THE ROLE OF CFP1 IN MURINE EMBRYONIC STEM CELL FUNCTION AND LIVER REGENERATION

Jyothi Mahadevan

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biochemistry and Molecular Biology,
Indiana University

August 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

David G. Skalnik, Ph.D., Chair

_____

Mark G. Goebl, Ph.D.

Doctoral Committee

_____

Maureen A. Harrington, Ph.D.

May 11, 2015

_____

B. Paul Herring, Ph.D

Dedication

To my loving parents and husband

I owe my success to their unconditional love and constant encouragement. I thank you for always being there for me and supporting me through this arduous journey.

Acknowledgements

Firstly, I would like to express my gratitude to my mentor Dr. David Skalnik for providing me with an opportunity to carry out my doctoral research in his laboratory. The scientific skills that I have imbibed under his excellent mentorship will remain with me for my lifetime. I appreciate his time, effort and interest in guiding me through my graduate career. I also thank him for his kindness during my difficult times.

I would like to thank the members of my committee, Dr. Mark Goebl, Dr. Maureen Harrington and Dr. Paul Herring for their ideas and suggestions regarding my thesis project. I am grateful to Dr. Simon Rhodes for his help in choosing a lab for my graduate work. I would like to express gratitude to Dr. Guoli Dai for his collaboration for my liver-related project and letting me use equipment in his lab. I am also grateful to Dr. Amber Mosley for her suggestions regarding my project and Dr. Jeanette McClintick for her help in bioinformatic analysis. I would like to thank Chris Konz, Julie Nigro and Vickie Birge from SARC for their training and help in maintaining my mouse colony. I am also grateful to all my teachers from India who have been instrumental in my success.

I am thankful to the former members of the Skalnik lab especially Erika Dobrota for her relentless support both inside and outside of the lab and making my doctoral journey an enjoyable one. I will never forget the engaging conversations we have had, both work related and personal. I also acknowledge the interactions I have had with Dr. Jeong-Heon Lee that made me understand the importance of critical reasoning and looking at the big picture in science. I would like to express my gratitude to Ryan Eller, Tammy Bowen, Jonathan Bendinger and Prathik Kini for providing me with an opportunity to excel at mentorship.

I would like to thank my family for being my greatest support at all times. This work would not have been possible without the love and nurturing I have received from my parents Lalitha and Mahadevan. They have constantly motivated me to do better and achieve greater heights. They have equipped me with everything I needed for reaching this point in my life. I would like to express my sincere gratitude to my loving husband,

Anish, who has always been there for me through every thick and thin of my life. I also appreciate his help in improving my manuscript writing skills. I look forward to spend the rest of my life in his company. I would like to thank my sister Latha, brother-in-law Krishnan and nephews Rahul and Gokul for bringing cheer in my life. I also acknowledge my parents-in-law Sandhya and Krishnamoorthy for trusting me and supporting my decisions and my sister-in-law Arti for her love and affection.

I would also like to extend thanks to my friends who have made my graduate journey more meaningful. I would like to thank Vidya, Kartik, Swapna, Lalitha, Purvi, Isha and Deepthi for their friendship and support. I would like to specially thank my flat-mate and friend Abhirami for always being there for me. I will always remember the good times we have spent together. I am grateful to my seniors Latha, Chandra and Sneha for their help and encouraging words. I am also thankful to Siddharth for giving me driving lessons. I would like to thank everyone who have made a difference in my life, small or big, with their kind words and/or deeds.

Jyothi Mahadevan

THE ROLE OF CFP1 IN MURINE EMBRYONIC STEM CELL FUNCTION AND LIVER REGENERATION

CXXC finger protein 1 (Cfp1), a component of the Set1 histone methyltransferase complex, is a critical epigenetic regulator of both histone and cytosine methylation. Murine embryos lacking Cfp1 are unable to gastrulate and Cfp1-null embryonic stem (ES) cells fail to undergo cellular differentiation *in vitro*. However, expression of wild type Cfp1 in Cfp1-null ES cells rescues differentiation capacity, suggesting that dynamic epigenetic changes occurring during lineage specification require Cfp1. The domain structure of Cfp1 consists of a DNA binding CXXC domain and an N-terminal plant homeodomain (PHD). PHDs are frequently observed in chromatin remodeling proteins, functioning as reader modules for histone marks. However, the histone binding properties and underlying functional significance of Cfp1 PHD are largely unknown. My research revealed that Cfp1 PHD directly and specifically binds to histone H3K4me1/me2/me3 marks. A point mutation that abolishes binding to methylated H3K4 (W49A) does not affect rescue of cellular differentiation, but, point mutations that abolish both methylated H3K4 (W49A) and DNA (C169A) binding result in defective *in vitro* differentiation, indicating that PHD and CXXC exhibit redundant functions.

The mammalian liver has the unique ability to regenerate following injury. Previous studies indicated that Cfp1 is essential for hematopoiesis in zebrafish and mice. I hypothesized that Cfp1 additionally plays a role in liver development and regeneration. To understand the importance of Cfp1 in liver development and regeneration, I generated a mouse line lacking Cfp1 specifically in the liver ($Cfp1^{fl/fl}$ Alb-Cre+). Around 40% of these mice display a wasting phenotype and die within a year. Livers of these mice have altered global H3K4me3 levels and often exhibit regenerative nodules. Most importantly, livers of these mice display an impaired regenerative response following partial hepatectomy. Collectively, these findings establish Cfp1 as an epigenetic regulator essential for ES cell function and liver homeostasis and regeneration.

David G. Skalnik, Ph.D., Chair

Table of Contents

List of Tables

List of Figures

Abbreviations

μ	micro
A	alanine
aa	amino acid
Alb	albumin
AP	Alkaline Phosphatase
bp	base pairs
BSA	bovine serum albumin
C	cysteine
cDNA	complementary DNA
CFP1	CXXC Finger Protein 1
ChIP-Seq	Chromatin Immunoprecipitation-Sequencing
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
DSB	double stranded break
DTT	dithiothreitol
E	glutamic acid
EDTA	ethylenediaminetetraacetic acid
ES	embryonic stem
ESC	embryonic stem cell
EZH1	enhancer of zeste homolog 1
FITC	fluorescein isothiocyanate
g	gram (s)
G	glycine
GST	glutathione-S-transferase
h	hour (s)
H	Histidine
H	histone
H&E	hematoxylin & eosin
HAT	histone acetyltransferase
HDAC	histone deacetylase complex

HEK	Human embryonic kidney
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HMT	histone methyltransferase
HP1	Heterochromatin Protein 1
Hygro	hygromycin
IB	immunoblot
IP	immunoprecipitation
K	lysine
LB broth	Luria Bertani
LIF	Leukemia inhibitory factor
M	molar, moles per liter
MBD	methyl-CpG binding domain
me	methylated
min	minute (s)
ml	milliliter (s)
mM	millimolar
Mnase	micrococcal nuclease (s)
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
ng	nanogram (s)
OD	optical density
P	proline
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PH	partial hepatectomy
PHD	plant homeodomain
PIPES	1,4-piperazinediethanesulfonic scid
PMSF	phenylmehtylsulfonyl fluoride
PRC	polycomb repressive complex
Pygo	Pygopus
R	arginine

RNA	ribonucleic Acid
RNAPII	RNA polymerase II
RNF	ring finger protein
rpm	revolutions per minute
RT-PCR	Reverse Transcriptase-polymerase chain reaction
S	serine
sdH_20	sterile distilled water
SDS	sodium dodecyl sulphate
SET/Set	Suppressor of Variegation, Enhancer of Zeste and Trithorax
SID	Set1-interaction domain
SUMO	small ubiquitin-related modifier
SUV	suppressor of variegation
Taq	Thermus aquaticus
TBE	tris-Borate EDTA
TBS	tris buffered Saline
TBS-T	tris buffered Saline –Tween 20
V	volts
vec	vector
W	tryptophan
WT	wild type
x g	centrifugal force, gravity
Y	tyrosine

Introduction

1. Fundamentals of Epigenetics

Almost every cell of a living organism contains identical genetic content, yet performs distinct functions. One of the ways this is made possible is by “epigenetic changes” in gene expression without involving changes to the underlying DNA sequence. These changes are heritable, self-perpetuating and reversible [1] and are mediated by processes that regulate the chromatin landscape, such as DNA methylation and histone modifications. The complexities of gene regulation, aging, cellular differentiation, cancer and other diseases can be better appreciated with an in-depth knowledge of their underlying epigenetic mechanisms.

2. Structure of chromatin

The eukaryotic genome is packaged into a compact form called chromatin to be accommodated into the nuclei of cells. Chromatin is composed of fundamental subunits called nucleosomes. Each nucleosome is comprised of 146 base pairs of genomic DNA wrapped around an octameric core consisting of two molecules each of histone protein H2A, H2B, H3 and H4. Core particles are connected by ~80 base pair long stretches of linker DNA. Unlike other histone proteins that form the core, H1 functions as a linker connects the exit/entry of the DNA strand on the nucleosome. Nucleosomes along with 20-60 base pairs of DNA form a 10 nm “beads on a string” structure [2]. Further, chromatin is condensed into 30 nm chromatin fibers consisting of chains of nucleosomes. These chromatin fibers undergo higher order condensation with the help of the linker H1 histone, ultimately leading to chromosome formation.

Chromatin is territorially divided into heterochromatin and euchromatin on the basis of states of compaction and transcriptional potential. Heterochromatin includes transcriptionally silent regions that are densely packaged and stained intensely, when subjected to dyes that bind DNA. These regions remain condensed throughout the cell cycle. Heterochromatin is further subdivided into constitutive and facultative heterochromatin. Constitutive heterochromatin is always compact, gene-deficient, repetitive and fulfills structural roles such as telomeres and centromeres. On the other

hand, facultative heterochromatin comprises of condensed, transcriptionally silent regions that can reversibly de-condense and allow transcription to occur on receiving suitable developmental signals. Heterochromatin is characterized by DNA hypermethylation, histone hypoacetylation and methylation of histone H3 at lysine 9 and lysine 27 in higher eukaryotes.

Euchromatin consists of loosely packaged regions of chromatin that stain lightly with DNA-binding dyes. This region is concentrated with genes and is often transcriptionally accessible. Euchromatin regions undergo cycles of condensation and de-condensation at various stages of cell cycle. These regions show hypomethylation of DNA along with a predominance of histone acetylation and trimethylation of H3K4, H3K36 and H3K79.

3. DNA cytosine methylation

DNA cytosine methylation involves the covalent transfer of a methyl group to the C-5 position of the cytosine ring of DNA by enzymes known as DNA methyltransferases (DNMTs). It is a heritable epigenetic mark associated with gene repression and plays a crucial role in X-inactivation, genomic imprinting and genome stabilization. DNA methylation is also essential for the transcriptional silencing of transposons and retroviruses that have accumulated in the mammalian genome [3]. Methylation of cytosine residues of DNA generally occurs in a symmetric CpG dinucleotide context [4]. The occurrence of CpG in the vertebrate genome is relatively rare ($1/5^{th}$ of the expected frequency) [5]. Unmethylated cytosine can undergo spontaneous deamination into uracil which is easily recognized as an inappropriate nucleotide and repaired by uracil-DNA glycosylase (UDG) [6]. On the other hand, methylated cytosines undergo spontaneous deamination to form thymidine thereby mutating CpG to TpG dinucleotides upon replication of DNA. This mutation is not easily recognized by the DNA excision repair enzymes thymine DNA glycosylase (TDG) or methyl-CpG-binding protein 4 (MBD4) and this results in an overall loss of CpG dinucleotide. This explains the relatively rare occurrence of the CpG dinucleotide in the vertebrate genome. However, the CpG-deficient genome is punctuated by CpG islands (CGIs) that are approximately 1000 base pair long, have an elevated G-C base composition, are rich in CpG dinucleotides and show a frequent absence of DNA methylation [7]. The CGIs have been shown to be functionally

important as sites for transcription initiation. Approximately 70% of the annotated gene promoters contain CGIs and about half of all CGIs track with promoters of annotated genes [7, 8].

Mammals express three major forms of DNA methyltransferases, DNMT1, DNMT3a and DNMT3b. DNMT1, is the first identified and most ubiquitous DNMT enzyme, that functions by preferentially methylating hemi-methylated DNA and is required for inheriting the pre-existing DNA methylation patterns from parent DNA strands to newly synthesized daughter DNA strands [9]. DNMT1 is therefore required for maintenance of methylation patterns following DNA replication. DNMT 3a and 3b are expressed highly in developmental stages. They have a preference for unmethylated CpG dinucleotides and are involved in *de novo* methylation during development [3, 10]. However, current research suggests that DNMT1 may be required for *de novo* methylation of genomic DNA [11] and that DNMT 3a and 3b might be required for maintenance of DNA methylation [12, 13].

The code of DNA methylation is interpreted partly by methyl-CpG binding proteins (MBDs). These proteins associate with histone modifiers to establish transcriptionally silent chromatin domains. MeCP2 is one such MBD that associates with methylated DNA to recruit histone deacetylase (HDAC) and repress transcription in a methylation-dependent manner [14, 15]. This is suggestive of the inter-relationships between DNA methylation-induced transcriptional silencing and histone modifications.

4. Posttranslational modifications of histones (PTMs)

Core histone proteins consist of globular carboxy-terminal domain that constitute the bead-like structure of the nucleosome and flexible amino- terminal tails protruding outward from the nucleosome. These tail domains provide a signaling platform for interaction with other proteins/protein complexes that function to remodel chromatin [16]. Histone tail domains undergo a number of PTMs including acetylation, ubiquitylation, sumoylation, phosphorylation and methylation. Several combinations of modifications can exist on single histones or single nucleosomes, that result in

establishing local and global patterns of chromatin marks thereby specifying unique downstream activities [17].

a. Acetylation

Histone acetylation involves the transfer of acetyl groups from acetyl-coenzyme A (acetyl-coA) onto the ε-amino groups of conserved lysine residues, catalyzed by histone acetyltransferases (HATs). Histone acetylation is the most extensively studied histone modification and is associated with transcriptionally permissible chromatin. Acetylation changes the overall charge on histone tails from positive to neutral. This causes a reduction in the affinity of N-terminal histone tail domains to the negatively charged DNA resulting in an increase in accessibility of DNA to transcription factors [18]. Thus, acetylation of histones results in increased expression of genes due to transcriptional activation [16, 19, 20].

The conserved transcriptional regulator Gcn5 was the first protein found to have histone acetyltransferase activity. Gcn5 and the partly homologous PCAF commonly modify lysine 14 of histone H3 and are known to be associated with co-activators or adaptors in multi-subunit complexes. The Gcn5 containing SAGA (Spt-Ada-Gcn5-acetyltransferase) complex is composed of 19 or more subunits in yeast [21]. This complex consists of a large number of co-activators that preferentially modify H3 and H2A [16, 19]. However, the core acetyltransferase activity of the SAGA complex is contained in the Gcn5 subunit [22]. There are several families of acetyltransferases including the GNAT (Gcn5-related N-acetyltransferases) superfamily that includes Gcn5, PCAF, Elp3 and Hat1; the p300/CBP HAT family that share sequence homology with GNATs and act as transcriptional co-activators; and the MYST family of HATs that derives its name from its founding members (**M**OZ, **Y**bf2/Sas3, **S**as2 and **T**ip60) [19, 23].

Another protein with HAT activity, known as TAF_I (TATA box associated factor), forms a part of the evolutionarily conserved TFIID complex along with TBP (TATA binding protein) [24]. TBP recognizes the TATA element found in the core promoter regions of many genes and functions to recruit other proteins to initiate gene transcription. This

suggests that HAT activity at the core promoter may be important for the interaction of transcription factors with nucleosomal DNA [23].

The reversible histone acetylation mark is removed by enzymes known as histone deacetylases (HDACs). HDACs are found in corepressor complexes and function by removing acetyl groups from histone tails resulting in condensation of chromatin structure. Eighteen different mammalian HDACs have been identified and have been placed into four distinct classes based on sequence similarity to yeast HDACs. HDACs requiring Zn^{2+} for deacetylase activity are included in Class I, II and IV. Class I HDACs possess sequence similarity to yeast RPD3 and comprise of HDAC1, HDAC2, HDAC3 and HDAC8. Class II HDACs are related to yeast HDA1 and are subdivided into class IIa (HDAC4, HDAC5, HDAC7 and HDAC9) and subclass IIb (HDAC6 and HDAC10). Class III HDACs contains 7 sirtuins which require NAD^+ as a cofactor and Class IV contains only HDAC 11 [25, 26]. Mammalian Class I HDACs (subtypes 1, 2, 3 and 8) have high sequence similarity and are abundantly expressed in mSin3A, NURD/Mi2/NRD and CoREST corepressor complexes [27]. HDAC3 associates with and is activated by the SMRT and NCoR co-repressors that function along with nuclear hormone receptors to regulate gene expression [27, 28]. These examples suggest that HATs and HDACs are two counteracting enzyme families that perform the dual functions of regulation of histone acetylation and association with DNA binding transcriptional activators to bring about chromatin remodeling to facilitate appropriate gene expression.

b. Ubiquitination

Ubiquitin, a 76 aa protein, is conjugated to substrate proteins in a three step reaction: an ATP-dependent activation reaction requiring ubiquitin activating enzyme (E1), followed by conjugation via a thioester linkage to a cysteine residue on a ubiquitin-conjugating enzyme (E2) and finally, ubiquitin is transferred from the enzyme E2 to a target lysine residue of the substrate protein by a ubiquitin-protein isopeptide ligase (E3) [29]. While polyubiquitination targets proteins for proteasome-mediated degradation, mono-ubiquitination occurs on histone H2A and H2B and is a widely studied chromatin modification. Occurrence of ubH2A is frequently associated with gene repression while ubH2B is mostly associated with transcriptional activation. H2A ubiquitination occurs at

lysine 119 through at least two different E3 ubiquitin ligases, Ring 1B and 2A-HUB that are associated with transcriptional silencing [30]. ubH2A represses transcription initiation and inhibits the deposition of euchromatin-associated markers H3K4me2/me3 [31]. Ring1B associates with Polycomb Repressive Complex 1 (PRC1). Moreover, ubH2A and PRC1 colocalize on the inactive X chromosome in mouse [32] [33]. The other H2A ubiquitin ligase, 2A-HUB, associates with the N-CoR/HDAC1/3 complex and represses transcription of some chemokine genes [34].

Monoubiquitination of H2B occurs at lysine 120 in vertebrates (lysine 123 in yeast) by two ubiquitin conjugating enzymes, homologous to yeast Rad6 enzyme: HR6A and HR6B (hRad6A/hRad6B). RNF20 and RNF40 function as ubiquitin ligases (homologous to yeast Bre1) and form a complex with HR6A and HR6B to catalyze H2B monoubiquitination [35]. In mammalian cells, RNF20/RNF40 associates with transcriptionally active genes [36]. H2B ubiquitination also affects di and tri-methylation of H3K4 and H3K79 [37]. Current research suggests that the COMPASS methyltransferase complex requires prior ubiquitination of H2B for di and trimethylation of H3K4 [38]. This provides an example of a crosstalk between histone ubiquitination and another histone mark, histone methylation.

c. Sumoylation

Sumoylation, like ubiquitination, requires covalent attachment of small ubiquitin like molecules to lysine residues on histones via the action of E1, E2 and E3 enzymes [39]. Sumoylation occurs on all core histones within chromatin and occurs with high frequency at sub-telomeres [40]. It is a mark that functions by antagonizing both lysine acetylation and ubiquitination on the same histone tail and is therefore frequently associated with transcriptional repression [41].

d. Phosphorylation

Phosphorylation (ph) of histones at serine or threonine residues is a dynamic process that is regulated by kinases and phosphatases that add and remove this mark respectively. A very well studied event is the phosphorylation of the histone variant H2AX, that plays a significant role in DNA damage response. This modification occurs at the serine (S) 139

of the H2AX variant and is commonly referred to as γH2AX. Phosphorylation of H2AX is one of the first events that follow DNA damage and it occurs in the vicinity of the double stranded break (DSB). This is followed by the bidirectional spreading of this mark over several megabases in the mammalian cells [42, 43]. The wide-spread nature of γH2AX around the DSB is thought to be required for the recruitment of DNA damage repair and signaling factors for facilitating downstream repair pathways [44, 45]. Phosphorylation of serine 1 of histone H4 is mediated by casein kinase II (CKII) in response to genotoxic stress induced by exposure to ultraviolet light and methyl methanesulfonate (MMS) [46] [47].

Phosphorylation of histone residues is also linked to transcriptional activation. Regulation of epidermal growth factor (EGF)-responsive genes involves phosphorylation of serines 10 and 28 of H3 and serine 32 of H2B [48, 49]. Occurrence of H3S10ph throughout the cell cycle is conserved in eukaryotes and hence H3S10ph has been used as a marker for mitotic cells in many studies [50, 51]. Phosphorylation of H3S10, T11 and S28 has been associated with Gcn5-dependent H3 acetylation. H3S10ph promotes acetylation of H3K14 and facilitates Gcn5-mediated transcriptional activation *in vivo* [52]. This coupling of H3S10ph and H3K14ac is also required for the regulation of gene expression during mesoderm formation during embryonic development [53]. Another mark, H3S28ph is found to displace the polycomb repressive complexes present at gene promoters to induce demethylation and acetylation of the adjacent H3K27 residues at these loci [47, 54]. Thus, histone phosphorylation is associated with the regulation of a number of varied cellular processes.

e. Methylation

All the four core histones can undergo methylation at arginine and lysine residues. Arginine methylation involves transfer of a methyl group from S-adenosyl methionine to a guanidino nitrogen of arginine resulting in the formation of methylarginine. Arginine residues can either undergo monomethylation and asymmetric or symmetric dimethylation catalyzed by protein arginine methyltransferases (PRMTs). PRMT1 is responsible for the bulk of arginine methylation activity and is required for the generation of asymmetric H4R3me2a, which functions as a transcriptional activation mark [55].

PRMT4 or CARM1 is a steroid receptor coactivator and is responsible for H3R17me2a and H3R26me2a [56]. On the contrary, PRMT5 acts as a strong transcriptional repressor [57, 58].

Histone lysine methylation involves the methylation of the ε-amino group of lysine residues by utilizing the cofactor S-adenosyl-L-methionine (SAM). The SET domain containing proteins form a major group of histone lysine methyltransferases (KMTs). The SET domain was initially identified in three epigenetic modifiers in *Drosophila*: **S**uppressor of variegation 3-9, **E**nhancer of Zeste and **T**rithorax and was named after these founding members. However, Dot1p is a non-SET domain containing protein that methylates H3K79 [59].

A majority of histone lysine methylation marks occur on lysines (K) K4, K9, K27, K36, K79 of H3 and K20 of H4. In addition, these lysine residues can be mono (me1), di (me2) or trimethylated (me3). This results in a wide range of functional diversity for each methylation site. The sites K4, K36 and K79 of H3 are associated with gene activity whereas K9 and K27 are related to transcriptional inactivity.

H3K4 KMTs are conserved from yeast to mammals. In yeast, Set1 is the sole H3K4 methyltransferase that catalyzes deposition of H3K4me1/me2/me3 [60] whereas mammals have six Set1-related proteins: Set1a, Set1b and MLL1, MLL2, MLL3 and MLL4. These enzymes require additional subunits for their catalytic activity and hence exist in the form of a complex with other proteins. The yeast Set1 is contained in a complex called COMPASS (Complex of proteins associated with Set1) along with other subunits [61]. All the mammalian H3K4 KMT complexes have common components: Ash2L, RbBP5, Wdr5 and Dpy30. In addition to these core components, mammalian Set1a/Set1b complexes consist of CFP1 (CXXC1) and Wdr82, MLL1/2 complexes contain MEN1 and MLL3/4 complexes contain NCOA6, PA1, PTIP and UTX [62].

The Set1a and Set1b proteins are the catalytic components of the respective complexes. These proteins exhibit 39% overall sequence identity and 56% similarity. The conserved RRM domain interacts with Wdr82 while the conserved N-SET domain interacts with

CFP1, RbBP5, Ash2 and Wdr5. Set1a/b also show the presence of a divergent central domain which exhibits only 29% sequence identity. A number of roles of Set1a and Set1b have been uncovered recently. Mammalian Set1a and Set1b are responsible for a majority of H3K4me marks [63]. Set1a and Set1b exhibit non-overlapping subnuclear distribution with euchromatin domains, suggesting that these enzymes are targeted at distinct genes [64]. A Set1a-interacting protein, Wdr82, is required for targeting the Set1A-mediated H3K4 trimethylation to transcription start sites by physically associating with the Ser5-phosphorylated C-terminal domain (CTD) of RNA polymerase II[65] [63]. Human Set1a has been implicated in the proliferation of colorectal tumor cells by regulating the Wnt-signaling target genes [66].

Set1b, on the other hand, interacts with the leukemogenic Rbm15-Mkl1 fusion protein associated with acute megakaryoblastic leukemia (AMKL) [67]. Although both Set1a and Set1b have almost identical proteomic profiles, they are essential at distinct stages of murine development [68]. Set1a appears to be important for gastrulation whereas Set1b-lacking embryos, although highly retarded, can gastrulate. Set1a is also required for derivation of induced pluripotent stem cells and for the maintenance of the transcriptional profile of genes governing cell-fate decisions in pluripotent embryonic stem (ES) cells [68]. The euchromatin associated yeast Set1 complex also mediates silencing at telomeres and represses RNA polymerase II activity within the rDNA locus [69, 70].

Mammalian MLL1 and MLL2 *Drosophila* homologs of trithorax (trx) and are involved in positive regulation of homeotic gene expression. MLL1 regulates the expression of the Hox gene, and is implicated in leukemic pathogenesis [71, 72]. However, MLL1 is required for the appropriate H3K4me3 levels of only 5% of the gene promoters that display this mark [73]. These genes include developmentally important Hox genes. MLL1 is required for the regulation of transcription initiation of a subset of Hox genes by RNA polymerase II [73]. MLL2, on the other hand, is essential for H3K4 trimethylation in oocytes [74]. The common component of the MLL1/2 complexes, Menin, is required for the majority of H3K4 methylation at the Hox loci [73]. Mammalian MLL3/4 is homologous to *Drosophila* trithorax-related (Trr) and functions as a major H3K4

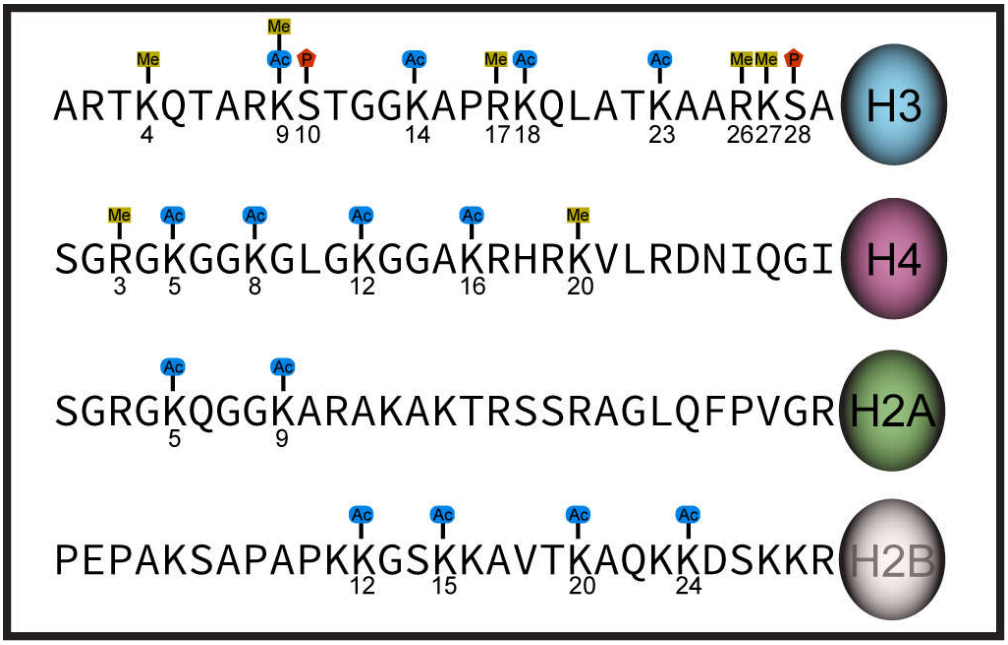


Figure 1: Many posttranslational modifications occur on N-terminal histone tail domains

After the process of translation, histone proteins undergo several modifications at their N-terminal tails. Some of these modifications are activating marks whereas others are repressive marks. The figure represents the four histone proteins that form the nucleosome and the posttranslational modifications (Me: methylation, Ac: Acetylation, Ph: Phosphorylation) occurring on specific amino acids. Figure adapted from http://www.psych.mpg.de/1496016

monomethyltransferase [75]. Other enzymes capable of H3K4 methylation include Smyd1, Smyd2, Smyd3 and Set7/9.

H3K4me3 marks concomitantly occur with active transcription whereas H3K4me2 can occur at poised or inactive genes [76]. H3K4me3 is widely accepted as a euchromatin marker [77, 78]. Monomethylation of H3K4 (H3K4me1) is predominant at the distal enhancer regions [79].

H3K36 methylation is carried out by the Set2 enzyme [80]. Initially, the Set2 enzyme and the H3K6 methyation mark were thought to be associated with transcriptional repression [80] . However, recent reports suggest that H3K36me3 occurs within the body of active genes. Moreover, in Caenorhabditis elegans, exonic regions are enriched with H3K36me3 as compared to intronic segments within the same gene [81]. Set2 is recruited to chromatin via association with RNA polymerase II and is also thought to participate in transcription elongation [82]. It might be possible that the Set2 is involved in gene activation or repression depending on gene context.

As mentioned earlier, methylation of H3K79 is carried out by the non-SET containing enzyme, Dot1 (disruptor of telomeric silencing) [83]. Dot1 preferentially methylates H3K79 in the context of nucleosomes rather than core histones [83]. H2B-K123 and H3K79 lie in close proximity on the same nucleosome surface [84] and it was recently found that monoubiquitination of H2B-K123 was required for H3K79 and H3K4 methylation [85]. Moreover, H3K79me occurs at actively transcribing regions and is also involved in transcriptional elongation [59, 86].

H3K9me3 is usually associated with transcriptionally inactive chromatin. Methylation of H3K9 is catalyzed by several enzymes: SETDB1 (ESET), SUV39H1, SUV39H2, EHMT1 (GLP) and EHMT2 (G9A). SETDB1, an enzyme highly expressed in ES cells, is required for maintaining pluripotency and self-renewal properties of ES cells [87, 88]. SUV39H1 and SUV39H2 govern H3K9-methylation at pericentromeric heterochromatin and are important for mammalian development [89]. On the contrary, EHMT1 and EHMT2 are responsible for deposition of H3K9me3 at euchromatic loci [90].

Unlike the existence of several enzymes that catalyze the deposition of H3K9me3, the repressive H3K27me mark is established by polycomb repressive complex 2 (PRC2), which consists of EZH2, EED, SUZ12 and RbAp46/48 [91]. EZH2 possesses catalytic activity, whereas EED performs an adaptor role by connecting the components [92]. EZH1 is a sequence homolog of EZH2 and forms the non-canonical EZH1-mediated PRC2 which complements the function of the canonical EZH2-mediated PRC2 in the deposition of H3K27me3 marks [93]. H3K27 methylation has been linked to many key processes including genomic imprinting, X-inactivation and silencing of Hox genes [94].

H4K20 methylation is a major and highly conserved mark found on the N-terminal tail on lysine 20 of H4 [95]. Tri-methylation of H4K20 is characteristic of heterochromatic regions, whereas H4K20me1/me2 marks play important roles in DNA replication and DNA repair processes [96, 97]. Monomethylation of H4K20 is performed by Set8/PR-Set7 and further di- and tri-methylation is catalyzed by SUV4-20H1 and SUV4-20H2 [98]. The dynamic control of histone lysine methylation requires the participation of demethylase enzymes that remove methyl groups from lysine residues. The Flavin Adenine Dinucleotide (FAD) requiring Lysine specific demethylase I (LSD1) was one of the first identified demethylase enzymes [99]. LSD1 requires a chromatin-associated transcriptional repressor CoREST and its action specific to H3K4me2 and H3K4me1 [100]. Another major group of histone lysine demethylases known as Jumonji C domain containing histone demethylases (JHDMs) have a characteristic Jumonji C domain [101]. Of these enzymes, JHDM1 specifically demethylates H3K36me2/me1 but H3K36me3 demethylation is catalyzed by JHDM3. On the other hand, H3K9me1/me2 are substrates of JHDM2 whereas H3K9me3 mark is demethylated by JHDM3 [102]. The activating histone mark H3K4me2/me3 is removed by another group of Jumonji C containing enzyme family termed JARID1 [103, 104].

5. Histone code and its complexity

Apart from the histone PTMs described in this study, many others including crotonylation, butyrylation and propionylation of lysine residues, citrullination and ADP-ribosylation of arginine residues, and glycosylation of serine and threonine residues have also been identified [105]. Some of these modifications are associated with active gene

transcription whereas others are related to gene repression. This pattern led to the idea that histone modifications could form a “code” that could be recognized by transcriptional factors and thereby dictate the transcriptional state of the gene. The histone code hypothesis was therefore put forth by Strahl and Allis which states “distinct histone modifications, on one or more tails, act sequentially or in combination to form a 'histone code' that is, read by other proteins to bring about distinct downstream events” [106].

A number of examples of histone crosstalk mechanisms have been described in literature. One of the well-studied crosstalk mechanisms is the ability of H3S10ph to stimulate the Gcn5-mediated acetylation of H3K14 and synergistically mediate gene transcription [52, 107]. In mammalian cells, H3S10ph also inhibits methylation of H3K9, a heterochromatin mark. This is an example of a ‘*cis*’ effect established by the communication of adjacent modifications in the same histone tail. Of higher interest, is the ‘trans’ effect that can occur by cross talk between modifications present on different histones either within a single nucleosome or across several nucleosomes [17]. For instance, monoubiquitination of H2B is required for methylation of H3K4 by the COMPASS complex [108]. Methylation of H4R3 by PRMT1 is inhibited by acetylation of H4 at K5, K8, K12 and K16 [17, 109]. Trimethylation of H4K20 at constitutive heterochromatin requires prior deposition of H3K9me3 [97].

Many histone crosstalk mechanisms are mediated by effector proteins that target PTMs. Specific recognition of PTMs by effector proteins is required for the recruitment of various signaling components, including transcription factors, to chromatin, for the function of key processes like gene regulation, DNA replication, DNA repair, formation of heterochromatin and chromatin remodeling. The catalytic action of chromatin modifying enzymes usually requires additional subunits. One or more of these subunits can specifically recognize distinct PTMs and mediate downstream signaling events. Soon after the histone code hypothesis was suggested, chromatin signaling components were classified into groups such as ‘writers’, ‘erasers’ and readers’ [110, 111]. Writers and erasers are proteins involved in deposition and removal of histone modifications respectively, and have been discussed in detail in earlier sections. Readers, on the other

hand, recognize/read these covalent signals specifically and relay the information to downstream players of the chromatin signaling pathway to establish distinct chromatin states.

A lot of interest has been generated recently in the study of reader modules because of the remarkable specificity and dynamic nature of the interaction of these domains with PTMs. These reader modules are known to provide an accessible surface for binding the incoming modified histone tail after determining the kind of modification (acetylation/methylation) and specificity of state (unmodified/mono/di or tri-methylation of lysine) [112]. One of the first identified readers is the bromodomain (BRD). BRDs are found in many chromatin remodeling complexes including SAGA, SWI/SNF, RSC, p300/CBP and TFIID. BRDs, in general, are thought to bind acetylated lysine residues and increase the affinity of these enzyme complexes to this mark [113]. BRD of Gcn5 that can read the monoacetylation marks on histone lysine residues [114]. In this case, Gcn5 can also catalyze acetylation of these residues, suggesting that this communication might be responsible for spreading of acetylation marks in chromatin. Taf1, the largest subunit of TFIID, consists of two BRDs with which it binds multiple acetylated lysines on the H4 tail [115]. Phosphorylated serine residue on H2AX (γH2A) is read by the BRCT domain of mediator of DNA damage-checkpoint1 (MDC1) and this regulates the repair of DNA double-strand breaks [45]. Moreover, mammalian 14-3-3 also recognizes H3S10ph and is recruited to c-fos and c-jun nucleosomes upon gene activation [116].

a. Readers of methylated and unmodified histone H3

Histone H3 can be methylated at either the arginine or lysine positions. The symmetric H4R3me2s mark is recognized by the ADD (ATRX, DNMT3, DNMT3L) domain of DNMT3a thereby linking this modification to gene inactivation [117]. In contrast, the Tudor domain of TDRD3 (Tudor domain containing 3) interacts with H3R17me2a H4R3me2a. TDRD3 functions as a transcriptional coactivator and links these modifications to gene activity [118].

Readers of methylated lysine residues on histone H3 are particularly well-studied. The N-terminal chromodomain (CHD) of HP1 (Heterochromatin associated protein1) binds to the H3K9 methyltransferase SUV39H1 and is found to read methylated H3K9, the product of this enzyme complex [119]. It was found that methylation of H3K9 results in creation of a binding site for interaction with HP1, which is suggestive of a role for this interaction in the propagation of heterochromatin [120]. Proteins that read methylated H3K4 and unmodified H3 show the presence of a characteristic plant homeodomain (PHD). However, PHDs are also known to interact with acetylated lysine residues. [121, 122].

The PHD finger consists of 50-80 aa residues and exhibits a zinc-binding motif [123]. This relatively small domain, consists of a two strand anti-parallel β-sheet and a C-terminal α-helix (present in few PHDs) that is stabilized by two zinc atoms scaffolded by the signature Cys4-His-Cys3 motif in a cross-brace topology [124]. The PHD of BPTF (Bromodomain and PHD domain transcription factor) was one of the first identified proteins within the PHD family proteins to read the histone code. BPTF is the largest component of the ATP-dependent chromatin remodeling NURF (Nucleosome remodeling factor) complex and is associated with actively transcribed genes [125]. The bromodomain-proximal PHD specifically recognizes H3K4me3 to stabilize the NURF complex at chromatin and improves the efficiency of NURF-catalyzed nucleosome sliding and activation of developmentally important gene promoters [126]. The ING family of tumor suppressors contains ING2 that plays a pivotal role in the regulation of cell replication, DNA repair and apoptosis [127]. ING2 forms a part of the repressive mSin3A/HDAC1 complex. Upon induction of DNA damage, ING2 associates with H3K4me2/3, recruits the mSin3A/HDAC1 complex to chromatin thereby negatively regulating gene transcription [128]. Pygopus 2 (Pygo2) is a downstream component of the canonical Wnt pathway in *Drosophila*. Pygo2 is required for the proper development of various tissues [129]. Pygo2 consists of a PHD finger that associates with di and trimethylated forms of H3K4 and also facilitates H3K4 methylation by associating with Wdr5, an essential and common component of the SET-domain containing histone methylase complexes [130]. PHF2, a histone demethylase enzyme belonging to the JHDM family, consists of a PHD domain and a Jumonji C-domain. The PHD finger of PHF2 specifically recognizes H3K4me3, and this is essential for PHF2-dependent

promoter H3K9-demethylation and expression of ribosomal DNA (rDNA) [131]. TAF2 associates with H3K4me3-enriched nucleosomes via its PHD finger and anchors the transcription factor TFIID to chromatin for gene activation [132]. Recombination-activating gene 2 (RAG2) encodes for the catalytic enzyme that plays a critical role in V(D)J recombination, a process that involves the assembly of antigen-receptor genes. RAG2 protein consists of a PHD finger that binds H3K4me3 at actively arranging gene segments and is necessary for efficient recombination to occur [133, 134].

A separate subset of PHD fingers recognizes unmodified H3. This suggests that unmodified H3 can act as distinct histone marks and be recognized by reader modules. The PHD fingers in this group lack an aromatic cage, a characteristic of H3K4me binders [124]. It is frequently observed that the interaction between these PHD fingers and unmodified H3 is ablated by methylation of H3K4 [124]. BHC80 is a component of the lysine-specific histone demethylase (LSD1) that removes H3K4me1/2 marks. BHC80 associates with unmodified H3, the product of LSD1-mediated enzyme action, and is essential for LSD1-mediated gene repression [135]. Auto immune regulator (AIRE) consists of two PHD domains (PHD1 and PHD2). The first PHD (PHD1) interacts with unmodified H3 and this binding is critical for appropriate expression of tissue- specific antigens in the thymus [136]. DNMT3L stimulates the DNA-methyltransferase activity of DNMT3a to achieve a repressive chromatin state. DNMT3L recognizes unmodified H3 and induces de-novo methylation by recruitment of DNMT3A [137].

6. PHD fingers in human diseases

PHD finger-containing proteins are associated with numerous human diseases including cancers and immunodeficiency disorders.

Loss or mutation of the ING family members, attributed with tumor suppressor roles, have been reported in a number of human cancers including esophageal squamous cell carcinoma, gastric cancer and breast cancer [138, 139]. The exact mechanism of occurrence of these cancers is unknown, however ING family of proteins are linked to critical functions such as cell cycle control, DNA repair, apoptosis and senescence [127].

APECED (Autoimmune Polyendocrinopathy Candidiasis Ectodermal Dystrophy) is an autosomal recessive disease associated with a number of autoimmune abnormalities including diabetes, thyroid disease and failure of adrenocorticoids and gonads [140]. Some of the APECED-linked mutations are found within the PHD domains of AIRE. Missense mutations in PHD1 have been found to destabilize the structure of AIRE-PHD1, resulting in loss of binding with unmodified H3 and in turn affecting gene transcription [136, 141]. Mutations in the Rag2 PHD finger have been found in patients with severe combined immunodeficiency (SCID) syndrome [142] . These patients completely lack B and T lymphocytes and suffer from severe immunological abnormalities [142, 143]. Mutations disrupting the aromatic cage of RAG2 PHD inhibit its binding to H3K4me3 and pathologically result in a milder form of SCID called Omenn's Syndrome. This occurs chiefly because of an impairment in V(D)J recombination [144, 145].

The involvement of PHD finger-containing proteins in a spectrum of human diseases, makes the idea of using PHDs as drug targets, attractive. Research in the future should be aimed at developing gene-therapeutic approaches to re-establish the disrupted gene regulation pathways and thereby treating the underlying diseases that occur due to aberrant PHD domain function.

7. Role of chromatin in embryonic development and ES cell differentiation

Embryonic development involves the generation of a multitude of different cell types each with distinct gene-expression patterns. Each of these cells originates from a single totipotent cell and retains its genetic information. Such a wide disparity between genetic material and phenotypic features can be attributed to epigenetic changes occurring during the process of embryonic development. Several developmental defects are observed when epigenetic modifications are blocked. For instance, targeted disruption of DNMT1 (maintenance DNA methylation) or DNMT3a and DNMT3b (de novo DNA methylation) results in embryonic lethality [10, 146]. Also, depletion of G9a, an enzyme responsible for H3K9me, results in early embryonic lethality [90].

DNA methylation plays a significant role in the process of genomic imprinting. It is an epigenetic mechanism in which genes are expressed in a parent-of-origin specific manner.

In mice, maternal and paternal chromosomes undergo a wave of demethylation after zygote formation [147, 148]. This occurs as a result of a passive mechanism that removes most, but not all of the methylation marks inherited from the gametes and is thought to be completed by the blastocyst stage [147, 148]. However, maternal and paternal imprints are preserved. Further, after implantation of the embryo, de novo methylation establishes embryonic DNA methylation patterns beginning from the inner cell mass of the blastocyst [147-149].

Dosage compensation in mammalian cells occurs by the silencing of one X-chromosome in female cells. Silencing of the X-chromosome is initiated early in the developmental process and involves the coating of the X-chromosome with the X-inactive specific transcript (*Xist*). This process requires the regulated action of several repressive changes including DNA methylation, inactivating histone modifications and recruitment of the polycomb group proteins to establish stable repression of a single X-chromosome [150, 151].

Embryonic stem cell differentiation is a very useful model to study chromatin changes occurring during this process. Embryonic stem cells are obtained from the inner cell mass of the embryo at the blastocyst stage. These cells can be cultured indefinitely *in vitro* and can be induced to differentiate through the formation of non-adherent spheroidal cell aggregates known as embryoid bodies. ES cells are often maintained in the presence of Leukemia Inhibitory Factor (LIF) to prevent spontaneous differentiation in culture. Removal of LIF from ES cell growth media pushes the cells towards differentiation. However, these cells do retain their proliferative potential [152].

ES cells exhibit an open chromatin configuration. Undifferentiated human ES cells possess less heterochromatin than differentiated ES cells as observed by electron microscopy [153]. Major architectural proteins such as HP1 and linker histone H1 display hyperdynamic plasticity in binding chromatin in these cells [154]. Undifferentiated ES cells exhibit widespread active chromatin domains consisting of activating histone marks including H3K4me and H3K9ac; and hypomethylated DNA.[4, 155-157].

Table 1: Epigenetic changes resulting in transcriptionally active chromatin

Epigenetic modification	Writers	Readers
Histone acetylation	HATs: GNAT, p300/CBP, MYST families, TAF_I	Bromodomain
H2B ubiquitination	HR6A, HR6B	
H3S10, H3S28 phosphorylation	Kinases	**H3S10ph:** 14-3-3
Arginine methylation	**H4R3me2a:** PRMT1 **H3R17me2a, H3R26me2a:** PRMT4	**H4R3me2a, H3R17me2a:** TDRD3
Histone methylation	**H3K4me:** Set1a, Set1b, MLL1-4 Smyd1-3, Set7/9 **H3K36me:** Set2 **H3K79me:** Dot1	**H3K4me3:** BPTF, ING2, Pygo2, Wdr5, PHF2, TAF2, RAG2
Histone demethylation	**H3K9me1/me2:** JHDM2 **H3K9me3:** JHDM3	

Table 2: Epigenetic changes resulting in transcriptionally inactive chromatin

Epigenetic modification	Writers	Readers
DNA methylation	DNMT1. DNMT3a, DNMT3b	MBDs
Histone deacetylation	HDACs: Class I-IV	
H2A ubiquitination	Ring 1B , 2A-HUB	
Histone methylation	**H3K9me:** SETDB1 (ESET), SUV39H1, SUV39H2, EHMT1 (GLP), EHMT2 (G9A) **H3K27me:** PRC2 **H4K20me:** Set8/PR-Set7 SUV4-20H1 and SUV4-20H2	**H3K9me:** HP1
Histone demethylation	**H3K4me1/me2:** LSD1 **H3K4me2/me3:** JARID1 **H3K36me1/me2:** JHDM1 **H3K36me3:** JHDM3	

ES cell differentiation is a highly dynamic event accompanied by global changes in the chromatin architecture resulting in a gradual change from an open and active to a more compacted and inactive state. Heterochromatin domains become more abundant, which is supported by the observation that well-differentiated ES cells exhibit an overall increase in repressive chromatin modifications such as H3K9me2 and H3K37me3 [153, 158, 159]. Moreover, HDAC activity, associated with gene repression, is required for embryonic stem cell differentiation [159].

8. CXXC finger protein 1 (CFP1)

Human CFP1, encoded by the CXXC1 gene, is a component of the euchromatin-specific Set1a and Set1b methyltransferase complexes that deposit H3K4me marks [64, 160, 161]. CFP1 homologues are found in other species such as *S. cerevisiae, S. pombe, D. melanogaster, C.elegans, X. laevis* and *M. musculus.*

CFP1 (656aa) is transcriptional activator that is ubiquitously expressed in all tissue-types [160]. The domain structure of CFP1 consists of a highly conserved, cysteine rich, CXXC domain that specifically binds unmethylated CpGs [160]. The consensus DNA sequence for CFP1 binding is (A/C)CpG(A/C) and methylation of cytosine residues within CpGs abolishes CFP1 interaction [162]. It was found that unmethylated CpG are sufficient to recruit CFP1 and thereby establish domains of H3K4me3 [163]. CXXC domains are also found in other chromatin-associating proteins including DNMT1, MBD1 and MLL2 [164-166]. CXXC domain is not conserved in *S. cerevisiae, S. pombe* and *C.elegans*, and these organisms lack cytosine methylation, suggesting that CXXC domain may play a role in cytosine methylation.

CFP1 also consists of a 47 aa long, N-terminal PHD domain that shows the presence of the conserved Cys4-His-Cys3 motif. A proteome-wide screen of proteins containing PHD fingers in S. cerevisiae established that the yeast CFP1 homolog, Spp1 interacts with H3K3me2/me3 [167]. It is currently unknown if human CFP1 PHD interacts with histones.

Moreover, CFP1 also consists of a cysteine rich Set1 interaction domain (SID) that mediates interaction of CFP1 with the Set1a and Set1b complexes [168]. The acidic domain is rich in aspartate and glutamate residues and is important for transcriptional activation [169]. On the contrary, the basic domain is enriched with lysine and arginine residues. Coiled coil domains facilitate homo- or hetero-dimerization of proteins and are frequently found in DNA binding proteins [160]. The precise function of the basic and coiled-coil domains in CFP1 is unknown. HEAT (Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1) domain is a solenoid protein domain known to facilitate protein-protein interactions in other proteins (Figure 2) [170, 171].

CFP1 is essential for vertebrate development as targeted disruption of *CXXC1* gene results in peri-implantation lethality (~4.5-6.5 days post coitum) in murine embryos [172]. This correlates temporally with a global change in DNA methylation patterns [149]. Antisense targeting of CFP1 in zebrafish embryos resultd in embryo runting, cardiac edema and a failure of primitive hematopoiesis [173]. It was observed that shRNA-mediated knockdown of CFP1 inhibits the growth, differentiation and viability of PLB-985 human myeloid cell line [174]. Due to early death of CXXC1$^{-/-}$ embryos, studies on the function of CFP1 during mammalian post-gastrulation development and homeostasis could not be performed. However, this was made possible by developing mice that are homozygous for the conditional CXXC1 allele and carry the Mx1-Cre recombinase transgene. On induction of Mx1-Cre recombinase with poly-IC, deletion of CFP1 occurred in all hematopoietic cell lineages and the liver [175]. Depletion of CFP1 in adult mice results in death within 9-13 days and reduced bone-marrow cellularity and mature peripheral blood cells. However, the Lin^{-}Sca-1^{+}c-Kit^{+} (LSK) population of cells that is enriched for hematopoietic stem cells and multipotential progenitor cells in the bone marrow remain unaffected and continue to proliferate in the absence of CFP1 during the two week time period. However, no pathology was observed in the quiescent liver tissue during this time period [176]. As CFP1 is ubiquitously expressed in all adult tissues including liver, CFP1 may be important for long-term liver homeostasis, an assessment of which is limited in this model system due to the short time interval (2 weeks) between gene deletion and death

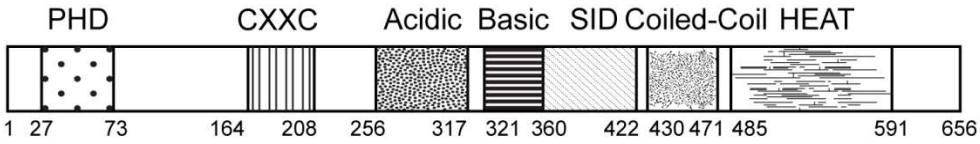


Figure 2: Domain structure of CFP1

This schematic representation depicts the various conserved domains of CFP1. The amino acid position of each domain within the primary sequence is denoted under the figure.

due to hematopoietic failure. These studies collectively suggest that CFP1 plays a critical role in embryonic as well as post-gastrulation development.

To obtain further insight into CFP1 function, ES cells were isolated from the inner cell mass of blastocysts lacking the CXXC1 gene ($CXXC1^{-/-}$) and heterozygous blastocysts ($CXXC1^{+/-}$) [177]. Heterozygous $CXXC1^{+/-}$ ES cells display a phenotype similar to wild type ($CXXC1^{+/+}$) ES cells. However, $CXXC1^{-/-}$ ES cells display a wide variety of defects. These cells have increased population doubling time due to a 3-fold increase in apoptosis but show normal cell-cycle distribution [177]. $CXXC1^{-/-}$ ES cells fail to differentiate upon removal of LIF from the growth medium. They form embryoid bodies but are unable to form the characteristic endodermal outgrowths and fail to induce lineage and developmental markers and fail to down-regulate the expression of markers of stem cells, Oct4 and alkaline phosphatase. A 60-80% defect in global cytosine methylation levels is observed in $CXXC1^{-/-}$ ES cells [177]. These cells exhibit a ~30% reduction in DNA methyltransferase activity towards a hemi-methylated oligonucleotide substrate but normal activity towards an unmethylated substrate. This suggests that $CXXC1^{-/-}$ ES cells have a deficiency in maintenance methyltransferase activity but normal *de novo* methylation activity [177]. $CXXC1^{-/-}$ ES cells also displayed a 15% decrease in global protein synthesis [178]. ES cells lacking CFP1 display altered histone methylation levels. A global increase in H3K4me2 levels (~43%) and reduction in H3K9me2 levels are observed in $CXXC1^{-/-}$ cells under normal conditions [161]. On induction of differentiation, wild type ES cells showed a gradual decline in H3K4me2 levels. However ES cells lacking CFP1 exhibit a steady 4-fold increase in H3K4me3, six days after induction of differentiation by removal of LIF [159]. The observation of increased H3K4 methylation levels in $CXXC1^{-/-}$ ES cells suggests that CFP1 restricts the activity of the Set1 H3K4 methyltransferase complexes. This was confirmed by confocal immunofluorescence. Wild type ES cells show almost no overlap between DAPI-bright (heterochromatin) regions and endogenous euchromatin-specific Set1a. However, in the absence of CFP1, ~25% of Set1a is detected inappropriately within DAPI-bright heterochromatin. Stable re-expression of CFP1 into $CXXC1^{-/-}$ ES cells results in the appropriate exclusion of Set1a from DAPI-bright heterochromatin. Similar results were obtained for H3K4me3, a product of the Set1 complex [179].

Genome-wide ChIP-seq analysis indicates that CFP1 was found along with H3K4me3 at ~80% of CGIs in mouse brain and siRNA-mediated knockdown of CFP1 in NIH-3T3 cells results in a drop of H3K4me3 levels at these sites [163]. This suggests that CFP1 may tether the Set1 complex to CGIs by binding to unmethylated CpGs. In the absence of CFP1 and the resultant lack of tethering function, Set1 may diffuse to inappropriate genomic locations.

The defects found in CXXC1$^{-/-}$ ES cells can be rescued by stable expression of CFP1 in CXXC1$^{-/-}$ ES cells. This feature was used to probe the structure-function relationships of CFP1. These studies demonstrated that expression of either the amino terminal half (1-367aa) or the carboxy terminal half (361-656aa) of CFP1 was sufficient to rescue defects in global cytosine methylation levels and *in vitro* differentiation. Introduction of a point mutation (C169A) within the amino terminal half (1-367aa) ablates the DNA binding function and a point mutation (C375A) within the carboxy-terminal half inhibits binding of CFP1 with the Set1 enzyme [162, 168]. Introduction of the DNA binding mutation within 1-367 aa or the Set1 interaction mutation within 361-656 aa of CFP1, fails to rescue cytosine methylation and *in vitro* differentiation defects. Introduction of either of these mutations within full-length CFP1 cDNA had no adverse effect on rescue activity. This suggests that neither of these domains is required for rescue activity. However, introduction of both of these mutations within full length CFP1 cDNA results in loss of rescue function. Therefore, retention of either the CXXC or the SID domain of CFP1 is required for the rescue of cytosine methylation and *in vitro* differentiation [180]. However, appropriate subnuclear localization of Set1 and H3K4me3 to euchromatin requires both the CXXC and SID domains of CFP1 [179].

Genome wide analysis of H3K4me3 levels in CXXC1$^{-/-}$ ES cells revealed that this chromatin mark was reduced at approximately one-half of the CGI-rich gene promoters. Moreover, absence of CFP1 also resulted in ectopic H3K4me3 peaks at regulatory regions devoid of CGIs or transcription start sites. Both of these defects could be rescued by the stable expression of full length CFP1 in CXXC1$^{-/-}$ ES cells [181]. Another interesting observation was that the DNA binding mutant form of full-length CFP1 could rescue appropriate H3K4me3 levels at most CGIs. However, this mutant could not rescue the

inappropriate occurrence of ectopic H3K4me3 peaks outside of CGIs. These results suggest that DNA binding domain is not necessary for the targeting of H3K4me3 to CGIs and that some other domain may perform this function [181]. As mentioned earlier, the yeast CFP1 homolog Spp1 binds H3K4me2/me3 marks on chromatin via its PHD finger. It is possible that human CFP1 PHD binds histones in a similar way and regulates Set1 targeting.

9. Liver homeostasis and regeneration

The mammalian liver is a vital organ required for several important metabolic functions. It receives a majority of blood from the small and large intestine through the portal vein and functions to detoxify harmful chemicals. The liver is responsible for the storage of carbohydrates in the form of glycogen and processes nutrients absorbed by the intestines and converts them to nutrient forms that can be used by the body. The liver produces secreted proteins like albumin and coagulation factors. The liver also synthesizes bile for emulsification and absorption of fat [182]. Moreover, the circadian clock system in the liver plays a critical role in regulation of liver function and homeostasis [183].

In mice, embryonic liver originates from the endoderm germ layer that is established during gastrulation. The initial morphological sign of the embryonic liver is observed adjacent to the developing heart at embryonic day of gestation 9 (e9.0). At e9.5, hepatic endoderm cells, known as hepatoblasts, are involved in formation of the liver bud. Between days e10-e15, the liver bud undergoes extensive growth, formation of blood vessels and colonization by hematopoietic cells making it the chief hematopoietic organ at this stage. Hepatoblasts have dual-potential, in that they can either result in the formation of biliary epithelial cells that line the intra-hepatic bile ducts or form mature functional hepatocytes. This gradual process begins at e13 and continues until after birth to ultimately result in a characteristic hepatic architecture to the liver [184, 185].

In the mature liver, a majority of functions are performed by the parenchymal cells or hepatocytes that constitute ~80% of the liver mass. Fully differentiated hepatocytes are quiescent and rarely undergo cell division in an adult animal. Therefore, active chromatin remodeling is unlikely to occur in adult liver tissue. The remaining 20% of the hepatic

cells are non-parenchymal and include endothelial cells that line hepatic blood vessels, lymphocytes that provide innate immunity against infection, Kupffer cells that phagocytose foreign particles and produce cytokines and stellate cells that produce factors leading to hepatic fibrosis in response to injury [186].

The most distinctive feature of the liver is its ability to regenerate in response to chemical or physical injuries. This property of the liver is widely studied in rodents by performing a surgical procedure called 'partial hepatectomy' that involves the removal of two-thirds of the liver [187]. In response to this injury, the residual hepatocytes re-enter the cell cycle and undergo one round of DNA synthesis which peaks at ~36 hours in the murine liver. A second, smaller wave of DNA synthesis occurs to re-establish the original number of hepatocytes. This is followed by a wave of apoptosis to precisely regulate the regenerative response [182, 188]. This process is thought to be completed in 7 days after surgery in mice. Thus, the process of liver regeneration is thought to be purely mitotic, without the involvement of liver-stem or progenitor cells. It is a compensatory mechanism to replenish the lost mass to restore proper functioning of the liver. Once the original mass of the liver is restored, the mitotic response of the remnant hepatocytes comes to a halt [186]. It is very likely that the dynamic process of liver regeneration is accompanied by active changes in chromatin structure to facilitate regulation of liver regeneration.

A number of growth factors and cytokines participate in the regulation of liver regeneration. For instance, growth factors that are required include insulin, glucagons, hepatocyte growth factor (HGF), epidermal growth factor (EGF) and transforming growth factors (TGF) and important cytokines include tumor necrosis factor-α (TNF-α) and interleukin-6 (IL6) [186, 189]. Besides these, regulation of liver regeneration also requires the interplay of a number of transcription factors. Furthermore, optimal telomerase activity is required for efficient liver regenerative response as cells with shortened telomeres are unable to replicate efficiently [190].

Recently, there has been a lot of focus on the expansion of the murine liver during pregnancy. Pregnancy results in global changes to the size and function of many organs including the brain, pancreas and liver [191-193]. The maternal liver is found to almost

double in size by the 18th day of pregnancy [193]. This substantial increase in liver size is a result of increased mitosis and hypertrophy of hepatocytes [193]. Pregnancy-dependent hepatomegaly closely resembles liver growth occurring in the regenerating liver. This is supported by the observation that a subset of genes regulating liver regeneration is activated during pregnancy-induced liver growth [193].

10. Role of chromatin in liver homeostasis and regeneration

Although there are numerous advances in the understanding of liver development, homeostasis and regeneration, studies on the capacity of chromatin to influence these processes are lacking. Few recent reports linking chromatin and liver homeostasis or regeneration are summarized below.

The circadian rhythm exists in the liver for the regulation of liver function and homeostasis. It was found that HDAC3 recruitment in the liver is required for the circadian rhythm. The circadian nuclear receptor Rev-erbα colocalizes with HDAC3 at genes governing lipid metabolism. Furthermore, depletion of HDAC3 results in hepatic steatosis [194].

Hepatic stellate cells are recognized for their role in fibrosis in response to injury. The transdifferentiation of these into cells similar to myofibroblasts is an essential process in wound healing and fibrosis. This process requires the recruitment of MeCP2 to methyl-CpG marks on genomic DNA resulting in silencing of essential genes. This silencing is required for hepatic stellate cells to acquire their fibrogenic phenotype [195-197].

EZH1 and EZH2 are homologous proteins that form a part of PRC2 complex and possess H3K27 methyltransferase activity. EZH1 and EZH2 redundantly maintain liver homeostasis and regulate liver regeneration [198]. In the absence of EZH1 and EZH2, mutant mice exhibit regenerative nodules and fibrosis accompanied with a depletion of H3K27me3 at genes governing cell survival, proliferation and fibrosis. These mice also displayed an impaired regenerative response to partial hepatectomy [198].

Focus of Dissertation

The work described in this dissertation focuses on understanding the significance of CFP1 in embryonic stem cell function and in the homeostasis and regeneration of the mammalian liver.

The first part of this work describes the interaction between CFP1 PHD and methylated H3K4. Critical residues required for this interaction were mutagenized and the functional significance of this interaction was assessed by expressing these CFP1 mutations in CXXC1$^{-/-}$ ES cells and assaying for rescue of defects that occur in the absence of CFP1. Subnuclear localization of endogenous H3K4me3 was studied by confocal immunofluorescence whereas the ability of CXXC1$^{-/-}$ ES cells expressing CFP1 mutations to differentiate *in vitro*, was assessed by morphological characteristics, alkaline phosphatase activity and RT-PCR analysis of stem cell and lineage-specific markers.
The second part of this work describes the interaction between CFP1 PHD and unmodified H3. Several point mutations were made within CFP1 PHD to try and ablate its interaction with unmodified H3 to determine the functional significance of this interaction. Further research is required to precisely identify amino acid residues that are critical for this binding event, mutagenize them and investigate the functional role of this binding event.

The third part of this work examines the importance of CFP1 in murine liver homeostasis and regeneration. This is the first study that suggests a role for CFP1 in long-term tissue homeostasis in an adult animal. A mouse line carrying the conditional CXXC1 allele and Cre-recombinase gene under the control of the liver-specific albumin promoter was established and their phenotypes were studied over an extended period of time. To understand the role of CFP1 in liver regeneration, CFP1-deficient mice were subjected to partial hepatectomy (surgical removal of 2/3 liver) and their liver growth response was assessed by comparing residual liver masses one week after the surgery.
The goal of this dissertation is to enhance the understanding of the functional role of the epigenetic regulator CFP1 in ES cells and to determine the significance of CFP1 in the post-development homeostasis of quiescent liver tissue and in the regenerating liver.

Gaining insight into these roles of CFP1 will further the knowledge of the regulation of epigenetic machinery required for these processes.

Methods

1. Cloning of plasmids

The plasmid encoding GST-tagged Pygo2 PHD was generously provided by Dr. Xing Dai, University of California-Irvine, CA and the plasmid encoding GST-tagged HP1α chromodomain was obtained from a non-profit, global plasmid repository called Addgene.

cDNA, encoding 1-656 amino acids and 1-102 amino acids of CFP1, was cloned into pcDNA 3.1/Hygro mammalian expression vector (Invitrogen, Carlsbad, CA) and pGEX-4T1 (GE Healthcare, United Kingdom) respectively by sequential restriction digests by Dr. Jeong-Heon Lee. The plasmid maps and cloning strategies used for pcDNA 3.1/Hygro and pGEX-4T1 plasmids are schematically represented in Figures 3 and 4 respectively.

pGEX-4T1 plasmids containing C-terminal and N-terminal truncations of CFP1 PHD domain were generated as follows. First, desired regions in CFP1 cDNA sequence were subjected to PCR amplification using specific oligonucleotide primers (Integrated DNA Technologies, San Jose, CA) (Table 3) using the parent pGEX-4T1 plasmid containing cDNA sequence for CFP1 (1-102 aa) as the template. These Taq-amplified PCR products (containing 3' A-overhangs) were directly ligated to the pCR2.1®-TOPO® TA vector provided in the TOPO® TA cloning kit (Invitrogen) as per manufacturer's instructions, followed by restriction digestion with XbaI and EcoRI (New England Biolabs, Ipswich, MA). The cDNA sequence encoding 1-102 aa of CFP1 was released from the parent plasmid by digestion using restriction enzymes XbaI and EcoRI to generate the empty vector. This was followed by ligation of the digested insert to the digested empty pGEX-4T1 vector using T4 DNA ligase (New England Biolabs). Sequences of these plasmids were verified by automated sequencing carried out at the Indiana University Biochemistry and Biotechnology Facility (IU BBF).

2. Site directed mutagenesis

Single, double or triple amino acid substitutions or deletions of certain amino acid residues were performed by site-directed mutagenesis using the Quick Change II Site-

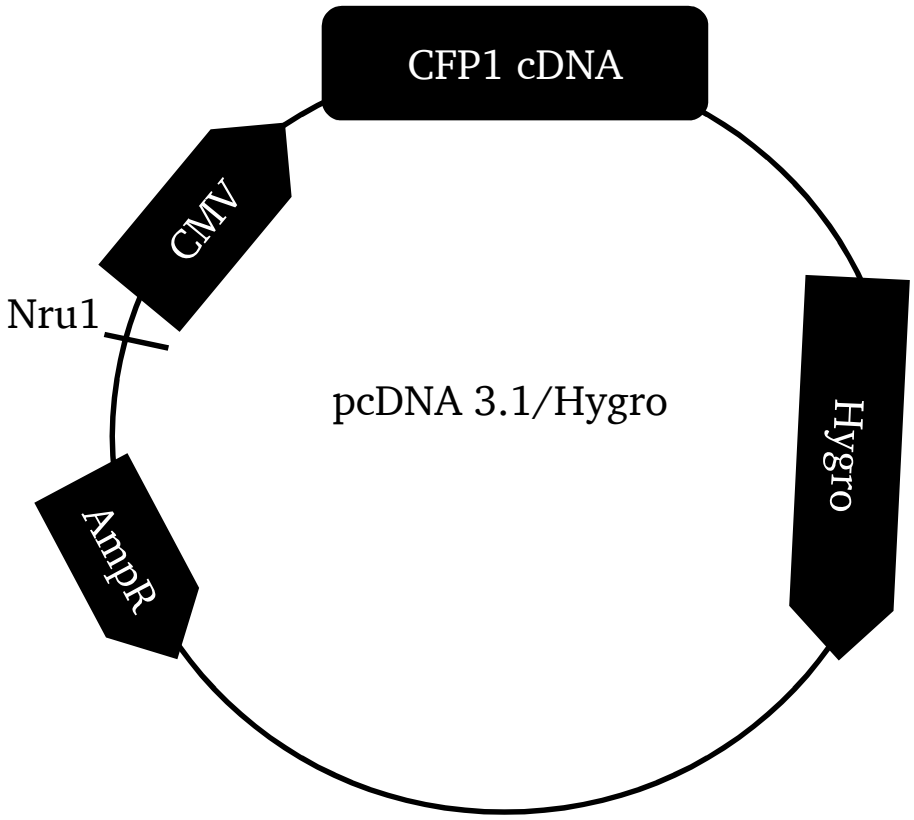


Figure 3: Cloning of a construct expressing FLAG-tagged CFP1 (1-656) protein. The human CFP1 cDNA (1968 bp) was ligated in the multiple cloning site in-frame between EcoRI and XbaI sites, downstream of the FLAG epitope cDNA sequence within the pcDNA 3.1 Hygro vector (5.6 kb) by Dr. Jeong-Heon Lee.

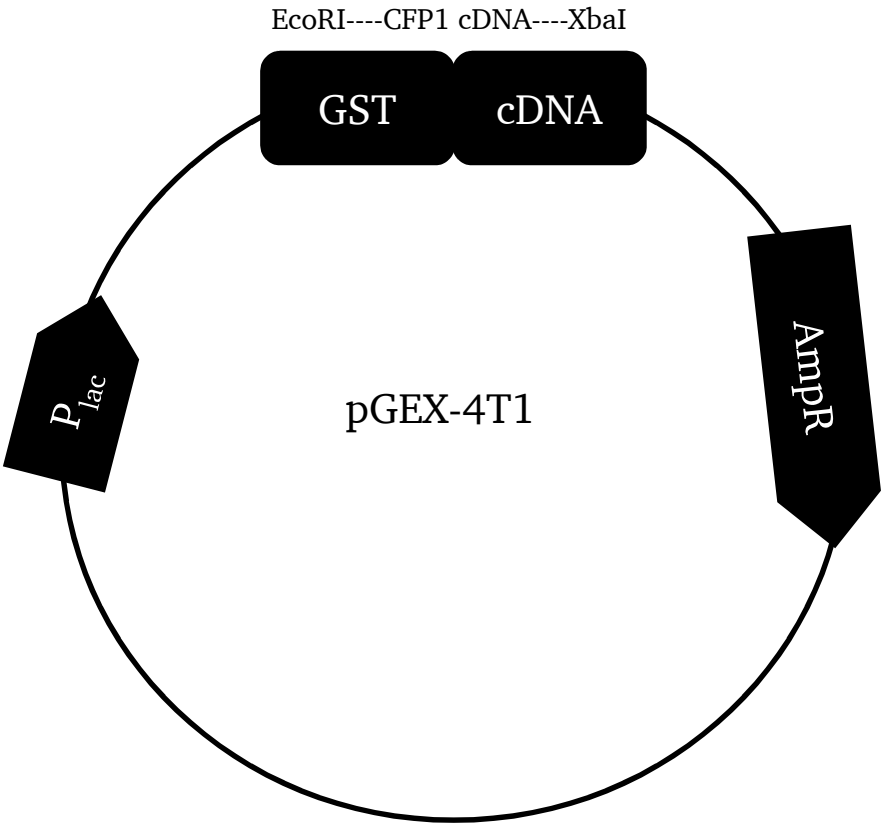


Figure 4: Cloning of constructs expressing recombinant GST-tagged CFP1 PHD
The human CFP1 cDNA encoding the CFP1 PHD domain (1-102) or its truncations 1-29, 1-49, 1-73, 50-102, 1-55, 1-60, 1-65, 1-70 used for the expression of recombinant GST proteins, were ligated into the pGEX-4T1 vector (4.9 kb) between EcoRI and XbaI.

Table 3: List of oligonucleotide primers used to generate CFP1 PHD truncations

Primers designed to generate CFP1 PHD truncations contained restriction site linkers for cloning of these constructs into pGEX-4T1. Forward and reverse primers are indicated by F and R/ Rev respectively. Numbers in the primer name suggest the location of the primer in terms of amino acids in CFP1. The EcoRI (GAATTC) restriction site sequence and the XbaI sequence (TCTAGA) was added to forward and reverse primers respectively (shown in bold).

1. CFP1 F: 5'-GTAC**GAATTC**ATGGAGGGAGATGGTTCAGACCCAGAG-3'
2. CFP1 Rev 29 aa: 5'-CATG**TCTAGA**GCAGTAGATGGGCGCATTCTCCCCATTC-3'
3. CFP1 Rev 49 aa: 5'-CATG**TCTAGA**CCACTCATTGCAGTTGTCACACCCG-3'
4. CFP1 Rev 73 aa: 5'-CATG**TCTAGA**GCACTCCCGACAGTACCACTCCCGG-3'
5. CFP1 50 to 102 F: 5'-GTAC**GAATTC**ATGTTCCATGGGGACTGCATCCGGATC-3'
6. CFP1 50 to 102 R: 5'-CATG**TCTAGA**GGGCTCACTGCTGTCCCGCTC-3'
7. CFP1 55 Rev: 5'-CATG**TCTAGA**GATGCAGTCCCCATGGAACC-3'
8. CFP1 60 Rev: 5'-CATG**TCTAGA**CTTCTCAGTGATCCGGATGCAG-3'
9. CFP1 65 Rev: 5'-CATG**TCTAGA**GATGGCCTTGGCCATCTTCTC-3'
10. CFP1 70 Rev: 5'-CATG**TCTAGA**ACAGTACCACTCCCGGATGG-3'

Directed Mutagenesis kit (Agilent, Santa Clara, CA) as per manufacturer's instructions. Fifty microliter reactions containing 50 ng of the template plasmid DNA, 125 ng of forward and reverse primers, 1 μl of 2'-deoxynucleoside 5'-triphosphate (dNTP) mix, 5 μl of 10X reaction buffer, 2.5 units of PfuUltra high fidelity DNA polymerase was subjected to thermocycling: 95°C for 30s followed by 18 cycles of 95°C for 30s, 55°C for 1 min, 68°C for 7 min. This was followed by digestion of the parent plasmid by 1 μl of DpnI for 1 hour at 37°C. Further, XL-1 Blue super competent cells were transformed with 1 μl of DpnI–treated DNA. Transformed bacteria were spread on LB agar plates containing 100 μg/ml of Ampicillin and grown overnight at 37°C. Individual clones were selected, grown in LB overnight at 37°C with shaking and these cultures were subjected to plasmid mini-preparations. The sequence of each plasmid was confirmed for integrity and the presence of the expected mutation by IU BBF or Eurofin MWG Operon sequencing services. The oligonucleotide primers used to generate various mutations are listed in Table 4.

3. Plasmid purification: Small scale and large scale

For small scale plasmid purification, 5-6 ml of LB media containing Ampicillin at 100 μg/ml was inoculated with a single transformed colony for overnight growth at 37°C with shaking at 240 rpm. Bacteria were pelleted and mini preparations (for upto 20 μg yield of plasmid DNA) were done using a plasmid miniprep kit (Sigma-Aldrich, St. Louis, MO) as per manufacturer's protocol.

To obtain large scale quantities of plasmid DNA (upto 500 μg yield of plasmid DNA), 3 ml of LB media supplemented with 100 μg/ml Ampicillin was inoculated with a single bacterial colony and grown at 37°C with shaking at 240 rpm. The 3 ml culture was added to 200 ml LB containing 100 μg/ml Ampicillin and grown overnight. Cultures were pelleted at 4000 rpm for 15 min at 4°C and plasmid DNA was isolated using a maxiprep kit (Qiagen, Valencia, CA) as per manufacturer's recommendations.

4. Transformation of plasmids:

To transform the plasmid of interest that presents ampicillin resistance property to the cells, 1 μl of chilled plasmid DNA was added to 15 μl of α-select competent cells (Bioline,

Table 4: List of oligonucleotide primers used for site-directed mutagenesis

The following primers used for site-directed mutagenesis contain nucleotide changes resulting in single, double or triple amino acid substitutions. F and R indicate forward and reverse primers respectively. Nucleotide changes for various mutations are shown after the primer name whereas missense codons are represented in bold in the forward primers.

1. Y28A F: TAC→GCA 5'-GAGAATGCGCCCATC**GCA**TGCATCTGCCGCAAAC-3'
2. Y28A R: 5' GTTTGCGGCAGATGCATGCGATGGGCGCATTCTC-3'
3. D44A F: GAC→GCC 5' CTTCATGATCGGGTGT**GCC**AACTGCAATGAGTGG-3'
4. D44A R: 5'-CCACTCATTGCAGTTGGCACACCCGATCATGAAG-3'
5. W49A F: TGG→GCC 5'-GTGTGACAACTGCAATGAG**GCC**TTCCATGGGGACTGCATCC-3'
6. W49A R: 5'-GGATGCAGTCCCCATGGAAGGCCTCATTGCAGTTGTCACAC-3'
7. D44A-W49A F: GAC→GCC, TGG→GCG
   5'-**GCC**AACTGCAATGAG**GCG**TTCCATGGGGACTG-3'
8. D44A-W49A R: 5'-CAGTCCCCATGGAACGCCTCATTGCAGTTGGC-3'
9. G42P F: GGG→CCG 5'-CAACTGCTTCATGATC**CCG**TGTGACAACTGCAATG-3'
10. G42P R: 5'-CATTGCAGTTGTCACACGGGATCATGAAGCAGTTG-3'
11. C29A F: TGC→GCA 5' GAGAATGCGCCCATCTAC**GCA**ATCTGCCGCAAACCGGAC-3'
12. C29A R: 5'-GTCCGGTTTGCGGCAGATTGCGTAGATGGGCGCATTCTC-3'
13. C73A F: TGC→GCG 5'-GAGTGGTACTGTCGGGAG**GCG**AGAGAGAAAGACCCCAAG-3'
14. C73A R: 5'-CTTGGGGTCTTTCTCTCTCGCCTCCCGACAGTACCACTC-3'
15. C70A F: TGT→GCT 5'-TCCGGGAGTGGTAC**GCT**CGGGAGTGCAGAG-3'
16. C70A R: 5' CTCTGCACTCCCGAGCGTACCACTCCCGGA-3'
17. C43A C46A F: TGT→GCT, TGC→GCC
    5'CAACTGCTTCATGATCGGG**GCT**GACAAC**GCC**AATGAGT GGTTCCATGGG3'
18. C43A C46A R: 5'CCCATGGAACCACTCATTGGCGTTGTCAGCCCCGATCATGAAGCAGTTG3'
19. C31A F: TGC→GCC 5'-GCCCATCTACTGCATC**GCC**CGCAAACCGGACATC-3'
20. C31A R: 5'-GATGTCCGGTTTGCGGGCGATGCAGTAGATGGGC-3'
21. H51A,C54AF: CAT→GCT, TGC→GCC
    5'-CGCCAATGAGTGGTTC**GCT**GGGGAC**GCC**ATCCGGATCACTGAG3'
22. H51A, C54A R: 5'-CTCAGTGATCCGGATGGCGTCCCCAGCGAACCACTCATTGGCG-3'
23. C70A, C73A F: TGT→GCG, TGC→GCG 5'-CGGGAGTGGTAC**GCG**CGGGAG**GCG**AGAG-3'
24. C70A, C73A R: 5'-CTCTCGCCTCCCGCGCGTACCACTCCCG-3'
25. R56A F: CGG→GCG 5'- CATGGGGACTGCATC**GCG**ATCACTGAGAAGATG-3'
26. R56A R: 5'-CATCTTCTCAGTGATCGCGATGCAGTCCCCATG-3'

27. E59A F: GAG→GCG 5’-CATCCGGATCACT**GCG**AAGATGGCCAAG-3’

28. E59A R: 5’-CTTGGCCATCTTCGCAGTGATCCGGATG-3’

29. K60A F: AAG→GCG 5’-CATCCGGATCACTGAG**GCG**ATGGCCAAGGCCATC-3’

30. K60A R: 5’-GATGGCCTTGGCCATCGCCTCAGTGATCCGGATG-3’

31. REK-AAA F: CGG→GCG, GAG→GCG, AAG→GCG
5’-CATGGGGACTGCATC**GCG**ATCACT**GCGGCG**ATGGCCAAGG CCATC-3’

32. REK-AAA R: 5’-GATGGCCTTGGCCATCGCCGCAGTGATCGCGATG
CAGTCCCCATG-3’

33. REK-GGG F: CGG→GGG, GAG→GGG, AAG→GGG
5’-TGGGGACTGCATC**GGG**ATCACT**GGGGGG**ATGGCCAAGGCC-3

34. REK-GGG R: 5’- GGCCTTGGCCATCCCCCCAGTGATCCCGATGCAGTCCCCA-3’

35. REK-PPP F: CGG→CCG, GAG→CCG, AAG→CCG
5’-CATGGGGACTGCATC**CCG**ATCACT**CCGCCG**ATGGCCAAGGCCATC-3’

36. REK-PPP R: 5’-GATGGCCTTGGCCATCGGCGGAGTGATCGGGATGCAGTCCCCATG-3’

37. DEL 55-60 F: 5’-TTCCATGGGGACTGCATGGCCAAGGCCATC-3’

38. DEL 55-60 R: 5’-GATGGCCTTGGCCATGCAGTCCCCATGGAA-3’

Taunton, MA) that was thawed on ice. This mixture of plasmid DNA/competent cells was incubated on ice for 20 min followed by heat shock at 42°C for 40 seconds. After addition of 300 μl of LB, the reaction was shaken at 37°C for an hour at 240 rpm. The competent cells were spread onto LB agar petri plates containing 100 μg/ml ampicillin for overnight growth at 37°C or until isolated colonies appeared on the plates.

5. Cell culture

Cells were routinely cultured in a CO_2 incubator (Thermo Scientific, Waltham, MA) at 37°C and 5% CO_2. The Human Embryonic Kidney-293 (HEK-293) cell line was grown on tissue culture dishes (Midwest Scientific, St. Louis, MO) in Dulbecco's Modified Eagle's medium (Gibco BRL Life Technologies, Grand Island, NY) enriched with 10% Hyclone Fetal Clone III Serum (GE Healthcare) and 100 units/ml of penicillin/streptomycin (Invitrogen). Murine ES cells (CCE 916) were cultured on tissue culture dishes coated with 0.1% gelatin in Dulbecco's Modified Eagle's medium enriched with 20% fetal bovine serum (Gibco BRL Life Technologies), 100 units/ml penicillin/streptomycin, 1.4% leukemia inhibitory factor (LIF)-conditioned media derived from chinese hamster ovary (CHO) cells (prepared by Erika Dobrota), 2mM L-glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.025% HEPES (pH 7.5, Invitrogen), 1% Hank's balanced salt solution (Invitrogen) and 100nM β-mercaptoethanol. Fresh ES media was replaced on cells every other day.

6. Transient transfection

Transient transfection was performed in HEK-293 cells using Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. Briefly, cells were grown in 10 cm dishes until they reached a confluency of 70-90%. Fresh media (10 ml) was added 4 hours before transfection. Plasmid DNA (24 μg) was diluted in Opti-MEM medium (Invitrogen) to a volume of 1.5 ml and incubated at room temperature for 5 min. Sixty microliters of Lipofectamine 2000 were diluted in Opti-MEM medium to a volume of 1.5 ml and incubated for 20 min. Diluted plasmid DNA (1.5 ml) and diluted Lipofectamine 2000 (1.5 ml) were mixed and the 3 ml mixture was added to the cells. Cells were harvested after two days of growth in 37°C and 5% CO_2.

7. Stable transfection

In order to stably express CFP1 or its mutants in CXXC1$^{-/-}$ ES cells, 25 μg of pcDNA 3.1/Hygro expression vector containing WT or mutant CFP1 cDNA sequence was electroporated into early passage (10-20) CXXC1$^{-/-}$ ES cells. The same expression vector without any CFP1 cDNA sequence was electroporated into CXXC1$^{-/-}$ ES cells to be used as vector control. 25 μg of plasmid DNA was linearized by overnight restriction digest using Nru1 restriction enzyme (New England Biolabs) at 37°C followed by inactivation of the enzyme at 65°C for 1 hour. Digested plasmid DNA was subjected to phenol-chloroform extraction followed by ethanol precipitation. ~200 ng of the purified digested plasmid was run on 1.5% agarose gels containing ethidium bromide to verify linearization of the plasmid by exposure to ultraviolet light.

Approximately, 10^7 CXXC1$^{-/-}$ ES cells growing at a confluency of ~70% were harvested by trypsinization, washed in PBS and resuspended in electroporation buffer [25mM HEPES (pH 7.5)]. These cells were transferred to sterile 4 mm gap electroporation cuvettes (Midwest Scientific) and mixed with ~25 μg digested and purified plasmid DNA. Pulse control setting was set to infinity on the Gene Pulser II (Bio-rad, Hercules, CA) and the mixture of cells and plasmid DNA was vortexed briefly and electroporated at 300V and 500 μF capacitance. After electroporation, cells were transferred to gelatin coated 15 cm tissue culture dishes containing fresh ES culture medium and were allowed to grow and recover for 48 hours at 37°C and 5% CO_2. Transfected cells were then supplemented with culture medium containing 200 μg/ml of hygromycin B (Sigma) and grown for ~14 days until individual clones appeared. These clones were picked and expanded thereby establishing stable cell lines. ES medium containing 200 μg/ml Hygromycin was replaced every other day throughout the procedure. However, for experiments, stably transfected ES cells were grown in ES medium containing 50 μg/ml Hygromycin.

Confirmation of protein expression of each stable clone was done by western blot analysis using antiserum against CFP1.

8. Recombinant GST-protein purification

A single bacterial colony containing the plasmid of interest was inoculated into 5 ml LB media containing ampicillin (100 μg/ml) for overnight growth at 37°C with shaking (240 rpm). Further, 250 ml of LB-Ampicillin media was inoculated with the 5 ml starter culture and grown with shaking at 37°C until the OD_{600} of the culture reached a value ~0.7. Cultures were induced with 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 hours at room temperature followed by high speed centrifugation (6000 rpm) at 4°C to pellet bacteria followed by storage at -80°C to allow for lysis of cells by a freeze-thaw cycle. Pelleted bacteria were resuspended and lysed by sonication in GST binding buffer [25 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 (+1mM phenyl methyl sulphonyl fluoride (PMSF), 1X cOmplete protease inhibitor (Roche, Indianapolis, IN) added fresh)]. After high speed centrifugation (15000 rpm) at 4°C, the supernatant containing GST fusion protein was incubated with a 50% slurry of glutathione-agarose beads (Sigma) (pre-washed in GST binding buffer) with rocking for 3 hours at 4°C. After 3 subsequent washes of the beads with GST binding buffer, GST fusion proteins were eluted using a buffer containing 10 mM reduced glutathione and 50 mM Tris-HCl (pH 8.0). Further, eluates were subjected to three sequential rounds of dialysis in 1X phosphate buffered saline (PBS). These proteins were separated by electrophoresis on an 8% polyacrylamide gel followed by staining of the gel with Simply Blue Safe Stain (Invitrogen) in order to observe and compare the intensity of GST-protein bands against BSA standards (0.0125-2.5 μg) to allow rough estimation of GST-protein concentration.

9. Micrococcal nuclease (MNase) treatment

GST-proteins were subjected to micrococcal nuclease treatment according to a previously published protocol [199]. Briefly, GST-proteins prepared as above were dialysed in TGMC buffer containing 20 mM TrisHCl (pH 7.9), 20% glycerol, 5 mM MgCl2, 5 mM CaCl2, 0.1% NP-40, 1 mM DTT, 0.2 mM PMSF and 0.1 M NaCl. MNase (New England Biolabs) was added to the GST-protein dialyzed in TGMC at a concentration of 0.033 U/μl and incubated at 30°C for 10 min after gentle mixing. Binding reactions were set up with MNase-treated GST-proteins as explained under 'Histone peptide pulldown'.

10. *In vitro* binding assays

a. Calf thymus histone pull-down assay

Two micrograms of GST-protein were incubated along with 1 μg of calf thymus total histones (Worthington, Lakewood, NJ) in binding buffer (50mM Tris-HCl, pH 7.5, 1M NaCl, 1% NP-40, 0.5 mM EDTA, 1mM PMSF, 1X cOmplete protease inhibitor [Roche]) at 4°C overnight followed by 1 hour incubation with glutathione-agarose beads. Following three washes in binding buffer, proteins bound to beads were denatured by 1X SDS sample buffer (Laemmli) (50 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 12.5 mM EDTA and 0.02 % bromophenol blue) and subjected to western blotting using antibodies against total H3, H3K4me1/me2/me3 and H3K9me3 (Abcam, Cambridge UK) [200].

b. Histone peptide pulldown

One microgram of biotinylated, N-terminal tail peptide (1-21 aa) of unmodified histone H3, H3K4me1/me2/me3 or H3K9me3 (AnaSpec EGT, Fremont CA) was incubated overnight at 4°C with 2 μg purified GST-protein in a binding buffer composed of 50 mM Tris pH 7.5, 300 mM NaCl and 0.05% NP-40. Streptavidin sepharose beads (GE Healthcare) were then added for 1 hour to immobilize the biotinylated histone peptides and pull down the interacting CFP1 PHD domain. After extensive washing, 1X SDS sample buffer was added to the beads and western blotting was performed on the eluates using an antibody directed against the GST tag (Thermo Scientific).

11. FLAG Immunoprecipitation

Cells were collected by scraping with a cell scraper and rinsed in cold 1X PBS followed by resuspension in hypotonic solution 4°C to allow swelling of cells. Cells were then lysed by Dounce homogenization 10 times. The resultant nuclear fraction was pelleted by centrifugation at 4000 rpm for 10 min at 4°C. This fraction was incubated with cytoskeletal (CSK) buffer [10 mM PIPES (pH 6.8), 100 mM NaCl, 300 mM sucrose, 3 mM MgCl2, 1mM EGTA, 0.5% Triton X-100, 1% protease inhibitor cocktail (Sigma), 1 mM PMSF and 1mM DTT] and lysed by Dounce homogenization 30 times. The lysate was treated with 50 units of benzonase nuclease (Millipore, Billerica, MA) for 1 hour on ice to enable degradation of DNA followed by centrifugation at 13000 rpm at 4°C for 15 min.

The nuclear extract was incubated for 3 hours with 50 μl of FLAG M2-conjugated agarose beads with rocking at 4°C. Prior to incubation, these beads were freshly prepared by washing twice in 1X TBS [20mM Tris-HCl (pH 8.0), 150 mM NaCl], twice with 0.1 M glycine-HCl (pH 3.5) and again thrice in 1X TBS. Following incubation, beads were washed thrice in CSK buffer, boiled in 1X SDS sample buffer and subjected to western blot analysis.

12. Preparation of Whole cell extract

a. From liver tissue

Approximately, 100 mg of liver tissue was weighed out and placed on dry ice. CSK buffer (300mM NaCl), containing protease inhibitors, was added to the liver tissue followed by homogenization using PTA 10S polytron (Dremel, Racine, WI) at the setting 'medium' for 20 seconds. The lysate was placed on ice throughout the procedure. The lysate was then centrifuged at 3000 rpm for 5 min. The supernatant was collected carefully to avoid the fat layer into another microfuge tube. This was again centrifuged at 13000 rpm at 4°C to remove the remnant cellular debris and the resulting supernatant was used as the protein extract.

b. From cells in culture

Approximately half a million healthy and exponentially growing cells were collected using a cell scraper, washed with 1X PBS and centrifuged at 1800 rpm for 3 min at 4°C. CSK buffer, containing protease inhibitors, was added to lyse these cells by vortexing at high setting for seven intervals of 30 seconds each (3.5 min total). The lysate was centrifuged at 13000 rpm at 4°C for 20 min to collect debris. The supernatant obtained was used as the protein extract. Protein concentration of the lysates was estimated by the Bradford method followed by western blotting [201].

13. Western Blot/Immunoblot analysis

Protein concentration of cellular extracts was determined by the spectroscopic Bradford assay [201]. Samples were denatured by boiling in 1X sample buffer and subjected to electrophoresis on 8% for CFP1 protein and on 15% polyacrylamide gels for histone marks. Proteins were transferred onto polyvinylidene fluoride (PVDF) membrane; either

0.45 μM (GE Healthcare) for CFP1 protein or 0.2 μM (Millipore) for histone proteins. The membrane was incubated with primary antibody diluted in 5% blotting grade blocking reagent (Bio-rad) in TBS-T (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1% Tween 20) for 1 hour at room temperature. The membrane was washed with TBS-T thrice for 10 min and then incubated with appropriate secondary antibodies linked to horseradish peroxidase (GE Healthcare). After three TBS-T washes, the membrane was incubated with ECL detection reagent (GE Healthcare) followed by autoradiography.

Photographic films were scanned on a flatbed scanner and bands were quantified by densitometry using ImageJ (NIH). In order to re-probe blots, they were washed twice in TBS-T followed by incubation in fresh, mild stripping buffer [200 mM glycine, 0.1% SDS, 1% Tween 20, pH 2.2] twice at room temperature. This was followed by two washes in 1X PBS and 1X TBS-T for 10 minutes each. Blots were checked for any remnant signal from the earlier probe by incubation with ECL reagents and exposing the X-ray film. After washing the blot in TBS-T, re-probing was done as explained above.

The following antibodies were used for western blot analyses:

a. Polyclonal rabbit CFP1 antiserum [64]
b. Polyclonal rabbit anti H3K4me1, H3K4me2, H3K4me3, H3K9me3 and anti-total H3 (Abcam)
c. Polyclonal rabbit anti-GST (Thermo Scientific)
d. Mouse monoclonal anti β-actin (Sigma)

14. Confocal microscopy

ES cells ($2x10^4$) were seeded onto sterile coverslips placed on each well of a 24 well tissue culture dish and grown for 2 days. ES medium was removed and cells were washed twice with cold 1X PBS. Further, cells were fixed using 4% paraformaldehyde for 20 min at room temperature followed by three washes with cold 1X PBS. Cells were then permeabilized in PBS-T (0.5% Triton X-100 in PBS) for 10 min followed by a single wash with 1X PBS. Blocking buffer (1% BSA, 5% Normal Donkey Serum (Santacruz Biotechnologies, Santa Cruz, CA in PBS-T) was prepared fresh and added for 1 hour at room temperature with gentle rocking. After a wash with 1X PBS, the primary antibody

for H3K4me3 (Abcam) was diluted in blocking buffer (1:100) and added to the cells for overnight incubation at 4ºC. The next day, cells were washed three times in 1X PBS and incubated with FITC conjugated secondary antibody (Santa Cruz) diluted 1: 200 in blocking buffer for 2 hours at room temperature. After 3 washes with 1X PBS, cells were stained with 0.1 μg/ml of 4', 6-diamidino-2-phenylindole (DAPI) solution for 20 min at room temperature. Coverslips were washed once in sdH20 and twice in 1X PBS before mounting on slides with Vectashield mounting medium (Vector Laboratories, Burlingame, CA). A Zeiss LSM 700 confocal imaging system (Department of Ophthalmology, Eugene and Marilyn Glick Eye Institute) was used to observe staining patterns of nuclei. This system utilized excitation wavelengths of 405 nm and 488 nm for visualizing DAPI staining and FITC staining respectively. Images obtained were analyzed using MetaMorph v7 (Molecular Devices, Sunnyvale, CA) with similar thresholding values for each image. Thresholding for DAPI was done so as to include DAPI bright regions whereas for FITC, thresholding was performed to exclude background and non-specific staining. Percent colocalization of H3K4me3 over DAPI was determined for at least 15 nuclei for every cell line using the MetaMorph colocalization module.

15. Morphological analysis of embryonic stem cell differentiation

ES cells growing at around 50-70 % confluency were trypsinized and washed thrice in ES media without LIF. The number of viable cells was determined using a hemocytometer by trypan blue dye exclusion and $3.5x10^5$ cells were plated on bacterial culture dishes in ES media lacking LIF. Around 7-8 ml of fresh ES media without LIF was replaced every day for 10 days. At the end of 10 days, morphology of cells was studied by microscopic analysis (Leica Microystems, Dr. Dai Lab) at 10X magnification.

16. Measure of Embryonic Stem Cell Differentiation : Leukocyte alkaline phosphatase assay

ES cells were processed as in Method 11. Further, ten days after initial plating, cells were trypsinized and disaggregated into a single cell suspension and plated on gelatin-coated tissue culture dishes in ES media lacking LIF and were allowed to recover overnight. Staining for alkaline phosphatase was done using a leukocyte alkaline phosphatase detection kit (Sigma). Adherent ES cells were washed in 1X PBS and fixed with citrate

fixative (18 mM citric acid, 9 mM NaCl, 12 mM surfactant buffered at pH 3.6) for 2 min at room temperature. Cells were washed with sterile distilled water (sdH2O) and alkaline phosphatase staining solution, prepared according to the manufacturer's protocol, was added to the cells. The cells were incubated with this staining solution for 15 min followed by a wash with sdH2O. These cells were immediately visualized under a microscope (Leica) and images were taken at 10X magnification. Cells positive for alkaline phosphatase activity are stained reddish purple in color while cells that are negative remained unstained. A minimum of 300 cells were scored for each dish.

17. RNA isolation, cDNA synthesis and semi-quantitative PCR for analysis of stem and lineage markers

Undifferentiated and differentiated ES cells were collected on Day 0 (day of seeding) and Day 10 (ten days after growth in ES media lacking LIF) for analysis of stem and lineage specific mRNA during *in vitro* differentiation. Total RNA was isolated from these cells using RNeasy Mini kit (Qiagen) as per manufacturer's directions. cDNA was generated using the iScript cDNA synthesis kit (Bio-rad). Briefly, 1 μg of RNA template was combined with 4 μl of 5X iScript reaction mix, 1 μl of iScript reverse transcriptase and nuclease free water to a total volume of 20 μl and was subjected to the following thermocycling conditions: 5 minutes at 25°C, 30 minutes at 42°C and 5 minutes at 85°C.

One microliter of single stranded cDNA obtained from the above step was amplified by semi-quantitative PCR in a 50 μl reaction mixture containing 50 pmole of specific forward and reverse primers (Table 5), 0.2 mM each deoxynucleotide triphosphate (dNTP), 0.75 unit Taq DNA polymerase (Roche) and 1X reaction buffer (Roche). The following thermocycling conditions were used: Initial denaturation at 94°C for 2 min followed by 30 cycles (24 cycles for Oct4) at 94°C for 30s, 55°C-60°C for 30s, 72°C for 30s and a terminal elongation step of 72°C for 10 min. After amplification, 25 μl of the reaction product was electrophoresed on 1.5% agarose gels containing ethidium bromide and were visualized by exposure to ultraviolet light.

18. Genomic DNA isolation

To isolate genomic DNA from tail-biopsy specimens or liver tissue samples, 600 μl of cell lysis buffer (10 mM Tris [pH 8.0], 100 mM EDTA, 0.5% SDS) was added along with 200 μg of Proteinase K (Roche) to allow digestion of the tissue overnight at 50^{o}C. The digests were then incubated with 6 μl of RNase A (Roche, 10mg/ml) at 37^{o}C for 15 min. The tubes were chilled on ice and 200 μl of 7.5 M Ammonium acetate solution was added to precipitate proteins followed by brief vortexing and incubation on ice for 10 min. The sample was then centrifuged at 13000 rpm at 4^{o}C to pellet proteins. The supernatant ($\sim$750 μl) was transferred to another microfuge tube and 525 μl of isopropanol was added and mixed thoroughly by tube inversion. The isolated DNA was centrifuged for 2 min at 13000 rpm at room temperature and supernatant was discarded. The DNA pellet was rinsed with 70% ethanol, air dried and re-suspended in a suitable amount of sdH2O.

19. Animal studies

a. Generation of CXXC1 $^{fl/fl}$ AlbCre+ mouse line

Studies involving mice were designed and executed in adherence to the Guide for the Care and Use of Laboratory Animals (Guide, NRC 2011). In order to understand the importance of CFP1 in an adult animal, a mouse strain of the C57BL/6 background, carrying a conditional (floxed) allele of the CXXC1 gene (CXXC1$^{fl/fl}$) was generated [176]. To further study the significance of CFP1 in liver development and regeneration, CXXC1$^{fl/fl}$ mice were bred with AlbCre mice [C57BL/6-TgN(AlbCre)21Mgn] purchased from The Jackson Laboratory (Bar Harbor, Maine) [202]. The AlbCre mice carry the Cre-recombinase gene under the control of the liver-specific albumin promoter. All pups born from the breeding of CXXC1$^{fl/fl}$ and AlbCre mice were heterozygous for the conditional allele (CXXC1$^{+/fl}$), with or without the AlbCre transgene (AlbCre$^{+/-}$). Further, CXXC1$^{+/fl}$ AlbCre+ mice were bred with CXXC1$^{+/fl}$ AlbCre$-$ mice to obtain pups homozygous for the conditional allele and carrying the Cre transgene (CXXC1$^{fl/fl}$ AlbCre+) to be used as experimental animals. This breeding scheme also resulted in a roughly equivalent number of CXXC1$^{fl/fl}$ AlbCre$-$ animals that were utilized as controls for experiments. Male CXXC1$^{fl/fl}$ AlbCre+ mice and female CXXC1$^{fl/fl}$ AlbCre$-$ mice were paired for breeding to prevent offspring from having two copies of the Cre transgene. Verification of genotypes was performed by PCR on genomic DNA obtained from tail-biopsy specimens using the

Table 5: List of Primers used for RT-PCR Analysis of stem cell and lineage development markers

1.	c-fms	: 5’-CTGAGTCAGAAGCCCTTCGACAAAG-3’
		: 5’-CTTTGCCCCAGACCAAAGGCTGTAGC-3’
2.	gp-II B	: 5’-AGGCAGAGAAGACTCCGGTA-3’
		: 5’-TACCGAATATCCCCGGTAAC-3’
3.	βH1	: 5’-AGTCCCCATGGAGTCAAAGA-3’
		: 5’-CTCAAGGAGACCTTTGCTCA-3’
4.	GATA-4	: 5’-CACTATGGGCACAGCAGCTCC-3’
		: 5’-TTGGAGCTGGCCTGCGATGTC-3’
5.	Oct-4	: 5’-GGCGTTCTCTTTGGAAAGGTGTTC-3
		: 5’-CTCGAACCACATCCTTCTCT-3’
6.	HPRT	: 5’-CACAGGACTAGAACACCTGC-3’
		: 5’-GCTGGTGAAAAGGACCTCT-3’

primers LOXP1F: 5′-TGTAGACACTTGTGGGAAGCC and CXXC1R: 5′-AGTTCACCCAGACCCTCTTC-3′ [176]. The presence of the LoxP sites in the CXXC1 conditional allele results in a PCR fragment slightly larger than that produced from the WT allele. The presence or absence of the Cre transgene was determined by PCR on the tail-DNA using primers oIMR 1084: GCGGTCTGGCAGTAAAAACTATC and oIMR 1085: GTGAAACAGCATTGCTGTCACTT (The Jackson Laboratories) (Figure 5A and 5B).

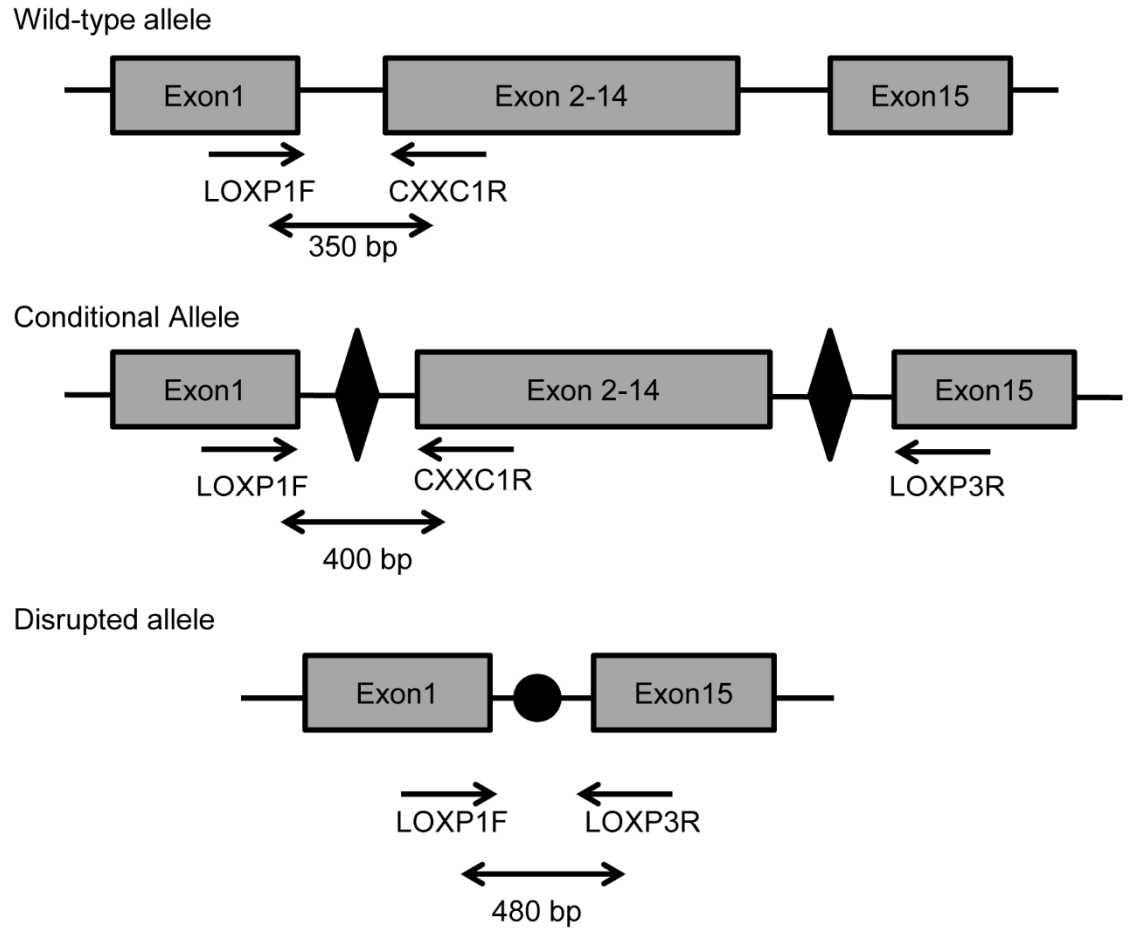
b. Confirmation of CXXC1 deletion specifically in the liver

Five and ten week old male mice, heterozygous for the conditional allele and carrying the Cre transgene (CXXC1$^{+/fl}$ AlbCre+), were sacrificed and tissues from the liver, lung, spleen, kidney were collected in microfuge tubes. Genomic DNA was isolated from these tissue samples as described above. In order to determine if CXXC1 was deleted specifically in the liver, genomic DNA obtained from tissues of various organs was used as templates in two distinct PCR reactions. In the first PCR reaction, primers LOXP1F and LOXP3R: 5'-GTTCCTGCTCAAAGAGCTCG-3' were utilized. In the other PCR reaction, primers LOXP1F and CXXC1R were used, as in PCRs performed for genotype determination. The location of the primers used and expected sizes of DNA fragments resulting from PCR amplification is shown in Figure 5A.

c. Partial hepatectomy (PH) surgery

Five male CXXC1$^{fl/fl}$ AlbCre+ and CXXC1$^{fl/fl}$ AlbCre−mice aged 2-3 months or 5-6 months were subjected to PH. This surgery was carried out by Dr. Guoli Dai, The Department of Biology, IUPUI according to a previously published protocol [203]. Briefly, mice were anesthetized by a 3% flow of isofluorane and an oxygen flow rate of 2L/min in an anesthesia chamber available in the surgery rooms at IUPUI Animal Facility. Anesthesia was maintained during the surgery at 2% flow of isofluorane and an oxygen flow rate of 0.8L/min. After mice were adequately anesthetized, their initial body mass was determined. Further, their upper abdomens were disinfected with 70% ethanol and were placed on a warming pad to prevent the risk of hypothermia during surgery. A midline incision was made on the upper abdomen and the three anterior lobes: right upper lobe, left upper lobe and left lower lobe (2/3 liver) were tied at the lobe origins and resected. The peritoneum was rejoined using a running 4-0 silk suture and skin was closed using

A)


Wild-type allele
Exon1
Exon 2-14
Exon15
LOXP1F
CXXC1R
350 bp
Conditional Allele
Exon1
Exon 2-14
Exon15
LOXP1F
CXXC1R
LOXP3R
400 bp
Disrupted allele
Exon1
Exon15
LOXP1F
LOXP3R
480 bp

B)

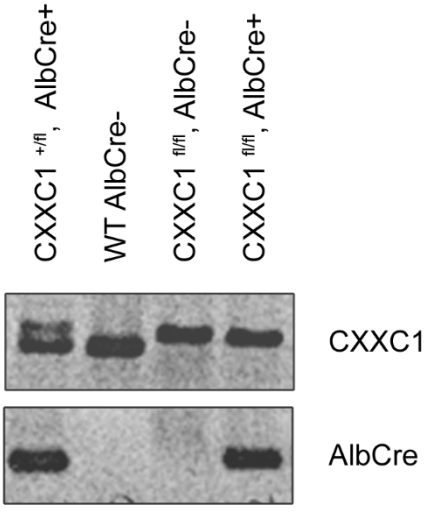

CXXC1 +/fl, AlbCre+
WT AlbCre-
CXXC1 fl/fl, AlbCre-
CXXC1 fl/fl, AlbCre+
CXXC1
AlbCre

Figure 5: Schematic representation of the linear sequence of the WT CXXC1 allele, conditional CXXC1 allele and the disrupted allele and genotyping analysis of transgenic mice

A) Arrows indicate the position and direction of oligonucleotide primers designed to evaluate recombination events occurring after induction of Albumin-Cre. Expected sizes of DNA fragments after PCR amplification are indicated.

B) DNA isolated from tail-biopsy specimens was amplified by PCR using primers LoxP1F and CXXC1R for the conditional allele; and oIMR 1084 and oIMR 1085 for the Cre transgene. The amplified products were subjected to electrophoresis on a 1% agarose gel containing ethidium bromide and visualized by exposure to ultraviolet light for analysis of genotype.

wound closure clips. This was followed by subcutaneous administration of sterile saline (3 ml) to replace any loss of fluid during surgery. The total time from start till finish of the surgery was limited to 15 min to reduce trauma. The mice were occasionally placed under warming lights to ease the process of coming out of anesthesia. Seven days after PH, mice were sacrificed and body mass and masses of the right kidney and remnant liver were recorded. A sample of the remnant liver was collected for protein analysis and another was placed in tissue embedding cassettes and fixed in 10% neutral buffered formalin solution for tissue sectioning. A small piece of tissue was mounted in OCT embedding compound (Tissue-Tek), placed on heptane on top of dry ice to allow gradual freezing and was stored at -80°C for future use.

20. Histone protein preparation from liver tissue

Liver tissue samples were weighed, placed on dry ice and cut into small pieces with a scalpel. These pieces were transferred into a Dounce homogenizer followed by addition of 1 ml of TEB buffer (PBS containing 0.5% Triton X 100, 2 mM PMSF and 0.02% NaN3) for every 200 mg of tissue. Pieces of liver tissue were disaggregated by 60 strokes, transferred to a microfuge tube and centrifuged at 13000 rpm for 5 min at 4°C. Supernatant was removed and the tissue pellet was resuspended in 3 volumes of extraction buffer (0.5 N HCl, 10% glycerol) to allow extraction of histones. This mixture was incubated on ice for 30 min followed by centrifugation at 13000 rpm at 4°C. Eight volumes of acetone were added to the resultant supernatant fraction followed by overnight incubation at -20°C. This fraction was centrifuged at 13000 rpm for 5 min and the histone pellet was air-dried and further dissolved in 100 μl of distilled water. Histone proteins were quantified by the Bradford method [201] and stored at -80°C in aliquots.

21. Histopathological analysis: Hematoxylin & Eosin (H&E) staining

Liver tissues were placed in embedding cassettes and chemically fixed by submerging in 10% neutral buffered formalin solution. These cassettes were transferred to 70% ethanol and sent to Department of Pathology and Laboratory Medicine, IUSM for routine processing service that includes paraffin embedding and sectioning of paraffin blocks into 5 μm slides followed by the standard histopathological H&E staining. Liver pathology was

examined by Dr. Harm HogenEsch, Purdue University College of Veterinary Medicine, West Lafayette, IN.

22. Statistical analysis

Data presented in this dissertation were evaluated for statistical significance by unpaired, two-tailed Student's t-test with equal variance. A difference of p-value ≤ 0.05 was considered to be statistically significant. Error bars represent Standard Error of the Mean (SEM). Graphical and statistical analyses were performed with Microscoft Excel and GraphPad Prism 6 softwares.

Results

1. Characterization of the interaction between CFP1 and methylated H3K4

a. The N-terminal CFP1 PHD binds H3K4me1/me2/me3

PHDs are frequently found in chromatin remodeling proteins and are known to interact with either modified or unmodified histone tails [123, 124, 126]. The domain structure of CFP1 consists of an N-terminal PHD that spans from amino acids 27-73. Comparison of the N-terminal PHD of CFP1 to other known histone H3K4me2/3 binders such as PHF2, ING2 and BPTF [125, 128, 131] revealed the conservation of crucial H3K4me3 binding residues; Y28, D44 and W49 (enclosed in boxes in Figure 6A). To examine its histone binding properties, human CFP1 PHD was bacterially expressed as a GST-fusion protein [GST-CFP1 PHD (1-102aa)] and pulldowns were performed with bulk calf thymus histones. Western blotting using antibodies against H3K4me1, H3K4me2 and H3K4me3 and total H3 revealed the association of GST-CFP1 PHD with these marks. However, GST-CFP1 PHD was unable to bind H3K9me3, a heterochromatin mark. (Figure 6B). This finding was further validated by performing histone peptide pulldown using biotin-labelled histone H3 N-terminal tail peptides (1-21aa) modified at various sites of methylation. It was found that GST-CFP1 PHD was specifically retained by biotinylated histone tail peptides H3K4me1, H3K4me2, H3K4me3 but not by H3K9me3 (Figure 6C). To ensure that the lack of binding of GST-CFP1 PHD to H3K9me3 was not a false-negative result, the ability of the H3K9me3 peptide tail to bind a known H3K9me3 reader, the N-terminal chromodomain of HP1α, was assessed. It was found that GST-tagged chromodomain of HP1α was able to interact with the H3K9me3 histone peptide tail (Figure 6D).

b. Mutations in CFP1 PHD ablate its binding to H3K4me3

When the putative H3K4me3 binding residues in GST-CFP1 PHD (Y28, D44 and W49) were mutated individually to alanine by site-directed mutagenesis and *in-vitro* histone peptide pulldowns were subsequently performed, a dramatic reduction in the binding of the GST-CFP1 PHD domain to the H3K4me3 tail peptide was observed (Figure 7A). Moreover, in order to understand the histone binding properties of CFP1 *in vivo*, the mammalian expression vector, pcDNA3.1 Hygro/Flag, containing WT or mutated forms

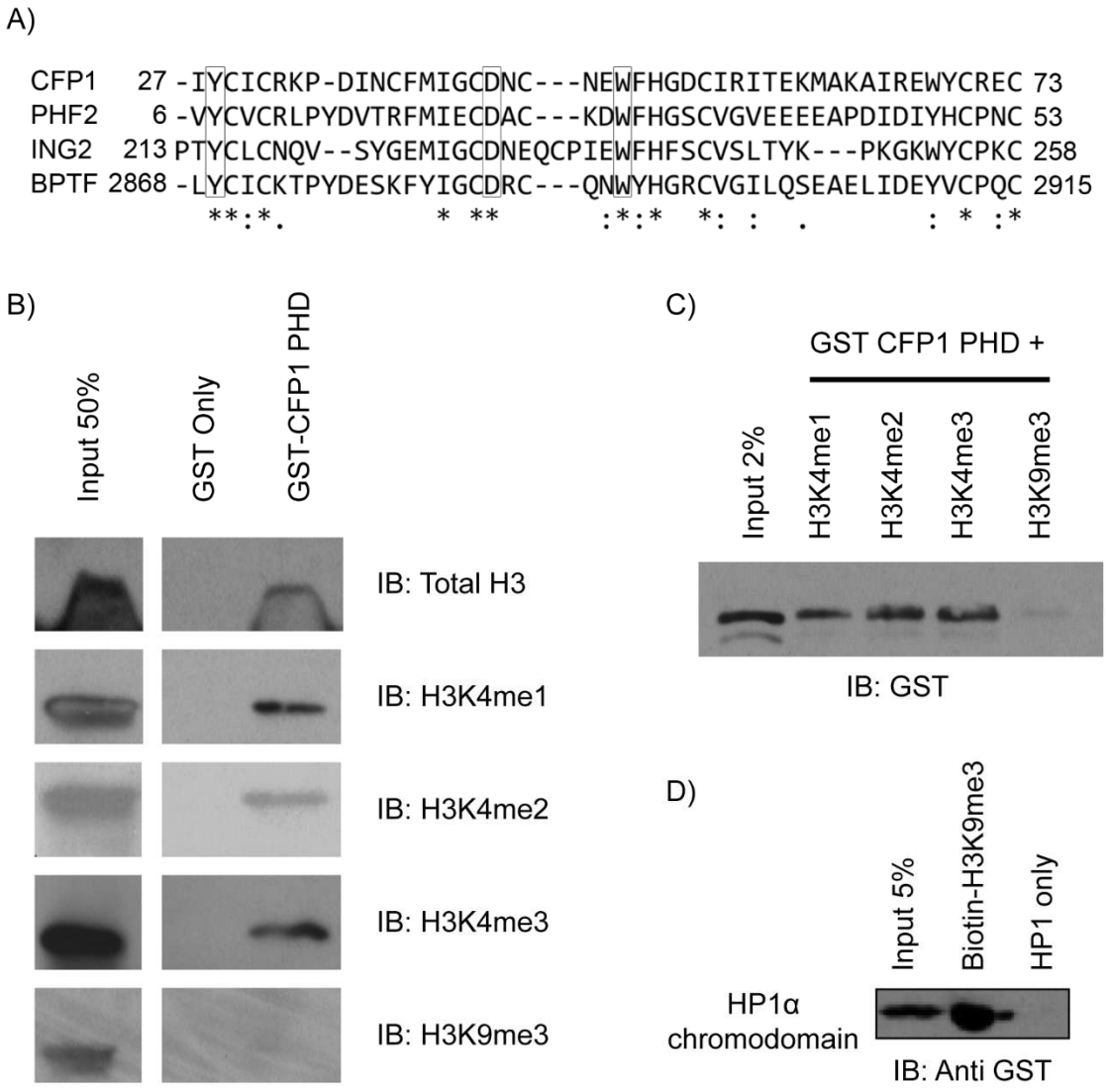


Figure 6: The N-terminal PHD of CFP1 binds H3K4me1/me2/me3

A) Sequence comparison of CFP1 PHD and PHDs in other known H3K4me3 binders: PHF2 (PHD finger protein 2), ING2 (Inhibitor of growth protein 2) and BPTF (Bromodomain PHD finger transcription factor) was performed using the Clustal O program. Conserved residues important for H3K4me3 binding are enclosed in boxes.

B) Purified GST-CFP1 PHD or GST was incubated with a mixture of calf thymus histones followed by addition of GSH agarose beads to immobilize the formed complex. Bound histone proteins were analyzed by immunoblotting (IB) using antisera against mono, di or trimethylated H3K4, trimethylated H3K9 and total H3.

C) Histone peptide pulldowns were performed wherein purified GST-CFP1 PHD was incubated with N-terminal, biotin-labelled histone H3 peptides (1-21aa) carrying the indicated modifications followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum.

D) Histone peptide pulldown was performed wherein purified GST-HP1 chromodomain was incubated with N-terminal, biotin-labelled histone H3K9me3 peptide (1-21aa) followed by addition of streptavidin sepharose beads. Bound protein was analyzed by immunoblotting (IB) with GST antiserum.

Y28A, D44A and W49A) of full length CFP1 cDNA were transiently transfected into HEK-293 cells. Nuclear extracts were prepared and Flag-CFP1 was immunoprecipitated from these extracts using Flag M2-conjugated agarose (Sigma). The immunoprecipitated material was subjected to western blot analysis for the endogenous H3K4me3 mark using antiserum against H3K4me3. It was found that the full length form of WT CFP1 was able to interact with endogenous H3K4me3 but the three point mutant forms of full length CFP1 completely abolished this interaction *in vivo* (Figure 7B).

c. Isolation of stable CXXC1$^{-/-}$ ES clones expressing CFP1 point mutations

CXXC1$^{-/-}$ ES cells are viable but exhibit a number of defects in comparison to CXXC1$^{+/+}$ ES cells. These pleiotropic defects can be rescued by introduction of a full length CFP1 expression vector (1-656aa) into CXXC1$^{-/-}$ ES cells [177]. This ability of wild type CFP1 to rescue defects in CXXC1$^{-/-}$ ES cells was used as a tool to probe for structure function relationships of CFP1 [180]. To understand the functional significance of the interaction of CFP1 with methylated H3K4, various mutated forms of FLAG-tagged CFP1 (1-656) were stably expressed in CXXC1$^{-/-}$ ES cells and were further analyzed for rescue function (Figure 8A). CXXC1$^{-/-}$ ES cells expressing full length FLAG-tagged CFP1 (1-656) and CXXC1$^{-/-}$ ES cells expressing only the vector backbone without the CFP1 sequence (VEC) were isolated to be utilized as positive and negative controls, respectively, in subsequent experiments. CFP1 protein levels were determined for every stable clone by western blot analysis using an antiserum against CFP1 (Figure 8B).

Previous studies indicate that ES cells that are heterozygous for the disrupted CXXC1 allele (CXXC1$^{+/-}$) do not show a defective phenotype despite expressing $\sim$ 50% CFP1 protein compared to CXXC1$^{+/+}$ ES cells [177]. Hence, stably transfected ES clones expressing at least 50% of the level of CFP1 in comparison to WT ES cells were used for analysis of rescue function. Western blots for CFP1 were performed at regular intervals during the course of rescue experiments to ensure that CFP1 expression in the stable clones was maintained above 50% compared to CXXC1$^{+/+}$ ES cells.

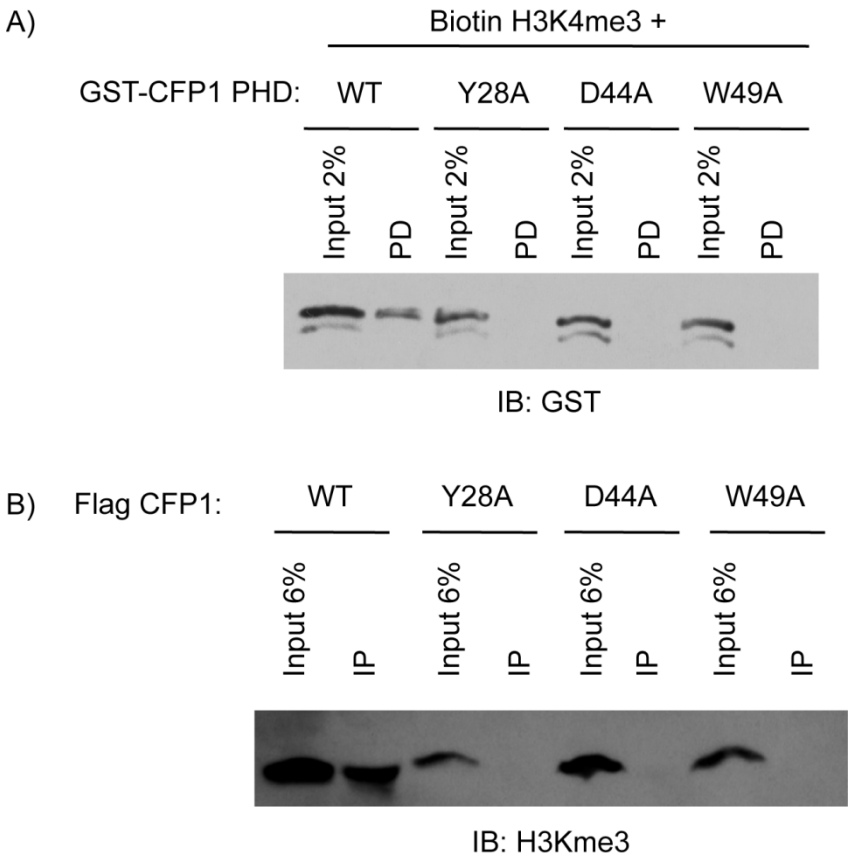


Figure 7: Mutations in CFP1 PHD ablate its binding to H3K4me3

A) GST-CFP1 PHD (WT or the indicated mutants) were incubated with biotin-labelled H3K4me3 peptide followed by pull down (PD) with streptavidin sepharose. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum.

B) Benzonase-treated nuclear extracts prepared from HEK-293 cells transiently expressing WT or mutant Flag CFP1 (1-656) were subjected to Flag immunoprecipitation. Binding of Flag-CFP1 to H3K4me3 was analyzed by immunoblotting with H3K4me3 antiserum.

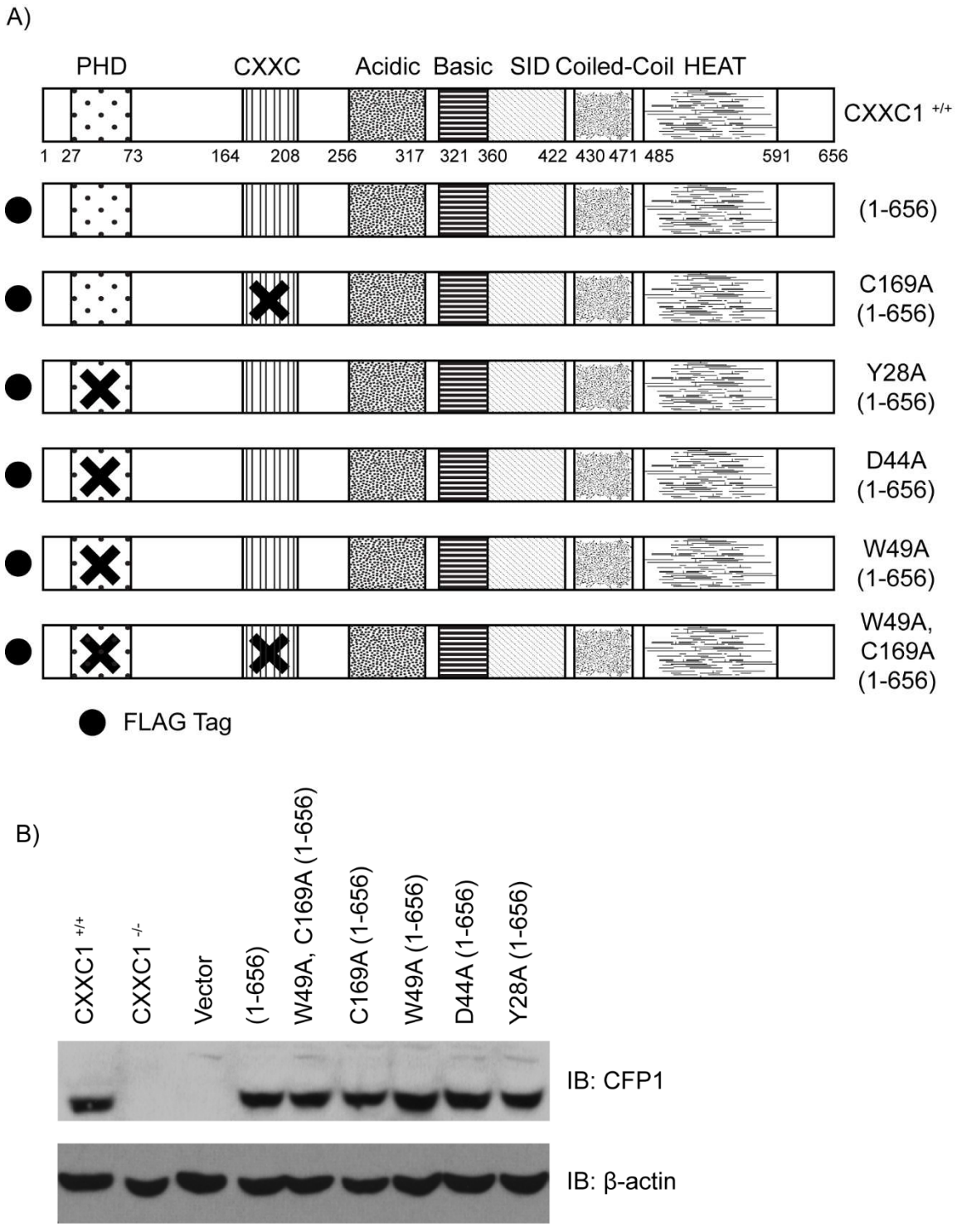


Figure 8: Isolation of stable CXXC1$^{-/-}$ ES clones expressing CFP1 point mutations

A) A schematic representation of full length CFP1 along with amino acid positions of each domain followed by CFP1 mutations that were stably expressed in CXXC1$^{-/-}$ ES cells. The "X" indicates the approximate location of the indicated point mutation within

full length CFP1. The black circle at the N-terminus of CFP1 represents the Flag tag. The nomenclature of each of the stably expressed constructs used in this dissertation is represented on the right of the respective schematic.

B) Immunoblot analysis was performed on whole cell lysates obtained from stable ES clones using CFP1 antiserum to determine the protein expression level of CFP1 which was compared to CFP1 expression in CXXC1$^{+/+}$ ES cells. The membrane was re-probed with β-actin antiserum for analysis as a loading control. A representative immunoblot of three independent experiments is shown.

Taking into account that the CFP1 transgene might have become incorporated into a functionally important region of the genome, thereby altering the interpretation of rescue activity, two to three independent clones were analyzed in subsequent rescue experiments.

In order to avoid the possibility of increased passage number (number of times the cells have been sub-cultured) from affecting results, stable clones were expanded and at least 5 vials were frozen for each independent clone immediately after stable selection (at passage 1-3) and thereafter, cells from lower passage numbers (10-15) were used for experiments.

d. Binding of CFP1 PHD to methylated H3K4 is dispensable for appropriate targeting of H3K4me3 to euchromatin in ES cells

Chromatin is divided into two distinct forms based on the level of transcriptional activity. A tightly packed form known as heterochromatin, is usually transcriptionally silent. DAPI (4', 6-diamidino-2-phenylindole) is a fluorescent stain that preferentially associates with double stranded DNA. Heterochromatic regions are characteristically observed as DAPI-bright regions after staining with DAPI. On the other hand, the less intensely stained form, known as euchromatin, has a relatively open conformation and is usually under active transcription.

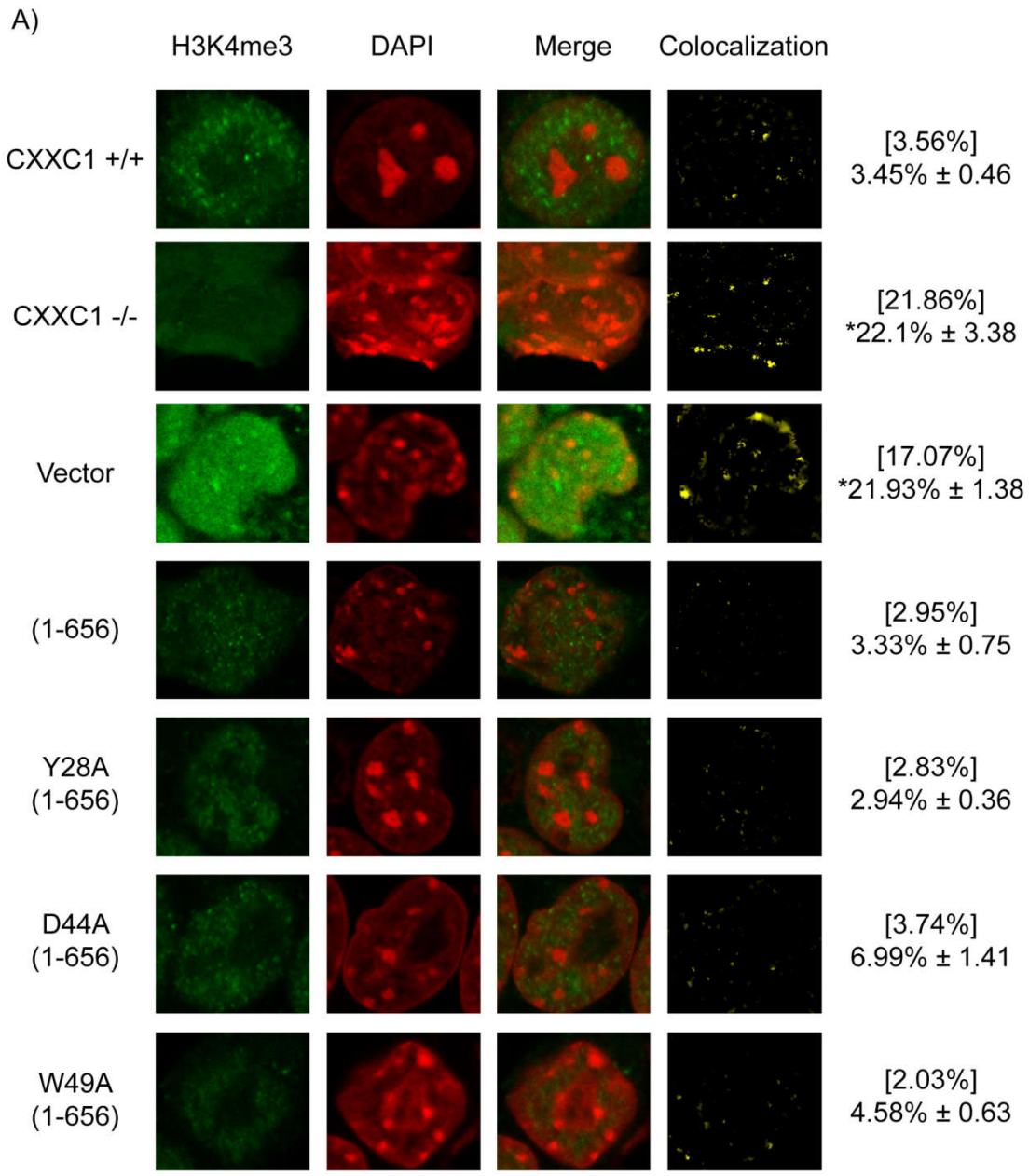
Previous studies in the lab suggest that CXXC1$^{-/-}$ ES cells show an inappropriate mislocalization of H3K4me3 to DAPI-bright heterochromatin region. Moreover, introduction of a DNA binding CFP1 mutant [C169A (1-656)] into CXXC1$^{-/-}$ ES cells did not rescue this defect suggesting that CFP1-DNA interaction is important for appropriate H3K4me3 targeting [179]. However, this interaction is not essential for the rescue of H3K4me3 targeting to the transcription start sites of housekeeping genes [181]. Hence, it was hypothesized that domains of CFP1 other than CXXC might be required for appropriate genomic targeting of H3K4me3 marks to euchromatin. To address the importance of the interaction of CFP1 PHD with methylated H3K4, for appropriate targeting of H3K4me3 marks to euchromatin, sub-nuclear localization of endogenous H3K4me3 was studied in CXXC1$^{-/-}$ ES cells expressing Y28A (1-656), D44A (1-656) and

W49A (1-656) by confocal immunofluorescence. Percent colocalization of H3K4me3 with DAPI-bright heterochromatin was significantly increased in CXXC1$^{-/-}$ ES cells and CXXC1$^{-/-}$ ES cells expressing vector compared to that in CXXC1$^{+/+}$ ES cells or CXXC1$^{-/-}$ ES cells expressing CFP1 (1-656) as observed earlier [179]. However, a rescue of colocalization of H3K4me3 was observed in CXXC1$^{-/-}$ ES cells expressing Y28A (1-656), D44A (1-656) and W49A (1-656) (Figure 9A and 9B). These results suggest that the interaction of CFP1 PHD with methylated H3K4 is not essential for the rescue of appropriate targeting of H3K4me3 to euchromatin.

e. Interaction of CFP1 with methylated H3K4 or DNA is necessary for efficient differentiation of embryonic stem cells

Pluripotent embryonic stem cells possess an ability to differentiate when cultured in suspension in a growth medium lacking LIF [204, 205]. The undifferentiated ES cells express high levels of alkaline phosphatase. Studies suggest that alkaline phosphatase can be utilized as a marker to quantitate the number of undifferentiated stem cells [206]. After the staining procedure, cells containing reddish deposits were considered positive for alkaline phosphatase activity.

CXXC1$^{-/-}$ ES cells fail to differentiate *in vitro* when cultured in growth medium lacking LIF. This defect can be reversed by introduction of a full length CFP1 expression vector [177]. To determine if the interaction of CFP1 PHD with methylated H3K4 is important for the rescue of cellular differentiation, CXXC1$^{-/-}$ ES cells expressing W49A (1-656) were subjected to growth in ES medium lacking LIF for a period of 10 days. These cells produced embryoid bodies (EBs) and subsequently formed extensive outgrowths characteristic of endoderm formation almost identical to CXXC1$^{+/+}$ ES cells and CXXC1$^{-/-}$ ES cells expressing full length CFP1 (1-656) cultured similarly (Figure 10). On staining these cells for alkaline phosphatase after a period of 10 days following LIF withdrawal, less than 5% of the cells were found to be positive for alkaline phosphatase (indicated by reddish deposits) suggesting that most of the CXXC1$^{-/-}$ ES cells expressing W49A (1-656) could undergo efficient differentiation (Figure 11).


A)
H3K4me3
DAPI
Merge
Colocalization
CXXC1 +/+
[3.56%]
3.45% ± 0.46
CXXC1 -/-
[21.86%]
*22.1% ± 3.38
Vector
[17.07%]
*21.93% ± 1.38
(1-656)
[2.95%]
3.33% ± 0.75
Y28A
(1-656)
[2.83%]
2.94% ± 0.36
D44A
(1-656)
[3.74%]
6.99% ± 1.41
W49A
(1-656)
[2.03%]
4.58% ± 0.63

B)

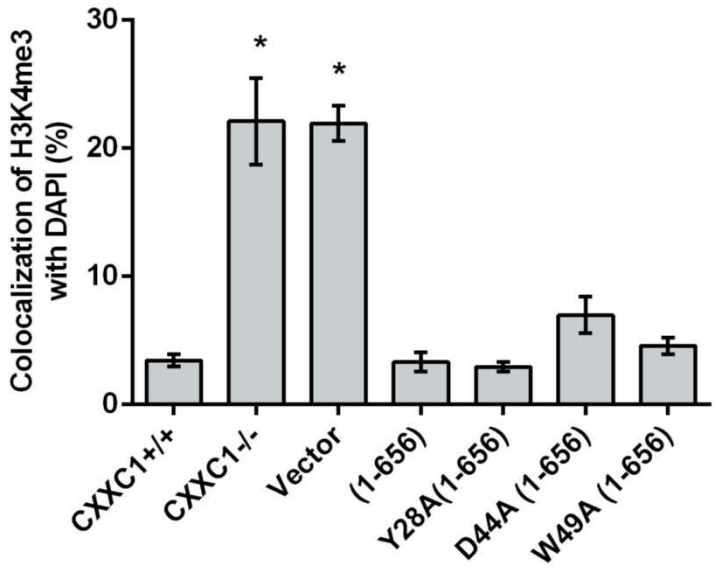


Figure 9: Binding of CFP1 PHD to methylated H3K4 is dispensable for appropriate targeting of H3K4me3 to euchromatin in ES cells

A) Sub-nuclear localization of endogenous H3K4me3 was detected in CXXC1$^{+/+}$, CXXC1$^{-/-}$, CXXC1$^{-/-}$ ES cells expressing vector, full-length CFP1 (1-656), or the indicated CFP1 mutations using a primary rabbit anti H3K4me3 antibody and bovine anti-rabbit IgG-fluorescein isothiocyanate (FITC)-conjugated secondary antibody. After counterstaining with DAPI, the nuclei were studied by confocal microscopy. Yellow regions in the merge and colocalized images indicate colocalization. The numbers within parentheses indicate the percent colocalization for that nucleus calculated using the MetaMorph colocalization module. The numbers not included within parentheses indicate the average percentage colocalization of H3K4me3 with DAPI bright heterochromatin and standard error for 7 to 37 nuclei. Asterisks represent statistically significant difference ($p<0.05$) compared to CXXC1$^{+/+}$ ES cells.

B) The results obtained in 9A are represented in the form of a bar graph. The error bars indicate standard error and asterisks denote statistically significant difference ($p<0.05$)

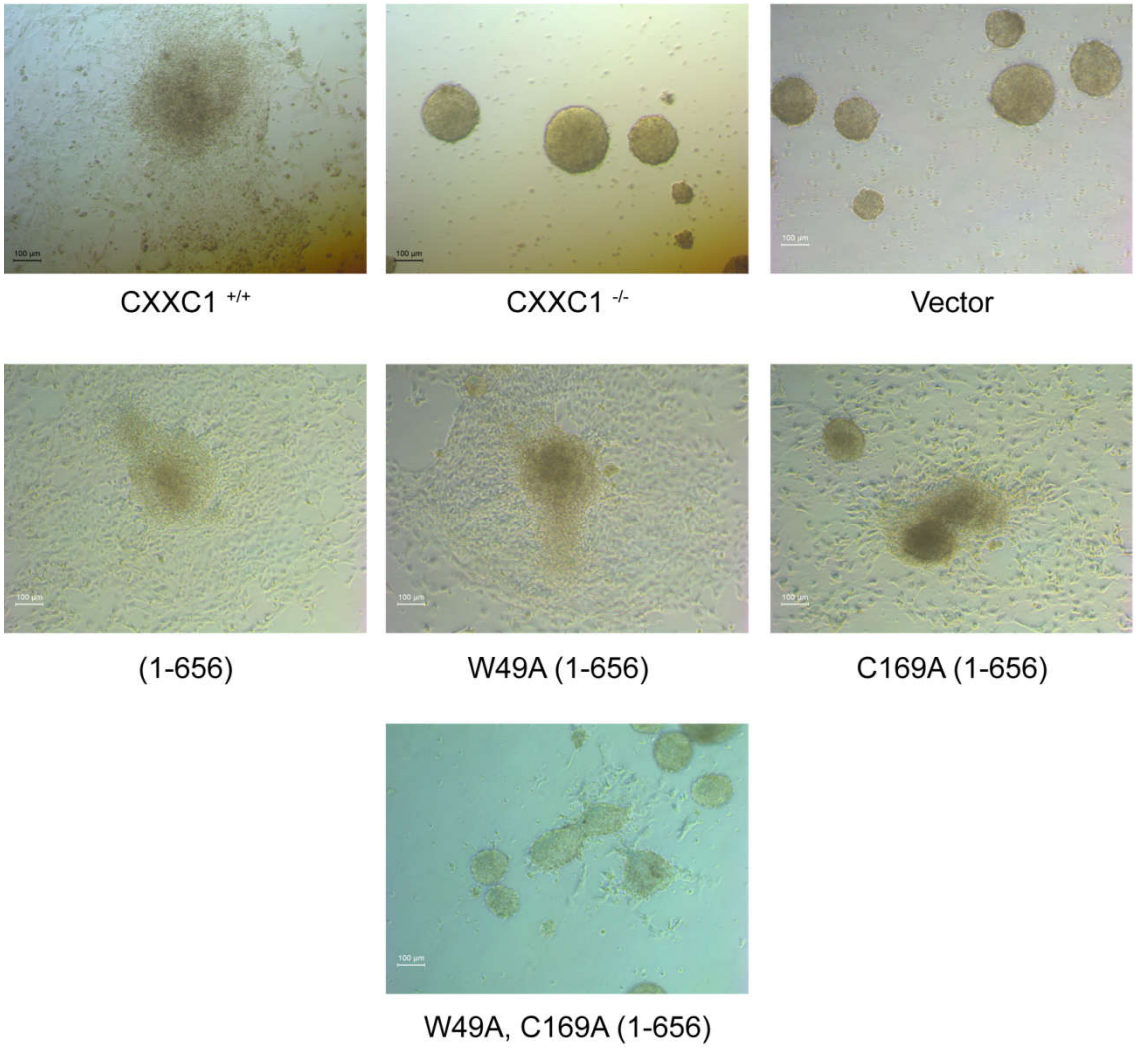


Figure 10: Interaction of CFP1 with H3K4me or DNA is necessary for formation of extensive outgrowths, characteristic of endoderm differentiation

CXXC1$^{+/+}$, CXXC1$^{-/-}$ and CXXC1$^{-/-}$ ES cells expressing vector, full length CFP1 (1-656), or the indicated mutations were grown in ES media lacking LIF for 10 days. Morphology of ES cell colonies was studied by microscopy ten days after induction of differentiation (10X magnification).

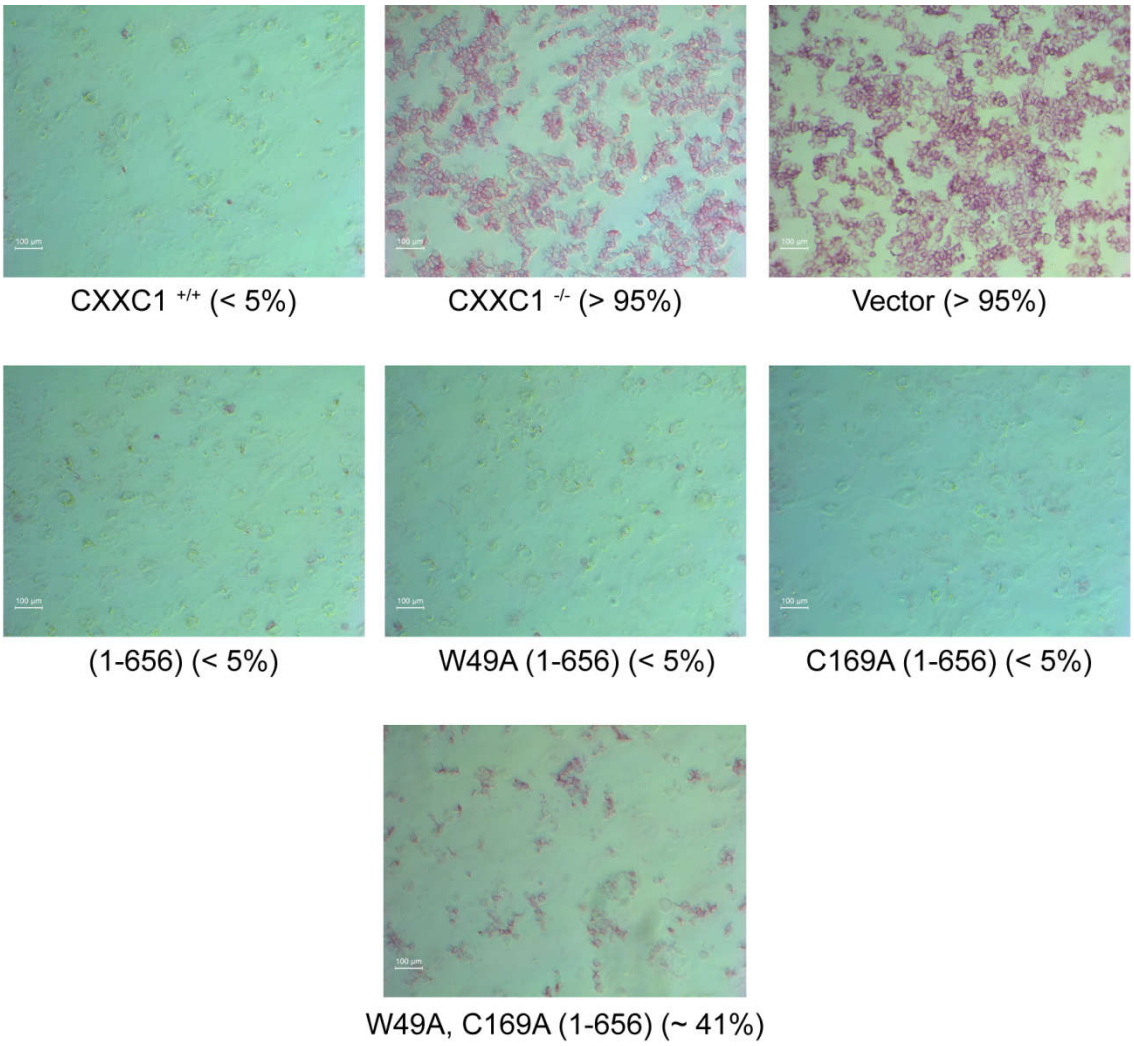


Figure 11: Interaction of CFP1 with H3K4me or DNA is necessary for down regulation of alkaline phosphatase activity during *in vitro* differentiation

CXXC1$^{+/+}$, CXXC1$^{-/-}$ and CXXC1$^{-/-}$ ES cells expressing vector, full length CFP1 (1-656), or the indicated mutations were grown in ES media lacking LIF for 10 days. After a period of 10 days, cells were collected from bacterial culture plates and trypsinized into a single cell suspension. These cells were re-seeded into tissue culture dishes coated with gelatin and were stained for alkaline phosphatase activity using a leukocyte alkaline phosphatase kit. Images were captured at 10X magnification. Cells containing reddish deposits were scored positive for alkaline phosphatase activity. A minimum of 300 cells were scored for each cell line. The numbers within the parentheses indicate the percentage of alkaline phosphatase-positive cells for the indicated cell line.

Introduction of a DNA binding point mutation (C169A) or the SID domain mutant (C375A) in full length CFP1 completely rescues cellular differentiation suggesting that these domains are redundant for rescue activity [180]. To determine if similar redundancy existed amongst the PHD domain and the DNA binding domain, point mutations W49A and C169A in the respective domains were introduced into full length CFP1. Further, CXXC1$^{-/-}$ ES cells stably expressing W49A, C169A CFP1 (1-656) were induced to differentiate by removal of LIF. Ten days after the withdrawal of LIF, examination of morphology revealed that only a few EBs could adhere and form outgrowths that remained small over time. This observation was strikingly different as compared to the well-differentiated morphology of CXXC1$^{+/+}$ ES cells, CXXC1$^{-/-}$ ES cells expressing full length CFP1 (1-656), W49A (1-656) or C169A (1-656). This suggests that CXXC1$^{-/-}$ ES cells expressing W49A, C169A CFP1 (1-656) are unable to undergo efficient differentiation (Figure 10). On staining for alkaline phosphatase, CXXC1$^{+/+}$ ES cells, CXXC1$^{-/-}$ ES cells expressing full length CFP1 (1-656), W49A (1-656) and C169A (1-656) contained less than 5% alkaline phosphatase positive cells. However, in CXXC1$^{-/-}$ ES cells expressing W49A, C169A (1-656), approximately 41% of the cells were found to be positive for this marker, indicating an inability to undergo complete differentiation, consistent with our morphological findings (Figure 11).

The process of ES cell differentiation entails a systematic and highly coordinated down regulation and up regulation of specific markers [207]. Oct4 is a master regulator of pluripotency of embryonic stem and germ cells [208]. Knockdown of Oct4 promotes differentiation of human and mouse ES cells [207, 209]. Induction of ES cell differentiation program results in down regulation of Oct4 expression. Previous studies indicate that CXXC1$^{-/-}$ ES cells are unable to down regulate mRNA expression of Oct4 after induction of differentiation and this defect can be rescued by introduction of a full-length CFP1 expression vector [177]. mRNA expression of Oct4 was determined by RT-PCR for undifferentiated ES cells (Day 0) and 10 days after induction of differentiation by growth in the absence of LIF. The ratio of Oct4 mRNA expression to HPRT was used to evaluate the change in expression of Oct4 after 10 days of induction of differentiation. Parallel to our morphological observations, while CXXC1$^{-/-}$ ES cells and vector containing ES cells retained Oct4 expression at Day 10, CXXC1$^{-/-}$ ES cells expressing

W49A (1-656) could significantly down regulate Oct4 expression similar to CXXC1$^{+/+}$ ES cells, CXXC1$^{-/-}$ ES cells expressing full length CFP1 (1-656) and C169A CFP1 (1-656) suggesting that these cells possess the ability to differentiate. On the other hand, Oct4 mRNA levels fail to get sharply down regulated in CXXC1$^{-/-}$ cells expressing W49A, C169A CFP1 (1-656) suggestive of defective differentiation (Figure 12).

WT but not CXXC1$^{-/-}$ ES cells, are able to upregulate lineage specific markers such as Gata-4 (endoderm), β-H1 (primitive erythroid), gp-IIb (megakaryocyte), c-fms (myeloid) and MHC-β (cardiac muscle) following differentiation by withdrawal of LIF [177]. On analyzing the mRNA expression of these lineage specific markers 10 days after removal of LIF, it was found that CXXC1$^{+/+}$ ES cells, CXXC1$^{-/-}$ cells expressing full length CFP1 (1-656), C169A (1-656), W49A (1-656) and W49A, C169A (1-656) possess the ability to activate markers of lineage restriction (Figure 12).

These observations collectively suggest that interaction of CFP1 with methylated H3K4 is dispensable for efficient ES cell differentiation. However, the loss of interaction of CFP1 with methylated H3K4 and DNA together by introducing a CFP1 PHD-H3K4me binding mutation (W49A) and a CFP1-DNA binding mutation (C169A) into CFP1 (1-656) results in an incomplete rescue of ES cell differentiation.

Summary

The data presented in this portion of the dissertation describes a direct and specific interaction of CFP1 with methylated H3K4 via its PHD domain. Transiently expressed Flag-tagged CFP1 was found to interact with endogenous H3K4me3 suggesting that this interaction occurs *in vivo*. Mutagenizing key residues necessary for binding (Y28, D44 and W49) resulted in ablation of this interaction *in vitro* and *in vivo*. Experiments performed to understand the functional significance of this interaction reveal that CXXC1$^{-/-}$ ES cells expressing CFP1 point mutations that abolish interaction of CFP1 with H3K4me (W49A) can rescue targeting of H3K4me3 to euchromatin and *in vitro* differentiation. However, point mutations that abolish the interaction of CFP1 with methylated H3K4 and DNA result in defective *in vitro* differentiation. Although CXXC1$^{-/-}$ ES cells expressing CFP1 point mutations within the PHD and DNA binding domain,

express lineage specific markers, they fail to downregulate stem cell markers like alkaline phosphatase and Oct4, on induction of differentiation.

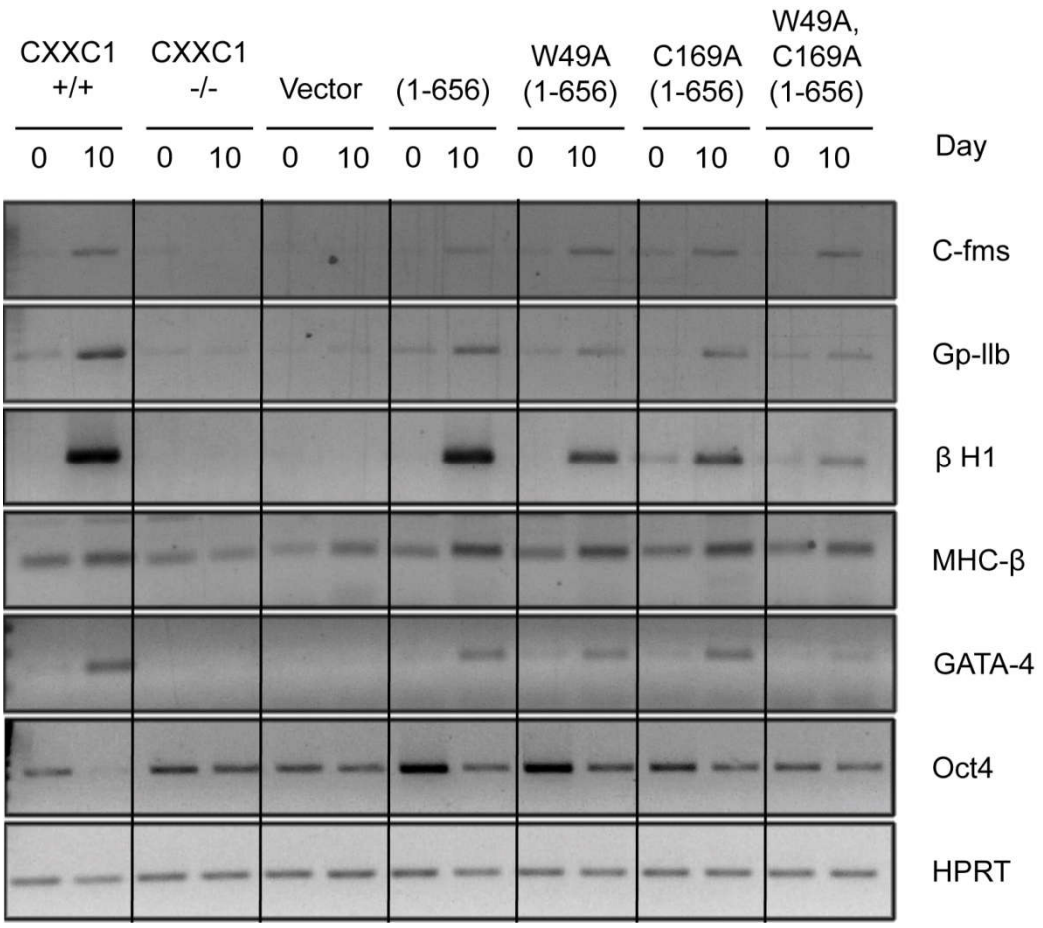


Figure 12: Interaction of CFP1 with H3K4me or DNA is necessary for down regulation of Oct 4 expression during *in vitro* differentiation

$CXXC1^{+/+}$, $CXXC1^{-/-}$ and $CXXC1^{-/-}$ ES cells expressing vector, full length CFP1 (1-656), or the indicated mutations were grown in ES media lacking LIF for 10 days. Collection of cells was done at Day 0 (undifferentiated) and Day 10 after growth in the absence of LIF. After isolation of total RNA, RT-PCR was performed for analyzing the expression levels of Oct 4 (a stem cell marker), HPRT (a constitutively active gene that was used as a control for RNA integrity and cDNA quantity), and lineage specific markers such as c-fms (myeloid), GpIIB (megakaryocyte), β-H1 (primitive erythroid), MHC- β (cardiac muscle) and GATA-4 (endoderm).

2. Characterization of the interaction between CFP1 and unmodified H3K4

a. CFP1 PHD directly binds unmodified H3

PHD domains present in AIRE, BHC80, UHRF1 recognize unmodified H3 (H3K4me0) with high specificity. [135, 136, 210] The conserved PHD finger consists of a two strand anti-parallel β sheet that is stabilized by two zinc atoms secured in its position by the Cys4-His-Cys3 motif in a cross brace topology [124, 125, 211]. The PHD domain in CFP1 extends from amino acids 27-73. To understand the specificity of the association of CFP1 PHD with the unmodified H3 tail, peptide pull-down assays were performed by incubating recombinant GST-CFP1 PHD (1-102aa) and biotin-labelled unmodified H3 histone N-terminal tail peptide (1-21aa) followed by recovery of the histone peptide/GST-protein complex using streptavidin beads. Negative control reactions containing only GST-CFP1 PHD or only the unmodified H3 peptide were included to ensure that the observed interaction was a specific one. On western blotting the eluates using an anti-GST antibody, it was observed that GST-CFP1 PHD directly binds to unmodified H3 suggesting that CFP1 PHD functions as a reader of this mark (Figure 13).

b. Binding affinity of CFP1 PHD to unmodified H3 is higher than that towards methylated H3K4

In order to compare the binding affinity of CFP1 PHD towards H3K4me3, H3K9me3 and H3K4me0, competition histone peptide pulldown experiments were performed. GST-CFP1 PHD was incubated with N-terminal, biotin labelled H3K4me3 peptide along with unlabeled H3K4me3, H3K9me3 and unmodified H3 as competitor peptides. This followed by the addition of streptavidin sepharose to immobilize the bound protein. It was found that an excess of unlabeled H3K9me3 did not affect the interaction between CFP1 PHD and H3k4me3. Thus, unlabeled H3K9me3 does not compete with biotin labelled H3K4me3 to interact with CFP1 PHD, suggesting that CFP1 PHD does not exhibit binding affinity towards the H3K9me3 tail peptide. However, it was found that a lesser amount of unlabeled H3K4me0 than unlabeled H3K4me3 peptide was required to compete with biotin labelled H3K4me3 interacting with CFP1 PHD. This suggests that CFP1 PHD binds with higher affinity to unmodified H3 than to H3K4me3 (Figure 14).

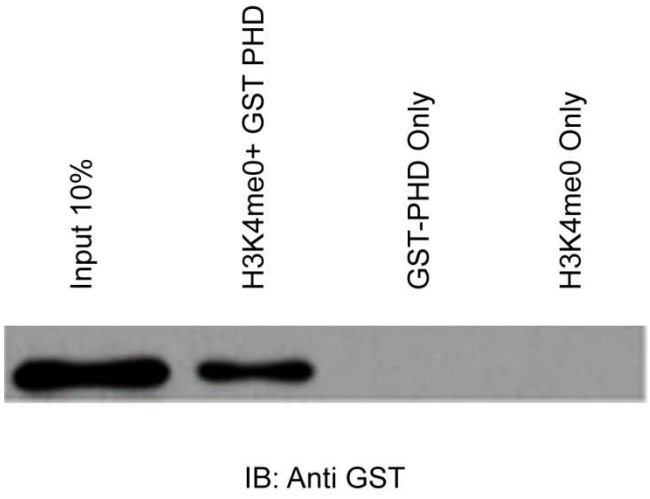


Figure 13: CFP1 PHD directly binds unmodified H3

Histone peptide pulldowns were performed by incubating GST-CFP1 PHD with biotin-labelled N-terminal (1-21aa) unmodified histone H3 peptide (H3K4me0) followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum. Pull down reactions containing only GST CFP1 PHD or only the unmodified H3 peptide served as controls for specificity of interactions.

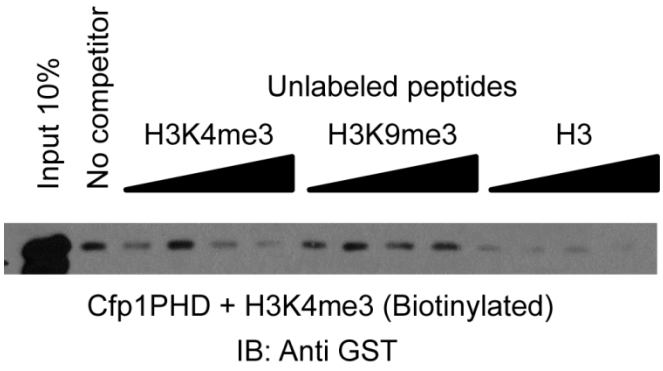


Figure 14: Binding affinity of CFP1 PHD to unmodified H3 is higher than that towards methylated H3K4

Competition histone peptide pulldowns were performed by incubating GST-CFP1 PHD with biotin-labelled N-terminal (1-21aa) histone H3K4me3 peptide and an excess of the indicated unlabeled histone peptides (10 μg-25 μg) as competitors, followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum.

c. Mutations in CFP1 PHD do not ablate CFP1 PHD-unmodified H3 interaction

In order to determine the functional significance of the recognition of unmodified H3 by CFP1 PHD, site-directed mutagenesis was performed to introduce point mutations into CFP1 PHD (1-102) for various conserved residues as follows:

i. Single mutations that ablate CFP1 PHD-H3K4me3 interaction Y28A, D44A, W49A:

ii. Double and triple mutations for the above residues: (Y28A, W49A), (Y28A, D44A), (D44A, W49A), and (Y28A, D44A, W49A).

iii. Single mutation, G42P that is thought to ablate hydrogen bonding between CFP1 PHD and unmodified H3. Structural prediction was performed by Dr. Thomas Hurley, Indiana University School of Medicine, using PyMOL, a molecular visualization software used to view and compare 3D structures of proteins.

iv. Single mutations of conserved cysteine residues: C29A, C73A

v. Mutation of multiple conserved cysteine residues:
   - I. (C29A, C73A)
   - II. (C29A, C70A)
   - III. (C29A, C43A, C46A)
   - IV. (C29A, C43A, C46A, C73A)
   - V. (C31A, C43A, C46A, C73A)

vi. Mutation of a conserved histidine and multiple cysteine residues:
   - VI. (C29A, C43A, C46A, C73A, H51A, C54A)
   - VII. (C31A, C43A, C46A, C73A, H51A, C54A)

vii. Mutation of conserved histidine and cysteine residues that define the signature Cys_4-His-Cys_3 motif:
   - VIII. (C29A, C43A, C46A, C70A, H51A, C54A, C73A)
   - IX. (C31A, C43A, C46A, C70A, H51A, C54A, C73A)

Addition of C70A mutation to the constructs prepared in (vi.) and subsequent histone peptide pulldowns for these two mutant proteins (vii.) were carried out by Ryan Eller, a rotation student from the Department of Biology, IUPUI.

Residues indicated within parentheses () were mutated within the same construct. Roman numerals (I-IX) indicate the corresponding nomenclature used in figures. Histone

peptide pulldowns followed by western blots against GST were performed to examine the interaction of these mutant proteins with unmodified H3 tail peptide. Results from these experiments suggest that these mutations do not ablate the association of CFP1 PHD and unmodified H3 (Figure 15A and 15B).

To examine if the binding between CFP1 PHD and unmodified H3 was actually mediated by direct contact and not indirectly by contaminating nucleic acids (RNA or DNA), leading to false-positive results, histone peptide pulldowns were performed after treating GST protein preparations with micrococcal nuclease (also referred to as S7 nuclease) [199]. Micrococcal nuclease cleaves single or double stranded DNA and RNA without any particular sequence specificity. WT GST-CFP1 PHD and the following mutant GST-CFP1 PHD proteins were subjected to treatment with micrococcal nuclease and subsequently used for peptide pulldowns with unmodified H3.

I. CFP1 PHD (C29A, C73A)

VI. CFP1 PHD (C29A, C43A, C46A, H51A, C54A, C73A)

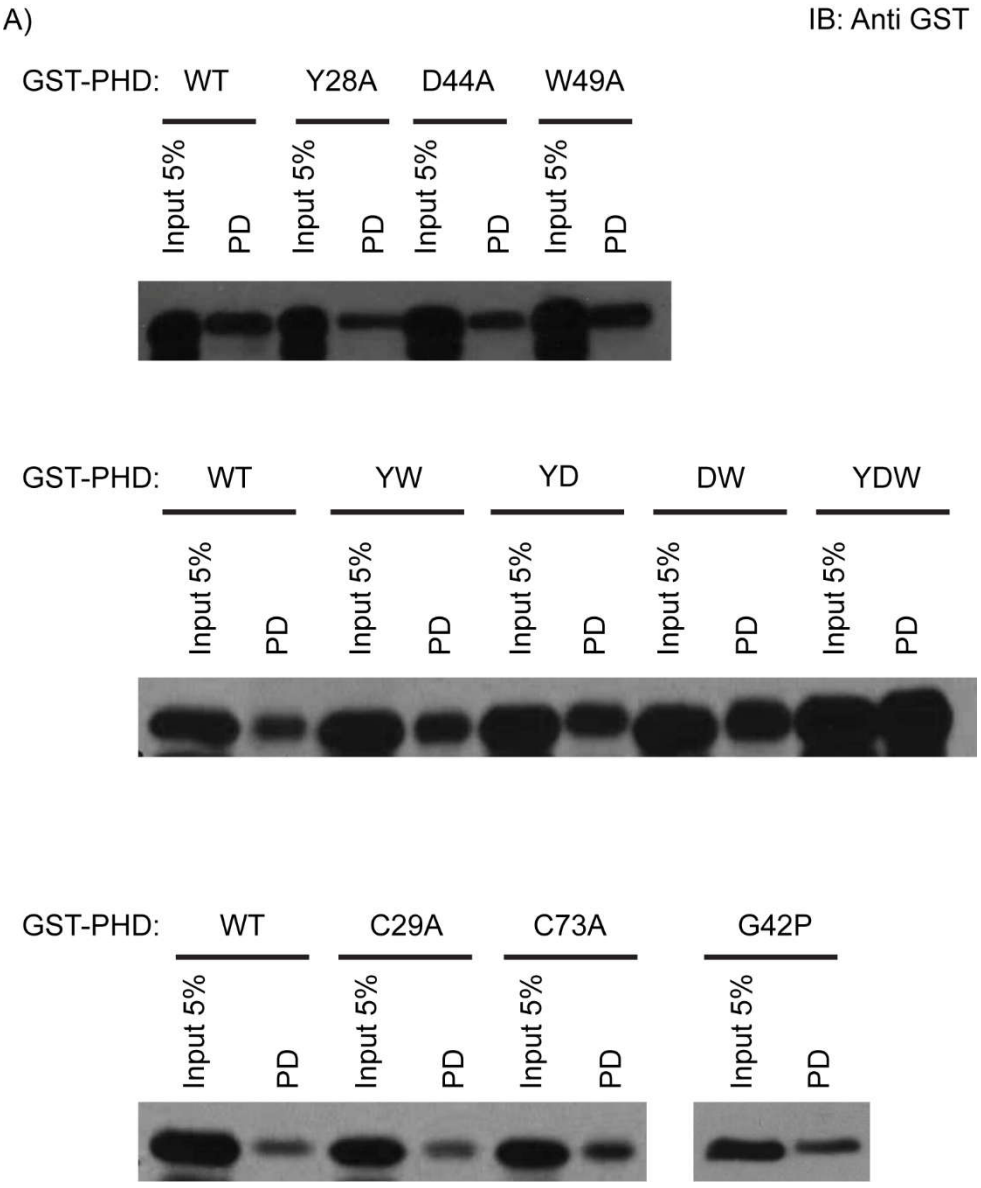
VII. CFP1 PHD (C31A, C43A, C46A, H51A, C54A, C73A)

These results suggest that interaction between mutant CFP1 PHD with unmodified H3 remains unperturbed even after the treatment with micrococcal nuclease (Figure 16) thereby suggesting that this binding is not mediated by contaminating nucleic acids present in the GST protein preparation and can be expected to be a direct one.

d. Analysis of various truncations of CFP1 PHD for unmodified H3 binding

In order to determine the stretch of amino acids in CFP1 PHD critical for interaction with unmodified H3, the following C- and N-terminal deletions of the domain were cloned into the pGEX-4T-1 vector.

A. C-terminal deletions:
   i. CFP1 PHD (1-29)
   ii. CFP1 PHD (1-49)


A)
IB: Anti GST
GST-PHD: WT
Y28A
D44A
W49A
Input 5%
PD
GST-PHD: WT
YW
YD
DW
YDW
Input 5%
PD
GST-PHD: WT
C29A
C73A
G42P
Input 5%
PD

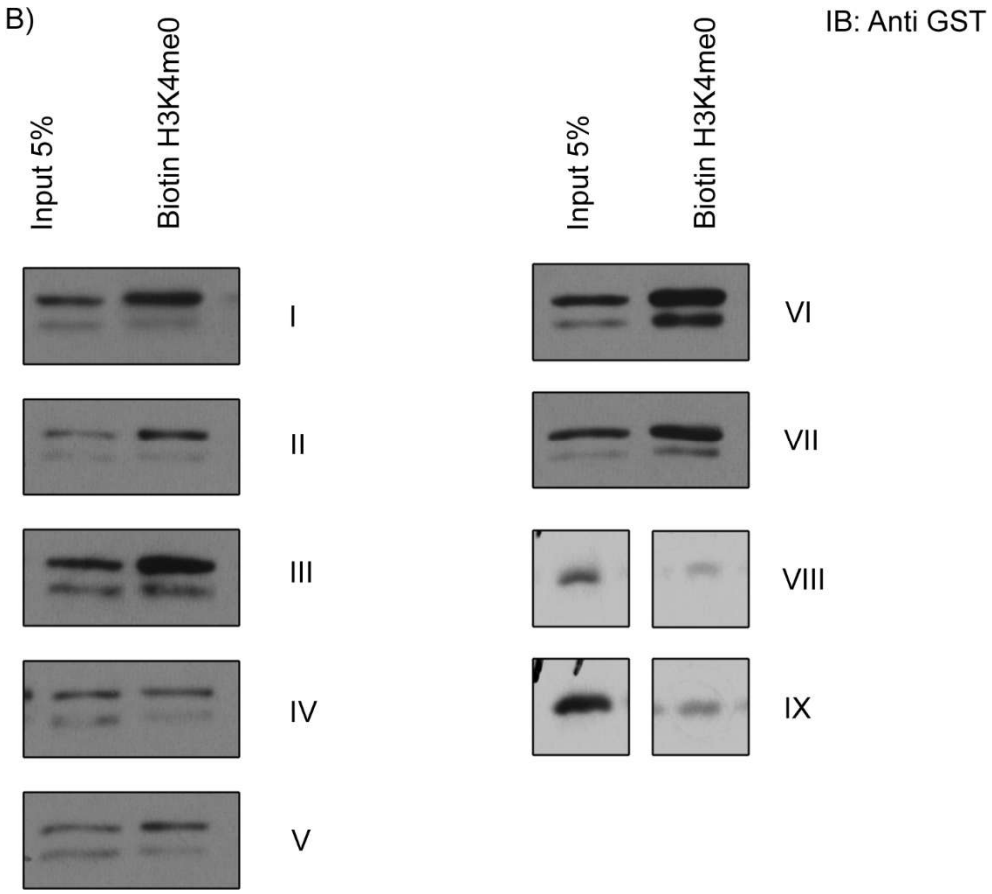


Figure 15: Mutations in CFP1 PHD do not ablate CFP1 PHD-unmodified H3 interaction

A) GST-CFP1 PHD (WT or the indicated mutants) were incubated with biotin-labelled unmodified H3 peptide followed by pull down (PD) with streptavidin sepharose. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum.

YW: (Y28A, W49A) CFP1 PHD,

YD: (Y28A, D44A) CFP1 PHD,

DW: (D44A, W49A) CFP1 PHD,

YDW: (Y28A, D44A, W49A) CFP1 PHD.

B) Mutated GST-CFP1 PHD proteins (as indicated in the figure and below) were incubated with biotin-labelled unmodified H3 peptide followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with

GST antiserum. The Roman numeral indicated to the right of each pulldown represents the mutations in each of the GST-CFP1 PHD fusion proteins as follows:

I. (C29A, C73A) CFP1 PHD

II. (C29A, C70A) CFP1 PHD

III. (C29A, C43A, C46A) CFP1 PHD

IV. (C29A, C43A, C46A, C73A) CFP1 PHD

V. (C31A, C43A, C46A, C73A) CFP1 PHD

VI. (C29A, C43A, C46A, C73A, H51A, C54A) CFP1 PHD

VII. (C31A, C43A, C46A, C73A, H51A, C54A) CFP1 PHD

VIII. (C29A, C43A, C46A, C70A, H51A, C54A, C73A) CFP1 PHD

IX. (C31A, C43A, C46A, C70A, H51A, C54A, C73A) CFP1 PHD

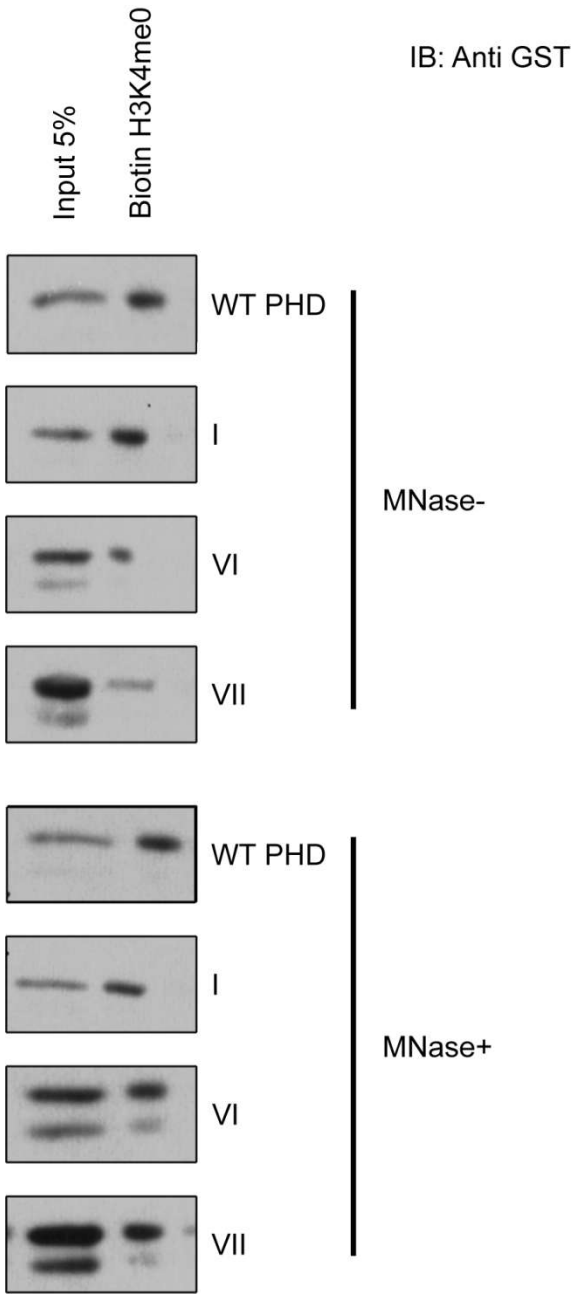


Figure 16: Binding of GST-CFP1 PHD is not mediated by contaminating DNA or RNA in GST-protein preparations

WT and mutant GST-CFP1 PHD proteins (as indicated in the figure and below) were either treated (+) or not treated (-) with micrococcal nuclease (MNase) and further used to perform histone peptide pull down in the presence of biotin-labelled unmodified histone H3 tail peptide. Bound proteins were analyzed by immunoblotting (IB) with GST

antiserum. The Roman numeral indicated to the right of each pulldown represents the mutations in each of the GST-CFP1 PHD fusion proteins as follows:

I. (C29A, C73A) CFP1 PHD

VI. (C29A, C43A, C46A, C73A, H51A, C54A) CFP1 PHD

VII. (C31A, C43A, C46A, C73A, H51A, C54A) CFP1 PHD

iii. CFP1 PHD (1-73)

B. N-terminal deletion:

i. CFP1 PHD (50-102)

This was followed by purification of truncated GST CFP1 PHD proteins and their use in histone peptide pulldowns with unmodified histone H3 tail peptide. This approach revealed that only the fragment constituting the complete PHD domain; CFP1 PHD (1-73) and not CFP1 PHD (1-29), CFP1 PHD (1-49) or CFP1 PHD (50-102) binds to unmodified H3. Hence, neither the N-terminal half; CFP1 PHD (1-49) nor the C-terminal half; CFP1 PHD (50-102) is sufficient for CFP1 PHD-unmodified H3 binding activity (Figure 17)

To further narrow down the amino acids critical for this binding event, following C-terminal deletions were cloned into the pGEX-4T-1 vector.

i. CFP1 PHD (1-55)

ii. CFP1 PHD (1-60)

iii. CFP1 PHD (1-65)

iv. CFP1 PHD (1-70)

Histone peptide pulldowns were performed using the above truncated CFP1 PHD protein fragments with unmodified histone H3 peptides. Results indicate that CFP1 PHD (1-55) shows diminished binding affinity to unmodified H3 whereas CFP1 PHD (1-60), CFP1 PHD (1-65) and CFP1 PHD (1-70) can efficiently bind this mark (Figure 17). This suggests that residues 55-60 aa of CFP1 PHD may be important for binding. Hence, point mutations for charged residues that lie between amino acid 55 and 60 were introduced as follows;

i. Single point mutations R56A, E59A, K60A

ii. Triple mutations REK-AAA, REK-PPP, REK-GGG

Simultaneously, a CFP1 PHD construct containing a deletion of amino acids 55-60 was made by site-directed mutagenesis (CFP1 PHD Del 55-60). Results from histone binding assays with unmodified H3 suggest that none out of the single point mutations R56A, E59A, K60A, the triple mutations REK-AAA, REK-PPP or CFP1 PHD Del 55-60 could

abolish the interaction between CFP1 PHD and unmodified H3 (Figure 18). Western blot analysis (Anti-GST) for CFP1 PHD containing REK-GGG mutation revealed that the protein resolved into multiple bands. This could be attributed to the increased flexibility imparted to the protein by replacement of charged residues to glycine, that might result in an unstable or a misfolded protein [212].

It is unclear if the drop in binding observed in REK-GGG is because of its instability. However, deletion of amino acids 55-60 (that includes R56, E59 and K60) did not ablate the binding of CFP1 PHD to unmodified H3.

Summary

The data presented in this part of the dissertation suggests that CFP1 PHD directly and specifically binds unmodified H3. Surprisingly, point mutation of several conserved residues did not result in ablation of the interaction between CFP1 PHD and unmodified H3. A summary of all the CFP1 PHD mutants/truncations analyzed for interaction with unmodified H3 is depicted in a tabular format (Table 6).

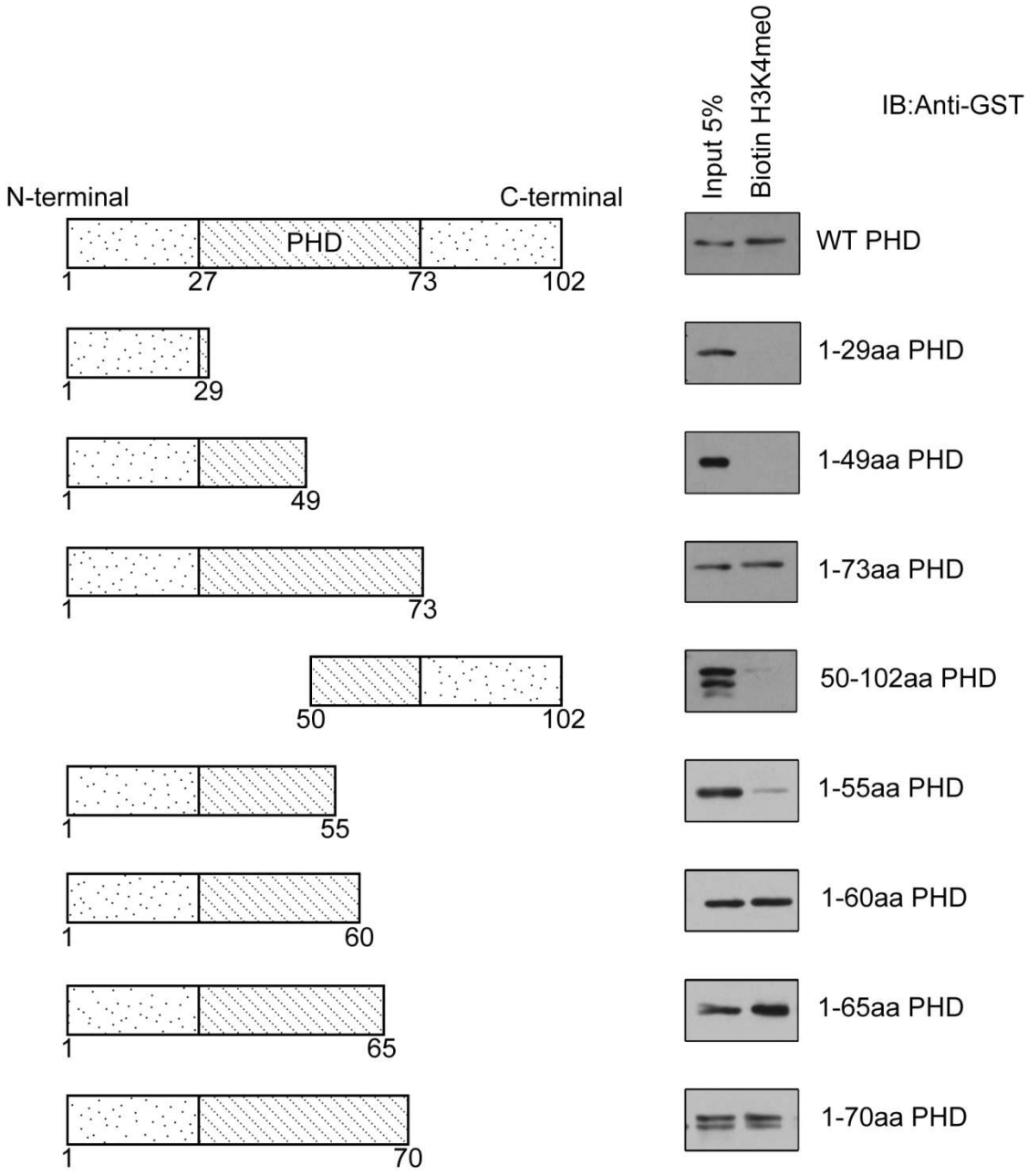


Figure 17: Analysis of various truncations of CFP1 PHD for unmodified H3 binding
WT GST-CFP1 PHD or the indicated GST-CFP1 PHD truncations were incubated with biotin labelled N-terminal unmodified histone H3 tail peptide followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum. A schematic representation of the WT CFP1 PHD and its truncations are depicted in the above figure.

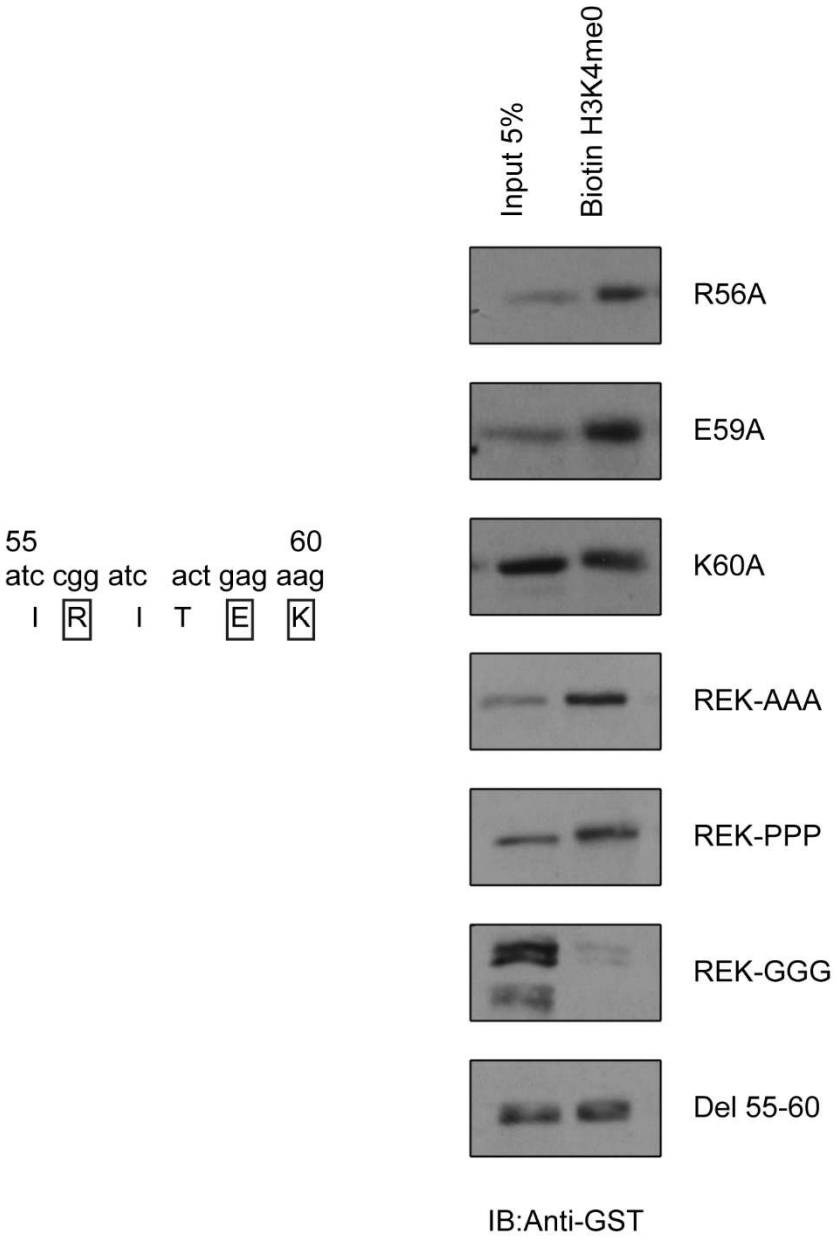


Figure 18: Analysis of point mutations of charged residues located between amino acids 55-60 and their deletion in CFP1 PHD

GST-CFP1 PHD proteins containing substitutions of charged residues located between amino acids 55 and 60 (as indicated in the figure) or showing a deletion of amino acids 55-60 (Del 55-60) were incubated with biotin-labelled unmodified H3 peptide followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum.

Table 6: Summary of the binding activity of various CFP1 PHD mutations and truncations towards unmodified H3

No.	Point mutations within CFP1 PHD	Interaction with H3K4me0 (+/-)
1	Y28A	(+)
2	D44A	(+)
3	W49A	(+)
4	Y28A, W49A	(+)
5	Y28A, D44A	(+)
6	D44A, W49A	(+)
7	Y28A, D44A, W49A	(+)
8	G42P	(+)
9	C29A	(+)
10	C73A	(+)
11	C29A, C73A	(+)
12	C29A, C70A	(+)
13	C29A, C43A, C46A	(+)
14	C29A, C43A, C46A, C73A	(+)
15	C31A, C43A, C46A, C73A	(+)
16	C29A, C43A, C46A, C73A, H51A, C54A	(+)
17	C31A, C43A, C46A, C73A, H51A, C54A	(+)
18	C29A, C43A, C46A, C70A, H51A, C54A, C73A	(+)
19	C31A, C43A, C46A, C70A, H51A, C54A, C73A	(+)
20	R56A	(+)
21	E59A	(+)
22	K60A	(+)
23	REK-AAA	(+)
24	REK-PPP	(+)
25	REK-GGG	(?)
	Truncations of CFP1 PHD	
1	(1-29)	(-)
2	(1-49)	(-)
3	(50-102)	(-)
4	(1-73)	(+)
5	(1-55)	Diminished
6	(1-60)	(+)
7	(1-65)	(+)
8	(1-70)	(+)
9	DEL (55-60)	(+)

The table represents a summary of the ability of various CFP1 PHD mutations and truncations to interact with unmodified H3. An ability to bind unmodified H3 is indicated by (+) whereas an inability to do so is indicated by (-). Further investigation is required to assess the ability of REK-GGG CFP1 PHD mutant to bind unmodified H3 (indicated by "?").

3. Delineation of the role of CFP1 in murine liver homeostasis and regeneration

CFP1 is ubiquitously expressed in various tissues including liver [160]. CXXC1$^{-/-}$ blastocysts are unable to gastrulate and murine ES cells cannot differentiate *in vitro* in the absence of CFP1 [172, 177]. In order to study the significance of CFP1 in adult mice, our lab generated mice (C57BL/6) containing a conditional "floxed" CXXC1 allele (CXXC1$^{fl/fl}$) which permits disruption of murine CXXC1 locus by exploiting the Cre-Lox system [176]. Induction of Cre recombinase by pI-pC in CXXC1$^{fl/fl}$, Mx1 Cre+ mice resulted in deletion of the CXXC1 gene in the liver and hematopoietic cells of adult mice. However, no visible pathology was observed in the liver of these mice by the end of 13 days when these mice die due to a severe defect in hematopoiesis [176]. It is unlikely that a function of CFP1 might manifest in the liver in this relatively short interval (13 days), between depletion of CFP1 and death due to hematopoietic failure.

To understand the role of CFP1 in long-term liver homeostasis and regeneration, a transgenic mouse line, carrying the conditional floxed CFP1 allele and the Cre recombinase gene under the control of the liver-specific albumin promoter (CXXC1$^{fl/fl}$ AlbCre+), was established. This mouse line permitted the depletion of CFP1 after the onset of expression of the albumin gene.

a. Confirmation of CFP1 depletion specifically in the liver at the age of 5 weeks

The AlbCre transgenic mice [C57BL6-TgN(Alb-Cre)21Mgn] that were used in this studywere previously used in another study wherein the AlbCre transgene in combination with a floxed glucokinase gene allele resulted in deletion of this gene specifically in the liver. The authors reported that the efficiency of recombination was only ~40% at birth, 75% at weaning (3 weeks) and appeared to be complete by 6 weeks of age even though albumin is expressed early during murine development (stage 7 to 8 somite). The authors suggest that this low level of recombination at birth might be because of a decreased expression of the Cre transgene [202]. Hence, experiments were performed to determine if deletion of CXXC1 in CXXC1$^{+/fl}$ AlbCre+ (heterozygotes) occurred at 5 or 10 weeks of age. Necropsies were performed for CXXC1$^{+/fl}$ AlbCre+ and control CXXC1$^{+/fl}$ AlbCre− mice and tissues from lung, liver, kidney and spleen were collected. Genomic DNA was

isolated from these tissues and semi-quantitative PCR was performed to amplify the recombined/disrupted allele and the conditional (floxed allele). Primers LOXP1F and LOXP3R were used to amplify the disrupted allele. The stretch of the CXXC1 gene between exon 2 and exon 14 is 4.2 kb long, thus making routine PCR amplification infeasible, resulting in the absence of a PCR product. However, following activation of Albumin Cre and disruption of the CXXC1 allele, exon 1 and exon 15 are only separated by a LoxP site, thus making routine PCR amplification possible and resulting in a PCR product. Therefore, formation of a PCR product in this reaction suggests deletion of CXXC1 gene (refer Figure 5A). On the other hand, primers LOXP1F and CXXC1R were used to amplify the conditional allele. The presence of a conditional allele that contains the LoxP site results in a band that is slightly larger than that produced from the wild type allele, resulting in the formation of a doublet band in mice that are heterozygous for the conditional allele. Results from these experiments indicate the presence of the recombined/disrupted CXXC1 allele specifically in the liver and not in other tissues such as lung, kidney and spleen as early as at 5 weeks of age. In parallel, a reduction in the conditional (floxed) allele, indicative of Cre recombination, was also observed only in the liver and was visible by 5 weeks of age. A reduction, and not a complete loss of the conditional allele might be because ~20% of the liver is composed of non-hepatocytic cells such as the macrophage Kupffer cells, stellate cells and endothelial cells [213]. These results together suggest that CXXC1 deletion occurs specifically in the liver within 5 weeks of birth (Figure 19) (This experiment was performed with the help of Erika Dobrota).

b. CFP1 protein expression is reduced in the livers of CXXC1$^{fl/fl}$ AlbCre+ mice

To determine if CFP1 protein levels were affected in CXXC1$^{fl/fl}$ AlbCre+ mice compared to control CXXC1$^{+/+}$ AlbCre+ mice whole cell extracts were prepared from liver tissues of these mice and western blotting was carried out using antiserum against CFP1. β- actin expression was used as the loading control. Results demonstrate that CFP1 protein expression is eliminated in CXXC1$^{fl/fl}$ AlbCre+ mice compared to that in CXXC1$^{+/+}$ AlbCre+ mice (Figure 20)

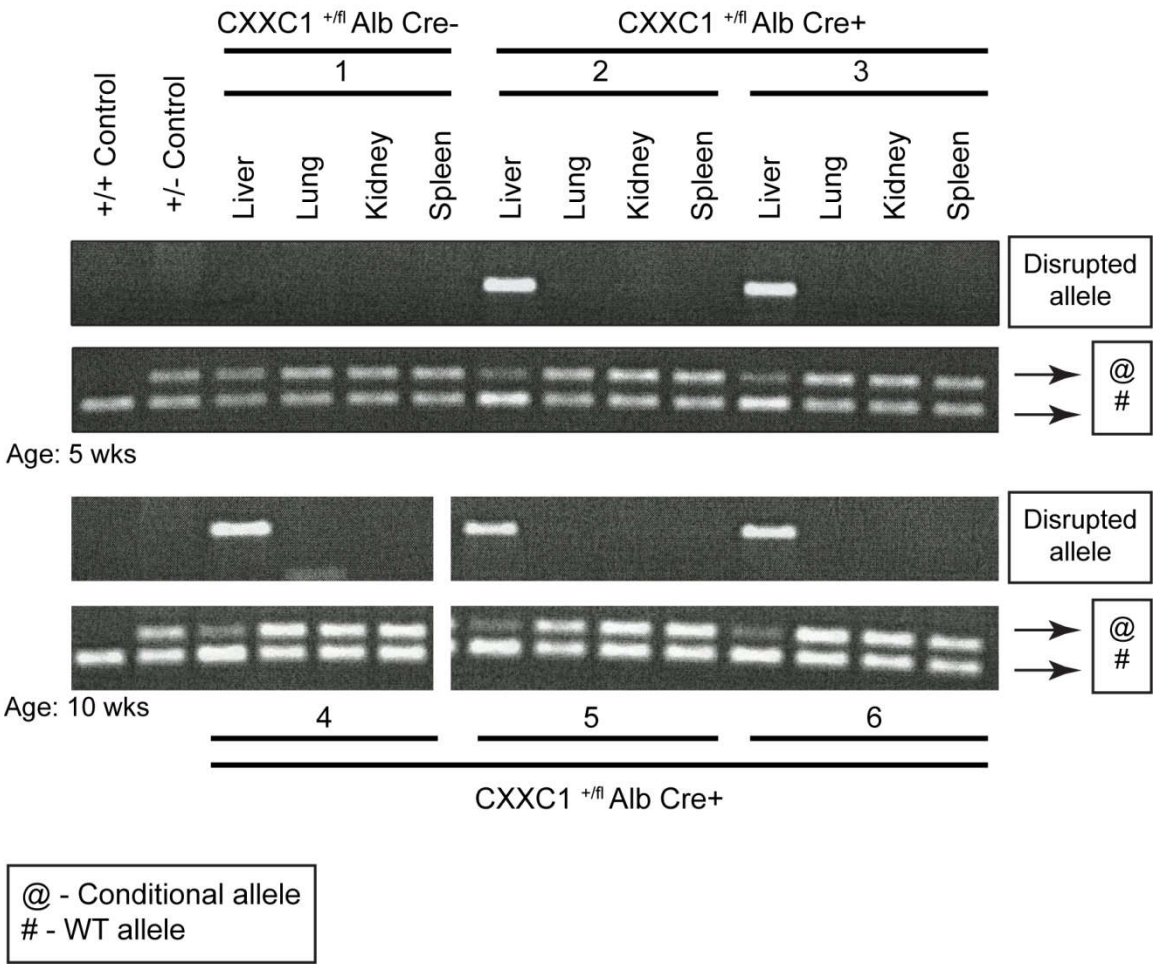


Figure 19: Confirmation of CFP1 depletion specifically in the liver at the age of 5 weeks

Following necropsy, liver, lung, kidney and spleen tissues were collected from 5 and 10 week old male CXXC1$^{+/fl}$ AlbCre+ and CXXC1$^{+/fl}$ AlbCre− mice. Genomic DNA isolated from these tissues was used in PCR reactions to (i) amplify the disrupted allele (top panels) using primers LOXP1F and LOXP3R and (ii) amplify the conditional allele (lower panels) using primers LOXP1F and CXXC1R (Refer figure 5A). One 5 week old CXXC1$^{+/fl}$ AlbCre− mouse (1), two 5 week old CXXC1$^{+/fl}$ AlbCre+ mice (2 and 3) and three 10 week old CXXC1$^{+/fl}$ AlbCre+ mice (4, 5 and 6) were used for this experiment. As shown in the figure, the top band in the doublet in the lower panels indicated by "@", corresponds to the conditional allele; whereas the lower band in the doublet, indicated by "#", corresponds to the wild type allele

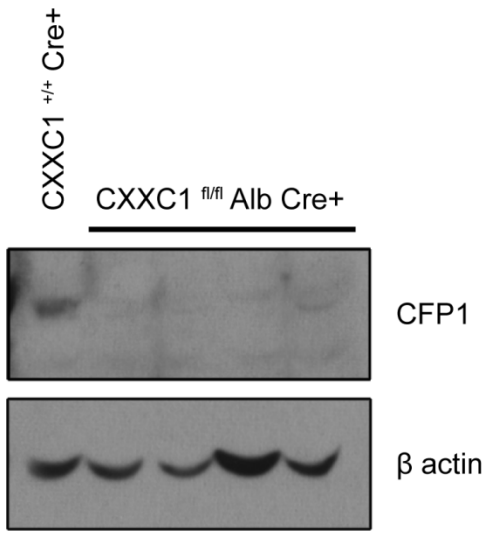


Figure 20: CFP1 protein expression is reduced in the livers of CXXC1$^{fl/fl}$ AlbCre+ mice

Whole cell extracts prepared from liver tissues of one CXXC1$^{+/+}$AlbCre+ and four CXXC1$^{fl/fl}$ AlbCre+ mice (ages 4-44 weeks) were subjected to western blot analysis. The representative blot was probed for CFP1 and β-actin as a loading control

c. Effects of CFP1 depletion in the mature, quiescent liver

Loss of CFP1 had no effect on the viability of CXXC1$^{fl/fl}$ AlbCre+ mice. There was no marked difference in the litter size of CXXC1$^{fl/fl}$ AlbCre+ mice compared to CXXC1$^{fl/fl}$ AlbCre− mice. However, ~40% of CXXC1$^{fl/fl}$ AlbCre+ mice become sick and die after weaning at various ages (4-52 weeks) (Figure 21) These sick mice exhibit signs of a wasting phenotype: smaller size, lack of vitality and hunched posture (Figure 22). H&E staining of formalin-fixed, paraffin embedded tissue sections of their livers reveal extensive areas of degenerating hepatocytes that appear to be attempting to regenerate (irregular foci of very large hepatocytes with abundant eosinophilic cytoplasm). These foci were larger in older experimental animals (~70 weeks of age) imparting a gross nodular appearance to the liver (Figure 23). Liver sections from older experimental mice also often display areas of fibrosis. Recently, an independent group reported that a combined loss of EZH1 and EZH2 (enzyme subunits responsible for H3K27me3 marks) in mouse hepatocytes resulted in the development of regenerative nodules and fibrosis at 8 months of age. This phenotype is very similar to that observed in CXXC1$^{fl/fl}$ AlbCre+ mice. Another observation was that the female CXXC1$^{fl/fl}$ AlbCre+ mice were unable to produce healthy and viable offspring. Eight breeding pairs consisting of female CXXC1$^{fl/fl}$ AlbCre+ and male CXXC1$^{fl/fl}$ AlbCre− mice were set up and monitored for a period of at least 6 months. Of the 8 female mice, only one was found to become pregnant (enlarged abdomen), but was unable to carry the pups to term. This female mouse suffered from complications during parturition. Hence, no viable pups were obtained from this breeding scheme.

CXXC1$^{-/-}$ ES cells exhibit increased (~40%) global H3K4me3 levels [161]. To examine if global H3K4me3 levels are altered in CXXC1$^{fl/fl}$ AlbCre+, histone proteins were extracted from liver tissues of these mice followed by western blot analysis for determining levels of global H3K4me3. It was found that there was a significant increase (~42%) in global H3K4me3 levels in CXXC1$^{fl/fl}$ AlbCre+ mice compared to CXXC1$^{fl/fl}$ AlbCre− mice, consistent with the observation in CXXC1$^{-/-}$ ES cells. (Figure 24A and B)

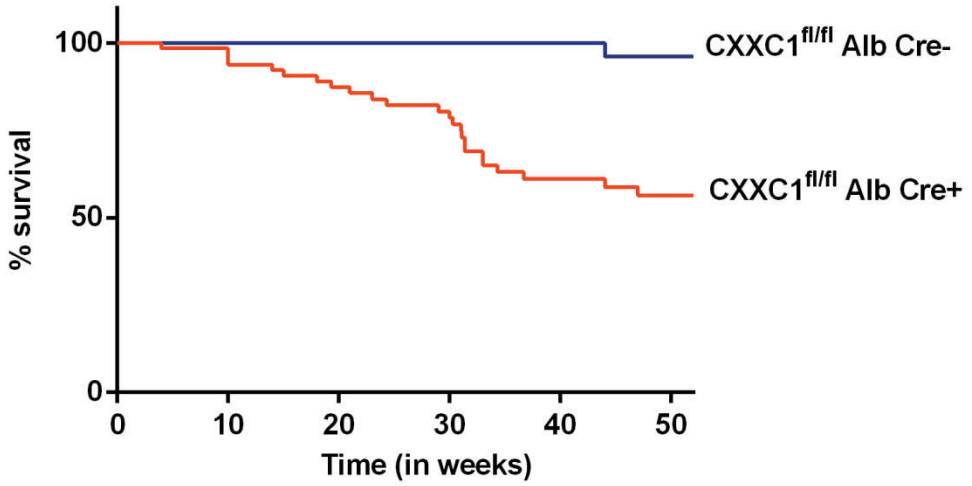


Figure 21: Loss of CFP1 results in premature death of CXXC1$^{fl/fl}$ AlbCre+ mice within one year

Percentage of survival of CXXC1$^{fl/fl}$ AlbCre+ mice and CXXC1$^{fl/fl}$ AlbCre− mice over a period of 52 weeks (~1 year) was studied by generating Kaplan-Meier survival curves using the software GraphPad Prism 6.

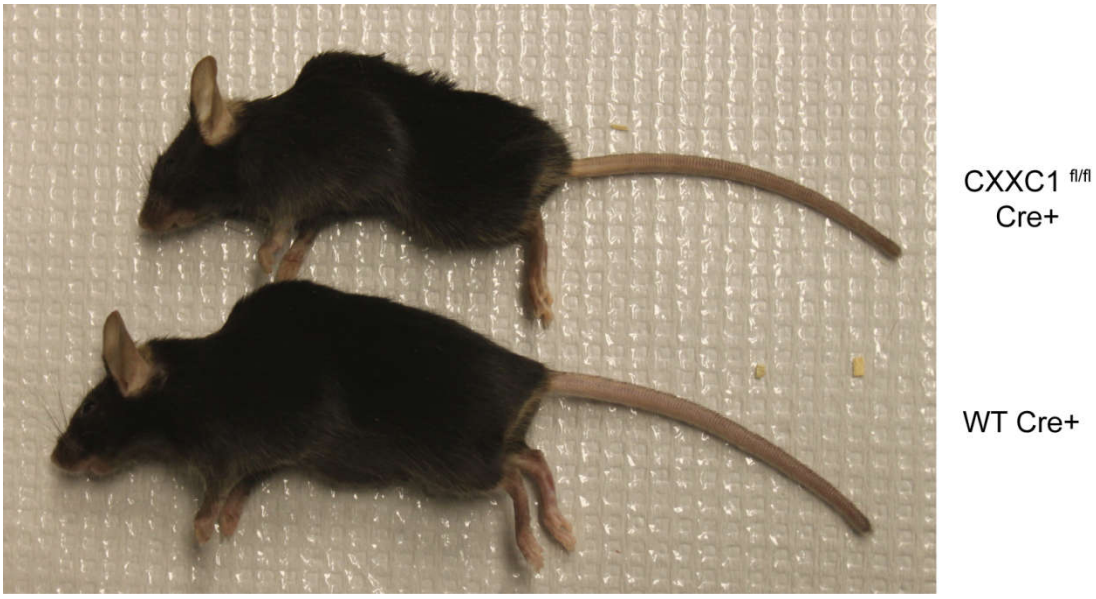


Figure 22: Sick CXXC1$^{fl/fl}$ AlbCre+ mice display a "wasting" phenotype

A photographic representation of the comparative sizes of a sick, moribund CXXC1$^{fl/fl}$ AlbCre+ and normal, healthy CXXC1$^{+/+}$ AlbCre+ mouse.

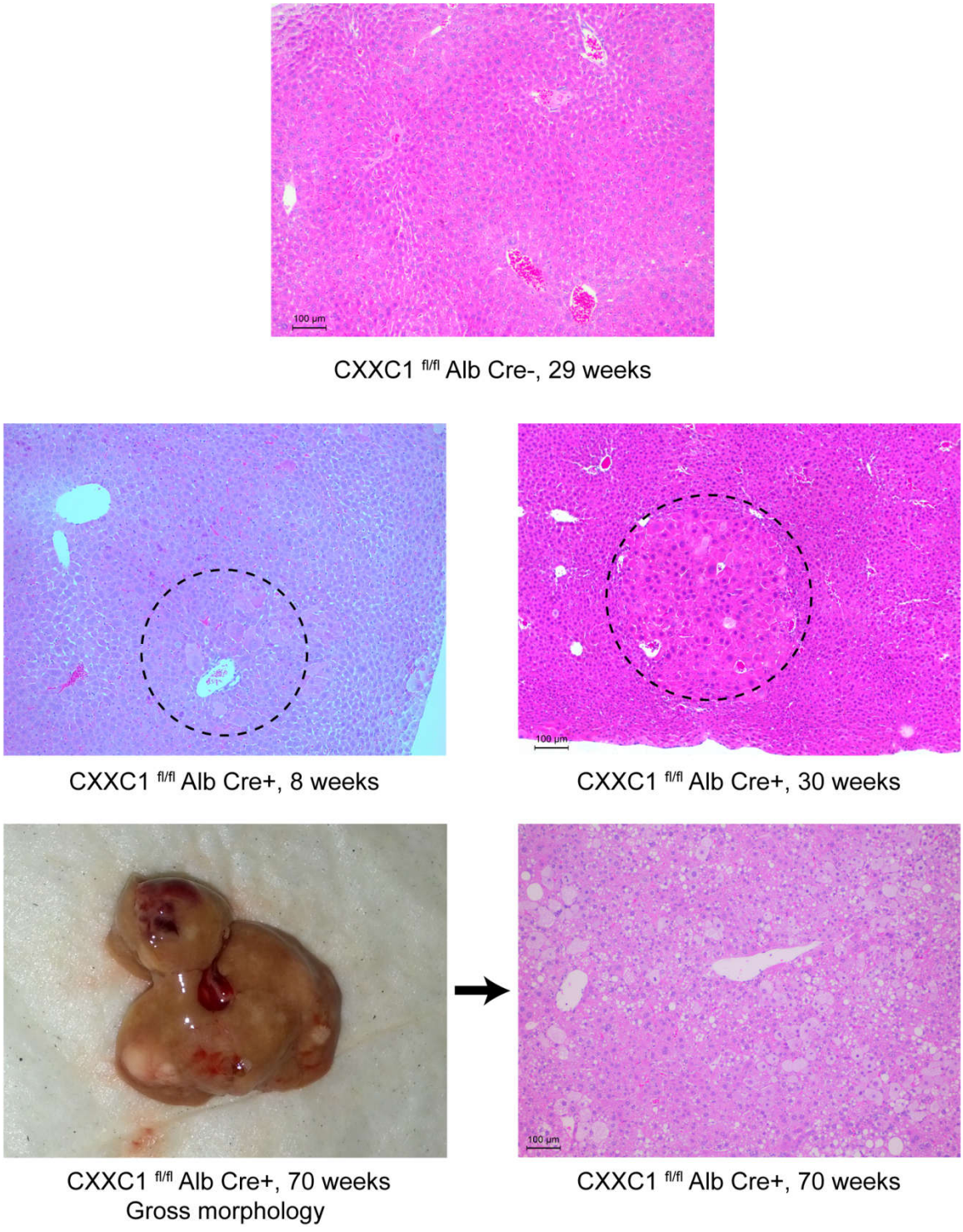


Figure 23: Older CXXC1$^{fl/fl}$ AlbCre+ mice often show the presence of regenerative liver nodules

Liver tissue collected from CXXC1$^{fl/fl}$ AlbCre+ and CXXC1$^{fl/fl}$ AlbCre− mice of the indicated ages were chemically fixed in 10% neutral buffered formalin, embedded in

paraffin and sectioned into 5 μm slides. These slides were subjected to the standard histopathological H&E staining to study the morphological characteristics of the liver tissue. Images were captured on a microscope at 10X magnification. Circles denote cells undergoing degeneration. The photograph on the bottom row shows the gross morphology of the liver excised from a representative 70 week old CXXC1$^{fl/fl}$ AlbCre+ mouse.

d. Effect of CFP1 depletion in the regenerating liver following partial hepatectomy

Hepatocytes in the mammalian liver are mitotically quiescent, highly differentiated cells with a very long half-life. When the liver undergoes injury, the remnant healthy hepatocytes enter the cell cycle to rapidly compensate for the injury. To illustrate the effects of loss of CFP1 in the regenerating liver, a surgical procedure called partial hepatectomy was performed to remove 2/3 of the liver mass in experimental ($CXXC1^{fl/fl}$ AlbCre+) and control ($CXXC1^{fl/fl}$ AlbCre−) male mice.

The gross liver sizes of the experimental and control mice appeared similar at the time of partial hepatectomy. These mice were sacrificed 7 days after surgery and the ratio of the mass of the remnant liver to total body mass was determined as a measure of regenerative response. Seven days after surgical resection, experimental mice aged 2-3 months show a statistically significant ~26% reduction in liver mass compared to age-matched control mice (Figure 25A). On the other hand, adult experimental mice aged 5-6 months showed a severe reduction (~40%) in liver mass compared to their control counterparts (Figure 25B). To ensure that the reduction in liver mass is not affected by an overall loss in body mass that often occurs after surgical procedures, the remnant liver mass was reported relative to the mass of the right kidney. There was a significant reduction (~34%) in the ratio of liver to right kidney mass in adult experimental mice compared to control mice (Figure 25C). These results collectively suggest that regeneration of murine liver after partial hepatectomy is impaired in the absence of CFP1.

Summary

This section provides a detailed analysis of the importance of CFP1 in liver homeostasis and regeneration after partial hepatectomy. The mouse line ($CXXC1^{fl/fl}$ AlbCre+) lacking CFP1 specifically in the liver was established and characterized. Around 40% of these mice exhibit a wasting phenotype followed by death at various ages. Livers of these mice often show the presence of regenerative nodules at senescence. Additionally, $CXXC1^{fl/fl}$ AlbCre+ mice show an increase in global H3K4me3 levels. Furthermore, $CXXC1^{fl/fl}$ AlbCre+ mice show a blunted regenerative response post partial hepatectomy suggesting that CFP1 is important for the process of liver regeneration.

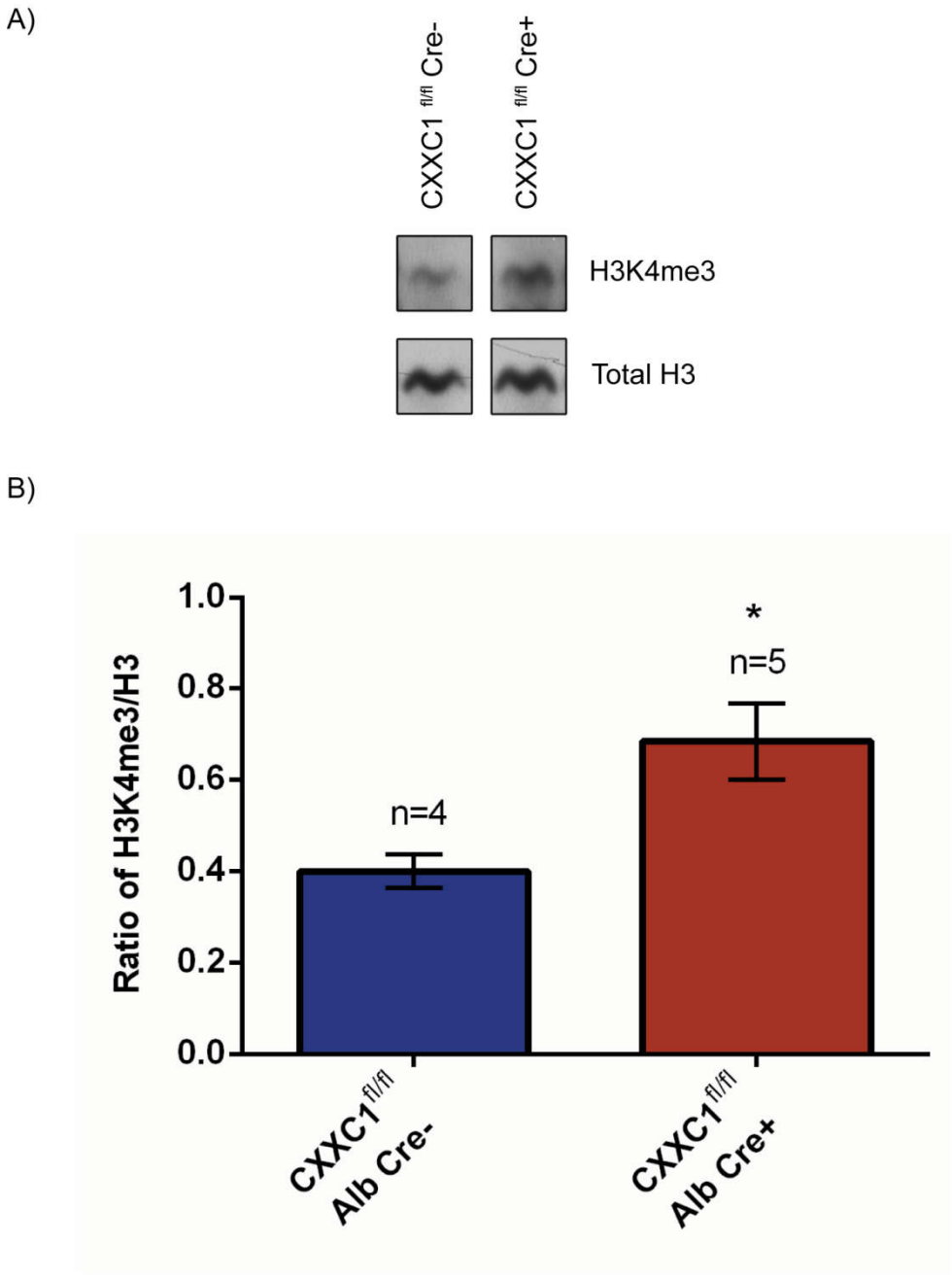


Figure 24: CXXC1$^{fl/fl}$ AlbCre+ mice show increased global H3K4me3 levels

A) Male mice that were aged 5-6 months were used for this experiment. Histone proteins were extracted from liver tissues of four CXXC1$^{fl/fl}$ AlbCre− and five CXXC1$^{fl/fl}$ AlbCre+ mice by acid extraction. These histone proteins were subjected to western blotting analysis using H3K4me3 antiserum. The membrane was re-probed for total H3 as

a loading control. Representative western blot results for one animal of each genotype ($CXXC1^{fl/fl}$ AlbCre− and $CXXC1^{fl/fl}$ AlbCre+) are shown.

B) The bar graph represents the quantitation of the average ± SE of H3K4me3 levels normalized to total H3 in $CXXC1^{fl/fl}$ AlbCre− (n=4) and $CXXC1^{fl/fl}$ AlbCre+ (n=5) mice. *$p<0.05$ represents a statistically significant difference compared to $CXXC1^{fl/fl}$ AlbCre− mice.

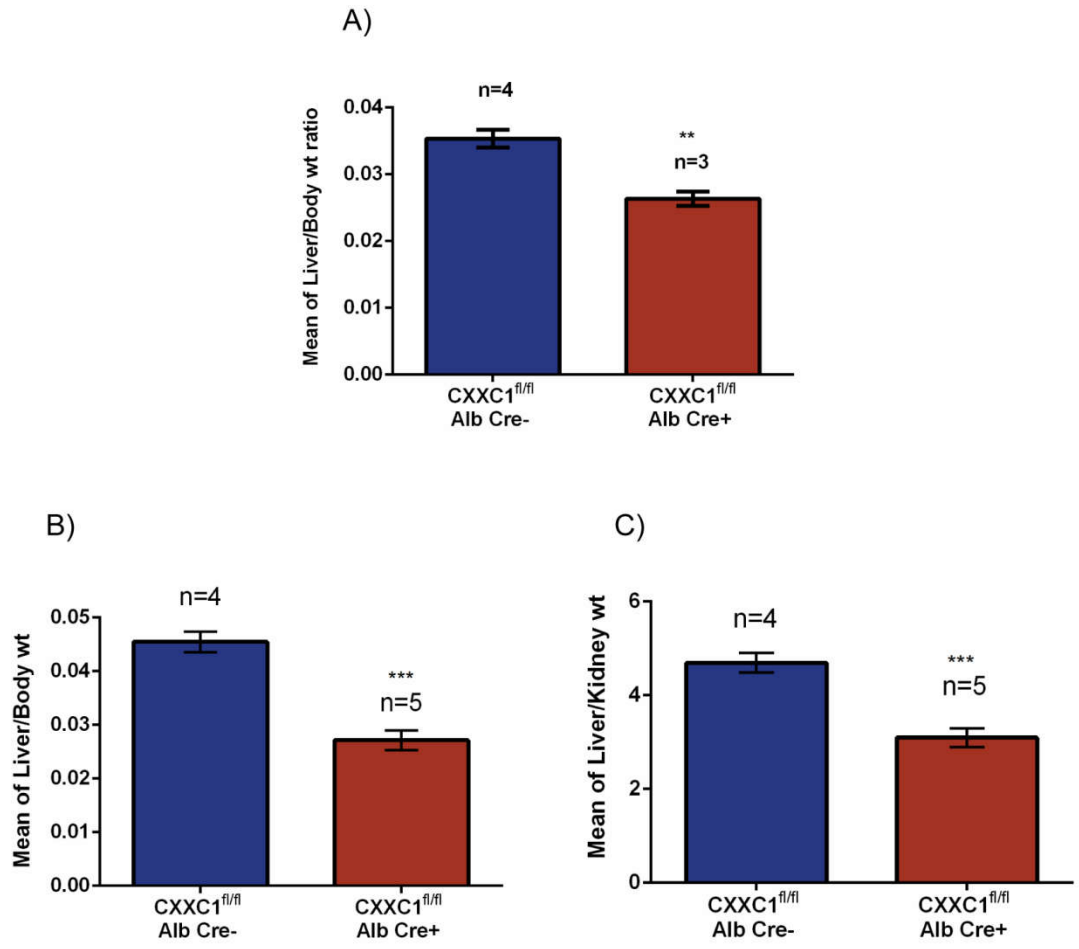


Figure 25: CFP1 depletion results in an impaired regenerative response to partial hepatectomy

A) Two to three month old CXXC1$^{fl/fl}$ AlbCre− (n=4) and CXXC1$^{fl/fl}$ AlbCre+ (n=3) mice were subjected to partial hepatectomy (surgical resection of 2/3 liver) and sacrificed seven days after surgery. Body mass and the mass of the remnant liver were noted. The bar graph represents the quantitation of the average ± SE of the remnant liver mass normalized to body mass in CXXC1$^{fl/fl}$ AlbCre− and CXXC1$^{fl/fl}$ AlbCre+ mice. **p<0.005 represents a statistically significant difference compared to CXXC1$^{fl/fl}$ AlbCre− mice.

B) Five to six month old CXXC1$^{fl/fl}$ AlbCre− (n=4) and CXXC1$^{fl/fl}$ AlbCre+ (n=5) mice were subjected to partial hepatectomy and sacrificed seven days after surgery. Body mass and the mass of the remnant liver were noted. The bar graph represents the quantitation of the average ± SE of the remnant liver mass normalized to body mass in

$CXXC1^{fl/fl}$ AlbCre− and $CXXC1^{fl/fl}$ AlbCre+ mice. ***$p<0.0005$ represents a statistically significant difference compared to $CXXC1^{fl/fl}$ AlbCre− mice.

C) Five to six month old $CXXC1^{fl/fl}$ AlbCre− (n=4) and $CXXC1^{fl/fl}$ AlbCre+ (n=5) mice were subjected to partial hepatectomy and sacrificed seven days after surgery. Mass of the right kidney and the remnant liver were noted. The bar graph represents the quantitation of the average ± SE of the remnant liver mass normalized to mass of the right kidney in $CXXC1^{fl/fl}$ AlbCre− and $CXXC1^{fl/fl}$ AlbCre+ mice. ***$p<0.0005$ represents a statistically significant difference compared to $CXXC1^{fl/fl}$ AlbCre− mice.

Discussion

1. CFP1 binds to unmodified and methylated histone H3K4 via its PHD domain

The studies presented in this part of the dissertation reveal that the N- terminal CFP1 PHD domain interacts with unmodified H3 (H3K4me0) and methylated histone H3K4 (H3K4me1/me2/me3). The human CFP1 PHD domain, expressed as a fusion with GST in bacterial cells, was able to recognize both unmodified H3 and methylated H3K4 histone tail peptides suggesting that these interactions were direct. Moreover, CFP1 PHD domain was unable to recognize H3K9me3, a heterochromatin marker, suggesting that this interaction was specific to the panel of histone-peptide tails examined in this work. Transiently expressed full length Flag-tagged CFP1, immunoprecipitated from the nuclear extract, binds to endogenous H3K4me3, suggesting that the interaction between CFP1 and H3K4me3 occurs *in vivo* within the nuclei of cells. Thus CFP1, a component of the euchromatin-specific Set1 histone methyltransferase complexes that catalyze the deposition of H3K4me marks, is involved in interacting with the same mark. Moreover, it is possible that methylation of H3K9 results in the blocking of the binding of H3K4me3 to CFP1 and this could explain how a repressive mark (H3K9me) prevents CFP1 recruitment. The loss of CFP1 results in a drop of H3K4me3 levels at CGIs and this defect is rescued by expression of full-length CFP1 [163, 181]. This suggests that CFP1 might participate in the reading of H3K4me marks and thereby facilitate a feedforward mechanism that aids the deposition of these marks at CGIs to maintain active chromatin status. Interestingly, many histone methyltransferase complexes adopt similar regulatory mechanisms for mediating activation or repression of specific genomic loci. For instance, WDR5, a common component of the MLL1, MLL2 and hSet1 complexes, functions as an effector protein by binding to H3K4-dimethylated nucleosomes and facilitating global and gene specific H3K4 trimethylation [214]. Another example is of HP1 that not only associates with H3K9me3 methyltransferase SUV39H1, but also recognizes the methylated H3K9 mark through its amino-terminal chromodomain. This phenomenon is thought to promote the establishment and spread of heterochromatin domains [215, 216]. Pygo2, a developmentally important protein, consists of a PHD domain that specifically recognizes H3K4me2/3 and this interaction is critical to mammary progenitor cell expansion. Additionally, Pygo2 also facilitates the generation of these marks at Wnt-

specific loci and in bulk chromatin thereby suggesting that a positive feedback loop might be in place for the establishment of active chromatin domains [130].

The residues corresponding to Y28, D44 and W49 of CFP1 PHD are conserved in many H3K4me3-binding proteins like BPTF, PHF2 and ING2 [125, 128, 131]. Of these, Y28 and W49 of CFP1 PHD correspond to Y10 and W32 of the BPTF PHD and are involved in formation of the aromatic amino acid cage that securely positions the trimethyl group of H3K4me3 [125]. Corresponding amino acids Y7/W29, present in PHF2 and Y215/W238, present in ING2 establish contacts with the trimethylammonium group of lysine 4 [131, 217]. The point mutations Y28A, D44A and W49A in CFP1 PHD and full length CFP1 resulted in ablation of their association with H3K4me3. This suggests that only the PHD domain of CFP1 is involved in binding to methylated H3K4. Moreover, it is also likely that the tri-methyl group of H3K4me3 is stabilized by aromatic residues Y28 and W49, similar to that in other known H3K4me3-binding proteins.

PHD domains of proteins capable of binding unmodified H3, usually lack an aromatic cage [124]. It is also found that methylation of H3K4 disrupts the interaction of these PHDs with unmodified H3 [124, 218]. However, CFP1 PHD was found to interact with both the unmodified and methylated forms of H3K4 efficiently. It can be surmised that the binding of CFP1, to the substrate unmodified H3, aids the initial docking of the Set1 enzyme complexes at the transcription start sites of active genes. This study revealed that point mutations of several conserved and charged amino acids were unable to abrogate the binding between CFP1 PHD and unmodified H3, suggesting that the amino acid residues critical for this binding activity might be other than the ones examined in this report. It might also be possible that point mutations within CFP1 PHD are unable to disrupt the overall tertiary structure of this domain to ablate its interaction with unmodified H3. The N- terminal half or the C-terminal half of CFP1 PHD is not sufficient to interact with unmodified H3, suggesting that this binding might be governed by the combinatorial participation of residues in both halves of the domain. Further investigation is required to identify a point mutation that ablates the interaction of CFP1 PHD to unmodified H3. Therefore, the functional significance of the interaction of CFP1 PHD with unmodified H3 still remains elusive.

2. Interaction of CFP1 PHD with methylated H3K4 is dispensable for appropriate H3K4me3 targeting

CFP1 restricts the sub-nuclear localization of H3K4me3 to euchromatin [179]. This study examined the importance of the binding of CFP1 to methylated H3K4 for the regulation of the targeting of these marks to DAPI-dim euchromatin. The loss of the binding of CFP1 to methylated H3K4 had no adverse effect on the targeting of H3K4me3 marks to euchromatin. However, it is possible that a CFP1 mutant that ablates binding to both unmodified H3 and methylated H3K4 will not be able to rescue appropriate genomic targeting of H3K4me3 to euchromatic regions. CFP1 regulates H3K4me3 levels at many CGI rich promoters of active genes [181]. It was also observed that the DNA binding function of CFP1 is not necessary for targeting of H3K4me3 to CGIs [181]. Therefore, it will be important to determine if the binding of CFP1 to unmodified H3 and/or to methylated H3K4 is required for the restoration of H3K4me3 levels at CGI rich promoters.

Many reports have linked H3K4me3 levels at transcription start sites with another euchromatin specific mark, histone H3 acetylation (H3K9ac and H3K14ac) [219-221]. The Gcn5 subunit of the SAGA complex that catalyzes the acetylation of histone H3 (H3K9ac, H3K14ac), can be targeted to the H3K4me3 nucleosomes via the adaptor subunit Sgf29 [21] [22] [222, 223]. Another group recently found that the lack of CFP1 also results in a reduction in H3K9ac and H3K14ac, along with a reduction in H3K4me3 levels at the transcription start sites of active genes in murine ES cells. [223]. It is possible that binding of CFP1 to unmodified and/or to methylated H3K4 is important for recruitment of Gcn5 to H3K4me3 nucleosomes to aid the accumulation of H3K9ac and H3K14ac marks at transcription start sites of active genes.

3. Interaction of CFP1 with methylated H3K4 or DNA is required for efficient differentiation of ES cells

This portion of the study explored the role of the interaction between CFP1 and H3K4me3 during the process of ES cell differentiation. Previous data indicates that either the DNA binding activity of CFP1 or interaction with the Set1 complexes is necessary for *in vitro* differentiation of ES cells, suggesting that the CXXC and SID domains of CFP1 are

redundant functional domains [180]. This study revealed that ablation of the interaction between CFP1 and methylated H3K4 does not adversely affect cellular differentiation. However, a combined disruption of both the H3K4me-binding and the DNA-binding activities of CFP1 resulted in compromised cellular differentiation. This suggests that neither the PHD nor the CXXC domains are required for *in vitro* differentiation and that the PHD and CXXC domains are redundant for this function. On induction of differentiation, embryoid bodies of CXXC1$^{-/-}$ ES cells expressing CFP1 containing point mutations in both PHD and CXXC domains fail to form extensive outgrowths and approximately 41% of these cells are unable to down regulate alkaline phosphatase (AP) activity. This indicates that approximately 41% of the ES cells containing mutated PHD and CXXC domains of CFP1 are unable to undergo lineage commitment.

RNA isolated from embryoid bodies expressing mutant PHD and CXXC domains of CFP1 was subjected to RT-PCR analysis of lineage and stem cell markers. When a population of mutant cells was examined, it was observed that markers of lineage commitment including c-fms, MHC-β and β-H1 were expressed along with the stem cell marker Oct4. Combined with the observations from the AP staining assay, it can be surmised that the ~59% AP-negative cells are responsible for the up-regulation of lineage development markers and can down-regulate Oct4 simultaneously. However, the drop in the levels of Oct4 is not sharp presumably due to the remaining ~41% cells that might be unable to down-regulate Oct4 along with AP. Oct4 is a critical factor required to maintain the pluripotency and self-renewal properties of murine ES cells [224]. Down regulation of Oct4 is essential for *in vivo* and *in vitro* differentiation of ES cells into the three germ layers; endoderm, ectoderm and mesoderm [225]. Persistence of Oct4 expression after induction of differentiation suggests that that the self-renewal capacity of ~41% of the mutant ES cell population remains unaffected and that this fraction of the population is unable to efficiently execute the differentiation program. One of the mechanisms of down-regulation of Oct4 during the process of differentiation requires the methylation of Oct4 promoter [226]. It is possible that the inability of CXXC1$^{-/-}$ ES cells expressing W49A, C169A CFP1 (1-656) to achieve *in vitro* differentiation is influenced by the defective methylation status of the Oct4 promoter.

The observation reported in this study that only a portion of the population of CXXC1$^{-/-}$ ES expressing W49A, C169A (1-656) remains undifferentiated 10 days after induction of differentiation could also be because of a delay in the differentiation program. A similar delay in differentiation is observed in the absence of Tcf3 in murine ES cells [227]. Tcf proteins are DNA binding transcriptional regulators that participate in the Wnt signaling pathway. Tcf3 is the most abundant Tcf factor in ESCs and the loss of Tcf3 by genetic ablation in ESCs results in a delay in differentiation without affecting their self-renewing properties. The authors found that, while Tcf3$^{+/+}$ ES cells expressed lineage specific markers including GATA4 and GATA6 as early as 3 days after induction of differentiation and retained their expression throughout the duration of the assay (8 days), Tcf3$^{-/-}$ ES cells, on the other hand, expressed low levels of lineage-specific markers even after 8 days of differentiation. Additionally, levels of a stem cell marker LRH-1 were down regulated in Tcf3$^{+/+}$ ES cells, but remained fairly persistent in Tcf3$^{-/-}$ ES cells after 8 days of differentiation. However, this did not prevent differentiated Tcf3$^{-/-}$ cell types from occurring during *in vitro* differentiation [227]. This defect was found to be due to the deregulation of the Tcf3's repressor activity on the Nanog promoter. Similarly, it is possible that the down-regulation of stem cell markers including Oct4 and AP is delayed in ES cells expressing PHD and CXXC domain mutations. Therefore, the ability of CFP1 to interact with methylated H3K4 and DNA might be required for the repression of other key genes governing cell fate following the induction of cellular differentiation.

Previous data revealed that either the N-terminal half (1-367aa) or the C-terminal half (361-656 aa) of CFP1 is sufficient for the rescue of *in vitro* differentiation in CXXC1$^{-/-}$ ES cells. However, the data reported in this dissertation suggests that either the PHD domain or the CXXC domain is required to rescue defective ES cell differentiation. Both of these domains are present in the N-terminal half of CFP1. The fact that the C-terminal half of CFP1 (not containing PHD or CXXC domain) completely rescues cellular differentiation could be possibly due to folding differences in truncated proteins resulting in differential exposure of key epitopes as compared to full length proteins. Another possibility could be that the C-terminal half of CFP1 lacks an uncharacterized 'inhibitory' domain that could inhibit the rescue of various functions.

The differentiation defect of the CXXC1$^{-/-}$ ES cells expressing W49A, C169A (1-656) is subtle as compared to the defect observed in CXXC1$^{-/-}$ ES cells. It is possible that the ability of the mutant CFP1 protein to rescue some pleiotropic mechanisms involved in the process of differentiation, is because of the presence of intact acidic, basic, SID, coiled-coil and HEAT domains of CFP1.

It could be argued that the differentiation defect observed in approximately 41% of the CXXC1$^{-/-}$ ES cells expressing W49A, C169A (1-656) (persistent AP activity) might be because of cellular variation in the levels of CFP1 protein. However, this is unlikely because stable ES cell clones resistant to hygromycin were individually isolated and expanded. Another possibility for the presence of a mixture of AP positive and AP negative cells upon differentiation could be because of the architecture of the embryoid bodies. Cells that form the exterior of the embryoid body are epithelial-like and display cell-cell tight junctions and form a dense layer of extra cellular matrix along with cells at the periphery as compared to the interior of the embryoid body [228]. It is possible that intact CFP1 PHD and CXXC domains are required differentially for regulating the interior and exterior microenvironment of embryoid bodies.

4. The histone binding property of CFP1 might be important for the resolution of bivalent promoters into active promoters upon ES cell differentiation

In mammalian cells, genomic domains are marked by distinct histone modifications. For example, H3K4me3 and H3K9/H3K27me3 are primarily found on active and repressed promoters respectively whereas active enhancers are marked by H3K4me1 and H3K27ac [229]. However, pluripotent embryonic stem cells contain a large group of gene promoters that display a co-occurrence of both activating (H3K4me3) and repressing (H3K27me3) marks [155, 230]. These bivalent domains are thought to maintain developmentally important genes in the “poised state”, enabling their rapid activation on receiving suitable developmental signals. These bivalent domains track with the unmethylated CpG islands that form a part of ~70% of all promoters in ES cells [7, 230]. CFP1 consists of a CXXC domain that specifically binds unmethylated CpGs [162]. Recent studies indicate that loss of CFP1 adversely affects H3K4me3 levels on the promoters of actively transcribed genes; however H3K4me3 levels on bivalent promoters are less

affected [181, 231]. This may suggest that recruitment of MLL1/2 might be more prevalent in these domains.

MLL2 consists of a CXXC domain that interacts with unmethylated CpGs [232]. There is evidence that the MLL2 enzymes mediate H3K4 trimethylation on bivalent gene promoters [233]. On induction of differentiation, most bivalent promoters lose one of the marks (H3K4me3 or H3K27me3), resulting in activation or repression of specific genes. However, the loss of MLL2 did not affect retinoic acid induced gene expression of bivalent promoters [233]. Thus, it is possible that the existing H3K4me3 marks on bivalent promoters bind CFP1 PHD and recruit the Set1 complex and result in a feedforward mechanism for deposition of more H3K4me3 marks thereby playing an active role in resolving bivalent promoters into active gene promoters enriched with H3K4me3 marks.

5. The proper targeting of Set1/H3K4me3 at gene promoters is regulated by many mechanisms

Appropriate targeting of the Set1 complex is thought to be coordinated by a number of mechanisms. Set1 complexes consist of two components Wdr82 and CFP1 that are lacking in the other COMPASS like complexes (MLL1-4) [63, 64, 161]. Wdr82 interacts with the C-terminal domain (CTD) of RNA polymerase II and recruits the Set1 complex to the transcriptional start sites of active genes [65]. Moreover, RNAi-mediated knockdown of Wdr82 results in a global reduction of H3K4 trimethylation with almost no effect on mono/di-methylation levels of H3K4, suggesting that Wdr82 is critical for H3K4 trimethylation [63]. Further, monoubiquitination of histone H2B has been found to be required for association of the yeast analog of Wdr82 (Cps35) to chromatin and the subsequent trimethylation of H3K4 [38]. Another study revealed that the Paf1 complex physically associates with the yeast Set1/COMPASS complex and recruits the COMPASS complex to the elongating form of RNA polymerase II [86].

The catalytic Set1A and Set1B proteins exhibit an overall sequence identity of 39% and a similarity of 56%. However, these proteins consist of an extensive central region that exhibits only 29% sequence identity and 41% similarity. Moreover, Set1a and Set1b are

found to have non-overlapping distributions within euchromatin regions as observed by confocal immunofluorescence [64]. Therefore, it can be expected that the central divergent domains of Set1A and Set1B confer distinct and non-redundant genomic targeting properties to these enzyme complexes thereby contributing to the complex nature of Set1 genomic targeting.

CGIs consisting of unmethylated CpGs have been shown to bind the CXXC domain of CFP1 and recruit the Set1 complexes to these loci for the deposition of H3K4me3. Artificial exogenous CpG rich insertions are sufficient for this targeting function [163]. Deletion of CFP1 in mouse ESCs results in the depletion of H3K4me3 from the CGI-rich promoters of actively transcribed genes with the simultaneous appearance of ectopic H3K4me3 at many regulatory regions. Expression of a CFP1 mutant lacking DNA binding activity results in the rescue of H3K4me3 levels at promoters but does not prevent aberrant H3K4me3 accumulation at the distal regulatory elements, suggesting that the DNA binding function of CFP1 is not required for the targeting of Set1 to the promoter regions [181]. The domain of CFP1 that mediates the targeting function still remains unidentified. The work done in this part of the dissertation suggests that CFP1 exhibits binding specificity to both the unmodified and the methylated form of H3K4. It is possible that CFP1 recruits the Set1 complex to its substrate unmodified H3 and facilitates the deposition of H3K4me3 marks at the promoter regions. Binding of CFP1 to H3K4me1/me2/me3, the product of the Set1-mediated methylation may additionally be required for targeting of the Set1 complex to active promoters having low levels of H3K4me3 in the absence of CFP1 [181]. It is therefore apparent that the molecular mechanisms involved in Set1 targeting are convoluted and further investigation is required to fully appreciate these complexities.

6. CFP1 is required for the appropriate epigenetic control of liver homeostasis

This portion of the study examined the role of CFP1 in murine liver homeostasis and regeneration. This involved the generation of a transgenic mouse line lacking CFP1 specifically in the liver. These mice carried the conditional (floxed) allele for the CXXC1 gene and the Cre-recombinase gene under the control of the liver-specific albumin gene promoter. Expression of albumin begins early (stage 7 to 8 somites) in murine

development [202, 234]. However, when the AlbCre transgene was used in combination with the conditional glucokinase gene allele, efficiency of recombination was found to be progressive with age, appearing to be complete by 6 weeks of age [202]. Previous research has indicated that the fetal liver is primarily a hematopoietic organ with a small population of hepatocytes [235]. However, at the perinatal stages of gestation, the function of hematopoiesis is undertaken by the bone marrow and majority of the liver is populated by hepatocytes, thus conferring hepatic functions to the liver [236]. Use of a marker that converts from red to green fluorescence following Cre-recombination revealed that the progressive increase in the efficiency of Cre-recombination is due to a progressive reduction in the proportion of non-hepatocytic cells in the liver, coupled with an increase in hepatocyte population [237]. This suggests that the AlbCre transgene is a useful tool for probing the functions of genes at various stages of liver development. Examination of AlbCre mediated recombination of the conditional CXXC1 allele revealed that deletion of the CXXC1 gene occurred at five weeks of age. Additionally, CFP1 protein levels were markedly reduced by the age of 4-5 weeks.

As observed in CXXC1$^{-/-}$ ES cells, global H3K4me3 levels in the mutant livers were significantly higher. This suggests that the loss of CFP1 results in aberrant activation and/or targeting of H3K4 methylases. It is also possible that the Set1/H3K4me3 targeting defect observed in CXXC1$^{-/-}$ ES cells [179] is also reflected in mutant livers lacking CFP1.

CFP1 physically interacts with DNMT1 and facilitates global cytosine methylation [168]. ES cells lacking CFP1 show reduced global cytosine methylation levels [177]. However, following depletion of CFP1, differentiated hematopoietic cells in adult mice do not show a decline in global cytosine methylation levels [176]. Hence, although adult livers lacking CFP1 were not examined for global cytosine methylation defects in this study, it is surmised that these livers may not display a defect in global cytosine methylation levels.

Around 40% of the mutant mice succumbed to a “wasting” phenotype within a year in the absence of liver CFP1. HNF4 α is a protein important for the morphological and functional differentiation of hepatocytes [238]. Depletion of HNF4α in the murine liver

resulted in an overall loss of body weight followed by death of ~70% of the mutant mice by 8 weeks of age [239]. The wasting phenotype associated with mice lacking liver CFP1, although subtle, is similar to that observed in mice lacking HNF4α in the liver. It possible that CFP1 and HNF4α interact to facilitate appropriate expression of genes required for liver homeostasis. Alternatively, lack of CFP1 might result in the down regulation of HNF4α expression, thus resulting in a wasting phenotype similar to that observed in HNF4α-depleted mice. Another possibility is that the loss of CFP1 results in epigenomic stress due to altered histone modifications, resulting in deregulation of liver homeostasis. Moreover, assessment of H&E stained liver sections of mutant mice often revealed large areas of hepatocytes undergoing apoptosis and degeneration. These hepatocytes have eosinophilic cytoplasm suggesting that they are prone to degenerate (communicated by Dr. HogenEsch). ES cells lacking CFP1 also undergo increased apoptosis and exhibit decreased global protein translation [178, 180]. Although the exact molecular mechanism for this phenomenon is not clear, the increased apoptosis could be attributed to the global translation defect [178] that might result in aberrant expression of pro-apoptotic proteins (Bad, Bak, Bid, Bax) or anti apoptotic proteins (Bcl-2 or Bcl Xl) in livers lacking CFP1.

The livers from older mutant mice lacking CFP1 formed regenerative nodules. A similar phenotype of the development of regenerative nodules has been reported in mice lacking EZH1 and EZH2 (enzymes catalyzing deposition of H3K27me3), suggesting that histone lysine methyltransferases play an important role in the maintenance of liver homeostasis [198].

Previous studies in the lab have elucidated the role of CFP1 in developmentally important cellular processes such as gastrulation and hematopoiesis [172, 173, 176]. However, this is the first study that suggests the participation of CFP1 in the maintenance of tissue homeostasis in an adult animal. This is supported by the fact that CFP1 is ubiquitously expressed in mature tissues [240]. However, at the same time, the role of CFP1 in the epigenetic regulation of tissue homeostasis is also intriguing because mature tissues do not undergo active chromatin remodeling like that occurring during the process of embryonic development.

7. CFP1 is critical for the process of murine liver regeneration

Liver regeneration is a highly coordinated process that entails ~95% of terminally differentiated hepatocytes to re-enter the cell cycle and reach peak DNA synthesis at 36 hours. Restoration of liver mass occurs after 1-2 rounds of cell divisions within the next 5-7 days. Genome-wide expression analysis in many cell lines have revealed that the promoters of most cell cycle-related genes are enriched with unmethylated CpGs and the active chromatin mark H3K4me3 [241]. It is possible that CFP1 binds to the unmethylated CpGs at these sites thereby recruiting the Set1 complex. Thus, epigenetic regulation might be critical for the process of liver regeneration. This study shows that CFP1-depleted livers fail to fully compensate the loss of liver mass suffered during partial hepatectomy. This suggests that CFP1 is important for the proper control of liver regeneration. Although our experimental design did not include a no surgery or sham surgery control group of age-matched male mice, the gross liver sizes of the experimental and control mice appeared similar before parial hepatectomy. The H3K27 methyltransferases, EZH1 and EZH2 are also known to play a crucial role in liver regeneration. Loss of EZH1/2 impairs regenerative response to liver injury due to partial hepatectomy and CCl_4 treatment [198]. It will be interesting to delineate the possible crosstalk between H3K4-and H3K27-methylases that might be a prerequisite for establishing proper epigenetic control required for liver regeneration.

There is very little known about the role played by various chromatin modifying enzymes in the epigenetic regulation of liver regeneration. This is the first study to our knowledge that illustrates the role played by a component of a H3K4-methylase in the regulation of liver regeneration. Further investigation is required to clearly elucidate the molecular mechanisms involved in this regulatory process. The knowledge of the molecular bases of liver regeneration in rodents can be extended to understand this process in humans to delineate the pathologies in which liver regeneration is impaired and ultimately provide new treatment options for patients suffering from liver damage. Nevertheless, this study provides preliminary information about the critical role played by CFP1 and the Set1 complexes in liver regeneration.

Future Directions

This dissertation focusses on understanding the role of CFP1 in two distinct systems. The first part examines the importance of the histone binding function of CFP1 PHD for appropriate targeting of H3K4me3 and during the process of *in vitro* differentiation in murine ES cells. The second part delineates the significance of CFP1 in liver homeostasis and regeneration. Further investigation is required for a clearer understanding of these functions of CFP1.

In vitro binding assays indicated that CFP1 PHD interacts with unmodified H3, although the inability to find point mutations that would ablate binding of CFP1 PHD to unmodified H3 prevented us from assessing the functional role of this interaction. In order to precisely determine the molecular basis of the read-out of unmodified H3 by the CFP1 PHD finger, crystal structures of CFP1 PHD in the free-state and in complex with H3 (1-21aa) can be determined. The crystals of the CFP1 PHD-unmodified H3 complex can be grown using the hanging drop vapor diffusion method. The crystals obtained can be analyzed for their diffraction properties [217]. This can provide information about the residues of CFP1 PHD and the unmodified histone H3 tail that make contacts with each other, thereby allowing subsequent mutagenesis for studying the functional significance of this binding event.

The mechanism and affinity of the binding of CFP1 PHD to unmodified H3 can additionally or alternatively be studied by isothermal titration calorimetry. Isothermal titration calorimetry can provide information about the number of peptide binding sites in CFP1 PHD and this might be important to generate a CFP1 mutant that is devoid of histone binding properties. Calorimetric experiments can be conducted using MCS-ITC instrument from MicroCal. The CFP1 PHD finger and lyophilized unmodified H3 peptides dialyzed in the same dialysis buffer can be used for titration [125, 217].

After identification of critical residues for binding unmodified H3 and understanding the binding mechanism, using the above techniques, point mutations of the critical residues can be generated by site-directed mutagenesis. These point mutations, most likely, can

also be expected to ablate binding with methylated H3K4. Subsequently, stable CXXC1$^{-/-}$ ES cells expressing mutant CFP1 (mutCFP1) can be isolated. These stable clones can further be used to understand the role of CFP1-histone binding function for the regulation of H3K4me3 levels at the transcription start sites of actively transcribed genes. Chromatin immunoprecipitation (ChIP) followed by qPCR can be performed to assess the levels of H3K4me3 at the transcription start sites of constitutively active genes such as Actb, Gapdh and Rpl3. Additionally, ChIP-qPCR can be performed to examine the levels of Set1 and CFP1 at the promoter regions of selected active genes. Further, analysis of genome-wide Set1 and H3K4me3 distribution in ES cells expressing mutCFP1 can be carried out by ChIP coupled to massively parallel DNA sequencing. Moreover, CXXC1$^{-/-}$ ES cells expressing mutCFP1 (mutant that is unable to bind histones) may exhibit a wide variety of defects in the rescue of functions including cytosine methylation, histone methylation, in vitro differentiation, DNA damage sensitivity, population doubling time and plating efficiency. These experiments collectively can provide a thorough understanding of the functional significance of CFP1-histone interaction.

It is currently unknown if loss of interaction of CFP1 with methylated H3K4 [W49A CFP1 (1-656)] or loss of interaction of CFP1 with methylated H3K4 and DNA together [W49A, C169A CFP1 (1-656)] affects the genome-wide levels of CFP1, Set1 and H3K4me3. Expression of C169A CFP1 in CXXC1$^{-/-}$ ES cells resulted in rescue of H3K4me3 levels at active promoters, but retained the inappropriate accumulation of H3K4me3 at regulatory regions outside CGIs [181]. This suggests that CXXC1$^{-/-}$ ES cells expressing W49A, C169A CFP1 may also display the defect of ectopic H3K4me3 accumulation apart from other defects that may be observed due to ablation of CFP1-H3K4me binding. Previous data indicates that the various mutations and truncations of CFP1 that do not rescue histone methylation and cytosine methylation levels are also unable to rescue *in vitro* differentiation of ES cells [180]. Based on this, it may be expected that CXXC1$^{-/-}$ ES cells expressing W49A, C169A CFP1 (1-656) might have aberrant histone and cytosine methylation levels. However, the extent of this aberration may be lower because a majority of the cells were able to undergo lineage commitment.

Ablation of binding of CFP1 to methylated H3K4 or DNA did not affect the rescue of *in vitro* cellular differentiation in ES cells. However, ablation of both these binding activities together resulted in compromised cellular differentiation. It is yet to be determined if ablation of binding of CFP1 to methylated H3K4 (W49A) and Set1 (C375A) affects the differentiation potential of ES cells. To understand this, site-directed mutagenesis can be performed to introduce the point mutations W49A and C375A into CFP1 (1-656). Further, this mutated form of CFP1 can be stably expressed in CXXC1$^{-/-}$ ES cells followed by induction of differentiation by growth in ES media lacking LIF. ChIP-qPCR followed by ChIP-Seq can be also be performed to assess the genome-wide levels of CFP1, Set1 and H3K4me3 in these mutant cells.

The second part of this dissertation examines the role of CFP1 in murine liver homeostasis and regeneration. There are many unexplored avenues to further the current level of understanding of this newly established role of CFP1. This study evaluated the effect of depletion of CFP1 in liver regeneration at the end of 7 days when the restoration of the lost liver mass is expected to be completed in the murine liver [186]. However, the defects in liver regeneration observed in the absence of CFP1 can be examined in more detail by performing a time-course study. Age matched, male CXXC1$^{fl/fl}$ AlbCre+ (experimental) and CXXC1$^{fl/fl}$ AlbCre− (control) mice undergoing partial hepatectomy or sham surgery can be sacrificed at 1, 2, 3 and 7 days following surgery. Blood can be drawn from the heart under isofluorane anesthesia [198]. Liver and body mass can be determined after euthanasia followed by preparation of formalin-fixed, paraffin embedded liver sections. The effect of the loss of CFP1 on liver integrity and function can be assessed by examining the levels of serum markers such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), and albumin; both in mature, quiescent and regenerating livers lacking CFP1 at various ages. Increase in serum levels of ALT and AST is a sensitive indicator of altered hepatocyte integrity whereas an increase in ALP levels signifies biliary obstruction. Furthermore, elevated serum albumin levels indicate an overall impairment of liver function [198].

Adult liver tissue consists of quiescent cells that remain in the G0 phase of the cell cycle. Upon liver regeneration following liver injury, these quiescent cells enter the cell cycle to

replenish the mass of the lost liver. The extent of proliferation of hepatocytes in the regenerating liver can be assessed by staining liver sections for Ki-67, a cellular marker of proliferation and appearance of mitotic figures.

It is yet to be determined if mRNA and protein expression of CFP1 is up/down regulated in wild type regenerating livers during the process of liver regeneration. Quantification and localization of CFP1 mRNA can be assessed by *in situ* hybridization (RNA ISH) on liver tissue sections whereas protein levels of CFP1 can be determined by western blotting using antiserum against CFP1. Additionally, examination of global levels of histone H3K4 and cytosine methylation in the regenerating liver can provide insight into the epigenetic dysregulation occurring in the absence of CFP1.

A genome-wide perspective of the alteration of H3K4me3 and DNA cytosine methylation marks can be achieved by performing high throughput ChIP-Seq on the hepatocytes derived from the experimental and control mice using antibodies against H3K4me3, 5-methyl cytosine or IgG control. Viable, primary hepatocytes can be isolated using a collagenase digestion method [242] followed by crosslinking with 1% formaldehyde for 10 min. Chromatin obtained from these hepatocytes can be sonicated to 200-300 bp. Further, the immunoprecipitated DNA can be blunt ended and ligated to Illumina indexed DNA adaptors (Illumina Inc, San Diego, CA) and sequenced using the Illumina HiSeq 2000 followed by downstream bioinformatic data analysis. This approach can provide valuable information regarding the aberrations found in the overall distribution of H3K4me3 and 5-methyl cytosine marks in the CFP1-deficient livers.

Additionally, RNA-sequencing (RNA-Seq) can be performed for a comparative analysis of the liver transcriptomes of the experimental and control mice. Briefly, following the isolation of total RNA from liver tissues of the experimental and control mice, cDNA can be synthesized using the Superscript (Invitrogen) and TruSeq RNA sample preparation kit (Illumina). This cDNA can be sequenced using HiSeq 2000 (Illumina). This experiment will provide information about the possible differences in gene expression patterns between the experimental and control livers. In combination with the ChIP-seq experiment describe earlier, it can be assessed if the genome-wide distribution of

H3K4me3 and 5-methylcytosine marks correlate with the gene expression patterns observed in CFP1-depleted livers. These studies will eventually be useful to identify key genes that undergo genetic and epigenetic modulation during liver homeostasis and/or regeneration. Furthermore, these high throughput experiments can also provide clues about the possible epigenetic cross talk mechanisms that may be required for supporting liver homeostasis and regeneration.

A small cohort of older experimental mice displayed fibrosis of liver tissue. To assess the extent of fibrosis, Masson's trichrome staining for collagen can be performed. Examination of the CFP1-deficient liver transcriptome can reveal the up-regulation of genes associated with fibrosis.

Carbon tetrachloride (CCl4)-induced liver injury is another widely used model for investigation of liver regeneration. Unlike the surgical process of partial hepatectomy, this mode of damage can be easily induced by 4 sub-cutaneous injections of CCl4 (2ml/kg, 2 times a week) [198]. The difference in the rate of proliferation of hepatocytes between experimental and control mice can be ascertained by Ki-67 staining at various time intervals. It is likely that CFP1-deficient mice will be unable to undergo compensatory regeneration and survive the CCl4 mediated toxicity.

The mammalian pancreatic tissue also possesses partial regenerative properties. Acinar cells, the most abundant pancreatic cell type, are the exocrine cells that produce and transport enzymes that assist with digestion of food and are capable of regeneration following pancreatectomy. Epigenetic mechanisms governing acinar cell regeneration are largely under-studied. The importance of CFP1 for acinar cell regeneration can be assessed by generating a mouse line lacking CFP1 in acinar cells by breeding $CXXC1^{fl/fl}$ mice with ElastaseCreERT2 mice. The ElastaseCreERT2 mice express Cre recombinase (Cre) as a fusion with a mutant estrogen receptor ligand-binding domain (ERT2) that is selectively responsive to tamoxifen [243]. The expression of CreERT2 is under the control of the strong and highly acinar cell specific elastase I promoter [243]. This will enable the deletion of CFP1 specifically in the acinar cell compartment. Acinar cells are known to undergo transdifferentiation into insulin-secreting β-cells *in vitro* [244]. A better

understanding of the epigenetic control mediated by CFP1 in acinar cell regeneration can be of immense value to further the utilization of these cells as surrogate insulin-secreting β-cells for the treatment of Type 1 diabetes.

The liver of the pregnant mouse undergoes dramatic changes during pregnancy. A two fold increase in liver weight is observed from the non-pregnant state to day 18 of pregnancy [193]. The features of liver growth occurring during pregnancy overlap with those observed during liver regeneration after partial hepatectomy [193]. In this study, it was observed that female CXXC1$^{fl/fl}$ AlbCre+ mice were unable to produce healthy and viable offspring. It is possible that lack of CFP1 results in the impairment of maternal liver growth occurring during pregnancy, thus resulting in abortions. This phenomenon can be studied in detail by setting up timed pregnancies in these mice. Both ovulation and mating typically occur during estrus, which occur after the onset of the dark cycle (absence of visible light) and lasts for 12-14 hours. Female mice can be housed in large groups in the absence of male mice in the vicinity. This results in the prolongation of diestrus and suppression of estrus due to the absence of male urinary pheromones (Whitten effect). Once the female mice are housed with male CXXC1$^{fl/fl}$ AlbCre− mice, estrous cycle will resume within 3 days. Vaginal plugs can be examined the following morning to check if mating has occurred. If the plugged female mouse becomes pregnant, the day after the plug is found is considered as the first day of gestation [245]. The pregnant female CXXC1$^{fl/fl}$ AlbCre+ mice can be euthanized at various stages of pregnancy to evaluate the change in liver mass and to identify the stage of pregnancy when abortions occurred. These experiments can potentially uncover additional roles that CFP1 might play in the regulation of liver growth during pregnancy.

Summary

The data presented in this dissertation describes the histone binding properties of the epigenetic regulator CFP1. It was observed that CFP1 directly binds to methylated H3K4 via its PHD finger. Results suggest that the interaction of CFP1 PHD with methylated H3K4 is not required for appropriate subnuclear localization of endogenous H3K4me3 and for the ability of ES cells to differentiate *in vitro*. However, studies revealed that introduction of a point mutation (C169A) that ablates the binding of CFP1 with DNA and a point mutation (W49A) that ablates binding of CFP1 with methylated H3K4 results in compromised cellular differentiation. These results indicate that the retention of either the CXXC or PHD domain is sufficient for the rescue of *in vitro* differentiation of ES cells, suggesting that these domains are redundant for this function. Moreover, CFP1 PHD also binds directly to unmodified H3. In order to study the functional significance of this interaction, point mutations for several conserved and charged amino acids were introduced into CFP1 PHD, however, binding of CFP1 PHD to unmodified H3 remained unaffected.

This dissertation also examines the importance of CFP1 in regulation of homeostasis in the quiescent adult murine liver. To this end, a mouse line was developed that carried the conditional (floxed) allele of CXXC1 and the Cre-recombinase gene under the control of the liver specific albumin promoter. The activation of Cre-recombinase resulted in depletion of CFP1 specifically in the liver by the age of 5 weeks. This model was used to study the role of CFP1 in long-term liver tissue homeostasis and regulation of regenerative response. Studies revealed that ~40% of the population of CFP1-deficient mice succumbed to a wasting phenotype within a year. Also, livers of CFP1-deficient mice often developed regenerative nodules and displayed altered global H3K4me3 levels. To understand the role of CFP1 in the process of liver regeneration, CFP1-deficient mice were subjected to partial hepatectomy. Comparison of the residual liver masses 7 days after the procedure revealed a severe impairment of compensatory liver growth in CFP1-deficient mice. This suggests that CFP1 is involved in the epigenetic regulation of liver regeneration. A clear understanding of the role of CFP1 in the molecular control of liver regeneration may prove useful for delineating pathologies involving impairment of liver

regeneration and ultimately provide new and better options for therapeutic intervention for patients with liver damage.

References

1. Bonasio, R., S. Tu, and D. Reinberg, *Molecular signals of epigenetic states.* Science, 2010. **330**(6004): p. 612-6.
2. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-71.
3. Bestor, T.H., *The DNA methyltransferases of mammals.* Hum Mol Genet, 2000. **9**(16): p. 2395-402.
4. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
5. Sved, J. and A. Bird, *The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4692-6.
6. Lindahl, T., P. Karran, and R.D. Wood, *DNA excision repair pathways.* Curr Opin Genet Dev, 1997. **7**(2): p. 158-69.
7. Deaton, A.M. and A. Bird, *CpG islands and the regulation of transcription.* Genes Dev, 2011. **25**(10): p. 1010-22.
8. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.* Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1412-7.
9. Lan, J., et al., *DNA methyltransferases and methyl-binding proteins of mammals.* Acta Biochim Biophys Sin (Shanghai), 2010. **42**(4): p. 243-52.
10. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.
11. Egger, G., et al., *Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival.* Proc Natl Acad Sci U S A, 2006. **103**(38): p. 14080-5.
12. Riggs, A.D. and Z. Xiong, *Methylation and epigenetic fidelity.* Proc Natl Acad Sci U S A, 2004. **101**(1): p. 4-5.
13. Jin, B., Y. Li, and K.D. Robertson, *DNA methylation: superior or subordinate in the epigenetic hierarchy?* Genes Cancer, 2011. **2**(6): p. 607-17.

14. Jones, P.L., et al., *Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription.* Nat Genet, 1998. **19**(2): p. 187-91.
15. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.* Nature, 1998. **393**(6683): p. 386-9.
16. Roth, S.Y., J.M. Denu, and C.D. Allis, *Histone acetyltransferases.* Annu Rev Biochem, 2001. **70**: p. 81-120.
17. Fischle, W., Y. Wang, and C.D. Allis, *Histone and chromatin cross-talk.* Curr Opin Cell Biol, 2003. **15**(2): p. 172-83.
18. Lee, D.Y., et al., *A positive role for histone acetylation in transcription factor access to nucleosomal DNA.* Cell, 1993. **72**(1): p. 73-84.
19. Grant, P.A., *A tale of histone modifications.* Genome Biol, 2001. **2**(4): p. REVIEWS0003.
20. Hong, L., et al., *Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA.* J Biol Chem, 1993. **268**(1): p. 305-14.
21. Grant, P.A., et al., *Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex.* Genes Dev, 1997. **11**(13): p. 1640-50.
22. Brownell, J.E., et al., *Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation.* Cell, 1996. **84**(6): p. 843-51.
23. Grant, P.A. and S.L. Berger, *Histone acetyltransferase complexes.* Semin Cell Dev Biol, 1999. **10**(2): p. 169-77.
24. Hahn, S., *The role of TAFs in RNA polymerase II transcription.* Cell, 1998. **95**(5): p. 579-82.
25. Delcuve, G.P., D.H. Khan, and J.R. Davie, *Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors.* Clin Epigenetics, 2012. **4**(1): p. 5.
26. de Ruijter, A.J., et al., *Histone deacetylases (HDACs): characterization of the classical HDAC family.* Biochem J, 2003. **370**(Pt 3): p. 737-49.
27. Yang, X.J. and E. Seto, *Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression.* Curr Opin Genet Dev, 2003. **13**(2): p. 143-53.

28. Gallinari, P., et al., *HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics.* Cell Res, 2007. **17**(3): p. 195-211.
29. Weake, V.M. and J.L. Workman, *Histone ubiquitination: triggering gene activity.* Mol Cell, 2008. **29**(6): p. 653-63.
30. Wang, H., et al., *Role of histone H2A ubiquitination in Polycomb silencing.* Nature, 2004. **431**(7010): p. 873-8.
31. Nakagawa, T., et al., *Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation.* Genes Dev, 2008. **22**(1): p. 37-49.
32. de Napoles, M., et al., *Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation.* Dev Cell, 2004. **7**(5): p. 663-76.
33. Fang, J., et al., *Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation.* J Biol Chem, 2004. **279**(51): p. 52812-5.
34. Zhou, W., et al., *Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation.* Mol Cell, 2008. **29**(1): p. 69-80.
35. Kim, J., S.B. Hake, and R.G. Roeder, *The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions.* Mol Cell, 2005. **20**(5): p. 759-70.
36. Zhu, B., et al., *Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation.* Mol Cell, 2005. **20**(4): p. 601-11.
37. Shahbazian, M.D., K. Zhang, and M. Grunstein, *Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1.* Mol Cell, 2005. **19**(2): p. 271-7.
38. Lee, J.S., et al., *Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS.* Cell, 2007. **131**(6): p. 1084-96.
39. Seeler, J.S. and A. Dejean, *Nuclear and unclear functions of SUMO.* Nat Rev Mol Cell Biol, 2003. **4**(9): p. 690-9.

40. Nathan, D., et al., *Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications.* Genes Dev, 2006. **20**(8): p. 966-76.
41. Shiio, Y. and R.N. Eisenman, *Histone sumoylation is associated with transcriptional repression.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13225-30.
42. Rogakou, E.P., et al., *Megabase chromatin domains involved in DNA double-strand breaks in vivo.* J Cell Biol, 1999. **146**(5): p. 905-16.
43. Iacovoni, J.S., et al., *High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome.* EMBO J, 2010. **29**(8): p. 1446-57.
44. Rossetto, D., et al., *Epigenetic modifications in double-strand break DNA damage signaling and repair.* Clin Cancer Res, 2010. **16**(18): p. 4543-52.
45. Stucki, M., et al., *MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks.* Cell, 2005. **123**(7): p. 1213-26.
46. Cheung, W.L., et al., *Phosphorylation of histone H4 serine 1 during DNA damage requires casein kinase II in S. cerevisiae.* Curr Biol, 2005. **15**(7): p. 656-60.
47. Rossetto, D., N. Avvakumov, and J. Cote, *Histone phosphorylation: a chromatin modification involved in diverse nuclear events.* Epigenetics, 2012. **7**(10): p. 1098-108.
48. Choi, H.S., et al., *Phosphorylation of histone H3 at serine 10 is indispensable for neoplastic cell transformation.* Cancer Res, 2005. **65**(13): p. 5818-27.
49. Lau, A.T., et al., *Phosphorylation of histone H2B serine 32 is linked to cell transformation.* J Biol Chem, 2011. **286**(30): p. 26628-37.
50. Wei, Y., et al., *Phosphorylation of histone H3 is required for proper chromosome condensation and segregation.* Cell, 1999. **97**(1): p. 99-109.
51. Gurley, L.R., et al., *Heterochromatin and histone phosphorylation.* Exp Cell Res, 1978. **111**(2): p. 373-83.
52. Lo, W.S., et al., *Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14.* Mol Cell, 2000. **5**(6): p. 917-26.
53. Lee, E.R., et al., *Dynamic changes in histone H3 phosphoacetylation during early embryonic stem cell differentiation are directly mediated by mitogen- and stress-*

activated protein kinase 1 via activation of MAPK pathways. J Biol Chem, 2006. **281**(30): p. 21162-72.

54. Lau, P.N. and P. Cheung, *Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2801-6.
55. Strahl, B.D., et al., *Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1.* Curr Biol, 2001. **11**(12): p. 996-1000.
56. Bedford, M.T. and S. Richard, *Arginine methylation an emerging regulator of protein function.* Mol Cell, 2005. **18**(3): p. 263-72.
57. Fabbrizio, E., et al., *Negative regulation of transcription by the type II arginine methyltransferase PRMT5.* EMBO Rep, 2002. **3**(7): p. 641-5.
58. Di Lorenzo, A. and M.T. Bedford, *Histone arginine methylation.* FEBS Lett, 2011. **585**(13): p. 2024-31.
59. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation.* Genes Dev, 2011. **25**(13): p. 1345-58.
60. Schneider, J., et al., *Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression.* Mol Cell, 2005. **19**(6): p. 849-56.
61. Miller, T., et al., *COMPASS: a complex of proteins associated with a trithorax-related SET domain protein.* Proc Natl Acad Sci U S A, 2001. **98**(23): p. 12902-7.
62. Herz, H.M., A. Garruss, and A. Shilatifard, *SET for life: biochemical activities and biological functions of SET domain-containing proteins.* Trends Biochem Sci, 2013. **38**(12): p. 621-39.
63. Wu, M., et al., *Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS.* Mol Cell Biol, 2008. **28**(24): p. 7337-44.
64. Lee, J.H., et al., *Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex.* J Biol Chem, 2007. **282**(18): p. 13419-28.
65. Lee, J.H. and D.G. Skalnik, *Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes.* Mol Cell Biol, 2008. **28**(2): p. 609-18.
66. Salz, T., et al., *hSETD1A regulates Wnt target genes and controls tumor growth of colorectal cancer cells.* Cancer Res, 2014. **74**(3): p. 775-86.

67. Lee, J.H. and D.G. Skalnik, *Rbm15-Mkl1 interacts with the Setd1b histone H3-Lys4 methyltransferase via a SPOC domain that is required for cytokine-independent proliferation.* PLoS One, 2012. **7**(8): p. e42965.
68. Bledau, A.S., et al., *The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation.* Development, 2014. **141**(5): p. 1022-35.
69. Briggs, S.D., et al., *Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae.* Genes Dev, 2001. **15**(24): p. 3286-95.
70. Nislow, C., E. Ray, and L. Pillus, *SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes.* Mol Biol Cell, 1997. **8**(12): p. 2421-36.
71. Yu, B.D., et al., *Altered Hox expression and segmental identity in Mll-mutant mice.* Nature, 1995. **378**(6556): p. 505-8.
72. Ayton, P.M. and M.L. Cleary, *Transformation of myeloid progenitors by MLL oncoproteins is dependent on Hoxa7 and Hoxa9.* Genes Dev, 2003. **17**(18): p. 2298-307.
73. Wang, P., et al., *Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II.* Mol Cell Biol, 2009. **29**(22): p. 6074-85.
74. Andreu-Vieyra, C.V., et al., *MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing.* PLoS Biol, 2010. **8**(8).
75. Herz, H.M., et al., *Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4.* Genes Dev, 2012. **26**(23): p. 2604-20.
76. Martin, C. and Y. Zhang, *The diverse functions of histone lysine methylation.* Nat Rev Mol Cell Biol, 2005. **6**(11): p. 838-49.
77. Schneider, R., et al., *Histone H3 lysine 4 methylation patterns in higher eukaryotic genes.* Nat Cell Biol, 2004. **6**(1): p. 73-7.

78. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nat Rev Genet, 2011. **12**(1): p. 7-18.
79. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.
80. Strahl, B.D., et al., *Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression.* Mol Cell Biol, 2002. **22**(5): p. 1298-306.
81. Sims, R.J., 3rd and D. Reinberg, *Processing the H3K36me3 signature.* Nat Genet, 2009. **41**(3): p. 270-1.
82. Xiao, T., et al., *Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast.* Genes Dev, 2003. **17**(5): p. 654-63.
83. Feng, Q., et al., *Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain.* Curr Biol, 2002. **12**(12): p. 1052-8.
84. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.
85. Briggs, S.D., et al., *Gene silencing: trans-histone regulatory pathway in chromatin.* Nature, 2002. **418**(6897): p. 498.
86. Krogan, N.J., et al., *The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation.* Mol Cell, 2003. **11**(3): p. 721-9.
87. Yuan, P., et al., *Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells.* Genes Dev, 2009. **23**(21): p. 2507-20.
88. Yeap, L.S., K. Hayashi, and M.A. Surani, *ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage.* Epigenetics Chromatin, 2009. **2**(1): p. 12.
89. Peters, A.H., et al., *Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability.* Cell, 2001. **107**(3): p. 323-37.
90. Tachibana, M., et al., *G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis.* Genes Dev, 2002. **16**(14): p. 1779-91.

91. Montgomery, N.D., et al., *The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation.* Curr Biol, 2005. **15**(10): p. 942-7.
92. Kim, J. and H. Kim, *Recruitment and biological consequences of histone modification of H3K27me3 and H3K9me3.* ILAR J, 2012. **53**(3-4): p. 232-9.
93. Shen, X., et al., *EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency.* Mol Cell, 2008. **32**(4): p. 491-502.
94. Cao, R. and Y. Zhang, *The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3.* Curr Opin Genet Dev, 2004. **14**(2): p. 155-64.
95. Yang, H. and C.A. Mizzen, *The multiple facets of histone H4-lysine 20 methylation.* Biochem Cell Biol, 2009. **87**(1): p. 151-61.
96. Sanders, S.L., et al., *Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage.* Cell, 2004. **119**(5): p. 603-14.
97. Schotta, G., et al., *A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin.* Genes Dev, 2004. **18**(11): p. 1251-62.
98. Jorgensen, S., G. Schotta, and C.S. Sorensen, *Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity.* Nucleic Acids Res, 2013. **41**(5): p. 2797-806.
99. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1.* Cell, 2004. **119**(7): p. 941-53.
100. Lee, M.G., et al., *An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation.* Nature, 2005. **437**(7057): p. 432-5.
101. Tsukada, Y., et al., *Histone demethylation by a family of JmjC domain-containing proteins.* Nature, 2006. **439**(7078): p. 811-6.
102. Klose, R.J. and Y. Zhang, *Regulation of histone methylation by demethylimination and demethylation.* Nat Rev Mol Cell Biol, 2007. **8**(4): p. 307-18.
103. Liang, G., et al., *Yeast Jhd2p is a histone H3 Lys4 trimethyl demethylase.* Nat Struct Mol Biol, 2007. **14**(3): p. 243-5.
104. Lee, N., et al., *The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase.* Nat Struct Mol Biol, 2007. **14**(4): p. 341-3.
105. Musselman, C.A., et al., *Perceiving the epigenetic landscape through histone readers.* Nat Struct Mol Biol, 2012. **19**(12): p. 1218-27.

106. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.
107. Cheung, P., et al., *Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation.* Mol Cell, 2000. **5**(6): p. 905-15.
108. Dover, J., et al., *Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6.* J Biol Chem, 2002. **277**(32): p. 28368-71.
109. Wang, H., et al., *Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor.* Science, 2001. **293**(5531): p. 853-7.
110. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
111. Ruthenburg, A.J., C.D. Allis, and J. Wysocka, *Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark.* Mol Cell, 2007. **25**(1): p. 15-30.
112. Yun, M., et al., *Readers of histone modifications.* Cell Res, 2011. **21**(4): p. 564-78.
113. Sanchez, R. and M.M. Zhou, *The role of human bromodomains in chromatin biology and gene transcription.* Curr Opin Drug Discov Devel, 2009. **12**(5): p. 659-65.
114. Owen, D.J., et al., *The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p.* EMBO J, 2000. **19**(22): p. 6141-9.
115. Jacobson, R.H., et al., *Structure and function of a human TAFII250 double bromodomain module.* Science, 2000. **288**(5470): p. 1422-5.
116. Macdonald, N., et al., *Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3.* Mol Cell, 2005. **20**(2): p. 199-211.
117. Zhao, Q., et al., *PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing.* Nat Struct Mol Biol, 2009. **16**(3): p. 304-11.
118. Yang, Y., et al., *TDRD3 is an effector molecule for arginine-methylated histone marks.* Mol Cell, 2010. **40**(6): p. 1016-23.
119. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-4.

120. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-20.
121. Lange, M., et al., *Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex.* Genes Dev, 2008. **22**(17): p. 2370-84.
122. Zeng, L., et al., *Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b.* Nature, 2010. **466**(7303): p. 258-62.
123. Aasland, R., T.J. Gibson, and A.F. Stewart, *The PHD finger: implications for chromatin-mediated transcriptional regulation.* Trends Biochem Sci, 1995. **20**(2): p. 56-9.
124. Sanchez, R. and M.M. Zhou, *The PHD finger: a versatile epigenome reader.* Trends Biochem Sci, 2011. **36**(7): p. 364-72.
125. Li, H., et al., *Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF.* Nature, 2006. **442**(7098): p. 91-5.
126. Wysocka, J., et al., *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.
127. Soliman, M.A. and K. Riabowol, *After a decade of study-ING, a PHD for a versatile family of proteins.* Trends Biochem Sci, 2007. **32**(11): p. 509-19.
128. Shi, X., et al., *ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression.* Nature, 2006. **442**(7098): p. 96-9.
129. Li, B., et al., *Developmental phenotypes and reduced Wnt signaling in mice deficient for pygopus 2.* Genesis, 2007. **45**(5): p. 318-25.
130. Gu, B., et al., *Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation.* J Cell Biol, 2009. **185**(5): p. 811-26.
131. Wen, H., et al., *Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation.* J Biol Chem, 2010. **285**(13): p. 9322-6.
132. Vermeulen, M., et al., *Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4.* Cell, 2007. **131**(1): p. 58-69.
133. Liu, Y., et al., *A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement.* Immunity, 2007. **27**(4): p. 561-71.

134. Matthews, A.G., et al., *RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination.* Nature, 2007. **450**(7172): p. 1106-10.
135. Lan, F., et al., *Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression.* Nature, 2007. **448**(7154): p. 718-22.
136. Chakravarty, S., L. Zeng, and M.M. Zhou, *Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator.* Structure, 2009. **17**(5): p. 670-9.
137. Ooi, S.K., et al., *DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.* Nature, 2007. **448**(7154): p. 714-7.
138. Campos, E.I., et al., *Biological functions of the ING family tumor suppressors.* Cell Mol Life Sci, 2004. **61**(19-20): p. 2597-613.
139. Baker, L.A., C.D. Allis, and G.G. Wang, *PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks.* Mutat Res, 2008. **647**(1-2): p. 3-12.
140. Musselman, C.A. and T.G. Kutateladze, *PHD fingers: epigenetic effectors and potential drug targets.* Mol Interv, 2009. **9**(6): p. 314-23.
141. Org, T., et al., *The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression.* EMBO Rep, 2008. **9**(4): p. 370-6.
142. Schwarz, K., et al., *RAG mutations in human B cell-negative SCID.* Science, 1996. **274**(5284): p. 97-9.
143. Villa, A., et al., *V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations.* Blood, 2001. **97**(1): p. 81-8.
144. Omenn, G.S., *Familial Reticuloendotheliosis with Eosinophilia.* N Engl J Med, 1965. **273**: p. 427-32.
145. Sobacchi, C., et al., *RAG-dependent primary immunodeficiencies.* Hum Mutat, 2006. **27**(12): p. 1174-84.
146. Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.
147. Howlett, S.K. and W. Reik, *Methylation levels of maternal and paternal genomes during preimplantation development.* Development, 1991. **113**(1): p. 119-27.
148. Rougier, N., et al., *Chromosome methylation patterns during mammalian preimplantation development.* Genes Dev, 1998. **12**(14): p. 2108-13.

149. Li, E., *Chromatin modification and epigenetic reprogramming in mammalian development.* Nat Rev Genet, 2002. **3**(9): p. 662-73.
150. Plath, K., et al., *Xist RNA and the mechanism of X chromosome inactivation.* Annu Rev Genet, 2002. **36**: p. 233-78.
151. Csankovszki, G., A. Nagy, and R. Jaenisch, *Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation.* J Cell Biol, 2001. **153**(4): p. 773-84.
152. Murray, P. and D. Edgar, *The regulation of embryonic stem cell differentiation by leukaemia inhibitory factor (LIF).* Differentiation, 2001. **68**(4-5): p. 227-34.
153. Park, S.H., et al., *Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells.* Ultrastruct Pathol, 2004. **28**(4): p. 229-38.
154. Meshorer, E., et al., *Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells.* Dev Cell, 2006. **10**(1): p. 105-16.
155. Azuara, V., et al., *Chromatin signatures of pluripotent cell lines.* Nat Cell Biol, 2006. **8**(5): p. 532-8.
156. Efroni, S., et al., *Global transcription in pluripotent embryonic stem cells.* Cell Stem Cell, 2008. **2**(5): p. 437-47.
157. Chen, T. and S.Y. Dent, *Chromatin modifiers and remodellers: regulators of cellular differentiation.* Nat Rev Genet, 2014. **15**(2): p. 93-106.
158. Ho, L. and G.R. Crabtree, *Chromatin remodelling during development.* Nature, 2010. **463**(7280): p. 474-84.
159. Lee, J.H., S.R. Hart, and D.G. Skalnik, *Histone deacetylase activity is required for embryonic stem cell differentiation.* Genesis, 2004. **38**(1): p. 32-8.
160. Voo, K.S., et al., *Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1.* Mol Cell Biol, 2000. **20**(6): p. 2108-21.
161. Lee, J.H. and D.G. Skalnik, *CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex.* J Biol Chem, 2005. **280**(50): p. 41725-31.

162. Lee, J.H., K.S. Voo, and D.G. Skalnik, *Identification and characterization of the DNA binding domain of CpG-binding protein.* J Biol Chem, 2001. **276**(48): p. 44669-76.
163. Thomson, J.P., et al., *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.* Nature, 2010. **464**(7291): p. 1082-6.
164. Sarraf, S.A. and I. Stancheva, *Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly.* Mol Cell, 2004. **15**(4): p. 595-605.
165. Bestor, T.H. and G.L. Verdine, *DNA methyltransferases.* Curr Opin Cell Biol, 1994. **6**(3): p. 380-9.
166. FitzGerald, K.T. and M.O. Diaz, *MLL2: A new mammalian member of the trx/MLL family of genes.* Genomics, 1999. **59**(2): p. 187-92.
167. Shi, X., et al., *Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36.* J Biol Chem, 2007. **282**(4): p. 2450-5.
168. Butler, J.S., J.H. Lee, and D.G. Skalnik, *CFP1 interacts with DNMT1 independently of association with the Setd1 Histone H3K4 methyltransferase complexes.* DNA Cell Biol, 2008. **27**(10): p. 533-43.
169. Fujino, T., et al., *PCCX1, a novel DNA-binding protein with PHD finger and CXXC domain, is regulated by proteolysis.* Biochem Biophys Res Commun, 2000. **271**(2): p. 305-10.
170. Perry, J. and N. Kleckner, *The ATRs, ATMs, and TORs are giant HEAT repeat proteins.* Cell, 2003. **112**(2): p. 151-5.
171. Groves, M.R., et al., *The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs.* Cell, 1999. **96**(1): p. 99-110.
172. Carlone, D.L. and D.G. Skalnik, *CpG binding protein is crucial for early embryonic development.* Mol Cell Biol, 2001. **21**(22): p. 7601-6.
173. Young, S.R., et al., *Antisense targeting of CXXC finger protein 1 inhibits genomic cytosine methylation and primitive hematopoiesis in zebrafish.* J Biol Chem, 2006. **281**(48): p. 37034-44.

174. Young, S.R. and D.G. Skalnik, *CXXC finger protein 1 is required for normal proliferation and differentiation of the PLB-985 myeloid cell line.* DNA Cell Biol, 2007. **26**(2): p. 80-90.
175. Kim, A., et al., *Beta common receptor inactivation attenuates myeloproliferative disease in Nf1 mutant mice.* Blood, 2007. **109**(4): p. 1687-91.
176. Chun, K.T., et al., *The epigenetic regulator CXXC finger protein 1 is essential for murine hematopoiesis.* PLoS One, 2014. **9**(12): p. e113745.
177. Carlone, D.L., et al., *Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein.* Mol Cell Biol, 2005. **25**(12): p. 4881-91.
178. Butler, J.S., et al., *DNA Methyltransferase protein synthesis is reduced in CXXC finger protein 1-deficient embryonic stem cells.* DNA Cell Biol, 2009. **28**(5): p. 223-31.
179. Tate, C.M., J.H. Lee, and D.G. Skalnik, *CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin.* FEBS J, 2010. **277**(1): p. 210-23.
180. Tate, C.M., J.H. Lee, and D.G. Skalnik, *CXXC finger protein 1 contains redundant functional domains that support embryonic stem cell cytosine methylation, histone methylation, and differentiation.* Mol Cell Biol, 2009. **29**(14): p. 3817-31.
181. Clouaire, T., et al., *Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells.* Genes Dev, 2012. **26**(15): p. 1714-28.
182. Michalopoulos, G.K., *Liver regeneration.* J Cell Physiol, 2007. **213**(2): p. 286-300.
183. Stokkan, K.A., et al., *Entrainment of the circadian clock in the liver by feeding.* Science, 2001. **291**(5503): p. 490-3.
184. Zorn, A.M., *Liver development*, in *StemBook*. 2008: Cambridge (MA).
185. Douarin, N.M., *An experimental analysis of liver development.* Med Biol, 1975. **53**(6): p. 427-55.
186. Taub, R., *Liver regeneration: from myth to mechanism.* Nat Rev Mol Cell Biol, 2004. **5**(10): p. 836-47.

187. Higgins, G.M.A., R. M, *Experimental pathology of the liver. I. Restoration of the liver of the white rat following partial surgical removal.* . Arch. Pathol., 1931. **12**: p. 186–202.
188. Sakamoto, T., et al., *Mitosis and apoptosis in the liver of interleukin-6-deficient mice after partial hepatectomy.* Hepatology, 1999. **29**(2): p. 403-11.
189. Michalopoulos, G.K., *Principles of liver regeneration and growth homeostasis.* Compr Physiol, 2013. **3**(1): p. 485-513.
190. Satyanarayana, A., et al., *Telomere shortening impairs organ regeneration by inhibiting cell cycle re-entry of a subpopulation of cells.* EMBO J, 2003. **22**(15): p. 4003-13.
191. Shingo, T., et al., *Pregnancy-stimulated neurogenesis in the adult female forebrain mediated by prolactin.* Science, 2003. **299**(5603): p. 117-20.
192. Kim, H., et al., *Serotonin regulates pancreatic beta cell mass during pregnancy.* Nat Med, 2010. **16**(7): p. 804-8.
193. Dai, G., et al., *Maternal hepatic growth response to pregnancy in the mouse.* Exp Biol Med (Maywood), 2011. **236**(11): p. 1322-32.
194. Feng, D., et al., *A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism.* Science, 2011. **331**(6022): p. 1315-9.
195. Mann, J., et al., *MeCP2 controls an epigenetic pathway that promotes myofibroblast transdifferentiation and fibrosis.* Gastroenterology, 2010. **138**(2): p. 705-14, 714 e1-4.
196. Mann, J., et al., *Regulation of myofibroblast transdifferentiation by DNA methylation and MeCP2: implications for wound healing and fibrogenesis.* Cell Death Differ, 2007. **14**(2): p. 275-85.
197. Zeybel, M., D.A. Mann, and J. Mann, *Epigenetic modifications as new targets for liver disease therapies.* J Hepatol, 2013. **59**(6): p. 1349-53.
198. Bae, W.K., et al., *The methyltransferases enhancer of zeste homolog (EZH) 1 and EZH2 control hepatocyte homeostasis and regeneration.* FASEB J, 2014.
199. Nguyen, T.N. and J.A. Goodrich, *Protein-protein interaction assays: eliminating false positive interactions.* Nat Methods, 2006. **3**(2): p. 135-9.
200. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.

201. Kruger, N.J., *The Bradford method for protein quantitation.* Methods Mol Biol, 1994. **32**: p. 9-15.
202. Postic, C. and M.A. Magnuson, *DNA excision in liver by an albumin-Cre transgene occurs progressively with age.* Genesis, 2000. **26**(2): p. 149-50.
203. Greene, A.K. and M. Puder, *Partial hepatectomy in the mouse: technique and perioperative management.* J Invest Surg, 2003. **16**(2): p. 99-102.
204. Williams, R.L., et al., *Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells.* Nature, 1988. **336**(6200): p. 684-7.
205. Keller, G., *Embryonic stem cell differentiation: emergence of a new era in biology and medicine.* Genes Dev, 2005. **19**(10): p. 1129-55.
206. O'Connor, M.D., et al., *Alkaline phosphatase-positive colony formation is a sensitive, specific, and quantitative indicator of undifferentiated human embryonic stem cells.* Stem Cells, 2008. **26**(5): p. 1109-16.
207. Hay, D.C., et al., *Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.* Stem Cells, 2004. **22**(2): p. 225-35.
208. Pesce, M. and H.R. Scholer, *Oct-4: gatekeeper in the beginnings of mammalian development.* Stem Cells, 2001. **19**(4): p. 271-8.
209. Zaehres, H., et al., *High-efficiency RNA interference in human embryonic stem cells.* Stem Cells, 2005. **23**(3): p. 299-305.
210. Lallous, N., et al., *The PHD finger of human UHRF1 reveals a new subgroup of unmethylated histone H3 tail readers.* PLoS One, 2011. **6**(11): p. e27599.
211. Kwan, A.H., et al., *Engineering a protein scaffold from a PHD finger.* Structure, 2003. **11**(7): p. 803-13.
212. Yan, B.X. and Y.Q. Sun, *Glycine residues provide flexibility for enzyme active sites.* J Biol Chem, 1997. **272**(6): p. 3190-4.
213. Baratta, J.L., et al., *Cellular organization of normal mouse liver: a histological, quantitative immunocytochemical, and fine structural analysis.* Histochem Cell Biol, 2009. **131**(6): p. 713-26.

214. Wysocka, J., et al., *WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development.* Cell, 2005. **121**(6): p. 859-72.
215. Elgin, S.C. and S.I. Grewal, *Heterochromatin: silence is golden.* Curr Biol, 2003. **13**(23): p. R895-8.
216. Aagaard, L., et al., *Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31.* EMBO J, 1999. **18**(7): p. 1923-38.
217. Pena, P.V., et al., *Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2.* Nature, 2006. **442**(7098): p. 100-3.
218. Tsai, W.W., et al., *TRIM24 links a non-canonical histone signature to breast cancer.* Nature, 2010. **468**(7326): p. 927-32.
219. Wang, Z., et al., *Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes.* Cell, 2009. **138**(5): p. 1019-31.
220. Hazzalin, C.A. and L.C. Mahadevan, *Dynamic acetylation of all lysine 4-methylated histone H3 in the mouse nucleus: analysis at c-fos and c-jun.* PLoS Biol, 2005. **3**(12): p. e393.
221. Edmunds, J.W., L.C. Mahadevan, and A.L. Clayton, *Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation.* EMBO J, 2008. **27**(2): p. 406-20.
222. Vermeulen, M., et al., *Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers.* Cell, 2010. **142**(6): p. 967-80.
223. Clouaire, T., S. Webb, and A. Bird, *Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells.* Genome Biol, 2014. **15**(9): p. 451.
224. Loh, Y.H., et al., *The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.* Nat Genet, 2006. **38**(4): p. 431-40.
225. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. **24**(4): p. 372-6.

226. Gidekel, S. and Y. Bergman, *A unique developmental pattern of Oct-3/4 DNA methylation is controlled by a cis-demodification element.* J Biol Chem, 2002. **277**(37): p. 34521-30.
227. Pereira, L., F. Yi, and B.J. Merrill, *Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal.* Mol Cell Biol, 2006. **26**(20): p. 7479-91.
228. Bratt-Leal, A.M., R.L. Carpenedo, and T.C. McDevitt, *Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation.* Biotechnol Prog, 2009. **25**(1): p. 43-51.
229. Calo, E. and J. Wysocka, *Modification of enhancer chromatin: what, how, and why?* Mol Cell, 2013. **49**(5): p. 825-37.
230. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
231. Voigt, P., W.W. Tee, and D. Reinberg, *A double take on bivalent promoters.* Genes Dev, 2013. **27**(12): p. 1318-38.
232. Risner, L.E., et al., *Functional specificity of CpG DNA-binding CXXC domains in mixed lineage leukemia.* J Biol Chem, 2013. **288**(41): p. 29901-10.
233. Hu, D., et al., *The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells.* Nat Struct Mol Biol, 2013. **20**(9): p. 1093-7.
234. Gualdi, R., et al., *Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control.* Genes Dev, 1996. **10**(13): p. 1670-82.
235. Martin, M.A. and M. Bhatia, *Analysis of the human fetal liver hematopoietic microenvironment.* Stem Cells Dev, 2005. **14**(5): p. 493-504.
236. Zaret, K.S., *Liver specification and early morphogenesis.* Mech Dev, 2000. **92**(1): p. 83-8.
237. Weisend, C.M., et al., *Cre activity in fetal albCre mouse hepatocytes: Utility for developmental studies.* Genesis, 2009. **47**(12): p. 789-92.
238. Parviz, F., et al., *Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis.* Nat Genet, 2003. **34**(3): p. 292-6.
239. Hayhurst, G.P., et al., *Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis.* Mol Cell Biol, 2001. **21**(4): p. 1393-403.

240. Carlone, D.L., et al., *Cloning and characterization of the gene encoding the mouse homologue of CpG binding protein.* Gene, 2002. **295**(1): p. 71-7.
241. Pena-Diaz, J., et al., *Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication.* Nucleic Acids Res, 2013. **41**(5): p. 2846-56.
242. Kobayashi, N., et al., *Hepatocyte transplantation in rats with decompensated cirrhosis.* Hepatology, 2000. **31**(4): p. 851-7.
243. Desai, B.M., et al., *Preexisting pancreatic acinar cells contribute to acinar cell, but not islet beta cell, regeneration.* J Clin Invest, 2007. **117**(4): p. 971-7.
244. Minami, K. and S. Seino, *Pancreatic acinar-to-beta cell transdifferentiation in vitro.* Front Biosci, 2008. **13**: p. 5824-37.
245. Mader, S.L., et al., *Refining timed pregnancies in two strains of genetically engineered mice.* Lab Anim (NY), 2009. **38**(9): p. 305-10.

Curriculum Vitae

Jyothi Mahadevan

Education

Doctor of Philosophy in Biochemistry & Molecular Biology 2009 – 2015

Indiana University, Indianapolis IN

Project: The role of CFP1 in murine embryonic stem cell function and liver regeneration

Mentor: Dr. David G. Skalnik

Current program GPA: 3.84/4

Minor: Life Sciences

Master of Science in Biochemistry Jun 2008

University of Mumbai, India

First Class

Bachelor of Science in Biochemistry & Microbiology Jun 2006

University of Mumbai, India

GPA: 3.95/4

Research Experience

Doctoral Research Aug 2010 – May 2015

Indiana University School of Medicine

Thesis Advisor: Dr. David G. Skalnik

- Studied the role of the plant homeodomain (PHD) of CFP1 in embryonic stem cell differentiation using *in vitro* and *in vivo* molecular and cellular biology techniques
- Generated and characterized the liver specific CFP1 knock out mouse line to study the importance of CFP1 in mouse liver regeneration after partial hepatectomy
- Studied genomic targeting of the Set1 histone methyltransferases Set1A and Set1B using ChIPqPCR and gained familiarity with high throughput ChIP-Sequencing

Graduate Research Aug 2007 – Jan 2008

CIPLA Pharmaceuticals Limited, Mumbai, India

Graduate Advisor: Dr. A.M. Varadkar

- Thesis title: "*Study of Stability of an Intramuscular Injectable (Amoxycillin Trihydrate) After 6 Months of Manufacture*"

Teaching & Mentoring Experience

Lecturer Apr 2008 - Apr 2009

Avalon Heights International School, Mumbai, India

- Taught Biology and Chemistry to High School students for the University of Cambridge International Examinations Curriculum.
- Worked as a Group Leader for Science section and conducted workshops for teacher training in the institution.

Mentor

Supervised and assisted in experimental design and data analysis

- Graduate students
  Department of Biology, IUPUI School of Science Aug 2012 – Dec 2014
- Summer Research Intern Apr 2012 – Aug 2012

Awards, Scholarships and Memberships

- *Best Graduate Student Poster Award*
  Biochemistry and Molecular Biology Research Day 2014, IUSM Nov 2014
- *Post Graduate Merit Scholarship*
  Awarded for being placed among the top 5 students at my undergraduate institution Apr 2006
- Shri. Sohrab Mistry Prize
  Awarded for having secured the highest marks in Chemistry in IInd Year BS Apr 2005
- Shri. Jayaram Mani endowment Prize
  Awarded for having secured the highest marks in Botany in Ist Year BS Apr 2004
- American Association for the Advancement of Science (AAAS) Since Dec 2010
- Secretary of Science Association, University of Mumbai 2004-2005

- Member of Student Council, University of Mumbai 2004-2005

Publications

- Mahadevan, J. and D.G. Skalnik, "Efficient differentiation of murine ES cells requires the binding of CFP1 to methylated H3K4 or DNA" (Manuscript in review)

Scientific Talks & Presentations

- *Annual Department of Biology Scientific Retreat*
  IUPUI School of Science, Indianapolis IN Oct 2012 & Oct 2013
- *Cancer Research Day*
  Indiana University School of Medicine, Indianapolis IN May 2013

Abstracts Presented

- Mahadevan, J. and D.G. Skalnik, *Plant Homeodomain (PHD) of CXXC finger protein 1(CFP1) is critical for efficient differentiation of murine embryonic stem cells.* Biochemistry and Molecular Biology Research Day, 2014, IUSM, Indianapolis IN
- Mahadevan, J. and D.G. Skalnik, *The role of CFP1 PHD domains in the genomic targeting of the Set1 histone methyltransferases*. Midwest Chromatin and Epigenetics Meeting 2014, Madison WI
- Mahadevan, J. and D.G. Skalnik, *Genomic targeting mechanisms of human Setd1a and Setd1b histone H3-Lys4 methyltransferases.* Biochemistry and Molecular Biology Research Day, 2012, IUSM, Indianapolis IN