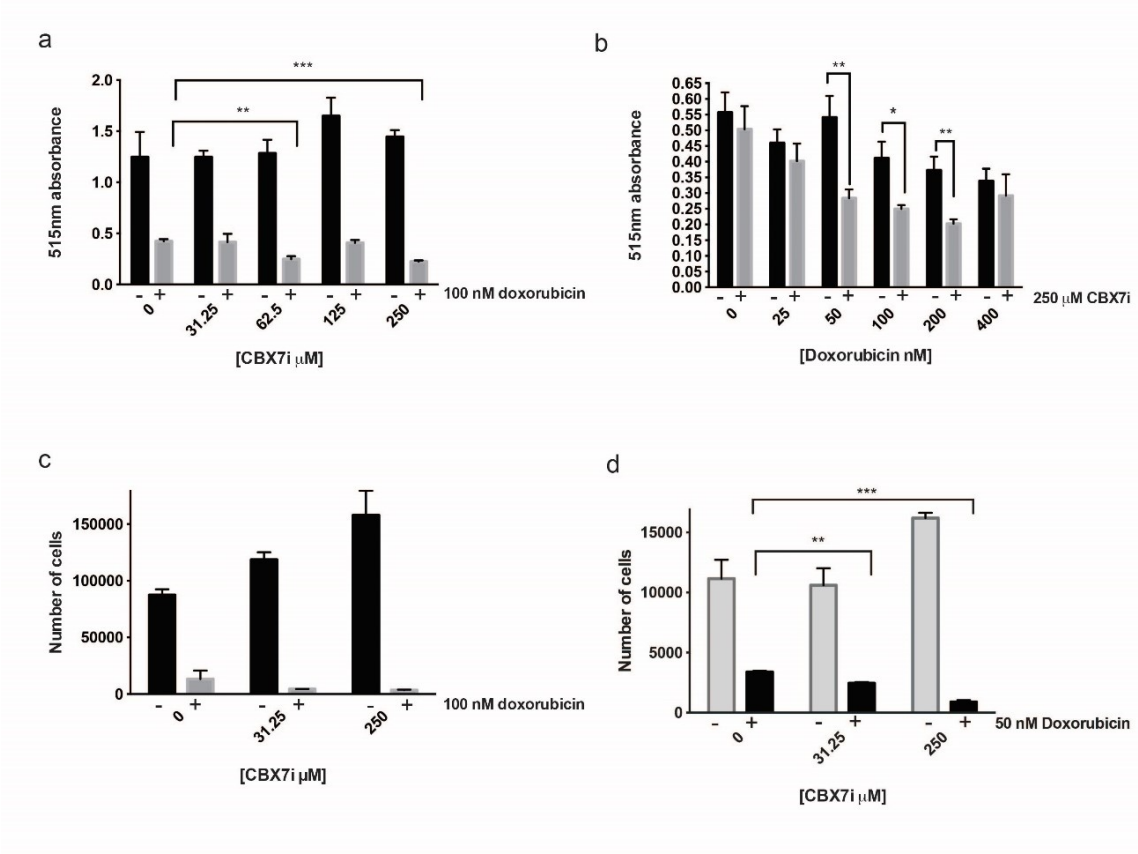


Figure 2.2 CBX7i with doxorubicin decreases cell viability.

a) U118MG cells treated with CBX7i in the presence (grey) or absence (black) of 100 nM doxorubicin total protein measured by sulforhodamine B 515 nm absorbance (n = 3) b) U118MG cells treated with doxorubicin cells treated with doxorubicin in the presence (grey) or absence (black) of 250 μM CBX7i bulk protein measured by sulforhodamine B 515 nm absorbance (n = 3) c) U118MG cells were treated with CBX7i in the presence (grey) or absence (black) of 100 nM doxorubicin and counted (n = 3) d) U118MG cells in a 96 well plate were treated CBX7i in the presence (grey) or absence (black) of 50 nM doxorubicin for five days and counted (n = 3) Data in a-d represented by mean ± SEM, p-values calculated with student's t-test: (*) < 0.05, p (**) < 0.01, p (***) < 0.001.

In glioblastoma, CBX7 is downregulated in over 80% of patients compared to normal brain tissue.[166] Thus, we found it important to understand the impact of the CBX7i on non-tumorigenic cells. Using our sulforhodamine assay with the SVGp12 astrocyte cell line, we observed that the astrocytes were not sensitive to the CBX7i alone (Figure 2.3d). As an intercalator and topoisomerase 2 inhibitor, doxorubicin primarily targets cycling cells. Since the SVGp12 cell line is highly proliferative, doxorubicin was, not surprisingly, toxic to the cells (Figure 2.3c). Similar to the GBM cell lines, this doxorubicin toxicity was exacerbated in the presence of CBX7i.

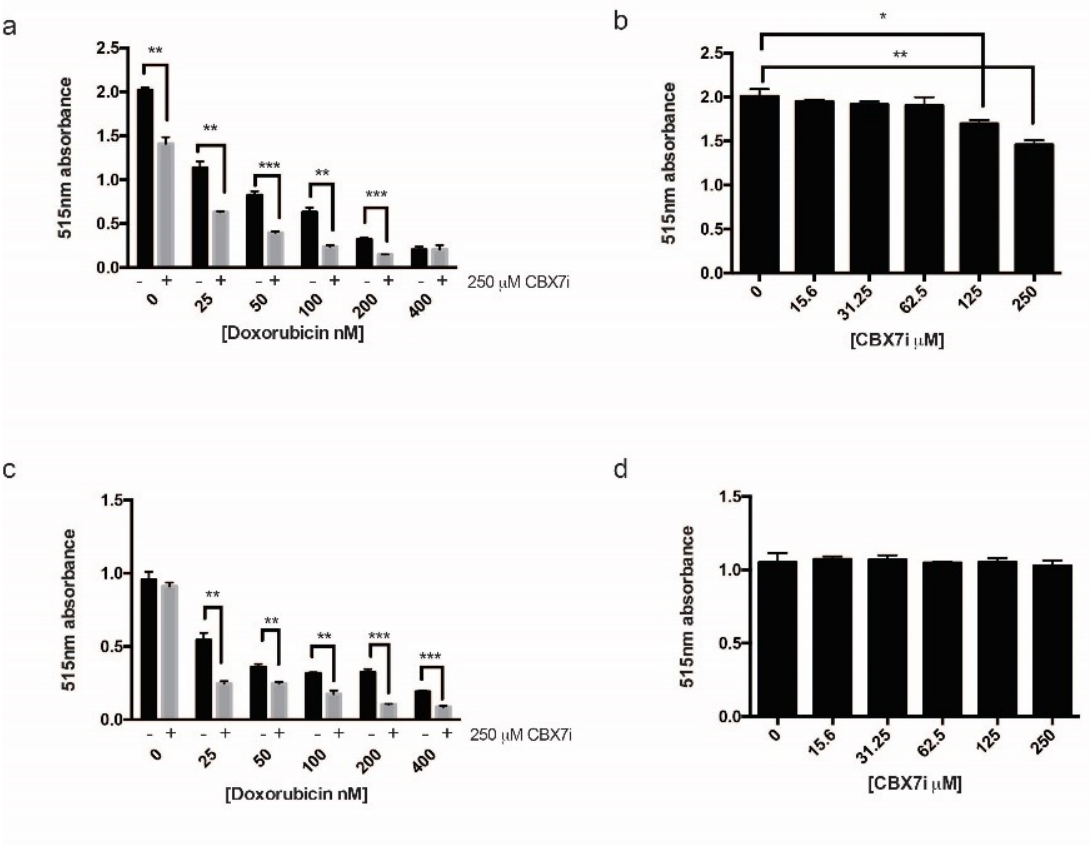


Figure 2.3 A172 GBM cells are sensitive to CBX7i, but astrocytes are not.

a) A172 cells treated with doxorubicin in the presence (grey) or absence (black) of 250 μM CBX7i bulk protein measured by sulforhodamine B 515 nm absorbance (n = 3) b. A172 cells were treated with CBX7i and bulk protein measured by sulforhodamine B 515 nm absorbance (n = 3) c. SVGp12 cells treated with doxorubicin in the presence (grey) or absence (black) of 250 μM CBX7i bulk protein measured by sulforhodamine B 515 nm absorbance (n = 3) d. SVGp12 cells were treated with CBX7i and bulk protein measured by sulforhodamine B 515 nm absorbance (n = 3) Data in a-d represented as mean ± SEM, p-values calculated by student's t-test: p (*) < 0.05, p (**) < 0.01, p (***) < 0.001.

The A172 cells' sensitivity to CBX7i is interesting and could be a result of the heterogeneity of GBM and the diversity of PRC1 composition. While CBX7 expression is generally low across GBM patients, it is possible that A172 cell lines have higher expression of CBX7 and are thus more sensitive to the CBX7i. The response of A172 cells also suggests that CBX7 may serve as a therapeutic target for a subset of glioblastomas that are reliant on CBX7 activity.

2.3.1.3 CBX7i disrupts CBX7 binding to H3K27me3 *in vitro*

To confirm the specificity of chromodomain inhibition, we performed peptide pulldown studies with unmethylated H3 and tri-methylated H3K27 peptides in presence of the inhibitor. The CBX inhibitor has previously been published to be specific for the CBX7 paralog chromodomain with a K_d of 28.9 μM.[155] As expected, we observed a decrease in CBX7 binding to the methylated peptide in the presence of the inhibitor, particularly at 250 μM (Figure 2.4a, b). CBX8 binding, however, in the presence of 250 μM was not affected (Figure 2.4a, b). The pulldown suggests that the effect seen at 250 μM is not due to inhibition of CBX8 chromodomain binding; however, inhibition of CBX4 (K_d of 95.8 μM) cannot be ruled out as our highest concentration used is 1.5-fold higher than the CBX4 K_d.

2.3.1.4 CBX7 inhibition increases doxorubicin-induced DNA damage and apoptosis

Doxorubicin acts through DNA intercalation and causes double stranded DNA breaks by inhibiting the re-ligation step of topoisomerase II.[176] To understand the role of CBX7i in enhancing doxorubicin response, we examined the protein expression level of DNA damage and apoptosis markers by immunoblot. When cells were treated with doxorubicin only (100 nM), we observe an induction of γH2A.X indicating DNA damage; however, in the presence of 250 μM CBX7i and doxorubicin, we observe a further 1.5-fold increase in γH2A.X induction (Figure 2.5a, c).

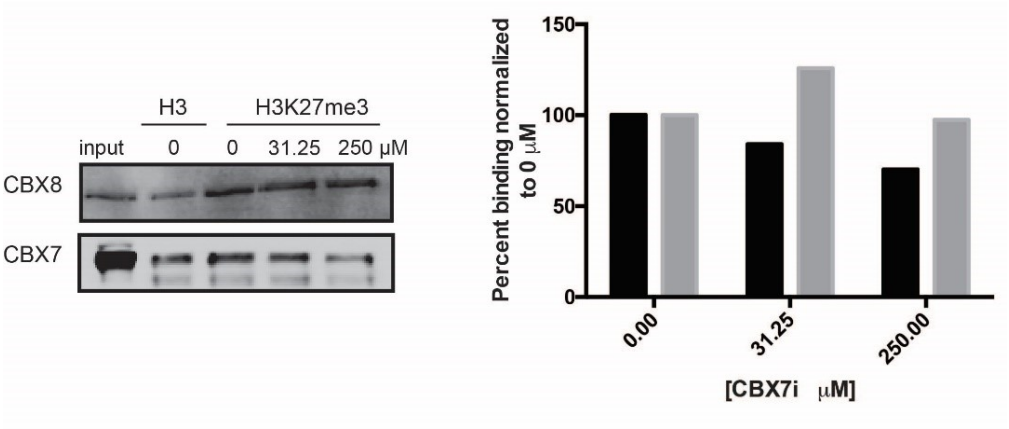


Figure 2.4 CBX7i disrupts CBX7 binding to H3K27me3.

Pulldowns with unmethylated H3 or H3K27me3 peptides confirm that increasing concentrations block CBX7 (T98G cells) binding but not CBX8 (U118MG cells). b. Quantitation of immunoblots normalized to H3K27me3 0 μM treatment. CBX7, black; CBX8, grey.

If DNA damage is not repaired, cells will undergo apoptosis. In order to determine if the dual treatment augments apoptosis, we examined cleaved PARP-1 levels. When cells undergo apoptosis, caspases cleave PARP-1, rendering it inactive. Our immunoblot analysis reveals that both doxorubicin and the combinatorial treated cells undergo apoptosis. Nevertheless, the CBX7i/doxorubicin combination increases cleaved PARP-1 two-fold compared to doxorubicin only (Figure 2.5a, b). This suggests that CBX7i enhances the impact of doxorubicin on DNA damage and apoptosis.

2.3.1.5 CBX7i and doxorubicin induce a G2/M block

We performed flow cytometry to examine the differences in cell cycle amongst the different treatments. Following a four-day treatment of either CBX7i or DMSO, we pulsed cells with 100 nM of doxorubicin for 16 hours and stained cells with Propidium Iodide (PI) for cell cycle analysis. Unsurprisingly, our DMSO and CBX7i treated cells had similar cell cycle phase distributions with a majority of the cell population in G1 and a smaller percent of cells in G2/M (Figure 2.5d) When we treated cells for 16h with a lower dose of doxorubicin (100 nM), we did not observe an increase in G2/M;[177] however, in accordance with an increase in DNA damage and apoptosis, the dual drug treatment increased the number of cells in G2/M 1.5-fold. (Figure 2.5d).

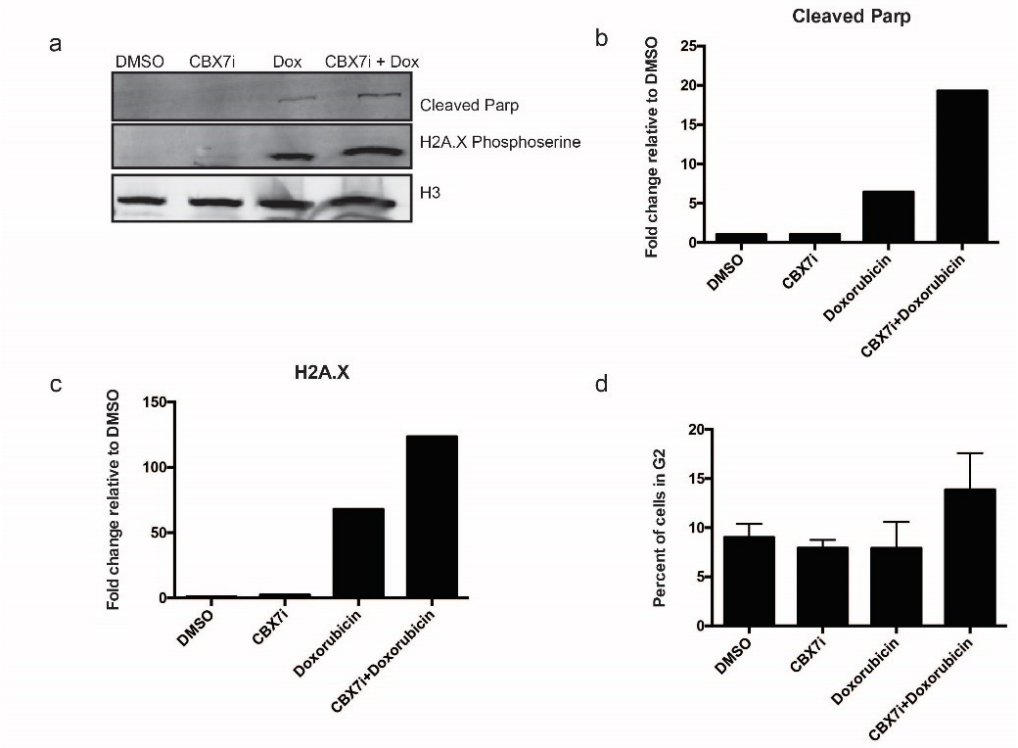


Figure 2.5 CBX7i and doxorubicin treatment increases DNA damage and apoptosis.

a) Western blot analysis of nuclear lysate staining for H2A.X serine 139 phosphorylation, cleaved PARP-1, and histone H3 (loading control) b. Quantitation of cleaved PARP-1 normalized to H3 loading control, fold change relative to DMSO-treated protein levels c. Quantitation of H2A.X normalized to H3 loading control, fold change relative to DMSO-treated protein levels d. Cell cycle analysis of cells treated with DMSO, CBX7i (250 μM), doxorubicin (100 nM) or CBX7i and doxorubicin (250 μM, 100 nM respectively), percent of cells in G2 plotted as a percent of total gated cells.

2.3.1.6 Presence of CBX7i prevents DNA damage repair

There are several potential mechanisms in which CBX inhibition can promote DNA damage in doxorubicin treated cells. Chromatin modulators, like the PcG proteins, are responsible for altering chromatin structure and thus DNA accessibility. It has been suggested that open chromatin is more susceptible to DNA damaging agents.[178] It is possible that inhibiting the CBX proteins allow for more open chromatin, as they can no longer be recruited to histones to compact chromatin and repress transcription. Additionally, previous studies have demonstrated that the recruitment of PcG proteins,

including the CBXs, to sites of DNA damage are important for the localization of other DNA damage machinery and subsequent DNA damage repair.[179,180]

To understand which mechanism may be contributing to the increased sensitivity to doxorubicin with CBX7i, we used H2A.X staining to investigate DNA damage repair after acute doxorubicin treatment (Figure 2.6). Immediately after doxorubicin treatment there was no difference in DNA damage with or without CBX7i (Figure 2.6). As cells had time for damage to accumulate and begin recovery, however, we observed a significant increase in DNA damage accumulation in the presence of CBX7i beginning at two hours (Figure 2.6). This increase in DNA damage was drastically extended as recovery time progressed (Figure 2.6). Together, our data suggests that treatment of GBM cells with a CBX7 chromodomain inhibitor improves the response to the chemotherapeutic doxorubicin by preventing DNA damage repair, although these results are still preliminary. Based on these data, we hypothesize that CBX7 chromodomain inhibition prevents DNA repair, resulting in a G2/M block and eventual apoptosis. Future studies will be needed to investigate this mechanism further.

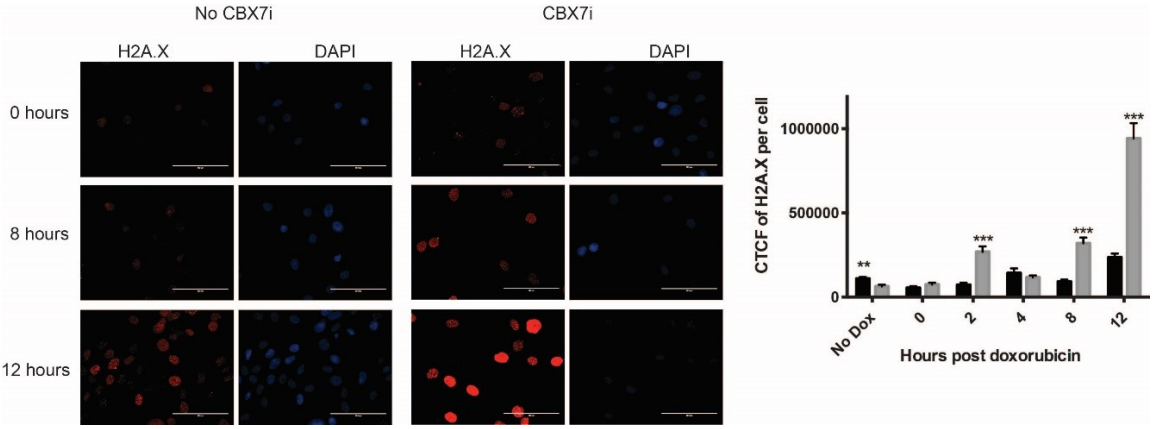


Figure 2.6 CBX inhibition prevents DNA damage repair.

U118MG cells were treated for four days in the presence or absence of CBX7i. On day five, cells were treated with doxorubicin (1 μM) for one hour. After the treatment, doxorubicin was removed and cells were allowed to recovery for either 0, 2, 4, 8, or 12 hours. If cells were treated with CBX7i, they remained on CBX7i throughout the entire process. Following recovery, cells were fixed and stained for phospho-H2A.X and DAPI. The cells were imaged and the mean total corrected cell fluorescence (CTCF) per cell was measured for each treatment type. Mean fluorescence per cell was quantitated and plotted (grey, CBX7i treatment; black, no CBX7i treatment) (n = 77, 111, 87, 64, 66, 64, 65, 80, 43, 78, 72, 88, left to right) Data represented as mean ± SEM, p-values calculated by student's t-test: p (**) < 0.01, p (***) < 0.001.

2.3.1.7 CBX7 inhibition increases doxorubicin toxicity in breast cancer

Doxorubicin is not a standard treatment for glioblastoma due to its inability to cross the blood brain barrier. However, it is often used in other cancers such as breast, lung, and leukemia.[176] In order to verify that this dual treatment is applicable to these cancers, we tested the combination of CBX7i and doxorubicin in the breast cancer MDA-MB-231 cell line. Consistent with our findings in GBM, the dual therapy enhances the toxicity of doxorubicin (Figure 2.7).

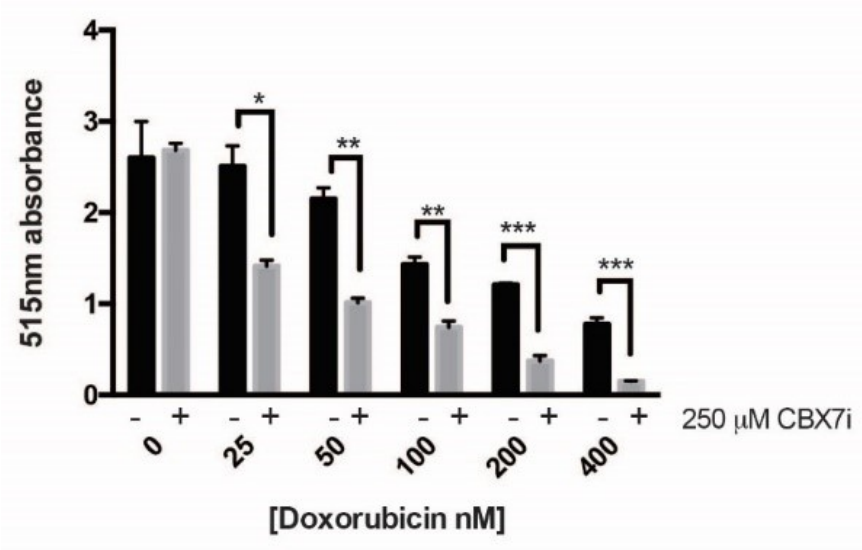


Figure 2.7 CBX inhibition enhances doxorubicin toxicity in breast cancer.

MDA-MB-231 cells were treated with a range of doxorubicin concentrations (0-400 nM) in the presence (grey) or absence (black) of 250 μM of CBX7i; total protein was measured by sulforhodamine B 515 nm absorbance (n = 3) Data represented as mean ± SEM, p-values calculated by student's t-test: p (*) < 0.05, p (**) < 0.01, p (***) < 0.001.

2.4 Discussion

With little improvement in glioblastoma survival over the last decade, it is critical to develop new therapeutic strategies. Although the exact role PcG proteins play in GBM progression and maintenance is unclear, their involvement is necessary.[162,163,166–169,171] In this study, we identified two therapeutic strategies, inhibition of CBX7 chromodomain binding and histone deacetylase activity, to improve GBM response to traditional chemotherapies. Although we saw a sensitivity to TMZ with the HDAC inhibitor SAHA,

there are already currently ongoing clinical trials with the drug combination[175] and previous *in vitro* studies have investigated potential mechanisms.[181–183] Therefore, we investigated the completely novel approach of inhibiting the CBX7 chromodomain to improve chemotherapeutic response to doxorubicin.

Our findings indicate that inhibition of the CBX7 chromodomain drastically increases DNA damage in response to doxorubicin. Our cell cycle analysis data indicates a G2/M block suggesting that damage cannot be properly repaired without CBX7.

2.4.1 PcG proteins in DNA damage response

Roles for PcG proteins in DNA damage response (DDR) have previously been identified. The majority of PcG proteins, including EZH2 and the CBXs, localize to regions of damaged chromatin in order to recruit DNA damage repair machinery.[179,180] However, efforts to inhibit PRC1 subunits with small molecule inhibitors, particularly with a focus on DNA damage, have been limited. We hypothesized that the inhibition of CBX proteins prevents PcG and other DNA damage repair machinery recruitment to DNA damage. Consequently, DNA damage accumulates, and the cells undergo apoptosis. Initial treatment of doxorubicin in both the presence and absence of CBX7i induced similar levels of H2A.X, suggesting that CBX7i is not improving response to chemotherapy by altering DNA accessibility. However, as the cells are allowed to recover from the initial DNA damage, the total DNA damage drastically accumulates, whereas cells with only doxorubicin have significantly less DNA damage. Our data suggests that inhibiting CBX enhances chemotherapeutic response by preventing DNA damage repair, allowing massive accumulation of DNA damage. Although CBX7 has been shown to localize to sites of DNA damage, its role in the process is unknown.[179] CBX4, however, has been studied and shown to be an important part of the DNA damage response.[91,184] These studies have demonstrated that CBX4 is recruited early to sites of DNA damage, and loss of CBX4 extenuates the presence of DNA damage.[182,183] While we used an inhibitor specific for CBX7, the concentration used in our experiments exceeded the K_d for CBX4.[155] Thus, it is still a possibility that the drastic increase in DNA damage in the presence of CBX7i and doxorubicin is a result of CBX4 inhibition. In addition, we cannot eliminate the possibility

that inhibition of CBX7 is playing a transcriptional role in improving the response to chemotherapy. Further studies are necessary to fully dissect the mechanism.

2.4.2 Future potential of CBX inhibitors

CBX7i is a first generation chromodomain small molecule inhibitor.[155] Although the CBX7i is not a potent small molecule for the clinic, it is does demonstrate very useful properties. It serves as a useful tool to study the biochemistry of the CBX proteins, particularly CBX7. Additional studies with the recently developed second generation CBX7 inhibitor, as well as the recently developed CBX4/7 inhibitor will be interesting to further understand if the effect observed in these studies can be extended to different CBX inhibitors.[146,185]

While we identified a novel therapeutic strategy that results in cell death *in vitro*, doxorubicin is not currently a viable chemotherapy for glioblastoma as it does not cross the blood brain barrier; however, understanding how inhibition of the CBX proteins can improve response to DNA damaging agents is an important area of research. DNA damaging agents are still the most clinically used chemotherapies, and we have demonstrated that CBX inhibition presents itself as a promising strategy for other cancers such as breast cancer where doxorubicin is a first line of treatment.[176] Finally, there is ongoing research to develop drug carriers that will improve blood-brain barrier penetrance, so that eventually drugs, like doxorubicin can be used in glioblastoma treatment.[186,187] Additionally, more in depth studies examining the role of CBX inhibition in combination with radiation treatment or TMZ may reveal similar enhanced toxicities.

Identifying therapeutic strategies as described in this paper not only will provide us new drug targets for cancers but will also allow clinicians to reduce the amount of chemotherapy necessary to kill cancer cells. This will reduce some of the toxic side effects of the anthracyclines. Though many studies need to be completed, the idea of targeting CBX proteins in combination with current treatments is novel and has great potential.

CHAPTER 3. CHARACTERIZING CBX8 IN GLIOBLASTOMA MULTIFORME

3.1 Introduction

Chromatin regulators are critical in the pathogenesis of cancer. Alterations to chromatin structure can result in genomic instability and misregulation of gene expression leading to cellular transformation and oncogenesis.[188,189] The first Polycomb Group protein identified as a proto-oncogene was BMI-1 (PCGF4).[190] It has been proposed that BMI-1 inhibits MYC-induced apoptosis through the repression of the *INK4A/ARF* locus that encodes the tumor suppressors p16 and p14.[191] Subsequently, additional studies identified a role for other PRC1 subunits at the *INK4A/ARF* locus to bypass senescence.[45,64,66] These studies were the first evidence of PRC1 cell cycle regulation in oncogenesis. However, repression of tumor suppressor genes is not the only plausible mechanism for PRC1 in oncogenesis. PRC1's role in stem cell maintenance and differentiation suggests that misregulation of PRC1 may redirect cells to a stem cell like state.[37] Unlike the mSWI/SNF complexes that are mutated in approximately 20% of cancers, canonical PRC1 (cPRC1) complexes are rarely mutated (~2%)[27] but rather are either upregulated or downregulated, generally at the transcript level.[192]

3.1.1 Glioblastoma multiforme

Glioblastoma multiforme (GBM), a grade IV glioma, is the deadliest form of brain cancer. Approximately 90% of all GBM cases are *de novo* primary tumors that aggressively infiltrate the brain making complete tumor resection difficult.[193] GBM patient survival is approximately 12 to 14 months following diagnosis. The tumor heterogeneity and blood brain barrier have posed challenges for therapeutic development leaving patients with limited treatment options that only improve survival by a few months. Currently, these treatment options are surgery, radiation, chemotherapy with TMZ, and/or a combination of these.[170]

As with many cancers, GBM tumors have a subpopulation of cells known as cancer stem cells (CSCs).[194] The Polycomb Group proteins are known for their role in stem cell maintenance,[195] thus it is not surprising that this family of proteins is important for neural

cancer stem cell self-renewal.[59,167,196] Furthermore, an assessment of Polycomb target expression in GBM revealed a global decrease compared to normal brain.[197] This finding is in accordance with the overexpression of EZH2, the H3K27 methyltransferase.[166,198] To date, there has been a lot of work understanding EZH2's role in GBM, and EZH2 inhibitors are currently in clinical trials for a variety of cancers. However, because EZH2 is the predominant H3K27me3 methyltransferase in every cell there are likely to be unanticipated side effects over time. A 2010 glioma gene expression study identified 71 genes that can discriminate between low grade (grades I and II) and high grade (grades III and IV) gliomas.[199] The PcG protein CBX8 was in the top ten genes for discriminatory power while only exhibiting a 1.5-fold change in gene expression between high and low grade gliomas.[199] Furthermore, Li et al., interrogated the Cancer Genome Atlas (TCGA) and found that CBX8 was overexpressed in over half of GBM patient tumors.[166] Paralog-specific inhibition of the CBX proteins is less likely to exhibit undesired effects as genome-wide studies suggest high gene target overlap and minimal post-developmental knockout phenotypes.[63,200]

3.1.2 CBX8 in cancer

Over the last several years, CBX8 has been found to be misregulated in a handful of cancers, including glioblastoma multiforme,[166] leukemia,[136] esophageal squamous cell carcinoma,[201,202] breast cancer,[51,203] and hepatocellular carcinoma.[204] To date, studies focused on understanding CBX8's role in oncogenesis have led to a variety of canonical and non-canonical mechanisms.

3.1.2.1 Canonical Roles of CBX8 in cancer

Several studies have detailed canonical CBX8 mechanisms in oncogenesis and metastasis. In both bladder cancer and colorectal cancer, high CBX8 expression correlates to a poor prognosis, and CBX8 knockdown reduces cell viability and tumor size.[205,206] These studies propose that CBX8 represses p53 expression to promote cell proliferation.[205,206] Interestingly, additional studies in colorectal and esophageal cancers outline CBX8 mechanisms in metastasis. Unlike CBX8's role in proliferation, CBX8 expression reduces cell migration and metastasis, indicating a paradoxical role for CBX8

in cancer.[202,206] The proposed molecular mechanisms suggest CBX8 represses genes critical for cell motility and epithelial to mesenchymal transition (EMT).[202,206]

3.1.2.2 Non-canonical roles of CBX8 in cancer

An early study in acute myeloid leukemias demonstrated that CBX8 interacts with the MLL-AF9 fusion protein, independent of PRC1, to drive hematopoietic stem cell transformation.[136] Quite interestingly, a subsequent study examined the role of CBX8 with the fusion protein MLL-ENL and demonstrated that CBX8's interaction with ENL, a homolog of AF9, is necessary for hematopoietic transformation.[122] The proposed mechanisms in which CBX8 drives hematopoietic transformation are unique. CBX8's interaction with MLL-AF9 allows for the recruitment of transcriptional activators to drive leukemogenesis.[136] On the other hand, Maethner et al., proposed a mechanism where MLL-ENL sequesters CBX8 to prevent Polycomb-mediated gene repression.[122] More recently, a non-canonical role for CBX8 in mammary tumorigenesis has been described where CBX8 interacts with the H3K4 methyltransferase subunit WDR5 to maintain the expression of *NOTCH* and the stem cell-like transcriptional program.[51]

In addition to transcriptional functions, CBX8 has been implicated in the DNA damage repair process. CBX8 and other PcG proteins are recruited to sites of DNA damage through PARP-dependent mechanisms.[179,207] These studies have demonstrated that loss of CBX8 increases the cells' susceptibility to DNA damage[201,207] and reduces cell viability.[207] Together, these studies suggest a role for CBX8 in sensitizing cancer cells to DNA damaging agents, such as chemotherapy. To understand the role of CBX8 in GBM, we utilized a variety of approaches including viability studies, transcriptional analysis, and proteomics.

3.2 Materials and Methods

3.2.1 Cell culture

All GBM cell lines (T98G, U118MG, U87MG, A172, U138MG) were obtained from ATCC. Cells were grown in Eagles Minimum Essential Medium (Corning), 10% Fetal Bovine Serum (Omega Scientific, JR Scientific), and 1% of the following: penicillin/streptomycin, sodium pyruvate, non-essential amino acids, and glutamine

(Corning). HEK293T cells for lentivirus production were grown in Dubecco's Modified Essential Media, 10% FBS, 1% penicillin/streptomycin, glutamine, sodium pyruvate. All cell lines were grown at 37 °C in 5% CO_2.

3.2.2 Lentivirus production

HEK293T cells were transfected with shRNAs, TetOFUW constructs, or pLenti CMV rtTA3 Hygro (w785-1) (a gift from Eric Campeau Addgene plasmid # 26730), and the viral packaging vectors (pMD2.G and psPAX2). Media was replaced 16 hours following transfection. Seventy-two hours following transfection, viral supernatant was harvested, filtered, and concentrated by ultracentrifugation at 17,300 rpm for 2 hours. Virus was resuspended in PBS. ShRNA constructs were added directly to cells, while TetOFUW and rtTA were co-added to T98G cells for doxycycline inducible cell lines. Following, plates were spun at room temperature for 1 hr at 200 x g. Forty-eight hours after infection, cells were selected with respective antibiotics: puromycin (2 µg/mL) and hygromycin (200 µg/mL) for a week to generate stable cell lines. For inducible cell lines, doxycycline was added every 48 hrs at 2 µg/mL to induce expression. CRISPR cell lines were generated using the px459 v2.0 vector (a gift from Feng Zhang, Addgene plasmid #62988)[208]. Briefly, 200,000 T98G cells were seeded 24 hr prior to transfection. The vector was transfected with Fugene 6 (Promega). Media was changed 24 hrs post-transfection and cells underwent puromycin selection (2 µg/mL) for 3 days, 48 hrs post-transfection.

3.2.3 Growth Curves

GBM cells for growth curves were plated for lentiviral infection. Cells were grown for seven days and counted.

3.2.4 Tumorsphere Formation

Six-well plates were coated with polyhema and dried at least 24 hours prior to cell seeding. T98G cells were trypsinized and harvested as a single cell suspension. Twenty thousand cells were plated and allowed to grow for a week, with additional media added on day 4. Following seven days, tumorspheres were harvested, washed with PBS, and trypsinized into a single cell suspension, counted, and replated at a density of 20,000 cells. This process

was repeated over several weeks. Cell number after each week was normalized to the sgControl tumorspheres.

3.2.5 Ammonium Sulfate Protein Precipitation

T98G cells were grown to confluency. Cells were harvested, washed with PBS, and homogenized with Buffer A (25 mM HEPES pH 7.6, 5 mM $MgCl_2$, 25 mM KCl, 0.05 mM EDTA, 10% glycerol, 0.1% NP-40) and protease inhibitor cocktails. Lysate was centrifuged for 3 min at 3000 rpm at 4 °C. The pellet was resuspended in Buffer C (10 mM HEPES pH 7.6, 3 mM $MgCl_2$, 100 mM KCl, 0.1 mM EDTA, 10% glycerol), protease inhibitor cocktail, and 3 M ammonium sulfate and rotated at 4 °C for 30 min. Following incubation, lysate was ultracentrifuged at 100,000 x g for 15 min. Subsequently, 0.3 mg/mL ammonium sulfate powder was added to precipitate protein. Sample was ultracentrifuged again at 100,000 x g for 10 min. The protein pellet was stored at -80 °C for later use.

3.2.6 Immunoprecipitation

Ammonium sulfate preparations were resuspended in Mass Spectrometry IP buffer (25 mM Tris pH 8, 1% NP-40, 300 mM NaCl, 1 mM EDTA). CBX8 and V5 antibodies (6 µg) were cross-linked to Dyna protein G beads following Abcam protocol. Protein (1 mg) was pre-cleared with Dyna protein G beads for 1 hr at 4 °C. Following pre-clear, protein was incubated with the crosslinked antibody beads for 2 hr at 4 °C. Beads were subsequently washed three times for 5 min in a high salt wash buffer (600 mM NaCl) followed by an ammonium bicarbonate wash. IP samples were run on a 4-12% SDS-page gel (Invitrogen) and silver stained (Thermo). The V5 IP on the silver stained gel was subdivided into 13 pieces and subjected to mass spectrometry analysis.

3.2.7 Immunoblot and antibodies

Cells were harvested, washed with PBS, and lysed in Buffer A on ice. Nuclei were pelleted at 1000 x g for 5 min. Following, nuclei were re-suspended in radioimmunoprecipitation assay buffer (RIPA) and incubated on ice for 15 min. Samples were centrifuged at 21,000 x g to pellet chromatin. LDS sample buffer with beta-mercaptanol (BME) was added to the sample. Samples were boiled and run on 4-12% SDS page gel. The gel was transferred to a PDVF membrane by electrophoresis. The membrane was blocked in 5% bovine serum

albumin (BSA) and incubated in primary antibody diluted in 5% BSA overnight at 4 °C. The next day membranes were washed with PBS-T and incubated with secondary antibody for 1 hr at room temperature. Immunoblots were washed again with PBS-T prior to imaging on the LiCOR Odyssey. Antibodies used CBX8, CBX7, CBX6, CBX4, CBX2, BRG1.

3.2.8 Size Exclusion Chromatography

Nuclear lysate for size exclusion was isolated as described in Connelly et al.[209] Briefly, cells are lysed with Buffer A on ice for 15 min. Nuclei were pelleted at 1000 x g. Nuclei were resuspended in a MS compatible buffer (25 mM Tris-HCl pH 7.6, 300 mM NaCl, 1 mM EDTA, and protease inhibitor cocktail) and incubated on ice for 45 min. Chromatin was pelleted at 21,000 x g for 5 min at 4 °C. Soluble nuclear proteins (~ 1mg) were fractionated into 500 µL fractions using the Superdex 200 10/300 GL column (GE Healthcare) using an AKTA fast protein liquid chromatography (FPLC) system at a flow rate of 0.2 mL/min. The SEC column was pre-equilibrated with 2 column volumes of buffer (20 mM Tris-HCl, pH 7.5, 0.5 mM DTT, 1 mM EDTA, 100 mM NaCl, 5% glycerol) at a flow rate of 0.2 mL/min. Following fractionation, samples were acetone precipitated with 5 volumes of cold acetone. Precipitated protein was used for subsequent immunoblot or mass spectrometry analysis.

3.2.9 RNA-seq

Short-term knockdown: T98G cells for RNA-seq transduced with pLKO (control) or shCBX8 (Thermo Fisher). Twenty-four hours following infection cells were split into a 6 well, 100,000 cells per well and selected with 2 µg/µL of puromycin. Cell were harvested 7 days after infection with Trizol and processed according to the product manual. Isolated RNA was cleaned up with RNA-Easy. Sequencing was performed with three biological replicates. Data was processed by Dr. Matthew Schipma at the Northwestern University Seq Core Facility.

Long-term knockout: T98G sgControl, T98G sgCBX8, T98G sgCBX8 +WT CBX8, and T98G sgCBX8+ CBX8 quad cells were plated at a density of 80,000 cells per well and induced with 2 µg/µL of doxycycline every 48-hours. Cells were harvested after 72 hours with Trizol and RNA was processed as described above. Sequencing was performed with

three biological replicates. All RNA samples were sent to BGI sequencing for downstream processing. RNA-seq data has approximately 20 million reads per sample following trimming using Galaxy trimmomatic function.[210] The paired-end reads were aligned to the hg38 genome using HiSat2 in Galaxy[210] and differentially expressed genes (FDR < 0.05) were identified using EdgeR[211,212] in RStudio. Gene ontology analysis was performed using the Gene Ontology Consortium.[213,214] Bonferroni correction was used and p-value < 0.05 was deemed significant.

3.2.10 Doxorubicin sensitivity assays

T98G CRISPR cell lines (sgcontrol, sgCBX8-1, sgCBX-2) were seeded in a 384 well plate at 300-500 cells/well. Cells were treated with doxorubicin 48 and 96 hours after plating. One week after plating, cells were lysed with Cell Titer Glo and luminescence intensity was measured. Doxorubicin concentrations were two-fold dilutions starting at 400 nM.

3.3 Results

3.3.1 CBX8 knockdown decreases GBM viability and tumorsphere formation

As previously mentioned, CBX8 is overexpressed approximately two-fold in over 50% of GBM patients.[166] We observe CBX8 overexpression relative to healthy astrocytes (Svg12) in five GBM cell lines (T98G, U118MG, U138MG, A172, and U87MG) by immunoblot (Figure 3.1a). Thus, to assess if CBX8 is important for GBM viability, we knocked down CBX8 expression with two small hairpin RNAs (shRNAs) (Figure 3.1b) in all five GBM cell lines. Cell count measurements revealed that both shRNAs against CBX8 reduced viability across all GBM lines compared to the empty vector control (Figure 3.1c). This decrease in viability, however, was not GBM specific, as we observe a decrease in Svg12 viability upon CBX8 knockdown (Figure 3.1d). This suggests that CBX8 may be important for cell cycle regulation. It is interesting to note that despite using lentivirus to generate stable knockdown cell lines, the CBX8 knockdown is lost after approximately 10 days suggesting the CBX8 knockdown is selected against. Therefore, we generated a T98G CBX8 CRISPR knockout line for subsequent analyses.

Tumorsphere formation is often used as a measure of tumor initiation/self-renewal capability.[215] The Polycomb Group proteins are critical for adult stem cell self-

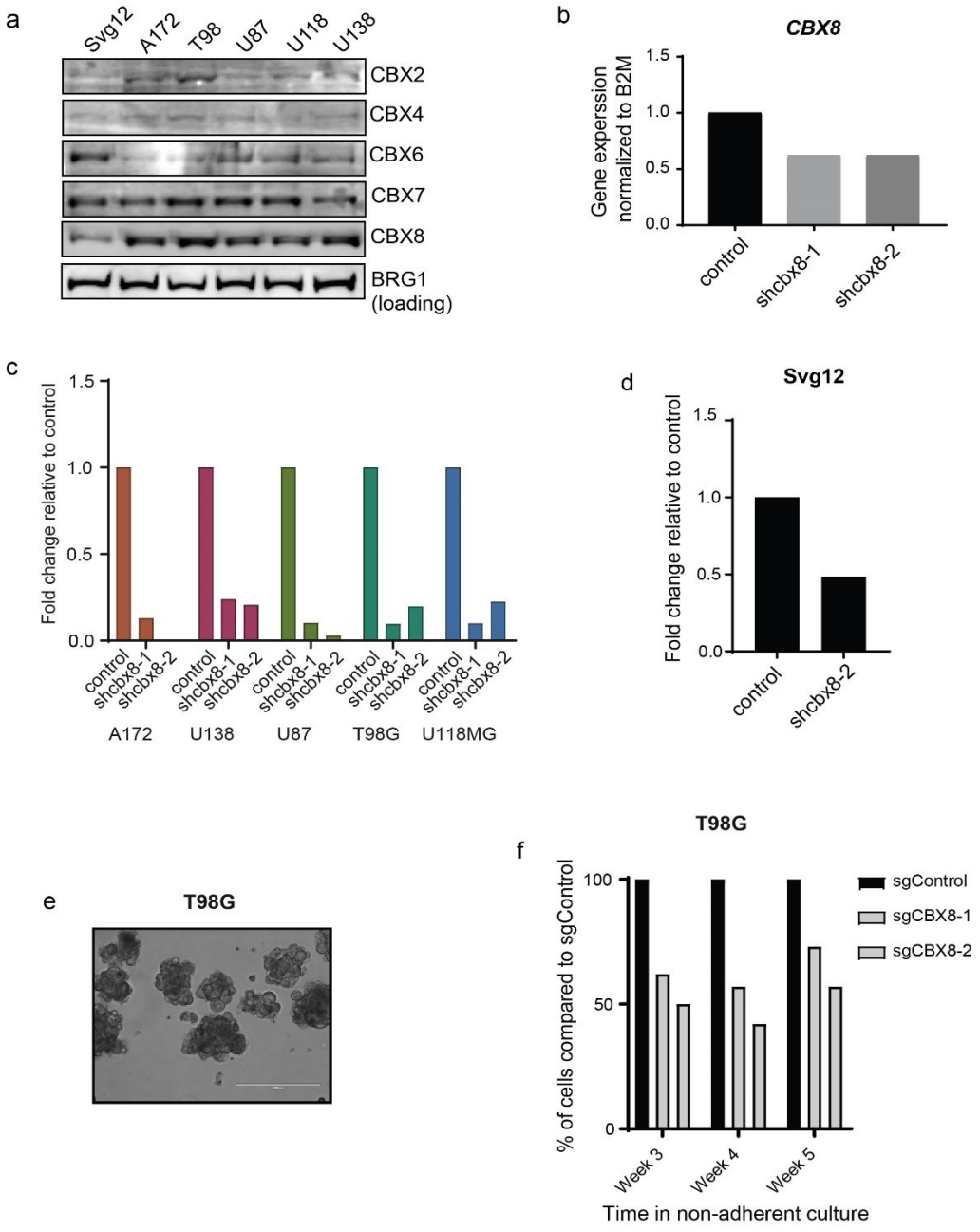


Figure 3.1 CBX8 is important for GBM viability

a) Immunoblot of CBX paralog expression levels across a panel of GBM cell lines and healthy astrocytes (svg12). Immunoblot performed by Aktan Alpsoy. b) validation of CBX8 knockdown using qRT-PCR. c) Viability of GBM cell lines upon CBX8 knockdown. Measurement was cell count. d) Viability of svg12 astrocytes upon CBX8 knockdown. Measurement was cell count. e) Image of T98G tumorsphere. f) Tumorsphere formation quantitation with the T98G CBX8 CRISPR knockdown over five weeks. Cell count was used as measurement. Assay performed by Dr. Emily Dykhuizen.

renewal,[37,195] and the BMI-1 (PCGF4) subunit of PRC1 sustains GBM self-renewal.[167] Thus, we sought to assess the role of CBX8 in GBM self-renewal using our CBX8 CRISPR cell lines and a tumorsphere assay. The GBM T98G cell line is capable of forming tumorspheres (Figure 3.1e). Upon loss of CBX8 with two independent guide RNAs (sgCBX8-1 and sgCBX8-2), we observed an approximately 50% reduction in tumorsphere formation compared to the control line (sgControl) (Figure 3.1f). This suggests that CBX8 is important for GBM self-renewal.

3.3.2 The CBX8 chromodomain is necessary for GBM survival

CBX8's importance in GBM viability and self-renewal suggests it may serve as a therapeutic target. In fact, the CBX chromodomain interaction with H3K27me3 has been deemed "druggable,"[216] and there has been success developing inhibitors against the CBX7 chromodomain as discussed in Chapter 2.[155,217] We hypothesized that the CBX8 chromodomain is a therapeutic target for GBM. To test this hypothesis, we generated a previously described chromodomain mutant (mutCBX8, K31A,W32A).[64] This chromodomain mutation reduces CBX8 localization to the *INK4A/ARF* locus and de-represses expression.[64] We re-expressed either wild-type CBX8 or mutCBX8 in a CBX8 knockdown background and performed a sequential salt extraction[218] to verify that mutCBX8's chromatin association was disrupted (Figure 3.2a). Following, both wild-type and mutCBX8 were re-expressed in the T98G shCBX8-2 cell line, and cell viability was assessed after six days. Wild-type re-expression rescued GBM cell growth, while the mutCBX8 failed to restore cell viability, suggesting the CBX8 chromodomain is important for CBX8's function in GBM progression and maintenance and may serve as a therapeutic target.

3.3.3 CBX8 knockdown does not enhance chemotherapy toxicity

The CBX paralogs share over 60% of the same gene targets and are thought to act redundantly in cases.[52,58,63] Thus, we sought to assess if CBX8 inhibition enhances doxorubicin toxicity as CBX7 does. We treated our CBX8 CRISPR knockdown cell lines with doxorubicin. As expected, we observed a dose response decrease in viability with

doxorubicin; however, loss of CBX8 did not enhance this response relative to the control cell line. Though further characterization is necessary, this data suggests CBX8 has

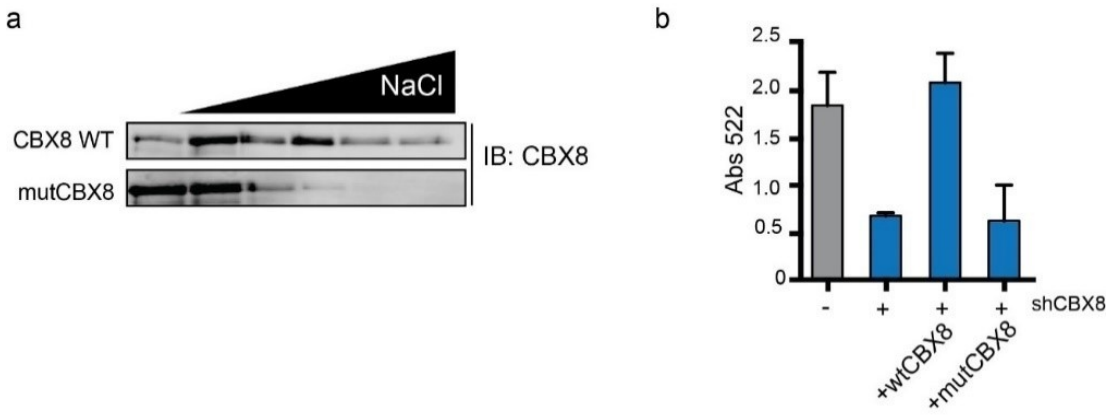


Figure 3.2 CBX8 chromodomain is important

a) Sequential salt extraction of WT or mutCBX8 re-expression in T98G cells to demonstrate chromatin association. b) CBX8 knockdown rescue study with WT or mutCBX8. Viability was measured at day 6 using a sulforhodamine assay to measure bulk protein levels.

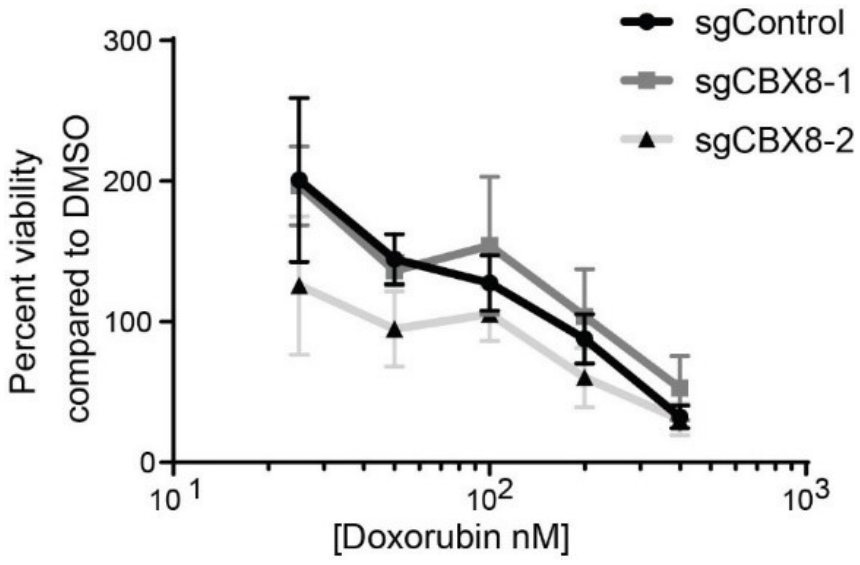


Figure 3.3 Loss of CBX8 does not enhance doxorubicin toxicity

T98G CBX8 CRISPR and control cells lines treated with increasing concentrations of doxorubicin. Viability measured using Cell Titer Glo, n =3.

different functions in GBM than CBX7 and that CBX8 does not play a significant role in DNA damage response in GBM.

3.3.4 RNA-seq analysis

To date, little is known about CBX8's gene targets, especially in cancer. The well-documented gene target of CBX8, *INK4A/ARF*, is deleted in many GBM cell lines.[219] Thus, we sought to understand CBX8's transcriptional role in GBM. To do so, we took two approaches to assess CBX8 gene regulation: short-term CBX8 knockdown and a long-term CBX8 knockout with or without CBX8 re-expressed.

3.3.4.1 Short-term CBX8 transcriptional profile

To assess CBX8's short-term transcriptional profile, we knocked down CBX8 in the T98G cell line for a week and performed next-generation RNA-sequencing. Our RNA-seq analysis revealed over 1000 differentially expressed genes (DEGs) in the CBX8 knockdown compared to control. Of these 1712 DEGs, 738 genes were upregulated upon loss of CBX8, while 974 genes were downregulated. As a member of the PcG family, we would expect CBX8 regulated genes to be upregulated upon CBX8 knockdown, however, without genome-wide localization studies we cannot conclude direct CBX8 targets. Gene ontology (GO) analysis of the upregulated genes revealed enrichment of genes involved in cell adhesion, cell differentiation, and development in the CBX8 knockdown cell line (Figure 3.4). Upregulation of genes involved in cell adhesion, more specifically, the positive regulation of the epithelial to mesenchymal transition, is in accordance with previously published mechanisms of CBX8 in cancer.[202,206] Additionally, the RNA-seq data suggests that CBX8 is involved in the regulation of neuronal development and differentiation processes which is consistent with the role of CBX8 in development and stem cell maintenance.[52] On the other hand, gene ontology analysis of the down regulated genes revealed an enrichment for various metabolic processes and the regulation of transcription which can have implications on tumor cell survival (Figure 3.4).

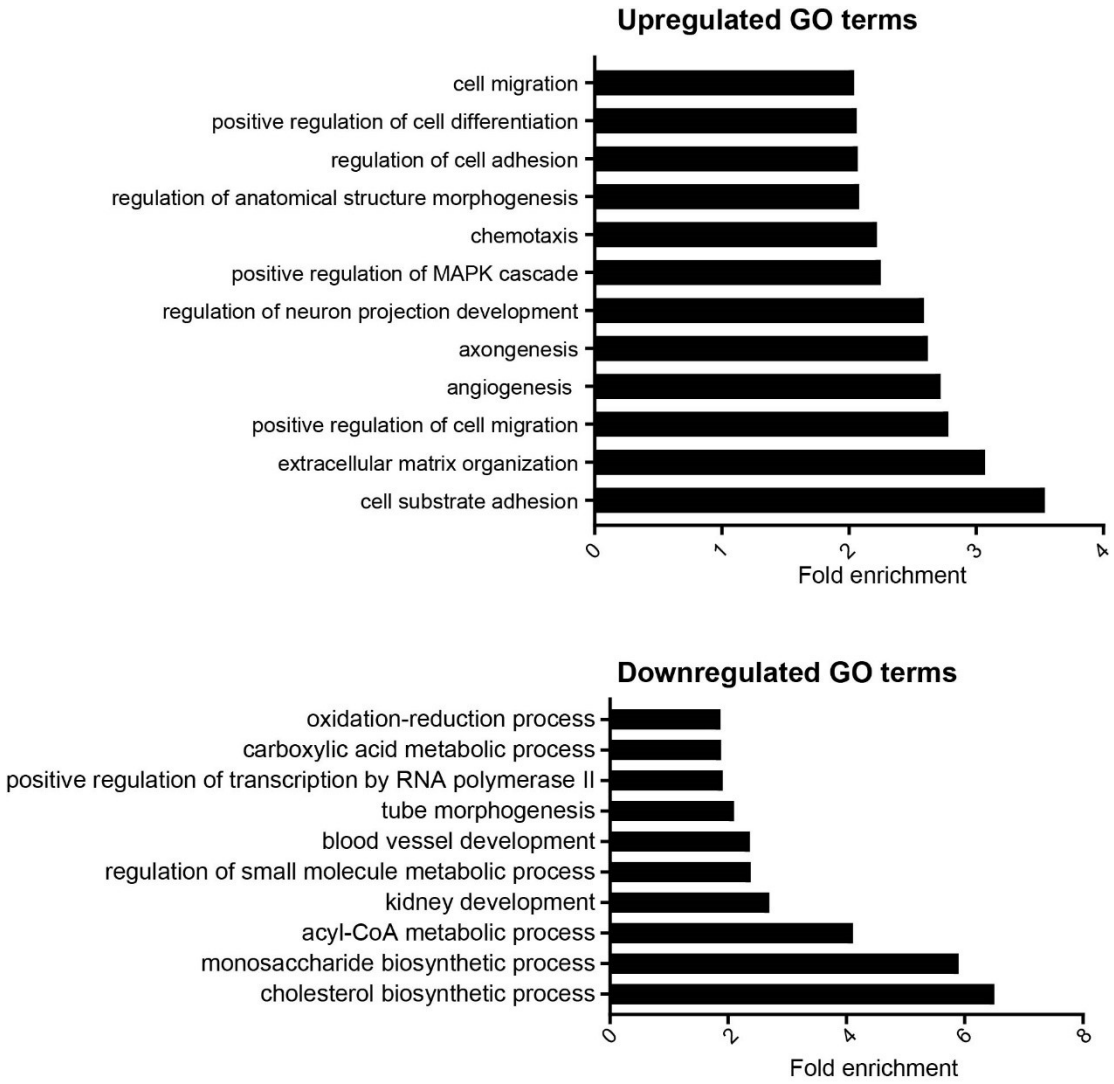


Figure 3.4 Gene ontology

Gene ontology analysis of significantly upregulated (top) or downregulated (bottom) genes from short term CBX8 knockdown. GO terms are significantly enriched ($p < 0.05$).

3.3.4.2 Long term CBX8 transcriptional profile

In addition to the CBX8 knockdown with shRNA, we assessed transcriptional changes in our T98G CBX8 knockout cell line. With this knockout line, we were able to assess long term transcriptional changes upon loss of CBX8. Additionally, we re-expressed either wild type CBX8 or a chromodomain-null mutant (quad, R19, 20, 22A and W32A) to examine transcriptional changes upon rescuing CBX8 expression. Cluster analysis demonstrates a difference between the control cell line and the CBX8 knockout as expected; however, the re-expression of WT CBX8 or the quad mutant are more similar to the KO than the control (Figure 3.5a). This suggests that our re-expression does not fully restore a CBX8 transcriptional profile, likely due to the short (72 hour) CBX8 re-expression time frame. In the CBX8 knockout, we identified 3245 differentially expressed genes compared to the control. Of the 3245 DEGs, 1532 genes were upregulated, while 1713 genes were downregulated. Gene ontology analysis of the upregulated genes upon CBX8 knockout revealed enrichment of genes involved in development of different tissues, cell adhesion, and differentiation (Figure 3.5b). This suggests that loss of CBX8 derepresses lineage specific genes. On the other hand, the downregulated genes demonstrated enrichment in metabolic processes similar to the short-term knockdown (Figure 3.5c). We then examined the differentially expressed genes in the CBX8 KO (relative to the control) and in the CBX8 WT re-expression (relative to CBX8 KO). Ideally, CBX8 regulated genes will demonstrate differential expression in the KO compared to the control as well as in the WT re-expression compared to the KO, if the re-expression restores CBX8's transcriptional program. In line with our cluster analysis, we observe a low overlap of differentially expressed genes in the CBX8 KO and CBX8 WT (~12%) further confirming that the CBX8 transcriptional program is not fully restored. A longer re-expression time-frame is likely necessary for CBX8's transcriptional profile to be more similar to the control than the knockout.

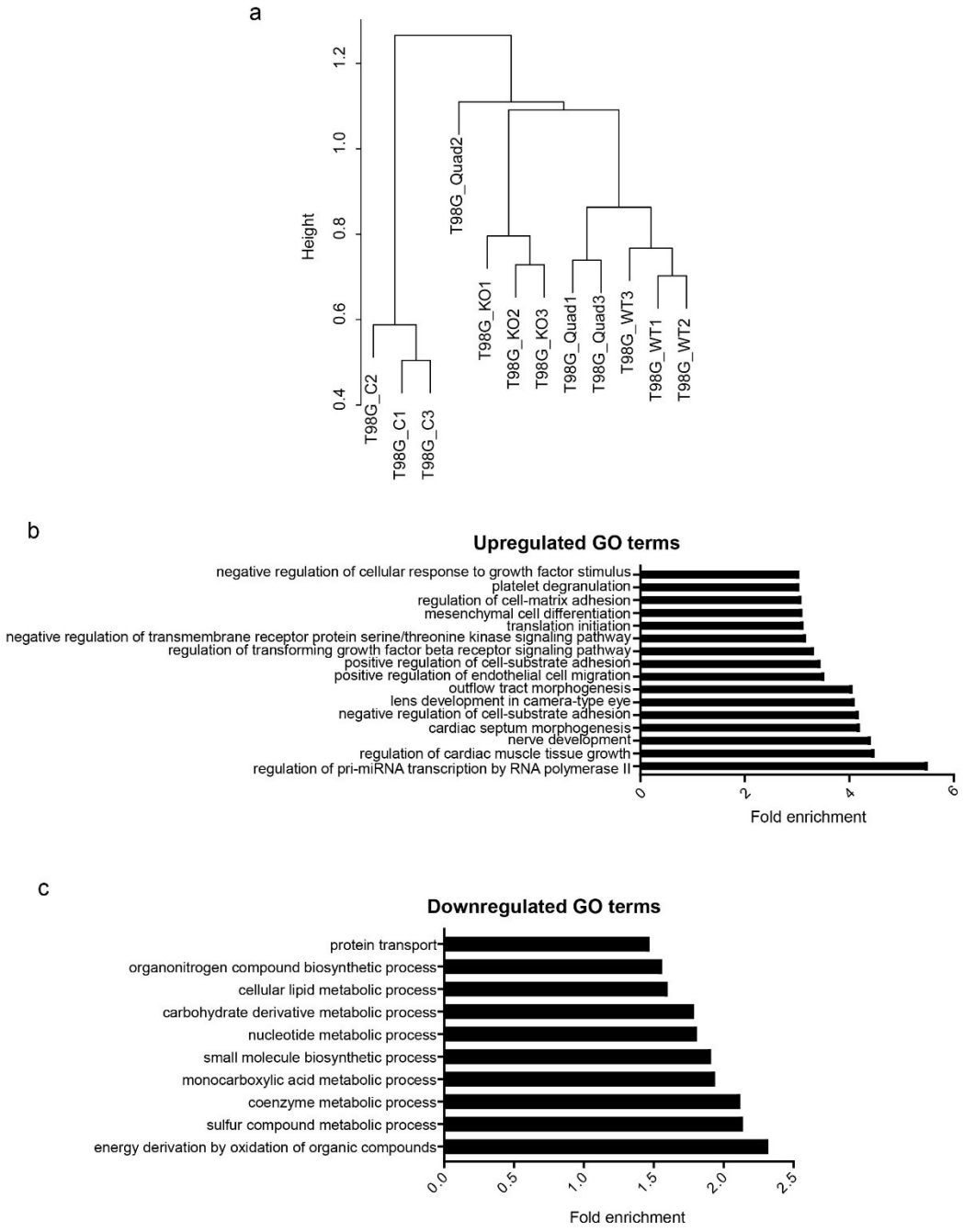


Figure 3.5 CBX8 knockout RNA-seq analysis.

a) H-cluster analysis of CBX8 knockout and CBX8 re-expression RNA-seq. b) Significantly enriched ($p < 0.05$) gene ontology terms for the significantly upregulated genes in the CBX8 KO compared to control, GO terms >3-fold enrichment shown. c) Significantly enriched GO terms for significantly downregulated genes; top 10 GO terms shown.

Further, we compared the differential gene expression between the CBX8 knockout (long-term) and the CBX8 knockdown (short-term, shRNA) and identified 556 genes that were differentially expressed in both sequencing data sets. Of these 556 genes, only 147 genes were upregulated in both the knockout and knockdown. The shared upregulated genes are involved in cell migration, adhesion, and development. Despite the low similarity of genes upregulated between the knockdown and knockout, the consistency in gene ontology terms suggests a role for CBX8 in regulating cell adhesion, migration and differentiation. Follow up studies are necessary to better understand CBX8's role in these processes. Further, genome-wide localization studies are necessary to identify which of the genes are direct CBX8 targets.

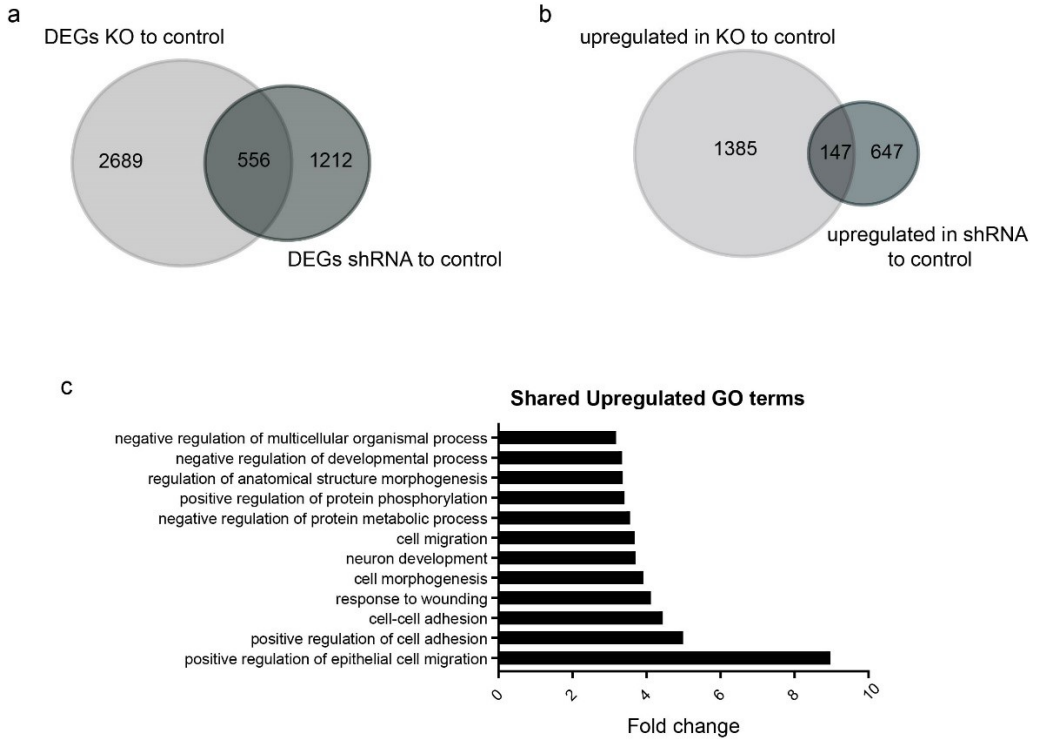


Figure 3.6 RNA-seq comparison.

a) Differentially expressed genes shared between the CBX8 KO and the CBX8 knockdown (shRNA) compared to the respective control. b) Venn diagram demonstrating the shared upregulated genes between the CBX8 KO and knockdown datasets c) Significantly enriched ($p < 0.05$) GO terms of the shared upregulated genes.

3.3.5 Understanding CBX8-containing complexes in GBM

With the 180 possible PRC1 complex compositions, many studies have focused on identifying the various PRC1 complexes that exist as well as cell-type specific complexes and their function. These studies have demonstrated that throughout the differentiation process to different cell lineages the predominant cPRC1 complex and its gene targets changes[52,54,58] (see Figure 1.3 for an example). In cancers, tumor sequencing has identified upregulation/downregulation of PRC1 subunits compared to normal tissue.[192] Based on these observations, it is likely that cPRC1 complexes undergo a composition change during oncogenesis. Further, little is known as to what the predominant PRC1 complexes that may have oncogenic functions are. Thus, we sought to understand the predominant CBX8-containing complex and its associating proteins in GBM.

Our initial approach to assessing CBX8-containing complexes in GBM was co-immunoprecipitation followed by mass spectrometry analysis (co-IP/MS). For these studies, however, it is important that the protein of interest (bait protein) is at endogenous expression levels. Introduction of the bait protein exogenously, usually with an epitope tag, can potentially alter complex stoichiometry and composition.[220] Thus, we performed co-immunoprecipitation in the T98G cell line using a CBX8 antibody and SDS-page gel purification for our mass spectrometry analysis. Unfortunately, we encountered high background masking detection of the Polycomb proteins. This is likely due to the low abundance of PRC1 proteins. We subsequently generated a doxycycline inducible cell line expressing V5-tagged CBX8 to improve the immunoprecipitation. We verified that the V5-tag did not impede complex formation. We then performed a V5 immunoprecipitation in parallel with a CBX8 immunoprecipitation for mass spectrometry. In order to detect proteins of interest, the SDS-page gel pieces had to be subdivided into 11 sections for mass spectrometry analysis. Unfortunately, the IP/MS still had a high protein background level. While it is difficult to identify novel CBX8-interacting proteins from our data, we were able to identify the composition of the CBX8-containing PRC1 complex. The predominant CBX8 complex consists of RING1/RING2, PHC2, and PCGF2/PCGF4 (Figure 3.5). In part, there is difficulty specifying which paralog is incorporated into the complex

PRC1 subunit	# of sig. sequences
CBX8	63
RING2	10
PHC2	8
RING1	8
PCGF2/4	1

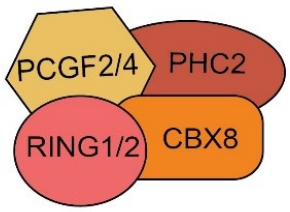


Figure 3.7 CBX8-containing PRC1 complex in GBM

PRC1 subunits and the number of significant sequences identified by co-IP/MS (left). Cartoon image of the CBX8-containing PRC1 complex in T98G GBM cells.

because of the high sequence similarity between paralogs, eg. PCGF2 and PCGF4 are 63% identical. Thus, the identified peptides can be for multiple paralogs but are only classified for one paralog. Additionally, there are caveats to our co-IP/MS because we performed the IP using an exogenously expressed bait protein. This exogenous expression can skew complex formation and may drive non-endogenous complex compositions.

To improve our proteomic characterization of PRC1 in GBM, we collaborated with the Purdue Proteomics Facility to develop a work flow to isolate and examine nuclear protein complexes (Figure 3.6).[209] Briefly, nuclear lysate was isolated and subjected to size exclusion chromatography (SEC) to separate protein complexes by size.[209] Following separation, the collected fractions are subjected to western blot analysis and liquid chromatography mass spectrometry (LC/MS) to identify the proteins within each fraction.[209] This work flow allows us to predict protein complex size and composition. Subsequently, protein complexes of interest can be validated by co-immunoprecipitation.[209]

We set out to assess PRC1 complex size and composition differences between the Svg12 astrocytes and the GBM T98G cell line. We performed size exclusion chromatography on nuclear lysates from each cell line. Following separation, we assessed separation reproducibility using immunoblot analysis against the CBX7 and CBX8 paralogs. Interestingly, in the astrocytes, CBX7 elutes in later fractions than CBX8 does in the T98G cell line (Figure 3.7). Based on the standard curve for the sizing column, CBX7 is predominantly in a complex that is approximately 648 kDa, while CBX8 is predominately in a complex over 1000 kDa. This is interesting as the core canonical PRC1 complex is approximately 230 kDa. The complex size difference between paralogs

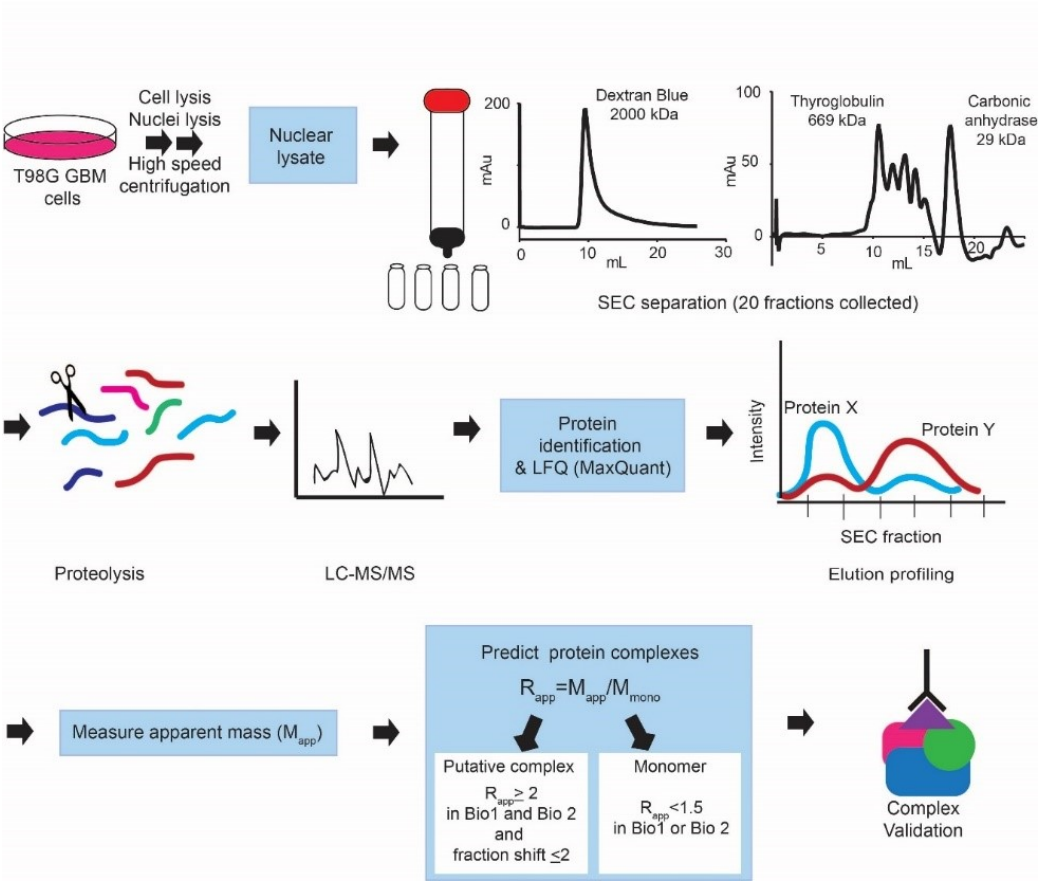


Figure 3.8 Nuclear protein complex workflow

Overview of workflow to identify nuclear complexes from nuclear protein isolation to post-mass spectrometry analysis. This figure is reproduced with permission from Connelly, K.E., Hedrick, V., Sobreira, T., Dykhuizen, E.C., Aryal, U.K. *PROTEOMICS.* 2018 May., 18(11). © 2018, *Wiley and Sons* [209]

suggests that PRC1 complexes associate with additional proteins that may help dictate the complex's function.

Following confirmation of successful separation, the T98G SEC samples were subjected to mass spectrometry analysis. Unfortunately, due to the low abundance of PRC1 subunits and high abundance of other nuclear proteins, PRC1 subunits were not detected Unfortunately, being unable to detect PcG proteins in our MS and not comparing CBX8-containing complex size between cell lines with immunoblot, we were unable to fully assess PRC1 complex composition changes upon oncogenesis. Future work is necessary to optimize conditions to enrich PRC1 subunits.

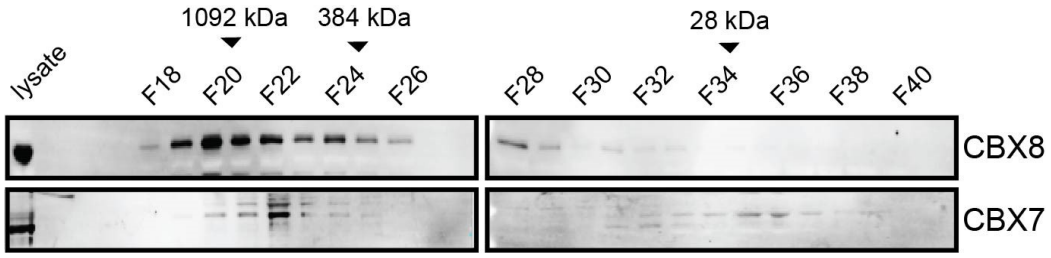


Figure 3.9 SEC CBX profile

Size exclusion chromatography separation of T98G (top) or svg12 (bottom) nuclear lysate demonstrates slight variation in CBX8 and CBX7 complex sizes.

3.4 Discussion and future directions

We and others have demonstrated that the CBX8 paralog is overexpressed in GBM.[166] Our phenotypic studies of CBX8 in glioblastoma multiforme suggest CBX8 plays an important role in GBM cell viability and that the chromodomain is necessary for CBX8's function. These data however raise a lot of questions regarding the function of CBX8. We took several different approaches to begin assessing the role CBX8 plays in GBM progression and maintenance.

As a member of the PcG family, CBX8 represses gene transcription. Thus, we first aimed to assess CBX8-mediated transcription. Our differential gene expression analysis suggests that CBX8 is important for maintaining the repression of genes involved in neuronal differentiation. This is not surprising considering PcG's role in stem cell maintenance and self-renewal[37] and suggests that CBX8 may play a role in maintaining a

stem like state. Further, the reduction of T98G tumorsphere formation in the CBX8 knockout cell line supports a role for CBX8 in GBM stem like cells and self-renewal. This is not without precedence as previous studies have demonstrated that Polycomb proteins such as BMI-1 (PCGF4) and EZH2 are necessary for GBM stem cell self-renewal.[167,221] Further, it is thought that the stem-cell like population is the source of GBM reoccurence.[163] Intriguingly, CBX8 is only upregulated two-fold in GBM cell lines and patient tumors.[166,199] A two-fold increase in CBX8 expression, however, does not necessarily account for the phenotypes we observe or the fact that CBX8 expression can distinguish low grade and high grade gliomas.[199] This is likely due to the heterogeneity for tumor samples. It is likely that CBX8 is highly upregulated in a small subset of GBM tumors/cells and when examined in a global context only demonstrates a two-fold increase in expression. Thus, we hypothesize that CBX8 is highly upregulated in the small stem-cell like population and is necessary for self-renewal. In order to explore this hypothesis, we can use flow cytometry to assess expression levels of the neural stem cell marker CD133. With flow cytometry, we can sort GBM cell populations based on both the CBX8 and CD133 expression levels. Previous studies have shown that expression of BMI-1 (PCGF4) and EZH2 correlate to high CD133 expression, a neural stem cell marker.[167] Similarly, we would expect to observe a $CBX8^{High}/CD133^{High}$ population, if CBX8 is upregulated in the stem cell like population. Additionally, single cell RNA sequencing can be performed to examine the transcriptional profile of individual cells in a heterogenous population. Specifically, transcriptional differences between a CBX8-expressing and a CBX8 KO cell can be determined. This would provide a more accurate understanding of CBX8 transcriptional regulation. If further studies confirm that CBX8 expression does correlate with the GBM stem like population, it suggests that CBX8 may be a good therapeutic target to reduce tumor reoccurrence.

Our findings that the CBX8 chromodomain may be a therapeutic target has led our lab in collaboration with Dr. Casey Krusemark's lab to develop CBX8 specific inhibitors. Through DNA-encoded libraries, Denton el al., identified structure activity relationships of peptidomimetics that specify the specificity between the CBX7 and CBX8 chromodomains.[222] These libraries identified a potent CBX8-specific chromodomain peptidomimetic (Figure 3.8). Although this inhibitor has shown great *in vitro*

characteristics, cell permeability has been a limitation and there is ongoing work to improve inhibitor properties. The identification of a CBX8-specific inhibitor or chemical probe will provide a tool to understand CBX8 chromodomain function in oncogenesis without having to use genetic approaches. The use of a chemical probe over genetic approaches allows us to assess the CBX8 function without the caveat of potentially altering PRC1 complex compositions. Further, ongoing work to develop a PROTAC system with a CBX8 chromodomain probe will improve the ability to assess immediate, CBX8 specific biological roles. This will aid in identifying direct gene targets of CBX8 and better understand the implications of CBX8 in GBM as well as in other systems.

Our proteomic studies sought to identify novel CBX8-associating proteins and define the CBX8-containing PRC1 complex in GBM to better understand CBX8's biological function. While we were able to characterize a CBX8-containing PRC1 complex, our work highlighted the challenges of studying CBX8-associating proteins. To date, many of the proteomic studies have utilized overexpression systems followed by IP or tandem affinity purification. Exogenously introducing the bait protein can alter protein complex formation and drive non-endogenous protein associations. While these studies have yielded CBX8-associating proteins, many of the identified proteins have not been validated in an endogenous system. Unfortunately, we experienced the limitations of using endogenous CBX8 for our co-IP/MS and had to utilize exogenous CBX8 to detect PcG proteins. The workflow we developed with the Purdue Proteomics Facility has promise in improving our ability to isolate CBX8-specific complexes. Though, further optimization is necessary, combining the biochemical separation of nuclear protein complexes with CBX8 immunoprecipitation of the CBX8-containing fractions may reduce the background. This workflow will allow us to characterize protein complex alterations between cell lines. Being able to understand the changes in PRC1 complex composition in between cell lines, particularly during the development and oncogenic processes, can shed a lot of information on the biological functions of particular complexes and how the composition can alter function.

CHAPTER 4. ENGAGEMENT OF DNA AND H3K27ME3 BY THE CBX8 CHROMODOMAIN DRIVES CHROMATIN ASSOCIATION

The following chapter has been provisionally accepted for publication in *Nucleic Acids Research*, published by Oxford University Press.

Connelly, K.E,* Weaver, T.M,* Gu, B.X., Musselman, C.A., and Dykhuizen, E.C. *Nucleic Acids Research.* 2018 *(in revision)*

*These authors contributed equally to this work

4.1 Introduction

Eukaryotic DNA is packaged into the cell nucleus in the form of chromatin. At its most basic level, chromatin is made up of repeats of nucleosome particles, which consist of ~147 base-pairs (bp) of DNA wrapped around an octamer of histones H2A, H2B, H3, and H4. The N-terminus of each histone (and the C-terminus of H2A) protrude from the nucleosome core and are collectively known as the histone tails. These tails can be heavily post-translationally modified, catalyzed by a group of proteins known as "writers". The modification of histone tails can alter chromatin structure directly and/or recruit or regulate chromatin associated proteins contributing to transcriptional changes.[223,224] These interactions are mediated through "reader" domains, that can recognize specific modification states.[225,226]

The Polycomb Group (PcG) proteins, which are critical for lineage specification and adult stem cell maintenance,[37] form two distinct histone writer complexes: Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2).[9,227,228] According to the canonical mechanism for Polycomb, PRC2 catalyzes the trimethylation of lysine 27 on histone H3 (H3K27me3),[227] and PRC1 binds this modification, where it monoubiquitinates H2AK119 and compacts chromatin to repress transcription.[10,80,229] In *Drosophila melanogaster*, recognition of H3K27me3 by dPRC1 is mediated by a chromatin modifier organization (chromo) domain, in the Polycomb (dPc) subunit.[9,118] There are five paralogous proteins to dPc in mammals, referred to as chromodomain-containing chromobox subunits (CBX2,4,6,7,8)[230], that form mutually exclusive PRC1 complexes.[66] The mammalian CBX chromodomains (CDs) demonstrate lower affinity for methylated histone peptides than

does the dPc CD, and not all display specificity for H3K27me3 over H3K9me3 histone peptides *in vitro*.[9,102,103] Of the five CBX CDs, the CBX8 CD demonstrates the weakest histone peptide binding (K_d>500 μM) in these studies and no measurable specificity for H3K27me3 peptides.[103] Meanwhile, live cell imaging studies suggest that H3K27me3 is important for the chromatin association of CBX8,[231] making the mechanism by which the CD contributes to CBX8 chromatin association and histone mark specificity unclear. A greater understanding of the CBX8 CD is not only desirable for understanding the fundamental mechanism of PRC1 function, but also for deciphering whether the CD is a good therapeutic target.[232] Specifically, CBX8 plays an oncogenic role in several cancers including breast cancer and acute myeloid leukemia[51,136,201] and is overexpressed in numerous others, such as glioblastoma multiforme (GBM).[166,205,233] Inhibition of CBX8 may be a viable path for treatment of these diseases, and the CD may be targetable with small molecule inhibitors.[222]

Here, we demonstrate that CBX8 association with chromatin is largely driven by the CD. Notably, we find that this is mediated through a combination of H3K27me3 binding, and a previously unrecognized interaction with DNA. We investigate the structural basis of both of these interactions, defining the root of moderate specificity for H3K27me3, and how histone and DNA binding integrate on multiple levels. Notably, despite the fact that *in vitro* histone tail binding is weak, and that nucleosome association *in vitro* is driven by DNA binding, we find that both DNA and H3K27me3 binding contribute to robust chromatin association *in vivo*, highlighting that the chromatin context is critical in determining ultimate reader domain function. These studies not only provide novel insight into reader domain and PRC1 function but will pave the way for the development of CBX8 inhibitors.

4.2 Methods

4.2.1 Plasmids and constructs

CBX8 chromodomain construct was a gift from Cheryl Arrowsmith (Addgene plasmid #62519). GST-CBX8 chromodomain was generated using Infusion (Clontech) and the pGSTag vector (a gift from Gerald Crabtree) WT CBX8 gene was obtained from the

Mammalian Gene Consortium (ID#4121509). The ΔCD (aa 77-389) was generated. QuikChange (Stravagene) was used to generate full length CBX8 W32A mutant. Additional full length CBX8 mutant constructs (R19A; R20A; R22A; R19,20,22A;) were purchased from GenScript and amplified. All CBX8 genes were cloned into the tet-inducible conditional lentiviral vectors TetO-FUW (a gift from Rudolf Jaenisch Addgene plasmid # 20323) containing the N-terminal V5 epitope with Infusion (Clontech). Infusion reactions were transformed in Stbl3 Competent cells (Invitrogen), DNA was isolated, and precipitated with 2 volumes ethanol and 0.1 volume sodium acetate for lentiviral transduction. CBX8 and control CRISPR guide RNAs were cloned into px459 v2.0 vector (a gift from Feng Zhang, Addgene plasmid #62988)[208] EZH2 guide RNAs were cloned into Lenti_sgRNA_EFS_GFP (a gift from Chris Vakoc, Addgene plasmid #65656.)[234] Guide RNA sequences: sgCBX8: GCATGGAATACCTCGTGAAA, sgControl: GTAGCGAACGTGTCCGGCGT sgEZH2: CTGGCACCATCTGACGTGGC.

4.2.2 Protein Expression and Purification

The recombinant CBX8 CD was expressed in BL21 (DE3) (New England Biolabs) *E. coli* cells. Cells were grown in LB-medium or M9-minimal media supplemented with ^{15}N-NH_4Cl or ^{15}N-NH_4Cl and ^{13}C-glucose. For unlabeled protein, cells were grown shaking at 215 rpm at 37°C until an OD~1.0 was reached and induced with 1 mM IPTG for 16-18 hours overnight. For isotopically-enriched protein, cells were grown in LB-medium until an OD~1.0, spun down at 4000 rpm for 10 minutes, and resuspended in M9-medium (4L LB cells per 1L M9) supplemented with either ^{15}N-NH_4Cl or ^{15}N-NH_4Cl/^{13}C-glucose. The cells were allowed to recover in M9 media for 1 hour shaking at 18°C and induced with 1.0mM IPTG for 16-18 hours overnight. Cells were subsequently collected by centrifugation at 6000 rpm for 20 minutes, frozen in N_2(l) and stored at -80°C.

For purification, cells were resuspended in a buffer containing 100 mM NaCl, 25 mM Tris (pH-7.5) supplemented with DNase I and lysozyme. Cells were then lysed using an Emulsiflex homogenizer (Avestin) or by sonication, and lysate cleared by centrifugation at 15,000 rpm for 1 hour at 4°C. The soluble supernatant was incubated with glutathione agarose resin (ThermoFisher Scientific) rotating at 4°C for 1 hour. The GST-tagged CBX8 CD was washed extensively with a high salt buffer containing 1 M NaCl and 25 mM Tris

(pH-7.5) before elution with a buffer containing 150 mM NaCl, 25 mM Tris (pH-7.5) and 50 mM reduced glutathione. The GST-CBX8 CD was concentrated using a 3,000 MWCO centrifugation filter unit to 2 mL and cleaved with TEV protease for 3 hours at 25°C. The cleaved CBX8 CD was further purified using a combination of cation-exchange and size exclusion chromatography (Superdex 75, GE Healthcare Life Sciences). For NMR studies, ^{15}N-CBX8 CD and ^{15}N/^{13}C-CBX8 CD were used in a final NMR buffer containing 100 mM NaCl and 40 mM phosphate buffer (pH 6.8). For EMSAs, the unlabeled CBX8 CD was used in a final buffer containing 25 mM phosphate buffer (pH 6.8), 25 mM NaCl, 1 mM EDTA, 1 mM DTT.

4.2.3 Histone Peptides

The unmodified H3 (1-44) and H3K27C (23-34) peptides were synthesized by GenScript. The H3K9me3 (1-21), H3K27me3 (23-34), Biotinylated H3 (21-44) and H3K27me3 (21-44) peptides were obtained from AnaSpec. For NMR studies, peptides were resuspended in H_2O to a final concentration of 20 mM and pH adjusted to ~7.0 with NaOH. For peptide pulldown experiments, peptides were resuspended to a final concentration of 1 μg/μL in 10% dimethyl sulfoxide (DMSO). Ethylcysteine alkylation of H3K27C peptide was performed as previously described.[235] Briefly, peptide (2 mg) was resuspended in 8 M guanidinium chloride, 1 M HEPES pH 7.5, 1 M DTT and incubated for 1 hr at 37 °C. Following incubation, (2-bromomethyl)-trimethylammonium bromide (20 mg) was added for 2 hr incubation at 50 °C in the dark. Reaction was quenched with BME and HPLC purified. Reaction success was confirmed by mass spectrometry. Expected mass:1174.6, identified mass at 1174.8.

4.2.4 DNA oligonucleotides

The single stranded DNA oligonucleotides (5'-GCGTTTAAGCG-3' and 5’ CGCTTAAACGC-3’) were obtained from Integrated DNA technologies. For annealing, oligonucleotides were resuspended in 100 mM NaCl and 40 mM phosphate buffer (pH 6.8), heated to 90°C for 5 minutes and cooled overnight. The annealed DNA oligonucleotides were further purified by S75 size-exclusion chromatography in a buffer containing 100

mM NaCl and 40 mM phosphate buffer (pH 6.8) and concentrated to 3.2 mM in a 3000 MWCO centrifugal concentrator.

4.2.5 Nucleosome Reconstitution

Unmodified human histones H2A.1, H2B.1, H3.2 and H4 (T71C) were transformed into Rosetta2 (DE3) or BL21 (DE3) (New England Biolabs) and expressed in LB media. Cells were induced at OD~0.4 with 0.2 mM IPTG for histone H4 or 0.4 mM IPTG for histones H2A, H2B, and H3 for 3–4 hr. The histones were extracted from inclusion bodies and purified by anion and cation exchange chromatography).[236]

Unmodified NCPs octamers were made largely following the protocol in[236]. In short, equimolar amounts of histones H2A, H2B, H3 and H4 were mixed in a buffer containing 6M Guanidine HCl, 20 mM Tris (pH 7.5) and 10 mM DTT then dialyzed multiple times into a buffer containing 2M KCl, 20 mM Tris (pH 7.5), 1 mM EDTA and 5 mM β-ME. The unmodified octamers were further purified using size exclusion chromatography (sephacryl S-200, GE Healthcare Life Sciences).

Unmodified nucleosome core particles (NCPs) were reconstituted using the unmodified octamers and the Widom 601 DNA (147 bp) through a desalting method.[236] Unmodified octamers and the Widom 601 DNA (147 bp) were combined in equimolar ratios and desalted using a linear salt gradient from 2M KCl to 150 mM KCl over 48 hours. After reconstitution, NCPs were heat shocked for 30 minutes at 37 $^{\circ}$C for homogenous positioning on the Widom 601 DNA (147 bp). The NCPs were further purified using a 10-40% sucrose gradient. The purity and proper formation of unmodified NCPs was confirmed using native polyacrylamide gel electrophoresis.

4.2.6 NMR Spectroscopy

For assignment of backbone amide resonances, HNCACB and HN(CO)CACB experiments were collected on a 0.275 mM ^{15}N/^{13}C-CBX8 CD sample on a Bruker Avance II 500 MHz spectrometer with a 5mm triple resonance probe at 25 °C. The triple resonance experiments were processed in NMRpipe[237] and further analysis carried out using CcpNmr.[238] Initial backbone amide resonance assignments were generated using the PINE server[239] and further curated using CcpNmr.[238]

Titrations experiments were performed by collecting ^{15}N-HSQC spectra on 0.05-0.10 mM ^{15}N-CBX8 CD in the apo state and upon addition of increasing molar ratios of the respective ligands. Titrations experiments were performed at 25 °C (with the exception of H3K$_c$27me3 that was performed at 20 °C) on a Bruker Avance II 800 MHz NMR spectrometer equipped with a 5mm triple resonance cryoprobe or a Unity Inova 600 MHz Oxford AS600 spectrometer equipped with a 5mm triple resonance probe. The data was processed using NMRPipe[237] and further analysis performed using CcpNmr.[238] To determine dissociation constants (K_d), GraphPad PRISM was used for nonlinear least-squares analysis and the data fit to a single-site binding model accounting for ligand depletion using the equation:

$$\Delta\delta = \Delta\delta_{max}\left(([L] + [P] + K_d) - \sqrt{([L] + [P] + K_d)^2 - 4[P][L]}\right)/(2[P])$$

where [P] is the concentration of the CBX8 CD, [L] is the concentration of ligand, $\Delta\delta$ is the normalized chemical shift change and $\Delta\delta_{max}$ is the normalized chemical shift change at saturation, calculated as:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (0.20\Delta\delta_N)^2}$$

where Δδ is the chemical shift in parts per million (ppm).

Global K_d values were determined by averaging the individual K_d values for all resonances significantly perturbed in the titration experiments and reported as the average +/- 1.00 standard deviations. A resonance was considered significantly perturbed if the Δδ value was greater than the average Δδ + 1.00 standard deviations after trimming 10% of residues (5 resonances) with the largest Δδ value. Individual resonances with K_d values greater than 2.00 standard deviations from the mean of the global K_d were removed from the analysis. If 50% or more of the individual residue K_d values were greater than 2 standard deviations from the global mean K_d, than the global K_d was reported as lower limit (for example >0.8mM).

4.2.7 Cell Culture

HEK293T cells were cultured in Dubecco's Modified Essential Media (DMEM), 10% fetal bovine serum (FBS, JR Scientific), 1% glutagro (Corning), 1%

penicillin/streptomycin (Corning), 1% sodium pyruvate (Corning). GBM T98G cells were cultured in Eagle's Modified Essential Media, 10% FBS, 1% non-essential amino acids (Corning), 1% glutagro (Corning), 1% penicillin/streptomycin (Corning) and 1% sodium pyruvate (Corning). All cells were grown at 37 °C and 5% CO_2. For generation of CBX8 and control CRISPR knockout lines, 200,000 T98G cells were plated in 6-well 24 hrs prior to transfection. The respective vector (3.3 µg) was co-transfected with 13 µL of Fugene 6 (Promega). Media was changed 24 hrs post-transfection. Transfected cells underwent puromycin selection (2 µg/mL) for 3 days, 48 hrs post-transfection. EZH2 knockout lines were transduced with the MSCV_Cas9_puro vector (a gift from Chris Vakoc, Addgene plasmid #65655)[234] followed by the guide RNA vector.

4.2.8 Lentiviral transduction

HEK293T cells were transfected with TetOFUW constructs or pLenti CMV rtTA3 Hygro (w785-1) (a gift from Eric Campeau Addgene plasmid # 26730), and viral packaging vectors (pMD2.G and psPAX2). After 72 hrs, viral supernatant was harvested, and virus was concentrated by ultracentrifugation at 17,300 rpm for 2 hours. Virus was resuspended in PBS and both TetOFUW and rtTA were co-added to T98G cells and spun at room temperature for 1 hr at 200 x g. Forty-eight hours after infection, cells were selected with puromycin (2 µg/mL) and hygromycin (200 µg/mL) for a week to generate stable cell lines. The doxycycline was added every 48 hrs at 1 µg/mL to induce expression.

4.2.9 Sequential Salt Extraction

Protocol was performed as previously described.[218] Briefly, 5 mill cells were harvested, washed with PBS and lysed in buffer A (25 mM Hepes pH 7.6, 5 mM MgCl2, 25 mM KCl, 0.05 mM EDTA, 10% glycerol, 0.1% NP40) with protease inhibitor cocktail for 10 minutes at 4 °C. Nuclei were pelleted at 5500 x g for 5 min at 4 °C. The nuclei pellet was resuspended in 500 µL of mRIPA 0 mM NaCl (50 mM Tris pH 8.0, 1% NP40, 0.25% sodium deoxycholate) and incubated on ice for ~10 min. Chromatin was pelleted at 6000 x g for 5 min at 4 °C. Supernatant was collected, and the pellet was resuspended in mRIPA with 100 mM NaCl. This was repeated up to 500 mM NaCl. 4X SDS loading dye with beta-mercaptoethanol (BME) and immunoblot was performed. Immunoblots were

quantitated using ImageJ[240] to calculate the percentage of CBX8 in each fraction, and two-tailed Student's t-tests were performed for each salt extraction comparison (* $p < 0.05$, ** $p < 0.01$, ***$p < 0.001$, **** $p<0.0001$) using PRISM GraphPad. For the EZH2i SSEs, cells were treated for 48 hrs with 1 μM GSK343 (Caymen Chemical) or DMSO prior to being harvested.

4.2.10 Immunoprecipitation

T98G cell lines were harvested and washed with PBS. Cell membranes were lysed with buffer A with protease inhibitor cocktail for 15 minutes on ice. Nuclei were pelleted at 1000 x g for 5 min at 4 °C. The nuclei pellets were resuspended in IP buffer (25 mM tris, 300 mM NaCl, 1 mM EDTA, 1% NP-40) and lysed on ice for 20 min. Chromatin was pelleted and supernatant was collected. Lysate was pre-cleared with 10 μL of Protein A/G Magnetic Dynabeads (Pierce) overnight. Two hundred micrograms (200 μg) of lysate was incubated with 2 μg of V5 antibody (mouse, Invitrogen) for 2 hours at 4 °C. Following, 10 μL of Protein A/G Magnetic Dynabeads equilibrated in IP buffer were added for 1 hour. IPs were washed 3x with IP buffer and resuspended in 4X SDS loading dye with BME and immunoblot was performed.

4.2.11 Peptide Pulldowns

T98G mutant cell lines were washed with PBS and harvested with Buffer A. Cells were incubated in Buffer A for 15 min on ice. Nuclei were pelleted at 1000 x g for 5 min at 4 °C. The nuclei were re-suspended in IP buffer and lysed for 20 min on ice. Chromatin was pelleted and supernatant was collected. Biotinylated H3 (21-44) and H3K27me3 (21-44) peptides (2 μg, Anaspec) were pre-bound to streptavidin agarose resin (TriLink) for 1hr at 4 °C. Following incubation, lysate (125 μg for WT, R19A, R20A, R22A or 375 μg for 3xRA and W32A) was added and incubated for two hours at 4 °C in peptide pulldown buffer (50 mM Tris pH 6.6, 150 mM NaCl, 1% NP-40, 0.5 mM DTT). Pulldowns were washed 3 times for 5 min with peptide pulldown buffer. Samples were resuspended in 4X SDS loading dye with BME and immunoblot was performed.

4.2.12 Immunoblot and antibodies

Lysates were boiled and loaded on a 4-12% SDS-page gel (Invitrogen). Gels were transferred to PDVF membranes (Millipore) and incubated in 5% bovine serum albumin (BSA) in PBS-t (PBS with 0.1% Tween-20) prior to primary antibody. Blots were incubated at 4 °C overnight in primary antibody. Blots were washed with PBS-t and incubated for an hour at room temperature in goat anti-rabbit or mouse conjugated to IR800CW or IRDye 680 (LI-COR) secondary antibody. Blots were imaged on the Licor Odyessy. Primary antibodies used: CBX8 (rabbit, 1:1000, Bethyl), V5 (mouse, 1:5000, Invitrogen), V5 (rabbit, 1:1000, Cell Signaling Technologies), EZH2 (mouse, 1:1000, BioRad), H3 (rabbit, 1:5000, Active Motif), H3K27me3 (mouse, 1:5000, Epigenetik), CBX4 (rabbit, 1:500, Bethyl), CBX7 (rabbit, 1:1000, Bethyl), CBX2 (mouse, 1:400, Santa Cruz).

4.2.13 Electrophoretic mobility shift assays

Linear 601 DNA or reconstituted biotinylated mononucleosomes (Epicypher) were incubated with CBX8 chromodomain in EMSA buffer (25 mM phosphate buffer pH 6.8, 25 mM NaCl, 1 mM EDTA, 1 mM DTT) at varying molar ratios for 15 min on ice. Samples (10 µL) containing 0.5 pmol nucleosome/0.25 pmol DNA, varying concentrations of chromodomain, and 2x sucrose loading dye were loaded onto a prewarmed 6% (59:1) non-denaturing acrylamide gel and run in 0.4 X TBE (pH 8.3) at 100 V on ice for 1 hr. Gels were stained with ethidium bromide for 5 min and imaged on BioRad ChemiDoc.

4.2.14 Genome-wide data analysis

Published annotated datasets were downloaded from Encyclopedia of DNA Elements (ENCODE) or Gene Expression Omnibus (GEO) as BED files.[241,242] Peaks for CBX8, H3K4me3, H3K9me3, H3K27me3, and DNAse hypersensitivity were called in reference to GrCh38. All data sets were imported into R Studio. Peak overlaps were determined using the ChIPpeakAnno package.[243,244] Overlaps were defined as being within 150 base pairs of each other and on the same strand. Accession numbers for CBX8: GSM830987, GSM1295078, GSM1295089, ENCFF001SYX; H3K27me3: GSM1295084,

GSM1295094, ENCFF001SZF; H3K4me3: GSM1295085, GSM1295095, ENCFF001SZJ; H3K9me3: ENCFF001SZN; DNase-seq: ENCFF100IJK.

4.3 Results

4.3.1 CBX8 associates robustly with chromatin *in vivo* through its chromodomain

Previous studies have revealed that the CBX8 CD demonstrates very weak affinity *in vitro* for H3K27me3 and H3K9me3 peptides with indeterminate specificity.[103] However, live cell imaging suggests that H3K27me3 is important for CBX8 chromatin targeting.[231] To further assess chromatin specificity *in vivo*, we interrogated available CBX8 and corresponding histone modification genome-wide data sets to determine if CBX8 binding correlates to either H3K27me3 or H3K9me3. In the chronic myeloid leukemia cell line K562, 93% of the CBX8 peaks were found to overlap (within 150 base pairs) with H3K27me3 peaks (Figure 4.1a).[50] Additional analysis in the neonatal foreskin fibroblast cell line Hs68 and breast fibroblast cell line (BF),[63] revealed an 84% and 71% overlap, respectively (Figure 4.1a). In contrast, only 33% of CBX8 peaks correlate with the presence of H3K9me3 in the K562 cell line (Figure 4.1b).[50] There was a similarly low correlation between CBX8 peaks and the activating mark H3K4me3 in K562 cells (~35%) as well as Hs68 and BF cells (~15%) (Figure 4.1c). This suggests that CBX8 preferentially associates with H3K27me3 *in vivo.* It also suggests that since CBX8 interacts with only a subset of total H3K27me3 peaks, additional chromatin targeting mechanisms must be involved.

To further assess chromatin association of CBX8, we investigated its interaction with bulk chromatin using a sequential salt extraction (SSE) assay, in which relative binding is determined by the salt concentration required to elute proteins from chromatin.[218] We initially compared the binding profile of CBX8 to several other CBX paralogs in two distinct cell lines, human embryonic kidney (HEK) 293T and GBM T98G. We observed that endogenous CBX8 from T98G cells associates with chromatin similarly to the other paralogs (Figure 4.1d, left). Interestingly, in HEK293T cells, CBX8 demonstrates a stronger chromatin association compared to CBX2, CBX4, and CBX7

(Figure 4.1d, right) suggesting there may be cell-type specific contributions to chromatin binding.

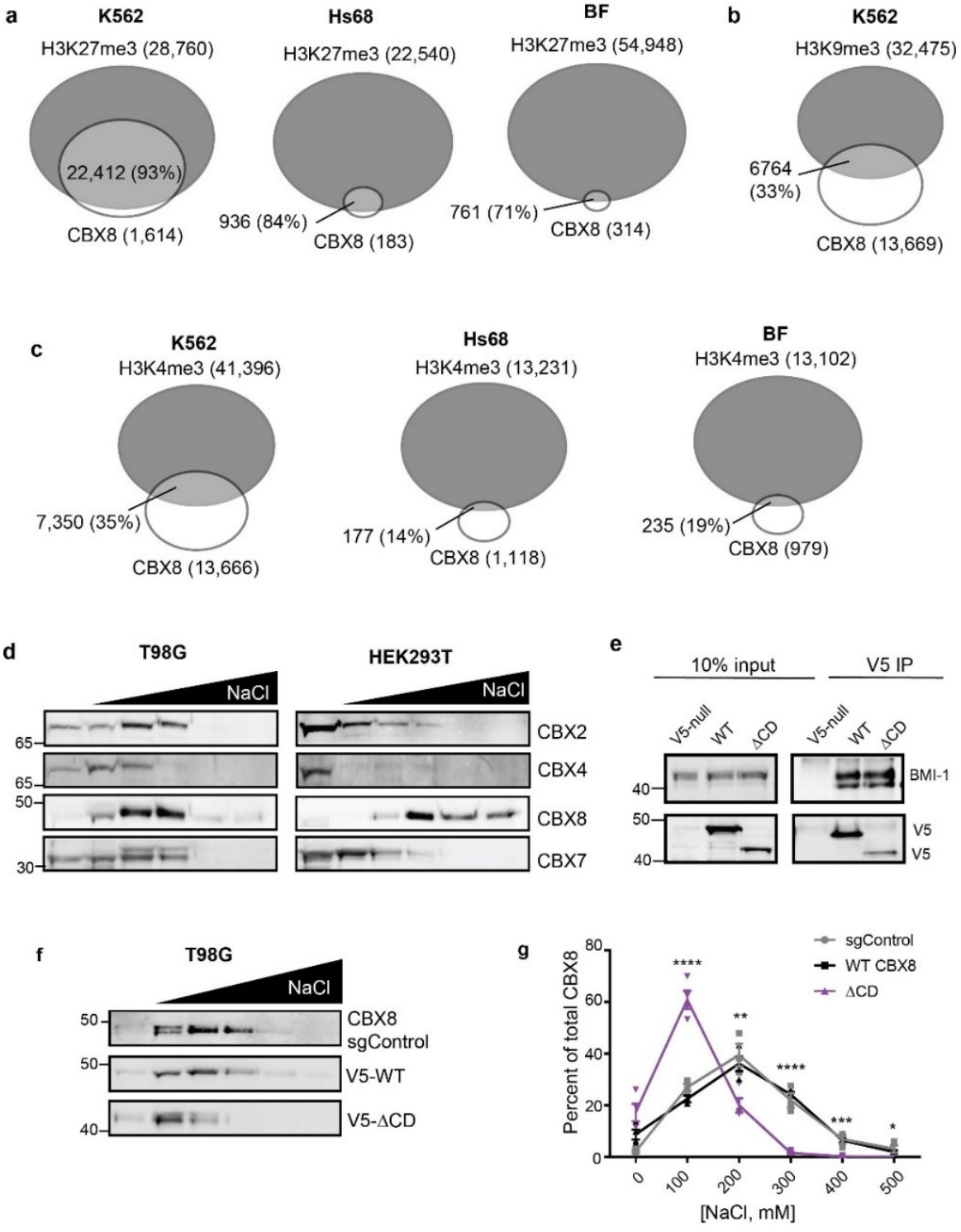
To evaluate the contribution of the CBX8 CD to this association, we deleted the corresponding amino acids 1-76 and re-expressed either this CD deletion (ΔCD) or wild-type (WT) CBX8 in a GBM T98G CBX8 knockout (sgCBX8) cell line using a doxycycline-inducible system. Importantly, both wild-type and ΔCD expressed CBX8 are incorporated normally into the PRC1 complex (Figure 4.1e). Using the SSE assay, we confirmed that WT CBX8 re-expression in T98G cells has a similar elution pattern to endogenous CBX8 (sgControl, Figure 4.1f) and that the ΔCD displays a reduction in bulk chromatin affinity compared to WT CBX8 (Figure 4.1f). In order to quantitatively assess binding differences, the amount of CBX8 in each fraction was assessed using ImageJ[240] and normalized to the total amount present (sum of all fractions), reported as the percentage of CBX8 eluting at each salt concentration. Quantitation revealed that the CD deletion significantly reduced the affinity of CBX8 for chromatin (Figure 4.1g) suggesting that the CD plays a critical role in CBX8 chromatin binding. Taken together, our data suggests that CBX8 specifies for H3K27me3 *in vivo* and that robust chromatin association is largely dependent on the CD.

4.3.2 The CD preferentially recognizes H3K27me3 *in vitro*

Previous *in vitro* studies demonstrated weak affinity of the CD for both H3K9me3 and H3K27me3 peptides. Because binding was beyond the limit of detection for the technique used, specificity was not distinguishable.[102,103] To further assess the CD specificity, we utilized nuclear magnetic resonance (NMR) spectroscopy, a technique well suited to study weak binding interactions. An initial ^{1}H-^{15}N-heteronuclear single quantum coherence (HSQC) spectrum of ^{15}N-CD reveals 49 main chain resonances of the 53 expected assuming fast conformational exchange. Resonances are well dispersed in both ^{1}H and ^{15}N dimensions, indicating that the CD is well folded and amenable to NMR studies. All 49 backbone amide resonances were assigned using HNCACB and HN(CO)CACB triple-resonance experiments (Appendix A Figure 1).

Figure 4.1 CBX8 robustly associates with chromatin via its chromodomain.

a) Genome-wide analysis of CBX8 and H3K27me3 peak overlaps in K562,[50] Hs68,[63] and BF[63] cell lines, respectively. Numbers indicate number of called peaks, overlap percent of total CBX8 peaks. b) Analysis of CBX8 and H3K9me3 overlaps in K562 cells c) CBX8 and H3K4me3 overlaps in K562, Hs68, and BF cells, respectively. d) Sequential salt extraction in T98G (left) and HEK293T (right) cells examining endogenous CBX paralog chromatin association. Paralog elution patterns were detected using immunoblot and antibodies against the respective paralog e) Anti-V5 co-immunoprecipitation of WT and ΔCD CBX8, with detection of the PRC1 subunit BMI-1 to verify PRC1 complex incorporation. f) Representative SSE of sgControl T98G cells (anti-CBX8), wild type and ΔCD re-expression (anti-V5) in T98G sgCBX8 cells g) Quantitation of SSEs displayed in (f), sgControl n= 3, WT and ΔCD n=4 biological replicates, errors bars represent standard error of the mean (SEM), p-values were calculated using Student's 2-tailed t-test * $p < 0.05$, ** $p < 0.01$,*** $p < 0.001$, **** $p < 0.0001$


a
K562
H3K27me3 (28,760)
22,412 (93%)
CBX8 (1,614)
Hs68
H3K27me3 (22,540)
936 (84%)
CBX8 (183)
BF
H3K27me3 (54,948)
761 (71%)
CBX8 (314)
b
K562
H3K9me3 (32,475)
6764
(33%)
CBX8 (13,669)
c
K562
H3K4me3 (41,396)
7,350 (35%)
CBX8 (13,666)
Hs68
H3K4me3 (13,231)
177 (14%)
CBX8 (1,118)
BF
H3K4me3 (13,102)
235 (19%)
CBX8 (979)
d
T98G
HEK293T
NaCl
NaCl
65
65
50
30
CBX2
CBX4
CBX8
CBX7
e
10% input
V5 IP
V5-null
WT
ΔCD
V5-null
WT
ΔCD
40
50
40
BMI-1
V5
V5
f
T98G
NaCl
50
50
40
CBX8
sgControl
V5-WT
V5-ΔCD
g
sgControl
WT CBX8
ΔCD
Percent of total CBX8
80
60
40
20
0
****
**
****
***
*
0
100
200
300
400
500
[NaCl, mM]

To investigate histone tail binding and CD specificity, sequential ^{15}N-HSQC spectra were collected on ^{15}N-labeled CD upon addition of peptides corresponding to unmodified H3 (residues 1-44), H3K9me3 (residues 1-21), or H3K27me3 (residues 23-34). Addition of increasing concentrations of the H3K9me3 and H3K27me3 peptides resulted in significant chemical shift perturbations (CSPs) in the CD spectrum, suggesting an interaction with both peptides (Figure 4.2a and Appendix A Figure 2a). In contrast, addition of the unmodified H3 peptide resulted in no significant CSPs, indicating that binding is methylation dependent (Figure 4.2a, Appendix A Figure 2a). Dissociation constants (K_d) were calculated for H3K9me3 and H3K27me3 by fitting normalized CSPs ($\Delta\delta$) to a one-site binding model accounting for ligand depletion (see methods for details). This yielded a K_d=0.7 mM for H3K27me3 and K_d>2.0 mM for H3K9me3 (Appendix A Figure 2b, Appendix A Figure 2c). Thus, though binding is weak to both modifications, the CD preferentially binds H3K27me3 *in vitro* (Figure 4.2b).

To determine the structural basis of complex formation and specificity, CSPs were further assessed for both peptides. Plotting normalized CSPs between the apo and peptide-bound CD as a function of primary amino acid sequence shows that both H3K9me3 and H3K27me3 titrations resulted in CSPs at residues in the N-terminal portion of the β1 strand and the β2/β3 loop of the CD (Figure 4.2c). When plotted onto the crystal structure of the CD previously solved in complex with H3K9me3 (PDBID:3I91), these residues cluster in and around the aromatic cage consistent with the canonical histone binding pocket (Figure 4.2d,e), as well as in residues at the N-terminus of the αA helix.[103,118] Notably, these primarily hydrophobic residues in the αA helix are substantially more perturbed upon binding H3K27me3 than H3K9me3, suggesting that these residues are important for specificity (Figure 4.2a). The two methylated lysines have significant sequence similarity in surrounding residues with sequences of AARKme3S (H3K27me3) and TARKme3S (H3K9me3). Notably, in the crystal structure of the CD bound to H3K9me3 (PDBID 3I91)[103], the polar H3T6 (-3 position) is in close proximity to these hydrophobic residues. Thus, it follows that the more hydrophobic alanine found at the -3 position of H3K27me3 (A24) would be preferred. Indeed, previous studies found mutating H3A24 to T24 results in significantly reduced affinity for H3K27me3.[103] In addition, a recent peptide inhibitor screen towards the CD identified high affinity peptide mimetics that utilized bulky

aromatic moieties at the -3 position.[222] The large CSPs around residues V10-A13 also support a conformational change seen in the previously published crystal structure. Specifically, the structure of the CD without H3K9me3, shows that these residues are disordered, resulting in an only partially formed aromatic cage with F11 positioned in the methyl-lysine binding pocket. In comparison, in the co-crystal structure of CBX8 bound to H3K9me3, V10-A13 are structured and form part of an extended β1 strand allowing for proper orientation of F11 and complete formation of the aromatic cage.[103] Our NMR data is consistent with this conformational change.

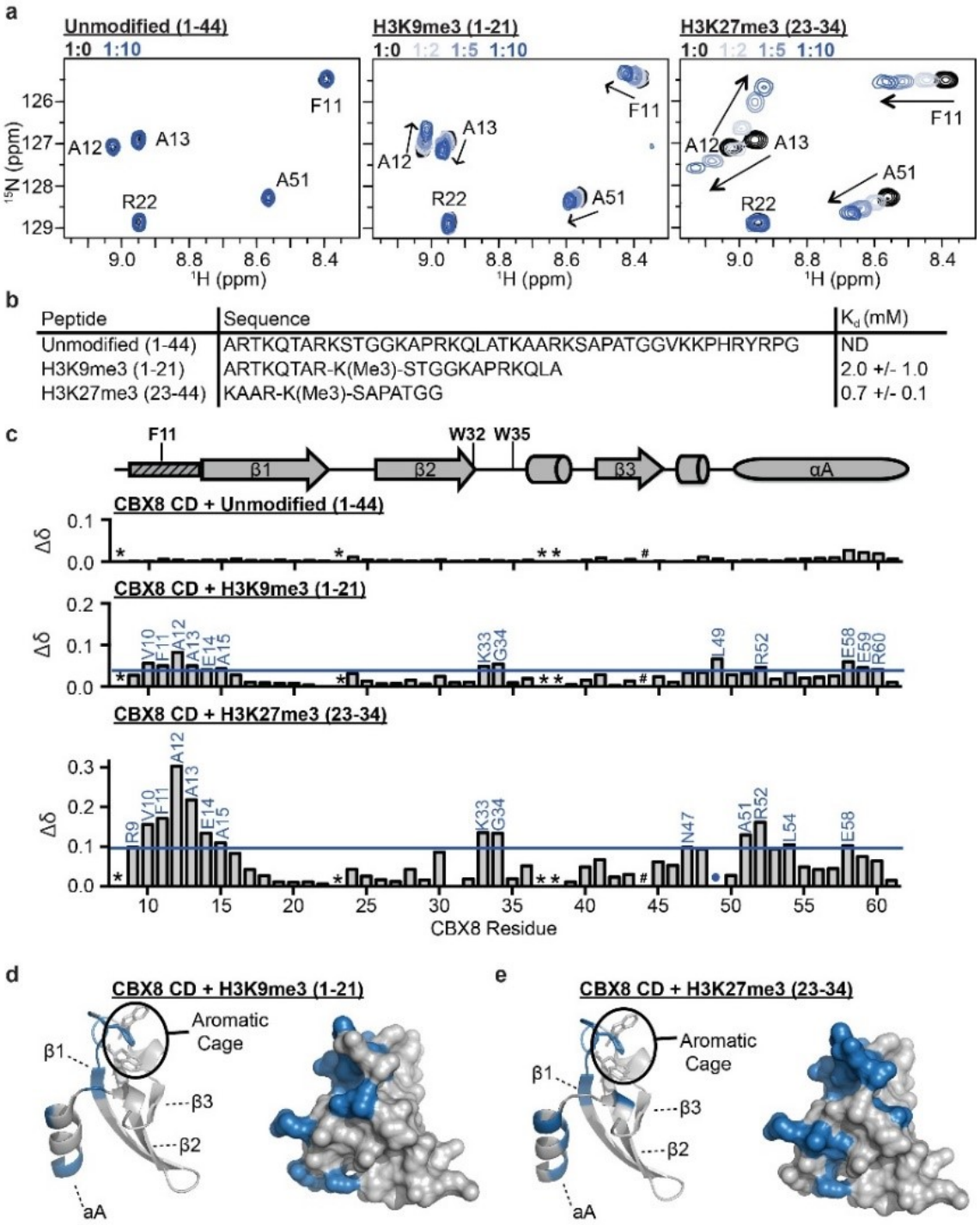
Taken together, our data reveals that though the CD binds weakly it has specificity for H3K27me3 *in vitro*, consistent with the *in vivo* data from genome-wide associations.

4.3.3 The CD association with nucleosomes is driven by interactions with DNA

Our NMR analysis reveals that the CD preferentially binds H3K27me3 *in vitro*; however, the generally weak affinity for histone tail peptides *in vitro* is inconsistent with this driving robust chromatin association. In order to determine if the CD makes additional contacts outside of the histone tail, we investigated binding in the context of methylated nucleosomes. Notably, a commonly utilized amino-ethylcysteine methyl-lysine analog (MLA) was not bound by the CD as has been seen with a subset of other reader domains (Appendix A Figure 2a).[245,246] Thus, electrophoretic mobility shift assays (EMSAs) were performed with nucleosome core particles (NCPs) reconstituted with the 147bp 601 DNA and unmodified H3 or methylated H3 (H3K9me3 or H3K27me3) generated by native chemical ligation. The addition of increasing concentrations of the CD resulted in significant changes in the mobility of all NCPs tested (Figure 4.3a, Appendix A Figure 3a), independent of methylation status. Notably, NCPs containing H3K9me3 and H3K27me3 resulted in changes in mobility at similar concentrations as the unmodified NCP, suggesting that robust association by the CD is driven primarily by interactions outside of the histone methyl mark. EMSAs of the CD with 601 DNA alone also indicated a robust interaction (Figure 4.3b, Appendix A Figure 3b). Together, this suggests that NCP binding by CD is driven primarily by interactions with DNA.

Figure 4.2 CD preferentially recognizes H3K27me3 *in vitro*. Data collected and analyzed by Tyler M. Weaver

a) ^{1}H-^{15}N-HSQC overlays for ^{15}N-CD upon addition of increasing concentrations of unmodified H3 (1-44, left), H3K9me3 (1-21, middle) or H3K27me3 (23-34, right) histone peptides. A selected region of each spectra is shown for clarity. Molar ratio of CD:peptide is color coded as shown in legend. b) Table of determined K_d values and associated standard deviations for histone peptides tested. K_d values are reported as the average K_d value determined from analysis of the CSPs for all CD residues significantly perturbed. ND indicates binding not detected. c) The normalized chemical shift perturbation ($\Delta\delta$) between the apo and bound states (1:10 ratio) of CD resonances plotted against CBX8 residue number for each histone peptide tested. CSPs were considered significant if greater than one the mean plus one standard deviation and are labeled in blue. The secondary structure of CD from the crystal structure PDBID 3I91 is diagramed above the $\Delta\delta$ plots with the aromatic cage residues labeled. The small rectangle with dashed lines represents the region of CD that undergoes a conformational change between apo and histone bound states in the crystal structure. * indicates missing resonances, # indicates proline residue and blue dots represent resonances that broaden beyond detection during the experiment. d) Residues with significant CSPs upon addition of H3K9me3 (left) or H3K27me3 (right) highlighted in blue on a cartoon representation of the CD (PDB 3I91).


a
Unmodified (1-44)
1:0 1:10
H3K9me3 (1-21)
1:0 1:2 1:5 1:10
H3K27me3 (23-34)
1:0 1:2 1:5 1:10
F11
A12
A13
A51
R22
15N (ppm)
1H (ppm)
b
Peptide | Sequence | Kd (mM)
Unmodified (1-44) | ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPG | ND
H3K9me3 (1-21) | ARTKQTAR-K(Me3)-STGGKAPRKQLA | 2.0 +/- 1.0
H3K27me3 (23-44) | KAAR-K(Me3)-SAPATGG | 0.7 +/- 0.1
c
F11
W32 W35
β1
β2
β3
αA
CBX8 CD + Unmodified (1-44)
CBX8 CD + H3K9me3 (1-21)
CBX8 CD + H3K27me3 (23-34)
Δδ
CBX8 Residue
d
CBX8 CD + H3K9me3 (1-21)
Aromatic Cage
β1
β3
β2
aA
e
CBX8 CD + H3K27me3 (23-34)
Aromatic Cage
β1
β3
β2
aA

4.3.4 The CD binds linear and nucleosomal DNA through an arginine-rich basic patch

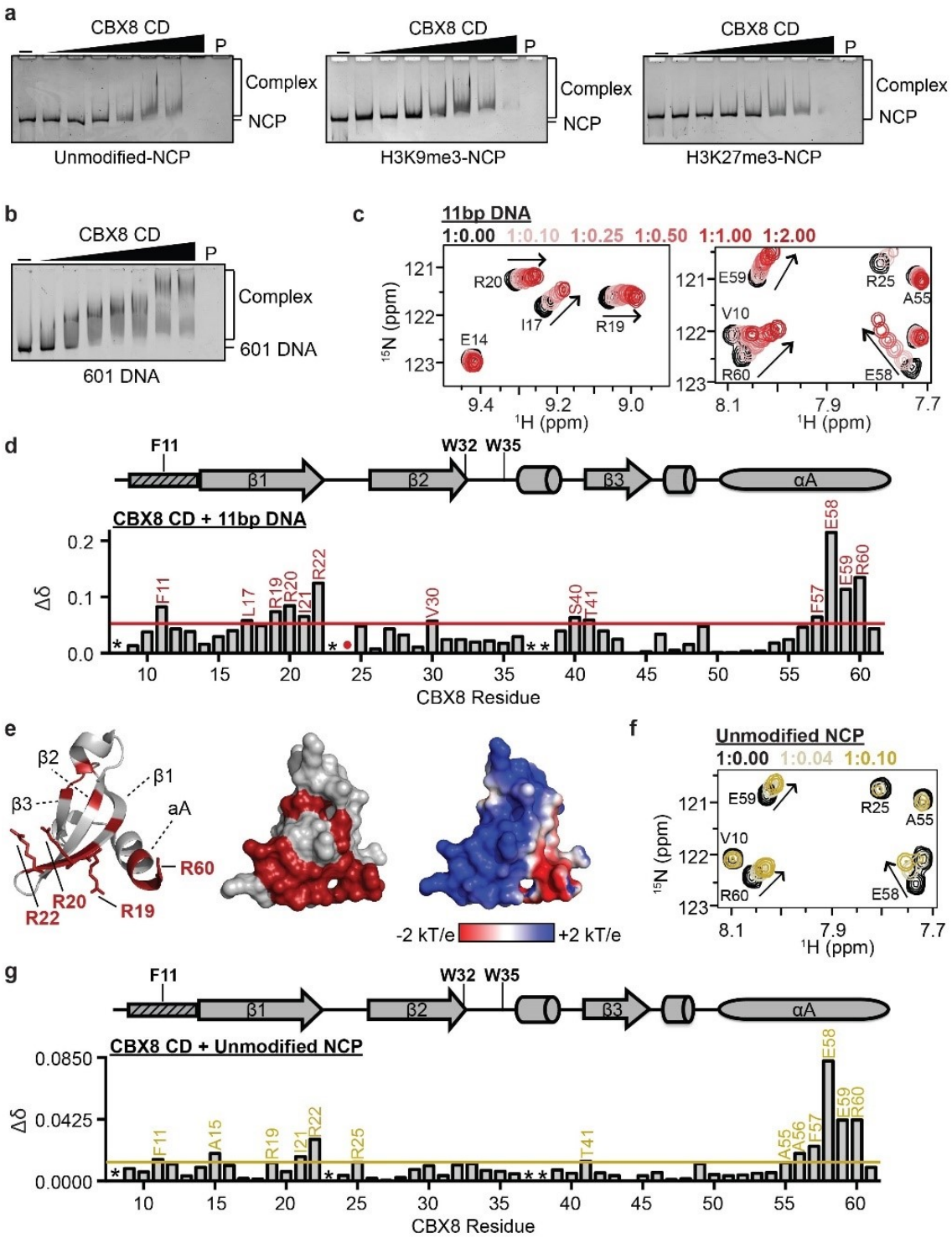
In order to further characterize the DNA binding by the CD, we performed NMR titration experiments with an 11bp DNA segment. Addition of increasing concentrations of unlabeled 11bp DNA into ^{15}N-CD resulted in significant CSPs for a subset of CD resonances (Figure 4.3c, Appendix A Figure 3c). Interestingly, for some resonances CSPs demonstrated curvature, suggesting binding that is more complex than a simple one-to-one complex. This makes the calculation of K_d values from the CSPs difficult, however as compared to the titration with H3K27me3 in which saturation is reached at 1:7 molar ratio, saturation with DNA is reached at ~1:1 molar ratio suggesting that the CD binds DNA with a substantially higher affinity than H3K27me3, likely with a K_d<10 μM.

To determine the structural basis for interaction with DNA, CSPs between the free and DNA-bound states were assessed as a function of CD sequence. Residues with significant CSPs localized to the C-terminal portion of the β1 strand and the αA helix (Figure 4.3d). Mapping the residues with significant CSPs onto the structure of the CD revealed that DNA binding is mediated through a cluster of arginine residues (R19, R20, R22 and R60) that form a basic patch on the opposite face of the histone binding pocket (Figure 4.3e).

To determine if the CD interaction with unmodified nucleosomes is indeed being mediated through DNA binding as suggested in the EMSAs, an NMR titration was performed using recombinant unlabeled, unmodified NCPs (reconstituted with the 147bp 601 sequence). Addition of increasing concentrations of unmodified NCPs resulted in significant CSPs in a subset of residues indicating binding, accompanied by global intensity decreases, which are expected due to the size of the complex (Figure 4.3f, Appendix A Figure 3c). Analysis of CSPs as a function of residue revealed an almost identical binding pocket as was observed with the 11bp DNA (compare Figures 4.3d and 4.3g), and notably the CSPs followed identical trajectories (compare Figure 4.3c and 4.3f) suggesting a similar binding mechanism. Together this confirms that NCP binding is driven through contacts with DNA, mediated by a basic patch on the CD surface.

Figure 4.3 CD association with nucleosomes is driven by interactions with DNA through an arginine-rich basic patch.

a) EMSAs performed with CD and unmodified (left), H3K9me3 (middle) and H3K27me3 (right) NCPs. Shown are representative gels from a triplicate of experiments. b) EMSAs performed with CD and the 147bp 601 DNA. Shown is a representative gel from a triplicate of experiments. c) 1H-15N-HSQC overlays for 15N-CD upon addition of increasing concentrations of an 11bp DNA. Molar ratio of CD:DNA are color coded as shown in legend. Two selected regions of the CD spectrum are shown for clarity. d) Normalized CSP (Δδ) between the apo and bound (1:2.00 ratio) spectra are plotted against CBX8 residue number. CSPs were considered significant if greater than the mean plus one standard deviation and are labeled in red. e) Residues with significant CSPs upon addition of the 11bp DNA plotted onto cartoon representation (left) or surface representation (middle) of CD (PDBID 3I91) and colored red. Arginine residues that form the basic patch are shown as sticks. APBS surface electrostatic representation of the CD (PDBID 3I91) is shown on the right. f) 1H-15N-HSQC overlays for 15N-CD upon addition of increasing concentrations of recombinant, unmodified NCPs. Molar ratio of CD:NCP are color coded as shown in legend. g) normalized CSP (Δδ) between the apo and bound (1:0.1 ratio) spectra are plotted against CBX8 residue number. CSPs were considered significant if greater than the mean plus one standard deviation and are labeled in gold. For d) and g) the secondary structure of CD from the crystal structure PDBID 3I91 is diagramed above the Δδ plot with the aromatic cage residues labeled. The small rectangle with dashed lines represents the region of CD that undergoes a conformational change between apo and histone bound states in the crystal structure. * indicates missing resonances, # indicates proline residue and red/gold dots represent resonances that broaden beyond detection during the experiment. c-g) collected and analyzed by Tyler M. Weaver


a
CBX8 CD
P
Complex
NCP
Unmodified-NCP
H3K9me3-NCP
H3K27me3-NCP
b
601 DNA
c
11bp DNA
1:0.00 1:0.10 1:0.25 1:0.50 1:1.00 1:2.00
15N (ppm)
1H (ppm)
R20
I17
R19
E14
E59
R25
A55
V10
R60
E58
d
F11
W32 W35
β1
β2
β3
αA
CBX8 CD + 11bp DNA
Δδ
CBX8 Residue
e
β2
β1
β3
aA
R60
R20
R22
R19
-2 kT/e
+2 kT/e
f
Unmodified NCP
1:0.00 1:0.04 1:0.10
g
CBX8 CD + Unmodified NCP

4.3.5 The CD can interact with DNA and H3K27me3 simultaneously

Comparison of CSPs for the CD upon binding to histone peptide or DNA revealed largely non-overlapping binding sites for DNA and H3K27me3, except for residues F11 and E58, which show significant CSPs upon binding either ligand (compare Figure 4.2c and 4.3d). This suggests that both may bind contemporaneously. To test this, NMR titrations were performed in which increasing concentrations of H3K27me3 peptide was added to ^{15}N-CD pre-bound to the 11bp DNA. This resulted in significant CSPs indicating that DNA-bound CD can interact with the H3K27me3 peptide (Figure 4.4a, compare red and purple spectra). Comparison of spectra for the CD in the presence of H3K27me3 alone, DNA alone, or both H3K27me3 and DNA revealed unique chemical shift values for most CD residues (Figure 4.4a, Appendix A Figure 4d). This is consistent with the formation of a ternary complex but notably suggests that the bound state is unique as compared to either binary complexes. Residues with significant CSPs upon addition H3K27me3 (23-34) to the CD pre-bound with DNA largely mapped to the determined histone binding pocket (compare Figure 4.2c and 4.4b). However, smaller CSPs were also seen in the DNA binding pocket. Notably, these CSPs are on a trajectory toward a state unique from the apo, DNA, or peptide bound states, suggesting that the DNA and histone binding are not completely independent of each other (Figure 4.4a middle, Figure 4.4b). Consistent with this finding, analysis of the CSPs for titration of H3K27me3 into the DNA-bound CD revealed an affinity of K_d=0.2 mM, ~3× tighter than that observed to the CD alone (Appendix A Figure 4f). This increase in affinity may in part be due to an increased local concentration of peptide, as histone tails are known to bind to DNA.[247–250] However, it may also be due to DNA mediated stabilization of the methyl-lysine binding pocket. As mentioned above, the CD aromatic cage is ill-formed in the absence of histone peptide, with residue F11 pointing into the aromatic cage, which must rearrange upon binding the methylated lysine. F11 is one of two residues observed to have significant CSPs upon binding of either DNA or H3K27me3, and notably, the perturbation upon DNA binding follows a similar trajectory to that seen upon titration of H3K27me3. This suggests that DNA binding may stabilize the aromatic cage, increasing the affinity for peptide. Importantly however, a ternary complex and increase in affinity are also seen with

H3K9me3, indicating that this does not lead to increased specificity for methyl marks (Appendix A Figure 4a-c,e). Taken together, our data indicated the CD can simultaneously bind DNA and H3K27me3 and that DNA binding enhances methyl-lysine binding.

4.3.6 The CD interaction with DNA and H3K27me3 are important for CBX8 chromatin association

To investigate the importance of the CD DNA and histone binding in CBX8 chromatin association *in vivo,* SSE assays were carried out upon disrupting either the histone or DNA interaction of the CD. To initially assess the importance of H3K27me3 binding on CBX8 chromatin association, we generated a stable EZH2 knockout (sgEZH2) in T98G cells to reduce global H3K27me3 levels (Figure 4.5a) and performed sequential salt extractions on CBX8. Upon knockout of EZH2, we observed a decrease in the affinity of CBX8 for bulk chromatin compared to the sgControl cell line (Figure 4.5b). Quantification of the salt extractions indicates a significant reduction in the affinity of CBX8 for chromatin upon loss of H3K27me3 (Figure 4.5c). We additionally performed SSE with the WT CBX8 re-expression cells treated for 48 hours with the EZH2 inhibitor (EZH2i) GSK343 (Appendix A Figure 5a). In the presence of GSK343, we observed a significant decrease in the affinity of CBX8 for chromatin (Appendix A Figure 5b) that mimics the EZH2 knockout (Appendix A Figure 5c).

To further examine the role of methyl-lysine binding in CBX8 chromatin association, we generated a previously published aromatic cage mutation, W32A, that disrupts CBX chromatin localization.[102,132] The V5-tagged W32A mutant was re-expressed in the T98G CBX8 knockout cell line and incorporated into the PRC1 complex (Appendix A Figure 5d). Peptide pulldowns confirm that W32A does not bind H3K27me3 (Appendix A Figure 5e). SSE with the V5-W32A expressing cells demonstrated a significant reduction in CBX8 chromatin binding (Figure 4.5d, Figure 4.5e). When compared to the binding profile of CBX8 in sgEZH2 cells, we observe a similar elution pattern albeit more pronounced likely due to the incomplete H3K27me3 depletion in the sgEZH2 cells (Appendix A Figure 5f). Importantly, the W32A mutant did not reduce CBX8 chromatin affinity to the same severity of the ΔCD (Appendix A Figure 5g). Together this data shows that despite the weak affinity of the CD for H3K27me3 in the context of peptides and mononucleosomes, this interaction is indeed important for CBX8 chromatin association.

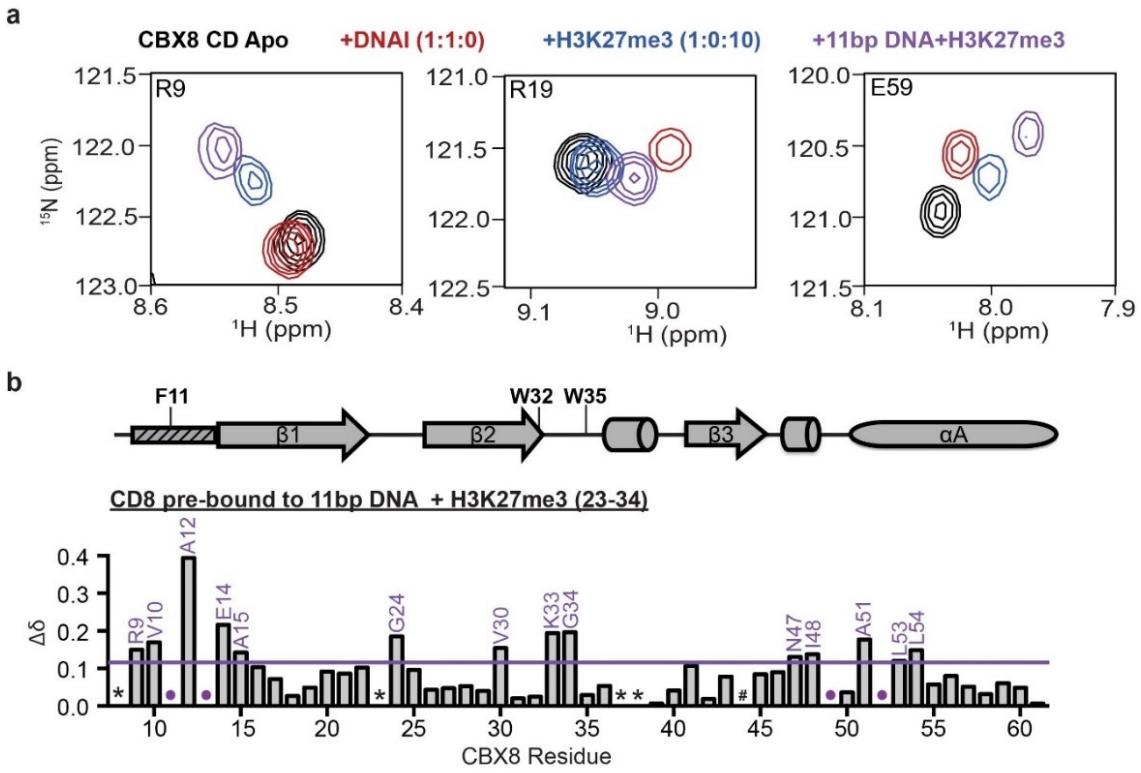


Figure 4.4 CD can interact with DNA and H3K27me3 simultaneously. Data collected and analyzed by Tyler M. Weaver

a) 1H-15N-HSQC overlays for 15N-CD in the apo (black, 1:0:0 ratio), bound to an 11bp DNA (red, 1:1:0 ratio), bound to H3K27me3 (blue, 1:0:10), or bound to both 11bp DNA and H3K27me3 (purple, 1:1:10 ratio). Shown are resonances for selected residues in the histone binding pocket (R9, left), DNA binding pocket (R19, middle) and a residue sensitive to both DNA and histone binding (E59, right) are shown. b) normalized CSP (Δ□) between the DNA-bound (1:1:0) and DNA and H3K27me3 bound (1:1:10 ratio) plotted against CBX8 residue number. CSPs were considered significant if greater than the mean plus one standard deviation and are labeled in purple. The secondary structure of CD from the crystal structure PDBID 3I91 is diagramed above the Δ□ plot with the aromatic cage residues labeled. The small rectangle with dashed lines represents the region of CD that undergoes a conformational change between apo and histone bound states in the crystal structure. * indicates missing resonances, # indicates proline residue and purple dots represent resonances that broaden beyond detection during the experiment.

These results also indicate that chromatin binding is not completely abrogated in the absence of methyl-lysine binding, suggesting that H3K27me3 is not the only contributing factor in CBX8 association with chromatin.

To assess the importance of the CD-DNA interaction in CBX8 chromatin association *in vivo,* we mutated residues identified as important for binding in the NMR

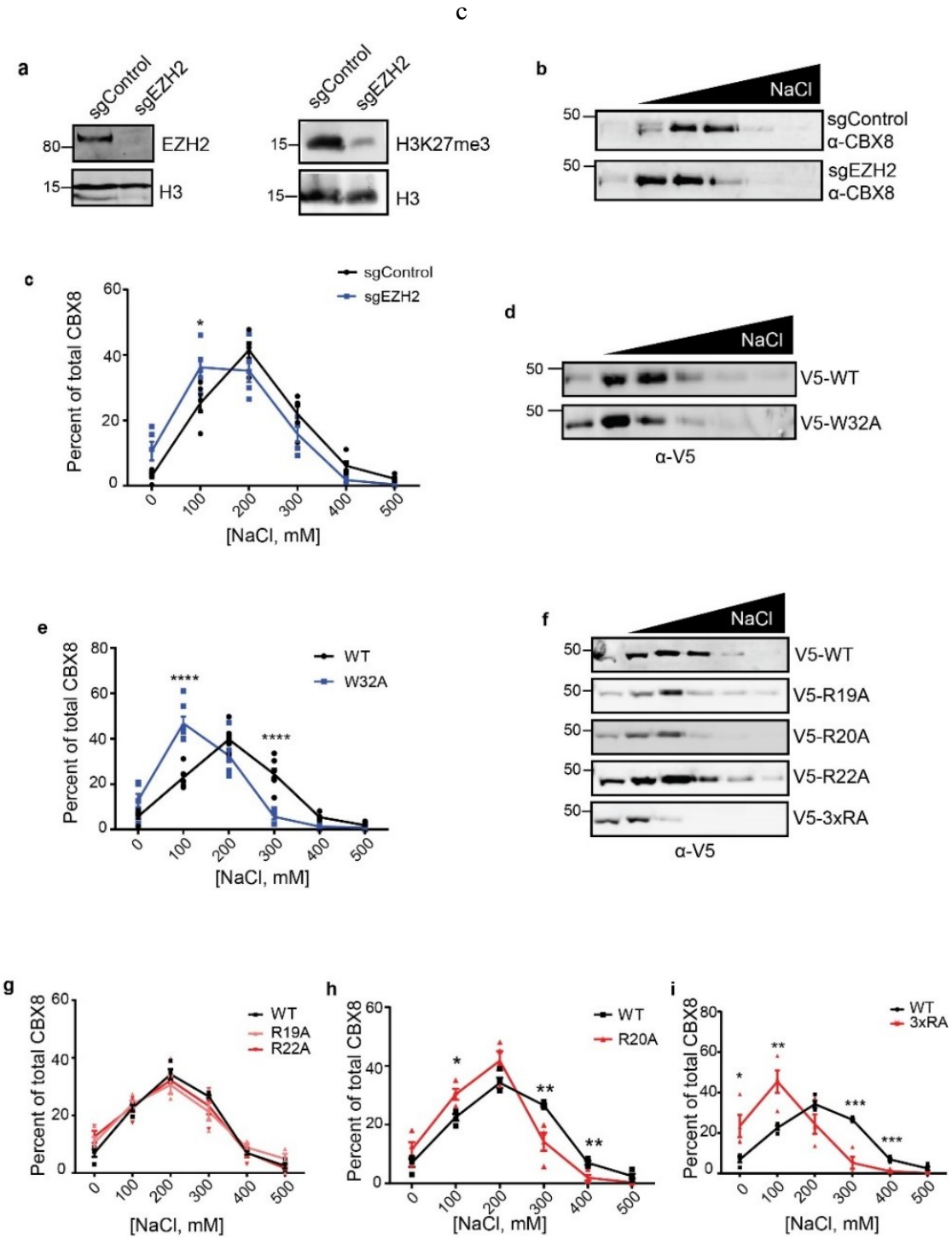
studies. Specifically, R19, R20, and R22 were mutated to alanine individually (R19A, R20A, R22A) or together (3XRA). The V5-tagged mutants were re-expressed in the T98G CBX8 knockout cell line earlier described and incorporated into the PRC1 complex as observed by immunoprecipitation (Appendix A Figure 5d). As expected, these mutants still engaged H3K27me3 in a peptide pulldown assay (Appendix A Figure 5e).

To examine the contribution to bulk chromatin binding affinity, SSE assays were carried out with all mutants (Figure 4.5f). No significant changes were observed in bulk chromatin binding for R19A or R22A (Figure 4.5g), however R20A led to a small but significant increase in the amount of CBX8 eluted at 100 mM NaCl compared to WT (Figure 4.5h). In comparison, the triple arginine mutant (3xRA) was consistently eluted at lower salt concentration as compared to WT (Figure 4.5i) indicating significantly weakened chromatin binding. Interestingly, when compared to the ΔCD SSE, there was not a significant difference in affinity for chromatin (Appendix A Figure 5h). This suggests that CBX8 chromatin binding is driven through DNA interactions. Together, these data suggest that both DNA and histone binding by the CD are critical for the association of CBX8 with chromatin.

Our findings reveal that the CBX8 CD engages both DNA and H3K27me3 to drive chromatin association of CBX8. The high affinity with which it associates with DNA as compared to H3K27me3 implies that DNA accessibility could be a major factor in determining the subset of H3K27me3 regions bound by CBX8. In support of this model, analysis of genome-wide data from K562 cells revealed that 82% of CBX8 peaks correspond to regions of DNase hypersensitivity[241,242] (Figure 4.6a) and that the majority of the H3K27me3 peaks with DNase hypersensitivity are bound by CBX8 (Figure 4.6a). Based on these findings, the DNA binding ability of the CD may lead CBX8 to be selectively involved in chromatin compaction and gene silencing at genes with H3K27me3 in regions of accessible DNA, whereas the histone binding activity would act to retain it at H3K27me3 enriched regions (Figure 4.6b).

Figure 4.5 Methyllysine and DNA binding contribute to CBX8 chromatin association.

a) Immunoblots of sgEZH2 KO and sgControl cell lysate demonstrating EZH2 knockout efficiency (anti-EZH2, anti-H3 loading control) (left) and reduced H3K27me3 levels (anti-H3K27me3, anti-H3 loading control) (right). b) Representative SSE of endogenous CBX8 in T98G sgEZH2 and sgControl cells (anti-CBX8). c) Quantitation of CBX8 eluted in each fraction as a percent of total CBX8, n= 5 biological replicates, blue denotes sgEZH2 cell line, black denotes sgControl cell line. d) Representative immunoblot of W32A SSE (anti-V5) in T98G sgCBX8 cells. e) Quantitation of amount of CBX8 in each fraction as a percent of total CBX8, n =4 ,7 biological replicates for WT and W32A, respectively. WT is denoted in black and W32A is denoted in blue. f) Representative SSE of DNA binding mutants (anti-V5) in T98G sgCBX8 cells. g,h,i) Quantitation of amount of CBX8 in each fraction as percent of total CBX8, n=4 biological replicates for each mutant compared to WT. DNA mutants are denoted in red (R19A, light red), WT is denoted in black. For all quantitation, error bars represent SEM, p-values were calculated using two-tailed Student's t-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$


c
a
sgControl
sgEZH2
80
EZH2
15
H3
sgControl
sgEZH2
15
H3K27me3
15
H3
b
NaCl
50
sgControl
α-CBX8
50
sgEZH2
α-CBX8
c
sgControl
sgEZH2
Percent of total CBX8
60
40
20
0
*
0
100
200
300
400
500
[NaCl, mM]
d
NaCl
50
V5-WT
50
V5-W32A
α-V5
e
WT
W32A
Percent of total CBX8
80
60
40
20
0
****
****
0
100
200
300
400
500
[NaCl, mM]
f
NaCl
50
V5-WT
50
V5-R19A
50
V5-R20A
50
V5-R22A
50
V5-3xRA
α-V5
g
WT
R19A
R22A
Percent of total CBX8
60
40
20
0
0
100
200
300
400
500
[NaCl, mM]
h
WT
R20A
Percent of total CBX8
60
40
20
0
*
**
**
0
100
200
300
400
500
[NaCl, mM]
i
WT
3xRA
Percent of total CBX8
80
60
40
20
0
*
**
***
***
0
100
200
300
400
500
[NaCl, mM]

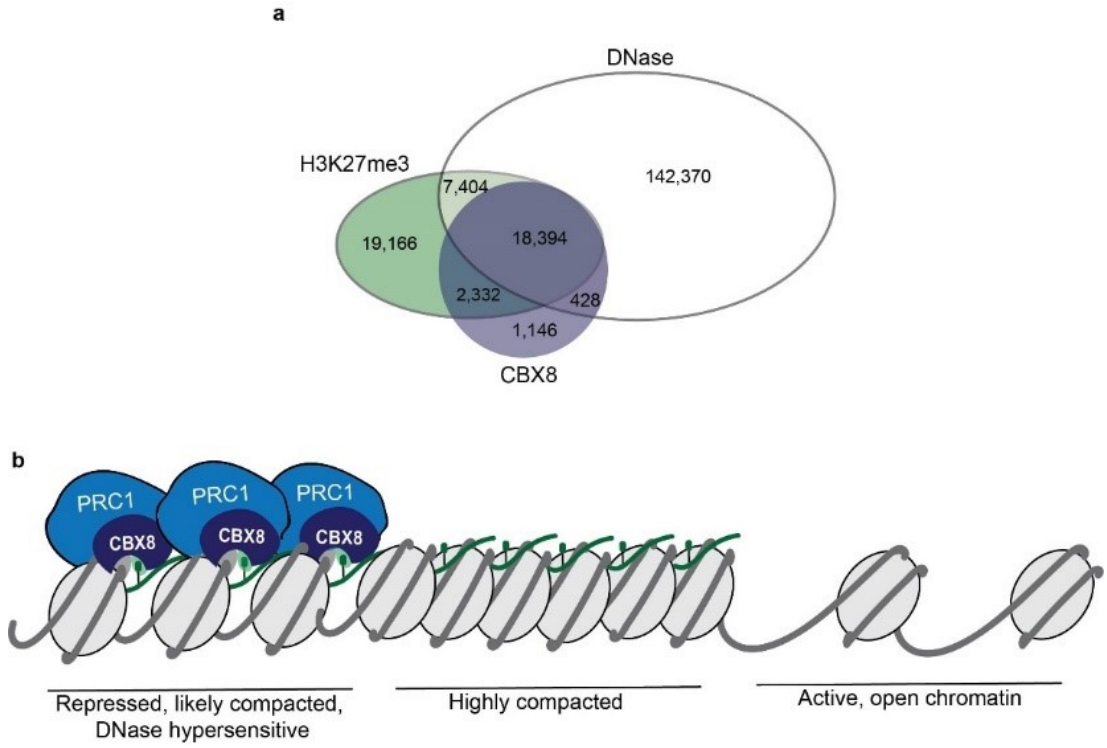


Figure 4.6 CBX8 and H3K27me3 associate with accessible DNA.

a) Overlap of DNase hypersensitive sites, H3K27me3, and CBX8 in K562 cells. b) Proposed model for the engagement of CBX8 with chromatin. PRC1 is shown in blue, with the CBX8 subunit in dark blue, engaging both H3K27me3 (green sphere) and DNA (gray). We propose that CBX8 will preferentially associate with DNase hypersensitive regions that are enriched in H3K27me3.

4.4 Discussion

Our findings in this study reveal a dual interaction mechanism for the CBX8 chromodomain, wherein engagement of both DNA and H3K27me3 mediate CBX8 association with chromatin. Importantly, we have shown that both binding modes are important for recovering full chromatin affinity *in vivo*. Our work bridges findings from two previous studies with conflicting conclusions regarding the importance of H3K27me3 for CBX8 chromatin association.[97,231] Namely, bimolecular fluorescence complementation work suggested H3K27me3 is not necessary for chromatin association,[97] while more recent kinetic studies conclude that H3K27me3 is important for CBX8 chromatin association.[231] We find that H3K27me3 is indeed important for chromatin association, but is not entirely necessary, and that DNA binding contributes much of the binding energy of CD to

nucleosomes. This is consistent with other studies, which have found that CBX proteins can bind nucleic acids, specifically RNA, via their chromodomains.[102,145] In addition, several other CDs have been found to bind a variety of nucleic acid substrates (reviewed in [251]). Two main structural mechanisms for nucleic acid binding have been determined including the utilization of a highly basic αA helix[252–254] and interfaces that are partially overlapping with the methyl-lysine binding site.[145,255] Notably, our data indicates a third structural mechanism of nucleic acid binding via an arginine-rich basic patch formed by the C-terminal portion of the β1 strand and αA helix.

There are two major implications from these results regarding the mechanism of reader domain mediated chromatin association. The first is the importance of DNA binding by canonical histone reader domains. Several studies have now demonstrated that histone reader domains, including chromodomains, bromodomains, PWWP domains, Tudor domains and PHD fingers, can interact with nucleic acids and that this binding is often much stronger than for histone peptides.[251] Though the importance of this nucleic acid binding is not yet clear for all of these domains, our results indicate that one potential function is chromatin association independent of histone binding. This highlights that it is critical to study reader domain function in the context of the nucleosome.

The second major implication is that the local chromatin environment is likely critical in determining the contribution of reader domain binding to chromatin affinity. Despite the very weak affinity and moderate specificity for H3K27me3 to CBX8 CD *in vitro*, this binding interaction contributes significantly to the affinity of CBX8 for chromatin *in vivo*. This could be in part related to the increased effective concentration of repressive modifications in large regions of heterochromatin. For example HP1α is retained at H3K9me3-containing nucleosome arrays upon array compaction, due to both multivalent binding and probability of rebinding.[256] This is similar to what is observed with bromodomains, which are often seen only to contribute substantially to chromatin affinity under conditions of hyper-acetylation (see Philpott et al., for example)[257] CBX proteins may be similarly retained in large heterochromatin domains marked enriched in H3K27me3 *in vivo*.[92] In addition, the weak specificity for H3K27me3 over H3K9me3 *in vitro* may be amplified by a combination of factors. For instance, HP1, which specifically associates with H3K9me3, will likely block CBX8 association this modification. In

addition, as discussed above, a higher level of accessible DNA in H3K27me3 enriched regions as compared to H3K9me3 enriched regions could also bias CBX8 binding.

In Drosophila, the dPc CD interacts with H3K27me3 and has been shown to target the dPRC1 complex to facultative heterochromatin. In vertebrates, the five CBX paralogs bind H3K27me3 with drastically reduced affinity and specificity than the dPc CD, suggesting that this classical targeting mechanism may be more complex in vertebrates than in Drosophila.[9,103] Additionally, our work highlights that these *in vitro* affinities of the CBX paralogs for H3K27me3 do not predict relative affinity for bulk chromatin *in vivo*, indicating that interactions outside of histone methylation drive chromatin association. We have identified that binding of the CBX8 CD to DNA is important for bulk chromatin association, which is likely also true for the other four CBX paralogs, as R20 and R22 are conserved across all five CBX paralogs and R19 and R60 are either R or K across all paralogs. In fact, the CBX2 CD has been shown to interact with DNA, although the structural basis for interaction and its role in chromatin association has not been assessed.[258] Additional studies are necessary to fully understand the relative importance for DNA binding on CBX paralog association with chromatin.

CBX paralog expression is misregulated in countless cancers, and a myriad of other studies propose oncogenic mechanisms involving the CBX paralogs for numerous cancers. Further, a systematic structural analysis of methyl-lysine readers predicted the CBX chromodomains to be "druggable."[216] As such, drug development initiatives have focused on developing paralog-specific CBX chromodomains inhibitors. To date, there has only been success in identifying CBX7 paralog-specific chromodomain inhibitors;[155,185,217] which may be in part due to our limited knowledge of CBX binding mechanisms. Finally, these studies pave the way for the development of CBX8-specific chromodomain inhibitors, which will be useful tools to assess paralog-specific function and define the therapeutic utility of CBX8 inhibitors. While our work suggests that disruption of methyl binding may be sufficient to disrupt *in vivo* CBX8 binding, it also suggests that inhibitors that also disrupt DNA binding would be more potent.

CHAPTER 5. FUTURE DIRECTIONS

5.1 Paralog roles in transcriptional regulation

The Polycomb Group proteins are known for their role in repressing gene transcription via H3K27me3, H2AK119 monoubiquitination, and chromatin compaction. In Drosophila, there is one possible canonical PRC1 complex to regulate gene transcription. However, as highlighted in Chapter 1, gene duplication has resulted in over 180 possible canonical PRC1 complexes that can regulate gene transcription.[230] Little is known about which complex compositions are present within cells and how the different paralogs and complex compositions contribute to transcriptional repression.

The targeting subunit of cPRC1, CBX, has five different paralogs, however, the accepted targeting mechanism is the same. Furthermore, the identification of cell-type specific complexes and paralog-specific phenotypes has raised the question as to paralog gene targets.[52,54] While genome-wide localization studies have identified paralog-specific gene targets, the majority of CBX paralog gene targets overlap. For example, Peters and colleagues profiled CBX6, 7 and 8 in two different cell lines, neonatal foreskin fibroblasts (Hs68) and breast fibroblasts (BF) and observe a large (>60%) overlap in gene targets (Figure 5.1a,b).[63] Further, Ram et al., also identified a similar trend with CBX2 and CBX8 in the chronic myelogenous leukemia cell line K562,[50] while studies in embryoid bodies and hematopoietic stem cells demonstrate high target overlap between CBX2/4 and CBX7/8 respectively.[52,58] This raises the question as to how do these paralogs function to repress gene transcription. The high target overlap suggests that the CBX paralogs may function redundantly and/or interdependently.

Interdependency of the PRC1 paralogs has been exhibited at the *INK4A/ARF* locus. ChIP studies revealed that both CBX7 and CBX8 simultaneously bind at this locus, and when one paralog is knocked down localization of the other paralog is also reduced.[66] While this has been demonstrated for one locus it has not been studied throughout the genome at overlapping gene targets. Additionally, it is possible that the CBX paralogs co-localize to the same region but exhibit different degrees of gene repression. This is supported by biochemical studies which have demonstrated varying compaction

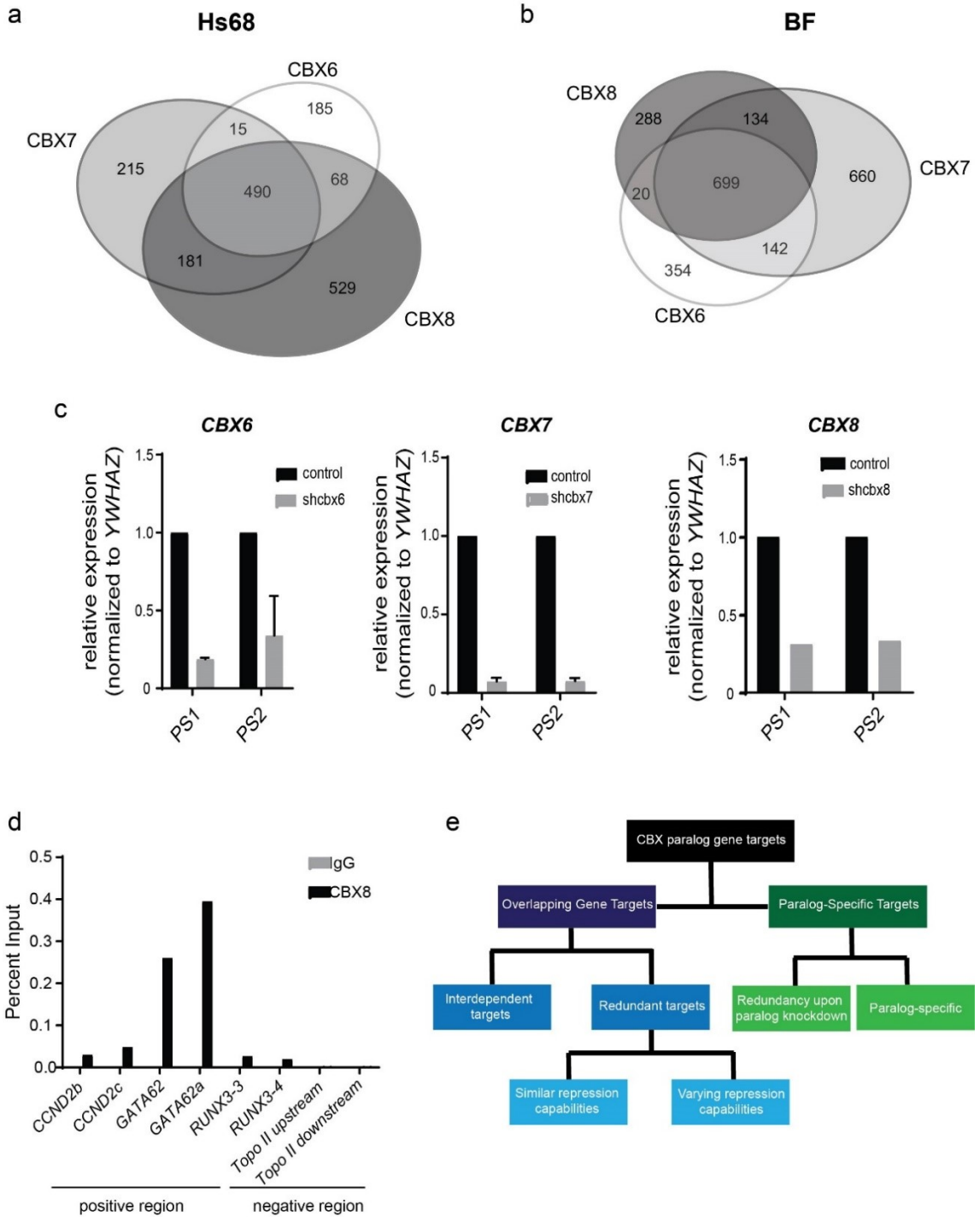


Figure 5.1 Paralog targeting

a,b) Overlap analysis of the CBX paralog target genes in Hs68 cells and BF cells, respectively. c) qPCR of CBX paralog knockdown in Hs68 fibroblasts for 7 days. d) CBX8 ChIP-qPCR at sites identified in Pemberton et al. e) Schematic of probable CBX paralog gene targeting mechanisms.

capabilities and H3K27me3 binding properties, two critical components in PRC1 repression.[12,102,103]

Pemberton et al., laid the groundwork and a system to understand paralog transcriptional regulation. Their genome-wide studies examining CBX6, CBX7, and CBX8 localization is one of few studies assessing multiple paralog genome-wide binding. To understand how each paralog contributes to gene regulation, we will knockdown CBX6, CBX7, and CBX8 paralogs individually with small hairpin RNAs in the Hs68 human neonatal foreskin fibroblast cell line (Figure 5.1c). Following knockdown, we will perform RNA sequencing to evaluate differential gene expression between control and knockdown cells. If the shared gene targets between the CBX paralogs are similarly differentially expressed with each knockdown, it is likely the paralogs work interdependently for repression. It is also possible that we may not observe gene expression changes at the previously identified CBX targets, suggesting functional redundancy. Further, it is possible that we may observe that one paralog contributes more to repression than the others. In addition to overlapping gene targets, our RNA-seq will provide insight into the types of genes that each paralog may regulate individually.

In addition to assessing paralog transcriptional changes, we can also assess if the paralogs are interdependent by assessing localization. The genome-wide co-localization of CBX paralogs from Pemberton et al. will provide the basis for asking about interdependency. We have successfully validated peaks identified in Pemberton et al., using ChIP-qPCR in the Hs68 fibroblasts (Figure 5.1d). ChIP-reChIP at shared gene targets will demonstrate if the paralogs are simultaneously present at the loci. Subsequently, ChIP-qPCR in the individual paralog knockdown lines will shed insight into how loss of one paralog affects additional CBX localization. If the paralogs are interdependent, we would expect to see a decrease in localization upon knockdown of a paralog. It is possible that the paralogs are not interdependent and that knockdown of one paralog will not affect the localization of others. Additionally, we can assess localization of paralogs at “paralog-specific” targets upon knockdown of that paralog. If there is redundancy between paralogs, the remaining paralogs may replace the lost paralog (outlined in Figure 5.1e).

5.2 Paralog binding properties

Though the CBX paralog chromodomains are highly homologous (Figure 5.2a), they exhibit varying *in vitro* histone peptide binding properties with only the CBX2 chromodomain demonstrating specificity for H3K27me3 (Figure 5.2b).[103] Despite this variance in *in vitro* binding between chromodomains, genome-wide studies demonstrate strong correlation of all the paralogs with H3K27me3.[50,52,63] Additionally, our *in vivo* bulk chromatin binding analysis demonstrates similar chromatin binding for the paralogs (See Figure 4.1d). As outlined in Chapter 4, these studies are not encompassing of the chromatin environment, and there are often additional mechanisms contributing to chromatin binding. For the CBX8 chromodomain, we have characterized a previously unidentified DNA binding mechanism that contributes to chromatin association (Chapter 4). This binding mechanism is mediated by an arginine-rich patch (R19, R20, R22, and R60). This basic patch is not unique to the CBX8 chromodomain; R20 and R22 are conserved across all five paralogs, while R19 and R60 are either lysine or arginine across the paralogs (Figure 5.2a). Thus, we hypothesize that all five CBX paralogs exhibit DNA binding, likely to varying degrees, for chromatin association.

To assess the DNA binding properties of the CBX paralogs, we propose electrophoretic mobility shift assays with linear 601 DNA, as described in Chapter 4. Using similar protein concentrations and experimental conditions, we will be able to qualitatively rank the DNA binding properties of the CBX chromodomains. Based on chromodomain sequences, we hypothesize CBX6/8 will demonstrate the strongest DNA binding, while CBX2/7 will demonstrate the weakest binding. A preliminary experiment with the CBX8 and CBX2 CDs supports this hypothesis. We observe that both CDs engage DNA, however, the CBX8 CD binds at lower protein concentrations, suggesting the CBX8 CD has more robust DNA binding than CBX2 (Figure 5.1c).

Additionally, we generated a baculovirus system to recombinantly express the entire PRC1 complex. When purifying PRC1 from mammalian cells, it is likely that the purification will result in heterogenous complexes as most of the paralogs are expressed in different cell types. The baculovirus system allows us to generate a homogeneous PRC1 complex with our subunits of interest. This system has been used previously to examine

a

b

Chromodomain	H3K9me3 (μM)	H3K27me3 (μM)
dPc	125	5
CBX2	>500	330
CBX4	70	205
CBX6	>500	>500
CBX7	55	110
CBX8	>500	>500

c

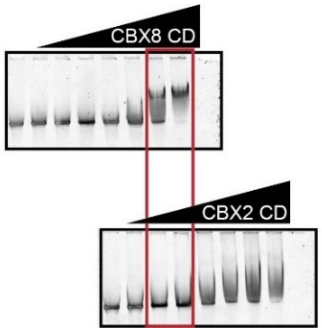


Figure 5.2 Paralog DNA binding

a) Sequence alignment of the human CBX chromodomains. Amino acids involved in CBX8 DNA binding are highlighted in blue; those shared in the other paralogs are in blue as well, and those with similarity are highlighted in green. The aromatic cage is depicted in red font. b) Binding affinities for the human CBX chromodomains for H3K9me3 or H3K27me3 peptides (table adapted from 103). c) EMSAs of the CBX8 CD or CBX2 CD with linear 601 DNA. Boxed lanes indicate similar protein concentration between the two experiments.

biochemical functions of PRC1 subunits.[8,10] Briefly, SF21 insect cells are transfected with the individual subunits to generate P1 viruses released into the media (Figure 5.3a). Because the P1 virus titer is often low, the SF21 cells are subsequently infected with P1 virus to generate a P2 virus (media) (Figure 5.3a). The P2 subunit viruses are then combined to generate a P3 expressing the complex (Figure 5.3a). The P3 cells are harvested, lysed, and then the protein is purified using affinity purification techniques for downstream applications (Figure 5.3a). We have generated constructs to form PRC1 complexes containing FLAG-BMI-1, RING1, PHC1, and His-CBX2/4/6/7/8. To date, we have focused on isolating a CBX8-containing PRC1 complex (Figure 5.3b) using a double affinity tag purification, 6x-His followed by FLAG. While we have had some success in isolating the CBX8-containing PRC1 complex, we have been limited by overall purity as well as quantity. Further optimization for purification scale, purity, and quantitation are needed.

The recombinant PRC1 complex can be used to assess complex binding in an *in vitro* setting. We performed preliminary binding studies with our CBX8-containing PRC1 complex and mononucleosomes. With recombinant PRC1 complex and unmodified nucleosomes, we observe a shifted band upon increasing concentrations of PRC1 complex in an EMSA (Figure 5.3c, top). Interestingly compared to the EMSAs using the CBX8 CD, we observe a less smeared band upon shifting. Subsequently, the shifted bands become even tighter and more defined upon the introduction of the methyl-lysine analog $H3K_c27me3$ (Figure 5.3c, bottom). We hypothesize that the reduction in smearing is due to the size of the PRC1 complex (~200 kDa) relative to the chromodomain (~5 kDa) reducing the number of ways binding occurs. While our *in vivo* sequential salt extractions are insightful for understanding bulk chromatin binding and our CD EMSAs highlight the role of the chromodomain in binding, using recombinant complex *in vitro* will provide insight into how these proteins work together rather than in isolation.

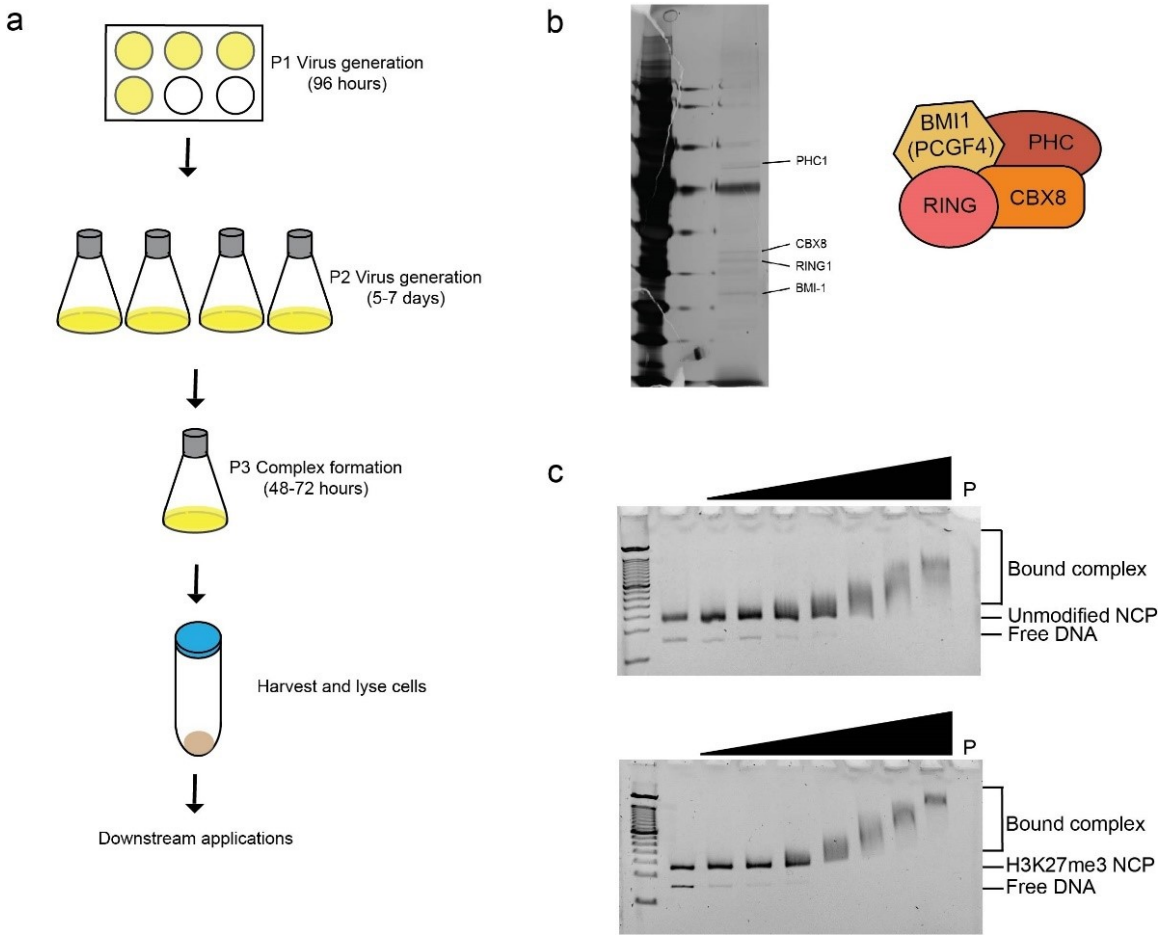


Figure 5.3 Baculovirus purification

a) Workflow of baculovirus generation for protein purification. b) Silver stain of purified recombinant PRC1. c) EMSAs using recombinant PRC1 and mononucleosomes (NCP), unmodified (top) and H3K$_c$27me3 (bottom)

5.3 CBX Self-association

The PcG proteins are involved in chromatin looping and condensation often through highly dense regions of Polycomb proteins known as Polycomb bodies.[92,93] These Polycomb bodies are meditated through self-association and oligomerization. To date, several PRC1 subunits have been shown to self-associate. The Polyhomeotic paralogs (PHC) self-oligomerize through their sterile alpha motif.[104] Additionally, a study has demonstrated that PCGF2 and PCGF4 homodimerize in solution.[106] For the CBX paralogs, however, little is known about CBX self-association. Studies from the late 1990s with the dPc chromodomain and the mouse CBX2 chromodomain suggest the chromodomain can self-associate and is supported by the dPc chromodomain crystal structure.[118,259,260] Furthermore, a recent proteomics study identified oligomerization of CBX4 isoforms,[75] however, this oligomerization has not been further studied. Thus, we hypothesized that CBX8 self-associates *in vivo*.

To determine if CBX8 self-associates *in vivo*, we generated two N-terminally tagged CBX8 constructs (V5 and FLAG) and co-transfected HEK293Ts. We performed co-immunoprecipitation with both V5 and FLAG antibodies. Immunoprecipitation of V5-CBX8 revealed an association with FLAG-CBX8 as seen by immunoblot (Figure 5.4 a). The reciprocal association was also observed (Figure 5.4a). Co-immunoprecipitation, however, was not observed when ARID-V5 and FLAG-CBX8 were co-transfected suggesting this is not an artifact of the epitope tags (Figure 5.4a). The PRC1 complexes interact with chromatin and RNA, so we sought to assess if CBX8 self-association was mediated through DNA or RNA interactions. To do so, we treated our immunoprecipitations with RNase or DNase. Immunoblot of the co-IPs demonstrated reciprocal immunoprecipitation of CBX8 regardless of treatment suggesting that this self-association is not mediated through nucleic acid interactions (Figure 5.4b). Additionally, CBX proteins are known to interact with post-translationally modified histones, which could also mediate CBX8 oligomerization. Using the K31A,W32A chromodomain mutation,[64] we assessed the role of histone binding in CBX8 self-association, however, these studies were inconclusive (data not shown).

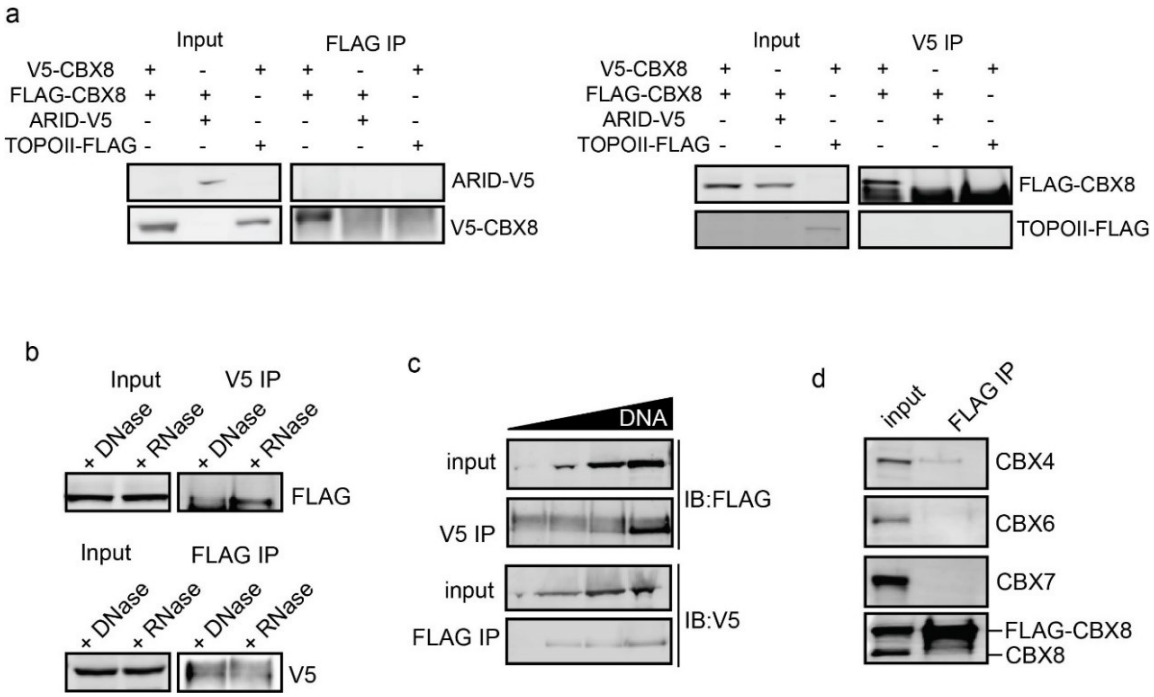


Figure 5.4 CBX8 self-association

a) FLAG (left) and V5 (right) co-IPs from co-transfected HEK293T cells as indicated above the immunoblots, immunoblots blotted for the V5 constructs (left) or FLAG constructs (right). b) co-IPs from co-transfected cells treated with either RNase or DNase, immunoblots stained for the reciprocal construct. c) co-IPs from HEK293T cells transfected with varying amounts of DNA, immunoblots stained for reciprocal tag. d) FLAG co-IP from FLAG-CBX8 transfected HEK293T cells examining endogenous CBX paralog association.

Self-association can sometimes be a result of a high concentration of protein, often a caveat of an overexpression system. To examine the effect of protein concentration on CBX8 self-association, we transfected HEK293Ts with varying amounts of DNA (3-12 μg). Altering the amount of DNA correlated to a similar change in protein expression (Figure 5.4c). We subsequently performed the FLAG and V5 co-IPs as done previously and observe self-association with low amounts of transfected DNA (6 μg) (Figure 5.4c). This suggests that CBX8 self-association is not an artifact of massive overexpression. Additionally, we expressed only FLAG-CBX8 to identify association with endogenous CBX proteins. Interestingly, FLAG-CBX8 co-immunoprecipitated endogenous CBX8 and CBX4 but not CBX6 or CBX7 (Figure 5.4d).

Next, we sought to identify the portions of CBX8 necessary for self-association. To do so, we generated various CBX8 truncations (Figure 5.5a). We deleted the chromodomain (ΔCD CBX8) and assessed its ability to self-associate with WT CBX8. Co-immunoprecipitation revealed that the chromodomain is not necessary for CBX8 self-association as it still co-immunoprecipitated endogenous and FLAG-CBX8 (Figure 5.5b). It is possible, however, that self-association of other PRC1 subunits are driving the observed self-association regardless of the chromodomain's contribution. Thus, we used the shortest CBX8 truncation that expressed, $CBX8^{1-91}$, to assess the minimal portion of CBX8 needed for self-association (Figure 5.5c). We co-transfected HEK293T cells with either FLAG-WT/V5-WT, V5-WT/FLAG-$CBX8^{1-91}$, V5 $CBX8^{1-91}$/FLAG-WT, and V5-$CBX8^{1-91}$/FLAG-$CBX8^{1-91}$ and performed V5 and FLAG immunoprecipitations (Figure 5.5d). Interestingly, $CBX8^{1-91}$ associates with WT CBX8 and itself (Figure 5.5d). $CBX8^{1-91}$ does not contain the Pc Box, which is required for CBX8 incorporation into the PRC1 complex,[107] suggesting that the other PRC1 subunits that have previously been shown to oligomerize are not driving this interaction.

To assess if the CBX8 chromodomain (amino acids 8-61) can self-associate, our collaborators at the University of Iowa assessed CBX8 chromodomain association in a concentration dependent manner using ^{1}H-^{15}N HSQC NMR spectroscopy. At high concentrations, chemical shift perturbations for a few amino acids were observed (data not shown). Interestingly, one of the residues that is highly perturbed was identified in the dPc chromodomain crystal structure as critical for self-association (L49).[118] We generated a

point mutant within the CBX8 chromodomain (L49D) and assessed oligomerization using co-IP. Unfortunately, the single mutation did not disrupt CBX8 self-association (data not shown). These studies however were performed in the context of the complex, and it is likely that single point mutation (L49D) within the chromodomain is not sufficient to disrupt self-association if additional PRC1 subunits are involved in self-association. To better assess the role of the chromodomain in self-association, this mutation would need to be generated in our $CBX8^{1-91}$ or the CBX8 chromodomain construct where CBX8 is not incorporated into the complex.

To assess the strength of CBX8's association, urea denaturation followed by co-immunoprecipitation can be used. This assay will allow us to confirm that the self-association is not an artifact of overexpression. It will also allow us to assess different contributors to the strength of the self-association, i.e., full complex compared to the chromodomain. Further, the PRC1 baculovirus system described earlier can be used to understand CBX8 self-association *in vitro*. Using purified PRC1 complex, we can use analytical ultracentrifugation to determine the oligomerization state of PRC1 and assess how different subunits, mutations, etc. alter this oligomerization. Further, cryo-EM can be used to examine PRC1 oligomerization with and without a nucleosomal template. Understanding CBX8 and PRC1 self-association will improve our overall understanding of PRC1 function.

5.4 Conclusions

Canonical PRC1 is essential for maintaining appropriate gene expression through chromatin compaction and histone ubiquitination.[230] Studies focusing on general PRC1 function have found that PRC1 represses lineage-specific, developmental regulators, and cell cycle regulatory genes making them critical for proper development, stem cell maintenance, and unsurprisingly, the development of cancer.[37] However, the functional contributions of PRC1 diversity in mammals is largely unknown. This dissertation highlights new potential therapeutic approaches for GBM among other cancers, describes a new binding mechanism for the CBX8 chromodomains, and lays the ground work for understanding CBX8's function *in vivo*.

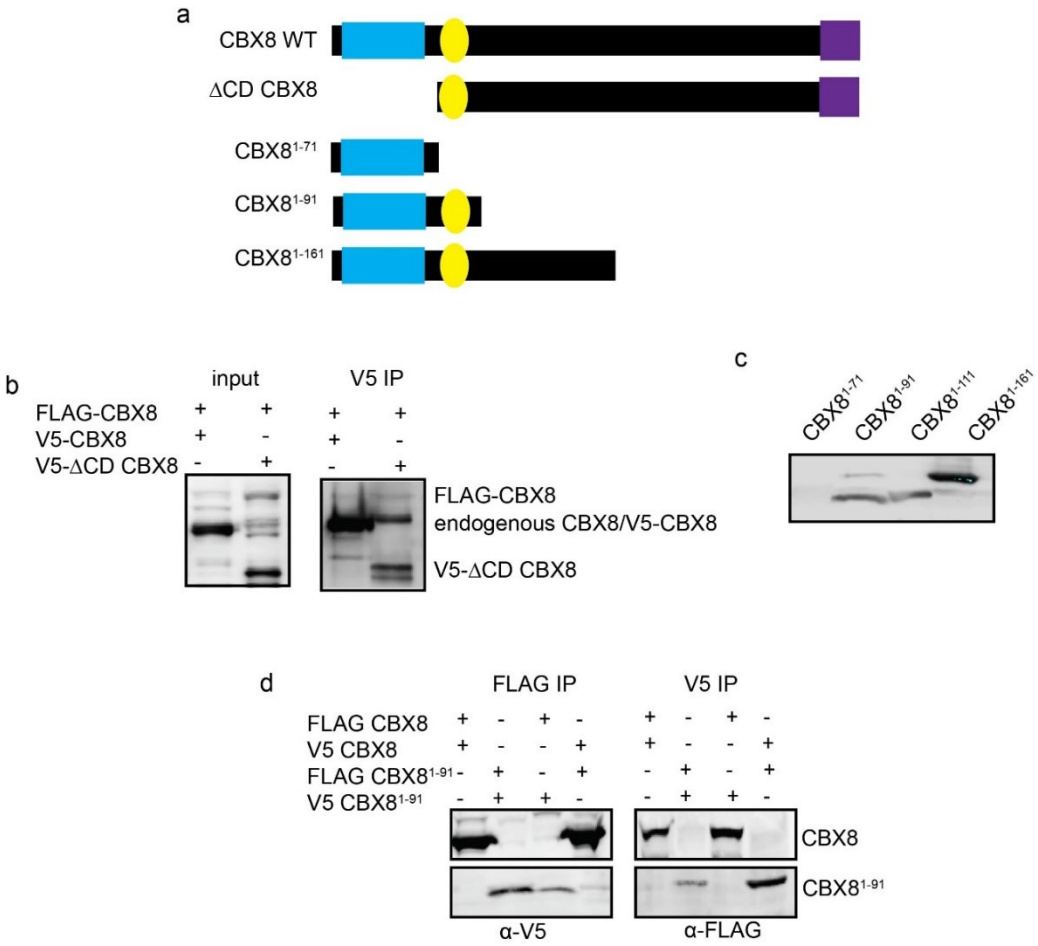


Figure 5.5 CBX8 chromodomain is necessary for self-association

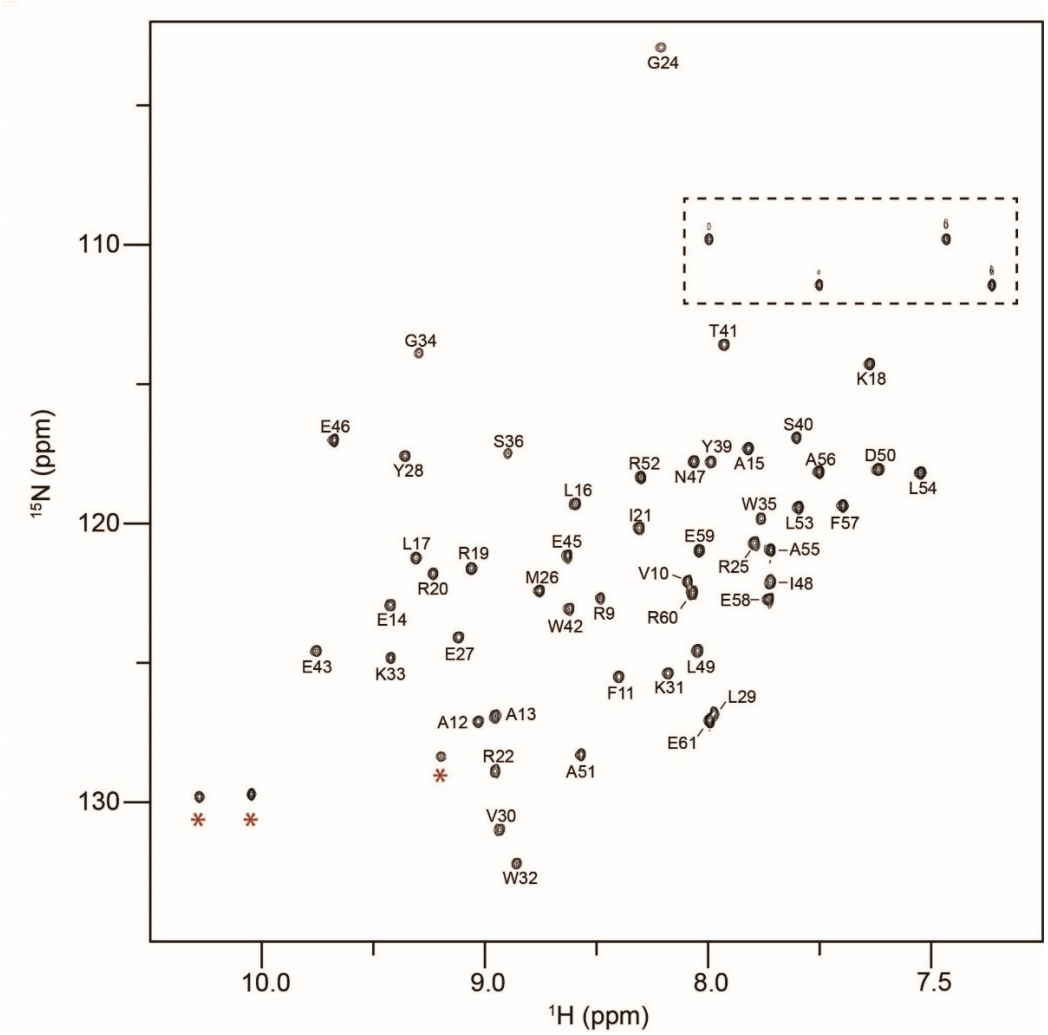
a) Depiction of CBX8 truncation constructs generated. b) V5 IP of WT CBX8 or ΔCD CBX8 co-transfected with FLAG-CBX8, immunoblot stained with anti-CBX8. c) Expression immunoblot of the CBX8 truncations, anti-V5. d) FLAG and V5 co-IPs with $CBX8^{1-91}$ and WT.

The work focused on GBM is the framework to better understand CBX paralogs in cancer and as a therapeutic target. In chapter two, we identified that CBX7 chromodomain inhibition enhances GBM cellular response to the chemotherapeutic doxorubicin. While previous studies have demonstrated inhibition of the EZH2 methyltransferase enhances radiation response,[169] we are the first to demonstrate that inhibition of a CBX paralog demonstrates a similar effect. Inhibition of a single CBX paralog will ensure that not all PRC1 function is lost and likely minimize toxicities. Further the combinatorial treatment strategy exhibits similar response using lower doxorubicin concentrations as compared to doxorubicin alone. Though these studies were performed with doxorubicin, which is not used in the treatment of glioblastoma multiforme, it is possible that further studies with varying doses of temozolomide or radiation may exhibit a similar enhancement. Further, inhibition of CBX7 with chemotherapy treatment may be applied to other cancers. In chapter three, we shifted our focus to the CBX8 paralog in glioblastoma as it is overexpressed in over half of patient samples and cell lines.[166] We found that CBX8 and its chromodomain are necessary for GBM viability suggesting it may serve as a therapeutic target. While previous studies have demonstrated an oncogenic role for CBX8 in a variety of cancers, the role of the CBX8 chromodomain has not been widely explored. Preliminary transcriptional analysis suggests CBX8 is involved in the regulation of cell adhesion, migration, and development genes. In addition, this work suggests that CBX7 and CBX8 demonstrate paralog specific roles in the context of GBM. This dissertation works outlines potential roles for both CBX7 and CBX8 in glioblastoma though further work is necessary.

In addition to understanding CBX paralog function in oncogenesis, this work focused on better understanding CBX8 biochemical functions. In chapter 4, we, with our collaborators, identified a new binding mode for the CBX8 chromodomain. The mammalian CBX chromodomains are thought to recognize and bind H3K27me3,[40] however, our work demonstrates that the CBX8 chromodomain engages both DNA and H3K27me3 for full chromatin association. This has large implications for our understanding of PRC1 targeting. Analysis of genome wide data revealed a high number of H3K27me3 peaks that did not correspond to CBX8 binding. Our findings suggest CBX8 binding is selective for regions with both H3K27me3 and accessible DNA. Additional studies are needed to understand the biological implications of this binding mechanism,

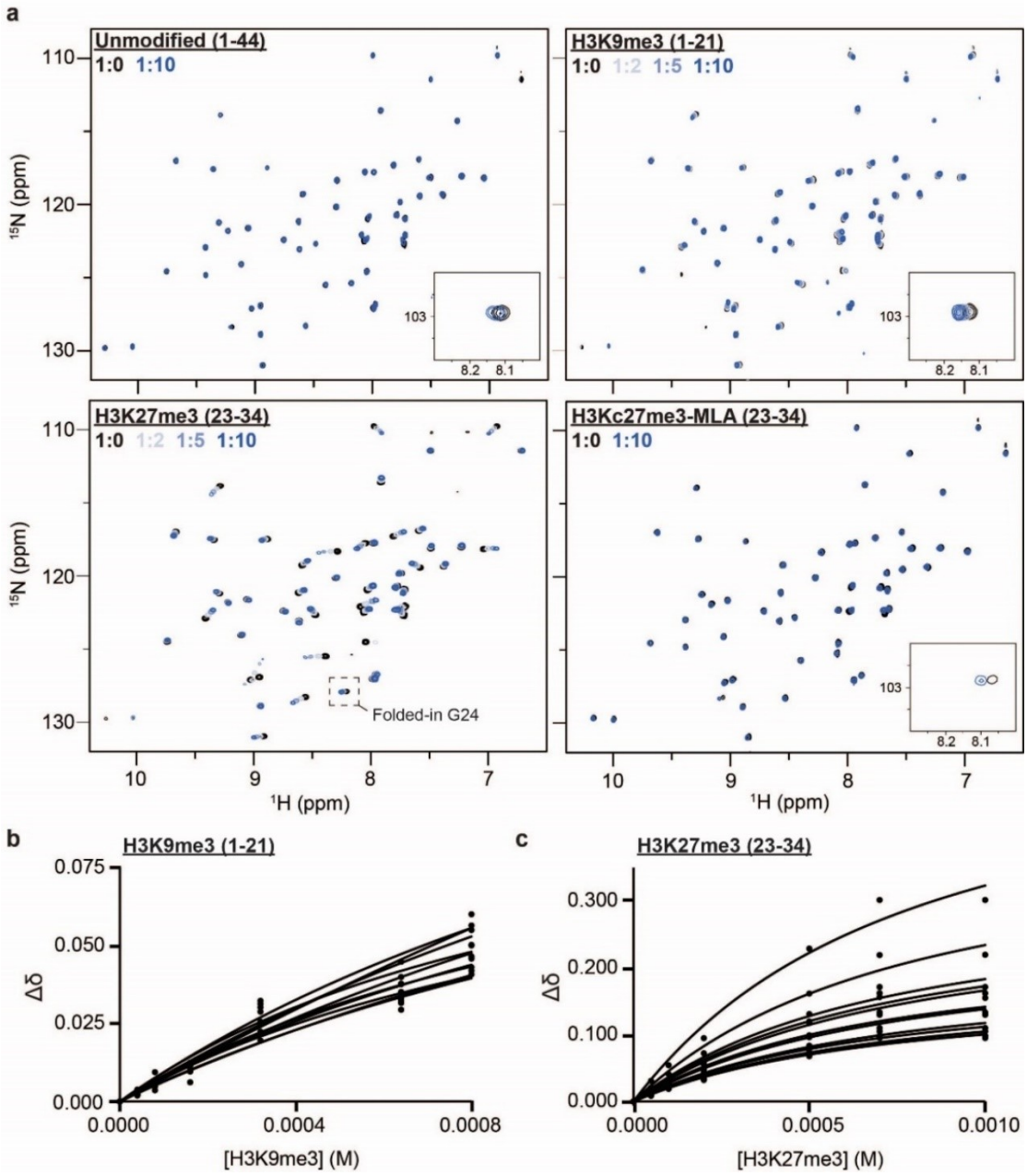
but this work begins to shift the CBX chromodomain targeting paradigm. Further this emphasizes the importance of studying chromodomains in the context of the chromatin environment. Additionally, these findings have implications on drug screening and development. Screening and validating inhibitors for only trimethyl-lysine binding may not demonstrate *in vivo* efficacy. Additionally, targeting CBX chromodomain DNA binding may also open avenues for drug development. Finally, the preliminary work in chapter five demonstrates CBX8 self-association. While additional work is necessary, CBX8 self-association may have implications in CBX8 function. The work in chapter five serves as a basis for future work and establishes the tools to further understand CBX8 self-association and its implications. This dissertation work adds to our understanding of PRC1 and CBX diversity and lays a foundation for future and ongoing studies.

APPENDIX A. SUPPLEMENTAL FIGURES



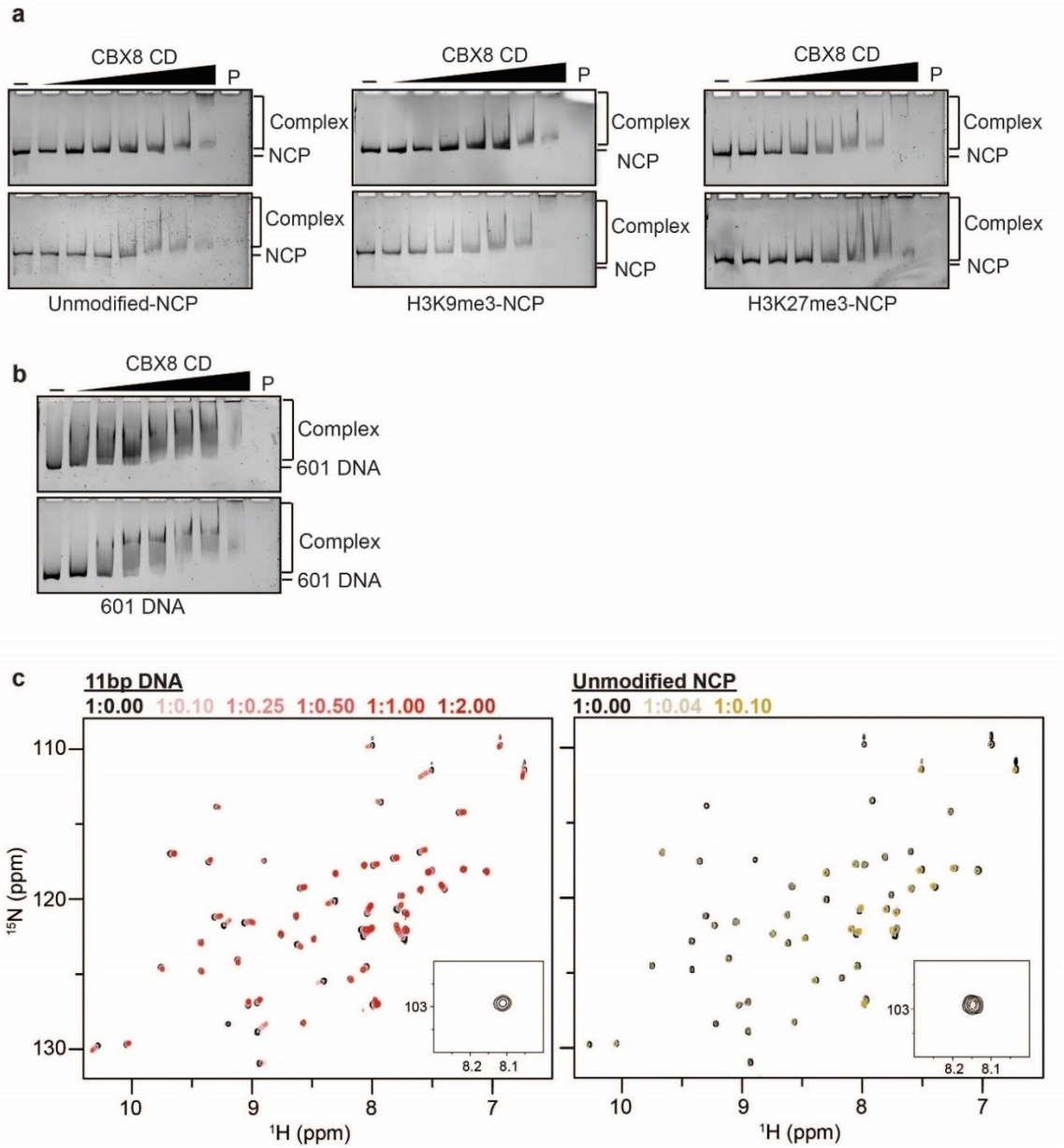
Appendix A Figure 1

a) ^{1}H-^{15}N-HSQC spectrum of the CD with assigned resonances labeled. Side chain NH peaks are enclosed in a dotted black square and tryptophan side chain resonances with a red star.



Appendix A Figure 2

a) Full ^{1}H-^{15}N-HSQC overlays for ^{15}N-CD upon addition of increasing concentrations of unmodified H3 (1-44, top left), H3K9me3 (1-21, top right), H3K27me3 (23-34, bottom left) or H3Kc27me3 (23-34, bottom right) histone peptides. Glycine 24 is shown as an inset for clarity. b) Binding curves for all resonances significantly perturbed in the H3K9me3 titration. c) Binding curves for all resonances significantly perturbed in the H3K27me3 titration.

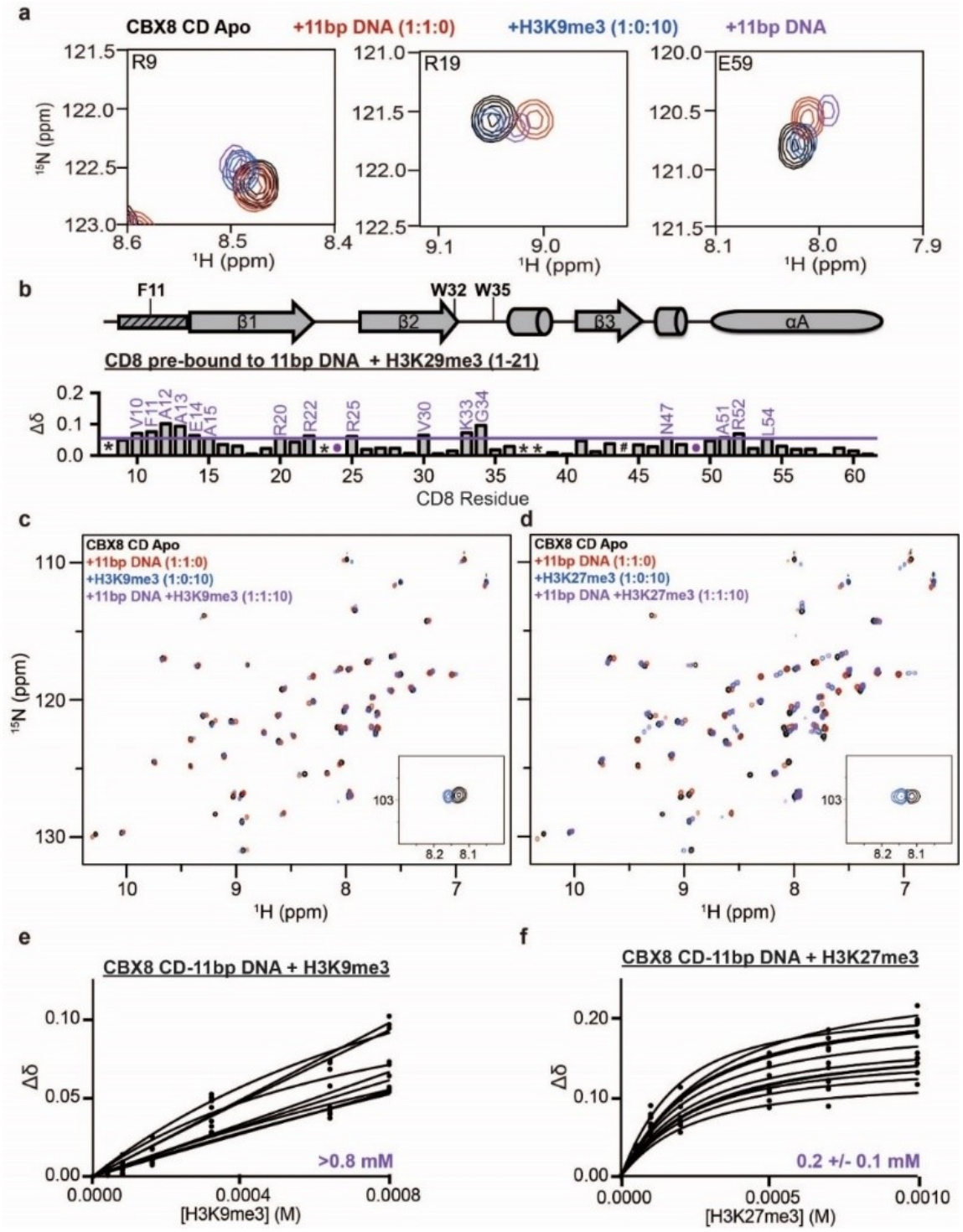


Appendix A Figure 3

a) EMSAs performed in triplicate with CD and unmodified (left), H3K9me3 (middle) and H3K27me3 (right) NCPs. b) EMSAs performed in triplicate with CD and the 147bp 601 DNA. c) Full 1H-15N-HSQC overlays for 15N-CD upon addition of increasing concentrations of the 11bp DNA (left) or unmodified NCP (right). Glycine 24 is shown as an inset for clarity.

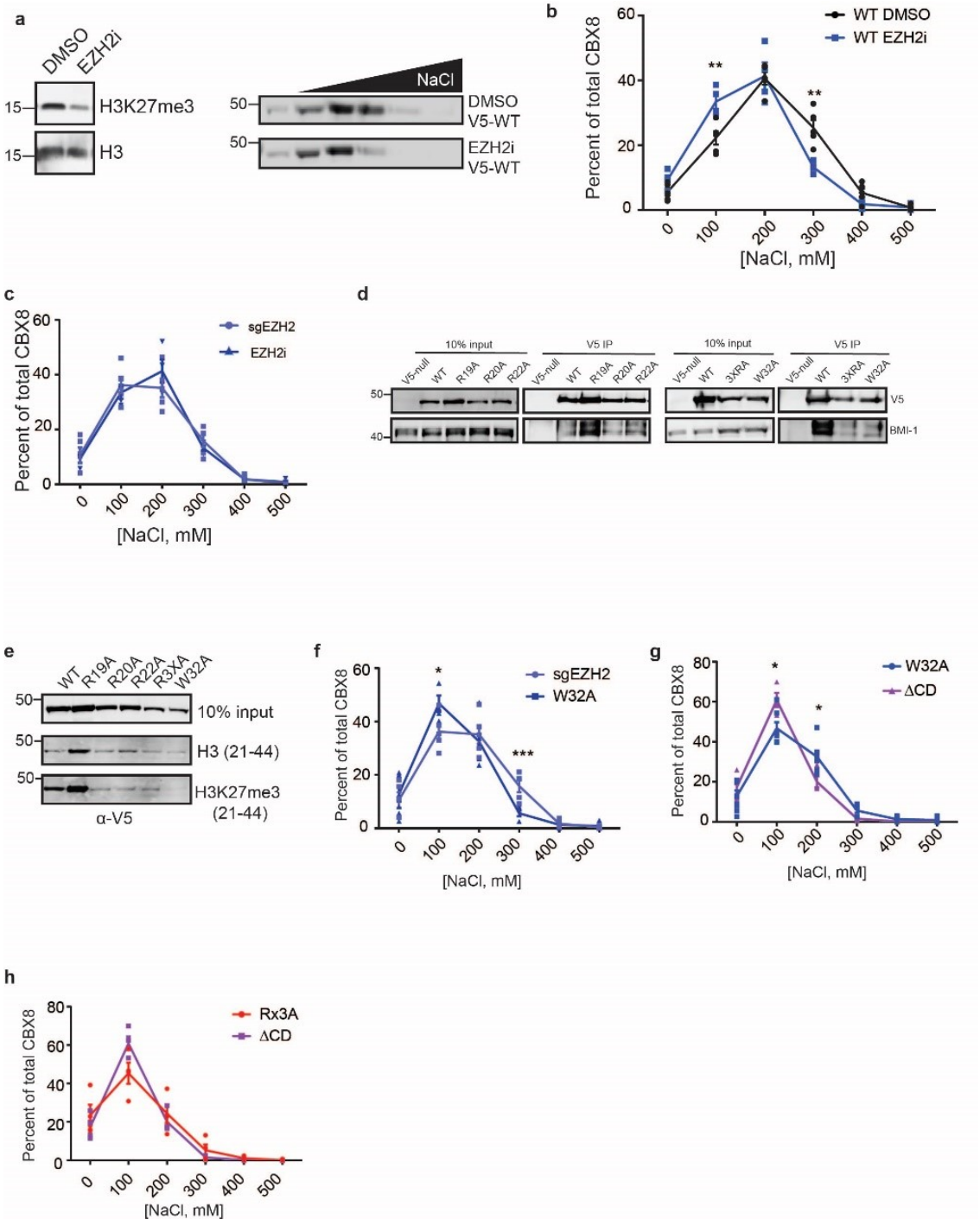
Appendix A Figure 4

a) ^{1}H-^{15}N-HSQC overlays for ^{15}N-CD in the apo (black, 1:0:0 ratio), bound to an 11bp DNA (red, 1:1:0 ratio), bound to H3K9me3 (blue, 1:0:10), or bound to both 11bp DNA and H3K9me3 (purple, 1:1:10 ratio). Shown are resonances for selected residues in the histone binding pocket (R9, left), DNA binding pocket (R19, middle) and a residue sensitive to both DNA and histone binding (E59, right) are shown. b) normalized CSP ($\Delta\delta$) between the DNA-bound (1:1:0) and DNA and H3K9me3 bound (1:1:10 ratio) plotted against CBX8 residue number. CSPs were considered significant if greater than the mean plus one standard deviation, and are labeled in purple. The secondary structure of CD from the crystal structure PDBID 3I91 is diagramed above the $\Delta\delta$ plot with the aromatic cage residues labeled. c) Full ^{1}H-^{15}N-HSQC overlays for ^{15}N-CD in the apo (black, 1:0:0 ratio), bound to an 11bp DNA (red, 1:1:0 ratio), bound to H3K9me3 (blue, 1:0:10), or bound to both 11bp DNA and H3K9me3 (purple, 1:1:10 ratio). to H3K27me3 (blue, 1:0:10), or bound to both 11bp DNA and H3K27me3 (purple, 1:1:10 ratio). Glycine 24 is shown as an inset for clarity. d) Full ^{1}H-^{15}N-HSQC overlays for ^{15}N-CD in the apo (black, 1:0:0 ratio), bound to an 11bp DNA (red, 1:1:0 ratio), bound to H3K27me3 (blue, 1:0:10), or bound to both 11bp DNA and H3K27me3 (purple, 1:1:10 ratio). Glycine 24 is shown as an inset for clarity. e) Binding curves for all resonances significantly perturbed in the H3K9me3 titration with the CD pre bound to 11bp DNA. f) Binding curves for all resonances significantly perturbed in the H3K27me3 titration with the CD pre bound to 11bp DNA.


a
CBX8 CD Apo
+11bp DNA (1:1:0)
+H3K9me3 (1:0:10)
+11bp DNA
R9
R19
E59
15N (ppm)
1H (ppm)
b
F11
W32 W35
β1
β2
β3
αA
CD8 pre-bound to 11bp DNA + H3K29me3 (1-21)
Δδ
CD8 Residue
c
CBX8 CD Apo
+11bp DNA (1:1:0)
+H3K9me3 (1:0:10)
+11bp DNA +H3K9me3 (1:1:10)
d
CBX8 CD Apo
+11bp DNA (1:1:0)
+H3K27me3 (1:0:10)
+11bp DNA +H3K27me3 (1:1:10)
e
CBX8 CD-11bp DNA + H3K9me3
>0.8 mM
[H3K9me3] (M)
f
CBX8 CD-11bp DNA + H3K27me3
0.2 +/- 0.1 mM
[H3K27me3] (M)

Appendix A Figure 5

a) Immunoblot of H3K27me3 levels following 48 h EZH2i treatment (left), immunoblot of a representative SSE of V5-WT ± EZH2i (anti-V5) (right). b) Quantitation of the SSEs from a) to compare CBX8 elution between EZH2i (blue) and DMSO (black) treated cells, n=4, 5 biological replicates, respectively. c) Quantitative comparison of EZH2i SSE (dark blue) and sgEZH2 SSE (light blue) shown previously, n= 4 ,5 biological replicates respectively. d) Co-immunoprecipitation of V5 mutants using anti-V5. Immunoblots were stained for V5 and the PRC1 subunit BMI-1. e) Peptide pulldowns of all mutants with unmodified H3 and H3K27me3 (anti-V5). f) Quantitative comparison of sgEZH2 (light blue) and W32A SSE (dark blue) quantitation n=5, 4 biological replicates, respectively. g) Quantitative comparison of W32A (blue) and ΔCD SSE (purple) previously shown, n=7, 4 biological replicates respectively. h) Quantitative comparison of 3xRA (red) and ΔCD SSEs (purple) shown previously, n=4 biological replicates for both. For all quantitations, the amount of CBX8 in each fraction is reported as a percentage of total CBX8. Two-tailed student's t-tests were used to determine significance; * $p<0.05$, ** $p< 0.01$, *** $p< 0.001$, **** $p< 0.0001$


a
DMSO
EZH2i
15
H3K27me3
15
H3
NaCl
50
DMSO
V5-WT
50
EZH2i
V5-WT
b
WT DMSO
WT EZH2i
Percent of total CBX8
[NaCl, mM]
c
sgEZH2
EZH2i
Percent of total CBX8
[NaCl, mM]
d
10% input
V5 IP
10% input
V5 IP
V5-null
WT
R19A
R20A
R22A
3XRA
W32A
50
40
V5
BMI-1
e
WT
R19A
R20A
R22A
R3XA
W32A
50
10% input
50
H3 (21-44)
50
H3K27me3
(21-44)
α-V5
f
sgEZH2
W32A
Percent of total CBX8
[NaCl, mM]
g
W32A
ΔCD
Percent of total CBX8
[NaCl, mM]
h
Rx3A
ΔCD
Percent of total CBX8
[NaCl, mM]

APPENDIX B. CELL LINES

Table B1: Cell lines generated and used throughout the dissertation

Cell line	**Description**
A172	glioblastoma cell line
A172 plko	Stable pLKO knockdown control A172 GBM line
A172 shCBX8-1	Stable CBX8 knockdown A172 GBM line, shCBX8-3 construct
A172 shCBX8-2	Stable CBX8 knockdown A172 GBM line, shCBX8-4 construct
U87MG	glioblastoma cell line
U87MG plko	Stable pLKO knockdown control U87MG GBM line
U87MG shCBX8-1	Stable CBX8 knockdown U87MG GBM line, shCBX8-3 construct
U87MG shCBX8-2	Stable CBX8 knockdown U87MG GBM line, shCBX8-4 construct
U118MG	glioblastoma cell line
U118MG plko	Stable pLKO knockdown control U118MG GBM line
U118MG shCBX8-1	Stable CBX8 knockdown U118MG GBM line, shCBX8-3 construct
U118MG shCBX8-2	Stable CBX8 knockdown U118MG GBM line, shCBX8-4 construct
T98G	glioblastoma cell line
T98G plko	Stable pLKO knockdown control T98G GBM line
T98G shCBX8-1	Stable CBX8 knockdown T98G GBM line, shCBX8-3 construct
T98G shCBX8-2	Stable CBX8 knockdown T98G GBM line, shCBX8-4 construct
T98G plko blast	Stable pLKO knockdown control, blasticidin selection T98G GBM line
T98G shCBX8-2 blast	Stable CBX8 knockdown, blasticidin selection T98G GBM line
T98G shCBX8-2 blast + WT CBX8	Stable CBX8 knockdown, blasticidin selection, FUW WT CBX8 re-expressed T98G GBM line
T98G shCBX8-2 blast + mutCBX8	Stable CBX8 knockdown, blasticidin selection, FUW mutCBX8 (K31A,W32A) re-expressed T98G GBM line
U138MG	glioblastoma cell line

Table B1 continued

Cell line	Description
U138MG plko	Stable pLKO knockdown control U138MG GBM line
U138MG shCBX8-1	Stable CBX8 knockdown U138MG GBM line, shCBX8-3 construct
U138MG shCBX8-2	Stable CBX8 knockdown U138MG GBM line, shCBX8-4 construct
SVG12	transformed astrocyte cell line
SVG12 plko	Stable pLKO knockdown control, normal astrocyte line
SVG12 shCBX8-2	Stable CBX8 knockdown in normal astrocytes, shCBX8-4 construct
HEK293Ts	
T98G pxsgcontrol	transient CRISPR-CAS9 control
T98G pxsgCBX8	transient CRISPR-CAS9 CBX8 knockout
T98G sgcontrol	CRISPR-CAS9 control
T98G sgEZH2	CRISPR-CAS9 EZH2 knockout against SET domain
T98G Empty	CBX8 pxsgCBX8 background with doxycycline inducible empty FUW vector
T98G V5-CBX8 WT	CBX8 pxsgCBX8 background with doxycycline inducible WT CBX8 expression (FUW vector)
T98G V5-CBX8 W32A	CBX8 pxsgCBX8 background with doxycycline inducible W32A mutant CBX8 expression (FUW vector)
T98G V5-CBX8 R19A	CBX8 pxsgCBX8 background with doxycycline inducible R19A mutant CBX8 expression (FUW vector)
T98G V5-CBX8 R20A	CBX8 pxsgCBX8 background with doxycycline inducible R20A mutant CBX8 expression (FUW vector)
T98G V5-CBX8 R22A	CBX8 pxsgCBX8 background with doxycycline inducible R22A mutant CBX8 expression (FUW vector)
T98G V5-CBX8 Rx3A	CBX8 pxsgCBX8 background with doxycycline inducible R19, 20, 22A mutant CBX8 expression (FUW vector)
T98G V5-CBX8 Quad	CBX8 pxsgCBX8 background with doxycycline inducible R19, 20, 22A and W32A mutant CBX8 expression (FUW vector)
T98G CBX8 K31A,W32A	CBX8 pxsgCBX8 background with doxycycline inducible K31A,W32A mutant CBX8 expression (FUW vector), no epitope tag

APPENDIX C. CONSTRUCTS

Table C1: Constructs generated

Construct name	**Description**	**Vector**	**Selection marker (cloning)**	**Selection marker (mammalian cells)**
FUW V5 CBX8 WT	Full length WT CBX8, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 R20A	Full length CBX8 R20A mutant in CD, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 R22A	Full length CBX8 R22A mutant in CD, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 R3A (3xRA)	Full length CBX8 triple mutant R19, R20, R22 to alanine, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 W32A	Full length CBX8 W32A mutant in CD, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 K31A, W32A	Full length CBX8 dual mutant K31A, W32A, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 62-389	CBX8 CD deletion (1-61), N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8, no RNAi resistance	WT CBX8, N-terminal V5 tag, doxycycline inducible lentiviral vector, no RNAi resistance mutation	FUW	ampicillin	puromycin
FUW CBX8 chromodomain mutation, RNAi resistant	Full length CBX8 with K31A, W32A mutation, and resistant to shcbx8-4, doxycycline inducible, N-terminal V5 tag	FUW	ampicillin	puromycin

Table C1 continued

Construct name	Description	Vector	Selection marker (cloning)	Selection marker (mammalian cells)
FUW V5 CBX8 R19A	Full length CBX8 R19A mutant in CD, N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
FUW V5 CBX8 72-389	CBX8 CD deletion (1-71), N-terminal V5 tag, doxycycline inducible lentiviral vector	FUW	ampicillin	puromycin
N106 FLAG CBX8 1-71	N-terminal FLAG tag, CBX8 CD + 10 amino acids	N106	ampicillin	puromycin
N106 FLAG CBX8 1-111	N-terminal FLAG tag, CBX8 CD + 50 amino acids	N106	ampicillin	puromycin
N106 FLAG CBX8 1-151	N-terminal FLAG tag, CBX8 CD + 100 amino acids	N106	ampicillin	puromycin
N106 FLAG CBX8 1-195	N-terminal FLAG tag, N-terminal half of CBX8	N106	ampicillin	puromycin
N106 CV5 CBX8 chromodomain/C-terminus deletion	C-terminal V5 tag, CBX8 CD deletion and C-term deletion	N106	ampicillin	puromycin
N106 CBX8 chromodomain deletion	N106 CBX8 72-389, C-terminal V5 tag	N106	ampicillin	puromycin
N106 NFLAG CBX8 chromodomain	N106 CBX8 72-389, N-terminal FLAG tag	N106	ampicillin	puromycin
N106 CV5 CBX8	C-terminal V5 tag, full length CBX8	N106	ampicillin	puromycin
N106 NFLAG CBX8 chromodomain + 30 amino acids	C-terminal V5 tag	N106	ampicillin	puromycin
N106 CV5 CBX8 C-terminal deletion	C-terminal V5	N106	ampicillin	puromycin

Table C1 continued

Construct name	**Description**	**Vector**	**Selection marker (cloning)**	**Selection marker (mammalian cells)**
N106 CV5 CBX8 chromodomain + 30 amino acids	C-terminal V5 CBX8 1-91	N106	ampicillin	puromycin
N106 NFLAG CBX8 dimer mutant	N-terminal FLAG tag, full length CBX8 with L49D point mutant	N106	ampicillin	puromycin
N106 NV5 CBX7	N-terminal V5 tag, full length CBX7	N106	ampicillin	puromycin
CBX2 chromodomain	bacteria protein expression, 6xHis tagged addgene: 25158	pET28	kanamycin	
CBX4 chromodomain	bacteria protein expression, 6xHis tagged addgene: 25237	pET28	kanamycin	
CBX6 chromodomain	bacteria protein expression, 6xHis tagged addgene: 25296	pET28	kanamycin	
CBX7 chromodomain	bacteria protein expression, 6xHis tagged addgene: 25241	pET28	kanamycin	
CBX8 chromodomain pET28	bacteria protein expression, 6xHis tagged addgene: 62514	pET28	kanamycin	
pFASTBAC His CBX8	full length CBX8; insect protein expression	pFASTBAC	kanamycin	
pFASTBAC His CBX2	mouse CBX2 (M33) from addgene:1986; insect protein expression	pFASTBAC	kanamycin	
pFASTBAC His CBX6	full length CBX6; insect protein expression	pFASTBAC	kanamycin	
pFASTBAC His CBX7	full length CBX7; insect protein expression	pFASTBAC	kanamycin	

Table C1 continued

Construct name	**Description**	**Vector**	**Selection marker (cloning)**	**Selection marker (mammalian cells)**
pFASTBAC His CBX4	mouse CBX4 from pTRIPZ (M)-HT-Cbx4; insect protein expression	pFASTBAC	kanamycin	
pFASTBAC BMI-1	mouse bmi-1 addgene: 1967 N-terminal FLAG tag; insection protein expression	pFASTBAC	kanamycin	
pFASTBAC RING1	mouse ring1a addgene:1994; insect protein expression	pFASTBAC	kanamycin	
pFASTBAC PHC1	mouse phc1 addgene: 1970; insect protein expression	pFASTBAC	kanamycin	
shCBX8-3	cbx8 hairpin,TRCN0000021895, 1-F-03	pLKO	ampicillin	puromycin
shCBX8-4	cbx8 hairpin,TRCN0000021896, 1-F-04	pLKO	ampicillin	puromycin
shcbx8-blast resistant	shCBX8-4 cloned into pLKO blast resistant line	pLKO	ampicillin	blasticidin
RTTA	lentiviral co-infection for doxycycline induced expression		ampicillin	hygromycin

REFERENCES

1. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science (80-.).* **291,** (2001).
2. Dumesic, P. A. *et al.* Product binding enforces the genomic specificity of a yeast polycomb repressive complex. *Cell* **160,** 204–18 (2015).
3. Shaver, S., Casas-Mollano, J. A., Cerny, R. L. & Cerutti, H. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics* **5,** 301–12 (2010).
4. Kennison, J. A. The Polycomb and Trithorax Group Proteins of *Drosophila* : Trans-Regulators of Homeotic Gene Function. *Annu. Rev. Genet.* **29,** 289–303 (1995).
5. Müller, J. *et al.* Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. *Cell* **111,** 197–208 (2002).
6. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431,** 873–8 (2004).
7. Cao, R., Tsukada, Y.-I. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol. Cell* **20,** 845–54 (2005).
8. Lavigne, M., Francis, N. J., King, I. F. G. & Kingston, R. E. Propagation of Silencing. *Mol. Cell* **13,** 415–425 (2004).
9. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17,** 1870–81 (2003).
10. Francis, N. J., Kingston, R. E. & Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **306,** 1574–7 (2004).
11. Gambetta, M. C. & Müller, J. O-GlcNAcylation Prevents Aggregation of the Polycomb Group Repressor Polyhomeotic. *Dev. Cell* **31,** 629–639 (2014).
12. Grau, D. J. *et al.* Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* **25,** 2210–2221 (2011).
13. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38,** 452–64 (2010).

14. Lehmann, L. *et al.* Polycomb Repressive Complex 1 (PRC1) Disassembles RNA Polymerase II Preinitiation Complexes. doi:10.1074/jbc.M112.397430
15. Hennig, L. & Derkacheva, M. Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* **25,** 414–423 (2009).
16. Berke, L. & Snel, B. The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. *BMC Evol. Biol.* **15,** 44 (2015).
17. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20,** 1147–55 (2013).
18. Sowpati, D. T., Senthilkumar, R. & Mishra, R. K. Expansion of the polycomb system and evolution of complexity. *Mech. Dev.* (2015). doi:10.1016/j.mod.2015.07.013
19. Gil, J. & O'Loghlen, A. PRC1 complex diversity: where is it taking us? *Trends Cell Biol.* (2014). doi:10.1016/j.tcb.2014.06.005
20. Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* **14,** 853–64 (2013).
21. Kloet, S. L. *et al.* The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. *Nat. Struct. Mol. Biol.* **23,** 682–690 (2016).
22. Blackledge, N. P. *et al.* Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* **157,** 1445–1459 (2014).
23. Schwartz, Y. B. & Pirrotta, V. Ruled by Ubiquitylation: A New Order for Polycomb Recruitment. *Cell Rep.* **8,** 321–325 (2014).
24. Voncken, J. W. *et al.* Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2468–73 (2003).
25. Pirity, M. K., Locker, J. & Schreiber-Agus, N. Rybp/DEDAF is required for early postimplantation and for central nervous system development. *Mol. Cell. Biol.* **25,** 7193–202 (2005).
26. Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A. & Di Croce, L. *RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic*

Stem Cells. *Cell Reports* **3,** (2013).

27. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45,** 592–601 (2013).
28. Ford, D. J. & Dingwall, A. K. The cancer COMPASS: navigating the functions of MLL complexes in cancer. *Cancer Genet.* **208,** 178–191 (2015).
29. Kennison, J. A. THE POL YCOMB AND PROTEINS OF DROSOPHILA : Trans-Regulators of Homeotic Gene Function. *Annu. Rev. Genet.* 289–303 (1995).
30. Kim, K. H. & Roberts, C. W. M. Targeting EZH2 in cancer. *Nat. Med.* **22,** 128–134 (2016).
31. Conway, E., Healy, E. & Bracken, A. P. PRC2 mediated H3K27 methylations in cellular identity and cancer. *Curr. Opin. Cell Biol.* **37,** 42–48 (2015).
32. Pasini, D. & Di Croce, L. Emerging roles for Polycomb proteins in cancer. *Curr. Opin. Genet. Dev.* **36,** 50–58 (2016).
33. Koppens, M. & van Lohuizen, M. Context-dependent actions of Polycomb repressors in cancer. *Oncogene* **35,** 1341–1352 (2016).
34. Scelfo, A., Piunti, A. & Pasini, D. The controversial role of the Polycomb group proteins in transcription and cancer: how much do we not understand Polycomb proteins? *FEBS J.* **282,** 1703–1722 (2014).
35. Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14,** 735–51 (2014).
36. Kerppola, T. K. Polycomb group complexes--many combinations, many functions. *Trends Cell Biol.* **19,** 692–704 (2009).
37. Sauvageau, M. & Sauvageau, G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* **7,** 299–313 (2010).
38. Pietersen, A. M. & van Lohuizen, M. Stem cell regulation by polycomb repressors: postponing commitment. *Curr. Opin. Cell Biol.* **20,** 201–7 (2008).
39. Richly, H., Aloia, L. & Di Croce, L. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis.* **2,** e204 (2011).
40. Ma, R.-G., Zhang, Y., Sun, T.-T. & Cheng, B. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *J. Zhejiang Univ. Sci. B* **15,**

412–428 (2014).

41. Katoh-Fukui, Y. *et al.* Male-to-female sex reversal in M33 mutant mice. *Nature* **393,** 688–92 (1998).
42. Baumann, C. & De La Fuente, R. Role of polycomb group protein cbx2/m33 in meiosis onset and maintenance of chromosome stability in the Mammalian germline. *Genes (Basel).* **2,** 59–80 (2011).
43. Forzati, F. *et al.* CBX7 is a tumor suppressor in mice and humans. *J. Clin. Invest.* **122,** 612–623 (2012).
44. Scott, C. L. *et al.* Role of the chromobox protein CBX7 in lymphomagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 5389–94 (2007).
45. Gil, J., Bernard, D., Martínez, D. & Beach, D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nat. Cell Biol.* **6,** 67–72 (2004).
46. Pallante, P. *et al.* Loss of the CBX7 gene expression correlates with a highly malignant phenotype in thyroid cancer. *Cancer Res.* **68,** 6770–8 (2008).
47. Zhang, X.-W. *et al.* Oncogenic role of the chromobox protein CBX7 in gastric cancer. *J. Exp. Clin. Cancer Res.* **29,** 114 (2010).
48. Bernard, D. *et al.* CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. *Oncogene* **24,** 5543–51 (2005).
49. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20,** 1123–36 (2006).
50. Ram, O. *et al.* Combinatorial Patterning of Chromatin Regulators Uncovered by Genome-wide Location Analysis in Human Cells. *Cell* **147,** 1628–1639 (2011).
51. Chung, C.-Y. *et al.* Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary Tumorigenesis. *CellReports* **16,** 1–15 (2016).
52. Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10,** 47–62 (2012).
53. Creppe, C., Palau, A., Malinverni, R., Valero, V. & Buschbeck, M. A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLoS Genet.* **10,** e1004851 (2014).
54. Morey, L. *et al.* Polycomb Regulates Mesoderm Cell Fate-Specification in

Embryonic Stem Cells through Activation and Repression Mechanisms. *Cell Stem Cell* **17,** 300–15 (2015).

55. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16,** 643–649 (2015).
56. Sánchez-Beato, M. *et al.* Variability in the expression of polycomb proteins in different normal and tumoral tissues. A pilot study using tissue microarrays. *Mod. Pathol.* **19,** 684–694 (2006).
57. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347,** 1260419 (2015).
58. Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell Biol.* **15,** 353–62 (2013).
59. Molofsky, A. V. *et al.* Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425,** 962–967 (2003).
60. O'Loghlen, A. *et al.* MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation. *Cell Stem Cell* **10,** 33–46 (2012).
61. Iwama, A. *et al.* Enhanced Self-Renewal of Hematopoietic Stem Cells Mediated by the Polycomb Gene Product Bmi-1. *Immunity* **21,** 843–851 (2004).
62. Fasano, C. A. *et al.* shRNA Knockdown of Bmi-1 Reveals a Critical Role for p21-Rb Pathway in NSC Self-Renewal during Development. *Cell Stem Cell* **1,** 87–99 (2003).
63. Pemberton, H. *et al.* Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. *Genome Biol.* **15,** R23 (2014).
64. Dietrich, N. *et al.* Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. *EMBO J.* **26,** 1637–48 (2007).
65. Guo, W.-J., Datta, S., Band, V. & Dimri, G. P. Mel-18, a polycomb group protein, regulates cell proliferation and senescence via transcriptional repression of Bmi-1 and c-Myc oncoproteins. *Mol. Biol. Cell* **18,** 536–46 (2007).
66. Maertens, G. N. *et al.* Several distinct polycomb complexes regulate and co-

localize on the INK4a tumor suppressor locus. *PLoS One* **4,** e6380 (2009).

67. Zhang, X.-W. *et al.* BMI1 and Mel-18 oppositely regulate carcinogenesis and progression of gastric cancer. *Mol. Cancer* **9,** 40 (2010).
68. Guo, W.-J. *et al.* Mel-18 acts as a tumor suppressor by repressing Bmi-1 expression and down-regulating Akt activity in breast cancer cells. *Cancer Res.* **67,** 5083–9 (2007).
69. de Nigris, F. Epigenetic regulators: Polycomb-miRNA circuits in cancer. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859,** 697–704 (2016).
70. Cao, Q. *et al.* Coordinated Regulation of Polycomb Group Complexes through microRNAs in Cancer. *Cancer Cell* **20,** 187–199 (2011).
71. O'Loghlen, A. *et al.* CBX7 and miR-9 are part of an autoregulatory loop controlling p16(INK) (4a). *Aging Cell* **14,** 1113–21 (2015).
72. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* **45,** 344–356 (2012).
73. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163,** 712–723 (2015).
74. Wang, R. *et al.* Polycomb group targeting through different binding partners of RING1B C-terminal domain. *Structure* **18,** 966–75 (2010).
75. Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P. & Angrand, P.-O. Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol. Cell. Proteomics* **10,** M110.002642 (2011).
76. Sanchez, C. *et al.* Proteomics Analysis of Ring1B/Rnf2 Interactors Identifies a Novel Complex with the Fbxl10/Jhdm1B Histone Demethylase and the Bcl6 Interacting Corepressor. *Mol. Cell. Proteomics* **6,** 820–834 (2007).
77. Yokoyama, A. *et al.* Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell. Biol.* **24,** 5639–49 (2004).
78. Béguelin, W. *et al.* EZH2 and BCL6 Cooperate to Assemble CBX8-BCOR Complex to Repress Bivalent Promoters, Mediate Germinal Center Formation and Lymphomagenesis. *Cancer Cell* **30,** 197–213 (2016).
79. Wang, W. *et al.* Polycomb Group (PcG) Proteins and Human Cancers:

Multifaceted Functions and Therapeutic Implications. *Med. Res. Rev.* (2015). doi:10.1002/med.21358

80. Endoh, M. *et al.* Histone H2A Mono-Ubiquitination Is a Crucial Step to Mediate PRC1-Dependent Repression of Developmental Genes to Maintain ES Cell Identity. *PLoS Genet.* **8,** e1002774 (2012).
81. Zhou, W. *et al.* Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol. Cell* **29,** 69–80 (2008).
82. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* **9,** 1428–1435 (2007).
83. de Napoles, M. *et al.* Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev. Cell* **7,** 663–676 (2004).
84. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135,** 1513–24 (2008).
85. Taherbhoy, A. M., Huang, O. W. & Cochran, A. G. BMI1–RING1B is an autoinhibited RING E3 ubiquitin ligase. *Nat. Commun.* **6,** 7621 (2015).
86. Ben-Saadon, R., Zaaroor, D., Ziv, T. & Ciechanover, A. The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its in vitro histone H2A ligase activity. *Mol. Cell* **24,** 701–11 (2006).
87. Buchwald, G. *et al.* Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J.* **25,** 2465–74 (2006).
88. Lagarou, A. *et al.* dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* **22,** 2799–810 (2008).
89. Kagey, M. H., Melhuish, T. A. & Wotton, D. The polycomb protein Pc2 is a SUMO E3. *Cell* **113,** 127–37 (2003).
90. Li, B. *et al.* Polycomb protein Cbx4 promotes SUMO modification of de novo DNA methyltransferase Dnmt3a. *Biochem. J.* **405,** 369–78 (2007).

91. Ismail, I. H. *et al.* CBX4-mediated SUMO modification regulates BMI1 recruitment at sites of DNA damage. *Nucleic Acids Res.* **40,** 5497–510 (2012).
92. Bantignies, F. & Cavalli, G. Polycomb group proteins: repression in 3D. *Trends Genet.* **27,** 454–464 (2011).
93. Pirrotta, V. & Li, H.-B. A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* **22,** 101–109 (2012).
94. Entrevan, M., Schuettengruber, B. & Cavalli, G. Regulation of Genome Architecture and Function by Polycomb Proteins. *Trends Cell Biol.* **26,** 511–525 (2016).
95. Cheutin, T. & Cavalli, G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* **25,** 30–37 (2014).
96. Ren, X., Vincenz, C. & Kerppola, T. K. Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell. Biol.* **28,** 2884–95 (2008).
97. Vincenz, C. & Kerppola, T. K. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 16572–7 (2008).
98. Ferraiuolo, M. A. *et al.* The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* **38,** 7472–84 (2010).
99. Kheradmand Kia, S. *et al.* EZH2-dependent chromatin looping controls INK4a and INK4b, but not ARF, during human progenitor cell differentiation and cellular senescence. *Epigenetics Chromatin* **2,** 16 (2009).
100. Ohno, K., McCabe, D., Czermin, B., Imhof, A. & Pirrotta, V. ESC, ESCL and their roles in Polycomb Group mechanisms. *Mech. Dev.* **125,** 527–541 (2008).
101. Herz, H.-M. *et al.* Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in Drosophila. *Mol. Cell. Biol.* **32,** 1683–93 (2012).
102. Bernstein, E. *et al.* Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* **26,** 2560–9 (2006).
103. Kaustov, L. *et al.* Recognition and specificity determinants of the human cbx

chromodomains. *J. Biol. Chem.* **286,** 521–9 (2011).

104. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26,** (2013).
105. Lo, S. M. & Francis, N. J. Inhibition of chromatin remodeling by polycomb group protein posterior sex combs is mechanistically distinct from nucleosome binding. *Biochemistry* **49,** 9438–48 (2010).
106. Fujisaki, S. *et al.* Dimerization of the Polycomb-group protein Mel-18 is regulated by PKC phosphorylation. *Biochem. Biophys. Res. Commun.* **300,** 135–140 (2003).
107. Völkel, P., Le Faou, P., Vandamme, J., Pira, D. & Angrand, P.-O. A human Polycomb isoform lacking the Pc box does not participate to PRC1 complexes but forms protein assemblies and represses transcription. *Epigenetics* **7,** 482–91 (2012).
108. Tatavosian, R. *et al.* Distinct Cellular Assembly Stoichiometry of Polycomb Complexes on Chromatin Revealed by Single-molecule Chromatin Immunoprecipitation Imaging. *J. Biol. Chem.* **290,** 28038–54 (2015).
109. Kassis, J. A. & Brown, J. L. Chapter Three – Polycomb Group Response Elements in Drosophila and Vertebrates. in *Advances in Genetics* **81,** 83–118 (2013).
110. Müller, J. & Kassis, J. A. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.* **16,** 476–484 (2006).
111. Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1,** e00205 (2012).
112. He, J. *et al.* Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat. Cell Biol.* **15,** 373–384 (2013).
113. van Kruijsbergen, I., Hontelez, S. & Veenstra, G. J. C. Recruiting polycomb to chromatin. *Int. J. Biochem. Cell Biol.* **67,** 177–87 (2015).
114. Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nat. Rev. Mol. Cell Biol.* **2,** 409–421 (2001).
115. Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev.* **21,** 525–30 (2007).
116. Maruyama, R. *et al.* Epigenetic regulation of cell type-specific expression patterns

in the human mammary epithelium. *PLoS Genet.* **7,** e1001369 (2011).

117. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
118. Min, J., Zhang, Y. & Xu, R.-M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17,** 1823–8 (2003).
119. Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142,** 967–80 (2010).
120. Barna, M. *et al.* Plzf Mediates Transcriptional Repression of HoxD Gene Expression through Chromatin Remodeling. *Dev. Cell* **3,** 499–510 (2002).
121. Hemenway, C. S., de Erkenez, A. C. & Gould, G. C. D. The polycomb protein MPc3 interacts with AF9, an MLL fusion partner in t(9;11)(p22;q23) acute leukemias. *Oncogene* **20,** 3798–3805 (2001).
122. Maethner, E. *et al.* MLL-ENL inhibits polycomb repressive complex 1 to achieve efficient transformation of hematopoietic cells. *Cell Rep.* **3,** 1553–66 (2013).
123. Malik, B. & Hemenway, C. S. CBX8, a component of the Polycomb PRC1 complex, modulates DOT1L-mediated gene expression through AF9/MLLT3. *FEBS Lett.* **587,** 3038–44 (2013).
124. Yu, M. *et al.* Direct Recruitment of Polycomb Repressive Complex 1 to Chromatin by Core Binding Transcription Factors. *Mol. Cell* **45,** 330–343 (2012).
125. Ren, X. & Kerppola, T. K. REST Interacts with Cbx Proteins and Regulates Polycomb Repressive Complex 1 Occupancy at RE1 Elements. *Mol. Cell. Biol.* **31,** 2100–2110 (2011).
126. Boukarabila, H. *et al.* The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. *Genes Dev.* **23,** 1195–1206 (2009).
127. Mulholland, N. M., King, I. F. G. & Kingston, R. E. Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes Dev.* **17,** 2741–2746 (2003).
128. Chong, J. A. *et al.* REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80,** 949–957 (1995).

129. Chen, Z.-F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required in vivo for repression of multiple neuronaltarget genes during embryogenesis. *Nat. Genet.* **20,** 136–142 (1998).

130. Appleford, P. J. & Woollard, A. RUNX genes find a niche in stem cell biology. *J. Cell. Biochem.* **108,** 14–21 (2009).

131. Tie, F. *et al.* Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc. Natl. Acad. Sci. U. S. A.* **113,** E744-753 (2016).

132. Li, Q. *et al.* Polycomb CBX7 Directly Controls Trimethylation of Histone H3 at Lysine 9 at the p16 Locus. *PLoS One* **5,** e13732 (2010).

133. Mueller, D. *et al.* A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood* **110,** 4445–4454 (2007).

134. García-Cuéllar, M. P. *et al.* The ENL moiety of the childhood leukemia-associated MLL–ENL oncoprotein recruits human Polycomb 3. *Oncogene* **20,** 411–419 (2001).

135. Slany, R. K. The molecular biology of mixed lineage leukemia. *Haematologica* **94,** 984–93 (2009).

136. Tan, J. *et al.* CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* **20,** 563–75 (2011).

137. Nagano, T. & Fraser, P. Emerging similarities in epigenetic gene silencing by long noncoding RNAs. *Mamm. Genome* **20,** 557–562 (2009).

138. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* **43,** 904–914 (2011).

139. Lee, J. T. Epigenetic Regulation by Long Non-coding RNAs. *Science (80-.).* **338,** 1435–1439 (2012).

140. Gieni, R. S. & Hendzel, M. J. Polycomb group protein gene silencing, non-coding RNA, stem cells, and cancer. *Biochem. Cell Biol.* **87,** 711–746 (2009).

141. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–23 (2007).

142. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322,** 750–6 (2008).

143. Pasmant, E. *et al.* Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res.* **67,** 3963–9 (2007).

144. Ray, M. K. *et al.* CAT7 and cat7l long non-coding RNAs Tune Polycomb Repressive Complex 1 Function During Human and Zebrafish Development. *J. Biol. Chem.* jbc.M116.730853 (2016). doi:10.1074/jbc.M116.730853

145. Yap, K. L. *et al.* Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol. Cell* **38,** 662–674 (2010).

146. Ren, C. *et al.* Structure-Guided Discovery of Selective Antagonists for the Chromodomain of Polycomb Repressive Protein CBX7. *ACS Med. Chem. Lett.* **7,** (2016).

147. Hatano, A., Matsumoto, M., Higashinakagawa, T. & Nakayama, K. I. Phosphorylation of the chromodomain changes the binding specificity of Cbx2 for methylated histone H3. *Biochem. Biophys. Res. Commun.* **397,** 93–9 (2010).

148. Wu, H. H. -a. *et al.* Mitogen-activated protein kinase signaling mediates phosphorylation of polycomb ortholog Cbx7. *J. Biol. Chem.* **288,** 36398–408 (2013).

149. Voncken, J. W. *et al.* Chromatin-association of the Polycomb group protein BMI1 is cell cycle-regulated and correlates with its phosphorylation status. *J. Cell Sci.* **112 (Pt 2,** 4627–39 (1999).

150. Voncken, J. W. *et al.* MAPKAP Kinase 3pK Phosphorylates and Regulates Chromatin Association of the Polycomb Group Protein Bmi1. *J. Biol. Chem.* **280,** 5178–5187 (2005).

151. Katoh-Fukui, Y. *et al.* Cbx2, a polycomb group gene, is required for Sry gene expression in mice. *Endocrinology* **153,** 913–24 (2012).

152. Isono, K. -i. *et al.* Mammalian Polyhomeotic Homologues Phc2 and Phc1 Act in Synergy To Mediate Polycomb Repression of Hox Genes. *Mol. Cell. Biol.* **25,** 6694–6706 (2005).

153. Akasaka, T. *et al.* Mice doubly deficient for the Polycomb Group genes Mel18 and

Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128,** 1587–97 (2001).

154. Park, I. *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423,** 302–305 (2003).
155. Ren, C. *et al.* Small-Molecule Modulators of Methyl-Lysine Binding for the CBX7 Chromodomain. *Chem. Biol.* **22,** 161–8 (2015).
156. Ismail, I. H., McDonald, D., Strickfaden, H., Xu, Z. & Hendzel, M. J. A small molecule inhibitor of polycomb repressive complex 1 inhibits ubiquitin signaling at DNA double-strand breaks. *J. Biol. Chem.* **288,** 26944–54 (2013).
157. Kreso, A. *et al.* Self-renewal as a therapeutic target in human colorectal cancer. *Nat. Med.* **20,** 29–36 (2014).
158. Ohgaki, H. & Kleihues, P. Epidemiology and etiology of gliomas. *Acta Neuropathol.* **109,** 93–108 (2005).
159. Urbańska, K., Sokołowska, J., Szmidt, M. & Sysa, P. Glioblastoma multiforme - an overview. *Contemp. Oncol. (Poznan, Poland)* **18,** 307–312 (2014).
160. Sarica, F. B. *et al.* Five-year follow-up results for patients diagnosed with anaplastic astrocytoma and effectiveness of concomitant therapy with temozolomide for recurrent anaplastic astrocytoma. *Asian J. Neurosurg.* **7,** 181–90 (2012).
161. Johnson, D. R. & O'Neill, B. P. Glioblastoma survival in the United States before and during the temozolomide era. *J. Neurooncol.* **107,** 359–64 (2012).
162. Cheng, L., Bao, S. & Rich, J. N. Potential therapeutic implications of cancer stem cells in glioblastoma. *Biochem. Pharmacol.* **80,** 654–65 (2010).
163. Jackson, M., Hassiotou, F. & Nowak, A. Glioblastoma stem-like cells: at the root of tumor recurrence and a therapeutic target. *Carcinogenesis* **36,** 177–185 (2014).
164. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–13 (2006).
165. Gil, J., Bernard, D. & Peters, G. Role of polycomb group proteins in stem cell self-renewal and cancer. *DNA Cell Biol.* **24,** 117–25 (2005).
166. Li, G. *et al.* Altered expression of polycomb group genes in glioblastoma multiforme. *PLoS One* **8,** e80970 (2013).

167. Abdouh, M. *et al.* BMI1 sustains human glioblastoma multiforme stem cell renewal. *J. Neurosci.* **29,** 8884–96 (2009).
168. Fan, T.-Y. *et al.* Inhibition of EZH2 reverses chemotherapeutic drug TMZ chemosensitivity in glioblastoma. *Int. J. Clin. Exp. Pathol.* **7,** 6662–70 (2014).
169. Kim, S.-H. *et al.* EZH2 protects glioma stem cells from radiation-induced cell death in a MELK/FOXM1-dependent manner. *Stem cell reports* **4,** 226–38 (2015).
170. Clarke, J. *et al.* Epigenetic pathways and glioblastoma treatment. *Epigenetics* (2013).
171. Hong, Y., Shang, C., Xue, Y. & Liu, Y. Silencing of Bmi-1 gene enhances chemotherapy sensitivity in human glioblastoma cells. *Med. Sci. Monit.* **21,** 1002–7 (2015).
172. Verma, S. K. *et al.* Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Med. Chem. Lett.* **3,** 1091–6 (2012).
173. Miranda, T. B. *et al.* DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol. Cancer Ther.* **8,** 1579–88 (2009).
174. Zhang, J., Stevens, M. F. G. & Bradshaw, T. D. Temozolomide: mechanisms of action, repair and resistance. *Curr. Mol. Pharmacol.* **5,** 102–14 (2012).
175. Lee, E. Q. *et al.* Phase I study of vorinostat in combination with temozolomide in patients with high-grade gliomas: North American Brain Tumor Consortium Study 04-03. *Clin. Cancer Res.* **18,** 6032–9 (2012).
176. Tacar, O., Sriamornsak, P. & Dass, C. R. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *J. Pharm. Pharmacol.* **65,** 157–70 (2013).
177. Ling, Y. H., el-Naggar, A. K., Priebe, W. & Perez-Soler, R. Cell cycle-dependent cytotoxicity, G2/M phase arrest, and disruption of p34cdc2/cyclin B1 activity induced by doxorubicin in synchronized P388 cells. *Mol. Pharmacol.* **49,** 832–41 (1996).
178. Zhou, Y., Feng, X. & Koh, D. W. Enhanced DNA Accessibility and Increased DNA Damage Induced by the Absence of Poly(ADP-ribose) Hydrolysis. *Biochemistry* **49,** 7360–7366 (2010).

179. Chou, D. M. *et al.* A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc. Natl. Acad. Sci.* **107,** 18475–18480 (2010).
180. Gieni, R. S., Ismail, I. H., Campbell, S. & Hendzel, M. J. Polycomb group proteins in the DNA damage response: A link between radiation resistance and 'stemness'. *Cell Cycle* **10,** 883–894 (2011).
181. Ryu, C. H. *et al.* Valproic Acid Downregulates the Expression of MGMT and Sensitizes Temozolomide-Resistant Glioma Cells. *J. Biomed. Biotechnol.* **2012,** 1–9 (2012).
182. Kitange, G. J. *et al.* Inhibition of histone deacetylation potentiates the evolution of acquired temozolomide resistance linked to MGMT upregulation in glioblastoma xenografts. *Clin. Cancer Res.* **18,** 4070–9 (2012).
183. Barker, C. A., Bishop, A. J., Chang, M., Beal, K. & Chan, T. A. Valproic acid use during radiation therapy for glioblastoma associated with improved survival. *Int. J. Radiat. Oncol. Biol. Phys.* **86,** 504–9 (2013).
184. Pelisch, F., Pozzi, B., Risso, G., Munoz, M. J. & Srebrow, A. DNA Damage-induced Heterogeneous Nuclear Ribonucleoprotein K SUMOylation Regulates p53 Transcriptional Activation. *J. Biol. Chem.* **287,** 30789–30799 (2012).
185. Stuckey, J. I. *et al.* A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 11. Stuckey, J. I. et al. A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. Nat. Chem. Biol. 12, 180–187 (2016). *Nat. Chem. Biol.* **12,** 180–187 (2016).
186. Qin, Y. *et al.* Liposome formulated with TAT-modified cholesterol for improving brain delivery and therapeutic efficacy on brain glioma in animals. *Int. J. Pharm.* **420,** 304–312 (2011).
187. Sardi, I. *et al.* Pharmacological modulation of blood-brain barrier increases permeability of doxorubicin into the rat brain. *Am. J. Cancer Res.* **3,** 424–32 (2013).
188. Ferraro, A. Altered primary chromatin structures and their implications in cancer development. *Cell. Oncol.* **39,** 195–210 (2016).
189. Morgan, M. A. & Shilatifard, A. Chromatin signatures of cancer. *Genes Dev.* **29,**

238–49 (2015).

190. Sparmann, A. & van Lohuizen, M. Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* **6,** 846–56 (2006).
191. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature* **397,** 164–8 (1999).
192. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45,** D777–D783 (2017).
193. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–77 (2013).
194. Singh, S. K. *et al.* Identification of a Cancer Stem Cell in Human Brain Tumors. *Cancer Res.* **63,** 5821–5828 (2003).
195. Sauvageau, M. & Sauvageau, G. Polycomb group genes: keeping stem cell activity in balance. *PLoS Biol.* **6,** e113 (2008).
196. Facchino, S., Abdouh, M., Chatoo, W. & Bernier, G. BMI1 confers radioresistance to normal and cancerous neural stem cells through recruitment of the DNA damage response machinery. *J. Neurosci.* **30,** 10096–111 (2010).
197. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40,** 499–507 (2008).
198. Crea, F., Hurt, E. M. & Farrar, W. L. Clinical significance of Polycomb gene expression in brain tumors. *Mol. Cancer* **9,** 265 (2010).
199. Vital, A. L. *et al.* Gene expression profiles of human glioblastomas are associated with both tumor cytogenetics and histopathology. *Neuro. Oncol.* **12,** 991–1003 (2010).
200. Smith, C. L. *et al.* Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.* **46,** D836–D842 (2018).
201. Weifan Xiao, Chao Ou, Jinlong Qin, Feng Xing, Yi Sun, Zhi Li, J. Q. CBX8, a novel DNA repair protein, promotes tumorigenesis in human esophageal carcinoma. *International Journal of Clinical and Experimental Pathology* **7,** 4817 (2014).
202. Wang, G. *et al.* CBX8 Suppresses Tumor Metastasis via Repressing Snail in

Esophageal Squamous Cell Carcinoma. *Theranostics* **7,** 3478–3488 (2017).

203. Lee, S. H., Um, S.-J. & Kim, E.-J. CBX8 suppresses Sirtinol-induced premature senescence in human breast cancer cells via cooperation with SIRT1. *Cancer Lett.* **335,** 397–403 (2013).
204. Zhang, C. Z. *et al.* CBX8 Exhibits Oncogenic Activity via AKT/β-Catenin Activation in Hepatocellular Carcinoma. *Cancer Res.* **78,** 51–63 (2018).
205. Yuan, G.-J. *et al.* Chromobox homolog 8 is a predictor of muscle invasive bladder cancer and promotes cell proliferation by repressing the p53 pathway. *Cancer Sci.* **108,** 2166–2175 (2017).
206. Tang, J. *et al.* Paradoxical role of CBX8 in proliferation and metastasis of colorectal cancer. *Oncotarget* **5,** 10778–10790
207. Oza, J. *et al.* A Novel Role of Chromodomain Protein CBX8 in DNA Damage Response. *J. Biol. Chem.* **291,** 22881–22893 (2016).
208. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
209. Connelly, K. E., Hedrick, V., Paschoal Sobreira, T. J., Dykhuizen, E. C. & Aryal, U. K. Analysis of Human Nuclear Protein Complexes by Quantitative Mass Spectrometry Profiling. *Proteomics* **18,** 1700427 (2018).
210. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46,** W537–W544 (2018).
211. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–40 (2010).
212. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40,** 4288–97 (2012).
213. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45,** D331–D338 (2017).
214. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–9 (2000).
215. Lee, C.-H., Yu, C.-C., Wang, B.-Y. & Chang, W.-W. Tumorsphere as an effective

in vitro platform for screening anti-cancer stem cell drugs. *Oncotarget* **7,** 1215–26 (2016).

216. Santiago, C., Nguyen, K. & Schapira, M. Druggability of methyl-lysine binding sites. *J. Comput. Aided. Mol. Des.* **25,** 1171–1178 (2011).
217. Simhadri, C. *et al.* Chromodomain antagonists that target the polycomb-group methyllysine reader protein Chromobox homolog 7 (CBX7). *J. Med. Chem.* **7,** (2014).
218. Porter, E. G., Connelly, K. E. & Dykhuizen, E. C. Sequential Salt Extractions for the Analysis of Bulk Chromatin Binding Properties of Chromatin Modifying Complexes. *J. Vis. Exp.* e55369–e55369 (2017). doi:10.3791/55369
219. Lines, G. C. *et al.* Frequent Co-Alterations of TP53, p16/CDKN2A, p14ARF , PTEN Tumor Suppressor Genes in Human Glioma Cell Lines. *Brain Pathol.* **9,** 469–479 (1999).
220. Werner, J. N. *et al.* Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 7858–63 (2009).
221. Suvà, M.-L. *et al.* EZH2 is essential for glioblastoma cancer stem cell maintenance. *Cancer Res.* **69,** 9211–8 (2009).
222. Denton, K. *et al.* Robustness of In Vitro Selection Assays of DNA-encoded Peptidomimetic Ligands to CBX7 and CBX8. *SLAS Discov.* (2018).
223. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21,** 381–95 (2011).
224. Bowman, G. D. & Poirier, M. G. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chem. Rev.* **115,** 2274–2295 (2015).
225. Musselman, C. A., Lalonde, M.-E., Côté, J. & Kutateladze, T. G. Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **19,** 1218–27 (2012).
226. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res.* **21,** 564–78 (2011).
227. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039–43 (2002).
228. Turner, S. A. & Bracken, A. P. A ‘complex’ issue: Deciphering the role of variant

PRC1 in ESCs. *Cell Stem Cell* **12,** (2013).

229. Gutiérrez, L. *et al.* The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* **139,** 117–27 (2012).
230. Connelly, K. E. & Dykhuizen, E. C. Compositional and functional diversity of canonical PRC1 complexes in mammals. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1860,** 233–245 (2017).
231. Zhen, C. Y. *et al.* Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. *Elife* **5,** (2016).
232. Teske, K. A. & Hadden, M. K. Methyllysine binding domains: Structural insight and small molecule probe development. *Eur. J. Med. Chem.* **136,** 14–35 (2017).
233. Mello, J. B. H. de *et al.* Genomic profile in gestational and non-gestational choriocarcinomas. *Placenta* **50,** 8–15 (2017).
234. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33,** 661–667 (2015).
235. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128,** 1003–1012 (2007).
236. Qiu, Y. *et al.* Solution structure of the Pdp1 PWWP domain reveals its unique binding sites for methylated H4K20 and DNA. *Biochem. J.* **442,** 527–38 (2012).
237. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–93 (1995).
238. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59,** 687–696 (2005).
239. Bahrami, A., Assadi, A. H., Markley, J. L. & Eghbalnia, H. R. Probabilistic Interaction Network of Evidence Algorithm and its Application to Complete Labeling of Peak Lists from Protein NMR Spectroscopy. *PLoS Comput. Biol.* **5,** e1000307 (2009).
240. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
241. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44,** D726-32 (2016).
242. ENCODE Project Consortium, T. E. P. An integrated encyclopedia of DNA

elements in the human genome. *Nature* **489,** 57–74 (2012).

243. Zhu, L. J. Integrative Analysis of ChIP-Chip and ChIP-Seq Dataset. in 105–124 (Humana Press, Totowa, NJ, 2013). doi:10.1007/978-1-62703-607-8_8
244. Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11,** 237 (2010).
245. Seeliger, D. *et al.* Quantitative Assessment of Protein Interaction with Methyl-Lysine Analogues by Hybrid Computational and Experimental Approaches. *ACS Chem. Biol.* **7,** 150–154 (2012).
246. Chen, Z., Notti, R. Q., Ueberheide, B. & Ruthenburg, A. J. Quantitative and Structural Assessment of Histone Methyllysine Analogue Engagement by Cognate Binding Proteins Reveals Affinity Decrements Relative to Those of Native Counterparts. *Biochemistry* **57,** 300–304 (2018).
247. Cutter, A. R. & Hayes, J. J. A brief review of nucleosome structure. *FEBS Lett.* **589,** 2914–22 (2015).
248. Stützer, A. *et al.* Modulations of DNA Contacts by Linker Histones and Post-translational Modifications Determine the Mobility and Modifiability of Nucleosomal H3 Tails. *Mol. Cell* **61,** 247–259 (2016).
249. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution. *J. Mol. Biol.* **319,** 1097–1113 (2002).
250. Morrison, E. A., Bowerman, S., Sylvers, K. L., Wereszczynski, J. & Musselman, C. A. The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. *Elife* **7,** (2018).
251. Weaver, T. *et al.* Reading More than Histones: The Prevalence of Nucleic Acid Binding among Reader Domains. *Molecules* **23,** 2614 (2018).
252. Shirai, A. *et al.* Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *Elife* **6,** (2017).
253. Johnson, W. L. *et al.* RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *Elife* **6,** (2017).
254. Ishida, M. *et al.* Intrinsic Nucleic Acid-Binding Activity of Chp1 Chromodomain Is Required for Heterochromatic Gene Silencing. *Mol. Cell* **47,** 228–241 (2012).

255. Kim, D. *et al.* Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat. Struct. Mol. Biol.* **17,** 1027–9 (2010).

256. Bryan, L. C. *et al.* Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions. *Nucleic Acids Res.* **45,** 10504–10517 (2017).

257. Philpott, M. *et al.* Assessing cellular efficacy of bromodomain inhibitors using fluorescence recovery after photobleaching. *Epigenetics Chromatin* **7,** 14 (2014).

258. Kawaguchi, T., Machida, S., Kurumizaka, H., Tagami, H. & Nakayama, J. Phosphorylation of CBX2 controls its nucleosome-binding specificity. *J. Biochem.* **162,** 343–355 (2017).

259. Cowell, I. G. & Austin, C. A. Self-association of chromo domain peptides. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1337,** 198–206 (1997).

260. Platero, J. S., Hartnett, T. & Eissenberg, J. C. Functional analysis of the chromo domain of HP1. *EMBO J.* **14,** 3977–86 (1995).

VITA

Katelyn Elizabeth Connelly, daughter of Scott and Evelyn Connelly, was born March 29, 1992 in Bay City, Michigan. She was raised in Grand Blanc, MI where she graduated from Grand Blanc High School in 2010. Katelyn went on to The Ohio State University to earn her Bachelor of Science in Pharmaceutical Science in 2013. Following completion of her degree, she joined the department of Medicinal Chemistry and Molecular Pharmacology at Purdue University's College of Pharmacy. In December 2013, she joined Dr. Emily C. Dykhuizen's lab as the first graduate student. Katelyn's work in the Dykhuizen lab has been focused on understanding the role of the CBX proteins, specifically CBX8, in cancer and chromatin targeting. During her PhD studies, Katelyn earned the Purdue Graduate School Ross Award, the Indiana Clinical and Translational Sciences Institute pre-doctoral fellowship, the Koo Travel Award, and the Graduate School's Bilsland Dissertation Fellowship. In addition to her time in the lab, Katelyn was actively involved in Purdue's Graduate Student Government and MCMP's graduate student organization to improve the graduate experience for all. Following the completion of her PhD, Katelyn will be joining the laboratory of Translational Genomics at the National Cancer Institute in the Division of Cancer Epidemiology and Genetics working as a post-doctoral fellow under Dr. Laufey Amundadottir.

PUBLICATIONS

Connelly, K.E.,* Weaver, T.M.,* Alpsoy, A., Gu, B.X., Musselman, C.A., Dykhuizen, E.C., Engagement of DNA and H3K27me3 by the CBX8 chromodomain. *Nucleic Acids Research*. 2018 *(under revision)*

*Authors contributed equally

Weaver, T.M., Liu, J., **Connelly, K.E.,** Coble, C., Varzavand, K., Dykhuizen, E.C., Musselman, C.A., The EZH2 SANT1 domain is a histone reader providing sensitivity to the modification state of the H4 tail. *Scientific Reports* 2018 *(under revision)*

Connelly, K.E., Hedrick, V., Sobreira, T., Dykhuizen, E.C., Aryal, U.K., Systematic analysis of human nuclear protein complexes by quantitative mass spectrometry profiling. *PROTEOMICS 2018.* doi: 10.1002/pmic.201700427

Porter, E.G., **Connelly, K.E.,** Dykhuizen, E.C., Sequential Salt Extractions for the Analysis of Bulk Chromatin Binding Properties of Chromatin Modifying Complexes. *J. Vis. Exp*. (128) e55369. doi: 10.3791/55369 (2017)

Connelly, K. E., Dykhuizen, E. C., Compositional and functional diversity of canonical PRC1 complexes in mammals. *Biochim Biophys Acta* 2017, 1860, 233-245.

Connelly, K.E., Martin, E.C., Dykhuizen, E.C., CBX Chromodomain Inhibition Enhances Chemotherapy Response in Glioblastoma Multiforme. *Yale J Biol Med*. 2016 Dec 23;89(4):431-440.